



in conjunction with the Medical Image Computing and Computer-Assisted Interventions Conference

2018 International MICCAI BraTS Challenge

September 16, 2018

Granada, Spain

(pre-conference proceedings)

Scope

BraTS has always been focusing on the evaluation of state-of-the-art methods for the segmentation of brain tumors in multimodal magnetic resonance imaging (MRI) scans. **BraTS 2018** utilizes multi-institutional pre-operative MRI scans and **focuses on the segmentation of** intrinsically heterogeneous (in appearance, shape, and histology) **brain tumors**, namely gliomas. Furthermore, to pinpoint the clinical relevance of this segmentation task, BraTS'18 also focuses **on the prediction of patient overall survival**, via integrative analyses of radiomic features and machine learning algorithms.

Clinical Relevance

Gliomas are the most common primary brain malignancies, with different degrees of aggressiveness, variable prognosis and various heterogeneous histological sub-regions, i.e., peritumoral edema, necrotic core, enhancing and non-enhancing tumor core. This intrinsic heterogeneity of gliomas is also portrayed in their imaging phenotype (appearance and shape), as their sub-regions are described by varying intensity profiles disseminated across multimodal MRI scans, reflecting varying tumor biological properties. Due to this highly heterogeneous appearance and shape, segmentation of brain tumors in multimodal MRI scans is one of the most challenging tasks in medical image analysis.

There is a growing body of literature on computational algorithms addressing this important task. Unfortunately, open data sets for designing and testing these algorithms are not currently available, and private datasets differ so widely that it is hard to compare the different segmentation strategies that have been reported so far. Critical factors leading to these differences include, but not limited to, i) the imaging modalities employed, ii) the type of the tumor (GBM or LGG, primary or secondary tumors, solid or infiltratively growing), and iii) the state of disease (images may not only be acquired prior to treatment, but also post-operatively and therefore show radiotherapy effects and surgically-imposed cavities). Towards this end, BraTS is making available a large dataset with accompanying delineations of the relevant tumor sub-regions.

Tasks

Task 1: Segmentation of gliomas in pre-operative MRI scans: The participants are called to address this task by using the provided clinically-acquired training data to develop their method and produce segmentation labels of the different glioma sub-regions. **The sub-regions considered for evaluation are: 1) the "enhancing tumor" (ET), 2) the "tumor core" (TC), and 3) the "whole tumor" (WT)** [see figure below]. The ET is described by areas that show hyper-intensity in T1Gd when compared to T1, but also when compared to "healthy" white matter in T1Gd. The TC describes the bulk of the tumor, which is what is typically resected. The TC entails the ET, as well as the necrotic (fluid-filled) and the non-enhancing (solid) parts of the tumor. The appearance of the necrotic (NCR) and the non-enhancing (NET) tumor core is typically hypo-intense in T1-Gd when compared to T1. The WT describes the complete extent of the disease, as it entails the TC and the peritumoral edema (ED), which is typically depicted by hyper-intense signal in FLAIR. The labels in the provided data are: 1 for NCR & NET, 2 for ED, 4 for ET, and 0 for everything else.

Task 2: Prediction of patient overall survival (OS) from pre-operative scans: Once the participants produce their segmentation labels in the pre-operative scans, they will be called to use these labels in combination with the provided multimodal MRI data to extract imaging/radiomic features that they consider appropriate, and analyze them through machine learning algorithms, in an attempt to predict patient OS. The participants do not need to be limited to volumetric parameters, but can also consider intensity, morphologic, histogram-based, and textural features, as well as spatial information, and glioma diffusion properties extracted from glioma growth models. Participants are expected to provide predicted survival status only for subjects with Gross Total Resection status. Three groups of survival are considered, i.e. long-survivors (e.g., >15 months), short-survivors (e.g., <10 months), and mid-survivors (e.g. between 10 and 15 months).

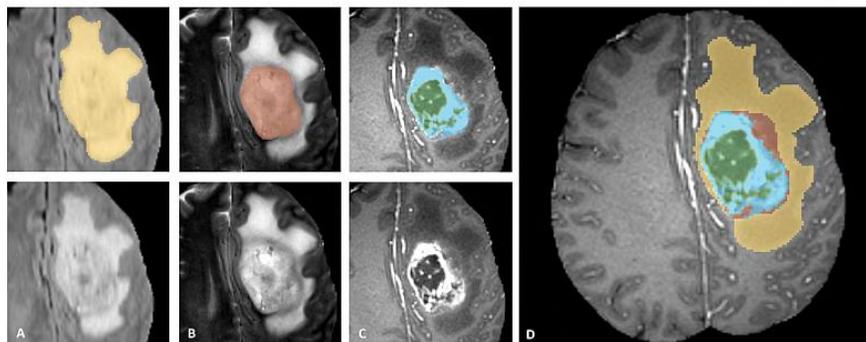


Fig.1: Glioma sub-regions. The image patches show from left to right: the whole tumor (yellow) visible in T2-FLAIR (Fig.A), the tumor core (red) visible in T2 (Fig.B), the enhancing tumor structures (light blue) visible in T1Gd, surrounding the cystic/necrotic components of the core (green) (Fig. C). The segmentations are combined to generate the final labels of the tumor sub-regions (Fig.D): edema (yellow), non-enhancing solid core (red), necrotic/cystic core (green), enhancing core (blue). (Figure taken from the [BraTS IEEE TMI paper](#).)

Data

The datasets used in this year's challenge have been updated, since BraTS'16, with more routine clinically-acquired 3T multimodal MRI scans and all the ground truth labels have been manually-revised by expert board-certified neuroradiologists.

Ample multi-institutional routine clinically-acquired pre-operative multimodal MRI scans of glioblastoma (GBM/HGG) and lower grade glioma (LGG), with pathologically confirmed diagnosis and available OS, will be provided as the training, validation and testing data for this year's BraTS challenge. These multimodal scans describe a) native (**T1**) and b) post-contrast T1-weighted (**T1Gd**), c) T2-weighted (**T2**), and d) T2 Fluid Attenuated Inversion Recovery (**FLAIR**) volumes, and were acquired with different clinical protocols and various scanners from multiple (n=19) institutions, mentioned as data contributors below. All the imaging datasets have been segmented manually, by one to four raters, following the same annotation protocol, and their annotations were approved by experienced neuro-radiologists. Annotations comprise the GD-enhancing tumor (ET — label 4), the peritumoral edema (ED — label 2), and the necrotic and non-enhancing tumor (NCR/NET — label 1), as described in the [BraTS reference paper](#), published in IEEE Transactions for Medical Imaging ([Fig.1](#)). The provided data are distributed after their pre-processing, i.e. co-registered to the same anatomical template, interpolated to the same resolution (1 mm³) and skull-stripped. The data provided since BraTS'17 differs significantly from the data provided during the previous BraTS challenges (i.e., 2016 and backwards). Specifically, since BraTS'17, expert neuroradiologists have radiologically assessed the complete original TCIA glioma collections and categorized each scan as pre- or post-operative. Subsequently, all the pre-operative TCIA scans were annotated by experts for the various sub-regions and included in this year's BraTS datasets. Participants are only allowed to use additional private data (from their own institutions) for data augmentation, if they also report results using only the BraTS'18 data and discuss any potential difference in the results. Validation data was released on July 1, allowing participants to obtain preliminary results in unseen data and also report it in their submitted papers, in addition to their cross-validated results on the training data. The ground truth of the validation data was not provided to the participants. Finally, all participants were presented with the same test data, for a limited controlled time-window (48h), before the participants are required to upload their final results in [CBICA's IPP](#) (ipp.cbica.upenn.edu).

Data Request

To request the training and the validation data of the BraTS 2018 challenge, please follow the steps below:

1. Create an account in CBICA's Image Processing Portal (ipp.cbica.upenn.edu) and wait for its approval. Note that a confirmation email will be sent so make sure that you also check your Spam folder.
2. Once your IPP account is approved, login to ipp.cbica.upenn.edu and then click on the application "BraTS'18: Data Request", under the "MICCAI BraTS 2018" group.
3. Fill in the requested details and press "Submit Job".
4. Once your request is recorded, you will receive an email pointing to the "results" of your submitted job. You need to login to IPP, access the "Results.zip" file, in which you will find the file "REGISTRATION_STATUS.txt" that will provide the links to download the BraTS 2018 data. The training data will include for each subject the 4 structural modalities, ground truth segmentation labels and accompanying survival information, age, and resection status, whereas the validation data will include on the 4 modalities.

Data Use

Please ensure that whenever you use and/or refer to the BraTS datasets in your manuscripts, you always cite the following papers:

[1] Menze BH, et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694

[2] Bakas S, et al. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117

In addition, if the place you submit your paper does not restrict you from citing "data citations" you might also cite the following:

[3] Bakas S, et al. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q

[4] Bakas S, et al. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF

Evaluation Framework

In this year's challenge, two reference standards are used for the two tasks of the challenge: 1) manual segmentation labels of tumor sub-regions, and 2) clinical data of overall survival.

For the segmentation task, and for consistency with the configuration of the previous BraTS challenges, we will use the "Dice score", and the "Hausdorff distance". Expanding upon this evaluation scheme, in BraTS'18 we will also use the metrics of "Sensitivity" and "Specificity", allowing to determine potential over- or under-segmentations of the tumor sub-regions by participating methods. Since the BraTS'12-'13 are subsets of the BraTS'18 test data, we will also calculate performance on the '12-'13 data to allow for a comparison against the performances reported in the BraTS TMI reference paper.

For the task of survival prediction, two evaluation schemes are considered. First, for ranking the participating teams, evaluation will be based on the classification of subjects as long-survivors (e.g., >15 months), short-survivors (e.g., <10 months), and mid-survivors (e.g. between 10 and 15 months). Predictions of the participating teams will be assessed based on accuracy (i.e. the number of correctly classified patients) with respect to this grouping. Note that participants are expected to provide predicted survival status only for subjects with resection status of GTR (i.e., Gross Total Resection). For post-challenge analyses, we will also compare both the mean and median square error of survival time predictions.

Organizing Committee

- [Spyridon \(Spyros\) Bakas, Ph.D.](#), — [Lead Organizer]
Center for Biomedical Image Computing and Analytics (CBICA), SBIA, UPenn, PA, USA
- [Bjoern Menze, Ph.D.](#),
Technical University of Munich (TUM), Germany
- [Christos Davatzikos, Ph.D.](#),
Center for Biomedical Image Computing and Analytics (CBICA), SBIA, UPenn, PA, USA
- [Keyvan Farahani, Ph.D.](#),
Cancer Imaging Program, National Cancer Institute (NCI), National Institutes of Health, USA



Data Contributors

- [John B. Freymann](#) & [Justin S. Kirby](#) - on behalf of [The Cancer Imaging Archive \(TCIA\)](#),
Cancer Imaging Program, NCI, National Institutes of Health (NIH), USA
- [Christos Davatzikos, Ph.D.](#), & [Spyridon \(Spyros\) Bakas, Ph.D.](#),
Center for Biomedical Image Computing and Analytics (CBICA), SBIA, UPenn, PA, USA
- [Hassan Fathallah-Shaykh, M.D., Ph.D.](#),
University of Alabama at Birmingham, AL, USA
- [Roland Wiest, M.D.](#),
University of Bern, Switzerland
- [Andras Jakab, M.D., Ph.D.](#),
University of Debrecen, Hungary
- [Rivka R. Colen, M.D.](#), & [Aikaterini \(Katerina\) Kotrotsou, Ph.D.](#),
MD Anderson Cancer Center, TX, USA
- [Daniel Marcus, Ph.D.](#), & [Mikhail Milchenko, Ph.D.](#), & [Arash Nazeri, M.D.](#),
Washington University School of Medicine in St. Louis, MO, USA
- [Marc-Andre Weber, M.D.](#),
Heidelberg University, Germany
- [Abhishek Mahajan, M.D.](#), & [Ujjwal Baid, M-Tech.](#),
Tata Memorial Centre, Mumbai, India





Table of Contents

Deep Learning and Radiomics for Glioblastoma Survival Prediction Mahbubul Alam, Lasitha Vidyaratne, Zeina Shboul, Linmim Pei, Thomas Batchelder and Khan M. Iftekharuddin	Page 11
Extending 2D deep learning architectures to 3D image segmentation problems Alberto Albiol, Antonio Albiol, and Francisco Albiol	Page 19
GBM Segmentation with 3D U-Net and Survival Prediction with Radiomics Ujjwal Baid, Abhishek Mahaja2, Sanjay Talbar, Swapnil Rane, Siddhesh Thakur, Aliasgar Moiyadi, Meenakshi Thakur, and Sudeep Gupta	Page 28
Multi-Planar Spatial-ConvNet for Segmentation and Survival Prediction in Brain Cancer Subhashis Banerjee, Sushmita Mitra, and B. Uma Shankar	Page 36
Deep Hourglass for Brain Tumor Segmentation Eze Benson, Michael P. Pound, Andrew P. French, Aaron S. Jackson1, Tony P. Pridmore	Page 46
Survival prediction using ensemble tumor segmentation and transfer learning Mariano Cabezas, Sergi Valverde, Sandra González-Villá, Albert Clérigues, Mostafa Salem, Kaisar Kushibar, Jose Bernal, Arnau Oliver, Joaquim Salvi, and Xavier Lladó	Page 54
Automatic Brain Tumor Segmentation Using a U-net Neural Network Eric Caver, Chang, Liu, Weiwei Zong, Zhenzhen Dai, Ning Wen	Page 63
Context Aware 3-D Residual Networks for Brain Tumor Segmentation Siddhartha Chandra, Maria Vakalopoulou, Lucas Fidon, Enzo Battistella, Theo Estienne, Roger Sun, Charlotte Robert, Eric Deutch, and Nikos Paragios	Page 74
Automatic segmentation of brain tumor from 3D MR images using a 2D convolutional neural network Yi-Ju Chang, Zheng-Shen Lin, Tsai-Ling Yang, Teng-Yi Huang	Page 83
S3D-UNet: Separable 3D U-Net for Brain Tumor Segmentation Wei Chen, Boqiang Liu, Suting Peng, Jiawei Sun, and Xu Qiao	Page 91
Segmentation of Brain Tumors using DeepLabv3+ Ahana Roy Choudhury, Rami Vanguri, Sachin R. Jambawalikar, and Piyush Kumar	Page 100
Automatic Brain Tumor Segmentation with Domain Transfer Lutao Dai, Tengfei Li, Hai Shu, Liming Zhong, Haipeng Shen and Hongtu	Page 111



Table of Contents

Three pathways U-Net for brain tumor segmentation Longwei Fang and Huiguang He	Page 119
Brain Tumor Segmentation using an Ensemble of 3D UNets Xue Feng, Nicholas Tustison and Craig Meyer	Page 127
Brain Tumor Detection and Segmentation Using Deep Learning U-Net on Multi Modal MRI Naomi Fridman	Page 135
Glioma Segmentation and a Simple Accurate Model for Overall Survival Prediction Evan Gates, J. Gregory Pauloski, Dawid Schellingerhout, and David Fuentes	Page 144
Semi-automatic Brain Tumor Segmentation by Drawing Long Axes on Multi-plane Reformat David Gering, Aaron Avery, Kay Sun, Lisa Kohli, Haley Knapp, Mark Gehring, Ajeet Vivekanandan, Jeff Hoffman, Brett Young-Moxon, Linda Peitzman ¹ , Roger Chylla, and Thomas Mackie	Page 153
Sirona: A Bio-Physics Based Deep Neural Network for Medical Image Segmentation Amir Gholami, Shashank Subramanian, Peter Jin, Naveen Himthani, Varun Shenoy, Xiangyu Yue, Sicheng Zhao, Pravin Ravishanker, Kurt Keutzer, and George Biros	Page 161
Neuromorphic Neural Network for Multimodal Brain Tumor Segmentation and Survival analysis Woo-Sup Han and Il Song Han	Page 171
BrainNet: 3D Local Re_ement Network for Brain Tumor Segmentation Xiaojun Hu, Weilin Huang, Deren Kong, Sheng Guo, and Matthew R. Scott	Page 179
Multi-level Activation for Segmentation of Hierarchically-nested Classes on 3D-Unet Xiaobin Hu, Marie Piraud	Page 188
Brain Tumor Segmentation on Multimodal MRI Using Multi-Level Upsampling in Decoder Yan Hu, Xiang Liu, Xin Wen, Chen Niu and Yong Xia	Page 196
Multimodal Brain Tumor Segmentation using Cascaded V-Net Rui Hua, Yu Zhang, Quan Huo, Yaozong Gao, Yu Sun, Feng Shi	Page 205
Pre and Post Processing Techniques for Brain Tumor Segmentation Vivek HV	Page 213



Table of Contents

No New-Net Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H. Maier-Hein	Page 222
Batch Normalized PixelNet for Brain Tumor Segmentation and Survival Prediction Mobarakol Islam, V Jeya Maria Jose, and Hongliang Ren	Page 232
Brain Tumor Segmentation and Tractographic Feature Extraction from Structural MR Images for Overall Survival Prediction Po-Yu Kao, Thuyen Ngo, Angela Zhang, Jefferson Chen, and B.S. Manjunath	Page 240
Brain Tumor Segmentation in Multimodal 3D-MRI of BraTS'2018 Datasets using Deep Convolutional Neural Networks Adel Kermi, Issam Mahmoudi and Mohamed Tarek Khadir	Page 252
3-D Densely Connected Fully Convolutional Neural Network for Brain Tumor Segmentation and Survival Analysis Avinash Kori, Mehul Soni, Pranjal B, Mahendra Khened, Varghese Alex, and Ganapathy Krishnamurthi	Page 264
Glioma Segmentation with Cascaded Unet: Preliminary results Dmitry Lachinov, Evgeny Vasiliev, and Vadim Turlapov	Page 272
Cascade of Random Forest Classifiers for Brain Tumor Segmentation Szidónia Lefkovits, László Szilágyi, and László Lefkovits	Page 280
Fused U-Net for Brain Tumor Segmentation based on Multimodal MR Images Xiaochuan Li	Page 290
Coarse-to-Fine Deep Convolutional Neural Networks for Multi-Modality Brain Tumor Semantic Segmentation Mingyuan Liu	Page 298
Automatic Brain Tumor Segmentation by Cascaded Lightweight CNN Architecture Jun Ma	Page 306
Automatic Brain Tumor Segmentation Using a Two-Stage Multi-Modal FCNN Michal Marcinkiewicz, Jakub Nalepa, Pablo Ribalta Lorenzo, Wojciech Dudzik, and Grzegorz Mrukwa	Page 314



Table of Contents

Ensembles of densely-connected CNNs with label-uncertainty for Brain tumor segmentation Richard McKinley, Raphael Meier, Roland Wiest	Page 322
3D U-net for Brain Tumour Segmentation Raghav Mehta and Tal Arbel	Page 331
Ensemble of Fully Convolutional Neural Networks for Brain Tumour Semantic Segmentation Miguel Monteiro, and Arlindo L. Oliveira	Page 341
3D MRI brain tumor segmentation using autoencoder regularization Andriy Myronenko	Page 349
ESPNNet for Glioma Segmentation Nicholas Nuechterlein and Sachin Mehta	Page 357
Ensemble Learning Models for Accurate Segmentation of Brain Tumor and Prediction of Patient Treatment Outcome: BraTS'2018 Challenge Alexander F. I. Osman	Page 365
Automatic Brain Tumor Segmentation using U-Net based 3D Fully Convolutional Network Anmol Popli, Manu Agarwal, and G.N. Pillai	Page 374
Brain Tumor Segmentation on multi-modal MR images with Deep Architectures Santi Puch, Irina Sánchez ¹ , Aura Hernández, Gemma Piella, and Vesna Prčkovska	Page 383
Glioma Segmentation In a Few Seconds Using Fully Convolutional Network and Transfer Learning Elodie Puybareau, Joseph Chazalon and Guillaume Tochon	Page 394
Ensembles of Multiple Scales, Losses and Models for Brain Tumor Segmentation and Overall Survival Time Prediction Task Xuhua Ren, Lichi Zhang, Dinggang Shen and Qian Wang	Page 402
Adversarial Framework for Learning Multiple Clinical Tasks Mina Rezaei and Christoph Meinel	Page 411
Brain Tumour Segmentation Method based on Sparse Feature Vectors J.P. Serrano-Rubio, Richard Everson, and Hugo Hutt	Page 420



Table of Contents

Brain Tumor Segmentation using 2D U-net Hyung Eun Shin and Moo Sung Park	Page 428
Leveraging a DenseNet Encoder Pre-Trained on ImageNet for Brain Tumor Segmentation Jean Stawiaski	Page 438
Multi-view 3D CNN with Dense CRF for Brain Tumor Segmentation and Survival Prediction Li Sun and Songtao Zhang	Page 448
End-to-End Deep Learning versus Classical Regression for Brain Tumor Patient Survival Prediction Yannick Suter, Alain Jungo, Michael Rebsamen, and Mauricio Reyes	Page 457
End-to-End Cascade Network for 3D Brain Tumor Segmentation in MICCAI 2018 BraTS Challenge Kuan-Lun, Tseng, Winston Hsu	Page 466
Brain Tumor Segmentation using Bit-plane and U-NET Tran Anh Tuan, Tran Anh Tuan and Pham The Bao	Page 474
Brain tumor segmentation with capsule networks versus fully convolutional neural networks Chiatse J. Wang, Yuhsiang M. Tsai , Chengen Lee, Yuehchou Lee , Anthony Costa , Cheyu Hsu , Eric Oermann and Weichung Wang	Page 482
Automatic Brain Tumor Segmentation using Convolutional Neural Networks with Test-Time Augmentation Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren	Page 492
Segmentation of Brain Tumors in 3D-MRI Data and Patient Survival Prediction: Methods for the BraTS 2018 Challenge Leon Weninger, Oliver Rippel, Simon Koppers, and Dorit Merhof	Page 500
Multimodal Brain Tumor Segmentation Using U-Net Shaocheng Wu, Hongyang Li and Yuanfang Guan	Page 508
A two-step cascaded strategy for automatic brain tumor segmentation in MICCAI 2018 BraTS Challenge Peiyuan Xu, Yifan Hu, Kai Ma, Yefeng Zheng	Page 516



Table of Contents

Brain Tumor Segmentation and Survival Prediction Based On Extended U-Net Model and XGBoost	Page 525
Xiaowen Xu, Xiangmao Kong, Guoxia Sun, Fengming Lin, Xiaomeng Cui, Shuang Sun, and Qiang Wu and Ju Liu	
Cascaded Mask V-Net	Page 534
Yanwu Xu, Mingming Gong, Huan Fu, Dacheng Tao, Kun Zhang, and Kayhan Batmanghelich	
Automatic Brain Tumor Segmentation with Contour Aware Residual Network and Adversarial Training	Page 542
Hao-Yu Yang and Junlin Yang	
Automatic Segmentation of Brain Tumor using 3D SE-Inception Networks with Residual Connections	Page 550
Hongdou Yao, Xiaobing Zhou and Xuejie Zhang	
3D Dense U-Nets For Brain Tumor Segmentation	Page 562
Xiaoyue Zhang, Weijian Jian, and Kun Cheng	
Learning Contextual and Attentive Information for Brain Tumor Segmentation	Page 571
Chenhong Zhou, Shengcong Chen, Changxing Ding, and Dacheng Tao	

Deep Learning and Radiomics for Glioblastoma Survival Prediction

Mahbubul Alam*, Lasitha Vidyaratne*, Zeina Shboul*, Linmim Pei, Thomas Batchelder and Khan M. Iftekharuddin

Vision Lab, Electrical & Computer Engineering, Old Dominion University
(malam001, lvidy001, zshbo001, lxpei001, tbatc002, kiftekha)odu.edu

Abstract. Glioblastoma is a high-grade invasive astrocytoma tumor. The highly invasive nature makes timely detection and characterization of the tumor critical for the survivability prediction of patients. This work proposes MRI- and clinical information-based automated pipeline that implements various state-of-the-art image processing, machine learning, and deep learning techniques to obtain robust tumor segmentation and patient survival estimation. We use 163 cases from the training dataset, and 66 cases from the validation dataset provided by the BraTS 2018 challenge for the evaluation of our model. We achieve a cross-validated accuracy of 76.07% and a mean squared error of 438.54 for the training data. Additionally, this model obtains an accuracy of 57.1% with a mean squared error of 382.96 for the validation dataset.

1 INTRODUCTION

High-grade glioblastoma (HGG) or glioblastoma represents tumors arising from the gluey or supportive tissue of the brain. HGG is considered as the most aggressive type of brain tumor. According to the American Brain Tumor Association (ABTA) [1] HGGs represent 74.6% of all malignant tumors and 24.7% of all primary brain tumors. World Health Organization (WHO) categorize HGGs as stage IV brain cancer [2]. Typically, the survival duration of patients with HGG tumor is less than two years [3, 4]. Therefore, accurate and timely detection of HGG tumor is essential for devising an appropriate treatment plan that may improve patient survival duration.

* The authors have equal contribution

Recent works have focused on developing automated survival prediction techniques for patients with HGG tumor. Several recent studies [5-10] have proposed automated brain tumor segmentation and patient survival prediction techniques using machine learning and deep learning. Our proposed deep learning and machine learning based survival prediction technique have shown the best performance in the BraTS 2017 survival prediction challenge with accuracy of 67% and 57% on the validation and test dataset, respectively [9]. In an effort to improve the performance of the survival prediction task, this work proposes sophisticated computational modelling based survival prediction method. Specifically, we approach the survival prediction task as a two-step process, brain tumor segmentation followed by survival prediction, considering the segmentation output as an input to the second step. The brain tumor segmentation task is performed by utilizing two state-of-the-art convolutional neural network (CNN) models, U-Net and fully convolutional neural network (FCN). The outcome of these two models are fused together to achieve the final segmentation output. This segmentation output along with the original MRI volumes are considered as input to the survival prediction step. Several radiomics features such as texture, topological, histogram etc. are extracted from the raw MRI sequences and the segmented tumor volume. Furthermore, a state-of-the-art 3D CNN architecture is utilized to extract additional features useful for the survival prediction task. These features are then processed using a gradient boosting based regression technique known as extended gradient boosting (XGBoost) to obtain the survival estimation of patients.

2 DATASET

In this study, we use MR images of 163 high grade GBM patients from BraTS18 training dataset [11-14]. The dataset provides the ground truth segmentation of tumor tissues which comprises of enhancing tumor (ET), edema (ED), and the necrosis and non-enhancing tumor (NCR/NET). In addition, the training dataset provides age and overall survival (in days) data. The available scans of the MRI are native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes. The dataset is co-registered, re-sampled to 1 mm^3 and skull-stripped.

3 METHODOLOGY

3.1 Brain tumor segmentation

This work utilizes two state-of-the-art CNN architectures, U-Net and fully convolutional neural network (FCN) to perform the brain tumor segmentation task. The following two sub-sections provides a very brief outline of these two models.

3.1.1 Brain tumor segmentaion using U-Net

This work utilizes a CNN based U-Net model [15, 16] to perform the brain tumor segmentaion task. Unlike patch based CNN segmentaion pipeline where the model only sees a localized region of the brain, the U-Net based segementation model captures the global information from different region of the brain tissues which is essential to achieve robust segmentaion performance. Moreover, U-Net based model allows to achieve an end-to-end segmentation framework rather than pixel wise classification technique. The U-Net architecture utilized in this work is implmented following [15]. More specifically, the architecture consists of a down-sampling (encoding) and an up-sampling (decoding) stage. Rather than using regular cross entropy based loss function, we utilize a soft dice metric based loss function to train the U-Net model [17]. The soft dice is shown to have a differentiable form of the original dice similarity coefficient (DSC) which is the most widely used metric to evaluate tumor segmentaion performance. The U-Net model is trained using mini-batch gradient descent (GD) technique which minimizes the soft dice cost function.

3.1.2 Brain tumor segementation using FCN

Fully convolutional networks (FCNs) has been successfully used for image processing on computer vision [18, 19]. It builds fully convolutional networks that take input of arbitrary size and produce a correspondingly sized output with efficient inference and learning. In comparison to generic patch-wise deep learning methods, FCNs are able to capture global information of the whole image. In our method, we adapt VGG-11 [20] as a pre-trained model. The overall FCN architecture contains an encode and a decode stage. The encoding stage contains convolution and max-pooling steps whereas the decoding stage contains deconvlolution step to obtain the same output size as the input. Unlike the U-Net model, the result of each convolutional layer is added as feature to the corresponding deconvolutional layer. Finally, we take the layer number which has the maximum probability as the segmentation result.

The final segmentation output is achieved by fusing the outcomes obtained from the above mentioned CNN based tumor segmentation pipelines as shown in Figure 1.

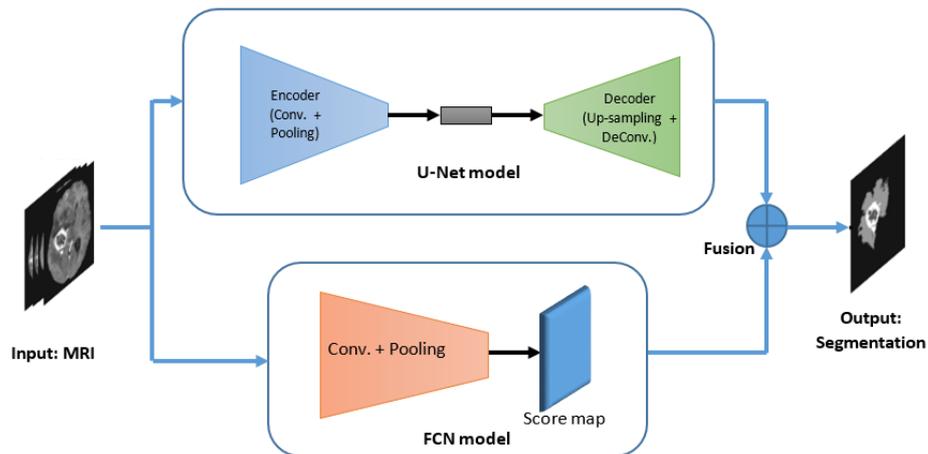


Figure 1. Brain tumor segmentation pipeline using U-Net and FCN

3.2 Survival Prediction

The proposed survival prediction pipeline essentially involves three stages, 1) Feature extraction stage, 2) feature selection stage, and 3) Feature classification/regression stage

3.2.1 Feature Extraction

Overall, approximately 31 thousand features are extracted from the tumor and the sub-regions (edema, enhancing tumor, and tumor core) as follows.

We extract forty one representative features [21] from three raw MRI sequences, and from eight texture representations of the tumor volume that includes several Texton filters [22], and fractal characterizations using algorithms such as PTPSA [23], mBm [24], and holder exponent [25]. The features obtained from these representations include histograms, co-occurrence matrix, grey-tone difference matrix, and several other statistical measures.

Furthermore, several histogram-based features are extracted from the tumor tissue regions. We also extract many representative volumetric features tumor tissues with respect to the brain and whole tumor regions.

In addition, we compute the Euler characteristics of the whole tumor, edema, enhancing and necrosis, for each slice as feature vectors.

To complement the above handcrafted features, we plan to add a set of learned features extracted from a deep learning model to boost our survival prediction performance. A few recent studies have attempted the use of transfer learning with deep CNN architectures for survival prediction task [5, 8]. In this study, we experiment with a novel 3D CNN architecture that is able to process entire 3D MRI volumes for the survival analysis. The architecture comprises of 10 layers, and takes full MRI volumes as well as the segmentation outcomes as inputs. The network is trained for survival duration estimate using the training dataset. The features are extracted from the second fully connected layer, and combined with the handcrafted features for overall survival estimation task.

3.2.2 Feature Selection and Classification

We perform recursive feature selection (RFS1) on the Euler features alone, another recursive feature selection (RFS2) on the other features (texture, volumetric, histogram-graph based), and a final feature selection on the features extracted from the 3D CNN.

3.2.3 Feature classification and Regression

Extreme Gradient Boosting (XGBoost) [26] is a tree boosting supervised machine learning technique that is highly effective. XGBoost is widely used in classification and regression tasks.

In our study, XGBoost is utilized for classification and regression overall survival prediction on the selected features. The trained models are tuned to their optimized hyper-parameters when a tuned grid (search grid) is created by the different combination of the hyper-parameters. The complete pipeline for classification and regression overall survival is illustrated in Figure 2.

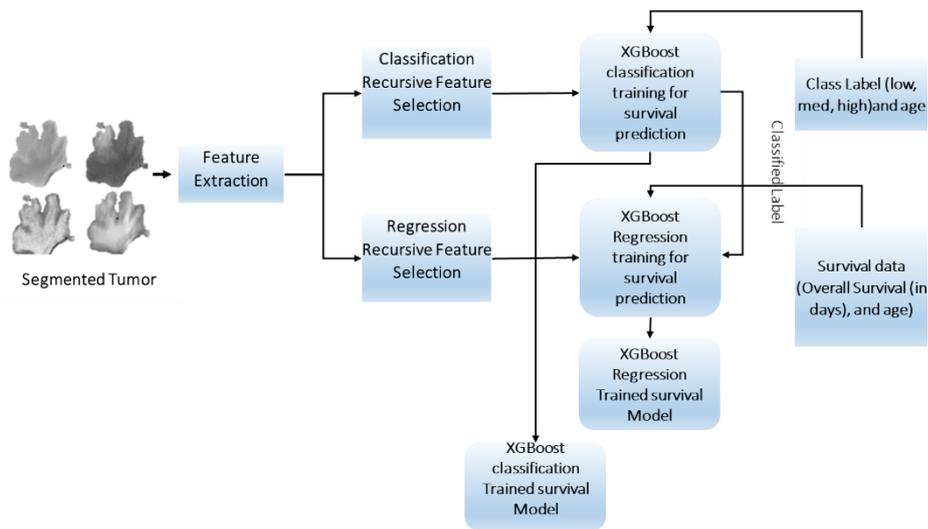


Figure 2. Survival prediction pipeline

4 EXPERIMENTAL RESULTS

We perform leave-one out cross-validation analysis on the BraTS 2018 training dataset using the proposed survival prediction pipeline. We evaluate the performance of our proposed method using the root mean square error (RMSE), and the classification accuracy for a three-class setting defined as follows: 1) long – more than 15 months, 2) Medium – between 10 to 15 months, 3) short – less than 10 months. The best model is picked from the analysis of training data to be used to process the validation dataset of the BraTS 2018 competition. The ground truth is considered for the segmented tumor in the survival analysis of the training dataset, while the segmented tumor obtained in stage one of the proposed pipeline is used as input in the evaluation of the validation dataset. The leave-one out cross-validation RMSE of the training dataset is 438.54, and the three-class cross-validated survival classification accuracy is 76.07%. With this model, we achieve a RMSE of 382.96 and a classification accuracy of 57.1% for the validation dataset.

5 CONCLUSIONS

This work proposes a robust automated glioblastoma survival prediction using state-of-the-art computational modeling techniques. The survival prediction

task is performed in two steps: tumor segmentation and survival prediction. The tumor segmentation task is performed by utilizing two state-of-the-art CNN architectures: U-Net and FCN. The survival prediction task is carried out by first extracting different types of representative features from the raw MRI data and the segmentation tumor output obtained from the previous step. In addition to the hand-crafted features, we also propose a novel 3D CNN based architecture for learning features from the raw MRI data and the segmentation output. The hand-crafted and learned features are then processed using popular XGBoost based regression technique to obtain the final survival prediction output. The performance of the proposed pipeline is evaluated using BraTS 2018 challenge training and validation datasets. Our results show a leave-on out cross-validated RMSE of 438.54, and a classification accuracy of 76.07% for the training dataset, and RMSE of 382.96 and classification accuracy of 57.1% for the validation dataset.

6 ACKNOWLEDGEMENTS

This work was funded by NIBIB/NIH grant# R01 EB020683.

References

- [1] A. B. T. Association, "Brain tumor statistics," *Available at: Accessed May*, vol. 2, 2016.
- [2] D. N. Louis *et al.*, "The 2007 WHO classification of tumours of the central nervous system," *Acta neuropathologica*, vol. 114, no. 2, pp. 97-109, 2007.
- [3] E. C. Holland, "Progenitor cells and glioma formation," *Current opinion in neurology*, vol. 14, no. 6, pp. 683-688, 2001.
- [4] H. Ohgaki and P. Kleihues, "Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas," *Journal of Neuropathology & Experimental Neurology*, vol. 64, no. 6, pp. 479-489, 2005.
- [5] J. Lao *et al.*, "A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme," *Scientific reports*, vol. 7, no. 1, p. 10353, 2017.
- [6] D. Nie, H. Zhang, E. Adeli, L. Liu, and D. Shen, "3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 212-220: Springer.
- [7] C. Huang, A. Zhang, and G. Xiao, "Deep Integrative Analysis for Survival Prediction," 2017.
- [8] L. Chato and S. Latifi, "Machine Learning and Deep Learning Techniques to Predict Overall Survival of Brain Tumor Patients using MRI Images," in *Bioinformatics and Bioengineering (BIBE), 2017 IEEE 17th International Conference on*, 2017, pp. 9-14: IEEE.
- [9] Z. A. Shboul, L. Vidyaratne, M. Alam, and K. M. Iftekharuddin, "Glioblastoma and Survival Prediction," in *International MICCAI Brainlesion Workshop*, 2017, pp. 358-368: Springer.

- [10] L. Vidyaratne, M. Alam, Z. Shboul, and K. Iftekharuddin, "Deep learning and texture-based semantic label fusion for brain tumor segmentation," in *Medical Imaging 2018: Computer-Aided Diagnosis*, 2018, vol. 10575, p. 105750D: International Society for Optics and Photonics.
- [11] S. Bakas *et al.*, "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific data*, vol. 4, p. 170117, 2017.
- [12] Bakas S *et al.*, "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection," ed, 2017.
- [13] S. Bakas *et al.*, "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection," *The Cancer Imaging Archive*, 2017.
- [14] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993-2024, 2015.
- [15] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks," in *Annual Conference on Medical Image Understanding and Analysis*, 2017, pp. 506-517: Springer.
- [16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234-241: Springer.
- [17] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision (3DV), 2016 Fourth International Conference on*, 2016, pp. 565-571: IEEE.
- [18] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431-3440.
- [19] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Fan, and Y. Zhang, "Brain tumor segmentation using a fully convolutional neural network with conditional random fields," in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 2016, pp. 75-87: Springer.
- [20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [21] M. Vallières, C. R. Freeman, S. R. Skamene, and I. El Naqa, "A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities," *Physics in medicine and biology*, vol. 60, no. 14, p. 5471, 2015.
- [22] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International journal of computer vision*, vol. 43, no. 1, pp. 29-44, 2001.
- [23] K. M. Iftekharuddin, W. Jia, and R. Marsh, "Fractal analysis of tumor in brain MR images," *Machine Vision and Applications*, vol. 13, no. 5-6, pp. 352-362, 2003.
- [24] A. Islam, K. M. Iftekharuddin, R. J. Ogg, F. H. Laningham, and B. Sivakumar, "Multifractal modeling, segmentation, prediction, and statistical validation of posterior fossa tumors," in *Medical Imaging 2008: Computer-Aided Diagnosis*, 2008, vol. 6915, p. 69153C: International Society for Optics and Photonics.
- [25] A. Ayache and J. L. V  hel, "Generalized multifractional Brownian motion: definition and preliminary results," ed: Springer, 1999.
- [26] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785-794: ACM.

Extending 2D deep learning architectures to 3D image segmentation problems

Alberto Albiol¹[0000-0002-1970-3289], Antonio Albiol¹[0000-0002-0679-912X], and Francisco Albiol²[0000-0003-2867-3525]

¹ iTeam, Universitat Politècnica de València, Spain
{alalbiol,aalbiol}@iteam.upv.es

² Instituto de Física Corpuscular (IFIC), Universitat de València, Consejo Superior de Investigaciones Científicas, Spain
kiko.albiol@ific.uv.es

Abstract. Several deep learning architectures are combined for brain tumour segmentation. All the architectures are inspired on recent 2D models where 2D convolution have been replaced by 3D convolutions. The key differences between the architectures are the size of the receptive field and the number of feature maps on the final layers. The obtained results are comparable to the top methods of previous Brats Challenges when median is use to average the results. Further investigation is still needed to analyze the outlier patients.

Keywords: Brain segmentation · Brats · 3D inception · 3D VGG · 3D Densely connected · 3D Xception.

1 Introduction

Brain tumour segmentation is an important problem which has received a considerable attention by the research community and particularly since advent of deep learning.

Gliial cells are the cause of gliomas that are the most common brain tumors. Gliomas are usually classified into low-grade gliomas (LGG) and high grade gliomas (HGG) which are malignant and more aggressive.

Brain tumors as usually imaged using several Magnetic Resonance (MR) sequences, such as T1-weighted, contrast enhanced T1-weighted (T1c), T2-weighted and Fluid Attenuation Inversion Recovery (FLAIR) images. From a pure pattern recognition point of view these modalities provide complimentary information and can be used as different feature input maps. Similar to different color planes for RGB natural images.

The Multimodal Brain Tumor Segmentation Challenge 2018 provided a set of MR sequences for training and evaluation of brain tumor segmentation algorithms. Ground truth for all the scans have been manually provided by expert board-certified neuroradiologists, so that every voxel is categorized into these classes [10]:

Layer name	kernel size	num filters
conv_1_1	$3 \times 3 \times 3$	30
conv_1_2	$3 \times 3 \times 3$	30
conv_2_1	$3 \times 3 \times 3$	60
conv_2_2	$3 \times 3 \times 3$	60
conv_3_1	$3 \times 3 \times 3$	120
conv_3_2	$3 \times 3 \times 3$	120
conv_4_1	$3 \times 3 \times 3$	240
conv_4_2	$3 \times 3 \times 3$	240
fc_1	$1 \times 1 \times 1$	400
fc_2	$1 \times 1 \times 1$	200
logits	$1 \times 1 \times 1$	4

Table 1. Description of the VGG-like architecture.

- Label 0: background
- Label 1: necrotic and non-enhancing tumor
- Label 2: edema
- Label 4: enhancing tumor

We did not pay much attention on the medical details of this problem. Our main contribution was to extend some of the recent approaches used for image classification: VGG, inception, Xception, densely connected models to be used with 3D data in a real segmentation problem.

2 Methods

Our approach uses an ensemble of deep neural networks with different architectures. The idea is that the ensemble provides a more robust solution with less variance. Also some architectures may compensate for other architectures weaknesses and thus improving the final results.

This section describes the different architectures used in our approach. All the architectures have in common that every pixel is independently labeled using a deep neural network architecture. We are aware that results might be better if some post processing that considered the spatial constraints is used similar to the CRF proposed in [9].

The key differences between the architectures are the number of parameters, the number of feature planes and the size of the receptive field for each voxel. These hyper-parameters were chosen as a trade-off usually limited by the memory of the GPU. More specifically, we mixed four different architectures in our final ensemble: VGG-like, inception-2, inception-3 and densely connected. These models are described in detail in the following subsections.

2.1 VGG-like model

This model is inspired on the well known VGG model proposed by [12]. The differences between our approach and the original VGG are:

- 2-D convolutions are replaced by 3-D convolutions
- Maxpool layers are not used
- The network is replicated in a convolutional way so that every pixel is labeled independently.

Table 1 describes in detail the layers used in this model. Note that all convolutional layers are preceded by batch normalisation and followed by a ReLU activation function, except the last layer which is followed by a softmax activation function.

2.2 Dense-like model

This architecture is inspired by the recent work [8]. The difference is that we have replaced 2D convolutions by 3D convolutions. The advantage of the densely connected network (compared to the VGG like) is that features are reused on subsequent layers and each layer adds a few new features only. This allows to increase the number of layers and therefore the size of the receptive field. This architecture also allows to combine features of relatively small receptive field (first layers) with features of large receptive field (last layers), which can be useful particularly in segmentation problems.

Table 2 summarizes the architecture of our densely connected network. Note that each layer concatenates all the output features of the previous layers, for this reason the number of input features grows steadily until layer conv_20. Then two fully connected layers similar to the VGG architecture are used.

2.3 inception-like model

This architecture is inspired on some of the ideas proposed in [14] and [13]. The key idea proposed by the inception model is to replace convolutional layer by several parallel structures with different kernel shapes. This reduces the number of parameters (regularization) and forces diversity on the output features of each layer.

We took these ideas and adapted them to the problem of brain segmentation. The main problem that we faced is the amount of memory available in the GPU. In the inception architecture the total number of layers is increased a lot, for instance some inception layers use 1-D convolutions for each spatial dimension, in the case of 2D convolution this option doubles the number of layers and the required memory used to store intermediate results and gradients. In our case this problem is even worse because an extension of the 1-D convolutions to volumes implies to use three times more memory.

For this reason we simplified the GoogLeNet model and added a few layers (named as *inception*) before the fully connected layers as detailed in Table 3.

Figure 1 shows the internal structure of the inception layers. As it can be seen, four different branches are used. The first layer extracts new features and reduces the dimensionality. The second and third branches introduce spatial convolution; and the fourth branch is an average layer without pooling. This structure is similar to the structure of fig 5. in [13].

Layer name	kernel size	num inputs	num output filters
conv_1	$3 \times 3 \times 3$	4	8
conv_2	$3 \times 3 \times 3$	12	8
conv_3	$3 \times 3 \times 3$	20	8
conv_4	$3 \times 3 \times 3$	28	8
conv_5	$3 \times 3 \times 3$	36	8
conv_6	$3 \times 3 \times 3$	44	8
conv_7	$3 \times 3 \times 3$	52	8
conv_8	$3 \times 3 \times 3$	60	8
conv_9	$3 \times 3 \times 3$	68	8
conv_10	$3 \times 3 \times 3$	76	8
conv_11	$3 \times 3 \times 3$	84	8
conv_12	$3 \times 3 \times 3$	92	8
conv_13	$3 \times 3 \times 3$	100	8
conv_14	$3 \times 3 \times 3$	108	8
conv_15	$3 \times 3 \times 3$	106	8
conv_16	$3 \times 3 \times 3$	114	8
conv_17	$3 \times 3 \times 3$	122	8
conv_18	$3 \times 3 \times 3$	130	8
conv_19	$3 \times 3 \times 3$	138	8
conv_20	$3 \times 3 \times 3$	146	8
fc_1	$1 \times 1 \times 1$	154	400
fc_2	$1 \times 1 \times 1$	400	200
logits	$1 \times 1 \times 1$	200	4

Table 2. Description of the Dense-like architecture.

Layer name	kernel size	num filters
conv_1_1	$3 \times 3 \times 3$	30
conv_1_2	$3 \times 3 \times 3$	30
conv_2_1	$3 \times 3 \times 3$	60
conv_2_2	$3 \times 3 \times 3$	60
conv_3_1	$3 \times 3 \times 3$	120
conv_3_2	$3 \times 3 \times 3$	120
inception	see. Fig. 1	240
inception	see. Fig. 1	240
fc_1	$1 \times 1 \times 1$	400
fc_2	$1 \times 1 \times 1$	200
logits	$1 \times 1 \times 1$	4

Inception2

Layer name	kernel size	num filters
conv_1_1	$3 \times 3 \times 3$	30
conv_1_2	$3 \times 3 \times 3$	30
conv_2_1	$3 \times 3 \times 3$	60
conv_2_2	$3 \times 3 \times 3$	60
conv_3_1	$3 \times 3 \times 3$	120
conv_3_2	$3 \times 3 \times 3$	120
inception	see. Fig. 1	240
inception	see. Fig. 1	240
inception	see. Fig. 1	240
fc_1	$1 \times 1 \times 1$	400
fc_2	$1 \times 1 \times 1$	200
logits	$1 \times 1 \times 1$	4

Inception 3

Table 3. Description of the inception architectures used in the final ensemble.

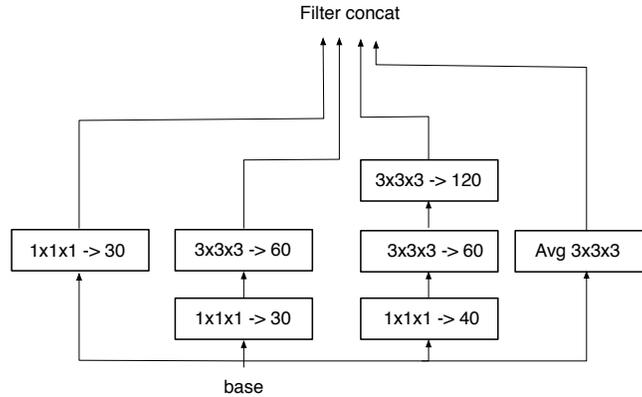


Fig. 1. Structure of the inception layers used in the models of table 3. Each box shows the kernel size and the number of output features.

2.4 Other architectures not in the final ensemble

We also made experiments with other architectures not included in the final ensemble for their lower performance in our training data using cross-validation.

The more innovative structure in this group was based in the Xception architecture presented in [6]. This architecture assumes that correlation in feature planes can be decoupled of spatial correlation, and therefore separability is applied. We implemented this separable 3D spatial filters from scratch in theano (the library only provides this feature for 2D images).

We also made experiments with other inception architectures similar to those presented in figs 6 and 7 of [13]. However, the results on our cross-validated training set were not good enough.

The problem of these other inception architectures and also the Xception layers is that they use more GPU memory compared to a more simple VGG architecture, because more layers are used and more intermediate data and gradients need to be stored. The main advantage of these architectures in 2D images is that they require a smaller number of parameters which help to regularize the model. However, we found that overfitting was not the problem in our cases (the training cost and training error was not negligible), and therefore models with many parameters (as the VGG) could be trained without overfitting.

Finally, we also made experiments with field bias correction of the input data [15]. In these experiments, we corrected the bias of the T1 and T1ce input modalities and compared the performance without the field bias correction and the same neural network architecture. The results with the bias correction were always worse compared to using the original raw data with the same model architecture, and for this reason we omitted field bias correction .

2.5 Number of parameters and receptive field size

Table 4 shows the number of parameters and receptive field size for the models used in our final ensemble. The model that requires more parameters is the VGG-like. This constraint limits the number of layers of the VGG model to avoid GPU memory problems. That is the reason why the VGG-like model has the smallest receptive field size.

The inception models halve the number of parameters (the latter layers are the ones with more parameters) and have a larger receptive field.

Finally the densely connected model is the model with less parameters and largest receptive field.

The idea of our ensemble is to be able to combine models with large receptive field (more context) as the densely connected model and very expressive models, i.e. models with many deep features (VGG-like) so that one models compensate for the weaknesses of the others.

Model	#parameters	receptive field size
VGG-like	3270252	$17 \times 17 \times 17$
inception2	1375872	$21 \times 21 \times 21$
inception3	1611882	$25 \times 25 \times 25$
Densely conected	494220	$41 \times 41 \times 41$

Table 4. Number of parameters and receptive field size for the models used in our enseble

3 Experiments and results

3.1 Data

Our system was evaluated on the data from the Brain Tumour Segmentation Challenge 2018 (BRATS) [11, 4, 5, 3]. As in previous editions, the training set consists of 210 cases with high grade glioma (HGG) and 75 cases with low grade glioma (LGG), for which manual segmentations are provided. The segmentations include the following tumour tissue labels: 1) necrotic core and non enhancing tumour, 2) oedema, 4) enhancing core. Label 3 is not used. The validation set consists of 66 cases, both HGG and LGG but the grade is not revealed. For each subject, four MRI sequences are available, FLAIR, T1, T1 contrast enhanced (T1ce) and T2. The datasets are pre-processed by the organisers and provided as skull-stripped, registered to a common space and resampled to isotropic 1mm3 resolution. Dimensions of each volume are $240 \times 240 \times 155$.

3.2 Implementation details

We implemented everything in python. Input/output data for MRI scans was handled with the nibabel library [7] and neural networks were implemented using tensorflow [1]. The code used in this work will be available here [2].

We did not try any bias field correction of the input scans. The only intensity normalization that we used was z-score normalization of the input scans using the mean and standard deviation of the brain volume only (so the mean and std deviation are not dependent of the brain size).

Models were trained using crops of the original MRI scans. As in [9], the size of each crop was larger than the size of the receptive field. More specifically, the size of the crop is set $(9 + rf) \times (9 + rf) \times (9 + rf)$, where rf is the size of the receptive field. Thus, each crop contributes to the cost function with $9 \times 9 \times 9$ voxels. This approach increases the computational efficiency (reuses many computations) and we think that it also acts as a regularizer, forces the model to be smooth during labeling. For each mini batch, we increased the number of crops to fill the GPU memory (12Gb in our machine). These crops were randomly sampled using a uniform distribution among the four classes: healthy, oedema, core and enhancing core. During evaluation the size of the crops were increased and consecutive crops had some overlap to handle the reduced size of the network output (we used convolutions with only valid support).

Training was done using gradient descent with the Adam optimizer using a constant learning rate of 0.0001 for about 40k steps. We did not observed any overfitting during training, and for this reason we did not investigate into adding any L2, L1 regularization, learning rate decay.... Perhaps one of the reason why we did not observed overfitting is because we implemented a strong data augmentation that generated affine 3D transformations of the MRI scans on the fly.

3.3 Training results

We split the training data in two random sets, so that one of the sets that contained 20% of the patients was used to evaluate the training progress. For each model architecture we generated two different training partitions using different random seeds, so that all training data was used by the models in the ensemble.

We ranked the model architectures using the Dice scores on our validation subset. Table 5 shows the Dice scores for our best models on our validation split. As it can be seen the differences among models are very small, however since the receptive field size and number of parameters is very different we think that the models might have captured complimentary information.

Model name	Dice_WT	Dice_TC	Dice_ET
VGG-like	0.880	0.771	0.689
inception2	0.882	0.792	0.685
inception3	0.880	0.789	0.695
densely connected	0.883	0.787	0.683

Table 5. Results of the selected model architectures on our validation split

3.4 Results on the validation set

We submitted the predicted labels for each of the described models and also for the ensemble model. The ensemble model averages the probabilities of 8 trained models (one for each architecture, and two random partitions of the training set).

Table 6 shows the results provided by the Brats evaluation platform on the blind validation dataset. The results are quite consistent with the results shown on Table 5, and hence we can conclude that we did not overfitted the training dataset and the models generalize quite well on new data. However, the evaluation on the Brats platform shows an interesting point, median values of the Dice scores are much larger than the mean values. This confirms the existence of image outliers.

Model name	mean Dice_WT	mean Dice_TC	mean Dice_ET	median Dice_WT	median Dice_TC	median Dice_ET
VGG-like	0.872	0.760	0.751	0.900	0.837	0.844
inception2	0.877	0.773	0.7533	0.909	0.866	0.858
inception3	0.873	0.776	0.781	0.907	0.852	0.858
densely connected	0.874	0.755	0.729	0.903	0.837	0.846
ensemble	0.881	0.777	0.773	0.912	0.873	0.860

Table 6. Results of the selected model architectures on our validation set

4 Discussion and conclusion

In this paper we have extended some well known architectures for 2D image classification to the problem of 3D image segmentation. This can be easily done by replacing 2D convolutions by their 3D counterparts and adjusting the number of layers and number of feature maps to more appropriate ranges so that models can be fitted in memory.

Each model architecture was chosen so that we had models with large/small receptive field, many/less parameters. The idea is that different configurations can capture complimentary information and an ensemble model can outperform each separate model.

The results on the validation set, show that there no exist many performance differences between the different model architectures, however the ensemble model outperforms each model. These results confirms our hypothesis and are also consistent with the results that we had previously obtained on the training data.

We also tried other models, not included in the final ensemble, such as the 3D Xception that assumes independence between spatial and feature dimensions. We also tried to use bias field correction however our results showed that this was not useful for our models.

A careful analysis of the obtained results shows the existence of image outliers that are not well classified. This issue severely drops our global performance as shown by the huge difference of using the mean or median metrics. We will investigate what is the cause of these outliers for the final evaluation on the test set.

References

1. Abadi, M., et al.: Tensorflow: Large-scale machine learning on heterogeneous systems (2015), <https://www.tensorflow.org/>, software available from tensorflow.org
2. Albiol, A., et al.: https://github.com/alalbiol/brain_segmentation (2018)
3. Bakas, S., others.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. The Cancer Imaging Archive
4. Bakas, S., et al.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Nature Scientific Data* **4**:170117 (2017)
5. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. The Cancer Imaging Archive (2017)
6. Chollet, F.: Xception: Deep learning with depthwise separable convolutions (2016), <http://arxiv.org/abs/1610.02357>, cite arxiv:1610.02357
7. developers, N.: nibabel 1.0.2 (Aug 2016). <https://doi.org/10.5281/zenodo.60861>, <https://doi.org/10.5281/zenodo.60861>
8. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017)
9. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical Image Analysis* **36**, 61 – 78 (2017). <https://doi.org/https://doi.org/10.1016/j.media.2016.10.004>, <http://www.sciencedirect.com/science/article/pii/S1361841516301839>
10. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging* **34**(10), 1993–2024 (Oct 2015). <https://doi.org/10.1109/TMI.2014.2377694>
11. Menze, B., et al.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging* **34**(10), 1993–2024 (2015)
12. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (09 2014)
13. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 2818–2826 (June 2016). <https://doi.org/10.1109/CVPR.2016.308>
14. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: *Computer Vision and Pattern Recognition (CVPR)* (2015), <http://arxiv.org/abs/1409.4842>
15. Tustison, N.J., et al.: N4ITK: improved N3 bias correction. *IEEE transactions on medical imaging* **29**(6), 1310–1320 (Jun 2010). <https://doi.org/10.1109/tmi.2010.2046908>, <http://dx.doi.org/10.1109/tmi.2010.2046908>

GBM Segmentation with 3D U-Net and Survival Prediction with Radiomics

Ujjwal Baid¹, Abhishek Mahajan², Sanjay Talbar¹, Swapnil Rane², Siddhesh Thakur¹, Aliasgar Moiyadi², Meenakshi Thakur², and Sudeep Gupta²

¹ Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, India
{baidujwal,sntalbar}@sggs.ac.in

² Tata Memorial Centre, Mumbai, India drabhishek.mahajan@yahoo.in

Abstract. Segmenting brain tumors from multi-modal imaging remains to be a challenging over decades. Heterogeneous appearance of brain tumors in MRI poses critical challenges in diagnosis, prognosis and survival prediction. In this paper proposed a 3D patch based U-Net model for segmentation task in Brain Tumour Segmentation(BRATS) challenge 2018. Radiomics feature extraction and classification is done on segmented tumor for overall survival(OS) prediction task. Preliminary results on BRATS 2018 validation dataset demonstrate that proposed method achieved good performance with Dice scores as 0.88, 0.83 and 0.75 for whole tumor, tumor core and enhancing tumor respectively. For survival prediction 57.1% accuracy is achieved.

Keywords: Brain Tumor Segmentation · Gliomas · Convolutional Neural Networks · Radiomics

1 Introduction

Glioma is the most common type of brain tumor arising from glial cells. Gliomas may have different degrees of aggressiveness, variable prognosis and several heterogeneous histological sub-regions that are described by varying intensity profiles across different Magnetic Resonance Imaging (MRI) modalities, which reflect diverse tumor biological properties [1]. Accurate segmentation and measurement of the different tumor sub-regions is critical for monitoring progression, surgery or radiotherapy planning and follow-up studies. However, the distinction between tumor and normal tissue is difficult as tumor borders are often fuzzy and there is a high variability in shape, location and extent across patients. Despite recent advances in automated algorithms for brain tumor segmentation in multimodal MRI scans, the problem is still a challenging task in medical imaging analysis [2][3][4].

Prior to BRATS challenge researchers tested their proposed algorithms on local datasets and there is no gold standard available for fair evaluation of methods globally. BraTS challenge have provided global platform for researchers to evaluate their proposed algorithms on publically available dataset with leaderboard. Additionally to the segmentation challenge, BraTS 2018 also required

participants to develop an algorithm for survival prediction based on radiomic imaging features.

Many computational methods based on texture analysis, probabilistic models, active contours have been proposed for tumor segmentation over decades[5]. Early approaches to this problem were based on the detection of abnormalities using healthy-brain atlases and probability models. Then, results were upgraded with the use of deformable registration fields coupled with Markov Random Fields (MRF). Subsequent approaches using machine learning techniques, such as Random Forest improved the results even more.

In recent years, Convolutional Neural Networks (CNN) have shown outstanding results in detection, classification and segmentation tasks, being able to match and sometimes outperform humans. Some of this success is due to the rapid improvement of machines computational power and to CNNs ability of abstracting features in different representations of an image. Fully convolutional networks (FCN) have proven to be an effective way to do pixel-by-pixel classification, obtaining a mean IoU of 67% in the PASCAL-VOC dataset in 2012. This method offers the advantage of combining coarse and shallow semantic information from images with an arbitrary input size.

U-Net based models have outperformed over traditional machine learning methods in bio-medical image segmentation [6]. Recently, there has been an increase in popularity of 3D CNNs which have is effective in segmentation task at the expense of additional computational complexity[7] compared to other state of the art algorithms.

2 Methods

We developed patch based 3D Unet model for tumor segmentation and evaluated efficiency of radiomic features for OS prediction. The BRATS dataset exhibits severe class imbalance, i.e. the tumor pixels are vastly outnumbered by the non-tumor pixels. This poses a problem when training the network, because the non-tumor pixels influence the total loss function much more strongly than the tumor pixels. To address this issue, we adopt a patch-based training approach. During training, we randomly sample patches from the images to form a batch. Each patch is exactly the size of receptive field of the network. Details of our approach are given in this section.

2.1 Dataset

This proposed method is trained and validated on BRATS 2018 training dataset and validation dataset[8][9][10]. The training dataset includes 210 High Grade Glioma(HGG) cases and 75 cases with Low Grade Glioma(LGG) while validation set consists of 66 cases. For each case, there are four MRI sequences, including the T1-weighted (T1), T1 with gadolinium enhancing contrast (T1c), T2-weighted (T2) and FLAIR. All cases have been segmented manually, by four raters and

marked annotations were approved by experienced neuro-radiologists into intra-tumor parts like tumor core, enhancing tumor and edema. The MRI data is collected from various institutions and were acquired with different protocols, magnetic field strengths and MRI scanners. Furthermore, to pinpoint the clinical relevance of this segmentation task, BraTS 2018 also focuses on the prediction of patient overall survival, via analysis of radiomic features. For this purpose the survival data (in days) of 163 cases is provided in training set and 54 in validation set. Reference segmentation and OS for validation set is hidden and evaluation is carried out via online evaluation portal.

2.2 Pre-processing

The pre-processing of dataset is done by the organizers which includes skullstrip, co-register to a common space and resample to isotropic 1mm x 1mm x 1mm resolution. The dimensions of each volume are $240 \times 240 \times 155$. Bias field correction is done on the MR data to correct the intensity in-homogeneity in each channel using N4ITK tool [11]. Each modality of images was normalized by subtracting the mean and dividing by the standard deviation of the intensities within the brain.

2.3 Patch Extraction and Training

Our model is modified version of 3D UNET with 3 downsampling and 3 up-sampling branches with two back to back convolution layers with kernel size 3. 3D voxels with size $64 \times 64 \times 64$ are extracted randomly from the training data and given input to first layer of the model. In patch extraction care is taken to include significant tumor area to avoid bias to background and non tumor pixels. This is done for all the four modalities and ground truth as well. Each layer is followed by ReLU activation and Batch Normalization. The model is trained for 50 epochs and need 48 hours for training on NVIDIA P100 GPU with 128GB system RAM. The framework is developed in Tensorflow[12]. At output 4 probability maps are generated for Necrosis, Edema, Enhancing Tumor and Background(including non tumor brain pixels). The label is assigned to the map with highest probability amongst all. It has been observed there are some False Positives present in the segmentation output. 3D Connected Component Analysis is done to remove false positives. Again, to reduce over-segmentation in certain cases a binary brain mask is generated from brain volume and logical AND operation is performed on segmentation output. This have improved the accuracy of the segmentation significantly.

2.4 Radiomic Feature Extraction

After segmentation of intra-tumor parts next task in BraTS 2018 is to predict the overall survival of the patients. For this task apart from ground truth the only detail organizers have provided is age which makes the task challenging.

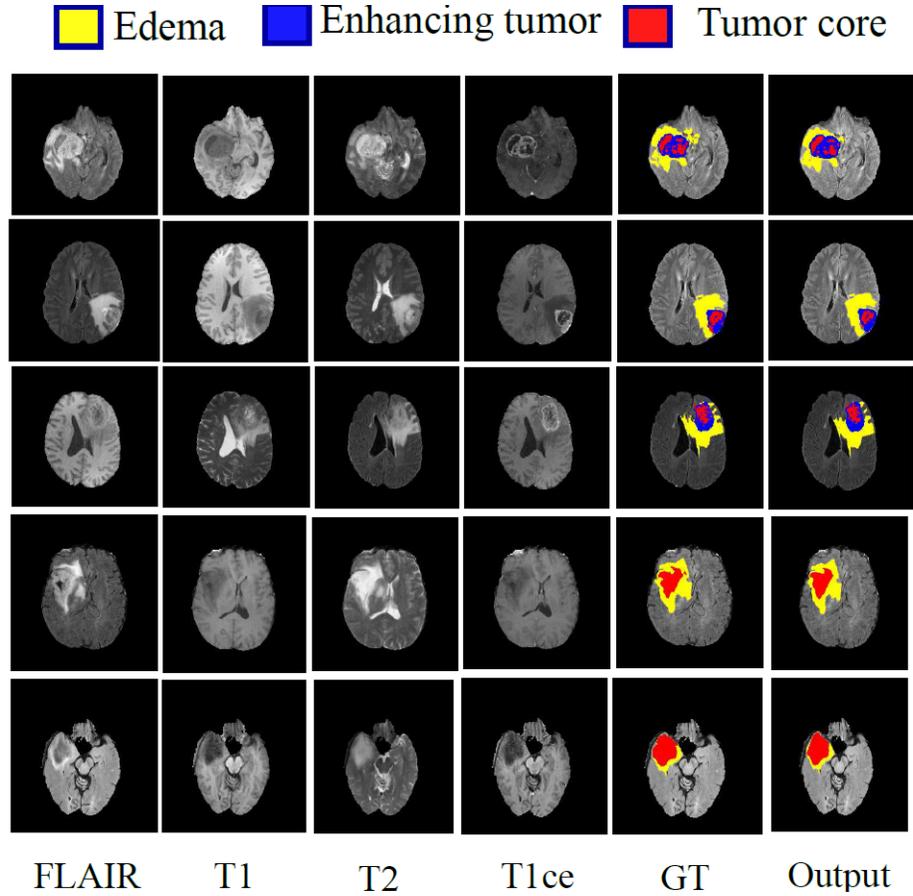


Fig. 1. Sample Segmentation Results. Each row represents one case. Columns from left to right: FLAIR, T1, T2, T1ce, GT and Output. Segmentation labels: Yellow for edema, Blue for enhancing tumor and Red for tumor core

From last few years researchers are working actively on Radiomic Feature extraction for tumor analysis and survival prediction task[13]. We compute first order statistics, shape features, Gray Level Co-occurrence Matrix and Gray Level Run Length Matrix features with pyradiomic package[14] and Cancer Imaging Phenomics Toolkit (CaPTk)[15]. We computed total 468 features for edema, tumor core and enhancing tumor. These features are used to train the regression model for survival prediction task.

3 Preliminary Results and Discussion

The performance of the proposed method is evaluated on BRATS2018 training data with 285 cases and validated on 66 cases for segmentation. Although the testing data has not yet been made available, the validation leaderboard still gives interesting information about the performance of the different teams' algorithms. Average performance of proposed method on training data and validation data is given in Table 1 and Table 2 respectively in terms of Dice Similarity Index and Sensitivity. Overall, our approach reached a superior result in the whole tumor segmentation task with an average dice coefficient of 93% over training dataset and 87% over validation dataset. Sample segmentation results for intra-tumor parts is given in Figure 1.

Next task is to predict the overall survival of the cases. For regression and classification we used WEKA software toolkit[16]. Various machine learning classifiers are available in WEKA out of which we used Multi-Layer Perceptron and Random Forests. We achieved 57.1% accuracy using MLP for OS prediction task and achieved second position in leader-board at the time of submission of the manuscript.

Table 1. Performance of proposed method on BraTS 2018 Training data

Evaluation Metrics	Dice			Sensitivity		
	ET	WT	TC	ET	WT	TC
Mean	0.8002	0.9324	0.9197	0.8951	0.9508	0.9359
StdDev	0.2746	0.1056	0.1327	0.1397	0.0859	0.1005
Median	0.9062	0.9613	0.9564	0.9313	0.9645	0.9529
25quantile	0.8421	0.9405	0.9303	0.8901	0.9451	0.9328
75quantile	0.9422	0.9728	0.9687	0.9622	0.9777	0.9697

4 Conclusion

In this study we proposed a 3D Unet network for brain tumor segmentation and efficacy of radiomic features for overall survival prediction. For segmentation task we extracted 3D patches for training. Radiomic features are extracted from all four MR modalities for OS prediction. The difference between mean and median

Table 2. Performance of proposed method on BraTS 2018 validation data

Evaluation Metrics	DICE			Sensitivity		
	ET	WT	TC	ET	WT	TC
Mean	0.7480	0.8780	0.8266	0.8266	0.9058	0.8186
StdDev	0.2659	0.1345	0.1828	0.2306	0.1413	0.2136
Median	0.8527	0.9179	0.8985	0.9035	0.9436	0.9167
25quantile	0.7325	0.8665	0.7771	0.8238	0.8953	0.7637
75quantile	0.8853	0.9419	0.9444	0.9514	0.9720	0.9571

Table 3. Performance of Random Forest and Multilayer perceptron for OS prediction on Validation Dataset

Method	Accuracy	MSE	medianSE	stdSE	SpearmanR
RF	0.375	109105.605	47545.125	143070.049	0.11
MLP	0.571	59550213.1	113611.616	128250465.8	0.427

in Table 2 indicates, that across all classes and measures few outliers caused a great impact on the scores and this needs further investigation.

Prediction of survival without more clinical data and treatment information is challenging and same is reflected through accuracy of the participants in the leader-board. As the number of cases for OS prediction are less there is need to develop efficient feature selection algorithm which will select potential features for accurate OS prediction.

Acknowledgement

This work was supported by Ministry of Electronics and Information Technology, Govt. of India under Visvesvaraya PhD scheme.

References

1. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging* 34(10), 1993–2024 (Oct 2015)
2. Mahajan, A., Moiyadi, A., Jalali, R., Sridhar, E.: Radiogenomics of glioblastoma: a window into its imaging and molecular variability 15, P14 (10 2015)

3. Mahajan, A., Goh, V., Basu, S., Vaish, R., Weeks, A., Thakur, M., Cook, G.: Bench to bedside molecular functional imaging in translational cancer medicine: To image or to imagine? 70 (07 2015)
4. Seow, P., Wong, J., Ahmad Annuar, A., Mahajan, A., Aniza Abdullah, N., Ramli, N.: Quantitative magnetic resonance imaging and radiogenomic biomarkers for glioma characterisation: a systematic review p. 20170930 (06 2018)
5. Gordillo, N., Montseny, E., Sobrevilla, P.: State of the art survey on mri brain tumor segmentation. *Magnetic Resonance Imaging* 31(8), 1426 – 1438 (2013), <http://www.sciencedirect.com/science/article/pii/S0730725X13001872>
6. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation pp. 1–8 (2015), <http://arxiv.org/abs/1505.04597>
7. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d cnn with fully connected ctf for accurate brain lesion segmentation. *Medical Image Analysis* 36, 61 – 78 (2017), <http://www.sciencedirect.com/science/article/pii/S1361841516301839>
8. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Nature Scientific Data* 170117(4) (2017)
9. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for pre operative scans of the tcga-lgg collection. *The Cancer Imaging Archive* 170117 (2017)
10. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for pre operative scans of the tcga-gbm collection. *The Cancer Imaging Archive* 170117 (2017)
11. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4itk: Improved n3 bias correction. *IEEE Transactions on Medical Imaging* 29(6), 1310–1320 (June 2010)
12. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), <https://www.tensorflow.org/>, software available from tensorflow.org
13. Kickingereder, P., Burth, S., Wick, A., Gtz, M., Eidel, O., Schlemmer, H.P., Maier-Hein, K.H., Wick, W., Bendszus, M., Radbruch, A., Bonekamp, D.: Radiomic profiling of glioblastoma: Identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. *Radiology* 280(3), 880–889 (2016), <https://doi.org/10.1148/radiol.2016160845>, pMID: 27326665
14. van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G., Fillion-Robin, J.C., Pieper, S., Aerts, H.J.: Computational radiomics system to decode the radiographic phenotype. *Cancer Research* 77(21), e104–e107 (2017), <http://cancerres.aacrjournals.org/content/77/21/e104>
15. Rathore, S., Bakas, S., Pati, S., Akbari, H., Kalarot, R., Sridharan, P., Rozycki, M., Bergman, M., Tunç, B., Verma, R., Bilello, M., Davatzikos, C.: Brain cancer

- imaging phenomics toolkit (brain-captk): An interactive platform for quantitative analysis of glioblastoma. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries - Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers*. pp. 133–145 (2017), https://doi.org/10.1007/978-3-319-75238-9_12
16. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: An update. *SIGKDD Explor. Newsl.* 11(1), 10–18 (Nov 2009), <http://doi.acm.org/10.1145/1656274.1656278>

Multi-Planar Spatial-ConvNet for Segmentation and Survival Prediction in Brain Cancer

Subhashis Banerjee^{1,2}, Sushmita Mitra¹, and B. Uma Shankar¹

¹ Machine Intelligence Unit, Indian Statistical Institute, Kolkata, INDIA

² Department of CSE, University of Calcutta, Kolkata, INDIA
mail.sb88@gmail.com, sushmita@isical.ac.in, uma@isical.ac.in

Abstract. A new deep learning method is proposed for the automatic delineation/segmentation of brain tumors from multi-sequence MR images. A Radiomic model for predicting the Overall Survival (OS) is designed, based on the features extracted from the segmented Volume of Interest (VOI). An encoder-decoder type ConvNet model is designed for pixel-wise segmentation of the tumor along three anatomical planes (axial, sagittal and coronal) at the slice level. These are then combined, using a consensus fusion strategy, to produce the final volumetric 3D segmentation of the tumor and its sub-regions. Novel concepts such as spatial pooling and unpooling are introduced to preserve the spatial locations of the edge pixels to reduce segmentation error around the boundaries. We also incorporate shortcut connections to copy and concatenate the receptive fields from the encoder to the decoder part, to help the decoder network localize and recover the object details more effectively. These connections allow the network to simultaneously incorporate high-level features with the pixel-level details. The proposed segmentation and OS models are trained and validated on the BRATS 2018 dataset.

Keywords: Convolutional Neural Network · Spatial-Pooling · Generalised Dice Loss · Glioblastoma Multiforme

1 Introduction

Gliomas are the most common and aggressive malignant brain tumors originating from the glial cells in the central nervous system. Based on their aggressiveness and infiltrative nature, they are broadly classified into two categories, viz. High-Grade Glioma or Glioblastoma Multiforme (HGG/GBM) and Low-Grade Glioma (LGG). over the last few decades, Magnetic Resonance Imaging (MRI) has been extensively employed in diagnosing brain and nervous system abnormalities mainly due to its improved soft tissue contrast. The MR sequences include $T1$ -weighted, $T2$ -weighted, $T1$ -weighted contrast enhanced ($T1C$), and $T2$ -weighed with fluid-attenuated inversion recovery (FLAIR). The rationale behind using these four sequences lies in the fact that different tumor regions may be visible in different sequences, thereby allowing for a more accurate demarcation of the tumor [2, 3].

Accurate delineation of tumor region in MRI sequences is of great importance since it allows: *i)* volumetric measurement of the tumor, *ii)* monitoring of tumor growth in the patient between multiple MRI scans, and *iii)* treatment planning with follow-up evaluation, including the prediction of overall survival (OS). Manual segmentation of tumors from MRI is a highly tedious, time-consuming and error-prone task, mainly due to factors such as human fatigue, overabundance of MRI slices per patient, and an increasing number of patients. Such manual operations often lead to inaccurate delineation, and the need for an automated or semi-automated Computer Aided Diagnosis thus becomes apparent [4, 5, 8]. The large spatial and structural variability among brain tumors make automatic segmentation a challenging problem. The distinctive segmentation of both HGG and LGG by the same model is also a difficult proposition.

Inspired by the success of Convolutional Neural Networks (ConvNets), we develop a novel ConvNet model with spatial-pooling called Spatial-ConvNet. This can preserve the edge information during automated segmentation of gliomas from multi-sequence MRI data. The segmented Volume of Interest (VOI) or tumor is used to extract two categories of Radiomic features, viz. “semantic” and “agnostic”, for predicting the OS of patients.

The rest of the paper is organized as follows. Section 2 provides details about the data, preparation of patch database for the ConvNet training, the proposed multi-planar ConvNet model with spatial-pooling layer, loss function for unbalanced segmentation, and radiomic analysis of the segmented VOI for OS prediction. Section 3 describes the experimental results of the segmentation and OS prediction, demonstrating their effectiveness both qualitatively and quantitatively. Finally, conclusions are provided in Section 4.

2 Material and Methods

In this section we discuss the BRATS 2018 data, and steps of the tumor segmentation and survival rate prediction. The proposed segmentation method comprises of extraction of patches, training and testing of the segmentation model, post-processing, radiomic feature extraction for overall survival prediction, followed by training and testing of classifier for OS.

2.1 Dataset

Brain tumor MRI scan datasets and patient Overall Survival (OS) data, used in this research, were provided by BraTS 18 Challenge [1, 6, 10, 11]. The dataset consists of 210 HGG/GBM and 75 LGG glioma cases as training dataset and 66 combined cases of HGG/GBM and LGG as validation dataset. The OS data was included with correspondences to the pseudo-identifiers of the GBM/HGG imaging data, with 163 and 53 as validation data points. Each patient MRI scan set consist of four MRI sequences or channels, encompassing native ($T1$) and post-contrast enhanced $T1$ -weighted ($T1C$), $T2$ -weighted ($T2$), and $T2$ Fluid-Attenuated Inversion Recovery ($FLAIR$) volumes, having 155 slices of 240×240

resolution. The data is already aligned to the same anatomical template, skull-stripped, and interpolated to 1mm^3 voxel resolution. The manual segmentation of volume structures was performed by experts following the same annotation protocol, and their annotations were revised and approved by board-certified neuro-radiologists. Annotation labels included were the GD-enhancing tumor (ET), the peritumoral edema (ED), and the necrotic and non-enhancing tumor (NCR/NET). The predicted labels are evaluated by merging three regions: whole tumor (WT: all the three labels), tumor core (TC: ET and NCR/NET) and enhancing tumor (ET).

The OS data is defined in terms of days, and also includes the age of patients along with their resection status. Only the subjects with the resection status GTR (Gross Total Resection) is considered for the evaluation of the OS prediction task. Based on the number of survival days, the subjects are grouped into three classes viz. long-survivors (> 15 months), short-survivors (< 10 months), and mid-survivors (between 10 to 15 months).

2.2 Multi-Planar ConvNet with Spatial-Pooling for segmentation

MRI scans are volumetric and can be represented in three-dimensions using multi-planar representation viz. axial (X - Z axes), coronal (Y - X axes), and sagittal (Y - Z axes) planes. Taking advantage of this multi-view property, we propose a deep learning based segmentation model that uses three separate ConvNets for segmenting the tumor in the three individual planes at the slice level. These are then combined using a consensus fusion strategy to produce the final volumetric 3D segmentation of the tumor and its sub regions. It is observed that the integrated prediction from multiple planes is superior, in terms of accuracy and robustness of decision, with respect to the estimation based on any single plane. This is perhaps because of utilizing more more information, while minimizing loss.

The ConvNet architecture, used for slice wise segmentation along each plane, is an encoder-decoder type of network. The encoder or the contracting path uses pooling layers to down sample an image into a set of high-level features, followed by a decoder or an expanding part which uses the feature information to construct a pixel-wise segmentation mask. The main problem with this type of networks is that, during the down sampling or the pooling operation network loses the spatial information. Up sampling in the decoder network then tries to approximate this through interpolation. This produces segmentation error around the boundary of the region-of-interest (ROI) or volume-of-interest (VOI). This is a major drawback in medical image segmentation, where accurate delineation is of utmost importance.

In order to circumvent this problem we introduce spatial-max-pooling layer, which can retain the max locations to be subsequently used in the unpooling operation through the spatial-max-unpooling layer. The procedure is illustrated in Fig. 1. We also incorporate shortcut connections to copy and concatenate the receptive fields (after convolution block) from the encoder to the decoder part to help the decoder network localize and recover the object details more effectively.

These connections allow the network to simultaneously incorporate high-level features with the pixel-level details. The entire segmentation model architecture is depicted in Fig. 2.

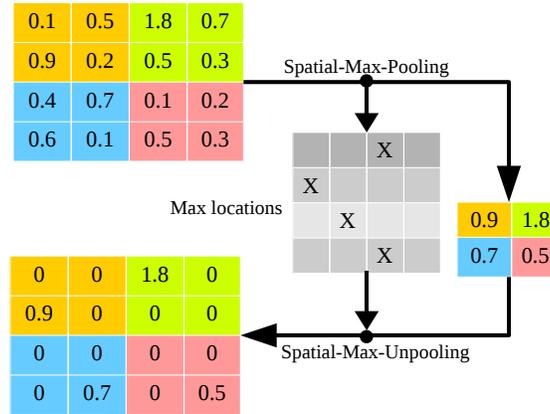


Fig. 1: Spatial Pooling and Unpooling operations.

Tumors are typically heterogeneous, depending on cancer subtypes, and contain a mixture of structural and patch-level variability. Applying a ConvNet directly to the entire slice has its inherent drawbacks. Since the size of each slice is 240×240 , therefore if we train the ConvNet on the whole image/slice then the number of parameters to train will be huge. Moreover, very little difference is observable in adjacent MRI slices at the global level; whereas patches generated from the same slice often exhibit significant dissimilarity. Besides, the segmentation classes are highly imbalanced. Approximately 98% of the voxels belong to either the healthy tissue or to the black surrounding area. The NCR/NET volumes are of the lowest size amongst all the three classes, as depicted in Fig. 3.

Each ConvNet trained on patches of size, $128 \times 128 \times 4$ extracted from all the four MRI sequences corresponding to a particular plane. A randomized patch extraction algorithm, developed in our lab, is employed. The patch selection is done using an entropy based criterion. The three ConvNets (along the three plane) are trained end-to-end/pixels-to-pixels based on the patches extracted from the corresponding ground truth images. During testing the stack of slices are fed to the model, to produce pixel-wise segmentation of the tumor along the three planes. The training performance is evaluated using Dice overlap score [7], for the three segmented sub-regions WT, ET and TC. Since the dataset is highly imbalanced therefore standard loss functions used in literature are not suitable for training and optimizing the ConvNet. This is because in such cases most classifiers focus on learning the larger classes, thereby resulting in poor classification accuracy for the smaller classes. Therefore, we propose a new loss

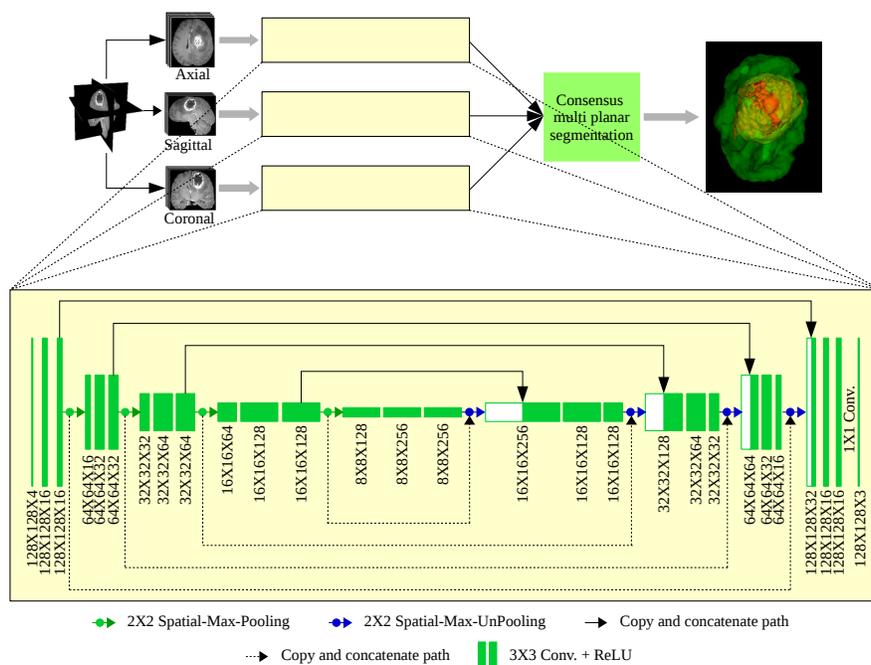
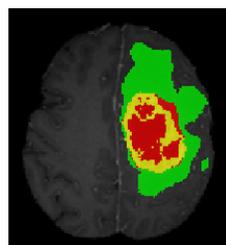


Fig. 2: Multi-planar ConvNet architecture for segmentation.



Class	Voxel Count	%
Background	8790699	98.46
ED	90806	1.02
ET	26264	0.29
CNE/NET	20231	0.23

Fig. 3: Tumor sub-class distribution for a sample patient.

function, which is a sum of two losses viz. – Generalized Dice loss [12] and Weighted Cross-entropy [9].

2.3 Overall Survival prediction based on radiomic features

For the OS prediction task we have extracted two types of Radiomic features, viz. “semantic” and “agnostic” [2]. The former includes attributes like size, shape, location, vascularity, spiculation, necrosis, and the latter attempting to capture lesion heterogeneity through quantitative descriptors like histogram, texture, etc. We extracted 33 semantic and 50 agnostic features from each segmented VOI. Then feed them to a Multilayer Perceptron (MLP) to predict the number of survival days which is further used to determine the survival class (short, mid or long).

3 Preliminary Results

The ConvNet models were developed using TensorFlow, with Keras in Python. The experiments were performed on the Intel AI DevCloud platform having cluster of Intel Xeon Scalable processors. Codes developed for our experiments will be available soon. The proposed segmentation model is trained and validated on the corresponding training and validation datasets provided by the BRATS 2018 [1, 10, 11] organizers.

Fig. 4 reports the preliminary quantitative evaluation results obtained by the proposed segmentation model on the BRATS 2018 validation dataset (on 66 patients). Quantitative metrics used for evaluating the segmentation results w.r.t. the gold standard (in case of training) and through the Leaderboard/blind testing (in case of validation) are (i) Dice score, (ii) sensitivity, (ii) specificity and (iii) Hausdorff distance computer for WT, TC and ET. Qualitative segmentation result obtained by the proposed method for a sample patient from the validation dataset is shown in Fig. 5.

Preliminary results of the proposed OS prediction method is reported in the Table 1. We used 80% of the training data (130 patients) for training, 20% (33 patients) for validation. The model was finally tested on the 28 patients having resection status GTR from the validation set through the Leaderboard blind testing.

Accuracy	MSE	medianSE	stdSE	SpearmanR
0.464	118343.214	32186.073	236471.312	0.242

Table 1: OS prediction result are reported on the BraTS 2018 validation dataset.

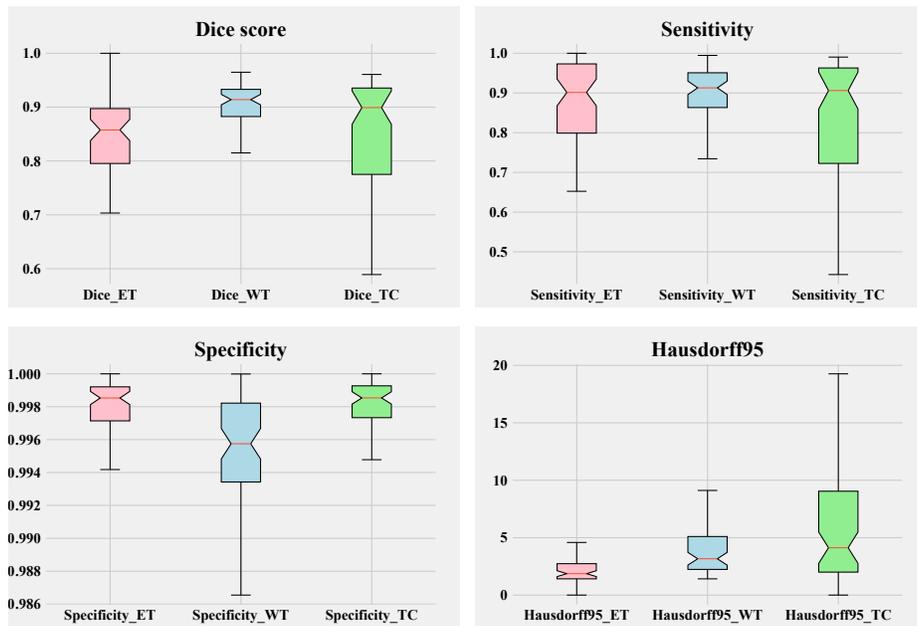


Fig. 4: Box-and-whisker plot of segmentation accuracy of the three sub-regions ET, WT and TC observers with Dice score, Sensitivity, Specificity and Hausdorff95.

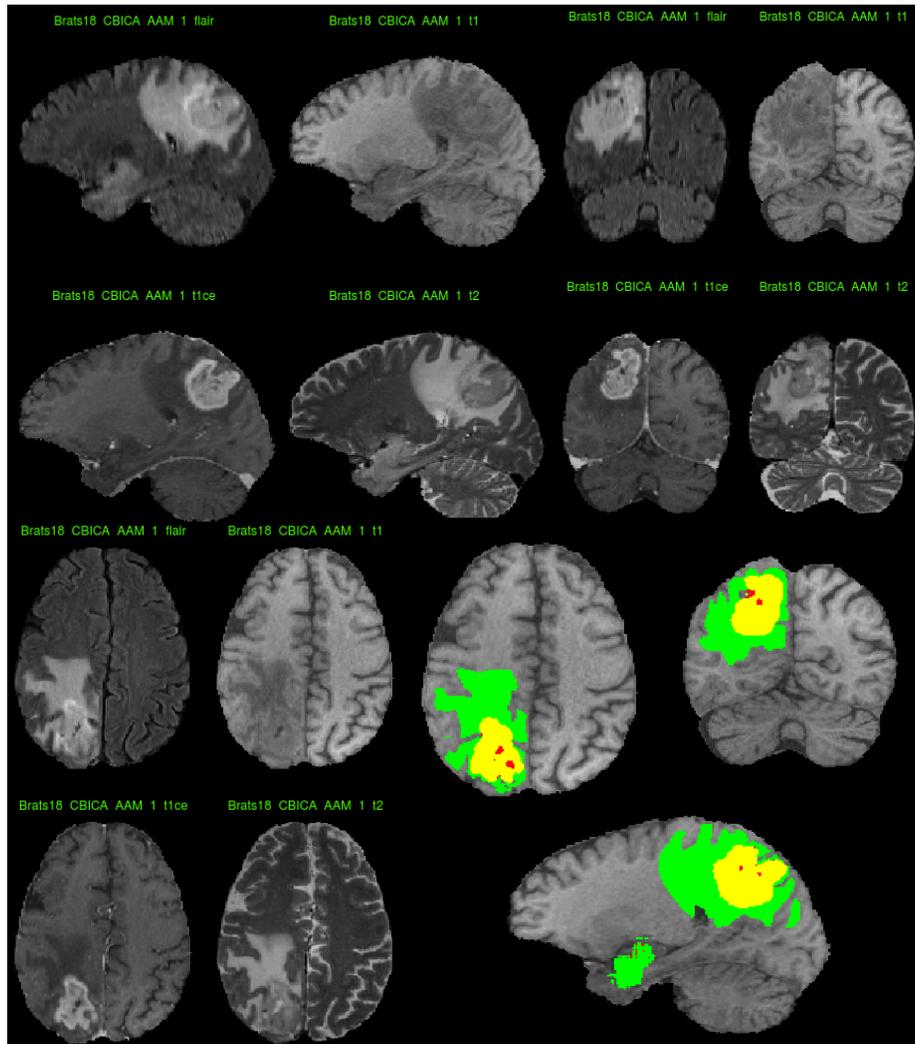


Fig. 5: Example segmentation result for a sample patient (ID: Brats18_CBICA_AAM_1) from BraTS 2017 Validation Data. The green label is edema, the red label is nonenhancing or necrotic tumor core, and the yellow label is enhancing tumor core. The Dice coefficients for the ET, WT and TC segmentation for this patient are 0.92, 0.90, 0.93 respectively

4 Conclusion

In this paper we proposed a deep learning method for the automatic delineation/segmentation of brain tumors from multi-sequence MR images. The encoder-decoder type ConvNet model for pixel-wise segmentation performs better than other patch based models. Integrated prediction from multiple anatomical planes (axial, sagittal and coronal) performs superior, in terms of accuracy and robustness of decision, with respect to the estimation based on any single plane. Novel concepts such as spatial pooling and unpooling reduced segmentation error around the boundary of the VOI. We also incorporate shortcut connections to copy and concatenate the receptive fields from the encoder to the decoder part, to help the decoder network localize and recover the object details more effectively. We are getting very good validation accuracy for the segmentation task, but for the OS task we are experimenting with some other new features and feature selection methods to improve the accuracy.

Acknowledgment

We gratefully acknowledge the support of Intel Corporation for giving access to the Intel AI DevCloud platform used for this work.

One of the authors, S. Banerjee, acknowledges financial support from the Visvesvaraya PhD Scheme by Ministry of Electronics and Information Technology (MeitY), Government of India.

References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data* **4**, sdata2017117 (2017)
2. Banerjee, S., Mitra, S., Shankar, B.U.: Synergetic neuro-fuzzy feature selection and classification of brain tumors. In: *Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. pp. 1–6 (2017)
3. Banerjee, S., Mitra, S., Uma Shankar, B.: Single seed delineation of brain tumor using multi-thresholding. *Information Sciences* **330**, 88–103 (2016)
4. Banerjee, S., Mitra, S., Uma Shankar, B.: Automated 3D segmentation of brain tumor using visual saliency. *Information Sciences* **424**, 337–353 (2018)
5. Banerjee, S., Mitra, S., Uma Shankar, B., Hayashi, Y.: A novel GBM saliency detection model using multi-channel MRI. *PLOS ONE* **11**(1), e0146388 (2016)
6. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal Brain Tumor image Segmentation benchmark (BraTS). *IEEE Transactions on Medical Imaging* **34**, 1993–2024 (2015)
7. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *3D Vision (3DV), 2016 Fourth International Conference on*. pp. 565–571. IEEE (2016)

8. Mitra, S., Banerjee, S., Hayashi, Y.: Volumetric brain tumour detection from MRI using visual saliency. *PLOS ONE* **12**, 1–14 (2017)
9. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 234–241. Springer (2015)
10. Spyridon, B., Hamed, A., Aristeidis, S., Michel, B., Martin, R., Justin, K., John, F., Keyvan, F., , Christos, D.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *Cancer Imaging Arch* (2017)
11. Spyridon, B., Hamed, A., Aristeidis, S., Michel, B., Martin, R., Justin, K., John, F., Keyvan, F., , Christos, D.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *Cancer Imaging Arch* (2017)
12. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 240–248. Springer (2017)

Deep Hourglass for Brain Tumor Segmentation

Eze Benson¹, Michael P. Pound¹, Andrew P. French^{1,2}, Aaron S. Jackson¹, Tony P. Pridmore¹

¹ School of Computer Science, University of Nottingham

² School of Biosciences, University of Nottingham

ezenwoko.benson@nottingham.ac.uk

michael.pound@nottingham.ac.uk

andrew.p.french@nottingham.ac.uk

aaron.jackson@nottingham.ac.uk

tony.pridmore@nottingham.ac.uk

Abstract. The segmentation of a brain tumour in an MRI scan is a challenging task, in this paper we present our results for this problem via the BraTS 2018 challenge, consisting of 210 HGG and 75 LGG volumes. We train and evaluate a CNN encoder-decoder network based on a singular hourglass structure. The hourglass network is able to classify the whole tumour, enhancing tumour and core tumour in one pass. We apply a small amount of preprocessing to the data before feeding it to the network but no post processing. We achieve an overall Dice coefficient of 92% on the training set. On the validation set our network achieves Dice coefficients of 0.59, 0.82 and 0.63 for ET, WT and TC.

Keywords: Convolutional Neural Network, Deep Learning, Hourglass, Glioma

1 Introduction

Identifying regions of the brain which are tumourous is a task often carried out by medical professionals. Manually classifying segments of the tumour is a subset of a group of problems commonly referred to as semantic segmentation. Semantic segmentation is the task of assigning a class to each pixel within an image, modern automated solutions to this problem often use convolutional neural networks (CNN). The introduction of fully convolutional networks (FCN) [1] established a convolutional neural network architecture that is widely used for the task of semantic segmentation. Architectures such as U-NET [2] achieved success in biomedical imaging by adopting a similar architecture.

We propose the use of an adapted hourglass [3] network to solve the problem of tumour segmentation. The hourglass network improves on U-NET by using bottleneck blocks and adding convolutions to the skip connections. Training a CNN for this problem is a natural choice as they have demonstrated state-of-the-art performance on semantic segmentation problems such as the widely used Pascal VOC2012 [9] and cityscapes [10] datasets.

2 Methods

2.1 Data

The dataset of BraTS 2018 [4-7] provides defined training and validation sets. The training set is composed of 210 MRI scans of high grade gliomas (HGG) and 75 MRI scans of low grade gliomas (LGG). Whilst the validation set is a group of 66 mixed HGG and LGG tumours. The MRIs are volumes with XxYxZ dimensions of 240x240x155. Each volume has four corresponding modalities FLAIR T1,T2 and T1CE.

2.2 PreProcessing

A high variance in intensity in both validation and training set was observed this lead us normalise the training set to be centred around zero with a standard deviation of one. By normalizing the data, we found that the required training time was reduced and the accuracy of the network was increased. The formula for normalization is given in figure 1. Each modality was normalized separately due to the variance in intensity profile between modalities.

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

Where x is the current intensity, μ is the mean of the modality and σ is the standard deviation of the modality.

2.3 Hourglass Architecture

Our approach is to handle 2D slices of each volume separately, a 2D semantic segmentation problem. We performed additional experimentation using a volumetric encoder-decoder but found that the benefit of an end-to-end volumetric approach was outweighed by the significant necessary drop in features at each layer due to memory restrictions.

We design our network using an encoder-decoder structure, adapted from an hourglass network, popularized in the domain of human-pose estimation [3] The structure of the hourglass is similar to other encoder-decoder networks, but contains a denser use of residual blocks throughout.

The encoder contains 7 residual bottleneck blocks [8], after each a max-pooling layer performs spatial downsampling. A further three residual blocks at the lowest spatial resolution derive higher-level features before a series of bilinear upsampling operations return the network to the original spatial resolution. As in the encoder, all upsampling operations of the decoder are interleaved with residual blocks. Skip layers are added between each matching resolution of the encoder and decoder, with each containing an additional residual block to learn an appropriate mapping.

We found that the choice of upsampling layer (e.g. bilinear, max-unpooling [11]) made little difference to the performance of the network. Unlike the original work [3] we chose not to stack hourglass networks sequentially and perform intermediate supervision, we found this too had a negligible effect on performance. The number of spatial-downsampling layers, 7 in total, were chosen based on the input resolution. Only one residual block is used at each depth because adding two at all depths immediately doubles memory consumption which surpasses current memory constraints.

2.4 Training

The training was split into two phases pre and post true validation set release. In the first phase the dataset was split into a test set, validation set and training set where each set was 10%,10% and 80% of the original training set respectively. The data provided is treated as though it is the entire dataset so that our training can be validated and tested in preparation for the true validation set. This allows the network to avoid overfitting and approximate the results expected on the release of the second dataset. Later the network is retrained using a 10% test set and 90% training set split in order to obtain test results on the original data whilst maximizing the training set size. The network is trained for the same number of epochs for all training. The second phase is conducted post true validation set release. In this phase the BraTS dataset is split into 10% validation and 90% training.

The network is trained using an identical training scheme for both the natural and augmented dataset.

The hourglass network implemented in this paper only uses spatial convolutions which means the data must be sliced along the depth dimension which effectively converts a volume into a series of 155 image with a spatial resolution of 256^2 . The volumes have a spatial resolution of 240^2 however for convenience we pad them to the new resolution. The dataset used is then 44175 images instead of 285 volumes. All four modalities are used to train the network and are inputted as different channels to the network.

The hourglass is trained using a cross entropy loss function with a learning rate of 10^{-5} which is decreased by a factor of 10 every 30 epochs. The network is trained for a total of 50 epochs therefore the learning rate is only adapted once. The adaptive gradient descent algorithm, RMSProp is used to train the network faster than the typical stochastic gradient descent.

2.5 Data Augmentation

Two methods of data augmentation are used in this paper vertical flipping and random intensity variation. Vertical flipping is used because it matches the natural symmetrical shape of the brain.

Random intensity variation is used because the intensity between MRI scans varies significantly. This is shown by the fact that the standard deviation of the FLAIR modality in the dataset is greater than the mean by almost a factor of 10. E.g. The standard deviation and mean for the FLAIR modality are 529.2 and 61.8 respectively. The T1, T1CE and T2 modalities have similar standard deviations. Intensity variation is performed on the normalised dataset by first rescaling the standard deviation of the dataset and then shifting the mean. This allows the dataset to include image intensities which are not present in the original dataset but could appear on an MRI volume. The range for randomly changing the standard deviation is between zero and two. The mean is shifted between values of 0.4 and -0.4. Values above a standard deviation of two were experimented with but lead to a significant decrease in accuracy. Shifting the mean by over 0.5 and under -0.5 were trialed but also caused an accuracy decrease.

3 Results & Discussion

The results are split into two sections, the results on the training data set and the results on the later released validation set. Results are shown for networks trained on the standard data and on augmented data.

3.1 Training Dataset

We trained the network on 90% of the data leaving 10% for testing purposes. The network achieved a Dice coefficient of 92% with an IOU of 86%. We find that IOU approximates the network's worst performance on the test set in contrast to Dice which gives an approximate representation of the average case.

3.2 Validation Dataset

The results presented in this section are those achieved when segmenting the validation set using the network trained in section 3.1. Table 1 shows the results of the segmentation without augmentation and table 2 shows the results with flipping and intensity variation. The metrics provided in both tables are the standard metrics output by the BraTS automatic online evaluation server. Some metrics have been omitted to save space, only the most important evaluation metrics have been included.

Table 1. The results of the hourglass network segmenting the unseen validation set without augmentation in the training data.

	Dice ET	Dice WT	Dice TC	Hausdorff ET	Hausdorff WT	Hausdorff TC
Mean	0.59102	0.81638	0.63479	18.11974	94.28005	130.6982
Std	0.28441	0.12274	0.24233	26.62022	50.15014	42.39722
Median	0.70712	0.86227	0.70526	5.73233	97.12933	132.5891
25 Quantile	0.47849	0.7832	0.50762	3.16228	52.71734	103.3565
75 Quantile	0.80193	0.89504	0.82568	20.02904	135.8139	163.3901

Table 2. The results of the hourglass network segmenting unseen validation set where the network has been trained with augmented data

	Dice ET	Dice WT	Dice TC	Hausdorff ET	Hausdorff WT	Hausdorff TC
Mean	0.56337	0.82204	0.61797	14.27762	13.57432	17.94668
Std	0.2888	0.12962	0.21713	23.25875	15.31909	18.13535
Median	0.66684	0.86688	0.6712	5.91548	6.59447	11.18034
25 Quantile	0.39652	0.78219	0.49556	2.82843	4.18205	8.29669
75 Quantile	0.79578	0.90147	0.79259	12.55482	14.96802	18.78738

After comparing the metrics between a dataset with augmentation and one without we find that in this challenge augmentation does not appear to give any significant increase in accuracy for Dice coefficient but improves the Hausdorff accuracies. It is likely the case that the frequency at which the network misclassifies pixels remains the same but the network's ability to localize the pixels is increased.

Overall the network segments the whole tumour more accurately than it does the core tumour or enhancing tumour, from the results in previous challenges this result is expected. Naturally the enhancing and core tumour are much more difficult to segment due to the similarity between all classes.

Table 1 and table 2 both show a large disparity between the median and mean accuracy especially with results for the enhancing tumour where the difference is around 10%. The difference is caused by the difficulty of detecting the enhancing tumour and core tumour in some volumes. In most volumes the Dice coefficients are well above the mean however some outliers achieve a score of 0 therefore dragging the mean down significantly. When removing these cases the mean Dice coefficient increases by 4% showing that the disparity can be explained by a few very difficult volumes. Some examples of the metrics achieved on these volumes are shown in table 3.

Table 3. Segmentation results for very difficult volumes using a network trained with augmented data

	Dice ET	Dice WT	Dice TC	Hausdorff ET	Hausdorff WT	Hausdorff TC
Brats18_TCIA09_248_1	0	0.79274	0.62938	0	14.17745	10.81665
Brats18_TCIA10_195_1	0	0.79941	0.62923	0	15.23155	25.98076
Brats18_TCIA11_612_1	0	0.73506	0.60061	0	52.7731	48.51804
Brats18_TCIA12_613_1	0	0.69355	0.2629	0	49.96999	9
Brats18_TCIA13_646_1	0	0.90187	0.3996	0	35.82736	6.48074

4 Conclusion

We propose a solution which achieves a 92% Dice coefficient on the training set and 0.59, 0.82 and 0.63 on the validation set. Although the network underperforms on Dice score it achieves a competitive Hausdorff distance.

Much of the networks underperformance is related to outliers in the validation set which could be mitigated in future with better preprocessing techniques. In the future we plan to add targeted residual blocks to increase performance. Memory consumption is often a problem when using CNNs, to combat this we plan to add residual blocks in depths which increase the overall accuracy of the network the most. We also plan to add skip connections with an inception block structure [12] as shown in [13] to increase accuracy further.

We show that 2D architectures can segment 3D volumes with success but require fine tuning and a deeper architecture to achieve better results. An approach to bridge the gap may between 2D and 3D may be required. 3D networks outperform 2D networks when depth context is key how much context is required in most tasks remains unclear. We plan to use a 2.5D approach where each slice has an accompanying adjacent slice either side to provide some depth context.

References

1. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition 2015* (pp. 3431-3440).
2. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention 2015 Oct 5* (pp. 234-241). Springer, Cham.
3. Newell A, Yang K, Deng J. Stacked hourglass networks for human pose estimation. In *European Conference on Computer Vision 2016 Oct 8* (pp. 483-499). Springer, Cham.
4. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging*. 2015 Oct;34(10):1993.
5. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific data*. 2017 Sep 5;4:170117.
6. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. *The Cancer Imaging Archive*. 2017;286.
7. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. *The Cancer Imaging Archive*. 2017;286.
8. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition 2016* (pp. 770-778).
9. Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes (voc) challenge. *International journal of computer vision*. 2010 Jun 1;88(2):303-38.
10. Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S, Schiele B. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition 2016* (pp. 3213-3223).
11. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *arXiv preprint arXiv:1511.00561*. 2015 Nov 2.

12. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition 2015 (pp. 1-9).
13. Bulat A, Tzimiropoulos G. Binarized convolutional landmark localizers for human pose estimation and face alignment with limited resources. In The IEEE International Conference on Computer Vision (ICCV) 2017 Oct 1 (Vol. 1, No. 2, p. 4).

Survival prediction using ensemble tumor segmentation and transfer learning

Mariano Cabezas¹, Sergi Valverde¹, Sandra González-Villà¹, Albert Clérigues¹,
Mostafa Salem¹, Kaisar Kushibar¹, Jose Bernal¹, Arnau Oliver¹, Joaquim
Salvi¹, and Xavier Lladó¹

Research Institute of Computer Vision and Robotics, University of Girona, Spain
mcabezas@eia.udg.edu

Abstract. Segmenting tumors and their subregions is a challenging task as demonstrated by the annual BraTS challenge. Moreover, predicting the survival of the patient using mainly imaging features, while being a desirable outcome to evaluate the treatment of the patient, it is also a difficult task. In this paper, we present a cascaded pipeline to segment the tumor and its subregions and then we use these results and other clinical features together with image features coming from a pretrained VGG-16 network to predict the survival of the patient. Preliminary results with the training and validation dataset show a promising start in terms of segmentation, while the prediction values could be improved with further testing on the feature extraction part of the network.

Keywords: convolutional neural networks · transfer learning · ensemble · pretrained · segmentation · prediction

1 Introduction

Gliomas are the most common primary brain malignancies, with different degrees of aggressiveness, variable prognosis and various heterogeneous histological sub-regions, i.e. peritumoral edema, necrotic core, enhancing and non-enhancing tumor core. This intrinsic heterogeneity of gliomas is also portrayed in their imaging phenotype (appearance and shape), as their sub-regions are described by varying intensity profiles disseminated across multimodal MRI scans, reflecting varying tumor biological properties. Due to this highly heterogeneous appearance and shape, segmentation of brain tumors in multimodal MRI scans is one of the most challenging tasks in medical image analysis.

The BraTS challenge [3, 2, 4, 6] started in 2012 and since then, it has always been focusing on the evaluation of state-of-the-art methods for the segmentation of brain tumors in multimodal magnetic resonance imaging (MRI) scans. The current iteration, BraTS 2018 utilizes multi-institutional pre-operative MRI scans and focuses on the segmentation of intrinsically heterogeneous (in appearance, shape, and histology) brain tumors, namely gliomas. Furthermore, to pinpoint the clinical relevance of this segmentation task, BraTS18 also focuses on

the prediction of patient overall survival using radiomic features and automatic machine learning algorithms.

On BraTS 2017, Kamnitsas et al. [5] obtained the best segmentation results with significant differences with respect to the other challengers. This approach introduced the use of an ensemble classifier composed of different deep convolutional neural networks (CNN) architectures. Moreover, in terms of survival prediction the best approach was presented by Shboul et al. [8] using a random forest approach after a previous feature extraction step where image features were extracted from a segmented image. In this paper we present a new approach based on the ideas of these two approaches where we first presegment the tumor region using a 3D unet [7] and then we refine this segmentation with a cascade approach using an ensemble of 4 different CNN architectures. Finally, using the VGG-16 [9] network pretrained on the imagenet dataset, we extract features on 20 tumor slices and combine them with the clinical data to obtain a survival prediction. The rest of the paper is structured as follows: in section 2 we present our approach for each task, followed by the preliminary results on the training and validation dataset and a discussion in section 3. Finally, our conclusions are presented in section 4.

2 Methods

2.1 Task 1: Segmentation of gliomas in pre-operative MRI scans

Following our approach for multiple sclerosis segmentation by Valverder et al. [10] and last's year's submission on the BraTS challenge by Wang et al. [1] we decided to implement a cascaded approach to deal with the segmentation task. However, unlike these two approaches, we used two different networks for each step.

Fully convolutional networks, and specifically unet ones, have a high accuracy when segmenting lesions with a small amount of data due to the capabilities of using large blocks of input data to train each convolutional kernel. However, they usually have poor results when trying to segment small subregions from a given mask. Therefore, we used a unet network to first delineate the tumor ROI mask.

Tumor delineation This network (see figure 1) uses 5 levels of convolutional and deconvolutional layers of 32 filters with a kernel size of $3 \times 3 \times 3$. Residual connections were also used to improve gradient decay, and the final output is given by a convolutional layer with kernel size of $1 \times 1 \times 1$ followed by a softmax activation. As input, patches of size $21 \times 21 \times 21$ were used for training, while the whole image was used for testing, since the network is fully convolutional and it speeds up the process.

Subregion segmentation With this initial ROI segmentation, we then apply a second network focused on segmenting the tumor subregions. Following Kamnitsas et al. [5], we implement a small ensemble net. However, instead of using

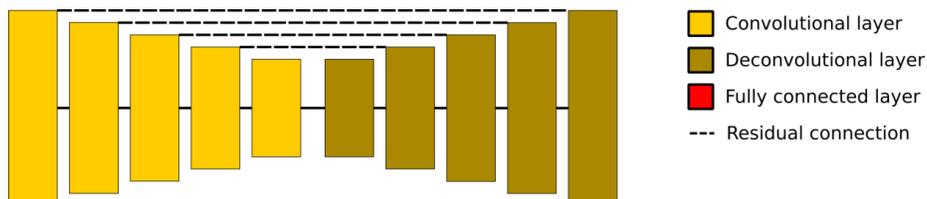


Fig. 1. 3D unet architecture for segmentation. This network uses only convolutional and deconvolutional layers of 32 filters with a kernel size of $3 \times 3 \times 3$. It also includes residual connections between convolutional and deconvolutional layers.

previous architectures, we propose a framework where each small subnet shares some metaparameters, while having different overall architectures and weights. The goal is to capture each network’s bias but keeping the same input information. The architectures include a unet with a dense output (UCNN), a unet with a fully convolutional output (UNET) and a dense (CNN) and fully convolutional networks (FCNN) using only convolutions (as illustrated by figure 2).

These networks were trained independently using the same input patches of size $13 \times 13 \times 13$ to guarantee that the final convolutional layer of both the CNN and FCNN networks were of size $3 \times 3 \times 3$ (to reduce the number of parameters of the final dense layer of the CNN network). Finally, a dense layer was trained with the output of the previous networks to give the ensemble results instead of using the average of all the networks. This training was performed on two steps, the first one for the independent networks (to avoid expanding their biases into each other) and then a second one for the last dense layer with the previous networks frozen.

2.2 Task 2: Prediction of patient overall survival (OS) from pre-operative scans

After segmenting the tumor subsections, we use their volumes as features together with the clinical data (age) as features for the survival task. Moreover, we decided to also include image features based on the surrounding area of the tumor to improve regression.

Due to promising results on transfer learning tasks using pretrained networks on natural images, we decided to use the VGG-16 network pretrained with the ImageNet dataset to compute features for 20 slices around the center of the tumor. These features are then passed through a fully convolutional layer (that shares weights among all the slices) of 156 units. Finally, the output of this layer is then combined with the clinical and volume features to obtain a final survival prediction as illustrated by figure 3.

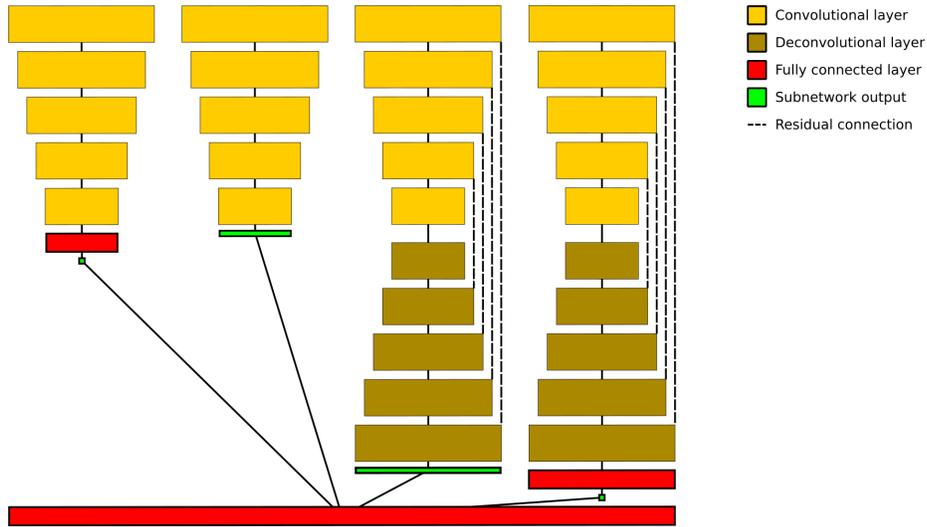


Fig. 2. Ensemble network for segmentation. The different networks use the same kind of convolutional and deconvolutional layers (represented in light yellow and dark yellow respectively) with 32 filters and kernels of $3 \times 3 \times 3$ and the fully connected layers all have 4 units (for the 3 tumor subregions + the background labels).

2.3 Implementation details

All the work was developed using the Keras library and python 2.7. Moreover, it was tested on a NVIDIA GTX Titan Xp and a Titan X PASCAL gpu with 12GB of RAM and a total system RAM of 256GB. Finally, all the code is publicly available at: <https://github.com/marianocabezas/challenges2018>.

3 Results and discussion

3.1 Task 1: Segmentation of gliomas in pre-operative MRI scans

In order to only evaluate the final results and avoid sending multiple submissions to the website, we only present the quantitative results of segmentation for the final ensemble in table 1.

Looking at the table, the results seem to be consistent between both datasets, even though the validation set provides slightly better results. This might be related to the fact that the number of cases (and most likely their variability) is lower in the validation dataset due to a lower number of cases.

Looking at the differences between the mean results and the median ones, we observe how the median is higher for both datasets. That also suggests, that while the overall performance is good, there are some outlier cases where the performance is worsened. For instance, this is the case for patient Brats18_CBICA_AOH_1

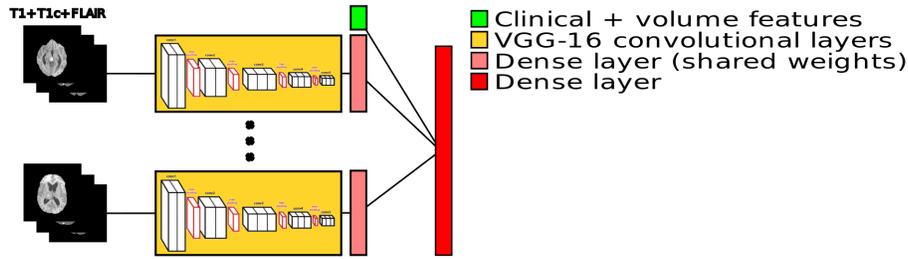


Fig. 3. Survival architecture composed by a pretrained VGG-16 network, a dense layer with weight shared for each slice and a final layer that takes also into account clinical and volumetric features.

Table 1. Summary of the results obtained from the IPP evaluation website for the segmentation task.

Metric	Training		Validation	
	Mean ($\pm\sigma$)	Median	Mean ($\pm\sigma$)	Median
Dice ET	0.66721 (0.29115)	0.78909	0.74034 (0.27735)	0.8339
Dice WT	0.84913 (0.13001)	0.89387	0.88928 (0.07497)	0.91197
Dice TC	0.71729 (0.23442)	0.80191	0.72644 (0.24267)	0.80026
Sensitivity ET	0.7343 (0.25561)	0.80397	0.76934 (0.26688)	0.86809
Sensitivity WT	0.80368 (0.17829)	0.8604	0.88843 (0.11746)	0.94084
Sensitivity TC	0.70196 (0.26795)	0.79212	0.76289 (0.2573)	0.87682
Specificity ET	0.99762 (0.00432)	0.99877	0.99803 (0.00285)	0.99886
Specificity WT	0.99649 (0.00416)	0.99793	0.99483 (0.00381)	0.99606
Specificity TC	0.99613 (0.00677)	0.99844	0.99529 (0.00607)	0.99735
Hausdorff95 ET	7.59116 (10.95449)	2.65791	5.3035 (9.96395)	2.23607
Hausdorff95 WT	7.76567 (10.008)	4.89898	6.95631 (11.9391)	3.31662
Hausdorff95 TC	9.82279 (8.50929)	8.06226	11.92386 (13.44799)	8

from the training dataset. For this patient, the DSC values for enhancing tumor (ET), whole tumor (WT) and tumor core (TC) are 0, 0.43607 and 0 respectively. In fact, for some cases of the training and validation dataset DSC value of 0 was obtained for the enhancing tumor region, proving to be one of the most difficult ones.

In order to show the improvement of the cascaded approach, some qualitative examples are presented in figure 4. As observed in this figure, the overall ROI of the tumor is similar between both networks. However, while the unet failed to properly differentiate between tumor subregions, the ensemble (which has two subnetworks dedicated to classify the central voxel of a patch) is capable of better delineate these subregions. Moreover, we can clearly see how the unet undersegmented the enhancing tumor area. This strengthens the notion that this is the most challenging tumor region to segment.

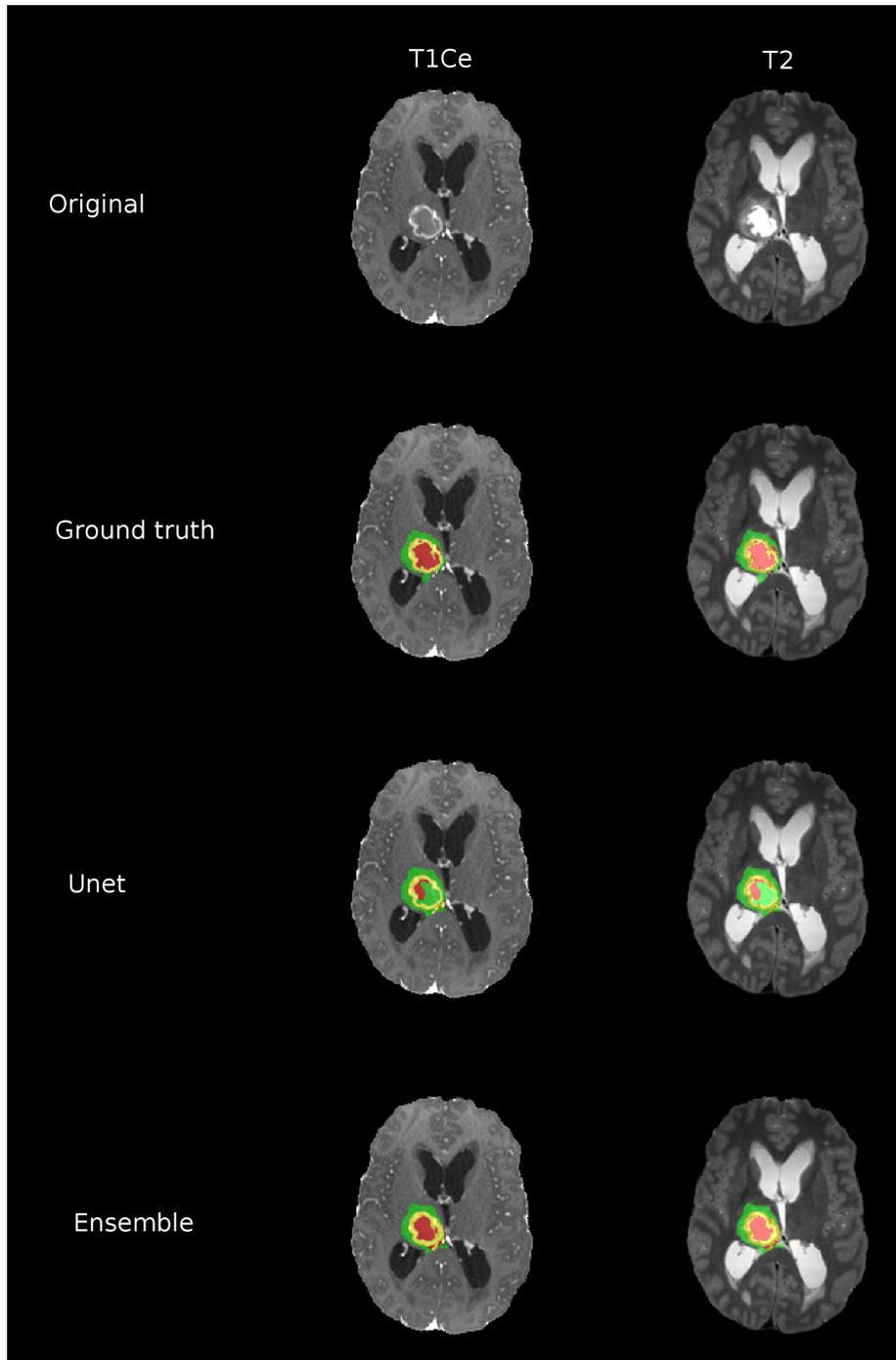


Fig. 4. Qualitative example of the segmentation of one slice. While the unet provides a good ROI segmentation, the subregions are not well delineated, while the ensemble improves this segmentation.

Table 2. Summary of the results obtained from the IPP evaluation website for the survival task.

Dataset	Cases	Accuracy	Mean SE	Median SE	Std. SE	Spearman R
Training	59	0.373	181387.051	64498.932	384906.463	-0.05
Validation	28	0.321	199081.864	73605.696	341473.991	-0.011

3.2 Task 2: Prediction of patient overall survival (OS) from pre-operative scans

The results obtained for the survival task are presented on table 2. As observed in the table, the results are fairly low. In fact the accuracy is below 0.5 for both validation and training. However, there are a few issues to take into account.

First, the number of evaluated cases is fairly lower when compared to the segmentation task. In fact, when training, only 164 cases had any survival information, while only 1/3 of these cases were taken into account for the evaluation of this task. It is well-known that deep learning strategies can fail to capture enough information in the presence of a small training dataset causing issues with generalisation. In fact, we observed how the survival prediction was usually centered in the range of 200-400, while the training dataset has a range of 5-1767. Our network could not generalise using cross-validation even though we had a low number of parameters.

Second, normalising the clinical features and the subregion volumes is not a trivial task. While using the zero mean approach, the network gave negative values as output, which have no meaning for this problem. Trying to normalise between 0 and 1 did not seem to have much of an effect. Probably, because the VGG network provides a large number of features and the network (again due to the low number of training samples) was not capable to compensate.

4 Conclusions

In this work, we have presented a combination of CNN architectures to first segment the tumor and its subregions and then provide a survival prediction estimate using this segmentation, clinical data and image information from FLAIR, T1 and T1 post-contrast. The results in terms of segmentation were promising, while there is room for improvement in terms of survival prediction. While using transfer learning seems to improve results over using only volume estimates, tweaking the features that are being used and how transfer learning is applied (maybe using new fully connected layers instead of the original VGG input) should improve results. We plan to further tweak our network to improve the results for the testing phase.

Acknowledgments

Mariano Cabezas holds a Juan de la Cierva - Incorporación grant from the Spanish Government with reference number IJCI-2016-29240. Jose Bernal and Kaiser

Kushibar hold FI-DGR2017 grants from the Catalan Government with reference numbers 2017FI B00476 and 2017FI B00372 respectively. This work has been partially supported by La Fundació la Marató de TV3, by Retos de Investigación TIN2014-55710-R, TIN2015-73563-JIN and DPI2017-86696-R from the Ministerio de Ciencia y Tecnología. The authors gratefully acknowledge the support of the NVIDIA Corporation with their donation of the TITAN-X PASCAL and TITAN-Xp GPU used in this research. The authors would like to thank the organisers of the Brats2018 challenge for providing the data.

References

- 1.
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection (2017). <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection (2017). <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Nature Scientific Data* **4**:170117 (2017). <https://doi.org/10.1038/sdata.2017.117>
5. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., Glocker, B.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. pp. 450 – 462 (2017)
6. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, ., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imag.* **34**(10) (2015). <https://doi.org/10.1109/TMI.2014.2377694>
7. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. pp. 234 – 241 (2015)
8. Shboul, Z.A., Vidyaratne, L., Alam, M., Iftekharuddin, K.M.: Glioblastoma and survival prediction. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.)

- Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 358 – 368 (2017)
9. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR **abs/1409.1556** (2014), <http://arxiv.org/abs/1409.1556>
 10. Valverde, S., Cabezas, M., Roura, E., González-Vilà, S., Pareto, D., Vilanova, J.C., Ramió-Torrentà, L., Rovira, A., Oliver, A., Lladó, X.: Improving automated multiple sclerosis lesion segmentation with a cascaded 3d convolutional neural network approach. NeuroImage **155**, 159 – 168 (2017)

Automatic Brain Tumor Segmentation Using a U-net Neural Network

Eric Caver, Chang, Liu, Weiwei Zong, Zhenzhen Dai, Ning Wen

Henry Ford Health System ,Detroit, MI 48202
nwen1@hfhs.org

Abstract. This study was designed to evaluate the ability of a u-net neural network to properly identify three regions of a brain tumor and evaluate a CNN to predict patient survivability. 210 patients were used for training while 66 patients were used for validation. There were four MRI modalities (T2WI, T1WI, T1CE, and FLAIR) for each patient and for each one, 1-4 physicians delineated the whole tumor, enhancing tumor, and tumor core regions. Multiple preprocessing steps were performed on each patient's data before loading them into the model. The segmentation model consists of three different u-nets, one for each region of interest. These created segmentations were then analyzed by use of quantitative metrics (dice similarity coefficient (DSC), Hausdorff distance (HD), sensitivity, and specificity) with respect to physician created contours. The average [s.d] DSC for whole tumor, enhanced tumor, and tumor core contours were 0.878 [0.123], 0.724 [0.284], and 0.760 [0.294], respectively. The average [s.d.] Hausdorff distance for whole tumor, enhanced tumor, and tumor core contours were 8.719 [18.196], 5.897 [10.413], and 10.089 [15.414], respectively. The average [s.d.] sensitivity for whole tumor, enhanced tumor, and tumor core contours were 0.888 [0.149], 0.767 [0.267], and 0.750 [0.323], respectively. The average [s.d.] specificity for whole tumor, enhanced tumor, and tumor core contours were 0.993 [0.005], 0.998 [0.003], 0.998 [0.002], respectively. This model was most effective in determining accurate location of the whole tumor as for successful patient results. These results show that the model did reasonably well defining the enhancing tumor and the tumor core. The models used for this study can be found at "<https://bitbucket.org/krogothliu/brats2018.git>". We approached overall survival prediction as a regression problem. Each 240X240 2D slice that displayed a lesion was represented by a CNN with 3 convolutional layers of random orthogonal kernels. Accuracy on training data is 0.864.

Keywords: Magnetic Resonance Imaging, Neural Network, U-NET.

1 Introduction

In 2017, about 23,000 adults were diagnosed with brain tumors. [3]. 20% of all primary malignant brain tumors are glioblastoma multiforme, which is the highest-grade form of astrocytoma. Astrocytoma has its name as it first appears in brain cells called astrocytes. These can be graded depending on growth and aggression. Low grade gliomas grow slowly, and are generally treated by surgery or radiation therapy, while high grade gliomas have extreme poor prognosis with a median survival of 12.6 months. [4] X-

rays, computed tomography (CT), and magnetic resonance imaging (MRI) images are important tools used by the medical community for cancer diagnosis, treatment and follow-up. MRI has been playing an important role in the neuroimaging due to its excellent soft tissue contrast and capability for functional imaging. [1] Conventional MR imaging techniques can give morphological information of the scanned area and they can generate images influenced by different types of tissue parameters including T1 weighted images (T1WI), T2 weighted images (T2WI) and fluid attenuation inversion recovery (FLAIR) Multi-parametric MRI allows for complementary images to come together to give the most accurate information for tumor delineation [2]. Post contrast T1 weighted (T1CE) images shows heterogeneous rim-enhancement in the tumor. FLAIR images suppress craniospinal fluids while enhance edema. In this study, MRI images were used to segment heterogeneous enhancing tumors, necrotic lesions and surrounding edema. It is imperative that the physician accurately identifies and localizes the brain tumor for diagnosis and optimal treatment planning. Over the last few years, deep neural networks used to aid in tumor delineation have evolved as more research breakthroughs have been accomplished. Our aims are to employ a deep neural network to do automatic image segmentation as well as predict life expectancy. Creating this program could potentially minimize the intra-observer variability in delineation as the uncertainties in programming are more readily understood and more predictable.

2 Methods

2.1 Patient Population

All data was downloaded from the BraTS multimodal Brain Tumor Segmentation Challenge 2018. A total of 210 patients were used during training, while 66 patients were used for validation. Each patient had four conventional MRI consisting of the following: T1-weighted, post contrast T1-weighted (T1CE), T2-weighted, and fluid attenuated inversion recovery (FLAIR). The MRI images came from 19 institutions and have different scanners and clinical protocols to ensure variety of data. One to four physicians segmented enhancing tumor, peritumoral edema, and necrotic (non-enhancing tumor core) on each MRI sequences for each patient. [5-8]

2.2 Image Preprocessing

All MRI imaging modalities (T1, T2, FLAIR) were rigidly registered with the T1 contrast enhanced image (T1CE) and resampled to a spatial resolution of $1 \times 1 \times 1\text{mm}^3$. In addition to the registration and resampling, skull stripping and normalization were performed. For each original 3D volume ($240 \times 240 \times 155$), we did the following preprocessing. To create a normal resolution, zero margins were added to pad each 2D image to 256×256 . To speed up training/validation/testing and eliminate non-enhancing and no surrounding edema region a patch was created. Our observation showed that all tumors were included in 64 or less adjacent slices, therefore the patch was created by cropping 64 slices containing the enhanced tumor and edema with a moderate number of margins. Considering the brain exhibits marked symmetry across the sagittal plane,

each slice was flipped left/right to decrease the dependence on location and increase the variability of the available dataset using data augmentation. In the validation phase, all 155 slices of each 3D volume went through the pipeline for segmentation. The final result contained only the largest connected component in each volume.

2.3 Neural Network Model (segmentation)

A U-net [20] was trained to learn the features of the segmented areas: whole tumor (WT), enhancing tumor (ET), and tumor core (TC) in the MRI images. The u-net used in this study can be downloaded at “<https://bitbucket.org/krogothliu/brats2018.git>”. All four modalities were used for segmenting WT, TC, and ET. This model was run three different times, each of which had one type of the reference contours as the input. A corresponding mask was generated as the output for each input structure. In this way, a whole tumor model, an enhancing tumor model, and a tumor center model were trained respectively. The whole tumor contour was generated by summing all three masks of each model. The tumor core contour was created by the combining the masks of the enhancing tumor and tumor core models. The enhancing tumor contour was the direct output of the enhancing tumor model. The difference between T1CE and T1WI was also taken into account. Four scaling layers were used following Pelt *et al.*'s [10] study as the accuracy was improved by using a U-net with 3 or 4 scaling layers. The number of features doubled for each consecutive convolutional layer and halved for each consecutive up-sampling layer as suggested by Ronneberger, *et al.* [11]. The merged layer was comprised of features from the up-sampling layers and the convolutional layers. The convolutional layers utilized batch normalization, which was placed right after every merged layer, as well as two more convolutional layers at the end of the model. [12] Soft dice was utilized as the loss function and is explained by formula 1.

$$\text{Dice Loss} = \frac{2 * \langle y_{\text{true}}, y_{\text{pred}} \rangle + c}{\langle y_{\text{true}}, y_{\text{true}} \rangle + \langle y_{\text{pred}}, y_{\text{pred}} \rangle + c}, \quad (1)$$

where y_{true} is the clinician's contour, y_{pred} is the model's output, and c (0.01) is a constant to avoid division-by-zero singularities. The best model for each type of contour was chosen according to the validation loss within 100 epochs run on GPU (Titan XP, nVidia, Santa Clara, CA).

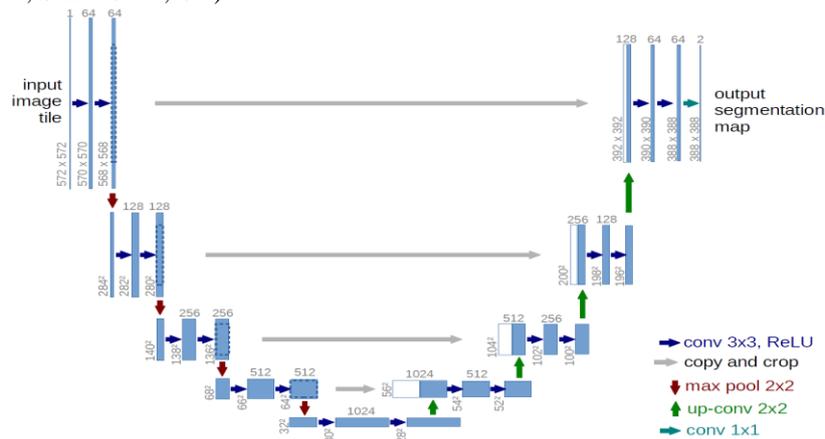


Fig. 1. U-net Model Architecture [11]

2.4 Statistics

The metrics used to evaluate the agreement between the created U-Net model and the given reference contours are the dice similarity coefficient (DSC), Hausdorff distance (HD), sensitivity, and specificity. Sensitivity is also known as a true positive. It is explained by equation 2. Specificity is also known as a true negative and explained by equation 3.

$$\text{Sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}} \quad (2)$$

$$\text{Specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}} \quad (3)$$

The DSC is explained in equation 4 and gives statistic regarding general overlap:

$$\text{DSC} = \frac{2 \cdot \langle y_{true}, \bar{y}_{pred} \rangle}{\langle y_{true}, y_{true} \rangle + \langle \bar{y}_{pred}, \bar{y}_{pred} \rangle}, \quad (4)$$

where \bar{y}_{pred} is binary prediction using 0.5 as the threshold.

Hausdorff Distance gives statistics regarding the amount of gross error between the created contour and reference contour. The Hausdorff distance is the maximum distance of a point in one contour to the nearest point of the other contour:

$$h(A, B) = \max_{a \in A} \{ \min_{b \in B} \{ d(a, b) \} \} \quad (5)$$

where a and b are points of sets A and B , respectively, and $d(a, b)$ is Euclidean metric between these points [13].

2.5 Overall Survival Prediction

The overall survival prediction was addressed as a regression problem. A 2D CNN with three convolution layers of random orthogonal kernels was employed. [14] The 240×240 2D slices and ground truth was input into this created CNN. More recent work has suggested that by initializing with orthogonal weights or kernels, the generated space tends to display rich features which in turn causes better performance. [15]. Resulting representation vector, with dimensionality of $12 \times 12 \times 30$, together with patients' age were fed into a regressor called "extreme learning machine (ELM)" [16]. ELM was proposed in 2006 as replacement of tedious parameter tuning in traditional shallow neural networks, where parameters of hidden nodes were randomly generated and only output weight connecting hidden nodes and output nodes were analytically calculated.

3 Results

3.1 Statistics (All Patients)

The average [s.d] DSC for the whole tumor, enhanced tumor, and tumor core contours were 0.878 [0.123], 0.724 [0.284], and 0.760 [0.294], respectively. The average [s.d.] Hausdorff distance f were 8.719 [18.196], 5.897 [10.413], and 10.089 [15.414], respectively. The average [s.d.] sensitivity for the whole tumor, enhanced tumor, and tumor core contours were 0.888 [0.149], 0.767 [0.267], and 0.750 [0.323], respectively. The average [s.d.] specificity was 0.993 [0.005], 0.998 [0.003], 0.998 [0.002], respectively.

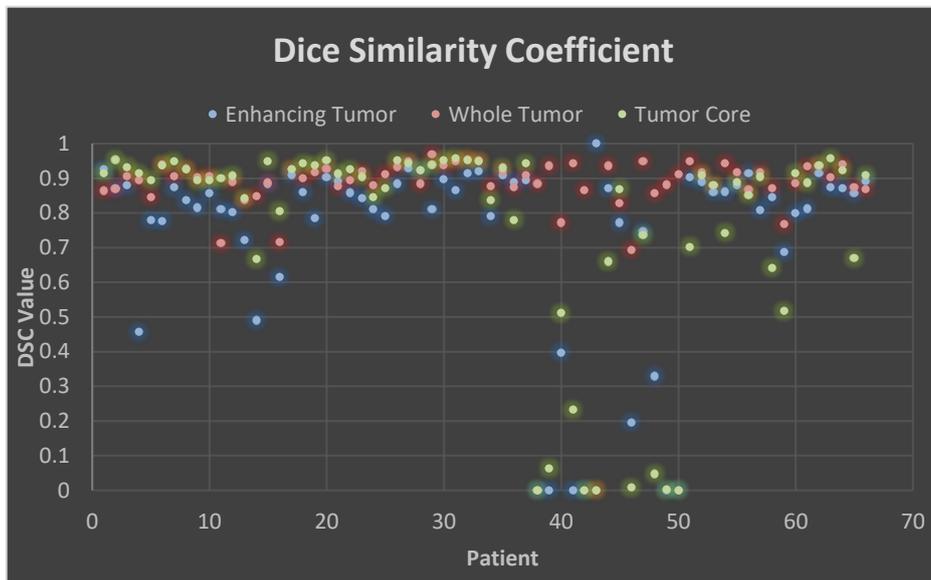


Fig. 2. DSC for whole tumor (WT), enhanced tumor (ET), and tumor core (TC) contours for each patient

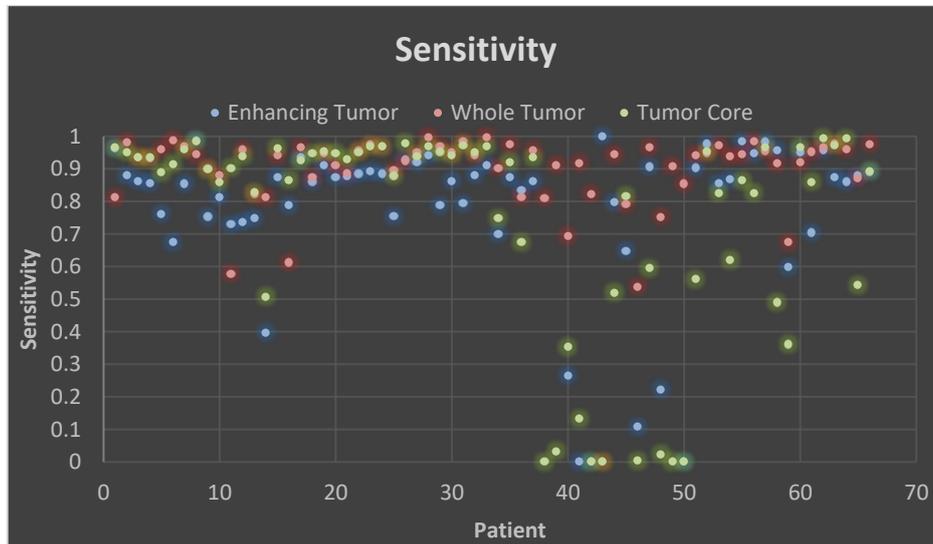


Fig. 3. Specificity for whole tumor (WT), enhanced tumor (ET), and tumor core (TC) contours or each patient

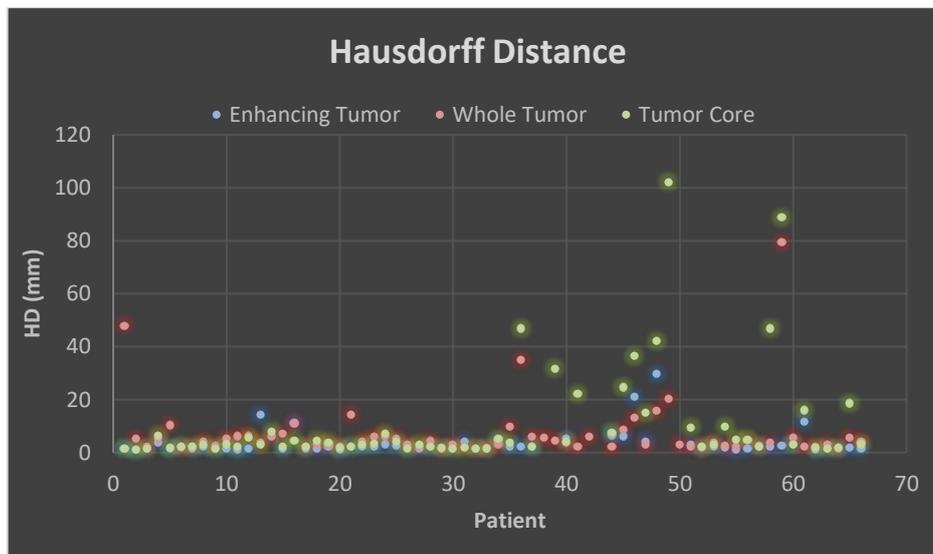


Fig. 4. Hausdorff Distance (mm) for whole tumor (WT), enhanced tumor (ET), and tumor core (TC) contours or each patient

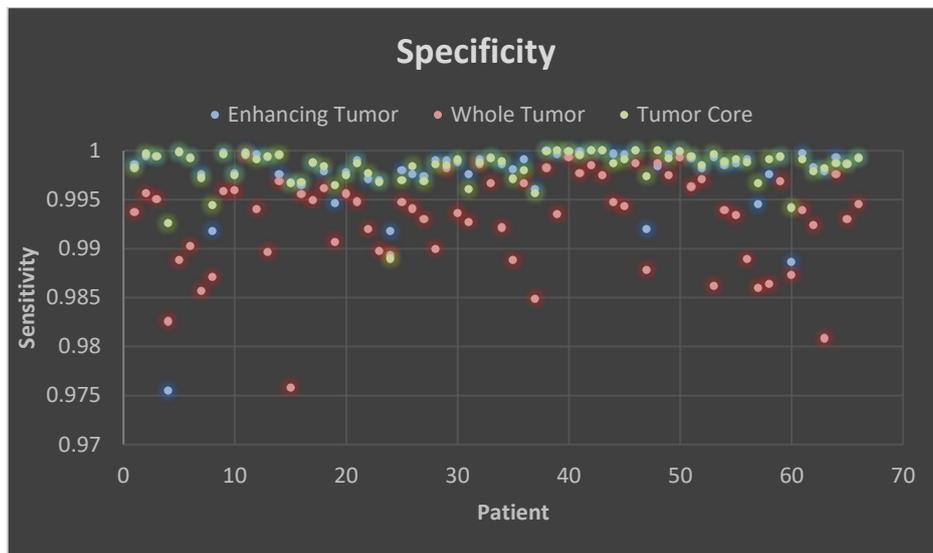


Fig. 5. Specificity for whole tumor (WT), enhanced tumor (ET), and tumor core (TC) contours or each patient

3.2 The validation of the Models

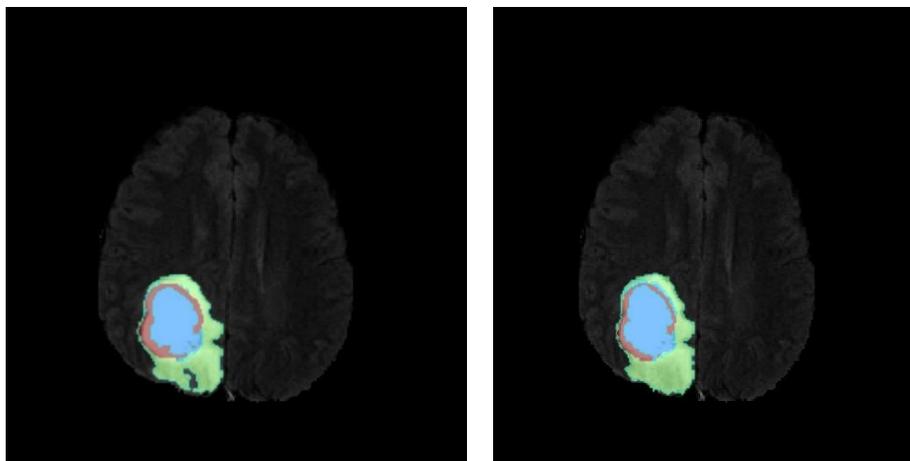


Fig. 6. One case from the validation dataset. The image on the left shows the reference contours, while the image on the right depicts the contours created by u-net. (Green=WT, Blue=TC, Red=ET)

The model had different levels of success depending on the tumor area it was locating. For the 66 patients in the validation dataset, the model correctly identified enhancing tumor location for 59 patients, tumor core location for 62 patients, and whole tumor

location for 65 patients. These results are shown in figure 7. A successful patient was defined as a patient where the DSC is greater than zero. In other words, the patient was deemed successful if the contour defined by the physician and the contour created by the model have any overlap.

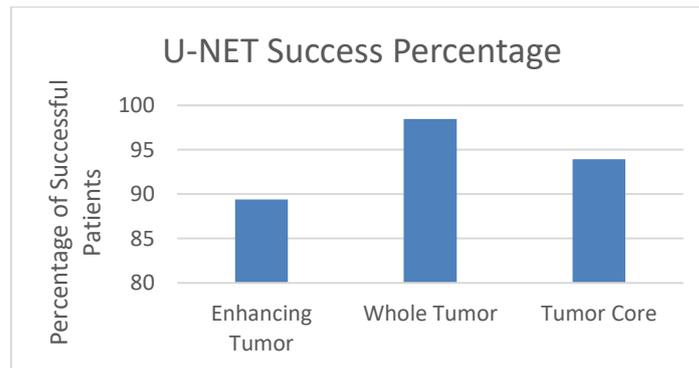


Fig. 7. Success of model on total validation patient population

The results can be further improved by only including the successful patients in the statistical analysis. These improved statistics for the successful patients are as follows. The average [s.d] value of whole tumor, enhanced tumor, and tumor core contours were 0.891 [0.056], 0.793 [0.171], and 0.809 [0.226], respectively for DSC, 6.992 [11.587], 4.121 [5.737], and 8.963 [14.664] respectively for Hausdorff distance, and 0.902 [0.099], 0.802 [0.208] and 0.798 [0.266], respectively for sensitivity. There was no change in specificity.

The 2D CNN was able to predict overall diagnosed patient survival. The training accuracy on training data is 0.864.

4 Discussion

There have been significant research breakthroughs in artificial intelligence and deep learning fields in recent years, which has the potential to transform the healthcare as well. The deep architecture of the deep neural network makes it powerful on the image analysis and processing. We have presented a deep U-Net architecture to correctly detect and segment the following areas of brain lesions: enhancing tumor, tumor core, and whole tumor. The results from our models were encouraging and can be viewed in Table 1.

		Successful Pt.	All Pt.	Change
DSC	ET	0.81±0.15	0.74±0.28	9.6%
	WT	0.89±0.06	0.88±0.12	1.5%
	TC	0.80±0.24	0.76±0.31	6.5%
Sens.	ET	0.81±0.18	0.78±0.25	4.6%
	WT	0.90±0.10	0.89±0.15	1.5%
	TC	0.79±0.27	0.74±0.33	6.5%
Spec.	ET	1.00±0.00	1.00±0.00	0.0%
	WT	0.99±0.00	0.99±0.00	0.0%
	TC	1.00±0.00	1.00±0.00	0.0%
HD (mm)	ET	3.53±4.87	6.47±16.34	-45.5%
	WT	7.05±11.64	8.77±18.22	-19.7%
	TC	10.57±19.07	12.99±23.53	-18.6%

Table 1. Statistical Results

The model was most effective in determining accurate location of the whole tumor with a DSC value of 0.891 ± 0.056 . The model did reasonably well in the segmentation of the enhancing tumor and the tumor core. Table 2 shows our results compared with the average DSC results according to the BraTS validation phase leaderboard 2018. From these preliminary results it appears that u-net has an above average potential for utilization in a clinical setting. However, there is still room for improvement as this study does not have the best reported results.

DSC	ET	WT	TC
U-Net Neural Network	0.74±0.28	0.88±0.12	0.76±0.31
Ave. Reported Results	0.64±0.15	0.82±0.06	0.73±0.24
% Difference	14.3%	7.2%	4.1%

Table 2. Results of this study compared with the preliminary reported results from BraTS 2018

One such improvement is to change the ET, WT, and TC contour definitions in the model. Currently, the whole tumor contour created by this model is defined as the sum of three masks created by the three neural networks; the enhancing tumor contour created is the sum of the tumor core mask and the enhancing tumor mask, and the tumor core contour is created by the tumor core mask. In an effort to improve the results, the definition of each created contour was modified. The purpose of this modification was to exclude any pixels in the contour that was outside of that specific mask. For example, the whole tumor contour was originally created from the sum of the three masks, but

now if any of the other masks were outside of the whole tumor mask, that mask's input in that area would be disregarded. However, the impact of this modification was moderate.

In this u-net, the input of each model consisted one type of structure contours along with all four MRI modalities. Further investigation is warranted to derive the optimal combination and weighting factor of each sequence. For example, FLAIR is best modality to segment the edema, but there was not enough data present for the model to derive this. Ideally, the weight values could be input as well to force the model to use the proper MRI modality for finding each structure. This could potentially aid the model as it must overcome a smaller training dataset.

Overall patient survivability training results were encouraging. However, there was a problem generating a ground truth mask for the validation data, which led to problems assessing the validation data.

5 Conclusion

U-net, a deep neural network, was used to automatically segment the brain into whole tumor, tumor center, and enhancing tumor. These created segmentations were then analyzed by use of quantitative metrics (dice similarity coefficient (DSC), Hausdorff distance (HD), sensitivity, and specificity) with respect to physician created contours. The results of these statistics were encouraging, especially for finding the whole tumor. Future study is still needed to investigate the impact of different scanners, protocols, sequences, and sizes of dataset on the performance of segmentation. Predicting overall patient survivability was successfully done when dividing the training dataset, but more work needs to be done to be able to use it in a clinical setting.

References

1. Magnetic Resonance Imaging (MRI), <https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri>, last accessed 2018/7/9.
2. Using Multiparametric MRI to detect Prostate Cancer, <https://prostate.net/articles/using-multiparametric-mri-detect-prostate-cancer/>, last accessed 2018/7/9.
3. Brain Tumor: Statistics, <https://www.cancer.net/cancer-types/brain-tumor/statistics>, last accessed 2018/7/9.
4. Brain Cancer and Gliomas, <https://www.webmd.com/cancer/brain-cancer/malignant-gliomas#1>, last accessed 2018/7/9.
5. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp I, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekeley G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation

- Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015)
DOI: 10.1109/TMI.2014.2377694
6. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
 7. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
 8. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
 9. F. Milletari, N. Navab, S.A. Ahmadi, V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), 2016, pp. 565-571.
 10. D.M. Pelt, J.A. Sethian, A mixed-scale dense convolutional neural network for image analysis, Proceedings of the National Academy of Sciences, 115 (2018) 254.
 11. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in: arXiv:1505.04597, 2015.
 12. F. Chollet, keras, in, GitHub, GitHub repository, 2015.
 13. A. Kumarasiri, F. Siddiqui, C. Liu, R. Yechieli, M. Shah, D. Pradhan, H. Zhong, I.J. Chetty, J. Kim, Deformable image registration based automatic CT-to-CT contour propagation for head and neck adaptive radiotherapy in the routine clinical setting, Medical physics, 41 (2014) 121712.
 14. Kasun, L.L.C., Zhou, H., Huang, G.B. and Vong, C.M., 2013. Representational learning with extreme learning machine for big data. IEEE intelligent systems, 28(6), pp.31-34.
 15. Aghajanyan, A., 2017. Convolution Aware Initialization. arXiv preprint arXiv:1702.06295.
 16. Huang, G.B., Zhu, Q.Y. and Siew, C.K., 2006. Extreme learning machine: theory and applications. Neurocomputing, 70(1-3), pp.489-501.

Context Aware 3-D Residual Networks for Brain Tumor Segmentation

Siddhartha Chandra¹, Maria Vakalopoulou^{1,2}, Lucas Fidon¹, Enzo Battistella^{1,2}, Theo Estienne^{1,2}, Roger Sun^{1,2}, Charlotte Robert², Eric Deutch², and Nikos Paragios^{1,3}

¹ CVN, CentraleSuplec, Universit Paris-Saclay, France

² Gustave Roussy Institute, Paris, France

³ TheraPanacea, Paris, France

Abstract. In this work we propose a novel deep architecture for the task of brain tumor segmentation. More specifically, we extend the 3-D ResNet-18 residual network towards the aim of inducing better context-awareness. To this end, we exploit an array of 3-D ResNet-18 networks for feature learning, one for each of the axial, sagittal and coronal views, and fuse the per-view features together to then train per-pixel linear classifiers for the whole tumor, tumor core and enhancing tumor categories. We reduce the network stride of the 3-D ResNet-18 network from 32 to 4 by introducing dilated convolutions using the ‘atrous’ pooling algorithm. Further, to enhance the context-awareness of the network, we introduce within the network a multi-scale feature pyramid via atrous spatial pyramid pooling. Finally, we use a fully-connected Conditional Random Field (CRF) to refine our predictions in a post processing step. The fully-connected CRF performs a bilateral filtering of the predictions delivered by our network. We report competitive results on the Brats 2018 validation set.

Keywords: Brain Tumor Segmentation · Residual Networks · Fully-Connected CRFs.

1 Introduction

Cancer is currently one of the leading causes of death worldwide with overall 14.1 million new cases and 8.2 million deaths in 2012 [10]. Brain tumors, with gliomas being one of the most frequent malignant types, are among the most aggressive and dangerous types of cancer [4]. According to recent classifications malignant gliomas are classified into four WHO grades. From these low grade gliomas (LGG), including grade I and II are considered as relatively slow-growing while high grade gliomas (HGG), including grade III and grade IV glioblastoma are more aggressive with the average survival time of approximately 1 year for patients with glioblastoma (GBM) [11, 19]. Besides being very aggressive, gliomas are very costly to treat, making their accurate diagnosis from the early stages, very important.

Multimodality magnetic resonance imaging (MRI) is the primary method of screening and diagnosis for gliomas. However, due to inconsistency and diversity of MRI acquisition parameters and sequences, there are large difference in appearance, shape and intensity ranges, adding variability to the one that gliomas can have between different patients. Currently, tumor regions are segmented manually by radiologists, but due to all these variations, the process is very time consuming while the inter-observer agreement between them is considerably low. In order to address all these challenges, the multimodal brain tumor segmentation challenge (BraTS) [20, 1–3] is organized annually, in order to highlight efficient approaches and the way forward for the accurate segmentation of brain tumors.

Currently, the emergence of deep learning as disruptive innovation method in the field of computer vision has impacted significantly the medical imaging community, with numerous architectures being proposed addressing task-specific problems. Among them the Fully Convolutional Networks (FCN) [18] and their extension to 3D [21, 12] are among the most commonly used architectures, boosting considerably the accuracies of the semantic segmentation. Inspired by these recent advances of deep learning, in these paper we adapt 3D CNNs with context aware, coupled with fully-connected Conditional Random Fields for the accurate segmentation of brain tumor.

In the next section, we discuss our contributions in detail, and we report our performance on the Training and Validation Dataset of BraTS 2018.

2 Context-Aware 3-D Residual Networks

We now describe in detail our deep-learning pipeline. Our pipeline in this work is based on the deep residual network architecture introduced by He in [14]. Residual networks (ResNets) ease the training of networks by adding ‘residual’ connections to the network architecture. These residual connections induce a short-cut connection of identity mapping without adding any extra parameters or computational complexity, thereby recasting the original mapping $F(x)$ as $F(x) = F'(x) + x$. We encourage the readers to refer to the original paper [14] for details.

ResNets are the building blocks of the majority of state of the art approaches on a variety of computer vision image segmentation benchmarks [9, 25, 24, 12], and thus are a natural starting point in this work. However, the vanilla ResNets lack certain desirable characteristics which make their application to the task of brain tumor segmentation challenging. In this work, we address these challenges by extending the 3-D ResNet architecture from [12]. We now discuss these challenges one by one and describe our strategies to overcome them.

2.1 Network Stride.

Approaches to semantic segmentation use ‘fully-convolutional networks’ (FCNs) [18, 17] which are networks composed entirely of stacks of convolution operations,

thereby producing per-patch outputs which spatially correspond to patches in the input image. A major challenge that presents itself in the use of FCNs is the *network stride*, also referred to as the *downsampling factor*. The output activations delivered by FCNs are smaller in spatial size than the input image due to the repeated max-pooling and convolutional strides. Thus, obtaining a labeling that is the same size as the input image requires upsampling of the output scores via interpolation, resulting in quantization and approximation errors and over-smooth predictions which do not capture the finer details in the input.

The downsampling factor of ResNets, like other popular network architectures such as [16, 22] is 32. This means that each output unit corresponds to a 32×32 patch in the input image. For the Brats 2018 data where the size of the input volume is $240 \times 240 \times 155$, the vanilla ResNet delivers outputs of the size $8 \times 8 \times 5$. A popular approach to reduce the downsampling factor is using a deconvolution filter which is a backwards convolution operation to upsample the output, as proposed by Long and Shelhamer in [18]. However, this results in an increased number of parameters, which will lead to overfitting for smaller datasets like the Brats 2018 dataset where obtaining pixel-accurate ground truth is tedious.

In this work, we use atrous convolutions proposed by Chen *et al.*. The atrous algorithm introduces holes in the convolution kernel, thereby allowing us to reduce the loss in spatial resolution without any increase in the number of parameters. Authors in [23] use the same operation, rebranding it as ‘dilated convolutions’. With a strategic use of atrous convolutions, we reduce the downsampling factor of ResNets to 4. This amounts to an output of size $60 \times 60 \times 39$ for the Brats 2018 data.

2.2 Context Awareness.

Standard deep networks do not have a built-in capacity to estimate the scale of the input [6]. This limitation becomes especially crippling for brain tumor segmentation where the scales of the whole tumor, tumor core and enhancing tumor categories depend on a variety of factors, therefore estimating the correct scale of tumors is a challenging task. State of the art approaches typically address this shortcoming by feeding the input to the network at different scales and averaging the network responses across scales [6, 7]. A number of recent methods have proposed using feature pyramids [25, 8, 5] which instead capture features at multiple scales. The feature pyramids are finally fused into a single feature map via element-wise maximization, averaging or concatenation. In this work, we use the atrous spatial pyramid pooling (ASPP) approach proposed in [8]. ASPP uses a stack of convolutional filters with increasing degrees of dilation, thereby simulating filtering at multiple sampling rates and receptive fields. This captures visual context at multiple scales and leads to performance boosts for a variety of segmentation benchmarks [5]. The features at different scales are fused via averaging. This strategy enhances the context-awareness of the network.

2.3 Richness of Features (Network depth) vs Training/Inference speed.

Deeper networks typically learn richer, more meaningful features as indicated by performance boosts over shallower networks [14, 13]. However, an increase in depth also increases training / inference time because the network represents a sequential directed acyclic graph and prior activations need to be computed before subsequent ones.

3-D FCNs are much slower than their 2-D counterparts. Unlike 2-D convolutions which have benefitted from both software and hardware level optimizations, 3-D convolutions still involve slow computations as the research into their optimization is in its infancy. To allow fast experimentation and validation, the network architecture design needs careful consideration.

The authors in [24] demonstrate that decreasing the depth and increasing the width of ResNets leads to both better accuracy and reduced training / testing time. Inspired by them, rather than using very deep ResNets, we use the smallest residual network ResNet-18 in our experiments. To increase the width of the network, we use a multi-view fusion architecture where our network has three branches, each computing features on one of axial, sagittal and coronal views. The features from the three views are transposed to a common view and concatenated, and linear classifiers for the three categories whole tumor, tumor core and enhancing tumor are trained on the fused features. This increases the speed at which the network operates as the activations of the three branches of the network can be computed in parallel. Further our preliminary experiments indicate that this multi-view fusion leads to better performance on a validation set, compared to deeper variants: ResNet-34 and ResNet-50. Our approach is described in Fig. 1.

2.4 Fully-Connected Conditional Random Fields.

As described in Sec. 2.1, the downsampling factor of our network is 4. This amounts to an output of size $60 \times 60 \times 39$ for an input of size $240 \times 240 \times 155$. To obtain predictions which have the same spatial resolution as the input volume, we use bilinear interpolation. This introduces quantization and approximation errors, and oversmooth predictions which do not capture the finer details in the input data. To address this limitation, we follow up the first pass of prediction using the network with a post-processing refinement pass. The refinement of the network prediction is done using a fully-connected Conditional Random Field (CRF). The fully-connected CRF performs bilateral filtering to refine the predictions made by our network, and uses the objective function proposed in [15]. Precisely, the CRF expresses the energy of a fully-connected CRF model as the sum of unary and pairwise potentials given by

$$E_I(\mathbf{l}) = \sum_i \psi_u(l_i) + \sum_i \sum_{j < i} \psi_p(l_i, l_j), \quad (1)$$

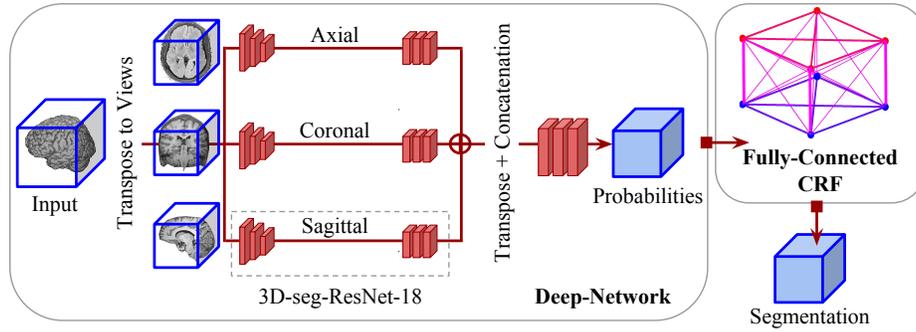


Fig. 1. A schematic overview of our approach. Our network consists of three parallel ResNet-18 branches, each computing mid-level features on one of the axial, sagittal and coronal views of the input. These mid-level features are fused by transposing to a common view and concatenation. Linear classifiers on top are trained to produce probabilities for each of whole tumor, tumor core and enhancing tumor categories. Further, we use a bilateral filtering performing fully-connected CRF to post-process our network predictions.

where

$$\psi_p(l_i, l_j) = \mu(l_i, l_j) \sum_{m=1}^K w_m^1 \exp\left(-\underbrace{\frac{|s_i - s_j|^2}{2\theta_\alpha^2} - \frac{|p_i - p_j|^2}{2\theta_\beta^2}}_{\text{appearance}}\right) + w_m^2 \exp\left(-\underbrace{\frac{|s_i - s_j|^2}{2\theta_\gamma^2}}_{\text{smoothness}}\right). \quad (2)$$

Here $\mathbf{l} = \{l_i\}$ denotes the labels for all the pixels indexed by i coming from a set of candidate labels $l_i \in \{1, 2, \dots, L\}$. ψ_u denotes the image dependent unary potentials, and the image dependent pairwise potentials $\psi_p(l_i, l_j)$ are expressed by the product of a label compatibility function μ and a weighted sum over Gaussian kernels. The pixel intensities are expressed using the 4 modalities in the input data $p_i = (flair, t1, t2, t1ce)$ and spatial positions are simply the coordinates in 3-D space $s_i = (x, y, z)$. These are used together to define the appearance kernel, and the spatial positions alone are used to define the smoothness kernel. The appearance kernel tries to assign the same class labels to nearby pixels with similar colour, and the hyperparameters θ_α and θ_β control the degrees of nearness and similarity. The smoothness kernel aims to remove small isolated regions. The model parameters $(\theta_\alpha, \theta_\beta, \theta_\gamma, w_m^1, w_m^2)$ are set by doing parameter sweeps using a validation set.

Having discussed our method in detail, we now delve into the experimental details and results in the next section.

3 Experiments and Results

We now describe our experiments in detail, covering our training protocol and evaluation.

3.1 Training Protocol

As described in Sec. 2 and in Fig. 1, our network is a 3 branch multi-view 3-D ResNet-18. The three branches are initialized from the 3-D ResNet-18 network from [12] which was trained for action recognition in videos. We augment the first convolutional layer (`conv1`) of the network from [12] with an additional input channel since we have 4 modalities (*flair*, *t1*, *t2*, *t1ce*), as opposed to 3 channels in natural images ($\mathbf{r}, \mathbf{g}, \mathbf{b}$).

Our implementation is based on the *Caffe2* library. We train our networks with randomly sampled input patches of size $97 \times 97 \times 97$, and our network outputs predictions of size $25 \times 25 \times 25$. The input brain volume is preprocessed by subtracting the per-image mean for each modality independently. We use the weighted Softmax Cross-Entropy loss to train our network for three classes: whole tumor, tumor core, and enhancing tumor. The weights for these three classes are 5, 10, 10 respectively. We use random flipping across the axial axis, and random scaling of the input between scales 0.25 – 2.5 for data augmentation. We use the standard stochastic gradient descent algorithm for training, with a weight-decay of $1 \times e^{-5}$ and momentum of 0.9. We use a polynomially decaying learning rate policy, with a starting learning-rate of $1 \times e^{-4}$ and we train for $100 - K$ iterations.

For results on the training set, we use a random train-test split of 200 – 85 samples respectively. For results on the validation set, we use all the 285 training samples to train, and evaluate on the 66 validation samples.

Our CRF parameters (Sec. 2.4) are estimated using the validation set of 85 samples.

3.2 Testing Protocol

We use multi-scale testing alongside flipping along the axial plane and average the probabilities delivered by the network. We use these probabilities as unary terms along hand-crafted pairwise terms (Sec. 2.4) to the CRF for post-processing.

3.3 Results

Our results on the train-test split are tabulated in Tab. 1, and the results on the Brats 2018 Validation dataset are tabulated in Tab. 2. These results were generated by the evaluation server on the official Brats 2018 website. Qualitative results are shown in Fig. 2.

Dice			Sensitivity			Specificity			Hausdorff95		
ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
0.661	0.85466	0.79064	0.69755	0.9442	0.84053	0.9968	0.98139	0.99396	6.83216	11.68252	10.91145

Table 1. Results on BraTS Training dataset.

Dice			Sensitivity			Specificity			Hausdorff95		
ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
0.74057	0.87189	0.79898	0.79469	0.82997	0.78876	0.99696	0.99639	0.99731	5.57574	5.0379	9.58839

Table 2. Results on BraTS Validation dataset.

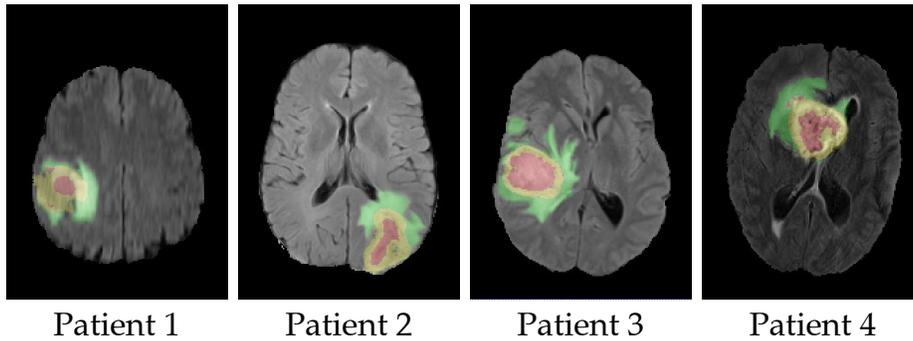


Fig. 2. Example segmentations on the Brats 2018 Validation set delivered by our approach for four patients. Green: edema, Red: non-enhancing tumor core; Yellow: enhancing tumor core.

4 Conclusions and Future Work

In this work, we have described a novel deep-learning architecture for automatic brain tumor segmentation. Our contributions address some of the limitations of standard deep networks. More specifically, we have extended the 3-D residual network to better capture the visual context in 3-D brain volumes. We devise strategies to enable the network deliver finer, more detailed output responses, incorporate multi-view fusion of predictions and capture multi-scale context via a feature-pyramid scheme. We also describe a fully-connected CRF to refine the network outputs in a post-processing step. Our approach delivers competitive results on the Brats 2018 dataset. In the future, we would like to incorporate bias correction, and more aggressive data augmentation strategies such as rotation and elastic deformation.

References

1. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos. Advancing the cancer genome atlas glioma mri

- collections with expert segmentation labels and radiomic features. *Scientific data*, 4, 9 2017.
2. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection, 07 2017.
 3. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection, 07 2017.
 4. E. C. Holland. Progenitor cells and glioma formation. 14:683–8, 01 2002.
 5. L. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587, 2017.
 6. L. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille. Attention to scale: Scale-aware semantic image segmentation. *CVPR*, 2016.
 7. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.
 8. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *arXiv:1606.00915*, 2016.
 9. L.-C. Chen, G. Papandreou, K. Murphy, and A. L. Yuille. Weakly- and semi-supervised learning of a deep convolutional network for semantic image segmentation. *ICCV*, 2015.
 10. J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. Parkin, D. Forman, and F. Bray. Cancer incidence and mortality worldwide, 2013.
 11. M. Hadziahmetovic, K. Shirai, and A. Chakravarti. Recent advancements in multimodality treatment of gliomas. *Future oncology*, 7 10:1169–83, 2011.
 12. K. Hara, H. Kataoka, and Y. Satoh. Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In *CVPR*, 2018.
 13. K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. *ICCV*, 2017.
 14. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016.
 15. P. Krähenbühl and V. Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *NIPS*, 2011.
 16. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.
 17. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, 1998.
 18. J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, pages 3431–3440, 2015.
 19. D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison. The 2016 world health organization classification of tumors of the central nervous system: a summary. *Acta Neuropathologica*, 131(6):803–820, Jun 2016.
 20. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M. A. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, G. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H. C. Shin, J. Shotton, C. A.

- Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. V. Leemput. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10), 2015.
21. F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 565–571. IEEE, 2016.
 22. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *ICLR*, 2015.
 23. F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. *ICLR*, 2016.
 24. S. Zagoruyko and N. Komodakis. Wide residual networks. In *BMVC*, 2016.
 25. H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. *CoRR*, abs/1612.01105, 2016.

Automatic segmentation of brain tumor from 3D MR images using a 2D convolutional neural network

Yi-Ju Chang, Zheng-Shen Lin, Tsai-Ling Yang, Teng-Yi Huang

Department of Electrical Engineering, National Taiwan University of Science and Technology,
Taipei, Taiwan
tyhuang@mail.ntust.edu.tw

Abstract. In the study, we used a two-dimensional convolutional neural network, to design an automatic segmentation of brain tumor from three-dimensional MR datasets. The training datasets were obtained from the BraTs 2018 challenge. In the network, we used four types of images, including T1, T1ce, T2 and FLAIR images, for input images. The segmentation network was SegNet. To boost the variability of the images, we randomly changed the angles, contrasts, and brightness of the training images. Based on the results, we learned that augmentation of enhancing variability of the images has better outcomes than images without augmentation. Also, we extracted two-dimensional slices from three slice orientations and used an ensemble method in the prediction stage. According to the results, the above strategies all contributed to the improvement of the accuracy of segmentation. During the validation phase, we applied the optimized segmentation method to the validation datasets. The average dice coefficients are 0.766, 0.896, 0.824 for the three subsets, the enhancing tumor, the tumor core, and the whole tumor, respectively.

Keywords: Gliomas, Deep learning, SegNet, Image segmentation

1 Introduction

Brain gliomas associated with glial cells are the major brain tumors among adults. Developing a precise diagnosis system for gliomas is the important issue nowadays. BraTs2018 hosts the competition for segmentation of brain gliomas with MR images [1-4]. Recently, deep learning methods based on convolutional neural networks (CNN), such as AlexNet, VGGNet, Google Inception Net, and ResNet [5] has shown efficient in image recognition tasks. These methods won ILSVRC competitions in recent years. Instead of constructing full connected “layers,” CNN used series of convolutional layers in the network. The extracted convolutional kernels are similar to the filters of image processing methods (e.g., Sobel filters, edge detectors). CNN-based techniques are especially suitable for image applications. On the other hand, graphics processing unit (GPU) accelerating the training stage, and worldwide competitions such as ImageNet challenge and BraTs, all contributes the fast development of deep-learning methods. Expanding the application of image recognition, the semantic segmentation methods recognize the class of each pixel in an image. Thus, it could be a suitable method for the segmentation task of the BraTs 2018 challenge. In the study, we attempt to use and compare the recently advanced convolutional neural network architecture for our task. Besides, we also compared different pre- and post-processing methods, such as normalization, augmentation, and voting, to improve the accuracy of the results.

2 Method

BraTS 2018 provided 285 MRI datasets with the Nifty file format, including subjects with high-grade gliomas (HGG):210, and low-grade gliomas (LGG): 75. Each dataset consists of four types of MR image contrasts, T1, T1 with contrast-enhancement, T2, and FLAIR and a true-label image with pixels labeled as the GD-enhancing tumor (ET), the peritumoral edema (ED), and the necrotic and non-enhancing tumor core (NCR/NET). They were all aligned and normalized to the same space, and the MR volume matrix was (X, Y, Z = 240×240×155). We padded zeros to each 3D volume, combined the four contrasts and the true-label image to construct a 4D training volume (256×256×256×4) and a 4D labeled volume (256×256×256×1) for each subject. For training and validation, we divided the 285 subjects into 255 training datasets and 30 testing datasets.

We used two-dimensional SegNet for pixel-wise semantic segmentation[6]. The architecture of our SegNet implementation is shown in Figure 1. The network consisted of the encoder and de-coder layers with 26 convolutional layers and an output layer connected a multi-class soft-max classifier. Figure 2 illustrated the convolution layer and the max pooling layer. It was implemented with the Tensorflow framework (v1.8) and the Python (v3.6) environment. We extracted 256×3 images from three slice orientations, including axial, coronal and sagittal slices. With 255 subjects in the training datasets, we then had 256×3×255 = 195840 images with a matrix size of 256×256. Subsequently, we normalized the images with the maximum intensity of each volume and created a new training dataset after removing the images without pixels labeled as gliomas[7]. Also, we used augmentation, including image flipping, rotation, transpose, and contrast adjustments, in the training procedures to enhance the variability of the images. We compared four different configurations of the training procedures, termed Seg-1 to Seg-4. In Seg-1, we trained the network directly from the raw data using only the axial slice orientations. The Seg-2 procedure was the same as Seg-1 albeit with normalized datasets. In Seg-3, the datasets were normalized, and the three slice orientations were used in training and predicting stages. The Seg-4 procedure was the Seg-3 procedure plus the data augmentation approaches.

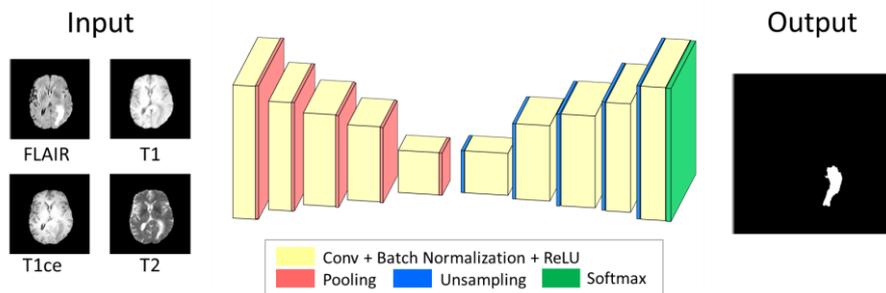


Figure 1 The SegNet architecture

We used the rest 30 datasets to evaluate the performance of the segmentation method. For each subject, we created different types of tensors as the input of the network. The configurations (batch, height, width, channels) of tensors were six combinations of dimensional permutations, (Z, X, Y, channels), (Z, Y, X, channels), (Y, X, Z, channels), (Y, Z, X, channels), (X, Y, Z, channels) and (X, Z, Y, channels). The channels were four MR image contrasts. We then produced the network softmax outputs of six tensors, merged the 2D slices into 3D volumes, removed the zero-padded parts of the volumes, and obtained six volumes with softmax values of each tumor subtypes.

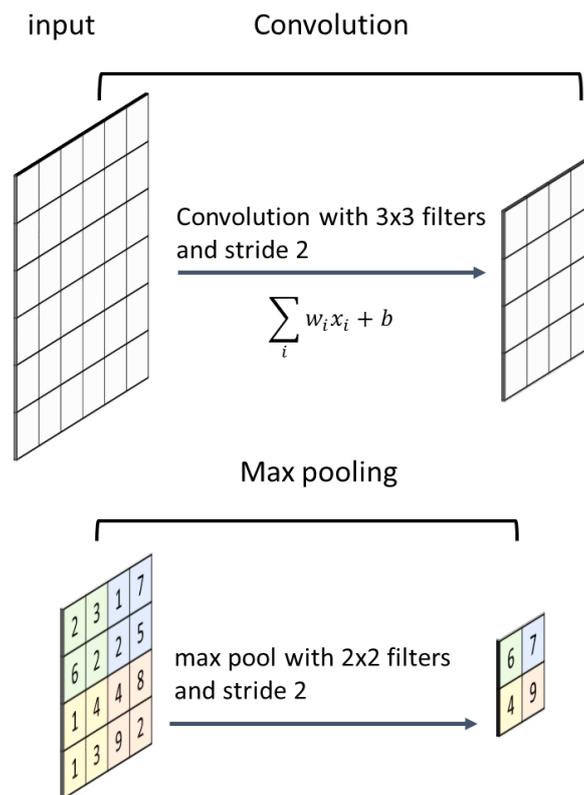


Figure 2 Convolution and max pooling

Finally, we implemented a voting method. Figure 3 displays the concept. We counted the four tumors types (non-tumor, ET, ED and NCR/NET) of the six volumes and identified the class of each voxel as the prediction of the tumor subtype of the voxel. The obtained volume with a matrix size of $240 \times 240 \times 155$ and pixels labeled with the tumor classes was stored as the Nifty file format for further evaluations and performance tests in the BraTS2018 website.

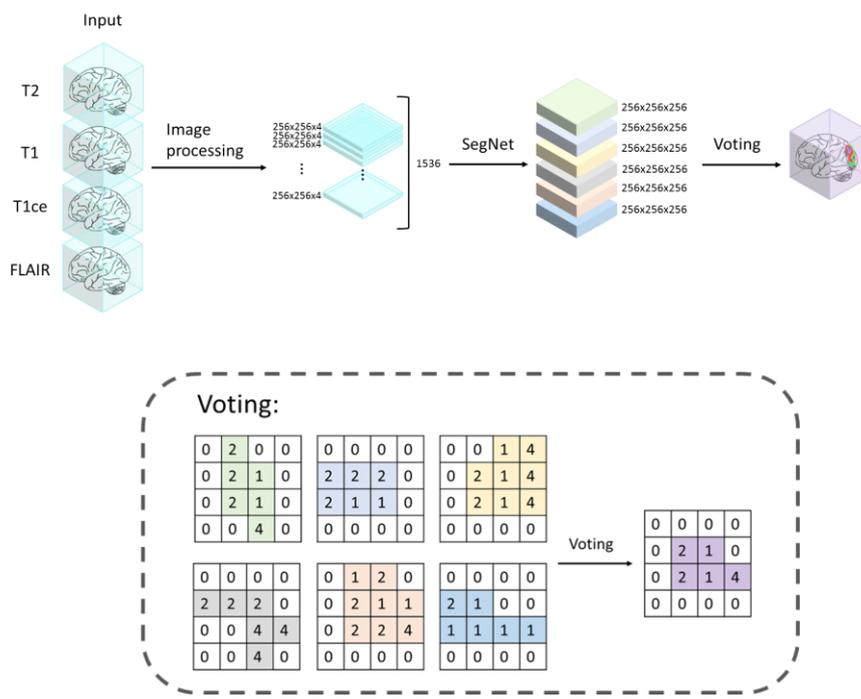


Figure 3 The voting method

Results

Figure 2 displays an example of the segmentation results. Figure 2(a) shows the true segmentation provided by BraTS2018 and Figure 2(b-e) shows the results obtained from SegNet with four different configurations of the training procedures, Seg-1 to Seg-4. Notice that the tumor regions are hardly recognized by the Seg-1 method. With data normalization, the Seg-2 results are slightly improved. Using the voting method, the accuracy of the Seg-3 method is prominently increased. The data augmentation approach of the Seg-4 method slightly improves the segmentation results. The results supported that normalization, voting, and augmentation are all beneficial to raise the accuracy of the segmentation results. The voting method using the six permutations of the 3D datasets played a major role to improve the segmentation accuracy.

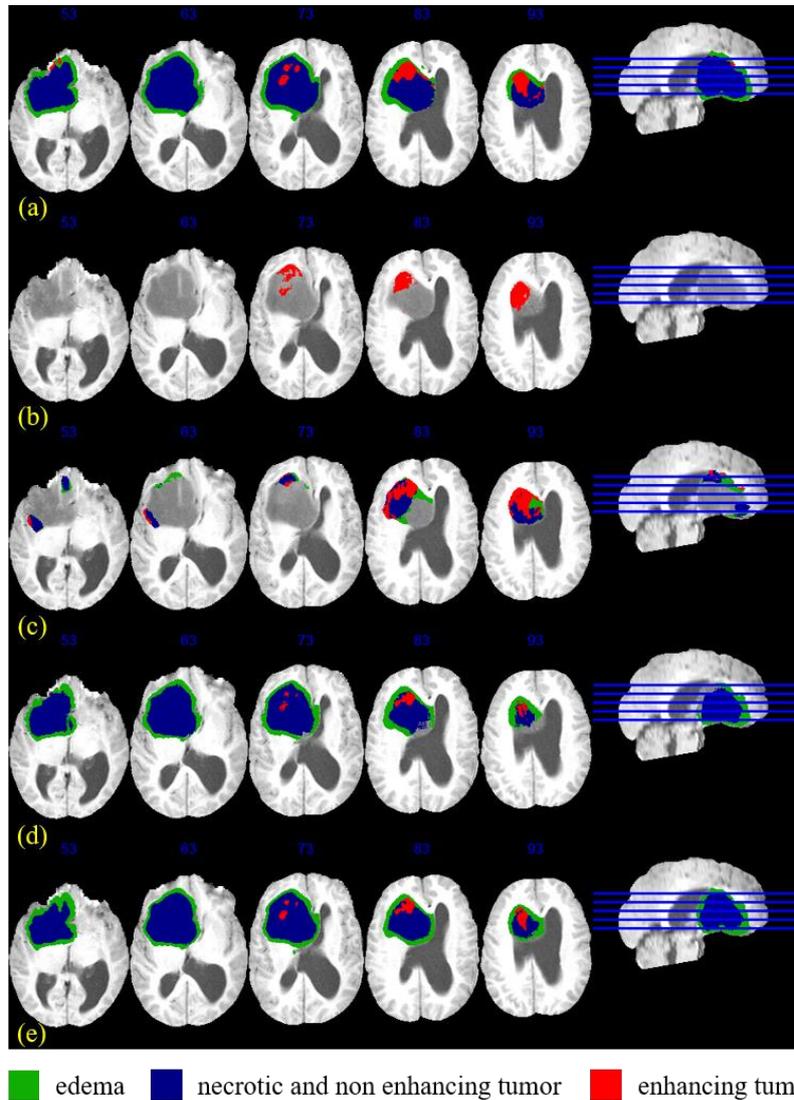


Figure 2. Results showing the effect of normalization, voting and augmentation. (a)Ground truth labels and the results of Seg-1 (b), Seg-2(c), Seg-3(d) and Seg-4(e).

Table 1 lists the average dice coefficients of the segmentation results of 30 testing subjects. The average Dice coefficients of the whole tumor region are 0.02, 0.75, 0.88, and 0.89 for the Seg-1, Seg-2, Seg-3 and Seg-4 methods, respectively. The group average results have the same trend as we observed in the example results shown in Figure 2. As a result, we adopted the Seg-4 method to conduct the segmentation challenge of Brats2018.

	Dice			Sensitivity			Specificity			Hausdorff_95		
	Enh.	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core
Seg-1	0	0.02	0.02	0	0.01	0.01	1	0.99	0.99	0	2.52	2.51
Seg-2	0.52	0.75	0.60	0.67	0.67	0.56	0.99	0.99	0.99	5.24	7.83	7.55
Seg-3	0.68	0.88	0.73	0.74	0.85	0.75	0.99	0.99	0.99	4.92	8.07	7.65
Seg-4	0.69	0.89	0.74	0.78	0.85	0.77	0.99	0.99	0.99	5.00	8.02	7.678

Table. 1. Average results affected by normalization, voting and augmentation

During the validation phase of the BraTs 2018, we received 66 datasets that were unseen by the neural network in the training stage. Then, we trained the SegNet using all the 285 training datasets with the Seg-4 configuration and used the obtained network to perform the tumor sub-type segmentation of the 66 datasets and produced the corresponding Nifty files of the segmentation results. The files were uploaded to the BraTs 2018 website to evaluate the accuracy of the results. Table 2 lists the accuracy indexes of the validation results. The average dice coefficients of the validation datasets are 0.766, 0.896, 0.824 for the three subsets, the enhancing tumor (ET), the tumor core (TC), and the whole tumor (WT), respectively.

	Dice			Sensitivity			Specificity			Hausdorff_95		
	Enh.	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core
Mean	0.766	0.896	0.824	0.78	0.873	0.793	0.998	0.996	0.998	3.285	4.292	5.946
StdDev	0.25	0.083	0.21	0.257	0.13	0.236	0.004	0.004	0.003	4.345	4.196	7.71
Median	0.854	0.922	0.9	0.862	0.907	0.89	0.999	0.998	0.999	1.732	2.828	3
25 quantile	0.788	0.889	0.803	0.748	0.865	0.762	0.999	0.993	0.999	1.414	2	1.494
75 quantile	0.896	0.944	0.938	0.944	0.949	0.937	0.999	0.999	0.999	3	4.899	5.911

66 subjects for validation

Table. 2. Results of 66 validation datasets

3 Discussions and Conclusions

In the study, we aimed to develop a fully automatic method to segment brain tumor regions and to identify the tumor subtypes of each voxel using four MRI volumes. We used a semantic segmentation method, SegNet for this task. We proposed four configurations of the training and prediction strategies. According to the segmentation results, we learned that normalization, the voting method, and data augmentations all contributed to the accuracy of the segmentation results. The accuracy of the validation results obtained with the optimized method was approximately 0.9 (the Dice coefficient of

WT) but the segmentation of the enhancing tumor region was not as accurate (the Dice coefficient of ET: 0.766). The CNN used in this study was implemented with 2D kernels. Thus, we spitted the 3D volumes into 2D slices and trained a 2D SegNet. The method may be not optimal for the 3D datasets since the accuracy of the segmentation may be improved by the 3D CNN for extracting 3D tumor features. Nonetheless, we introduced the voting method combined with permutations of image dimensions which prominently improved the results. Using 2D slices, the amount of training datasets is significantly larger than the amount of 3D datasets (195840 vs. 285). The amount of datasets used in the training stage is crucial for the deep neural networks. During the competition period of BraTS2018, we are continuing to develop the method. We attempt to combine other neural networks, such as ResNet [8] and U-Net [9] and evaluate whether ensemble methods [10] could improve the segmentation results. In conclusion, the 2D semantic segmentation method combined with the voting method could be a practical method for the segmentation of brain gliomas.

References

1. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJ1Q
2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
3. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015)
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
5. Mishkin, D., N. Sergievskiy, and J. Matas, Systematic evaluation of convolution neural network advances on the Imagenet. *Computer Vision and Image Understanding*, 2017. 161: p. 11-19.
6. Badrinarayanan, V., A. Kendall, and R. Cipolla, SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *Ieee Transactions on Pattern Analysis and Machine Intelligence*, 2017. 39(12): p. 2481-2495.

7. Yang TL, Ou YN., Huang TY, Automatic segmentation of brain tumor from MR images using SegNet: selection of training data sets, in BraTS 2017. 2017: Quebec, Canada. p. 309-312.
8. He, K., et al. Deep residual learning for image recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
9. Ronneberger, O., P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. in International Conference on Medical image computing and computer-assisted intervention. 2015. Springer.
10. Kamnitsas, K., et al., Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation. arXiv preprint arXiv:1711.01468, 2017.

S3D-UNet: Separable 3D U-Net for Brain Tumor Segmentation

Wei Chen, Boqiang Liu*, Suting Peng, Jiawei Sun, and Xu Qiao

Department of Biomedical Engineering, School of Control Science and Engineering,
Shandong University, China
chenypic@mail.sdu.edu.cn
bqliu@sdu.edu.cn

Abstract. Brain tumor is one of the leading causes of cancer death. Accurate segmentation and quantitative analysis of brain tumor are critical for diagnosis and treatment planning. Since manual segmentation is time consuming, tedious and error-prone, a fully automatic method for brain tumor segmentation is needed. Recently, state-of-the-art approaches for brain tumor segmentation are built on fully convolutional neural networks (FCNs) using either 2D or 3D convolutions. However, 2D convolutions cannot make full use of the spatial information of volumetric medical image data, while 3D convolutions suffer from high expensive computational cost and memory demand. To address these problems, we proposed a novel Separable 3D U-Net architecture, which exploits the intra-slice and inter-slice representations separately. In this way, the number of parameters of the network can be greatly reduced, thus reducing the risk of overfitting. Preliminary results of our method on validation data have mean dice coefficients of 0.73, 0.88, and 0.80 for enhancing tumor, whole tumor, and tumor core, respectively.

Keywords: Separable, Segmentation, BraTS, Convolutional Neural Networks

1 Introduction

Image segmentation, especially semantic segmentation, is a fundamental and classic problem in computer vision. It refers to partitioning an image into several disjoint semantically meaningful parts and classifying each part into a pre-determined class. In the application of brain tumor segmentation, the task includes the division of several sub-regions, such as GD-enhancing tumor, peritumoral edema, and the necrotic and non-enhancing tumor core [1]. Accurate segmentation and quantitative analysis of brain tumor are critical for diagnosis and treatment planning. Generally, manual segmentation of brain tumor is known to be time consuming, tedious and error-prone. Therefore, there is a strong need for a fully automatic method for brain tumor segmentation. However, brain tumor segmentation is a challenging task because MR images are typically acquired using various protocols and magnet strengths, which results in the non-standard range of MR images. In addition, brain tumors can appear

anywhere in the brain, and their shape and size vary greatly. Furthermore, the intensity profiles of tumor regions are largely overlap with healthy parts. Due to the challenge of brain tumor segmentation and the broad medical prospect, many researchers have proposed various methods to solve the problem of brain tumor segmentation.

Brain tumor segmentation methods can be divided into different categories according to different principles [2]. Broadly, these methods can be divided into two major categories: generative methods and discriminative methods. Generative methods typically rely on the prior information about the appearance of both healthy tissues and tumors. The proposed models often regard the task of segmentation as a problem of a posteriori distribution estimation. On the contrary, discriminative methods use very little prior information and typically rely on a large number of low-level image features to learn the distribution from the annotated training images.

More recently, due to the success of convolutional neural networks (CNNs), great progress has been made in the field of computer vision. At the same time, many deep learning based brain tumor segmentation methods have been proposed and achieved great success. Havaei et al. [3] proposed a two-pathway architecture with a local pathway and a global pathway, which can simultaneously exploit both local features and more global contextual features. Kamnitsas et al. [4] proposed an efficient fully connected multi-scale CNN architecture named deepmedic that uses 3D convolution kernels and reassembles a high resolution and a low resolution pathway to obtain the segmentation results. Furthermore, they used a 3D fully connected conditional random field to effectively remove false positives. Isensee et al. [5] proposed 3D U-Net, which carefully modified the popular U-Net architecture and used a dice loss function to cope with class imbalance. They achieved competitive results on the BraTS 2017 testing data. Kamnitsas et al. [6] introduced EMMA, an ensemble of multiple models and architectures including deepmedic, FCNs and U-Net. Due to the heterogeneous collection of networks, the model is insensitive to independent failures of each component and has good generalization performance. They won first place in the final testing stage of the BraTS 2017 challenge among more than 50 teams.

Although so many achievements have been made, the progress of medical image analysis is slower than that of static images, and a key reason is the 3D properties of medical images. This problem also occurs in the tasks of video understanding. To solve this problem, Xie et al. [7] proposed S3D model by replacing 3D convolutions with spatiotemporal-separable 3D convolutions. This model significantly improved on the previous state-of-the-art 3D CNN model in terms of efficiency.

Inspired by S3D architecture for video classification and the state-of-art U-Net architecture for medical image segmentation, we propose a novel framework named S3D-UNet for brain tumor segmentation. We decouple a 3D convolution with one 2D filter on intra-slice and one 1D filter on inter-slice. To make full use of 3D volumes, we divide each 3D convolution into three branches in a parallel fashion, each with a different orthogonal view, i.e. axial, sagittal and coronal. In this way, the number of parameters of the network can be greatly reduced, thus reducing the risk of overfitting in the case of a small training set.

Methods

1.1 Dataset

The brain tumor MRI dataset used in this study are provided by BraTS'2018 Challenge [1, 8–10]. The training dataset includes multimode brain MRI scans of 285 subjects, of which 210 are GBM/HGG and 65 are LGG. Each subject contains four scans: native T1-weighted (T1), post-contrast T1-weighted (T1c), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR). All the subjects in the training dataset are provided with ground truth labels, which are segmented manually by one to four raters. Annotations comprise the GD-enhancing tumor (ET - label 4), the peritumoral edema (ED - label 2), and the necrotic and non-enhancing tumor core (NCR/NET - label 1). The validation dataset includes multimode brain MRI scans of 66 subjects which are similar to the training dataset but have no expert segmentation annotations and the grading information.

1.2 Data Pre-processing

To remove the bias field caused by the inhomogeneity of the magnetic field and the small motions during scanning, the N4ITK bias correction algorithm [11] is first applied to the T1, T1c and T2 scans. The multimodal scans in BraTS 2018 were acquired with different clinical protocols and various scanners from multiple institutions[1], resulting in a non-standardized intensity distribution. Therefore, normalization is a necessary stage of processing multi-mode scanning by a single algorithm. We use histogram matching algorithm [12] to transform each scan to a specified histogram to ensure that all the scans have a similar intensity distribution. We also resize the original image of 240x240x155 voxels to 128x128x128 voxels by removing as many zero background as possible. This processing not only can effectively improve the calculation efficiency, but also retain the original image information as much as possible. In the end, we normalize the data to have a zero mean and unit variance.

1.3 Architecture

Our framework is based on the U-Net structure proposed by Ronneberger et al. [13] which consists of a contracting path to analyze the whole image and a symmetric expanding path to recovery the original resolution. The U-Net structure has been widely used in the field of medical image segmentation and has achieved competitive performance. Several studies[14, 15] have demonstrated that the 3D versions of U-Net architecture using 3D volumes as input can produce better results than entirely 2D architecture. Although 3D U-Net has good performance, it has more parameters than 2D version, and the computational complexity of 3D model is much higher than that of 2D model. Spatiotemporal-separable 3D convolutions can greatly improve the efficiency of the 3D CNN model. An effective model using spatiotemporal-separable 3D convolution is S3D [7] for video understanding, which significantly improved on the previous state-of-the-art 3D CNN model in terms of efficiency. The most promi-

nent feature of this model is that each 3D convolution is decoupled into a 2D convolution layer to learn spatial features and a 1D convolution layer to learn temporal features.

Inspired by the recent successes of S3D and U-Net in many computer vision tasks such as video understanding and medical image segmentation, we develop a new model named S3D-UNet, which can make full use of 3D volumes while reducing the number of parameters and computational complexity. The schematic representation of our proposed network is shown in Fig. 1. Just like the original U-Net, our network consists of two parts: the left part corresponds to the contracting path that encodes the increasingly abstract representation of the input, and the right part corresponds to the expanding path that restores the original resolution.

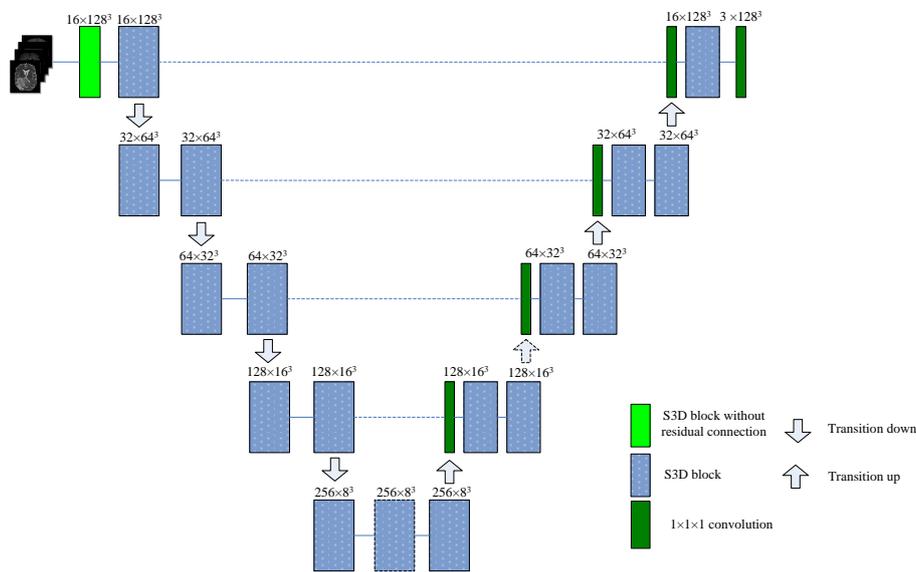


Fig. 1. Schematic representation of our proposed network.

The original separable 3D convolution architecture is to replace a 3D convolution with two consecutive convolution layers: one 2D convolution to learn spatial features and one 1D convolution to learn temporal features. This structure is very effective for video understanding tasks. For the task of brain tumor segmentation, the original S3D convolution structure may not be optimal due to the isotropic high resolution (1 mm^3) of MRI scans. In the processing of MRI scans, we refer to the corresponding spatial convolution and temporal convolution in original S3D convolution as intra-slice convolution and inter-slice convolution. Naturally, we divide a 3D convolution into three branches in a parallel fashion, each with a different orthogonal view, i.e. axial, sagittal and coronal, as shown in Figure 2. After each convolution, an instance

normalization [16] and LeakyReLU activation is applied. We call this architecture as S3D block, which is the main contribution of our proposed method.

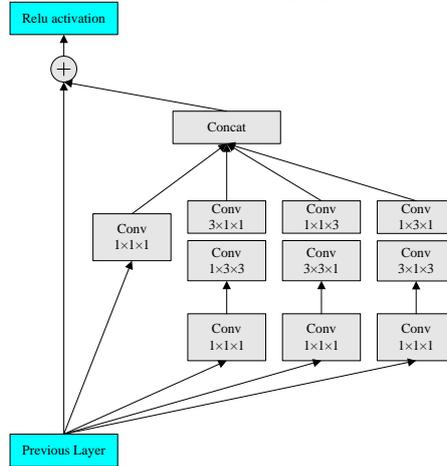


Fig. 2. The detailed illustration of S3D block

Each level of the network comprises one to three S3D blocks of different resolutions. As the data proceeds through different level along the contracting path, its resolution is reduced. This is performed through transition down, which includes a 3D convolution with a kernel size of $2 \times 2 \times 2$ and a stride of 2. On the contrary, the transition up, which includes a 3D deconvolution with a kernel size of $2 \times 2 \times 2$ and a stride of 2, is used to expand the resolution in the expanding path.

1.4 Loss Function

Table 1. The distribution of the classes in the training data of BraTS 2018.

	Background	NCR/NET	ED	ET
Percentage	98.88	0.28	0.64	0.20

The performance of neural network depends not only on the choice of network structure but also on the choice of the loss function[17]. Especially for severe class imbalance, the choice of loss functions becomes more important. Due to the physiological characteristics of brain tumors, the segmentation task has an inherent class imbalance problem. Table.1 illustrates the distribution of the classes in the training data of BraTS 2018. Background (label 0) is overwhelmingly dominant. According to [5], we apply a multiclass Dice loss function to approach this issue. Let R be the one hot coding ground truth segmentation with voxel values r_n^k , where $k \in K$ being the class at voxel $n \in N$. Let P be the softmax output the network with voxel values p_n^k , where $k \in K$ being the class at voxel $n \in N$. The multiclass Dice loss function can be expressed as

$$DL = 1 - \frac{2}{K} \sum_{k \in K} \frac{\sum_n p_n^k r_n^k}{\sum_n p_n^k + \sum_n r_n^k} \quad (1)$$

2 Results

The network is trained on one GeForce GTX 1080Ti GPU with a batch size of 1 using PyTorch toolbox. Adam [18] is used as the optimizer with an initial learning rate 0.001 and a 12 weight decay of 1e-8. We evaluate all the cases for training data and validation data using online CBICA portal for BraTS 2018 challenge. The segmentation results of validation data are presented in Table 1 and 2. Table 3 and 4 show the results of training data. Examples of the segmentations obtained from the training data using our method are shown in Figure 3.

Table 2. Dice and Sensitivity for BraTS 2018 validation dataset

	Dice			Sensitivity		
	ET	WT	TC	ET	WT	TC
Mean	0.73348	0.88787	0.80781	0.77863	0.90126	0.80024
StdDev	0.2289	0.07533	0.18744	0.21653	0.08534	0.20449
Median	0.80157	0.90601	0.87509	0.84971	0.92149	0.88745
25quantile	0.69969	0.87748	0.75545	0.68999	0.88176	0.72666
75quantile	0.85716	0.93244	0.9243	0.93181	0.95833	0.94249

Table 3. Specificity and Hausdorff95 for BraTS 2018 validation dataset

	Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC
Mean	0.99752	0.99427	0.99817	4.64261	5.50541	8.14015
StdDev	0.00328	0.00414	0.00219	7.24967	7.10154	13.1969
Median	0.99871	0.99581	0.99868	2.23607	3.16228	4.24264
25quantile	0.99731	0.99228	0.99793	1.41421	2.23607	2.23607
75quantile	0.99943	0.99725	0.99946	4.24264	5	9.43398

Table 4. Dice and Sensitivity for BraTS 2018 training dataset

	Dice			Sensitivity		
	ET	WT	TC	ET	WT	TC
Mean	0.73953	0.88809	0.84419	0.78628	0.88069	0.83281
StdDev	0.23324	0.05405	0.11202	0.20056	0.08275	0.11848
Median	0.80752	0.89999	0.88117	0.83615	0.90545	0.86713
25quantile	0.71495	0.86768	0.80397	0.7227	0.83986	0.78056
75quantile	0.86075	0.92713	0.91793	0.90571	0.939	0.92181

Table 5. Specificity and Hausdorff95 for BraTS 2018 training dataset

	Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC
Mean	0.99791	0.99481	0.9972	4.63102	5.88769	5.66071
StdDev	0.00224	0.00391	0.00466	6.53892	5.00074	6.60362
Median	0.99854	0.9957	0.99878	2.23607	4.89898	4.12311
25quantile	0.99712	0.99315	0.99732	1.41421	3.16228	2.23607
75quantile	0.99949	0.99736	0.99944	4.32834	7.09921	7.2111

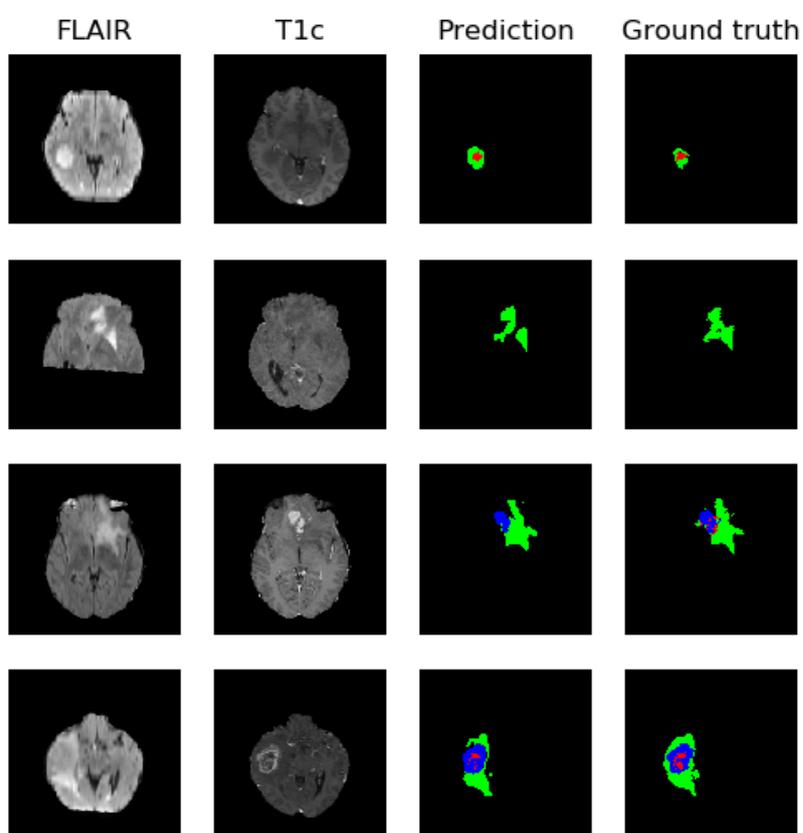


Fig. 3. Examples of segmentation from the of BraTS 2018 training data. red: NCR/NET, green: ED, blue: ET.

3 Discussion

We have presented S3D-UNet architecture which aims to learn intra-slice and inter-slice representations in deep networks. Particularly, we decouple a 3D convolution

with one 2D filter on intra-slice and one 1D filter on inter-slice. To make full use of 3D volumes, we divide each 3D convolution into three branches in a parallel fashion, each with a different orthogonal view, i.e. axial, sagittal and coronal. This model can significantly reduce the number of parameters and reduce the risk of over-fitting. In the future work, we will continue to improve the structure of the network and use some post-processing methods such as fully connected conditional random field to remove false positives.

Acknowledgment

This work was supported by the Department of Science and Technology of Shandong Province (Grant No. 2015ZDXX0801A01, ZR2014HQ054, 2017CXGC1502), National Natural Science Foundation of China (grant no. 61603218).

References

1. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*. 34, 1993–2024 (2015).
2. Gordillo, N., Montseny, E., Sobrevilla, P.: State of the art survey on MRI brain tumor segmentation. *Magnetic Resonance Imaging*. 31, 1426–1438 (2013).
3. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., Larochelle, H.: Brain tumor segmentation with Deep Neural Networks. *Medical Image Analysis*. 35, 18–31 (2017).
4. Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis*. 36, 61–78 (2017).
5. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge. In: Crimi, A., Bakas, S., Kuijff, H., Menze, B., and Reyes, M. (eds.) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. pp. 287–297. Springer International Publishing, Cham (2018).
6. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., Glocker, B.: Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation. In: Crimi, A., Bakas, S., Kuijff, H., Menze, B., and Reyes, M. (eds.) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. pp. 450–462. Springer International Publishing, Cham (2018).

7. Xie, S., Sun, C., Huang, J., Tu, Z., Murphy, K.: Rethinking Spatiotemporal Feature Learning For Video Understanding. arXiv:1712.04851 [cs]. (2017).
8. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data*. 4, 170117 (2017).
9. Spyridon Bakas, C.D., Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. *The Cancer Imaging Archive*. (2017).
10. Spyridon Bakas, C.D., Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. *The Cancer Imaging Archive*. (2017).
11. Tustison, N.J., Avants, B.B., Cook, P.A., Yuanjie Zheng, Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*. 29, 1310–1320 (2010).
12. L. G. Nyul, J. K. Udupa, Xuan Zhang: New variants of a method of MRI scale standardization. *IEEE Transactions on Medical Imaging*. 19, 143–150 (2000).
13. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 234–241. Springer (2015).
14. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. arXiv:1606.06650 [cs]. (2016).
15. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge. arXiv:1802.10508 [cs]. (2018).
16. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance Normalization: The Missing Ingredient for Fast Stylization. arXiv:1607.08022 [cs]. (2016).
17. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. arXiv:1707.03237 [cs]. 10553, 240–248 (2017).
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. (2014).

Segmentation of Brain Tumors using DeepLabv3+

Ahana Roy Choudhury¹, Rami Vanguri^{2,3}, Sachin R. Jambawalikar², and
Piyush Kumar¹

¹ Department of Computer Science, Florida State University, Tallahassee, FL, USA.
{roychoud,piyush}@cs.fsu.edu

² Department of Radiology, Columbia University Medical Center, New York, USA.
sj2532@cumc.columbia.edu

³ Data Science Institute, Columbia University, New York, NY, USA.
rv2368@cumc.columbia.edu

Abstract. Multi-modal MRI scans are commonly used to grade brain tumors based on size and imaging appearance, which play an important role in the diagnosis and treatment administered to patients. Deep learning based approaches in general, and convolutional neural networks in particular, have been utilized to achieve superior performance in the fields of object detection and image segmentation. In this paper, we propose to utilize the DeepLabv3+ network for the task of brain tumor segmentation. For this task, we consider only the T1CE, T2 and FLAIR images to identify the whole tumor, the tumor core and the enhancing core of the brain tumor for 57 subjects. We use 80% of the MICCAI BraTS training data to train our network and use the remaining 20% for validation purposes. Our method involves the segmentation of individual slices in the three orientations and a majority voting-based combiner. We achieve Dice coefficients of 0.8154, 0.8016, and 0.9059 for the enhancing tumor, the tumor core and the whole tumor respectively.

Keywords: Image segmentation, Convolutional neural networks, DeepLab, MRI, tumor, enhancing tumor, tumor core

1 Introduction

Gliomas are a type of tumor that start in glial cells and occur in the brain and spinal cord. Differentiating High-grade gliomas (HGG) from Low-grade gliomas (LGG) is essential in determining patient treatment path, since 80% of malignant brain tumors are considered HGGs. By developing a segmentation algorithm to automatically identify tumor regions using multimodal MRI scans, there is potential to improve diagnostics and follow-up treatment. Due to the heterogeneity in the size, location, and shape of gliomas, developing algorithms for automatic segmentation is challenging.

To accomplish automatic segmentation, we propose to use the DeepLab v3+ framework[6], which has been shown to successfully identify objects in natural

images. To do this, we utilize the pre-trained Xception network of DeepLabv3+ and further train on multimodal MRI, treating the individual modes as image channels. We use the sagittal, coronal and axial slices separately to train a pre-trained DeepLabv3+ model, thus obtaining three separate models trained on the slices in the three different orientations. For validation, we use the sagittal, coronal and axial slices as the inputs to the models trained on the corresponding slices and then employ a basic majority voting-based combiner to combine the predictions.

2 Methodology

2.1 Data

The training dataset for the MICCAI BraTS 2018 challenge consists of 285 sets of NIfTI image files, with 210 sets of image files for HGG type of tumor and 75 sets for LGG. Each set of image files consists of MRI scans using four different modes: T1, T2, T1CE and FLAIR. Also, the corresponding segmentation map is provided for each of the 285 patients. The segmentation labels are assigned as follows:

- Label 4 is assigned to the enhancing tumor region.
- The edema is denoted by label 2.
- The NCR+NET region is identified using label 1.

Our task is to identify the enhancing tumor, the tumor core, which consists of the enhancing tumor and the NCR+NET regions, and the whole tumor, which consists of all three of the above regions. Details about the BraTS challenge and the annotated dataset are provided in [10] and [3]. The relevant data sources are [4] and [2]

2.2 Data Preprocessing

In order to perform segmentation of the different regions of a brain tumor, we slice the 3-D NIfTI brain images along 3 orientations: sagittal, coronal, and axial. By doing so, we get 155 axial slices and 240 slices each in the sagittal and coronal directions. In the original image, the intensity of each voxel is stored as a 16 bit integer and as a result, the maximum possible value of intensity of a voxel is greater than 255. In order to use DeepLabv3+ for segmenting the slices, we store each slice as a separate `.png` file. However, the intensity of each pixel in a `.png` file can only range from 0 to 255. Thus, we scale the intensities in the original NIfTI image to a value between 0 to 255 before saving each slice as a `.png` file. Each NIfTI image file stores the intensities of $240 \times 240 \times 155$ voxels. Thus, each axial slice consists of 240×240 pixels, while the coronal and sagittal slices have 240×155 pixels each. We use zero padding to pad each of the coronal and sagittal slices to increase their size to 240×240 pixels.

We use only the T1ce, T2 and FLAIR images to perform our experiments. We combine the corresponding slices from our data and create RGB images, where the FLAIR image occupies channel 1 (R), T1ce occupies channel 2 (G) and T2 occupies channel 3 (B).

2.3 Deeplabv3+ Architecture

Convolutional neural networks typically apply convolution with different strides as well as pooling. These methods cause a reduction in the sizes of the feature maps that are operated on by the succeeding layers, and as a result, the last feature map lacks comprehensive information associated with object boundaries. Hence, in case of semantic image segmentation, such architectures fail to provide precise delineation of objects. In order to overcome this challenge, an encoder-decoder pathway is used in several architectures. While the U-Net[11] decoder concatenates the feature maps at corresponding scales in the encoder and decoder pathways, SegNet[1] saves and uses the max pooling indices from the encoder pathway. On the other hand, Deeplabv3[5] uses parallel atrous convolutions at different rates in order to capture the information at different scales. Effectively, this technique, called the Atrous Spatial Pyramid Pooling (ASPP) allows the encoder network to use different fields of view. Thus, while encoder-decoder based networks use the decoder pathway to progressively recover spatial context, networks that use ASPP can store multi-scale information within their feature maps. DeepLabv3+[6] combines these two techniques and uses an encoder-decoder pathway as well as ASPP in order to achieve precise delineation of object boundaries.

DeepLabv3+ has a similar encoder architecture as the one used in DeepLabv3, which utilizes ASPP. However, due to the computation cost and complexity of performing ASPP as well as the size constraints of GPU memories, it is not possible to extract features that have a resolution greater than $\frac{1}{8}^{th}$ of the input image. An output stride of 16 (at most 8) is used in the convolution layer in order to perform image segmentation. The last feature map generated by the encoder network, which has 256 channels, is used as the input to the decoder network. Besides this, DeepLabv3+ involves the use of depth-wise separable convolution, which divides a traditional convolution operation into a depth-wise convolution and a point-wise convolution. This results in lower computation cost and uses fewer number of parameters, while maintaining the same level of accuracy.

To overcome the fact that the encoder cannot give an output that has the same resolution as the input image, DeepLabv3+ introduces a decoder network. The decoder network first uses bilinear upsampling to increase the resolution of the feature maps that are received from the encoder. Then, a 1×1 convolution is performed before the features are concatenated with the corresponding features from the encoder network. Finally, the decoder performs 3×3 convolutions and upsamples the output to the resolution of the original image. The decoder network uses depth-wise separable convolution and an output stride of 16 to strike a balance between the accuracy and the computation speed.

DeepLabV3+ modifies the Xception model[7] by making the network deeper, using depth-wise separable atrous convolution instead of max pooling, and introducing additional batch normalization and non-linear activation layers. The resulting architecture[8], which is implemented using Tensorflow, does not include any form of post-processing and achieves an mIoU of 89% on the Pascal VOC dataset. At present, DeepLabv3+ heads the Pascal VOC leaderboard for image segmentation, and thus, we decided to use this network in order to develop our automatic brain tumor segmentation software.

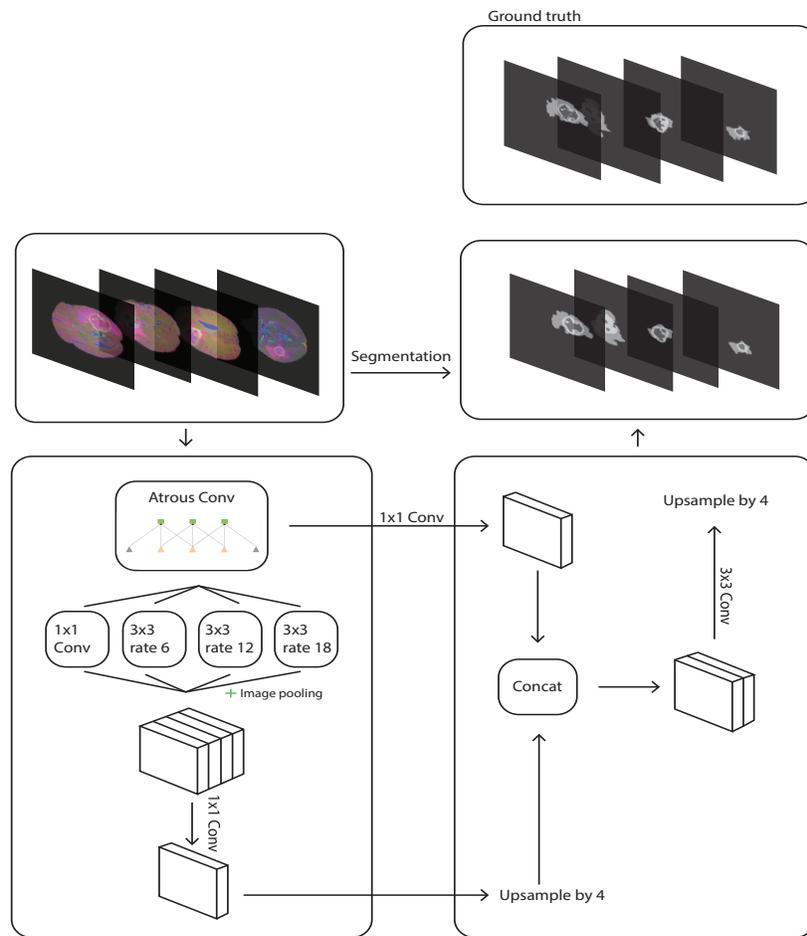


Fig. 1. DeepLabv3+ involves the use of atrous convolution at multiple scales in the encoder network as well as a decoder network

2.4 Brain tumor Segmentation using DeepLabv3+

For the purpose of our preliminary experiments, we have divided the BraTS training data into two sets, one consisting of the MRI scans of 80% of the patients and the other one consisting of the remaining 20% data. We use the first set to train our network and the second set for validation. Thus, the validation set consists of the MRI scans from 57 patients and is not seen by our model during training.

During training, we only use the slices in our training set which actually have tumors. On an average, in the training data provided, about 1.07% of the brain is occupied by tumor in case of HGG and about 1.24% in case of LGG. As a result, the portion of the brain that contains tumors is significantly smaller than the healthy part. So, the number of voxels that have label 0 is much greater than those that are labeled 1, 2, or 4. Thus, to minimize the class imbalance, and to ensure that the sections of the brain that have tumors can be successfully recognized, we only use the slices that have tumors during the training phase. However, the validation set consists of all the slices in a specific direction (axial, sagittal or coronal) and are segmented using the classifier trained on the slices in the corresponding orientation.

We train our model using an existing DeepLabv3+ pre-trained model. Specifically, we use a checkpoint of the Xception model[9], which is pre-trained on the augmented Pascal VOC dataset. We use a learning rate of 0.005 (the recommended learning rate for training this particular model is 0.007) and train for a total of 75000 steps. We use poly decay and a decay rate of 0.1 to decrease the learning rate after every 30000 iterations. While a larger batch size could have helped us to train our model more successfully, we had to use a batch size of 8 due to GPU memory constraints.

Using the above method, we train our model separately on the axial, sagittal and coronal slices. Thus, we get three separate trained models and use each model to predict the label of each voxel using the corresponding validation slices. In this way, after performing training and validation on the axial, sagittal and coronal slices, we get three predictions for each voxel in the validation data.

Finally, we use majority voting in order to assign a label to each voxel based on the predictions obtained in the previous step. Our simplistic majority voting scheme works by first recombining all the prediction slices in the axial direction to get the 3-D brain segmentation. We also perform the same operation along different axes to reconstruct the segmented image using the sagittal and coronal slices. Then, after we have reconstructed the predictions in order to get three 3-D brain images, one from the slices in each of the three orientations, we apply majority voting. Here we refer to the prediction for voxel at position i, j, k considering the sagittal direction slices as $p_{i,j,k}^{sagittal}$, the axial direction slices as $p_{i,j,k}^{axial}$ and coronal direction slices as $p_{i,j,k}^{coronal}$. If, for any voxel, any two of these predictions are the same, that is either $p_{i,j,k}^{sagittal} = p_{i,j,k}^{coronal}$ or $p_{i,j,k}^{sagittal} = p_{i,j,k}^{axial}$ or $p_{i,j,k}^{axial} = p_{i,j,k}^{coronal}$, then we assign the label predicted by the majority to that voxel. However, if all the three predictions differ, we give priority to the prediction that uses the slices in the orientation that achieves the best performance. From our

experiments, we found that the highest average mIoU is achieved by the model that uses the cross-sectional slices.

Our preprocessing codes and the majority voting code can be found at <https://www.dropbox.com/s/efidw7km7j4te4q/BraTS18FSUML.zip?dl=0>

3 Result

We evaluate the performance of our network by training the model separately on our training sets for the sagittal, coronal and axial slices. The models are used to perform predictions on the validation slices of the corresponding orientations. We report the Dice coefficient and sensitivity of each of the trained models. We also report the Dice coefficient and the sensitivity of our combiner model which combines the predictions from each orientation by using majority voting.

We first provide a few samples in order to visualize our image segmentation results for slices in the three orientations. The figures provided in this section depict the level of accuracy that we have achieved by using slices in each of the three orientations. However, in some cases the uneven boundaries are not identified accurately. By using the prediction combiner, we are able to handle some special cases where any one of the three models can not achieve the best segmentation, but a combination of the predictions from the three models can enhance the accuracy of the prediction.

In the Figures 2, 3 and 4, the image in the top left is the ground truth, our prediction is depicted by the image in the middle of the top row, while the input FLAIR, T1ce, and T2 images are located at the bottom left, middle and right respectively.

In the Figure 5, we illustrate how the basic majority voting-based combiner improves the accuracy of brain tumor segmentation. The image on the left of the top row is the ground truth, the middle of the top row shows the prediction after using the combiner and the blank image on the top right shows the axial prediction. Thus, even the axial model, which has the best performance can make wrong predictions that can be corrected using majority voting.

Table 1. Comparison of the Dice Coefficient achieved by the three models trained on the sagittal, coronal, axial slices and the basic majority voting-based combiner

Class	Sagittal Prediction	Coronal Prediction	Axial Prediction	Majority prediction
WT	0.8836	0.8939	0.8950	0.9059
CT	0.7937	90.7727	0.8000	0.8016
ET	0.7936	0.7926	0.7954	0.8154

In Table 1 and Table 2 we provide the sensitivity and the Dice coefficient achieved for the three classes: Whole tumor (WT), tumor core (TC), and enhancing tumor

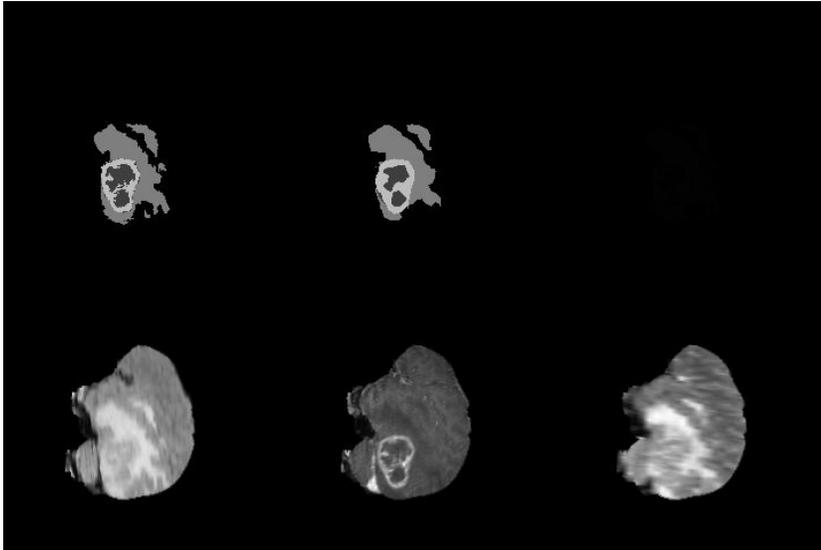


Fig. 2. Segmentation using sagittal slices: The image depicts the ground truth and the predictions using sagittal slices using the FLAIR, T1ce and T2 images as the inputs.

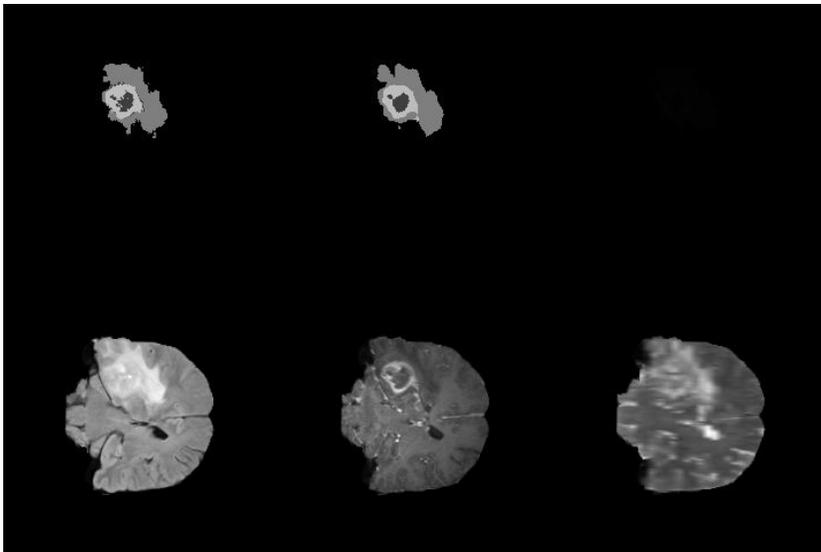


Fig. 3. Segmentation using coronal slices: The image depicts the ground truth and the predictions using coronal slices using the FLAIR, T1ce and T2 images as the inputs.

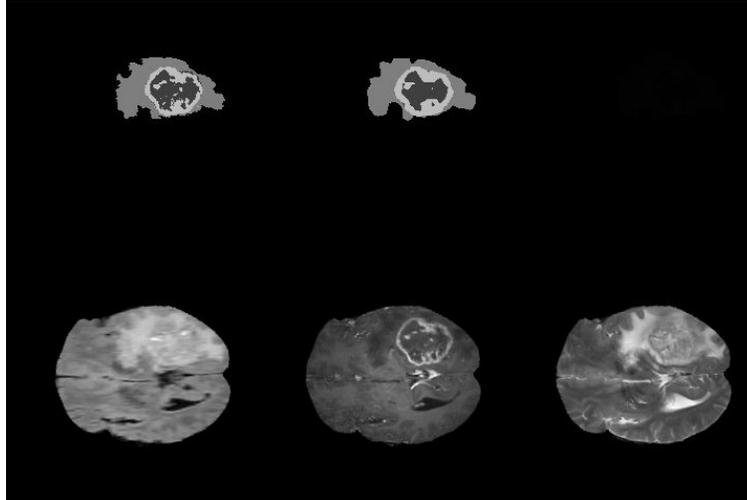


Fig. 4. Segmentation using axial slices: The image depicts the ground truth and the predictions using axial slices using the FLAIR, T1ce and T2 images as the inputs.

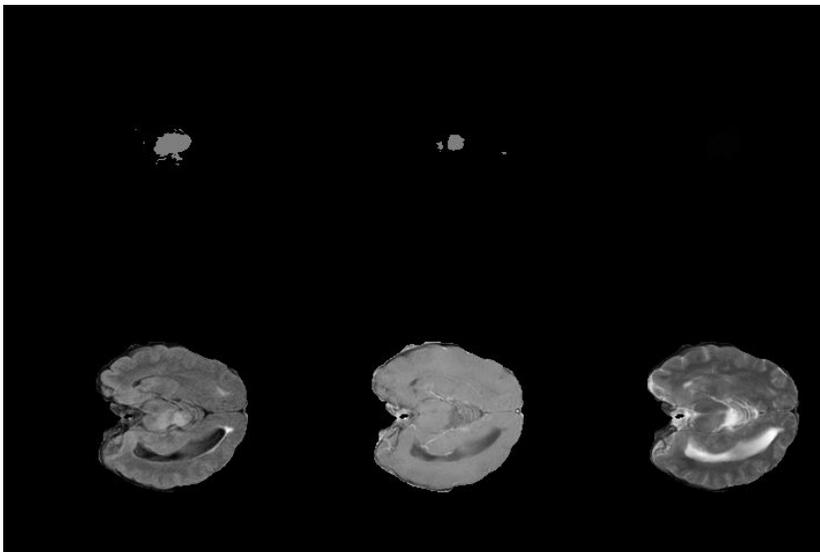


Fig. 5. Accuracy improvement using the basic majority voting-based combiner: The tumor can not be predicted by our model by using the axial slice but can be identified using majority voting.

(ET). For each model, we provide the results for the 20% of the BraTS 2018 training set, which we use for validation only.

Table 2. Comparison of the sensitivity achieved by the three models trained on the sagittal, coronal, axial slices the basic majority voting-based combiner.

Class	Sagittal Prediction	Coronal Prediction	Axial Prediction	Majority prediction
WT	0.8513	0.8661	0.8558	0.8628
CT	0.7106	0.6878	0.7311	0.7135
ET	0.7578	0.7738	0.7788	0.7926

From the above results, we see that the use of the basic majority voting-based combiner improves the Dice coefficient, while it has a mixed impact on the overall sensitivity.

4 Discussion

In the next month, we plan to make improvements on our current methods for both improving as well as implementing new features in our system. The main improvements we plan to implement is for the problem of combining classifiers to do better segmentations. The main feature that we plan to implement is a new predictor for survivability. We elaborate on both these plans in the following two subsections

4.1 Boosting

We currently can compute segmentation using multiple deep classifiers that act on multiple slices around a voxel and then employ a basic majority voting-based combiner. We plan to combine the output segmentations using weighted majority voting, and random forests. One of the challenges in this problem is the size of data. Optimizing Random forests on 100s or 1000s of datasets, where each dataset is millions of voxels, is too slow in real life. We plan to avoid this using a majority voting that can be done using one pass on the data.

We explain the majority voting using a concrete example. Currently our classifiers output 2 bits of information (4 different types of segmentation). For three axes (xy, yz, zx), we have 3 different classifiers that output a total of 6 bits of information. We know the ground truth segmentation for each voxel (2 bits). We plan to construct a matrix of size $2^6 \times 2^2$, where each entry (i, j) corresponds to the number of times the classifiers output $i \in (0, 2^6 - 1)$ and the actual segmentation of that voxel was j . Using this table we can now compute the probability of each segmentation given the output of the classifiers. We can now just output the highest probability segmentation for each input 6 bit classification output.

An easy extension of the above mentioned algorithm is to use the 6 bit output of the classifiers from neighboring voxels to decide the segmentation for the current voxel. The table size now increases depending on the neighbors included. For example, if there are 6-neighbors, the classifiers output a total of seven 6 bit numbers. This increases the table size to 7×2^6 rows. Now for each row, we can either do the most probable segmentation or design a random forest for computing the output 2 bit segmentation.

4.2 Survival Analysis

Prediction of overall patient survival is one of the most clinically relevant and challenging parts of the BraTS challenge. We will develop a separate deep learning classifier which uses features generated from the DeepLabv3+ framework to predict overall patient survival. As the data is limited we may also try traditional machine learning approaches like support vector regression or other regression techniques using hand crafted features from the tumor. We are currently extracting texture and shape features from the segmented tumor data for survival prediction from these features. These features will then be used with machine learning algorithms including logistic regression, random forests and support vector machines to predict survival. Additionally we will attempt an end-to-end deep learning solution that performs segmentation and applies a convolutional deep neural network to the segmented tumor volume to predict survival. However, the small amount of data may make this approach inviable.

5 Conclusion

We use the DeepLabv3+ deep learning framework to perform brain tumor segmentation and glioma classification. We train our network separately on the coronal ,axial and sagittal slices and use the networks to make predictions on the corresponding slices in the validation set. Using a basic majority voting-based combiner to combine the predictions from the three models, we achieve Dice coefficients of 0.9059, 0.8016, and 0.8154 for the whole tumor, core tumor and enhancing tumor respectively. With further improvements and additions to this network, we expect to use trained DeepLabv3+ outputs to build another network that can predict overall patient survival. If successful, this will form a basis for further research and can be interrogated to discover new tumor biomarkers related to survival.

References

1. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. CoRR **abs/1511.00561** (2015), <http://arxiv.org/abs/1511.00561>

2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. (2017), <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific Data* **4**, 170117 EP – (Sep 2017), <http://dx.doi.org/10.1038/sdata.2017.117>, data Descriptor
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection (2017), <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>
5. Chen, L., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. *CoRR* **abs/1706.05587** (2017), <http://arxiv.org/abs/1706.05587>
6. Chen, L., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. *CoRR* **abs/1802.02611** (2018), <http://arxiv.org/abs/1802.02611>
7. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. *CoRR* **abs/1610.02357** (2016), <http://arxiv.org/abs/1610.02357>
8. Google: Deeplabv3+ (2018), <https://github.com/tensorflow/models/tree/master/research/deeplab>
9. Google: `deeplabv3_pascal_train_aug` (2018), http://download.tensorflow.org/models/deeplabv3_pascal_train_aug_2018_01_04.tar.gz
10. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, ., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftexharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging* **34**(10), 1993–2024 (Oct 2015). <https://doi.org/10.1109/TMI.2014.2377694>
11. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. *CoRR* **abs/1505.04597** (2015), <http://arxiv.org/abs/1505.04597>

Automatic Brain Tumor Segmentation with Domain Transfer

Lutao Dai^a, Tengfei Li^b, Hai Shu^c, Liming Zhong^d, Haipeng Shen^a and Hongtu Zhu^b

Faculty of Business and Economics, The University of Hong Kong^a
Department of Biostatistics, University of North Carolina at Chapel Hill^b
Department of Biostatistics, University of Texas, MD Anderson Cancer Center^c
Guangdong Provincial Key Laboratory of Medical Image Processing,
School of Biomedical Engineering, Southern Medical University, Guangzhou, China^d

Abstract. Deep convolution neural networks, in particular, the encoder-decoder networks, have been extensively used in image segmentation. To improve the segmentation accuracy, we divide training samples into two domains according to preliminary segmentation results, and then equip such networks with domain transfer to learn a domain invariant feature representation. Our proposed segmentation network is applied to BraTS 2018 challenge for brain tumor segmentation, and achieves the dice score of 0.90482, 0.84144 and 0.76493 for whole tumor, tumor core and enhancing tumor, respectively, on the challenge's validation data set.

Keywords: Confusion loss, domain transfer, encoder-decoder network

1 Introduction

Image segmentation plays an important role in the accurate diagnosis and efficient treatment of brain tumors. However, segmenting brain tumors such as glioblastomas and gliomas is difficult because of poor tissue contrast, irregular shapes and various appearing locations. Moreover, manual segmentation can be very time-consuming and may have large intra/inter-expert variability. This creates a great need to develop reliable automatic approaches for brain tumor segmentation.

The Brain Tumor Segmentation (BraTS) challenge [2, 7, 11, 10] is one of such platforms to evaluate the state-of-the-art methods on a large data set of annotated high grade glioblastomas and lower grade gliomas. To foster accurate segmentation, the BraTS 2018 challenge provides multimodal MRI scans of each patient, including native T1-weighted, post-contrast T1-weighted, T2-weighted, and T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes.

Modern deep convolutional networks have exhibited exceptional competitiveness in image segmentation, becoming industrial benchmarks [5, 13, 4]. A large family of such networks is the encoder-decoder networks with U-shaped architectures, such as DeconvNet [8], SegNet [1], and U-Net [9, 3]. These networks

are composed of a convolutional encoder to extract salient features and a deconvolutional decoder to recover image details. Such architecture has advantages, including flexible input image sizes, consideration of spatial information, and an end-to-end prediction, leading to lower computational cost and higher representation power.

Despite the excellent performance in the 2017 challenge, the state-of-the-art encoder-decoder network of [4] in our model exploration still loses significant segmentation accuracy for part of the BraTS 2018 training set. This is probably because the network majorly captures the key features of well-segmented samples but misses those of the others. From the transfer learning perspective as in [12], if treating the well-segmented samples as samples in the “source” domain and the poor-segmented samples in the “target” domain, then the network fails to learn a domain invariant feature representation. This can hence be viewed as the so-called domain transfer problem that aims to match the marginal feature distributions of source and target. Inspired by the domain transfer technique of [12], we add a domain classifier to the modified U-net of [4] together with a confusion loss to learn a domain invariant feature representation for the brain tumor segmentation task. Our proposed network with domain transfer significantly enhances the segmentation accuracy on the validation set with dice score of [0.90482](#), [0.84144](#), [0.76493](#) for whole tumor, tumor core and enhancing tumor respectively.

2 Data Description

The BraTS 2018 challenge data [2, 7, 11, 10] are collected from three different resources that are denoted as “2013”, “CBICA”, and “TCIA”, respectively. The training data set includes 20 high-grade gliomas subjects (HGGs) from “2013”, 88 HGGs from “CBICA”, and 102 HGGs from “TCIA”, and also includes 10 “2013” and 65 “TCIA” subjects with low-grade gliomas (LGG) that are less aggressive and infiltrative.

Each subject has four modalities of MRI scans, including native T1-weighted (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 FLAIR volumes. All MRI images are registered to a common template with the volume size of $240 \times 240 \times 155$ voxels resampled to 1 mm isotropic resolution. The tumor regions are annotated into three classes: the GD-enhancing tumor (ET, labeled 4), the peritumoral edema (ED, labeled 2), and the necrotic and non-enhancing tumor (NCR/NET, labeled 1).

A validation data set of 66 subjects is also provided for each participating team but with no HGG/LGG status or tumor labels. The segmentation results on the validation set are allowed to submit multiple times at an online evaluation platform. The final evaluation of a segmentation approach will be conducted on an independent test data set released later.

3 Encoder-decoder Network with Domain Transfer

In this section, we discuss the details of our framework for brain tumor segmentation. Specifically, we use domain transfer to regulate the feature learning so that the proposed network model can capture more general features and therefore improve generalizability. We also discuss our data preprocessing and post-processing procedures that smooth and optimize the segmentation results.

3.1 Data Preprocessing

The main purpose of data preprocessing is to bring intensities to similar distributions to avoid any initial bias. This is important for data-driven approaches such as deep learning algorithms. The provided data had already been skull stripped, co-registered and resized to uniform resolution. On top of that, we removed the top and bottom 1 intensity percentiles, and then normalized the intensities by subtracting the mean and dividing the standard deviation. The preprocessing was conducted on brain regions only and independently across modalities and individuals.

3.2 Network Architecture

Our proposed network consists of two components: a modified U-Net responsible for segmentation, and a domain classifier that regulates the feature learning of the U-Net. The network architecture is illustrated in Fig. 1.

The modified U-Net is inspired by [4]. In our model, each level of the encoding pathway consists of a residual block with the same structure. The first convolution layer of each residual block halves the spatial dimension with a stride of 2 (except for the first residual block), and increases the number of channels to 8×2^n with n the level counting from 1. As a result, the stack of 5 residual blocks progressively reduces the spatial dimension of the input tensor by a factor of 16, expands the receptive field, and learns increasingly abstract feature representations. To increase the prediction resolution, the decoding pathway progressively doubles the spatial dimension on each level by an upsampling layer of scale 2, eventually recovering the original spatial dimension. The feature maps generated by the first four residual blocks are concatenated to decoding pathway of the same level, further improving the gradient flow. Moreover, the idea of deep supervision [6] is adopted, where output maps of different levels are combined sequentially through element-wise addition to constitute the network's final prediction via a softmax function.

The domain classifier is appended to the end of encoding pathway, where the feature representation is the most abstract. A $1 \times 1 \times 1$ convolution layer is first applied to significantly reduce the number of channels from 256 to 32, followed by alternating three fully connected layers of length 256, 32 and 1, and two leaky ReLU rectifiers.

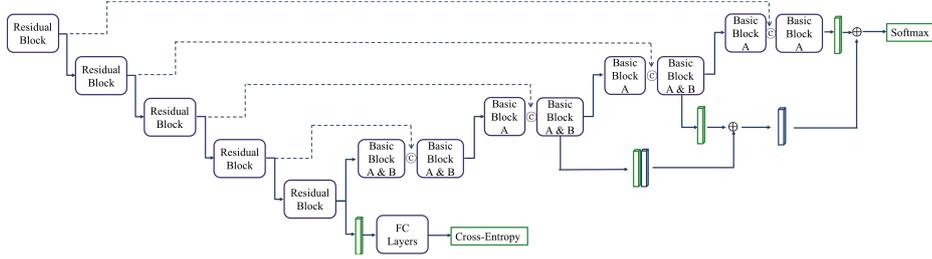


Fig. 1. Network architecture

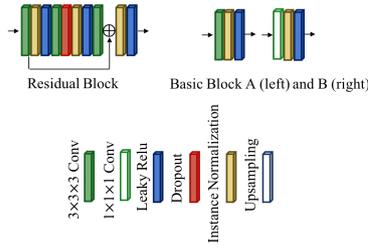


Fig. 2. Network blocks

3.3 Domain Division

We performed two sets of five-fold cross-validation using the modified U-Net of [4] (one of which has additional N4 bias correction) and obtained two full sets of the dice coefficients on the training samples. Although N4 bias correction had a minimal effect on most samples, it did yield significantly different results for some cases. We picked 75 subjects with the most variations and classified them to a different domain from the rest. The full list of those 75 subjects can be found in [Appendix](#).

3.4 Training Procedure

The U-Net and domain classifier are trained alternatively. In the $2n$ -th epoch, the parameters of the domain classifier are kept unchanged and only parameters of the U-Net are updated. The objective function is composed of the multiclass dice loss function as in [4] and a confusion loss that is the cross entropy of the predicted domain label and a uniform distribution:

$$L_{2n} = -\frac{2}{|K|} \sum_{k \in K} \frac{\sum_i u_i^k v_i^k}{\sum_i u_i^k + \sum_i v_i^k} - \lambda \sum_d \frac{1}{D} \log q_d,$$

where K is the set of prediction classes, D is the set of domain categories, u and q are the probability maps respectively output by the U-Net and the domain classifier, v is the one-hot encoding of the ground truth, and i is the voxel index.

The hyperparameter λ controls the degree of regulating effect of the domain classifier to the U-Net.

In the $(2n + 1)$ -th epoch, all parameters except those of the domain classifier are frozen. The domain classifier aims to discriminate samples according to the feature representation output by the encoding pathway. The cross-entropy loss is computed with domain labels as follows:

$$L_{2n+1} = - \sum_d \mathbb{I}[y_D = d] \log q_d.$$

By training the model iteratively, both the U-Net and the domain classifier are optimized. The best domain classifier learned by L_{2n+1} is expected to still perform poorly on the final domain prediction due to the confusion loss in L_{2n} . With such a domain classifier, the encoding pathway has incentives to capture the domain-invariant features. This helps to improve generalizability since differences in MRI data representation are usually significant.

Two decreasing pattern types of loss functions are observed in our experiments (see Fig. 3). Type I refers to a pattern with a sudden drop. The sudden drop is majorly contributed by the cross-entropy loss from the domain classifier. Interestingly, such a drop is synchronized regardless the cross-entropy loss is computed with domain labels or the uniform distribution. Type II behaves similarly to the one without domain transfer, where training loss decreases steadily. Given the same training data and sufficient training epochs, cross-entropy losses of either type reach similar magnitude eventually.

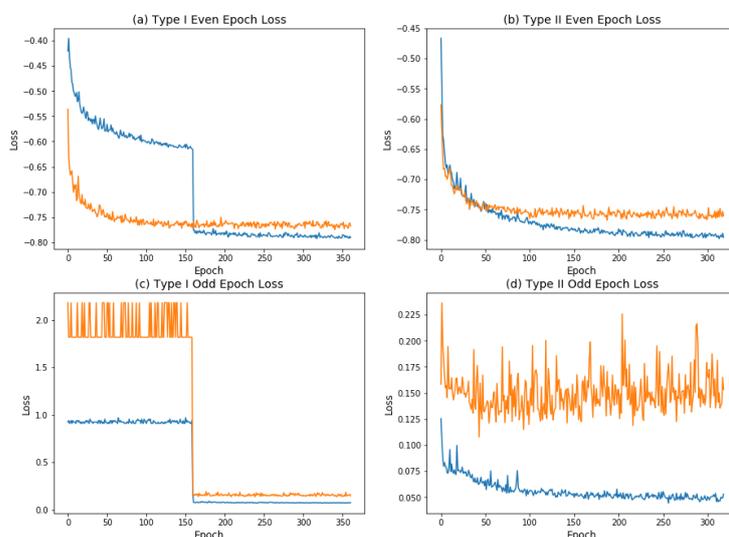


Fig. 3. Two pattern types of loss functions (blue: training set; orange: validation set). The validation loss in even epochs refers to the dice loss only.

The input tensors of size $128 \times 128 \times 128$ are randomly sampled from brain areas and augmented by random flipping and transpose during each epoch. The training is implemented by PyTorch using the Adam optimizer with the learning rate initially set to be 8×10^{-4} and exponentially decaying at a rate of 0.98 over every two epochs. The proposed network with each different value of tuning parameter λ was trained for about 600 epochs, and the one with the minimum dice loss on the validation stage was chosen as the output model.

At the test time, the domain classifier is dropped. The whole brain regions whose dimensions are padded to the nearest multiple of 16 are served as inputs to the modified U-Net, and the returned segmentation maps are subsequently padded to be the original dimension.

3.5 Post-processing

We employed the following post-processing techniques [14] to fill in the holes and delete the small isolated clusters:

1. Segment the tumor mask into all connected components/clusters. Voxels in clusters whose volume is less than 0.2 times the largest connected cluster volume will be reclassified as non-tumor.
2. Segment the enhancing core mask into all connected components/clusters. Voxels in clusters whose volume is less than 0.01 times the largest connected cluster volume will be reclassified as the necrosis.
3. Fill in the holes within the tumor mask and assign voxels within the holes to necrosis area.

By comparing the dice-ratios of the defined tumor sub-regions by our TPCNN model with and without the three-step post-processing, we find that the performance could be improved by applying the post-processing on the existing TPCNN results after the two-phase prediction.

4 Results

To investigate the optimum value of λ in L_{2n} , we conducted multiple trials with λ as the only varying parameter. Within a certain range of λ , there is a clear enhancement of the average dice coefficient for whole tumor and tumor core, whose optimum values are achieved at $\lambda = 0.1$ and $\lambda = 0.075$ respectively. Passing over the optimum point, we can see a clear decline in average dice coefficient for both. The average dice coefficient of enhancing tumor fluctuates with λ , but its highest peak is at $\lambda = 0.1$, which is very close to the optimal λ 's of whole tumor and tumor core. We hence choose $\lambda = 0.1$ for our proposed network.

Detailed segmentation results are shown in table 5.

5 Appendix

See table 1.

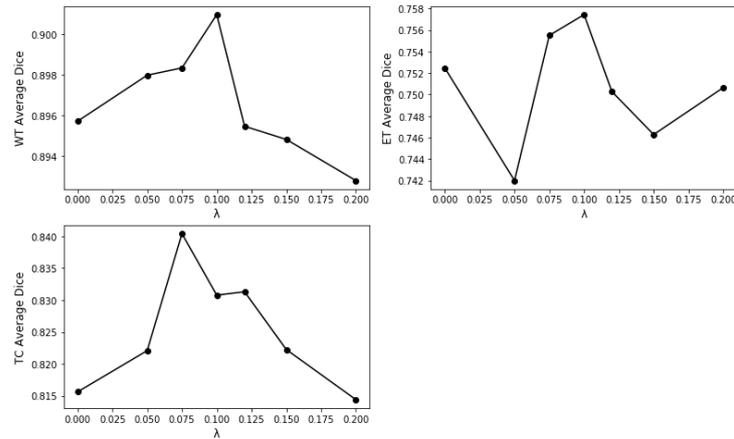


Fig. 4. Dice coefficients with varying λ .

	DSC			Sensitivity			Specificity			Hausdorff_95		
	Enh.	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core
Training	0.6965	0.88815	0.81974	0.76504	0.87444	0.81115	0.99834	0.99526	0.9969	5.92345	6.42145	7.03793
Validation	0.76493	0.90482	0.84144	0.79687	0.89929	0.80553	0.9984	0.99589	0.9988	3.77781	5.19277	7.23815

Fig. 5. Segmentation result.

References

1. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561 (2015)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Nature Scientific Data* 4, 170117 (2017)
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 424–432. Springer (2016)
4. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In: *International MICCAI Brainlesion Workshop*. pp. 287–297. Springer (2017)
5. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., et al.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: *International MICCAI Brainlesion Workshop*. pp. 450–462. Springer (2017)
6. Kayalibay, B., Jensen, G., van der Smagt, P.: Cnn-based segmentation of medical imaging data. arXiv preprint arXiv:1701.03056 (2017)
7. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor

ID	Name	Type
1	BraH182013_111	HGG
2	BraH182013_151	HGG
3	BraH182013_221	HGG
4	BraH182013_271	HGG
5	BraH182013_41	HGG
6	BraH182013_411	HGG
7	BraH182013_4111	HGG
8	BraH182013_41111	HGG
9	BraH182013_411111	HGG
10	BraH182013_4111111	HGG
11	BraH182013_41111111	HGG
12	BraH182013_411111111	HGG
13	BraH182013_4111111111	HGG
14	BraH182013_41111111111	HGG
15	BraH182013_411111111111	HGG
16	BraH182013_4111111111111	HGG
17	BraH182013_41111111111111	HGG
18	BraH182013_411111111111111	HGG
19	BraH182013_4111111111111111	HGG
20	BraH182013_41111111111111111	HGG
21	BraH182013_411111111111111111	HGG
22	BraH182013_4111111111111111111	HGG
23	BraH182013_41111111111111111111	HGG
24	BraH182013_411111111111111111111	HGG
25	BraH182013_4111111111111111111111	HGG
26	BraH182013_41111111111111111111111	HGG
27	BraH182013_411111111111111111111111	HGG
28	BraH182013_4111111111111111111111111	HGG
29	BraH182013_41111111111111111111111111	HGG
30	BraH182013_411111111111111111111111111	HGG
31	BraH182013_4111111111111111111111111111	HGG
32	BraH182013_41111111111111111111111111111	HGG
33	BraH182013_411111111111111111111111111111	HGG
34	BraH182013_4111111111111111111111111111111	HGG
35	BraH182013_41111111111111111111111111111111	HGG
36	BraH182013_411111111111111111111111111111111	HGG
37	BraH182013_4111111111111111111111111111111111	HGG
38	BraH182013_41111111111111111111111111111111111	HGG
39	BraH182013_411111111111111111111111111111111111	HGG
40	BraH182013_4111111111111111111111111111111111111	HGG
41	BraH182013_41111111111111111111111111111111111111	HGG
42	BraH182013_411111111111111111111111111111111111111	HGG
43	BraH182013_4111111111111111111111111111111111111111	HGG
44	BraH182013_411	HGG
45	BraH182013_4111	HGG
46	BraH182013_411	HGG
47	BraH182013_4111	HGG
48	BraH182013_411	HGG
49	BraH182013_4111	HGG
50	BraH182013_411	HGG
51	BraH182013_4111	HGG
52	BraH182013_411	HGG
53	BraH182013_4111	HGG
54	BraH182013_411	HGG
55	BraH182013_4111	HGG
56	BraH182013_411	HGG
57	BraH182013_4111	HGG
58	BraH182013_411	HGG
59	BraH182013_4111	HGG
60	BraH182013_411	HGG
61	BraH182013_4111	HGG
62	BraH182013_411	HGG
63	BraH182013_4111	HGG
64	BraH182013_411	HGG
65	BraH182013_4111	HGG
66	BraH182013_411	HGG
67	BraH182013_4111	HGG
68	BraH182013_411	HGG
69	BraH182013_4111	HGG
70	BraH182013_411	HGG
71	BraH182013_4111	HGG
72	BraH182013_411	HGG
73	BraH182013_4111	HGG
74	BraH182013_411	HGG
75	BraH182013_4111	HGG

Table 1. Selected 75 subjects.

image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging* 34(10), 1993 (2015)

8. Noh, H., Hong, S., Han, B.: Learning deconvolution network for semantic segmentation. In: *Proceedings of the IEEE international conference on computer vision*. pp. 1520–1528 (2015)
9. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 234–241. Springer (2015)
10. S, B., Akbari H, S.A., Bilello M, R.M., J, K., J, F., K, F., C., D.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive* (2017)
11. S, B., Akbari H, S.A., Bilello M, R.M., J, K., J, F., K, F., C., D.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *The Cancer Imaging Archive* (2017)
12. Tzeng, E., Hoffman, J., Darrell, T., Saenko, K.: Simultaneous deep transfer across domains and tasks. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 4068–4076 (2015)
13. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: *International MICCAI Brainlesion Workshop*. pp. 178–190. Springer (2017)
14. Zhou, F., Li, T., Li, H., Zhu, H.: Tpcnn: Two-phase patch-based convolutional neural network for automatic brain tumor segmentation and survival prediction. In: *International MICCAI Brainlesion Workshop*. pp. 274–286. Springer (2017)

Three pathways U-Net for brain tumor segmentation

Longwei Fang^{1,2} and Huiguang He^{1,2,3}

¹ Research Center for Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Sciences

² University of Chinese Academy of Sciences

³ Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences

huiguang.he@ia.ac.cn

Abstract. Multi-modality images are widely used in brain tumor diagnose. Different modalities clearly show different sub-regions of the tumor. For instance, the whole tumor is easy to be segmented from FLAIR image, the necrotic components and the enhancing core can be easily distinguished from T1 post contrast image. In this paper, we present a novel three pathways U-Net structure to segment the brain tumor. Each modality is processed in a single pathway, flair pathway for whole tumor segmentation and t1ce pathway for the enhance tumor and necrotic components segmentation. At the end of the model, these two pathways are fused together by the fusion pathway to get the final segmentation. We adopt a weighting scheme at the loss layer to alleviate the problem of label imbalance. The preliminary results on BRATS-2018 validation dataset show that the average dice similarity coefficient of the whole brain, tumor core and enhancing part is 85.6%, 72.6%, and 72.2%, respectively.

Keywords: Segmentation, Brain Tumor, Convolution Neural Network, U-Net.

1 Introduction

Brain tumor segmentation is a fundamental problem in medical image analysis field[1, 2]. But it is labor-intensive and impractical to manually segment a large set of 3D MR images. Thus, recent developments mainly focused on automatic segmentation. However, there are multiple challenges in automatic segmentation: 1) the tumor has a low contrast with the surrounding tissue 2) large variation in tumor shape across different subjects.

In the past few years, many attempts have been made to address these challenges in tumor segmentation. The convolutional neural network has become the most promising technique for image segmentation. In particular, U-Net shape[3, 4] architectures have dominated this field. U-Net shape architecture is composed of convolutions and deconvolutions. A novel bridge strategy is used to connect the shadow layer features with the deep layer features. This bridge can help the neural network to better learn the mapping from raw data to the segmentation label and also can help to alleviate the gradient vanish problem in training stage. However, it's still difficult to segment the sub-regions of

the tumor from a single modality image by just using the U-Net. The main reason is that the low contrast in image intensity. For example, the whole tumor is easy to be segmented from FLAIR image, but the necrotic components of the tumor cannot be distinguished from this modality. Necrotic components can be easily segmented from T1 post contrast image, but the peritumoral edema of the tumor is hard to be distinguished from the normal tissue. As shown in **Fig. 1**.

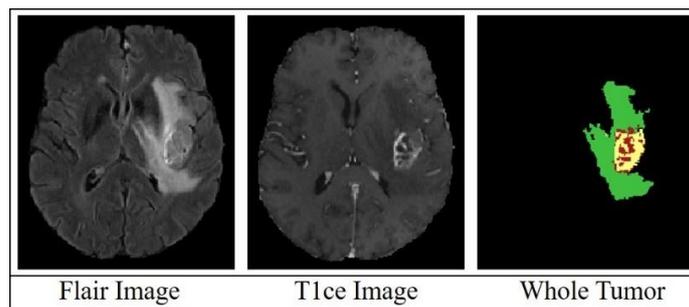


Fig. 1. Different modality images show different parts of the tumor. Flair image clearly shows the whole tumor but cannot distinguish necrotic components. While the T1ce image can distinguish necrotic components and enhance tumor but the peritumoral edema is hard to be distinguished from the normal tissue. The green color in the whole tumor is peritumoral edema, the yellow color is the enhance tumor and the red color is the necrotic components.

In this paper, a multi-modality strategy is applied to segment the brain tumor. We propose a novel three pathways U-Net architecture by using FLAIR image and T1 post contrast image. Each modality image is processed in a single pathway. Flair-pathway receives Flair image for whole tumor segmentation and t1ce-pathway receive the T1ce image for the enhance tumor and necrotic components segmentation. We fuse these two pathways together in the fusion pathway at the end of the model to get the final segmentation. The main novelty of our method is two folds:

- 1) Two pathways to process different modality data. Each pathway segment specific region of the tumor.
- 2) A novel weight scheme is used to alleviate the problem of label imbalance. The weights of each label are calculated from the training data.

2 Methods

In this section, we detail the proposed three pathways U-Net framework for automatic brain tumor segmentation. Our method comprises training and testing stages. In the training stage, a 2d slices select strategy is used to prepare the training slices, and then the weights for each label is calculated by the label amount of all training slices. Then all the training slices are inputted into the proposed model for training. In the testing stage, every slice from the image volume is processed one by one to get the final

segmentation result. After the testing stage, a post-processing procedure is applied to further refine the segmentation result.

2.1 Three pathways U-Net Framework

The flowchart of the proposed framework is summarized in **Fig. 2**, which comprises three components: 1) flair-pathway, 2) t1ce-pathway and 3) fusion pathway. T1ce-pathway is used to segment the enhance tumor and necrotic components from the T1ce image, while flair-pathway is used to segment the whole tumor and brain from flair image. Both pathways use U-Net structure that can combine the low-level feature with the high-level feature by concatenating operation. For flair-pathway and t1ce-pathway, we add two auxiliary supervisions to force the network to encode more semantic concepts, separately. Specifically, we add t1ce loss at the end of t1ce-pathway and flair loss at the end of the flair-pathway. The ground truth of the flair-pathway contains two labels, which includes the healthy brain tissue and the whole tumor. The ground truth of the t1ce-pathway contains two labels, which includes the enhance tumor and necrotic components. For fusion-pathway, the input is the sum of the last layers of flair-pathway and t1ce-pathway. The ground truth of the fusion-pathway contains all labels need to be segmented, which includes whole brain, peritumoral edema of the tumor, enhance tumor and necrotic components. The parameters for each layer are listed in **Tab. 1**. For convolution operation, the kernel size is 3x3 and the step size is 1, while for deconvolution operation, the kernel size is 4x4 and the step size is 2.

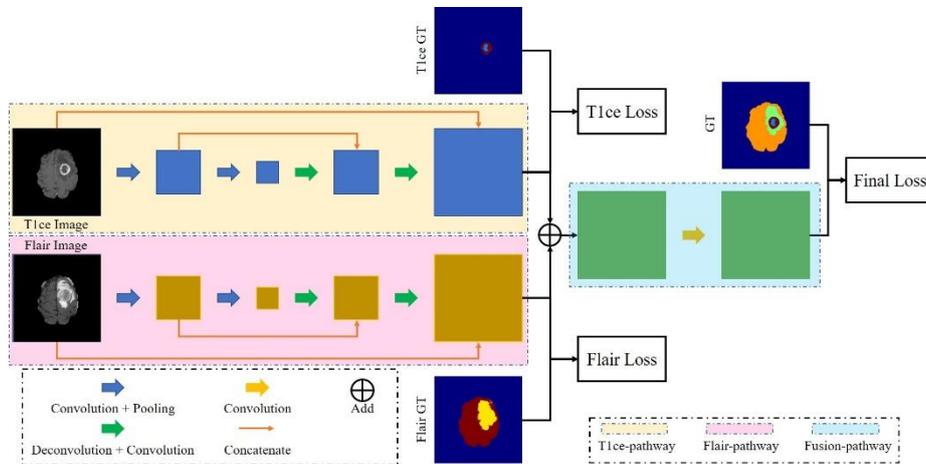


Fig. 2. The flowchart of three pathways U-Net architecture. The three pathways are highlighted in light-pink, light-yellow, and light-blue bands. GT represents the ground truth. The batch normalization layer and the ReLU layer are followed by the convolution and deconvolution operations.

Table 1. parameters for each layer of two pathways U-Net

Name	Type	Output Size
T1ce-pathway Input		240x240x1
Conv1 (repeat 2 times)	Convolution + BatchNorm + ReLu	240x240x64
Pooling1	Pooling	120x120x64
Conv2 (repeat 2 times)	Convolution + BatchNorm + ReLu	120x120x128
Pooling2	Pooling	60x60x128
Conv3 (repeat 2 times)	Convolution + BatchNorm + ReLu	60x60x256
Pooling3	Pooling	30x30x256
Conv4 (repeat 2 times)	Convolution + BatchNorm + ReLu	30x30x512
Pooling4	Pooling	15x15x512
Conv5 (repeat 4 times)	Convolution + BatchNorm + ReLu	15x15x1024
Deconv4	Deconvolution + BatchNorm + ReLu	30x30x512
Concat4 (Deconv4 + Conv4)	Concatenate	30x30x1024
Deconv4_1	Convolution + BatchNorm + ReLu	30x30x512
Deconv3	Deconvolution + BatchNorm + ReLu	60x60x256
Concat3 (Deconv3 + Donv3)	Concatenate	60x60x512
Deconv3_1	Convolution + BatchNorm + ReLu	60x60x256
Deconv2	Deconvolution + BatchNorm + ReLu	120x120x128
Concat2 (Deconv2 + Donv2)	Concatenate	120x120x256
Deconv2_1	Convolution + BatchNorm + ReLu	120x120x128
Deconv1	Deconvolution + BatchNorm + ReLu	240x240x64
Concat1 (Deconv1 + Donv1)	Concatenate	240x240x128
Deconv1_1	Convolution + BatchNorm + ReLu	240x240x64
T1ce_last	Convolution + BatchNorm + ReLu	240x240x5
Flair-pathway uses the same parameter with t1ce-pathway		
Add (T1ce_last + Flair_last)	Add	240x240x5
Conv6	Convolution + BatchNorm + ReLu	240x240x32
Fusion_last	Convolution + BatchNorm + ReLu	240x240x5

2.2 Training Slices Preparation

We only use high-grade gliomas (HGG) images to train our model[5-7]. The first step is to select the proper slices from the images volumes. In here, slices are selected according to the number of the label in the label volume. Specifically, we go through the label volume slice by slice. For each slice, the number of the label (peritumoral edema of the tumor, enhance tumor and necrotic components) is counted and only the counted value large than 300 is selected. Then we extract the corresponding intensity slice from the FLAIR and T1 post enhance images. In this paper, we use the flip operation to augment the training data. We should note that all the images are normalized to a range from 0 to 255 before training slices preparation.

2.3 Label Weights Calculation

The proposed architecture has three pathways. These three pathways segment different components of the tumor and each pathway has its unique supervision label. In this paper, the background is set to 0, the necrotic component is set to 1, peritumoral edema

of the tumor is set to 2, the healthy brain tissue is set to 3 and the enhance tumor is set to 4. For flair-pathway, we segment health brain tissue and whole brain, so we set label value 4 and 1 to value 2. Then the weight of each label is calculated by **Eq. 1**.

$$\omega_i = \frac{\min\{N_i | i \in L_p\}}{N_i} \quad (1)$$

Where ω_i is the weight of label i , N_i is the number of label i in all the training data, L_p is the collection of ground truth label value. For flair-pathway, $L_p = \{0, 2, 4\}$, for t1ce-pathway, $L_p = \{0, 1, 3\}$, for fusion pathway, $L_p = \{0, 1, 2, 3, 4\}$. For the weights of the absent value, we set them to 1. In the similar way, we can calculate the weights for t1ce-pathway and fusion-pathway.

2.4 The Network Training

The network is trained by Keras[8]. The kernel weights were initialized by Xavier function, and stochastic gradient descent (SGD) was used for backpropagation. We set the start learning rate to 0.01 and used inverse learning policy, where the weight decay is set to 0.00005, other parameters use the default setting.

Our proposed method was implemented on GPU server (GeForce GTX TITAN X, RAM 12GB, 8 Intel(R) Core(TM) i7-6700K CPU@4.00GHz).

2.5 Post Process

Post-processing is applied to refine the segment result. We remove the incorrect small islands produced from the network inference stage. Specifically, only the largest volume is kept from the final segmentation, all other small fragments are removed.

3 Experiments and Results

The proposed model is evaluated on BRATS-2018 validation dataset[6, 9] and training dataset. For the training dataset, we randomly select 30 subjects as the testing data, 20 subjects as the validation data and the rest 160 subjects as the training data.

We used the Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) [10] to evaluate the segmentation accuracy. DSC measures the degree of overlap between two ROIs. HD measures the distance between segmentation result and the ground truth. The DSC is calculated by Eq. 2, where $|\cdot|$ denotes the volume of an ROI, S_1, S_2 are two regions in the brain, and \cap denotes the intersection operator. The HD between set A and B is calculated using Eq. 3, where $\|a - b\|$ is Euclidean Distance.

$$DSC(S_1, S_2) = \frac{2 \times |S_1 \cap S_2|}{(|S_1| + |S_2|)} \quad (2)$$

$$HD(A, B) = \max(h(A, B), h(B, A)) \quad (3)$$

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$$

We have submitted preliminary results to the competition leaderboard. **Tab. 2** shows the results on 30 cases of training dataset and 66 cases of validation dataset. The DSC of the whole tumor on both training dataset and validation dataset is same, while the DSC of enhancing tumor and tumor core on validation is 10% and 14.7% lower than the DSC on the training dataset. For the Hausdorff Distance, the value of the whole tumor is smaller on the validation dataset, but the value of enhancing tumor and tumor core is larger on the validation dataset. The main reason for the bad performance on validation dataset is that in some cases of validation dataset, the regions of tumor core and enhance tumor is very small, a small wrong prediction will cause a shape decrease in DSC. In some cases, the DSC even become zero.

As we do not have the ground truth of validation data, we just show the segmentation on training dataset in **Fig. 3**. Since ground truth is manually labeled, the discontinuity error might occur between adjacent slices. However, our segmentation is smoother than the ground truth, this is more biological feasible. Our method doesn't reproduce this discontinuity error.

The average time for processing one image is about one minute. This is a reasonable processing time for clinical usage.

Table 2. The validation result on 30 cases of training dataset and 66 cases on the validation dataset. DSC and 95th percentile of HD for whole tumor (WT), enhance tumor (ET) and tumor core (TC)

	<i>DSC(%)</i>			<i>HD(mm)</i>		
	WT	ET	TC	WT	ET	TC
Training	85.5	82.3	87.3	9.5	5.0	6.9
Validation	85.6	72.0	72.6	7.5	5.7	9.5

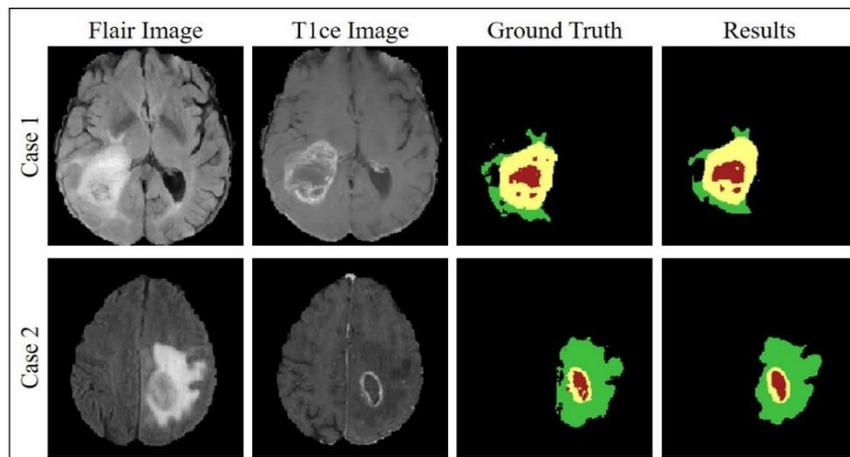


Fig. 3. The segmentation results in the training dataset. Our segmentation results are smoother than the ground truth. The green color in the whole tumor is peritumoral edema, the yellow color is the enhance tumor and the red color is the necrotic components.

In the proposed method, we remove the incorrect small islands produced from the network inference stage to refine the segmentation result. The average DSC is improved 4% in whole tumor and almost 2% in both tumor core and enhance tumor on the training dataset. **Fig. 4** shows the results before post-process and after post-process. We can clearly see that the post-processing procedure can remove the significant mis-segmentation.

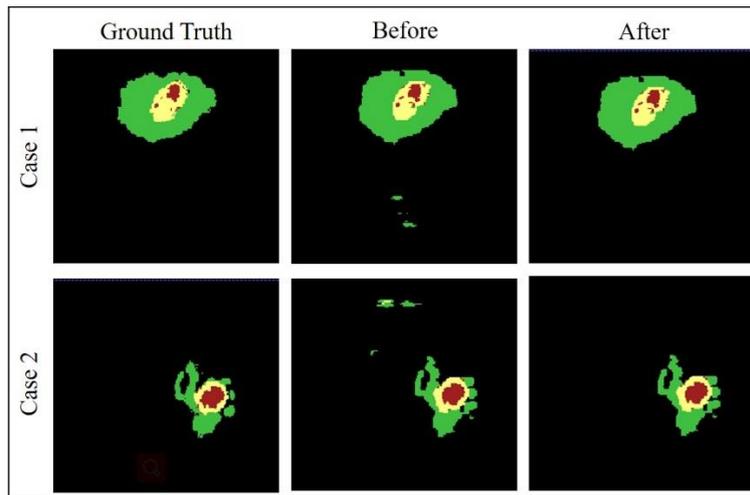


Fig. 4. The comparison between before post-processing and after post-processing with ground truth. The green color in the whole tumor is peritumoral edema, the yellow color is the enhance tumor and the red color is the necrotic components.

4 Conclusion

In this paper, we propose a fully automatic three pathways U-Net for brain tumor segmentation from two modality MRI data. Our model receives 2D slices extracted from the dataset volumes. A novel weights calculation method is used to help to train the model and a post-processing procedure helps to refine the segmentation results. The average DSC of the whole brain, tumor core and enhancing part on BRATS-2018 validation dataset is 85.6%, 72.6%, and 72.2%, respectively. While the average DSC of the whole brain, tumor core and enhancing tumor on training dataset can reach 85.5%, 82.3%, and 87.3%. In the future work, we will put more efforts on improving the segmentation ability in small regions so that the segmentation accuracy will improve in the validation dataset.

Reference

1. Havaei, M., et al., *Brain tumor segmentation with deep neural networks*. Medical image analysis, 2017. **35**: p. 18-31.

2. Pereira, S., et al., *Brain tumor segmentation using convolutional neural networks in MRI images*. IEEE transactions on medical imaging, 2016. **35**(5): p. 1240-1251.
3. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *International Conference on Medical image computing and computer-assisted intervention*. 2015. Springer.
4. Çiçek, Ö., et al. *3D U-Net: learning dense volumetric segmentation from sparse annotation*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2016. Springer.
5. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "*Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection*", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
6. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "*Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features*", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
7. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Γ, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "*The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)*", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
8. Chollet, F., *Keras*. 2015.
9. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "*Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection*", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
10. Taha, A.A. and A. Hanbury, *Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool*. BMC medical imaging, 2015. **15**(1): p. 29.

Brain Tumor Segmentation using an Ensemble of 3D U-Nets

Xue Feng^{1[0000-0002-2181-9889]}, Nicholas Tustison² and Craig Meyer^{1,2}

¹ Biomedical Engineering, University of Virginia, Charlottesville VA 22903, USA

² Radiology & Medical Imaging, University of Virginia, Charlottesville VA 22903, USA
xf4j@virginia.edu

Abstract. Accurate segmentation of different sub-regions of gliomas including peritumoral edema, necrotic core, enhancing and non-enhancing tumor core using multimodal MRI scans has important clinical relevance in diagnosis and treatment. However, due to the highly heterogeneous appearance and shape, segmentation of brain tumors is very challenging. Recent development using deep learning models has proved its effectiveness in the past few brain segmentation challenges as well as other semantic and medical image segmentation problems. Most models in brain tumor segmentation use a 2D/3D patch to predict the class label for the center voxel. Variant patch sizes and scales are used to improve the model performance. U-Net is a widely used network structure for end-to-end segmentation and can be used on entire image or extracted patches to provide classification labels over the entire input voxels. Instead of picking the best network structure, an ensemble of multiple models, trained on different dataset or different hyper-parameters, can generally improve the segmentation performance. In this study we propose to use an ensemble of 3D U-Nets with different hyper-parameters for brain tumor segmentation. Preliminary results showed effectiveness of this model. In addition, we developed a linear model for survival prediction using extracted imaging and non-imaging features, which, despite the simplicity, can effectively reduce overfitting and regression errors.

Keywords: Brain Tumor Segmentation, Ensemble, 3D U-Net, Deep Learning, Survival Prediction

1 Introduction

Gliomas are the most common primary brain malignancies, with different degrees of aggressiveness, variable prognosis and various heterogeneous histological sub-regions, i.e. peritumoral edema, necrotic core, enhancing and non-enhancing tumor core. The Multimodal Brain Tumor Segmentation Challenge (BraTS) 2018 utilizes multi-institutional pre-operative MRI scans and focuses on the segmentation of intrinsically heterogeneous brain tumors [1-2]. The dataset used in this challenge includes multiple-institutional clinically-acquired pre-operative multimodal MRI scans of glioblastoma

(GBM/HGG) and low-grade glioma (LGG) containing a) native (T1) and b) post-contrast T1-weighted (T1Gd), c) T2-weighted (T2), and d) Fluid Attenuated Inversion Recovery (FLAIR) volumes [3-4]. 285 training volumes with annotated GD-enhancing tumor, peritumoral edema and necrotic and non-enhancing tumor and the overall survival data defined in days are provided. 46 volumes are used for validation with the following two goals: 1) provide pixel-by-pixel label maps for the three sub-regions and background; 2) estimate the survival days.

Convolutional neural network (CNN) based models have proven their effectiveness and superiority over traditional medical image segmentation algorithms and are quickly becoming the mainstream in BraTS challenges. Due to the highly heterogeneous appearance and shape of brain tumors, small patches are usually extracted to predict the class for the center voxel. To improve model performance, multi-scale patches with different receptive field sizes are often used in the model [5]. In contrast, U-Net is a widely used convolutional network structure that consists of a contracting path to capture context and a symmetric expanding path that enables precise localization with 3D extension [6-7]. It can be used on the entire image or extracted patches to provide class labels for all input voxels when padding is used. Instead of picking the best network structure, an ensemble of multiple models, trained on different dataset or different hyper-parameters, can generally improve the segmentation performance due to the averaging effect. In this study we propose to use an ensemble of 3D U-Nets with different hyper-parameters trained on non-uniformly extracted patches for brain tumor segmentation. During testing, a sliding window approach is used to predict class labels with adjustable overlap to improve accuracy. With the segmentation labels, we will develop a linear model for survival prediction using extracted imaging features and provided non-imaging features since the linear models can effectively reduce overfitting and thus regression errors.

2 Methods

For the brain tumor segmentation task, the steps in our proposed method include pre-processing of the images, patch extraction, training multiple models using a generic 3D U-Net structure with different hyper-parameters, deployment of each model for full volume prediction and final ensemble modeling. For the survival task, the steps include feature extraction, model fitting, and deployment. Details are described as follows.

2.1 Image Pre-processing

To compensate for the MR inhomogeneity, the bias correction algorithm based on N4ITK was first applied to the T1, T1Gd images, T2 and flair images [8], followed by an adaptive non-local means denoising method [9]. The implementations on ITK [10] were used with a Python wrapper from Nipype [11]. Python-based parallel execution

with multiple threads execution was used to accelerate the two steps. Fig. 1 shows the original T1 image (left), image with only bias correction (center) and image with bias correction and denoising (right). The signal-to-noise ratio (SNR) of the image is increased with the denoising method.

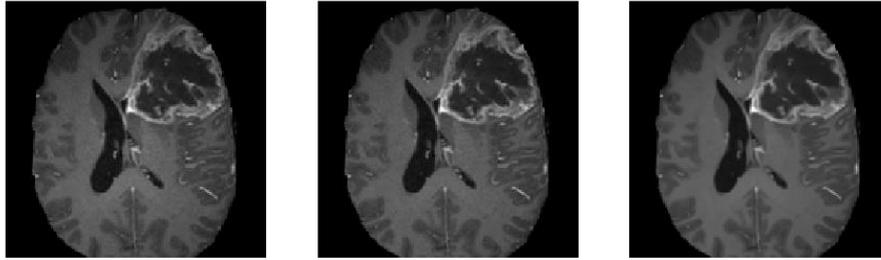


Fig. 1. Original T1 image (left), image with only bias correction (center) and image with bias correction and denoising (right). The right image has improved SNR.

As MR images do not have standard pixel intensity values, to reduce the effects from different contrasts and different subjects, each 3D image was normalized to 0 to 1 by subtracting the min values and divided by the pixel intensity range. For each subject, images of all contrast were fused to form the last dimension so that the whole input image size becomes $155 \times 240 \times 240 \times 4$.

2.2 Non-uniform Patch Extraction

For simplicity, we will use foreground to denote all tumor pixels and background to denote the rest. There are several challenges in directly using the whole images as the input to 3D U-Net: 1) the memory of a moderate GPU is often 12 Gb so that in order to fit the model into the GPU, the network needs to greatly reduce the number of features and/or the layers, which often leads to a significant drop in performance; 2) the training time is greatly prolonged since more voxels contribute to calculation of the gradients at each step and the number of steps cannot be proportionally reduced during optimization; 3) as the background voxels dominate the whole image, the class imbalance will cause the model to focus on background if trained with uniform loss, or prone to false positives if trained with weighted loss that favors the foreground voxels. Therefore, to more effectively utilize the training data, smaller patches with size $64 \times 64 \times 64 \times 4$ were extracted from each subject. As the foreground labels contain much more variability and are difficult to segment, more patches from the foreground voxels should be extracted.

In implementation, during each epoch, a random patch was extracted from each subject using non-uniform probabilities. The valid patch centers were first calculated by removing edges to make sure each extracted patch was completely within the whole

image. The probability of each valid patch center $p_{i,j,k}$ was calculated using the following equation:

$$p_{i,j,k} = \frac{s_{i,j,k}}{\sum_{i,j,k} s_{i,j,k}} \quad [1]$$

in which $s_{i,j,k} = 1$ for all voxels with maximal intensity lower than 1 percentile, $s_{i,j,k} = 6$ for all foreground voxels and $s_{i,j,k} = 3$ for the rest. The patch center was randomly selected based on the calculated probability and the corresponding patch was extracted. Since normal brain images are symmetric along the left-right direction, a random flip was made after patch extraction. No other augmentation was applied.

Before training, the per-input-channel mean and standard deviation of extracted patches were calculated by running the extraction process 400 times, with each time using a randomly selected training subject. The extracted patches were then subtracted with the mean and divided by the standard deviation along each input channel.

2.3 Network Structure and Training

A 3D U-Net based network was used as the general structure. Zero padding was used to make sure the spatial dimension of the output is the same with the input. For each encoding block, a VGG like network with two consecutive 3D convolutional layers with kernel size 3 followed by the activation function and batch norm layers were used. The parametric rectilinear function (PReLU), given as:

$$f(x) = \max(0, x) - \alpha \max(0, -x) \quad [1]$$

was used with trainable parameter α as the activation function. The number of features was doubled while the spatial dimension was halved with every encoding block, as in conventional U-Net structure. To improve the expressiveness of the network, 96 features were used in the first encoding block. Dropout with ratio 0.5 was added after the last encoding block. Symmetric decoding blocks were used with skip-connections from corresponding encoding blocks. Features were concatenated to the de-convolution outputs. The extracted segmentation map of the input patch was expanded to the multi-class the ground truth labels (3 foreground classes and the background). Weighted cross entropy was used as the loss function.

The number of encoding/decoding blocks and the weights in the loss function were used as the two tunable hyper-parameters when constructing multiple models. In current implementation, due to constraint in computational resources, only three models were trained: 3 encoding/decoding blocks with uniform weighting, 3 encoding/decoding blocks with weights to be 1.0 for background and 2.0 for each class of foreground voxels, 4 encoding/decoding blocks with the same non-uniform weights.

Training was performed on a Nvidia Titan Xp GPU with 12 Gb memory. 640 epochs were used. As mentioned earlier, during each epoch, only one patch was extracted every subject. Subject orders were randomly permuted every epoch. Tensorflow framework was used with Adam optimizer. Batch size 1 was used during training. A learning rate of 0.0005 was used without further adjustments during training. The total training time was about 60 hours.

2.4 Volume Prediction Using Each Model

Due to the fact that the model cannot fit into the memory, a sliding window approach needs to be used to get the output of the whole input. However, as significant padding was made to generate the output label map with the same size as the input, boundaries were expected to cause problems when directly sliding across the whole image without overlaps. To alleviate this problem, a stride size at a fraction of the window size was used and the output probability was averaged. The window size was chosen to be the same as the training window size $64 \times 64 \times 64$, and the stride was chosen as $32 \times 32 \times 32$. For each window, the original image and left-right flipped image were both predicted, and the average probability after flipping back the output of the flipped input was used as the output. Therefore, each voxel, except for a few on the edge, will be predicted 16 times when sliding across all directions. Although smaller stride sizes can be used, the deployment time will be greatly increased. For the given image size, it takes 1 minutes to generate the output for the entire volume on the same GPU. The probability output was saved for each model.

2.5 Ensemble Modeling

The ensemble modeling process was rather straightforward. The probability output of all classes from each model was obtained and the final probability was calculated via simple averaging. The class with the highest probability was selected as the final segmentation label of each voxel.

2.6 Survival Prediction

To predict the post-surgery survival time measured in days, extracted images features and non-image features were used to construct a linear regression model. 6 image features were calculated from the ground truth label maps during training and the predicted label maps during validation. For each foreground class, the volume by summing up the voxels and the surface area by summing up the magnitude of the gradients along three directions were obtained. Age and resection status were used as non-image features. As there were two classes of resection status and many missing values of this status, a two-dimensional feature vector was used to represent the status, given as GTR: (1, 0), STR: (0, 1) and NA: (0, 0). A linear regression model after normalizing the input features to zero mean and unit standard deviation was fit with the training data.

3 Results

3.1 Brain Tumor Segmentation

All 285 training subjects were used in the training process. 66 subjects were provided as validation. The dice indexes, sensitivities and specificities, 95 Hausdorff distances of the enhanced tumor (ET), whole tumor (WT) and tumor core (TC) were automati-

cally calculated after submitting to the CBICA’s Image Processing Portal. With multiple submissions, we were able to compare the performances of each individual model and the final ensemble.

Table 1. Performances of each individual model and the ensemble

	Dice_ET	Dice_WT	Dice_TC	Dist_ET	Dist_WT	Dist_TC
L3, uniform	0.7688	0.90148	0.82365	4.12703	4.54368	5.5226
L3, nonuniform	0.76773	0.90662	0.82484	4.22177	6.46369	8.85934
L4, nonuniform	0.77068	0.89902	0.81043	3.14539	6.00811	6.98141
Ensemble	0.78734	0.90587	0.83359	3.96414	4.01843	5.34027

Table 1 shows the mean dice scores and 95 Hausdorff distances of ET, WT and TC for the three individual models and the ensemble. Sensitivity and specificity are highly correlated with the dice indexes so that they are not included. Comparing the non-uniform weighted with the uniform weighted loss while the network structures are the same (L3), as the foreground labels have higher weights, the WT dice score increased from 0.901 to 0.907, but the distance also increased from 4.54 to 6.46. The ensemble of all models has the overall best performances as compared with each individual model.

3.2 Survival Prediction

All 163 training subjects with survival data were used in the training process. 28 cases were evaluated after submitting to the CBICA’s Image Processing Portal. The accuracy was 0.321, MSE was 97997.5, median SE was 76560.9, std SE was 102916.3 and Spearman Coefficient was 0.278.

4 Discussion and Conclusions

In this paper we developed a brain tumor segmentation method using an ensemble of 3D U-Nets. Bias correction and denoising were used as pre-processing. Three networks were trained with different number of encoding/decoding blocks and different weights for loss. The preliminary results showed an improvement with ensemble modeling. In addition, we used a simple linear regression for survival prediction by combining imaging and non-imaging features.

It is noted that the median metrics are significantly higher than the mean metrics. For example, the median dice indexes are 0.867, 0.924 and 0.904 for ET, WT and TC. It makes sense in that the theoretical maximum dice index is 1 and minimum dice index is 0. However, we noted that in several cases, the dice indexes are 0 for ET and TC and 0.6 for WT. It is mostly due to the low sensitivity meaning that the model is not able to recognize the tumor regions. The possible reason for these failed regions is that their characteristics deviate a lot from the training dataset.

In the 3D U-Net model, we found that the batch norm layer was helpful in improving the model stability and performance. However, different with the canonical application of the batch norm layer, in which the batch statistics is used in training and the global

statistics is used in deployment, it performed much better with batch statistics in deployment than global statistics. Since the batch size is 1, a per-channel normalization is actually performed by subtracting its own mean. One possible explanation could be that by doing such normalization, the model focuses on the differences of neighboring pixels in one channel and ignores the absolute values, which may help the segmentation process. However, further investigation is needed to figure out the exact reason.

Compared with the patch-based model that only predicts the center pixel, when predicting the segmentation label maps for the full patch, different pixels are very likely to have different effective receptive field sizes due to the zero padding in the edge. We argue that a pixel should still be able to be predicted even based on partial receptive field, which, for the very edge pixel, corresponds to only half of the receptive field. Furthermore, the U-Net structure learns an optimized receptive field with multiple encoding and decoding paths and the connections in between and thus is superior than the multi-scale model. Furthermore, the significant overlap in prediction sliding windows can improve the prediction accuracy with more averages.

In the current implementation, only three networks were trained due to limitations in computation time. It is expected with more networks, the results can be further improved, although the marginal improvement is expected to decrease.

For the survival task, since it is very likely to overfit with such a small dataset, we used a linear regression model to minimize the test errors, although at the cost of the expressiveness of the model. Further exploration of imaging features is expected to improve the predicting power of the model.

In conclusion, we developed an ensemble of 3D U-Nets for brain tumor segmentation. The network hyper-parameters are varied to obtain multiple trained models. A linear regression model was also developed for the survival prediction task. The code is available at <https://github.com/xf4j/brats18>.

References

1. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftikharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015)
2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, (2017) [In Press]
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q

4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
5. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B. "Efficient Multi-Scale 3D CNN with fully connected CRF for Accurate Brain Lesion Segmentation", Medical Image Analysis 36, 61-78 (2017)
6. Ronneberger O, Fischer P, Brox T. "U-Net: Convolutional Networks for Biomedical Image Segmentation", arXiv:1505.04597 (2015)
7. Cicek O, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation", arXiv:1606.06650 (2016)
8. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC. "N4ITK: Improved N3 Bias Correction", IEEE Trans Med Imaging. 29(6), 1310-1320 (2010)
9. Manjon JV, Coupe P, Marti-Bonmati L, Collins L, Robles M. "Adaptive Non-Local Means Denoising of MR Images With Spatially Varying Noise Levels", J Magn Reson Imaging. 31(1), 192-203 (2010)
10. ITK Homepage, <https://itk.org/>
11. Nipype Homepage, <https://nipype.readthedocs.io/en/latest/>

Brain Tumor Detection and Segmentation Using Deep Learning U-Net on Multi Modal MRI

Naomi Fridman

Afeka Academic College of Engineering, Tel Aviv, Israel
Naomi.Fridman@s.afeka.ac.il

Abstract. In this study, we propose a fully automatic method for brain tumor segmentation, which is developed using U-Net based deep convolutional networks. U-net is a Deep Neural network architecture, based on fully connected Convolutional Neural Network. U-net has been proven to be a dynamic robust tool for biomedical image segmentation. U-net can be built as 3D [11][12] neural network or 2D. Our approach to the segmentation problem is to feed a 2D U-net network with 2D planes along the z axis. The classification to tumor classes is achieved with an architecture of 4-channel output, where a softmax is applied to reassure probabilities. I report preliminary, results obtained using BraTS 2018 Training dataset [1-4]. The conclusion of this work is that U-net have potential to perform well on the segmentation task. A combinations of few models, is needed to improve the results even further.

Keywords: Segmentation, BraTS challenge, U-net, Convolutional neural networks, Deep learning

1 Introduction

In this project, I investigate the potential 2D U-net to segment the Glioma tumor on an MRI modality scan. Gliomas are a type of brain tumor that can be deadly, and accurate segmentation of the tumor and its sub regions, is needed for treatment planning and surgery. The large amount of data of one MRI scan, meaning 4 models of 155 images of 240x240 pixels. To analyze and annotate this amount is time consuming task for human expert. Automatic tumor segmentation can improve medical treatment by providing an efficient and annotation of tumor and its sub-region location. Deep learning U-net neural network, has proven lately as efficient tool for many biomedical segmentation tasks. The nature of most medical images, that have high correlation among near voxel, make them good candidate for analyzing with convolution sliding window.

2 Methods and Materials

2.1 MRI Dataset

The Training dataset of BraTS 2017 [1,2,3,4] comprises multimodal MRI scans of glioblastoma (GBM/HGG, 135) and lower grade glioma (LGG, 108). For each patient, there are four MRI scans as described: T1-, contrast enhanced T1- (T1c), T2- and T2-weighted FLAIR. The multimodal MRI scans, were acquired with different clinical protocols and various scanners from multiple (n=19) institutions, all images were pre-processed, meaning: co-registered to the same anatomical template, interpolated to the same resolution and skull-stripped.

All Images have been segmented and approved manually by experts. Annotations comprise the GD-enhancing tumor (ET - label 4), the peritumoral edema (ED - label 2), necrotic (NCR/NET - label 1) and non-enhancing tumor (label 3).

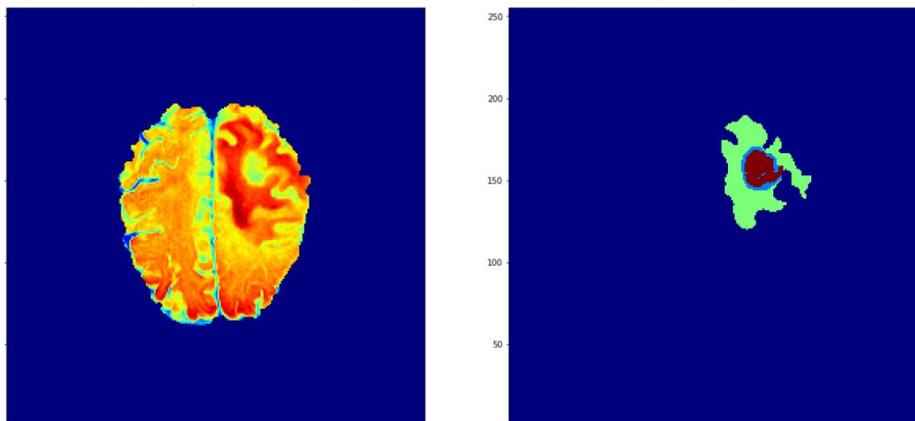


Fig. 1: MRI slice, Flair img with 4 class labeling

2.2 MRI Pre-processing

MRI data often contain artifacts produced by inhomogeneity in the magnetic field or small movements made by the patient during scan time, this often creates a bias across the resulting scans, which can affect the segmentation results particularly made by computer-based models. To correct that, I employed SimpleITK [9] N4 bias field correction filter [10] on all T1 and T1C images in the dataset. Since MRI intensities are expressed in arbitrary units and may vary between different machines, additional image pre-processing was made to standardize the voxel intensities, so each sequence was transformed to have zero mean and unit standard deviation.

2.3 U-Net: Convolutional Networks

U-Net is a very popular end-to-end encoder-decoder network for semantic segmentation [7]. It was originally invented and first used for biomedical image segmentation. U-net's have multi channel architecture, which suits well the multi channel input of MRI images and the multi-class classification task.

Medical MRI images feature a high similarity and correlation in the intensities among neighboring voxels, so they are good candidates for the convolution blocks constructing the U-net.

Essentially, U-net is a deep-learning framework based on fully convolutional networks [14], it comprises two parts:

1. A contracting path similar to an encoder, to capture context from a compact feature representation..
2. A symmetric expanding path similar to a decoder, which allows accurate localisation. This step is done to retain boundary information (spatial information) despite down sampling and max-pooling performed in the encoder stage.
3. Instead of up-sampling methods,in U-net, we concatenate the suitable transformed layer from the decoder path.

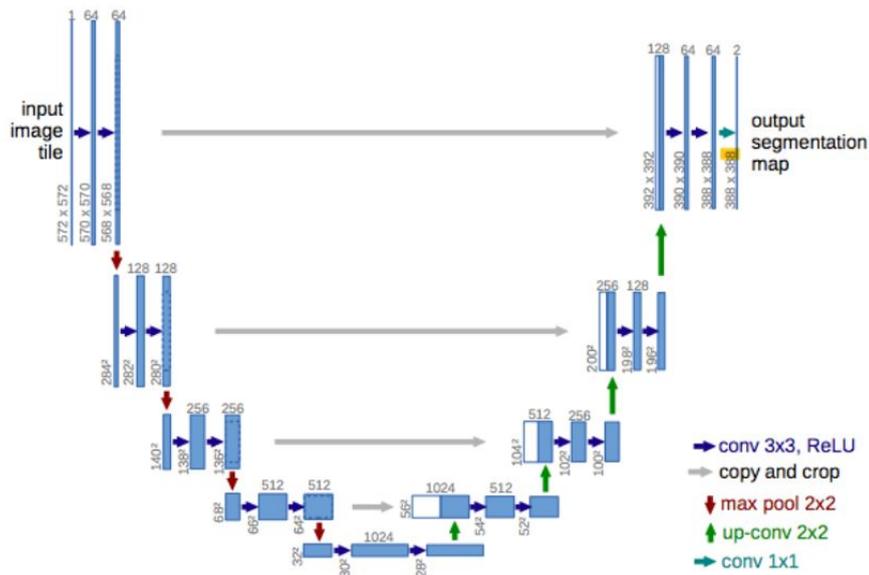


Fig.2: Architecture of U-Net based on the paper by [Olaf Ronneberger\[7\]](#) et.al

2.4 Model Training

The 2D U-net convolutional network, we used 4 channels for the input, holding all MRI modalities. For output we tried 4 channels, each stand for classification of the tumor. It didn't perform well and overfit the train data, probably from lack of enough data. So we used separate learning process for each label, and decide the right label by max on he predictions. Input data was 2D planse along Z-axis, and output data was segmented planes along Z-axis. Training converged slowly and steady, in the first epochs, improvement was around 0.002 in dice coefficient,where each epoch consisted 4800 slices.

Loss function

During the training process, Dice score lost function was used. Sørensen's original Dice formula was intended to be applied to binary data. Given two sets, X and Y, it is defined as:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

where $|X|$ and $|Y|$ are the cardinalities of the two sets (i.e. the number pixels segmented as belonging to the class, in each set). The Sørensen index equals twice the number of pixels common to both sets divided by the sum of the number of elements in each set. To use Dice score for a multi-class classification, we flatten the prediction and the ground truth, to a 1D binary vector, where each section of the sequence, stands for a different class.

Sampling

We used random batch generator,that chose samples according to labels. To stain single label segmentation, we used simple planes with then 4 pixels of that label. For training of full label segmentation, we used batch's, each contained 20% slices with enhancing tumor (ET - label 4), and 30% slices including peritumoral edema (ED - label 2) and 4 enhancing tumor (ET - label 4). This was chosen due to label distribution, which is 11% of label 1, 54 % of label 2 and 35% of label 3.

Loss function behaves similarly in all training process. It decrease very slowly in the first step (3-5), then few decreasing leaps, then steady slow decrease of around 0.003 per epoch. For example, first few steps of Label 2 segmentation training:

epoch 1: loss:

-0.0436 - dice_coef: 0.0436 - binary_crossentropy: 0.6943

val_loss: -0.0531 - val_dice_coef: 0.0531 - val_binary_crossentropy: 0.6951

epoch 2

loss: -0.0454 - dice_coef: 0.0454 - binary_crossentropy: 0.7070

val_loss: -0.0588 - val_dice_coef: 0.0588 - val_binary_crossentropy: 0.7433
 epoch 3
 loss: -0.0604 - dice_coef: 0.0604 - binary_crossentropy: 1.0244
 val_loss: -0.0801 - val_dice_coef: 0.0801 - val_binary_crossentropy: 1.0861
 epoch 4
 loss: -0.1291 - dice_coef: 0.1291 - binary_crossentropy: 0.8191
 val_loss: -0.3246 - val_dice_coef: 0.3246 - val_binary_crossentropy: 1.0624
 Epoch 5
 loss: -0.4431 - dice_coef: 0.4431 - binary_crossentropy: 0.4999
 val_loss: -0.4656 - val_dice_coef: 0.4656 - val_binary_crossentropy: 0.5125
 We used 64 images per batch, and 40 batches per epoch. Tests on smaller batches, like 32, improves fast, but reach plateau very fast as well.. After 120 epoch, we got around 0.5 dice coefficient, and we can clearly see in Fig 3. below, that the segmentation captures correctly around half the of pixels.

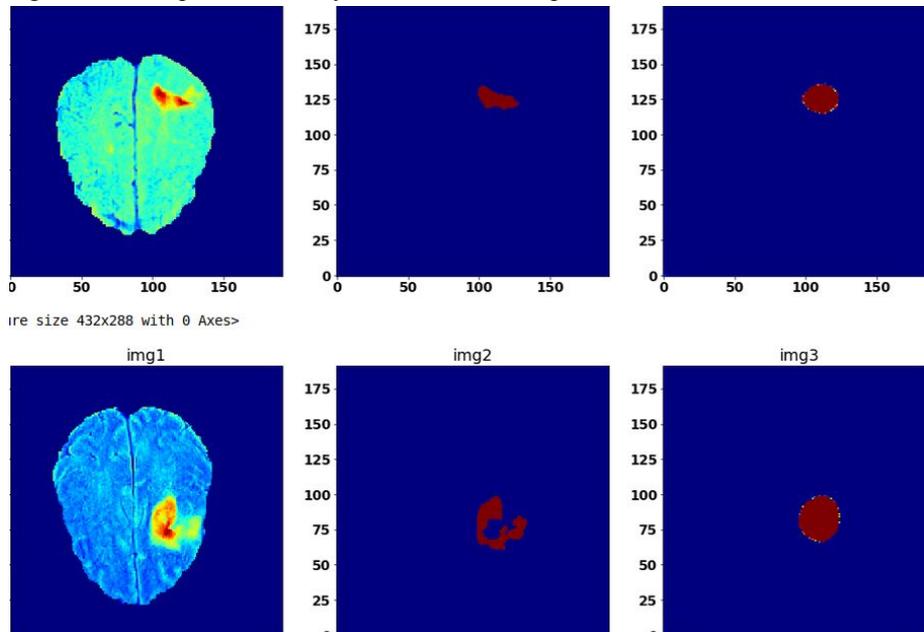


Fig 3. Prediction of label 2 after 120 epoch. Dice coefficient is around 0.5.

3 Results

The performance of U-net's was mixed for this segmentation task. While segmenting peritumoral edema (ED - label 2) very well, the distinction between (ET - label 4) and (NCR/NET - label 1), should be improved.

The U-net perform well, when segmentation had “blobe” structure, rather then sparse structure.

The U-net did not separate well between label 4 and 1, and had bad performance on sparse pixel segmentation.

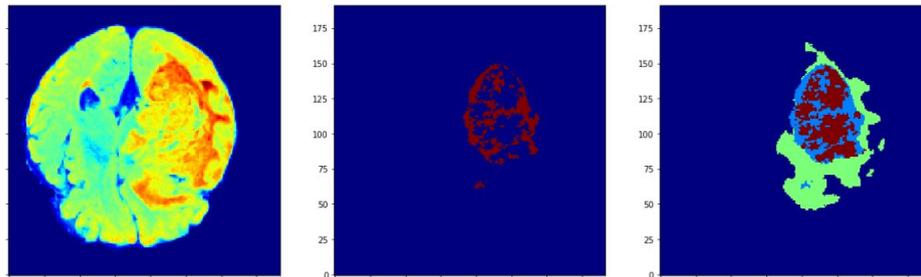
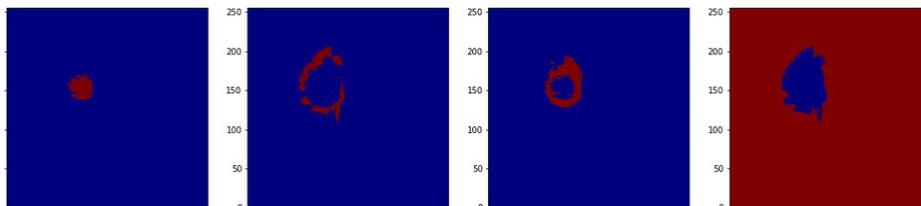


Fig 4. Sparse labeling, perform worse in U-net convolutional network

4 Discussion

U-nets model yield good result in MRI image segmentation task. To improve the model, we suggest to explore three main paths: Preprocessing the data and resampling techniques, the other is optional loss functions, and data augmentation. Improvement of the non-sufficient segmentation results on non-enhancing tumor (label – 3), is needed.

The U-net was not sensitive enough, different architectures, that encodes to smaller size can improve the sensitivity of the segmentation..



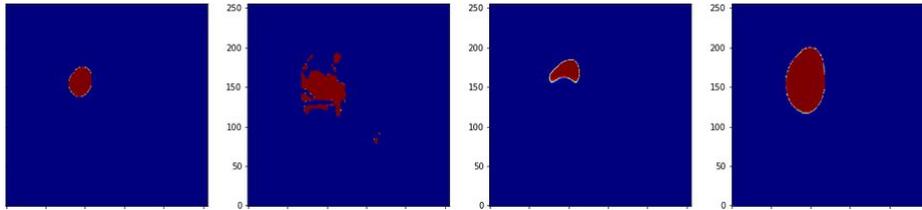


Fig 5. Segmentation of Labels 1, 2, 4 and all tumor. Top line true labels, bottom line are predictions.

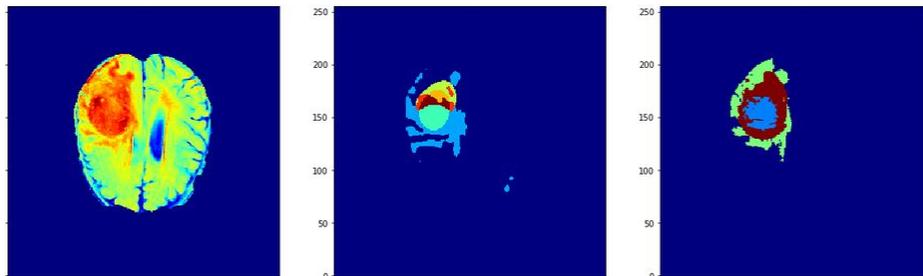


Fig 6. Simple addition of Labels 1, 2, 4 predictions. and True label

5 Conclusion

The performance of U-net's was mixed for this segmentation task. While segmenting peritumoral edema (ED - label 2) very well, the distinction between (ET - label 4) and (NCR/NET - label 1), should be improved. U-net performs well on large blobs of labeled pixels, but performance decrease when the classified pixels create small sparse areas. Combining different technique, as local pixel segmentation, can be refine the results.

Sizes of tumors varies a lot. Large tumors are segmented better than small tumors. Specialized U-net should be built and trained to improve Segmentation results on small tumors.

Choice of loss function and resampling techniques, influence learning significantly.

6 References

1. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117

2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
4. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftikharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015)
5. IL.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L.Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. In ICLR, 2015.
6. Havaei, M. et. al, Brain Tumor Segmentation with Deep Neural Networks. arXiv preprint arXiv:1505.03540, 2015.Author, F., Author, S., Author, T.: Book title. 2nd edn. Publisher, Location (1999).
7. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351: 234--241, 2015, available at arXiv:1505.04597 [cs.CV]
8. Theano Python library, <https://github.com/Theano/Theano>
9. N4BiasFieldCorrectionImageFilter Class Reference, https://itk.org/SimpleITKDoxygen/html/classitk_1_1simple_1_1N4BiasFieldCorrectionImageFilter.html
10. J.G. Sled, A.P. Zijdenbos and A.C. Evans. "A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in Data" IEEE Transactions on Medical Imaging, Vol 17, No 1. Feb 1998.
11. Chengjia Wang, Tom MacGillivray, Gillian Macnaught, Guang Yang, David Newby. "A two-stage 3D Unet framework for multi-class segmentation on full resolution image". 2016. arXiv:1606.06650 [cs.CV]
12. Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, Olaf Ronneberger. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation". 2016. arXiv:1606.06650 [cs.CV]

13. Sørensen, T. (1948). "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons". *Kongelige Danske Videnskabernes Selskab*. **5** (4): 1–34.
14. Jonathan Long, Evan Shelhamer, Trevor Darrell. "Fully Convolutional Networks for Semantic Segmentation". 2015 arXiv:1411.4038 [cs.CV]
15. Keras: The Python Deep Learning library, <https://keras.io/>

Glioma Segmentation and a Simple Accurate Model for Overall Survival Prediction

Evan Gates¹, J. Gregory Pauloski¹, Dawid Schellingerhout², and David Fuentes¹

¹ Department of Imaging Physics

² Department of Cancer Systems Imaging and Diagnostic Radiology,
University of Texas MD Anderson Cancer Center, Houston TX, USA

EGates1@mdanderson.org

DTfuentes@mdanderson.org

Abstract. Brain tumor segmentation is a challenging task necessary for quantitative tumor analysis and diagnosis. We apply a multi-scale convolutional neural network based on the DeepMedic to segment glioma subvolumes provided in the 2018 MICCAI Brain Tumor Segmentation Challenge. We go on to extract intensity and shape features from the images and cross-validate machine learning models to predict overall survival. Using only the mean FLAIR intensity, nonenhancing tumor volume, and patient age we are able to predict patient overall survival with reasonable accuracy. Preliminary challenge results are presented.

Keywords: Glioblastoma · Segmentation · Neural Network · Quantitative Imaging

1 Introduction

Gliomas are highly malignant primary brain tumors that carry a dismal median overall survival of 15 months for high grade tumors [1]. One characteristic that contributes to this poor survival is the substantial heterogeneity and differences between distinct glioma subvolumes. Successful and automated detection of these subvolumes is a key step in quantitative analysis towards patient risk stratification and computer aided diagnosis. In recent years, convolutional neural networks (CNNs) are the undisputed champions of biomedical segmentation tasks [2]. Quantitative measurements of these subvolumes are likely to provide insight into patient’s prognosis.

In this work we use a multi-scale convolutional neural network to segment glioma sub-volumes in multi-contrast MRI images. We go on to extract shape and intensity features from the sub-volumes to predict patient overall survival. Preliminary results on the 2018 MICCAI Brain Tumor Segmentation Challenge [3–6] are provided and compared to previous challenges.

2 Segmentation

2.1 Network Structure

Data Preprocessing The BraTS 2018 Training set contains 285 multi-contrast MRI (T1, T1ce, T2, FLAIR) scans of high and low-grade gliomas. 75 of the 285 patients are labeled low-grade (LGG) and the remaining are high-grade (HGG). The imaging data is brain extracted, registered, and resampled to 1 mm isotropic voxel size. Each subject has a ground truth segmentation with four labels, non-tumor (label 0), necrotic and non-enhancing tumor core (label 1), peritumoral edema (label 2), and Gadolinium-enhancing tumor (label 4). The BraTS 2018 Validation set contains a mix of 66 HGG and LGG patients equivalently pre-processed and does not have ground truth segmentations.

All MRI scans were normalized by subtracting the mean intensity and dividing by the standard deviation. A binary brain mask for each patient was also created using the T1 scan, and this mask is used by the CNN to focus sampling on only the brain. The same preprocessing steps were also applied to the validation data set before segmenting.

Convolutional Neural Network We used a 3-dimensional CNN built using the DeepMedic architecture created by Kamnitsas et al [7]. DeepMedic has consistently produced high performing image segmentations in previous BraTS challenges. Our CNN implementation contains a low and normal resolution pathway consisting of 11 layers each.

We trained the CNN on 95% of the BraTS 2018 Training Data. The remaining 5% was used to validate the training progress of the model. The network was trained for 50 epochs with a batch size of 10, and the RMSprop optimizer was used with an initial learning rate of 0.001 and lowered throughout training. Training on a Nvidia Kepler Titan 6GB took 96 hours and, using the trained CNN, we performed a full inference on the BraTS 2018 Validation Data to segment the images.

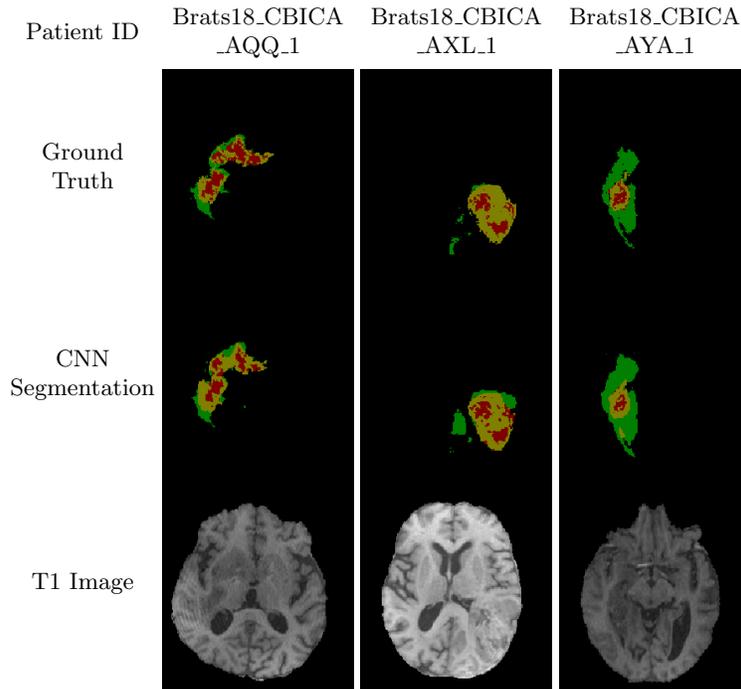
2.2 Training and Validation Set Results

The segmentation results from the full inference on the validation set were uploaded to the BraTS Challenge portal where the Dice score, sensitivity, specificity, and Hausdorff distance were calculated. The preliminary results for the 66 patients in the validation set are shown in Table 1. We also performed a full inference on the 5% of training set cases which were excluded from the model training process, so that we can compare the performance of the model against the ground truth segmentations. This comparison is shown in Table 2. From these samples, we can see that the CNN classifies the overall tumor well but has greater difficulty classifying regions with a dense mix of lower and higher grades.

Table 1. Mean values for metrics from segmentations on the training and validation data sets.

Data Set	Label	Dice	Sensitivity	Specificity	Hausdorff
Training	Enhancing Tumor	0.7332	0.84265	0.99781	6.20545
	Whole Tumor	0.89633	0.88636	0.99531	5.14866
	Tumor Core	0.75292	0.73297	0.99833	8.47618
Validation	Enhancing Tumor	0.67831	0.72923	0.99611	14.52297
	Whole Tumor	0.80558	0.81374	0.98703	14.415
	Tumor Core	0.6852	0.68018	0.99619	20.01745

Table 2. Ground truth versus CNN segmentation samples from the validation data excluded from training. Red is non-enhancing tumor and necrosis, green is edema, and yellow is enhancing tumor. The T1 weighted image for each patient is shown for reference.



3 Survival Analysis

Of the 285 training cases, 163 cases had age (range 19-86 years) and overall survival data (range 5-1767 days) provided. We used this clinical data and features extracted from segmentations to predict the overall survival in days.

After determining the best model, we applied it to the 28 challenge validation set cases with patient age and gross total resection status.

The challenge assesses the predictions based on the accuracy: total number of cases correctly assigned a survival < 10 months, between 10 and 15 months, and > 15 months. The mean-square-error (MSE) is also used as a performance metric.

3.1 Image Processing

The format of the provided imaging data is described in section 2.1. For the survival task, we further pre-processed the data by normalizing based on reference tissue intensities. Note, this is different than the normalization used in the segmentation task. Creating a consistent intensity scale between patients allows images features to discriminate short and long survival patients. To apply this normalization, we placed small regions of interest for each patient in the gray matter (GM) of the lentiform nucleus, the cerebrospinal fluid (CSF) of the ventricles, and the normal appearing white matter (WM). Using the mean intensity for a pair of reference tissues, each voxel in the image was linearly scaled to map the mean intensities to 0 and 1 respectively. For example, in the FLAIR image each voxel value x was transformed according to

$$y_{CSF/WM} = \frac{x - \overline{CSF}}{\overline{WM} - \overline{CSF}}$$

For the FLAIR image CSF and WM were chosen because they were the darkest and brightest reference tissues respectively. For sequences T1 and FLAIR we normalized using the CSF/WM pair, for T2 we used the WM/CSF pair, and for the T1 contrast enhanced image we used a CSF/GM pair. This procedure is similar to other methods presented in the literature [8] Although we performed the normalization semi-automatically with manually placed ROIs, this procedure can be performed fully automated using brain tissue segmentation software applied to the non-tumor regions.

3.2 Image Features

To predict patient overall survival we calculated image features for each of the available image sequences and segmentation labels from section 2. We also computed the union of the three regions (nonenhancing, enhancing, and edema) to generate a whole-tumor ROI for each patient. For each image and region pair we calculated the mean intensity using the *Pyradiomics* software package [9]. We also calculated the volume of each region, so the total number of features considered is.

$$\begin{aligned}
4 \text{ regions} \times 4 \text{ images} &= 16 \text{ mean values} \\
&+ 4 \text{ region volumes} \\
&= 20 \text{ image features per patient}
\end{aligned}$$

We initially experimented with features quantifying higher order histogram statistics (quantiles, skewness, etc) and complex shape descriptors (i.e. flatness) but found these features did not improve the performance of predictive modeling beyond using just mean values. Similarly, we quantified image texture using gray level co-occurrence matrices and gray level run length matrices, and nearest gray tone difference matrices but again found that including these features did not substantially increase model performance.

3.3 Survival Task

Our approach to overall survival follows a standard machine learning approach to select the best performing model and estimate the generalization error using the training data. An overview is shown in Figure 1. The best model was used to make predictions on the provided validation data. Modeling for the survival task was implemented in R version 3.4.0.

We partitioned the training data into 80% training and 20% testing data with an approximately equal proportion of short, medium, and long survivors in each set. Using the 80% partition, we performed variable selection and trained several classes of predictive models using leave-one-out cross validation. We selected the model with the highest Pearson correlation (R^2) between predicted and observed overall survival within the cross validation and made predictions on the testing set to see how well the model generalized.

For feature selection we consecutively applied univariate, multivariate, and step-wise feature elimination. After each selection step the resulting variables were stored as a possible set of inputs to predictive models. First, we used a Cox model to individually determine which image features were significantly associated with overall survival. Any feature with $p > 0.05$ for the Wald-test was discarded. Next, the remaining features were fed into a multivariate Cox model to reduce redundancy. Features with $p < 0.05$ in the multivariate Cox model were retained. Lastly, we further reduced the set of inputs using step-wise elimination based on Akaike Information Criteria (AIC) [10]. Starting with all variables, the algorithm eliminates or replaces variables one at a time to maximize the AIC.

In addition, we applied the Boruta method [11] to select variables predictive of overall survival. The Boruta method is based on variable importance from the random forest algorithm, which has traditionally been a top performing machine learning model.

After selecting the best model and variable combination, we trained a final model on all the training data, made predictions on the test set, and compared the performance to the leave-one-out cross validation. In particular, we checked

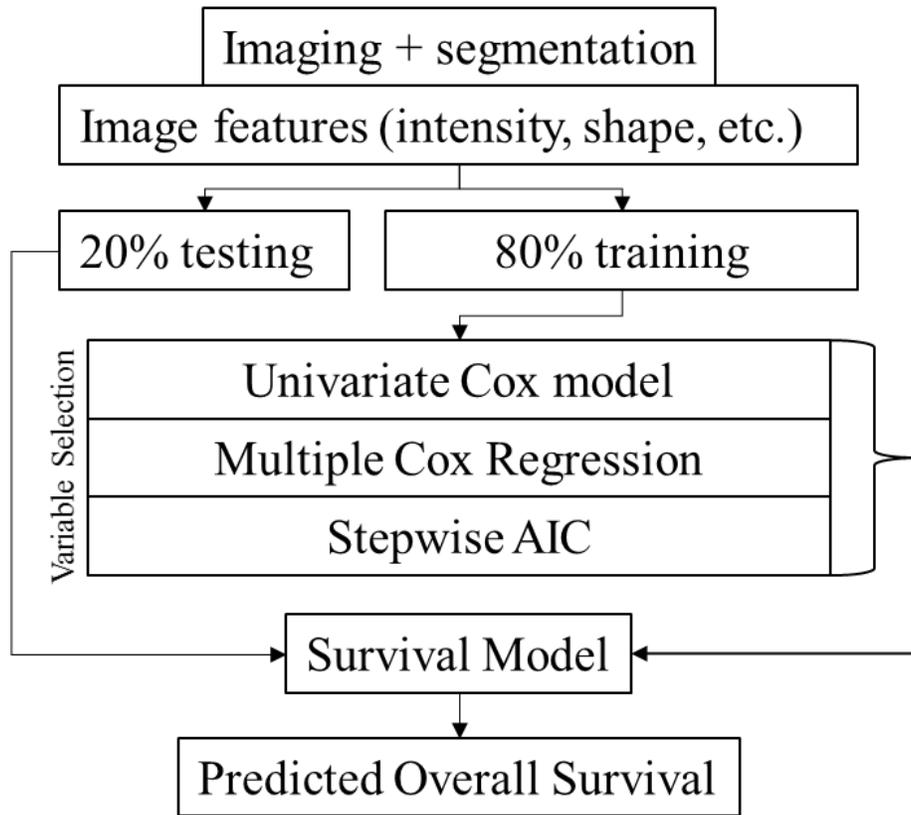


Fig. 1. Flowchart depicting the modeling process and variable selection methods for predicting overall survival. Several variable selection methods based on the cox model are used to generate input sets for predictive models. Before any variable selection 20% of the training data is held out for testing and the remaining 80% is used to select the best variable and model combination. We tested linear models, neural networks, and random forests to predict overall survival. Patient age, an important clinical factor, was always included.

for substantially decreased performance on the test data that would indicate model overfitting.

3.4 Results

Landmark normalization was successfully applied to all patients. One case had poor fluid suppression on the FLAIR image and could not be effectively landmark normalized. This case was excluded from the training.

Among the mean intensities for each image over each region, the region volumes, and the patient age, we found five features were significantly associated with overall survival in the cox model. They are: mean FLAIR intensity over the nonenhancing and necrotic region, mean T1ce intensity over the whole tumor, the volumes of the nonenhancing and enhancing regions, and age. With these variables input into a multivariate Cox model only age, the FLAIR nonenhancing mean, and nonenhancing volume were independently significant. Applying stepwise AIC did not change the variable selections any further.

Among the candidate models we tested (random forest, neural network, linear model) the linear model performed best with $R^2 = 0.134$ and mean-square-error 114994 using the three inputs selected by the multivariate cox model. With the same model parameters fit to all 162 evaluable challenge cases the model to predict overall survival is given by.

$$\text{Survival} = 926.8 - 10.5 \cdot \text{Age} + 91.6 \cdot \text{FNNV} - 55.1 \cdot \text{NNV}$$

Where Age is the age in years, FNNV is the “FLAIR Nonenhancing and Necrosis Mean” value on the landmark normalized scale, and NNV is the “Nonenhancing and Necrosis Volume” in units of $\text{mm}^3/10000$. This volume scaling makes the range of values comparable to the other features. Surprisingly, this simple linear model performed substantially better on the testing data and on the challenge validation dataset. This strongly suggests the model is not over fitting the data. The metrics are provided in table 3

Table 3. Performance metrics for our linear model on the training data: (80% of 163 provided cases), testing data (20% of 163 provided cases), and validation data (26 cases without known survival). The Pearson R^2 for the validation data is not provided.

	R^2 , predicted vs observed	Accuracy	MSE
Training Data	0.134	44.5%	114994
Testing Data	0.399	38.2%	55193
Validation Data	-	53.6%	87998

At the time of this writing, this model has the lowest mean square error of any participant team posted on the validation leaderboard. The accuracy (53.6%) is also the third highest accuracy posted.

4 Discussion

Brain tumor segmentation and prediction of overall survival are both challenging tasks. Despite good results, our segmentation model did not perform as well as the implementation of DeepMedic by Kamnitsas et al. that won the BraTS 2017 challenge [12]. Their model achieved better segmentation results by averaging results across an ensemble of six different models. The single model we used is not as robust as their ensemble method but provides satisfactory results without the high computational cost.

In the task of predicting patients as short, medium, or long survivors we achieved a validation accuracy of 54% with a MSE of 87998. In the training data the most frequent class is short survivors at 65 of 163 (39.9%) which means our models are performing better than chance. The root mean square error for continuous prediction is on the order of 300 days, which is comparable to the range seen among all patients. Overall survival is impacted by several factors, including age, treatment, and performance status (not provided) and the accuracy and MSE reflects the complexity of this task even when some variables are controlled for.

We were able to produce good results using two highly primitive image measurements (mean intensity and volume) and a linear regression model. Although vast numbers of higher-order texture features and nonlinear models are commonly employed in quantitative imaging tasks, we found they were not useful in predicting overall survival for this task. We suspect this is because these features are more sensitive to tumor segmentation (and segmentation error) as well as other variations in image quality and processing. Since predicting overall survival is already a highly uncertain task, it is easy for models to over-fit the higher order features. In other words, the simple and robust features more easy to generalize.

5 Conclusion

We found we could segment glioma tumors with high accuracy using a multi-scale convolutional neural net. Using these segmentations and simple image features we were able to predict overall survival with reasonable accuracy.

References

1. Roger Stupp, Warren P Mason, Martin J van den Bent, Michael Weller, Barbara Fisher, Martin J B Taphoorn, Karl Belanger, Alba A Brandes, Christine Marosi, Ulrich Bogdahn, Jürgen Curschmann, Robert C Janzer, Samuel K Ludwin, Thierry Gorlia, Anouk Allgeier, Denis Lacombe, J Gregory Cairncross, Elizabeth Eisenhauer, and René O. Mirimanoff, "the European Organisation for Research and Treatment of Cancer Brain Tumor and Radiotherapy Groups", and "the National Cancer Institute of Canada Clinical Trials Group". Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *N. Engl. J. Med.*, 352(10):987–996, 2005.

2. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation BT - Medical Image Computing and Computer-Assisted Intervention MICCAI 2015. pages 234–241, Cham, 2015. Springer International Publishing.
3. Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, and Others. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging*, 34(10):1993–2024, 2015.
4. Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data*, 4:170117, sep 2017.
5. S Bakas, H Akbari, A Sotiras, M Bilello, M Rozycki, J Kirby, J Freymann, K Farahani, and C Davatzikos. “Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection”, The Cancer Imaging Archive, 2017, DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q.
6. S Bakas, H Akbari, A Sotiras, M Bilello, M Rozycki, J Kirby, J Freymann, K Farahani, and C Davatzikos. “Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection”, The Cancer Imaging Archive, 2017, DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF.
7. Konstantinos Kamnitsas, Christian Ledig, Virginia F.J. Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis*, 36:61–78, 2017.
8. Kelvin K Leung, Matthew J Clarkson, Jonathan W Bartlett, Shona Clegg, Clifford R Jack Jr, Michael W Weiner, Nick C Fox, Sébastien Ourselin, and the Alzheimer’s Disease Neuroimaging Initiative. Robust atrophy rate measurement in Alzheimer’s disease using multi-site serial MRI: Tissue-specific intensity normalization and parameter selection. *NeuroImage*, 50(2):516–523, 2010.
9. Joost J M van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina G H Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, and Hugo J W L Aerts. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research*, 77(21):e104 LP – e107, nov 2017.
10. H Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.
11. Miron B Kurba and Witold R Rudnicki. Feature Selection with the Boruta Package. *Journal Of Statistical Software*, 36(11):1–13, 2010.
12. Konstantinos Kamnitsas, Wenjia Bai, Enzo Ferrante, Steven G. McDonagh, Matthew Sinclair, Nick Pawlowski, Martin Rajchl, Matthew C. H. Lee, Bernhard Kainz, Daniel Rueckert, and Ben Glocker. Ensembles of multiple models and architectures for robust brain tumour segmentation. *CoRR*, abs/1711.01468, 2017.

Semi-automatic Brain Tumor Segmentation by Drawing Long Axes on Multi-plane Reformat

David Gering¹, Aaron Avery¹, Kay Sun¹, Lisa Kohli¹, Haley Knapp¹, Mark Gehring¹, Ajeet Vivekanandan¹, Jeff Hoffman¹, Brett Young-Moxon¹, Linda Peitzman¹, Roger Chylla¹, and Thomas Mackie¹

¹ HealthMyne, Madison WI 53717, USA
david.gering@healthmyne.com

Abstract. A semi-automatic image segmentation method based on workflow familiar to clinical radiologists is proposed. The user initializes 3D segmentation by drawing a long axis on a multi-plane reformat. As the user draws, a 2D segmentation updates in real-time for interactive feedback. When necessary, additional long axes, short axes, or other editing operations may be drawn on one or more MPR planes. The method learns probability distributions from the drawing to perform the MPR segmentation, and in turn, it learns from the MPR segmentation to perform the 3D segmentation. As a preliminary experiment, a batch simulation was performed where long and short axes were automatically drawn on each of the 285 multispectral MR brain scans of glioma patients in the BraTS 2018 online training data. Average Dice score for tumor core was 0.86218, and the Hausdorff 95% distance was 4.37 mm.

Keywords: Brain tumor, image segmentation, semi-automatic, machine learning

1 Introduction

Evidence from cancer researchers suggests that extraction of quantitative variables from medical images can contribute more information for decision support in management of cancer patients. Specifically, quantitative metrics can improve both 1) diagnostic and prognostic accuracy; as well as 2) longitudinal monitoring of patient response [1]. Criteria for monitoring radiographic brain tumor progression include the Macdonald criteria [2], Response Evaluation Criteria in Solid Tumors (RECIST) [3, 4], WHO criteria [5], and RANO criteria [6].

Currently, radiological studies are generally limited to detection and staging along with qualitative descriptions. Quantitative descriptors are not yet in the standard of care primarily due to a lack of infrastructure and tools to derive, test, and deploy these quantitative metrics at the point-of-care for all patients. Currently available tools to do this are limited to research or clinical trials, and have not been widely deployed as they lack the speed, precision and consistency required for wider clinical use [7]. The amount of time required to delineate lesion boundaries correctly could be intrusive to the radiologist's workflow. Delineation can be performed by manually drawing the

tumor boundary on each image slice, or by semi-automatically guiding an algorithm whose results require editing. Although manual delineation offers complete control to the user, humans exhibit great variability and the process is very time-consuming. Even if an automatic or semi-automatic method were to suffer a shortcoming in accuracy, as long as there is consistency in defining the boundary, then the volume change or change in a quantitative feature can be tracked more dependably.

For MR brain tumor segmentation, recent research with fully automated segmentation, especially based on deep neural networks, has been promising [8]. Our approach differs from a Computer Aided Detection (CAD) approach because we rely on a radiologist to make an indication. The motivation is adoption by clinical radiologists who desire full-control over the segmentation, real-time feedback, an algorithm that is ready to run immediately without the need to first be trained on a large database from their site, and an algorithm whose rationale behind decisions is explainable.

The vital part of any measurement tool is an interface that is both familiar and effortless. Drawing the longest axis across a lesion is a natural choice for initiating contours because radiologists are already accustomed to drawing the long axis. Oncologists participating in clinical trials follow published international criteria for objectively gauging the extent and progression of disease. The Macdonald, RECIST, and WHO criteria each incorporate long axis measurements. However, inherent challenges with axis-based criteria have been reported for aggressive brain tumors [9], thus motivating the discovery of volumetric-based criteria with similar familiarity as axis-based criteria.

Besides familiarity, there are several more goals of volumetric contouring. One goal is to achieve inter-observer consistency, while also catering to individual preferences for accuracy and style. Consistency results from initialization strategies that are reproducible, such as generating 3D volumetric contours from a straight stroke rather than free-form drawing. Tailoring to individual preferences is accomplished by editing tools prepared for whenever the initial contours may be unsatisfactory. Another goal is to provide a contingency plan in case the radiologist is both unsatisfied with the contours, and unwilling to invest the requisite time to edit them. Radiologists should be given the choice of confirming either the contours (thereby enabling volumetric measures), or just the long axis, which has already been drawn, and is held in reserve as an instant alternative. Yet another goal is automatic, large-scale, quantitative validation. Given hundreds of datasets that have been manually contoured, batch processing can be implemented by calculating the long axis from each expert's contours, and employing the long axis as the simulated user input. Yet another goal is to alleviate the need to select tools from a confusing suite of options. Ideally, there is exactly one tool in a reading room, generally applicable to all organs, yet simultaneously specialized with organ-specific features. The organ is automatically identified upon tool initialization.

Our software aims to satisfy all the aforementioned goals, namely familiarity, consistency, individualism, contingency, automatic validation, and general applicability yet specialization. While advancements in processing speed have propelled deep learning (DL) in various fields, medical image analysis is missing the mass quantities of new labelled data needed for training artificial intelligence networks [10]. The multimodal Brain Tumor Segmentation challenge (BraTS) represents a pioneering step in this direction [11, 12]. One of our goals is to develop software which can be

used to generate such contours on new scans, at the point of read, which in turn, can serve as the labeled image data for DL in subsequent clinical application.

2 Methods

The algorithm is implemented as a probabilistic framework with efficient user control.

2.1 Interactive MPR Segmentation

The user initializes the segmentation process by drawing a long axis on one plane of a Multi-Plane Reformat (MPR). As the user draws the long axis, a 2D segmentation updates in real-time for interactive feedback. The feedback has proven to be very helpful for the user to know precisely where to place the endpoint of the axis. Upon release of the mouse, 2D segmentation occurs immediately on the other MPR planes.

When the 2D contour is unsatisfactory, an optional short axis may be drawn perpendicular to the long axis. Other editing operations are available, such as a “ball tool” for drawing with a digital brush. A correct 2D segmentation is important since probability distributions are learned from the 2D segmentation to be employed in segmenting the other MPR planes.

When the contours on other MPR planes are unsatisfactory, then the user can draw there with the same editing tools, and the option for drawing a long axis and short axis. This is especially useful for lesions which are irregularly shaped or oriented obliquely.

2.2 3D Segmentation

Multivariate Bayesian classification [13] labels image voxels as belonging to one of two classes, Background or Foreground. Classification combines the likelihood of class membership based on voxel brightness, with the probability of membership prior to observing brightness. The likelihoods are conditional probability distributions that do not vary across the image, while the prior probabilities are spatially varying, and a function of distance from region boundaries.

The user directly drives the segmentation process by manipulating four types of regions, where some regions govern the likelihoods, while some regions govern the prior probabilities. Various regions are described in Table 1, and illustrated in Figure 1.

Table 1. Regions which drive probabilities.

Region	Color in Fig. 1	Description
Inclusion	Green	All voxels within belong to the Foreground class, and statistically sample it.
Containment	Yellow	All voxels outside belong to Background class.
Background	Blue & pink	Statistically typify Background class.
Avoidance	Not shown	Spatially prohibit Foreground without affecting statistics.

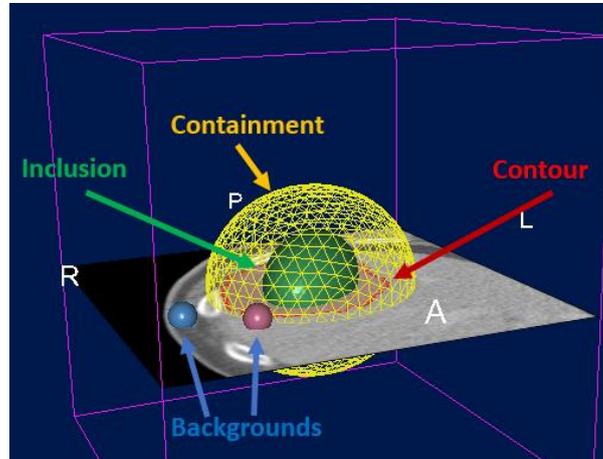


Fig. 1. The regions are initially configured as ellipsoids, and may become warped.

The sizes and poses of regions are automatically derived from the long axis. While the long axis describes lesion extent along one dimension, the initialization stage estimates lesion extent along other dimensions by analyzing orthogonal scout planes given statistical sampling along the long axis. Probability distributions are modeled parametrically as Gaussian Mixture Models (GMM) [14] while placing the Inclusion region, and as non-parametric distributions thereafter. Background regions are automatically placed by searching the vicinity outside the Containment region, and within the body outline, while maximizing the Mahalanobis distance [13] from the Inclusion region. Once Background and Inclusion regions are initialized, the voxels within are used to perform Parzen windowing [13] to estimate the likelihoods for Bayesian classification.

Noise and artifacts in CT vary by dose and choice of reconstruction kernel, and in MR by field strength and protocol parameters, so Bayesian classification is augmented with a Markov Random Field [15] with 3 iterations of mean-field approximation. The output is a 3D mesh fit to voxel classification by adapting vertices connected by virtual springs to their neighbors to provide a regularizing force that smooths the surface.

2.3 Preliminary Experiment

As a preliminary experiment, a batch simulation was performed where long and short axes were automatically drawn on each of the 285 multispectral MR brain scans of glioma patients in the BraTS 2018 online training data. To achieve this, the ground-truth was analyzed to find the largest slice in the central third, and an automatic process drew the long-axis on that slice. The center of that axis was used for the center of sagittal and coronal planes to comprise a 3-plane MPR. In order to simulate the type of long axis that a human user might draw, the axis position was favored to be more medial than the true longest axis. Therefore, on each plane, an ellipse was fit by Principle Component Analysis (PCA) [13] to the segmentation on that slice. The long axis with the same orientation as the major axis of the ellipse was found. The short axis was then found as the longest axis perpendicular to this, as shown in Figure 2.

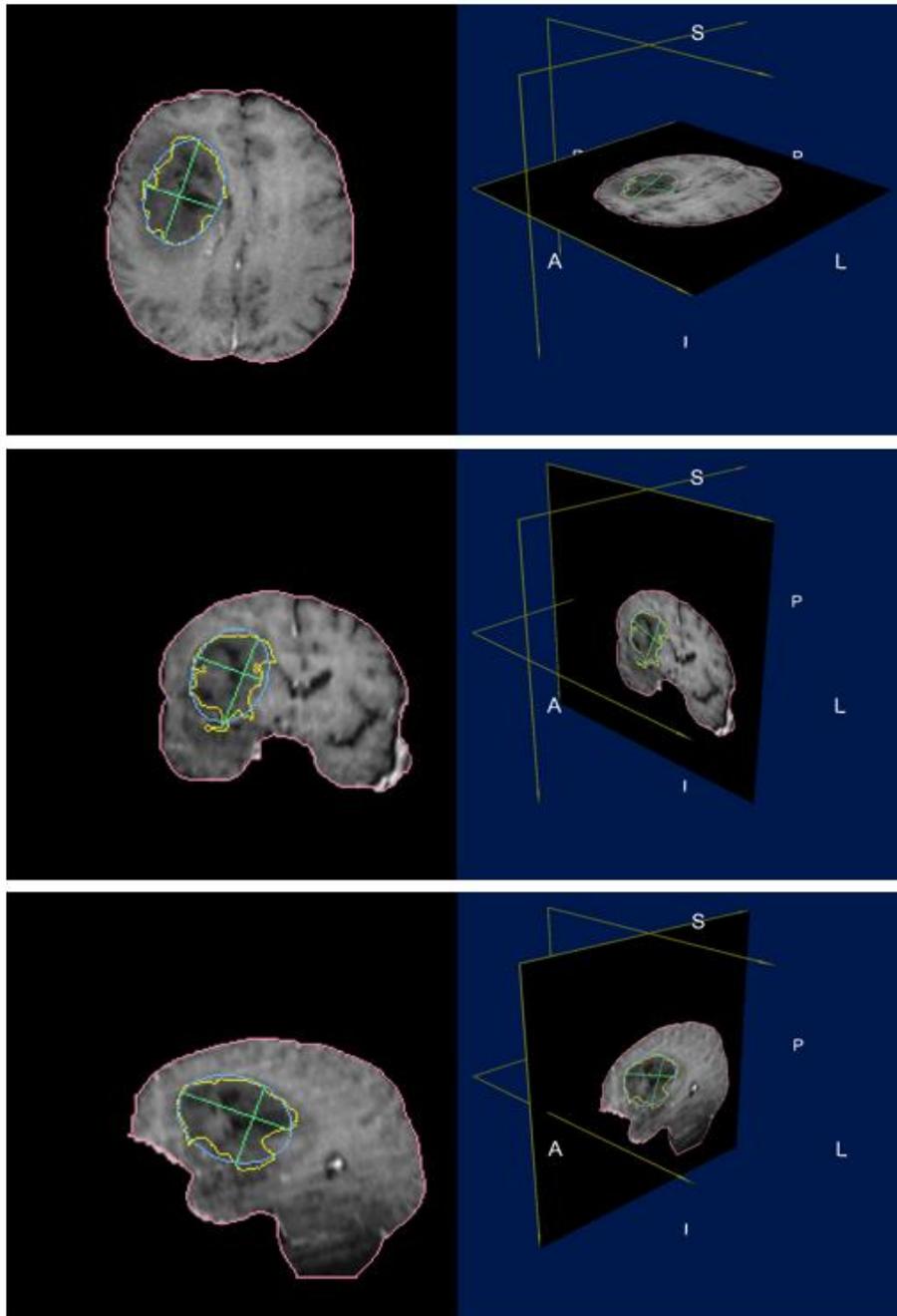


Fig. 3. Axial, coronal, and sagittal planes of MPR are shown from top to bottom. The blue ellipse was fit to the yellow contour of ground-truth in order to generate the green long and short axes. This process simulated a human user manually drawing on MPR.

The drawn axes precipitate MPR segmentation, and Figure 3 shows a few examples.

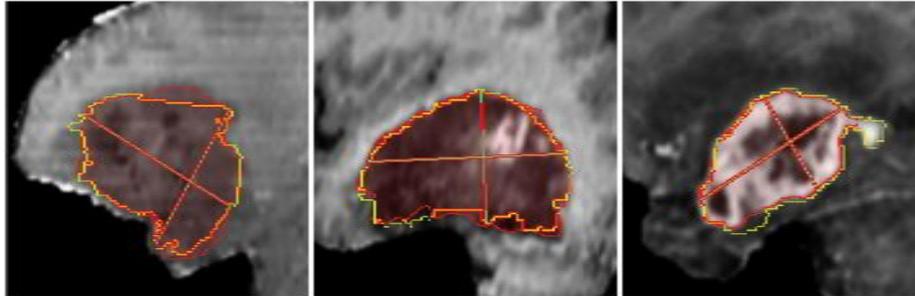


Fig. 3. MPR segmentation (red) depicted relative to ground-truth contours (yellow) and long/short axes (green) on a reformatted sagittal slice. MPR segmentations (not final 3D) were measured to have 0.901 average Dice, compared to ground-truth, for 855 planes of 285 cases.

3 Results

Multi-institutional, routine clinically-acquired pre-operative multispectral MR scans were provided by the 2018 BraTS challenge [16, 17]. The data had been preprocessed to be co-registered to the same anatomical template, interpolated to the same resolution (cubic mm), and skull-stripped.

The T1-weighted post-contrast scan was combined with the T2-weighted scan to create a dual-spectra image that was input to our algorithm. Since long axes were drawn on core tumor, this experiment segmented only that structure, and not the whole tumor and enhancing tumor structures which are also included in the 2018 BraTS challenge. We intend to add those structures by the testing phase of the challenge, as well as make use of the currently unused spectra, T1-weighted and FLAIR.

Segmentation accuracy was computed by the CBICA Image Processing Portal [18]. Average Dice score for the core tumor was 0.86218, with standard deviation 0.06009. Median score was 0.87252. Specificity was 0.89267, and sensitivity was 0.99832. The Hausdorff 95% distance was 4.37 mm.

4 Discussion

In comparison with other semi-automatic tools, products from Invivo [19] and Mirada [20] feature initialization by a single click, whereas the additional information contained in our long axis bolsters reliability relative to a click. Perhaps the most similar algorithm to ours is the GrowCut algorithm [21, 22] implemented in the 3D Slicer [23]. Both have general applicability, and a concept of Background and Foreground regions. However, GrowCut is not initiated as quickly as a drag across the long axis, and one study measured lung lesion contouring to require an average of 10 minutes [24], whereas our ambition is sub-minute. Perhaps the most similar initialization method is

[25] for the Random Walker algorithm [26], because a clicked point or stroke commences 2D segmentation from which Background and Foreground seeds are generated for 3D segmentation. However, our approach intentionally seeks statistical separation rather than a simple circumscribed shape for Background. GrowCut and the Random Walker both lack the two additional regions that ours adds, Containment and Avoidance, which make editing expeditious. Furthermore, ours differs by its Bayesian framework, which in conjunction with the added regions, make it possible to seamlessly incorporate organ-specific processing, and in the future, to employ DL-based CAD to derive additional spatially-varying prior probability maps.

Quantitative results were promising, while also leaving ample opportunity for near-term improvement. Since the algorithm is currently general-purpose, and well-suited for CT lung and liver lesions, we will add some MR-specific and brain-specific enhancements between the time of this writing and the test phase of the BraTS challenge. These improvements will include the partitioning of the tumor into its constituent parts: edema, necrosis, and actively enhancing regions.

During the test phase of the challenge, the long axis will be drawn manually by a human user, with the guidance of real-time MPR segmentation as constructive feedback. We anticipate that the advantage of feedback will produce better quantitative scores than the batch-generated long and short axes of this preliminary experiment.

5 References

1. Nordstrom, R. The Quantitative Imaging Network in Precision Medicine, *Tomography*. 2(4), 239–241 (2016).
2. Macdonald D.R., Cascino T.L., Schold S.C., Jr., Cairncross J.G. Response criteria for phase II studies of supratentorial malignant glioma. *J. Clin. Oncol.* 8, 1277–1280 (1990).
3. Therasse, P., Arbuck, S.G., Eisenhauer, E.A., Wanders, J., Kaplan, R.S., Rubinstein, L., Verweij, J., Van Glabbeke, M., van Oosterom, A.T., Christian, M.C. and Gwyther, S.G. New guidelines to evaluate the response to treatment in solid tumors. *Journal of the National Cancer Institute* 92(3), 205-216 (2000).
4. Sorensen A.G., Batchelor T.T., Wen P.Y., Zhang W.T., Jain R.K. Response criteria for glioma. *Nat. Clin. Pract. Oncol.* 5, 634–644 (2008).
5. Suzuki C., Jacobsson H., Hatschek T., Torkzad M.R., Boden K., Eriksson-Alm Y., Berg E., Fujii H., Kubo A., Blomqvist L. Radiologic measurements of tumor response to treatment: Practical approaches and limitations. *Radiographics* 28, 329–344 (2008).
6. Wen P.Y., Macdonald D.R., Reardon D.A., Cloughesy T.F., Sorensen A.G., Galanis E., Degroot J., Wick W., Gilbert M.R., Lassman A.B. Updated response assessment criteria for high-grade gliomas: Response assessment in neuro-oncology working group. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* 28, 1963–1972 (2010).
7. Yankeelov, E., Mankoff, D., Schwartz, L., Rubin, D. Quantitative Imaging in Cancer Clinical Trials. *Clinical Cancer Research*. 22, 284-290 (2016). 10.1158/1078-0432.CCR-14-3336
8. BraTS 2017 Proceedings Short papers, https://www.cbica.upenn.edu/sbia/Spyridon.Bakas/MICCAI_BraTS/MICCAI_BraTS_2017_proceedings_shortPapers.pdf
9. Mehta A.I., Kanaly C.W., Friedman A.H., Bigner D.D., Sampson J.H. Monitoring Radiographic Brain Tumor Progression. *Toxins* 3(3), 191-200 (2011). doi:10.3390/toxins3030191

10. Freedman, D.H. A reality check for IBM's AI ambitions. MIT Technology Review (2017). <https://www.technologyreview.com/s/607965/a-reality-check-for-ibms-ai-ambitions/>.
11. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Γ, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharruddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekeley G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
12. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
13. Duda, R.O., Hart, P.E. and Stork, D.G.: Pattern classification. 2nd edn. John Wiley & Sons, New York (2001).
14. Press, W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P.: Numerical recipes: The art of scientific computing. 3rd edn. Cambridge University Press, Cambridge (2007).
15. Li, S.Z.: Markov random field modeling in image analysis. Springer Science & Business Media (2009).
16. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
17. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
18. CBICA Image Processing Portal, <https://ipp.cbica.upenn.edu/>.
19. Invivo DynaCAD, <http://www.invivocorp.com/solutions/lung-cancer-screening/>.
20. Mirada XD3, http://www.mirada-medical.com/_public/documents/1425036206_mm3540-1-usaxd3forrecisthr.pdf.
21. Vezhnevets, V. and Konouchine, V. GrowCut: Interactive multi-label ND image segmentation by cellular automata. In proc. of Graphicon, vol. 1, pp. 150-156 (2005).
22. Zhu, L., Kolesov, I., Gao, Y., Kikinis, R. and Tannenbaum, A. An effective interactive medical image segmentation method using fast growcut. In: MICCAI workshop on interactive medical image computing, Boston (2014).
23. 3D Slicer. <http://www.slicer.org>.
24. Velazquez, E.R., Parmar, C., Jermoumi, M., Mak, R.H., Van Baardwijk, A., Fennessy, F.M., Lewis, J.H., De Ruyscher, D., Kikinis, R., Lambin, P. and Aerts, H.J. Volumetric CT-based segmentation of NSCLC using 3D-Slicer. Scientific reports, 3, p. 3529 (2013).
25. Jolly, M.P. and Grady, L. 3D general lesion segmentation in CT. In: Biomedical Imaging: From Nano to Macro. 5th IEEE Intern. Symposium on ISBI, pp. 796-799. IEEE, (2008).
26. Grady, L. Random walks for image segmentation. IEEE transactions on pattern analysis and machine intelligence 28(11), 1768-1783 (2006).

Sirona: A Bio-Physics Based Deep Neural Network for Medical Image Segmentation

Amir Gholami¹, Shashank Subramanian², Peter Jin¹, Naveen Himthani²
Varun Shenoy¹, Xiangyu Yue¹, Sicheng Zhao¹, Pravin Ravishanker¹
Kurt Keutzer¹, and George Biros¹

¹ University of California, Berkeley

² University of Texas at Austin

Abstract. We present Sirona, a new framework for fully automatic segmentation of brain MR images. Sirona combines the following novel approaches: (i) fine grained segmentation of tumor and healthy tissues including Gray Matter, White Matter, Cerebra Spinal Fluid, and Glial Matter through a new image registration method; (ii) a novel domain adaptation method to automatically enrich the training dataset using a new multi-species tumor growth model along with a generative adversarial model to transfer the simulation results to match brats intensity histogram; These approaches are not particular to a particular model and could be used in conjunction to existing solutions to increase segmentation performance. Here we briefly discuss the formulation and provide initial results on BRATS'18 training and validation datasets.

Keywords: Segmentation, Neural Network, Machine Learning, Glioma, Tumor, Sirona

1 Introduction

Gliomas are a type of brain tumor arising from glial cells of the brain. They account for about 80% of all malignant brain tumors diagnosed in the United States [4]. Gliomas can be graded according to their aggressiveness. While low-grade gliomas can be benign and carry a reasonable prognosis, high-grade gliomas are malignant and have a tendency to grow quickly and invade other parts of the brain [16]. A combination of surgery and chemotherapy is the usual course for treatment, but the prognosis is generally poor. The median survival rate for patients with high-grade gliomas is less than a year. Hence, the early diagnosis of the tumor is crucial for improving treatment options for the patient.

Medical imaging is an efficient way of obtaining information about the location and structure of the brain tumor. Magnetic resonance imaging (MRI) is a common imaging technique used to visualize the extent of tumor regions. The standard clinical routine involves accurately delineating the different tumor structures in the brain. However, the manual segmentation of 3D MRI scans is

time-consuming and prone to inaccuracies and variability due to the highly heterogeneous nature of tumor appearance. Automatic segmentation methods have the potential to provide an alternative to obtaining accurate segmentation leading to improved tumor prognosis and treatment planning, especially for cases where access to expert radiologist is limited. Moreover, the widespread usage of such a system would provide standardization to a field with high discrepancies among different radiologists regarding brain segmentation.

A brain can be visualized using different MRI modalities. Different imaging modalities provide different information about the scanned region. For example, a particular feature visible in one modality may be hidden in another. Common brain MRI modalities include post-Gadolinium T1 (used to enhance contrast and visualization in the blood-brain barrier), T2 and FLAIR (to highlight different tissue fluid intensities), and T1. Together, the complementary information from these modalities enable a more robust segmentation of a tumor-afflicted brain. Recently new imaging modalities such as Diffusion Tensor Imaging (DTI) and Positron Emission Tomography (PET) have been used in cases with gliomas.

Deep learning approaches using convolutional neural networks (CNNs) have recently demonstrated excellent performance in semantic segmentation tasks in medical imaging. A seminal work for segmentation stem from fully-convolutional network (FCNs) [11]. U-nets [15] are another popular architecture in medical segmentation tasks, which merge feature maps from the contracting path of an FCN to its expanding path to preserve local contextual information. Multi-scale information is often incorporated by using parallel convolutional pathways of various resolutions [10] or by using dilated convolutions and cascading network architectures [17]. Post-processing using Conditional Random Fields (CRF) [10] has proven to be useful in fine-tuning segmentation and removing noisy patches as well. The winning algorithm of the Multimodal Brain Tumor Image Segmentation Benchmark (BraTS) challenge in 2017 was based on Ensembles of Multiple Models and Architectures (EMMA) [9], which bagged a heterogeneous collection of networks (including DeepMedic (winner of ISLES 2015 [12]), U-nets and FCNs) to build a robust and generic segmentation model. Due to the limited availability of data, standard data augmentation methods involving affine transformations, reflections and elastic deformations are frequently employed [15] to regularize the model.

In this paper, we present Sirona which could be used in conjunction to the above solutions. Our new contributions include fine grained segmentation of the brain, including healthy tissues. Creating a dataset for healthy brain segmentation is a very tedious task, but we propose an automated approach based on image registration to delineate healthy tissue in tumor-bearing medical images. Moreover, we propose a domain adaptation strategy to enrich the training dataset using Generative Adversarial Networks (GANs) in conjunction with a novel multi-species tumor model with mass effect.

The outline of the paper is as follows. In §2, we discuss the methodology for whole brain segmentation (§2.1), and domain adaptation (§2.2). Then in §3 we present preliminary results on the BraTS'18 challenge [1-3, 12].

2 Methods

2.1 Whole Brain Segmentation With Healthy Tissues

The healthy structure of the brain changes in the presence of abnormalities such as tumors. For instance, the delineation of the tissues could be compressed due to tumor growth in the confined space of the brain. Therefore, providing this information to the classifier should increase the accuracy of the segmentation. However, such data is not available in the BRATS training data, mainly due to the fact that labelling a full 3D brain is an extremely time-consuming task. However, here we propose an automated approach to compute this information. Our method for segmenting healthy brain tissue uses multiple healthy brains as atlas that have been segmented a priori. The entire process of segmentation is divided into three main steps, which are then explained briefly in the following sections,

1. Preprocessing: affine registration of each atlas image to the brats image,
2. Diffeomorphic registration of each atlas image to the brats image, and using the resulting deformation to transport the atlas segmentation, and
3. Performing majority voting to fuse labels of all deformed atlases to get the final healthy tissue segmentation.

2.1.1 Preprocessing: Rigid+Affine Registration We follow the strategy presented in [5]. First, all the healthy brain atlases are re-sampled to match the brats image resolution ($240 \times 240 \times 155$) using MATLAB. This is followed by a rigid and affine registration (using NMI similarity measure; 12 degrees of freedom; 3 multi-resolution levels; and a 3D linear interpolation model) of the atlas T_1 image to the brats T_1 image. The resulting transformation is applied to the atlas segmentation (using nearest neighbour interpolation). We use MIRTk (<https://mirtk.github.io>) to do the affine registration step. This is then repeated for all the remaining atlases.

2.1.2 Diffeomorphic Registration We use large-scale diffeomorphic registration to deform the atlas so that it would be deformed to match the brats image structure (for details please see [13, 14]). The forward registration problem is defined as follows: given a velocity field $\mathbf{v}(\mathbf{x})$ and image vector ρ_T , the deformed image is computed by a solving the advection equation in $t \in (0, 1]$

$$\partial_t \rho(x, t) + \mathbf{v} \cdot \nabla \rho(x, t) = 0, \quad (1)$$

with $\rho(x, 0) = \rho_T$ in Ω . We denote this operation by the forward operator \mathcal{F} :

$$\rho_T(x, 1) = \mathcal{F}(\mathbf{v}, \rho_T) \quad (2)$$

On the other hand, the inverse problem is formulated as follows: given two vector fields ρ_T and ρ_R , we seek to find a velocity $\mathbf{v}(\mathbf{x})$ such that the L^2 distance

between $\mathcal{F}(\mathbf{v}, \boldsymbol{\rho}_T)$ and $\boldsymbol{\rho}_R$ is minimal. This is an ill-posed problem and needs regularization to place smoothness constraints on the velocity field. We again refer to [13, 14] for details on how we solve this inverse problem to find this velocity.

The affinely registered atlases T_1 images are then diffeomorphically registered to the brats T_1 image. The image intensities are rescaled to $[0,1]$ before performing the registration. The resulting velocity output is used to transport the atlas segmentation. The resulting segmentation is a crude estimation of the healthy tissue in the brats image.

2.1.3 Majority Voting and Label Fusion We use a simple majority voting technique based on the quality of diffeomorphic registration. We compute the L_2 norm of the residual between each deformed atlas and brats image,

$$r = \|(\boldsymbol{\rho}_T(x, 1) - \boldsymbol{\rho}_R)\|_2$$

before (r_0) and after (r_1) the registration. Based on the residual ratio $\frac{r_1}{r_0}$, weights are assigned to each atlas. We set the weights equal to the inverse of the residual ratio which implies that a deformed atlas which is closest to the BraTS image will get the highest weight in the fusion process. The weights are then normalized and the label probabilities are computed using:

$$p_j(x) = \sum_{i=1}^{N_a} d_{ij} * \tilde{w}_i$$

$$\tilde{p}_j(x) = \frac{1}{\sum_{j=1}^{N_l} p_j(x)}$$

where, x is target voxel to be labelled, N_a is the number of atlases, N_l is the number of labels, \tilde{w}_i are the normalized weights, j is the label id, $\tilde{p}_j(x)$ is the probability for voxel x being label j and

$$d_{ij} = \begin{cases} 1, & \text{if } atlas_i(x) = j \\ 0, & \text{otherwise} \end{cases}$$

The final segmentation is computed using the max probability for a given voxel,

$$label(x) = \arg \max_j \tilde{p}_j(x)$$

. We finally overlay, the BraTS segmentation for the tumorous regions.

2.1.4 Limitations In general, the quality of BraTS T1 images is very poor, they are flat and lose contrast near the brain boundaries. This makes it difficult to create good quality healthy segmentation using the atlas based method. This

is mainly due to the fact that the probabilistic atlases are only available in T1 modality. However, this limitation could be addressed by creating one or more segmentation of T1ce or Flair images.

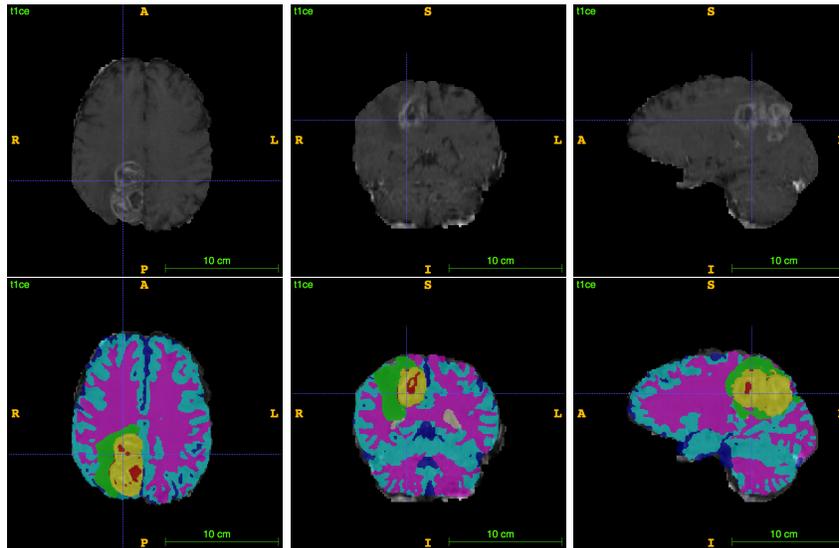


Fig. 1: (top row): The original T1 image for Brats18_TCIA02_135_1 training data is shown for different views (Axial, Coronal, and Sagittal), (bottom row): the corresponding fine grained segmentation for healthy cells computed by solving a 3D registration problem with a segmented atlas. We overlay BRATS tumor segmentation with the registered segmentation to get the final results.

2.2 Domain Adaptation

One of the main limitations in medical imaging, and in particular brain tumors, is the scarcity of training data. To address this issue, we use a novel domain adaptation strategy and generated synthetic BRATS data to enrich the training dataset. This is performed by first solving a PDE based multi-species tumor model using an atlas brain. However, this data could not be used directly due to the domain shift between actual BRATS images and simulated results. Directly using this synthetic data during the training process may guide the neural network to learn some features that only exist in synthetic data, resulting in poor performance on the real dataset. To address this issue, we use CycleGAN [18] to do the domain adaptation from the generated synthetic data to the real BRATS data. The goal of CycleGAN [18] is to learn a mapping $G : X \rightarrow Y$ such that the distribution of images from $G(X)$ is indistinguishable from the distribution Y using an adversarial loss. Because this mapping is highly under-constrained,

it is coupled with an inverse mapping $F : Y \rightarrow X$ and a cycle consistency loss is introduced to enforce $F(G(X)) \approx X$ (and vice versa). Samples of our adaptation results are shown Fig. 2. In order to visualize each brain data sample, we normalize the values to $[0, 1]$ and treat the channels as CMYK channels of an image.

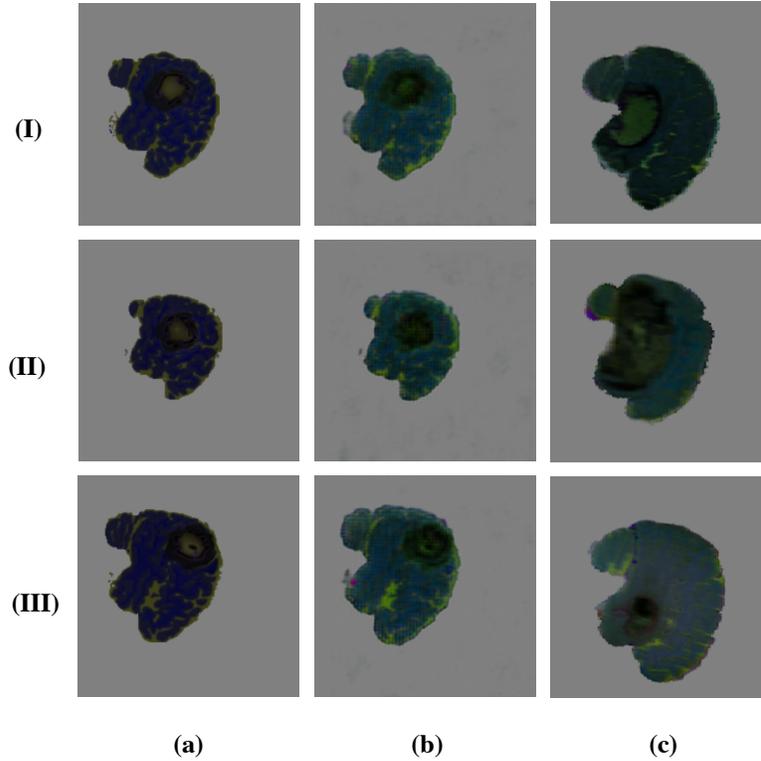


Fig. 2: Domain Adaptation results: I, II, III represents 3 different samples. (a) represents synthetic brain images, (b) represents adapted synthetic images, (c) represents real brain images. Images are shown in CMYK mode, with each channel mapped from T1, T1CE, T2, and FLAIR channel. As we can see from the colors, the values in the adapted images are closer to the real images.

We generate synthetic images by simulating tumor growth in healthy brain atlases. We use a multi-species tumor growth model based on the works of [6, 16], which provides us with the time evolution of enhancing and necrotic tumor concentrations along with tumor-induced brain edema. The governing equations for the model are reaction-diffusion-advection equations for the tumor species along with a diffusion equation for oxygen and other nutrients. We couple this model with linear elasticity equations with variable elasticity material properties

to simulate the deformation of surrounding brain tissue due to tumor growth, also known as “mass effect”.

Tumor and healthy cell concentration maps are converted into synthetic MRI images by sampling from the intensity distribution of real MRI scans. We use the ground truth tumor segmentation of the BraTS’18 dataset in conjunction with our healthy tissue segmentation to transfer intensities.

2.3 Parallel Framework

We utilize a prototype framework for parallelized automatic differentiation called **anode**. The **anode** framework is designed to accelerated differentiable dataflow models, such as deep convolutional networks, via generalized data parallelism by partitioning tensors across batch elements and across the spatial dimensions [7]. Generalized data parallelism across batch and spatial axes enables the training of large models which may not easily fit on modern GPUs [8]. For example, a UNet architecture [15] in single floating point precision applied to input tensors with dimensions $128 \times 128 \times 128 \times 4$ uses over 4.4 GB of memory for just the forward computation of a single input.

Preliminary microbenchmarks of **anode** on UNets suggest nearly linear strong scaling to at least 32 Tesla P100 GPUs.

3 Preliminary Results

3.1 Fine Grained Healthy and Tumor Segmentation

Here we report preliminary results on BRATS’18 validation dataset, with an exemplary segmentation shown in Figure 4. We used a 10 layer UNet with 3D convolution as the baseline network and trained for 500 epochs using an 20% validation split of the training data with ADAM optimizer. We used TensorFlow/Keras framework for these preliminary tests. For the preprocessing step, we first solve a 3D image registration problem to find the fine grained healthy segmentation of the brain given a set of segmented healthy atlases. Then using this information we crop the validation data to crop background. Then we downsample the cropped image to 128^3 resolution. We achieve a dice score of (87.15,66.56,77.18) for (WT,TC,EN), respectively.

4 Conclusion

We presented preliminary results for Sirona, a new scalable framework for biophysics-based medical image segmentation, and in particular gliomas. Our contributions include an automatic healthy segmentation of the BRATS dataset, and a novel Generative Adversarial Network to enrich the training dataset. We demonstrated that our approach yields promising results, and showed preliminary results on BRATS’18 validation dataset.

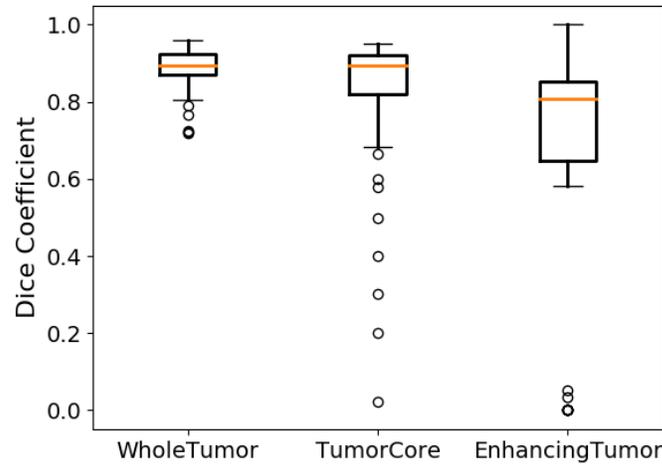


Fig. 3: Boxplot for the baseline model's validation score on the BraTS'18 training data is shown. This model achieves a dice score of (87.15,66.56,77.18) percent for (WT,TC,EN), respectively.

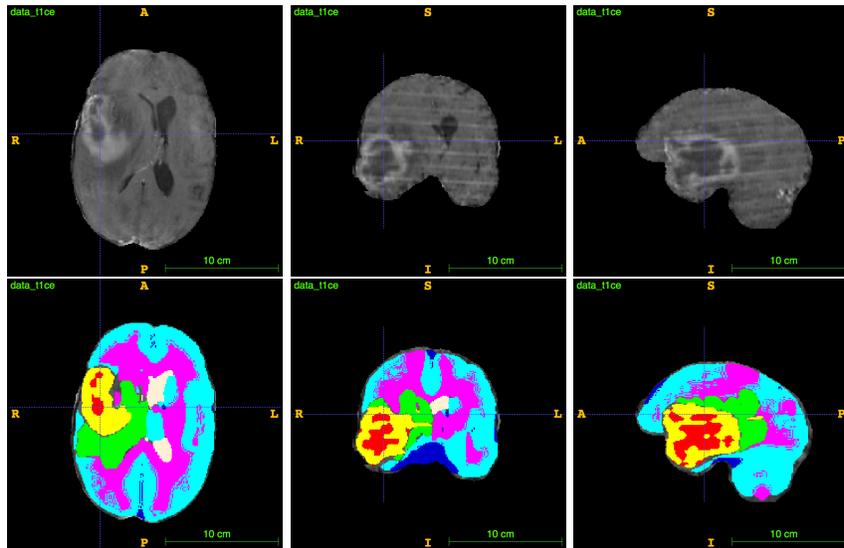


Fig. 4: (top row): The original T1ce image for Brats18-TCIA02.230.1 validation data is shown for different views (Axial, Coronal, and Sagittal), (bottom row): the corresponding segmentation result for healthy cells computed by the neural network visualized using ITK-Snap. We achieve a dice score of (91.855,94.48,94.026) for (EN,WT,TC), respectively.

References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Nature Scientific Data* (2017), in press
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels for the pre-operative scans of the TCGA-GBM collection (2017), <http://doi.org/10.7937/k9/tcia.2017.klxwj1q>
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels for the pre-operative scans of the TCGA-LGG collection (2017), <http://doi.org/10.7937/k9/tcia.2017.gjq7r0ef>
4. Gholami, A., Mang, A., Biros, G.: An inverse problem formulation for parameter estimation of a reaction-diffusion model of low grade gliomas. *Journal of Mathematical Biology* 72(1), 409–433 (2016)
5. Gooya, A., Pohl, K.M., Bilello, M., Cirillo, L., Biros, G., Melhem, E.R., Davatzikos, C.: Glistr: Glioma image segmentation and registration. *IEEE Transactions on Medical Imaging* 31(10), 19411954 (2012)
6. Hawkins-Daarud, A., Rockne, R.C., Anderson, A.R.A., Swanson, K.R.: Modeling tumor-associated edema in gliomas during anti-angiogenic therapy and its impact on imageable tumor. *Front Oncol* 3(66) (2013)
7. Jin, P., Ginsburg, B., Keutzer, K.: Spatially parallel convolutions (2018)
8. Jin, P., Ginsburg, B., Keutzer, K.: Spatially parallel convolutions. *ICLR 2018 Workshop* (2018), <https://openreview.net/forum?id=S1Yt0d1vG>
9. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S.G., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M.C.H., Kainz, B., Rueckert, D., Glocker, B.: Ensembles of multiple models and architectures for robust brain tumour segmentation. *CoRR abs/1711.01468* (2017), <http://arxiv.org/abs/1711.01468>
10. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical Image Analysis* 36, 61 – 78 (2017), <http://www.sciencedirect.com/science/article/pii/S1361841516301839>
11. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. *CoRR abs/1411.4038* (2014), <http://arxiv.org/abs/1411.4038>
12. Maier, O., Menze, B.H., von der Gabelentz, J., Hni, L., Heinrich, M.P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., Christiaens, D., Dutil, F., Egger, K., Feng, C., Glocker, B., Gtz, M., Haeck, T., Halme, H.L., Havaei, M., Iftekharuddin, K.M., Jodoin, P.M., Kamnitsas, K., Kellner, E., Korvenoja, A., Larochelle, H., Ledig, C., Lee, J.H., Maes, F., Mahmood, Q., Maier-Hein, K.H., McKinley, R., Muschelli, J., Pal, C., Pei, L., Rangarajan, J.R., Reza, S.M.S., Robben, D., Rueckert, D., Salli, E., Suetens, P., Wang, C.W., Wilms, M., Kirschke, J.S., Krmer, U.M., Mnte, T.F., Schramm, P., Wiest, R., Handels, H., Reyes, M.: Isles 2015 - a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral mri. *Medical Image Analysis* 35 (2017)
13. Mang, A., Biros, G.: A semi-lagrangian two-level preconditioned newton–krylov solver for constrained diffeomorphic image registration. *SIAM Journal on Scientific Computing* 39(6) (2017)

14. Mang, A., Gholami, A., Biros, G.: Distributed-memory large deformation diffeomorphic 3d image registration. SC16: International Conference for High Performance Computing, Networking, Storage and Analysis (2016)
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. CoRR abs/1505.04597 (2015), <http://arxiv.org/abs/1505.04597>
16. Saut, O., Lagaert, J.B., Colin, T., Fathallah-Shaykh, H.M.: A multilayer grow-or-go model for gbm: effects of invasive cells and anti-angiogenesis on growth. *Bulletin of mathematical biology* 76(9), 2306–2333 (2014)
17. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. CoRR abs/1709.00382 (2017), <http://arxiv.org/abs/1709.00382>
18. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Computer Vision (ICCV), 2017 IEEE International Conference on* (2017)

Neuromorphic Neural Network for Multimodal Brain Tumor Segmentation and Survival analysis

Woo-Sup Han and Il Song Han

Odiga, London, UK
ishan@odiga.co.uk

Abstract.

Image analysis of brain tumors is one of key elements for clinical decision, while manual segmentation is time consuming and subjective. In this paper, we examine the neuromorphic convolutional neural network on this task of multimodal images, using a down-up sizing network structure. We use a controlled rectifier neuron function incorporated in neuromorphic neural network, which we think is important for a successful segmentation of noisy data. Experiment results show the effectiveness and feasibility of our proposed methods, from segmentation to overall survival analysis.

Keywords: Convolutional Neural Network, Neuromorphic Processing, Brain Tumor, Image Segmentation, Survival Analysis

1 Introduction

The assessment of brain tumors delivers valuable information and becomes one of key elements of clinical diagnosis. Therefore, the automatic brain image segmentation emerges as a critical technology, as there are advantages of faster, more objective and potentially desirable accuracy. Due to the irregular nature of tumors as well as noisy 3D MRI images, the development of practical solution is still challenging throughout the BRATS Challenge [1-4].

Overall survival(OC) analysis of patients has been also the subject of interests, which is evaluated from the baseline to the time of last available follow-up. A time threshold of 18 months was defined to differentiate the patients into 2 groups, those with short- or long-term survival [5]. In this paper, the segmentation algorithm is proposed and applied to evaluate OC based on BRAT 2018 high grade glioblastomas (HGG) data set and the survival data.

Since the convolutional networks like U-net have been widely in use [5-6], CNN and its derivatives have attracted more attentions on segmentation tasks. Recently, the neuromorphic neural network introduced its feasibility of segmentation of 3D brain CT images, by the successful 3D dental tooth segmentation including roots in the gum [7].

In this paper, we will give an experimental study of the effectiveness of neuromorphic neural network on multimodal brain tumor segmentation. This paper is intended for Tumor Segmentation Challenge 2018[1-4]. Since the available dataset is low, we

utilized the convolutional filters useful for prior segmentation analysis of medical images and other noisy ones. Experiment results show the prior techniques of bio-inspired convolutional filters and controlled linear rectifier neurons can boost the performance of the segmentation tasks.

2 Methods

The neuromorphic convolutional neural network in Fig. 1 is inspired by a neuromorphic neuron of simple cell, with the various orientation selective features. The system has the process of orientation feature extraction using neuromorphic processing mimicking the simple cell of visual cortex, based on the convolution with the filter banks. The introduction of down-up resizing in Fig. 1 enables the abstract feature extraction, with the robust saliency map generation combined with the controlled rectifier neuron.

Our network is based on the integer computation, while the convolutional filters are 13X13 of unsigned 8 bit integers. The fully connected neural networks in Fig. 1 is the feedforward network of 2x20x20x5x1, with two input variables of aggregated pixel numbers of each segmented images. The output represents the OS prediction of the short survival period ('0': less than 18 months) or the long survival period ('1': more than 18 months).

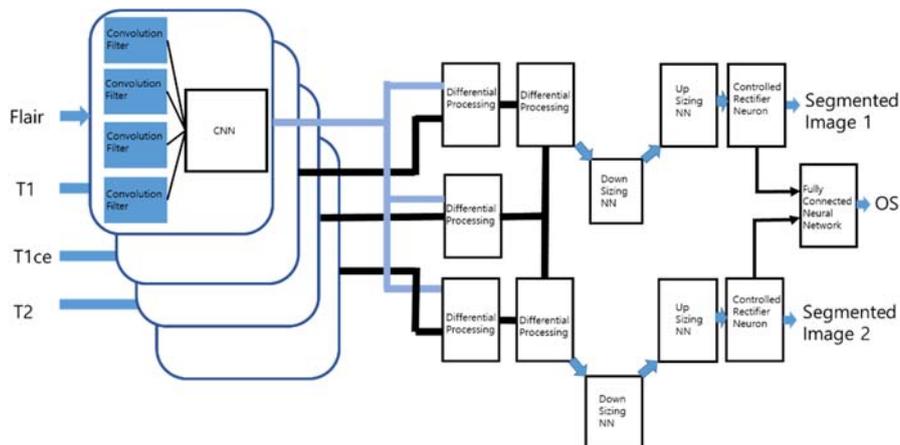


Fig. 1. Network architecture. Our architecture is inspired by the earlier neuromorphic convolutional recurrent neural networks of 3D dental segmentation, based on X-ray CBCT [7].

3 Results

The preliminary results are illustrated in Fig. 2-4. The neuromorphic orientation enhanced features are observed at the outputs of 1st stage processing, which can reduce the illumination change of individual image. The abstraction features enable the possible effectiveness in pattern recognition or clustering, which become advantageous for the limited size of training data,

asedoo

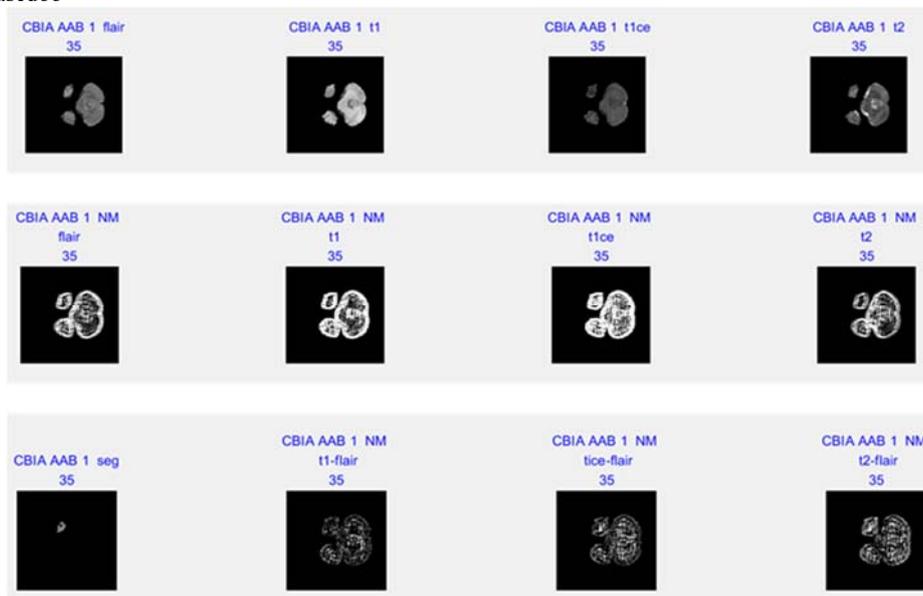


Fig. 2. Outcomes of proposed neuromorphic neural network using BRATS 2018 training data set, (HGG, CBIA_AAB_1 A, layer no= 35). Top: multimodal brain MRI images, middle: intermediate outputs of 1stage CNN, bottom: segmentation results (left end) and intermediate results.

The segmentation procedure utilizes the averaging and threshold process during down-up sizing neural network operation. The bottom left object illustrated the segmented neuron 1 in Fig. 2, which is some way close to the enhancing tumor in the provided ground truth. The similar functions were observed in dental tooth segmentation [7], which suggests the automatic segmentation capacity of Neuromorphic convolution filters and sow-up sizing neural networks. The controlled linear rectifier neuron is employed to improve the OS prediction, with both the pegged threshold value and the fixed threshold value for Fig. 3 and Fig. 4.

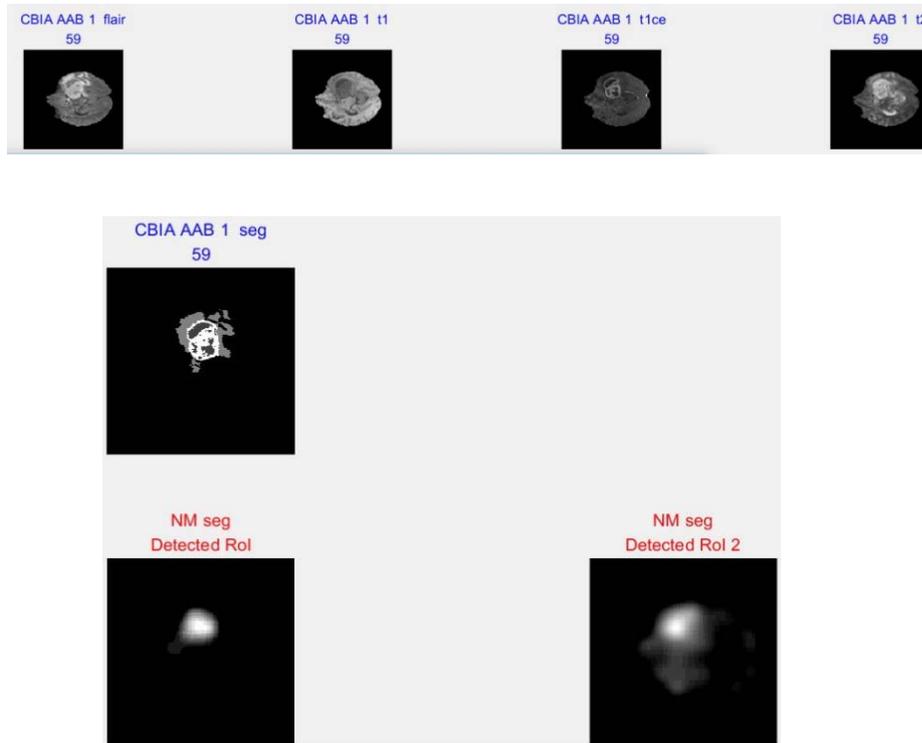


Fig. 3. Outcomes of proposed neuromorphic neural network using BRATS 2018 training data set, (HGG, CBIA_AAB_1 A, layer no= 59). Top: Multimodal brain MRI images, middle: segmentation results of BRATS dataset, bottom: two parallel outcomes of image segmentation intermediate results.

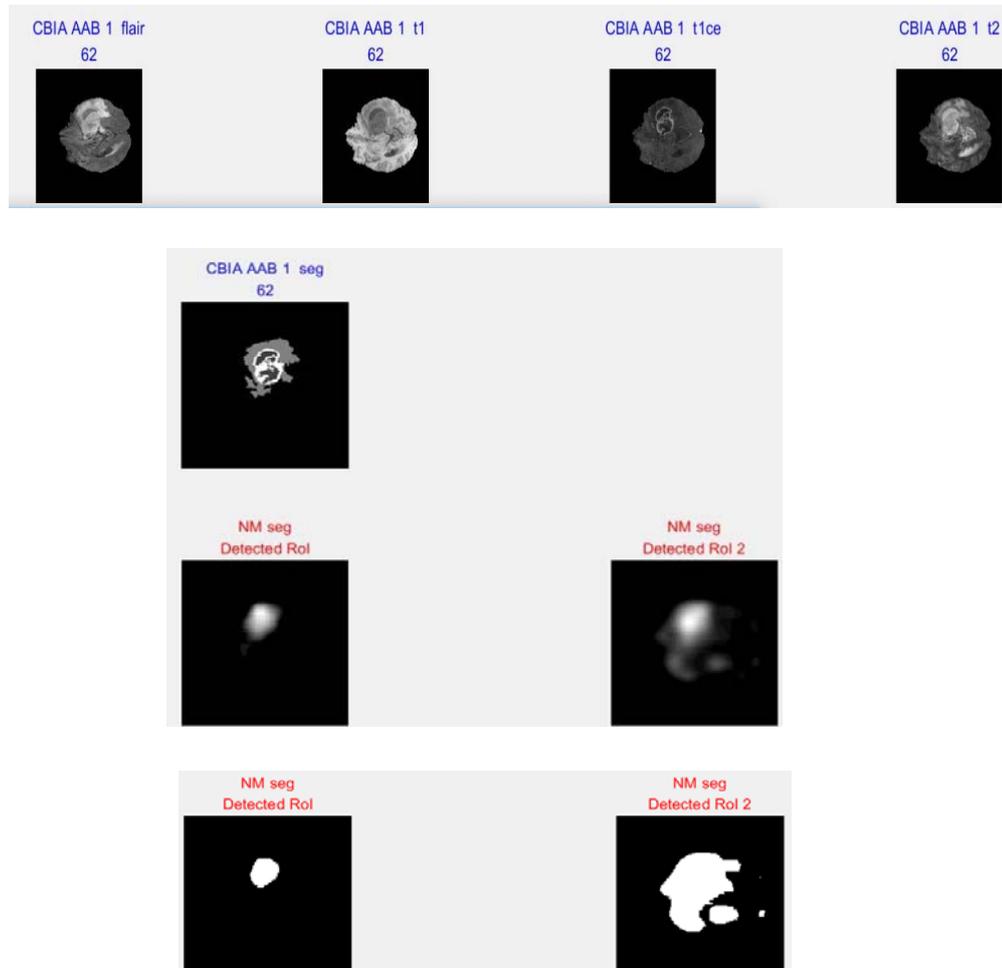


Fig. 4. Outcomes of proposed neuromorphic neural network using BRATS 2018 training data set, (HGG, CBIA_AAB_1 A, layer no= 62). Top: Multimodal brain MRI images, upper middle: segmentation results of BRATS dataset, lower middle: two parallel outcomes of image segmentation intermediate results, bottom: segmented images converted by controlled rectifier neurons...

The fully connected network is trained by the limited number of 161 data sets, and the result of Table 1 is summarized. Since there is a substantial difference among human experts of tumor segmentation, it would be challenging to implement the system with the definite result better human experts. The estimated performance is around 71% of accuracy.

Table 1. OS evaluation of BRATS 2018 HGG dataset by neuromorphic neural networks

Accuracy and sensitivity	Cases
Correct recognition	71% (115 cases among 161 cases)
Positive failure (mistaken as the long OS period : more than 18months)	11%
Negative failure (fail to predict the short OS period: less than 18 months)	18%

The contents of Table 2 illustrate the OS prediction of validation data released for BRATS 2018, without the ground truth of tumor segmentation. Images of Fig. 5 represent the result using validation dataset, without the ground truth of segmentation.

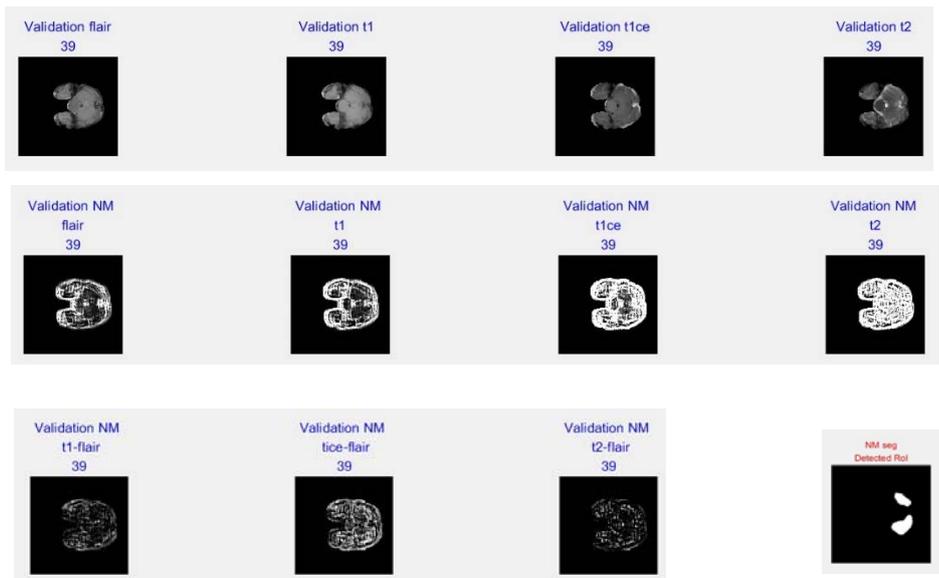


Fig. 5. Outcomes of proposed neuromorphic neural network using BRATS 2018 validation data set, . Top: Multimodal brain MRI images. middle: intermediate outputs of 1stage CNN, bottom: intermediate results and segmentation result (right end: image 1 output of Fig. 1)

4 Discussion

Our proposed algorithm has the feature of mimicking human visual recognition process as an artificial way, where the neuromorphic convolutional neural network unlikely incurs the heavy computing resources for learning. The decent size of neural network is more favorable to the fast operation or real-time operation for enhanced medical imaging instrumentations.

In this paper we presented the network architecture using the subset configuration of existing neuromorphic convolutional neural networks, due to the limited time and data set scale. We could expect to find the better one with more consistent training data sets, with further optimized feature processing neural networks.

References

1. Menze B., Jakab A., Bauer S., Kalpathy-Cramer J., Farahani K., Kirby J., Burren Y., Porz N., Slotboom J., Wiest R., Lanczi L., Gerstner E., Weber M., Arbel T., Avants B., Ayache N., Buendia P., Collins D., Cordier N., Corso J., Criminisi A., Das T., Delingette H., Demiralp C., Durst C., Dojat M., Doyle S., Festa J., Forbes F., Geremia E., Glocker B., Golland P., Guo X., Hamamci A., Iftekharuddin K., Jena R., John N., Konukoglu E., Lashkari D., Mariz J, Meier R., Pereira S., Precup D., Price S., Raviv R., Reza S., Ryan M., Sarikaya D., Schwartz L., Shin H., Shotton J., Silva C., Sousa N., Subbanna N., Szekely G., Taylor T., Thomas O., Tustison N., Unal G., Vasseuer F., Wintermark M., Ye D., Zhao L., Zhao B., Zikic D., Frastawa M., Reyes M., Leemput K.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans, Medical Imaging* 34(10), 1993-2024 (2015)
2. Bakas S., Akbari H., Sotiras A., Bilello M., Rozycki M., Kirby J., Freymann J., Farahani K., Davatzikos C.: Data Descriptor: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Nature Scientific Data*, 117 (2017)
3. Bakas S., Akbari H., Sotiras A., Bilello M., Rozycki M., Kirby J., Freymann J., Farahani K., Davatzikos C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection, The Cancer Imaging Archive, DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q (2017)
4. Bakas S., Akbari H., Sotiras A., Bilello M., Rozycki M., Kirby J., Freymann J., Farahani K., Davatzikos C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection, The Cancer Imaging Archive, DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF (2017)
5. Zacharaki E. Morita N. Bhatt P. O'Rourke D, Melhem E. Davatzikos C.: Survival Analysis of Patients with High-Grade Gliomas Based on Data Mining of Image Variables. *AJNR Am J Neuroradiol* 33(6), 1065-1071 (2012)
6. Cao S. et al: 3D U-Net for Multimodal Brain Tumor Segmentation. In 2017 International MICCAI BraTS Challenge Proceedings, pp. 30-33. Quebec City (2017).
7. Isensee F. et al: Brain Tumor Segmentation and Radiomics Survival Prediction ; Contribution to the BRATS 2017 Challenge. Pp 10-107. Quebec City (2017).
8. Han W., Han I.: Object Segmentation for Vehicle Video and Dental CBCT by Neuromorphic Convolutional Recurrent Neural Network, *Intelligent Systems and Application*, (ISBN 978-3-319-69266-1, 264-284. Springer (2018)

Table 2. OS(Overall Survival) evaluation of BRATS 2018 validation dataset by neuromorphic

	OS		OS
Brats18_MDA_1012_1	0	Brats18_TCIA13_646_1	0
Brats18_MDA_1015_1	0	Brats18_TCIA13_636_1	0
Brats18_MDA_1081_1	0	Brats18_TCIA09_248_1	1
Brats18_MDA_907_1	1	Brats18_TCIA02_230_1	1
Brats18_MDA_922_1	1	Brats18_TCIA02_400_1	1
Brats18_CBICA_BHN_1	0	Brats18_TCIA03_216_1	1
Brats18_CBICA_BLK_1	0	Brats18_TCIA03_288_1	1
Brats18_CBICA_AAM_1	1	Brats18_TCIA03_604_1	1
Brats18_CBICA_ABT_1	1	Brats18_TCIA03_313_1	1
Brats18_CBICA_ALA_1	0	Brats18_TCIA04_212_1	0
Brats18_CBICA_ALT_1	1	Brats18_TCIA04_253_1	0
Brats18_CBICA_ALV_1	1	Brats18_TCIA07_602_1	0
Brats18_CBICA_ALZ_1	1	Brats18_TCIA07_601_1	0
Brats18_CBICA_AMF_1	0	Brats18_TCIA07_600_1	0
Brats18_CBICA_AMU_1	0	Brats18_UAB_3446_1	1
Brats18_CBICA_ANK_1	1	Brats18_UAB_3448_1	1
Brats18_CBICA_APM_1	1	Brats18_UAB_3449_1	1
Brats18_CBICA_AQE_1	1	Brats18_UAB_3454_1	1
Brats18_CBICA_ARR_1	0	Brats18_UAB_3455_1	1
Brats18_CBICA_ATW_1	0	Brats18_UAB_3456_1	1
Brats18_CBICA_AUC_1	1	Brats18_UAB_3490_1	1
Brats18_CBICA_AUE_1	0	Brats18_UAB_3498_1	1
Brats18_CBICA_AZA_1	1	Brats18_UAB_3499_1	1
Brats18_TCIA13_652_1	0	Brats18_WashU_S037_1	0
Brats18_TCIA13_638_1	0	Brats18_WashU_W033_1	1
Brats18_TCIA13_617_1	0	Brats18_WashU_W038_1	1
		Brats18_WashU_W047_1	1

neural
networks
(long pe-
riods:
over 18
months)

BrainNet: 3D Local Refinement Network for Brain Tumor Segmentation

Xiaojun Hu, Weilin Huang, Deren Kong, Sheng Guo, and Matthew R. Scott

Malong Technologies

Abstract. We present an efficient yet conceptually simple 3D residual framework for automatic brain tumor segmentation. Our model is able to directly predict dense voxel segmentation of tumor regions in 3D brain MRIs *in a single shot*. The contributions of this paper are three-fold: (i) we propose a novel multi-level 3D refinement module that automatically aggregates both local details and spatial-temporal context information within 3D convolutional layers, leading to clear performance boost; (ii) we incorporate recent Focal loss into our framework, enabling our model to naturally cope with data imbalance that previously attempted in dense training and sampling; (iii) a new training strategy is introduced by leveraging curriculum learning, where we design the curriculum by incorporating data augmentation and the Focal loss. These technical developments are integrated seamlessly into a single 3D segmentation model, resulting in a highly-compact and end-to-end trainable model that can run at about 0.5s per MRIs - about 50 times faster than previous approaches [13, 9]. The proposed BrainNet is evaluated on the BRATS 2015 benchmark, where we achieve the state-of-the-art results, by surpassing recently published results reported in [13, 9]. It also obtains preliminary results on the training and validation data of the BRATS 2018 Challenge.

1 Introduction

Accurate brain tumor segmentation from MR images is of great importance for improving cancer diagnosis, surgery planning and prediction of patient outcome. However, manual segmentation of brain tumors is highly expensive, time-consuming and subjective. Efforts have been devoted to developing automatic method for this task, but it is still challenging to precisely identify some tumors (e.g., gliomas and glioblastomas), which are often diffused, poorly contrasted, and their boundaries are easily confused with healthy tissues. Furthermore, structural tumor regions, such as necrotic core, oedema and enhancing core, can appear in any location of the brain with various size and shape, making it particularly difficult to segment them accurately. To improve the performance, multiple MRI modalities, such as T1, T1-contrast, T2 and Fluid Attenuation Inversion Recovery (FLAIR), are often utilized to provide richer visual information, and automatic methods are developed to explore the multiple MRI modalities.

Past work in the literature has been dominated by approaches that pose brain tumor segmentation as the problem of semantic segmentation, which produces dense classification at the pixel level. Generally, hand-crafted features are designed by incorporating with a classifier learned separately, where the classifier does not impact the nature of the designed features. Recent deep convolutional neural networks (CNNs) have been applied to this task by advancing feature extraction. These approaches compute deep, hierarchical and learned features from brain MRIs, allowing the features to be learned jointly and collaboratively with an integrated classifier. This results in more meaningful features that lead to the state-of-the-art performance [7, 12, 9, 5, 13, 6]. However, recent CNN approaches for brain tumor segmentation often suffer from several common limitations that negatively impact their performance. First, CNNs are powerful to compute high-level context features with hierarchical designs, but multiple pooling operations give rise to a significant issue of down-scaling of the feature maps, leading to loss of fine structures and detailed information through the hierarchical convolutional layers. This local detailed information is critical to accurate segmentation. Second, segmentation often involves dense training and inference where training samples are highly correlated with neighboring pixels, significant data imbalance between various classes and background often happens. These limitations make it difficult to train a high-performance 3D segmentation model. Third, effectively aggregating meaningful spatio-temporal information from multi-modality 3D MRIs within a single CNN model is also challenging.

1.1 Contributions

In this work, we present a new BrainNet which is a 3D segmentation network built on deep residual architecture [8]. The approach we describe in this paper is to our knowledge the first to develop a single-shot 3D segmentation CNN for this challenging task. Our model is able to directly output dense voxel-level segmentation results of various tumor tissues from 3D brain MRIs, *without any post-processing*. The closely related work are that of [9] which inspired the current work, but design a different 3D convolutional architecture that combines multi-scale MRIs, with a CRF for post-processing. The current paper describes a method that integrates a number of key technical designs into a single CNN. Our contributions are described as follows:

- We propose a new 3D refinement module capable of aggregating rich fine-scale spatio-temporal features over multiple 3D convolutional layers. This allows it to explore both local detailed features and high-level context information in both spatial and temporal domain, which is critically important to achieving accurate segmentation.
- We introduce a new training strategy that incorporates curriculum learning and recent Focal loss into our 3D segmentation networks. This allows our model to learn more efficiently with the designed curriculum, where the issue of dense training and class imbalance are handled effectively and naturally.

- We integrate these technical improvements into a single model which allows for a direct prediction of dense voxel-level segmentation in a single shot. This results in a highly efficient model running at about 0.5s per MRIs - allowing for real-world application of BrainNet. The BrainNet obtains the state-of-the-art results on the BRATS 2015 [11]. It also obtains preliminary results on the training and validation data of the BRATS 2018 Challenge [11, 3, 1, 2].

1.2 Related Work

Convolutional neural networks have recently been applied to brain tumor segmentation in brain MRIs [7, 12, 9, 5, 13, 6]. For example, Havaei *et. al.* [7] explored fully convolutional networks (FCN) for brain tumor segmentation on 2D MRI slice, where two-phase training procedure and cascaded architecture were developed to tackle the issue of class imbalance. Similarly, boundary-aware FCN was developed to improve segmentation accuracy by explicitly incorporating boundary information[12]. These approaches were mainly designed for 2D segmentation of individual MRI slices, while our work focuses on proposing an efficient 3D module able to learn rich spatio-temporal features directly from 3D MRIs.

Our work is closely related to that of [9, 5], where 3D CNNs were developed for this task. Chen *et. al.* developed VoxResNet [5] for brain lesion segmentation, which focuses on integrating the features learned from multiple MRI modalities. Kamnitsas *et. al.* [9] also proposed a dual pathway 3D CNN for aggregating multi-level features, and CRF was applied to refine the results. Our BrainNet learns multi-mobility information and aggregates multi-level 3D convolutional features with a single model that produce inference in one shot, providing a more principled solution that work more efficiently.

2 BrainNet: 3D Local Refinement Network

We describe details of the BrainNet, including the proposed 3D refinement module and a new training strategy that integrates curriculum learning [4], Focal loss [10] and data augmentation. Our goal is to accurately predict the label of four tumor tissues at each voxel, by using four 3D MRI modalities: T1, T1-contrast, T2 and FLAI, as shown in Fig. 1. The four tumor tissues are necrotic core, oedema, non-enhancing and enhancing core, as defined in [11]. An exemplar image is shown in Fig. 1. In this work, we cast ResNet, originally designed for image classification, into 3D dense segmentation for brain MRIs. Essentially, segmentation task can be considered as a dense classification problem implemented on each image patch corresponding to a dense pixel at the output layer. Specifically, we make two main modifications on the original ResNet consisting of four convolutional blocks by using 3×3 convolutional kernels. First, we introduce 3D convolutions with the kernel size of $3 \times 3 \times 3$ in all convolutional layers, as shown in Fig. 1. The 3D convolutional filters naturally take all four MIR modalities as an input, by considering each modality as a 3D channel like the R, G, or

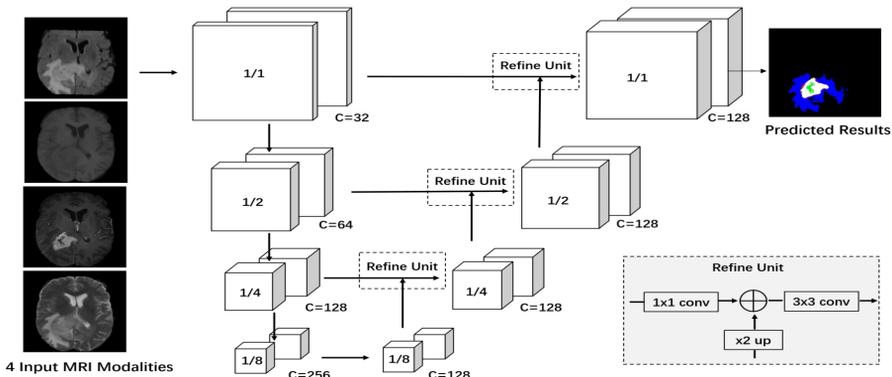


Fig. 1: Architecture of the proposed BrainNet. It has four convolutional blocks containing 19 layers with residual units.

B channel in a RGB image. Second, the output prediction maps should have an exactly same spatial size as the input MRIs, to ensure a dense prediction. The output layer is built on the last convolutional layer by computing a 5-class soft-max function at each spatial location. Therefore, the spatial resolution of the last convolutional layer should be amplified and aligned to that of input volume. This is achieved by employing the proposed 3D refinement module which is described next.

2.1 3D Refinement Module

For dense 3D segmentation, one of the key steps is to align the spatial shape of the convolutional features to the input volume; a straightforward approach is to use an up-sample operation. However, the up-sample operation commonly suffers from a significant issue of downscaling of the feature maps, leading to loss of fine structures and local details through the convolutional layers. Both fine structures and high-level context information are important to segmentation task. To encode both of them in the convolutional features, we propose a novel 3D refinement module, with the goal of aggregating local details and multi-level high-level context information in both spatial and temporal domains.

Specifically, BrainNet has four convolutional blocks, with down-sample factors of $\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}\}$ and 3D channel numbers of $\{32, 64, 128, 256\}$, as shown in Fig. 1. The proposed 3D refinement module is applied to each convolutional block, and encodes high-level context information in both spatial and temporal domains recursively, from the higher layers to the lower ones. It works as follows: (1) we design an adaptive layer which is connected to the output of each convolutional block. The adaptive layer reshapes the convolutional maps by changing the number of 3D channels to a same number of 128, while keeping their original shape of the convolutional features in each channel. (2) multi-level convolutional features are fused recursively from the top convolutional block to the bottom one, where

element-wise summation is used, with an $\times 2$ up-sample operation implemented on the lower-resolution maps. This results in 128-channel 3D convolutional maps in the last refined convolutional layer, where each 3D convolutional maps have a same shape with the input 3D volume. The final 5-class voxel-wise prediction is computed on the 128D features at each voxel location of the 3D features maps. We adopt $1 \times 1 \times 1$ kernels for all adaptive layers for simplicity, but more advanced configurations, such as non-local operation, dilated convolutions, or inception architecture, are readily applicable, and has the potential to improve the performance. Our goal here is to present a fundamental framework of the 3D refinement module. Finally, the 3D refinement module generates high-resolution 3D semantic feature maps, which enrich feature representation. It can be integrated seamlessly into our 3D segmentation network.

2.2 Training 3D BrainNet with Designed Curriculum

An efficient training approach is critically important to exploring the full potential of our 3D segmentation model, with a very limited of MRIs. We describe a new training strategy for end-to-end learning of 3D BrainNet, by incorporating with curriculum learning [4], Focal loss [10] and data augmentation. We emphasize that high performance of our 3D model is not only based on our model design, but also due to our new training strategy. As mentioned, our segmentation task can be considered as dense classification problem, and then training loss of BrainNet is computed *densely* over all spatial-temporal locations in 3D MRI volume, gave rise to a number of significant issue. Firstly, dense 3D training generates a large number of training samples which are significantly redundant by learning from neighboring locations in spatial and temporal domains. These samples are closely relevant with less diversity, and thus are less informative. Secondly, training would be highly inefficient when most sampling locations have easy samples which would contribute less useful signal for learning. This happens in dense 2D image detection [10], but would become more significant for our 3D segmentation task. Our training approach is developed to cope with these issues.

Focal Loss As stated in [10], using automatically-selected meaningful samples is critical to learn a high-capability model for a dense training task. Focal loss was originally introduced in [10] for detection task. It encourages the model to learn from a sparse set of hard samples. This naturally alleviates negative impact from the vast amount of easy samples, leading to performance boost. Formally, the Focal loss is defined by introducing a modulating factor (γ) to the cross entropy loss, with a parameter (α) for class balancing [10]: $FL(p_t) = -(1 - p_t)^\gamma \log(p_t)$. $p_t = p$ if $y = 1$, and otherwise $p_t = 1 - p$. $y \in \{-1, +1\}$ is the ground-truth class, and $p \in [0, 1]$ is the estimated probability for the class with label $y = 1$. γ is focusing parameter, and the Focal loss is equal to original cross entropy loss with $\gamma = 0$, and the training focuses on hard samples when $\gamma > 0$. It down-weights easy samples smoothly which have a high value of p_t , indicating a high estimated probability for the correct class. A larger value of γ means more contribution

from the hard samples to the training process. In our BrainNet, we cast Focal loss into our 3D segmentation framework by replacing soft-max loss, providing a simple formulation that allows the model to automatically select a spare set of meaningful samples for learning.

Data Augmentation The training data is 3D volumetric MRIs, and a straightforward approach to increase the amount of training data is through data augmentation. This allows us to generate massive scale training data with increasing diversity. Our data augmentation is produced as follows. First, a simple slice-level augmentation is implemented by randomly amplifying color values. Second, we produce volume-level augmentation where random operations are implemented through all slices within a volume: (i) all slices are rotated with a random orientation from $[90, 90]$ degrees, and are re-scaled by using a random ratio from $[0.7; 1.3]$; (ii) horizontal and vertical flipping are implemented sequentially, with a probability of 50% for each operation; (iii) a random spatial cropping is further produced, and each cropped region should include a whole region of tumor if it is presented in current slice.

Learning with Designed Curriculum Both Focal loss and data augmentation encourage the model to learn from data with more diversity and complexity. However, as shown in our experiments, directly applying these technologies to our 3D segmentation model is difficult to deliver obvious performance gains, and an efficient learning scheme is critical. Our training scheme is inspired from the intuition of curriculum learning, which encourages the learning to start from an easier task, and then takes more complex tasks gradually during training process [4]. We propose a three-stage learning curriculum where data complexity is increased gradually. This allows for an efficient implementation of curriculum learning designed for our task. (i) Our 3D model is trained on the original data without data augmentation and Focal loss. (ii) Then we increase data complexity by applying data augmentation to original videos with probability of 50%. (iii) the model is learned from harder samples by employing Focal loss with stronger data augmentation applied to 75% of training volumes. The three-stage curriculum allows the model to learn properly from augmented data, leading to stronger generalization capability with clear performance boost.

2.3 Implementation Details

The BrainNet is implemented by using Pytorch. In training stage, we equally sample 3D patches with size of $12*128*128$ from lesions and background. The BrainNet was trained on 4 Titan Xp GPUs for 500k iterations, with a batch size of 40 (about 8G memory for each GPU used). We adopt the Adam to optimize our model and set the learning rate to $1e-4$. It takes about 8 hours to train a single model. In inference stage, we split one sample into five parts with size of $31*240*240$ in order to produce inference with a single GPU, due to the memory limitation. We predict each part individually and concatenate

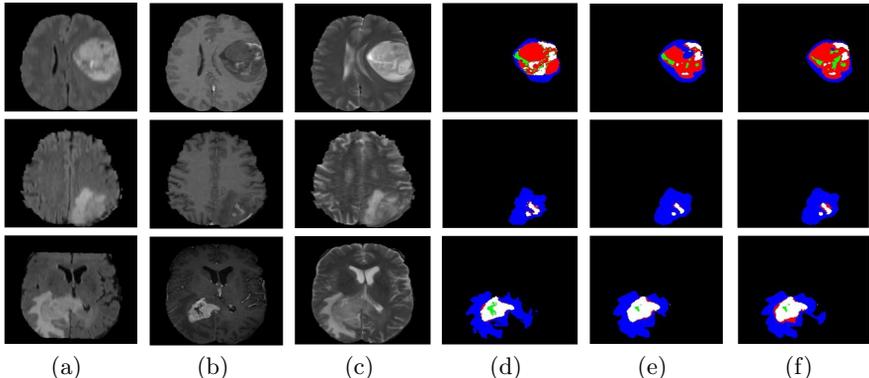


Fig. 2: Segmentation results on several examples. (a) FLAIR, (b) T1-contrast, (c) T2, (d) ground truth (e) BrainNet, and (f) BrainNet with Focal loss. Colors: necrotic core (red), oedema (green), non-enhancing core (blue), and enhancing core (white).

their outputs to form the final results. Each MRI scanning takes about 500ms to predict, including the time of data preparation. We get the final predict labels by taking the maximum probability of each class, without any post-processing.

3 Experimental Results

Table 1: Evaluation results of 110 testing cases in BRATS 2015 testing dataset, with comparisons with the most recent results reported in [9, 13]

	Dice			Precision			Sensitivity		
	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core	Enh.
Zhao et. al.[13] + CRF	82.0	72.0	62.0	84.0	78.0	60.0	83.0	73.0	69.0
Zhao et. al.+3D CRF [13]	84.0	73.0	62.0	89.0	76.0	63.0	82.0	76.0	67.0
DeepMedic [9]	83.6	67.4	62.9	82.3	84.6	64.0	88.5	61.6	65.6
DeepMedic+CRF [9]	84.7	67.0	62.9	85.0	84.8	63.4	87.6	60.7	66.2
BrainNet_no_FL	85.0	69.0	64.0	88.0	86.0	63.0	85.0	63.0	69.0
BrainNet	86.0	72.0	64.0	87.0	85.0	64.0	87.0	68.0	66.0

we evaluate our 3D BrainNet on the BRATS 2015 [11] and BRATS 2018 Challenge [11, 3, 1, 2]. (1) For BRATS 2015 [11], the training set has 220 cases with high grade (HG) and 54 cases with low grade (LG) glioma, and segmentations GT is provided. The task is to segment four tumor tissues: necrotic core, oedema, non-enhancing and enhancing core. The test set includes 110 cases. We follow [9] by merging the four predicted labels into different sets of *whole* tumor

(all classes), the *core* (class 1,3,4), and the *enhancing* tumor (class 4). FLAIR, T1, T1-contrast and T2 modalities are available. (2) On BRATS 2018 Challenge [11, 3, 1, 2], 285 cases are included in the training set, and the validation set has 66 cases. More details of the BRATS 2018 data are described in [11, 3, 1, 2].

Experimental results. Our model is evaluated on the full test set of the BRATS, in the terms of standard *Dice* (f-measure), *Precision* and *Sensitivity*, which are computed on the online evaluation platform. The results and comparisons are presented in Table 1, and the predicted results on several examples are demonstrated in Fig. 2. As shown in Table 1, our BrainNet with Focal loss achieved the best performance on the whole tumor and the enhancing tumor in the term of *Dice*, which is the most important measure that balances *precision* and *sensitivity*. BrainNet with Focal loss leads to clear performance improvements, particularly for the core tumor. As demonstrated in Fig. 2, Focal loss allows the model to find finer-grained details on the boundary of tumor. By comparing with recent state-of-the-art results reported in [9, 13], BrainNet obtained improvements of about 2%, which are significant for this challenging task. Furthermore, as reported in [13] and [9], both of them employed CRF as a post-processing to improve the performance, resulting in about 1-2% performance gains. Our BrainNet is a single-shot model without any post-processing, and thus is a more efficient approach that can run at about 0.5s per MRI scanning using a single GPU, which can be applied to real-world applications. This is about $\times 50$ and $\times 240$ faster than that of [9] (30s) and [13] (2-4 mins).

On the BRATS 2018 Challenge, we evaluate BrainNet on both training set and validation set. On the training set, we compute a mean value of 5-fold cross-validations on the total 285 cases (228 cases for training and the remained 57 cases for testing for each validation). The mean *Dice* scores for *Whole*, *Core*, *Enhancing* are 87.8%, 81.9% and 72.28%, respectively. On the validation set, we train the whole training set, and test BrainNet on the validation set, we obtain mean *Dice* scores of 88.2%, 81.2% and 76.0%, for *Whole*, *Core*, *Enhancing* respectively. As can be found, the evaluation results on the training set and validation set are generally consistent. Comparing to the BRATS 2015, the mean *Dice* scores of all three terms are improved on the BRATS 2018, due to the increased number of the training samples. The performance on the BRATS 2018 will be further optimized, and the results on the test set will be reported when it is available.

4 Conclusions

We have presented a new 3D segmentation network, BrainNet, for automatic segmentation of brain tumor in 3D MRIs. we proposed a novel 3D refinement module that directly aggregates both local details and spatial-temporal context information within 3D convolutional layers. We introduced a new training strategy that incorporates curriculum learning, Focal loss and volumetric data augmentation, allowing for a better generalization of the trained model. These technical improvements are integrated elegantly into a single 3D segmentation

model, resulting in a highly-compact and end-to-end trainable model that can run at about 0.5s per MRI volume. The BrainNet obtained the state-of-the-art performance on the BRATS 2015, and preliminary results on the BRATS 2018 was reported.

References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., C, C.D.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive* (2017)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., C, C.D.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *The Cancer Imaging Archive* (2017)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Nature Scientific Data* (2017)
4. Bengio, Y., Louradour, J., Collobert, R., Weston, J.: Curriculum learning (2009), in *ICML*
5. Chen, H., Dou, Q., Yu, L., Qin, J., Heng, P.A.: Voxresnet: Deep voxelwise residual networks for brain segmentation from 3d mr images. *NeuroImage* (2017)
6. Fidon, L., Li, W., C., L., Garcia-Peraza-Herrera, Ekanayake, J., Kitchen, N., Ourselin, S., Vercauteren, T.: Scalable multimodal convolutional networks for brain tumour segmentation (2017), in *MICCAI*
7. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H.: Brain tumor segmentation with deep neural networks. *Medical Image Analysis* 35, 18–31 (2017)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2016), in *CVPR*
9. Kamnitsas, K., Lediga, C., Newcombe, V.F., Simpson, J.P., Kaneb, A.D., Menon, D.K., Rueckert, D., Glockera, B.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical Image Analysis* 36, 61–78 (2017)
10. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection (2017), in *ICCV*
11. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Y. Burren, N.P., Slotboom, J., R. Wiest, e.a.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans. on Medical Imaging* (2015)
12. Shen, H., Wang, R., Zhang, J., McKenna, S.J.: Boundary-aware fully convolutional network for brain tumor segmentation (2017), in *MICCAI*
13. Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., Fan, Y.: A deep learning model integrating fcnn and crfs for brain tumor segmentation. *Medical Image Analysis* 43, 98–111 (2018)

Multi-level Activation for Segmentation of Hierarchically-nested Classes on 3D-Unet

Xiaobin Hu^{1,2}, Marie Piraud¹

¹*Department of Computer Science, Technische Universität München, Munich, Germany*

²*Xiaobin.hu@tum.de*

Abstract

For a plenty of segmentation tasks, especially for the biomedical image, the topological prior is vital information which is useful to achieve a better segmentation result. The containment/nesting is a typical inter-class geometric relationship, and it is comparably more important to be considered. In this challenge, the nested classes relationship is introduced into the 3D-dilated-Unet architecture. The network comprises a context aggregation pathway and a localization pathway, which encodes increasingly abstract representation of the input as going deeper into the network, and then recombines these representations with shallower features to precisely localize the interest domain via a localization path. The nested-classes-prior is combined by proposing the multi-class activation function and its corresponding loss function. The model is only trained on the training dataset of Brats2018, and 20% dataset is regarded as the validation dataset. The performance of validation process on the leaderboard is 86%,78% and 70% Dice score for whole tumor, enhancing tumor and tumor core, respectively.

Key words: Topological prior, nested classes, 3D-dilated-Unet, multi-class activation function

1. Introduction

There exist many different models for segmentation tasks, but very few CNN models are encoded the nested-topologic-prior. From the literature [1-7], three main avenues are illustrated as following: (1) cascaded geometries, the network model is modified to accommodate the hierarchical information via training successive segmentation network for hierarchical segmentation target. (2) Comprising the nested-classes information into the loss function, and penalize the result which is not respect the nested geometry relationship. (3) The conditional and Markov Random Fields are employed as post-processing ways to integrating label context prior. All aforementioned methods handle the nesting of classes in a rather indirect way, and consequently we propose a directly new activation function to segment the hierarchically nested labels.

2. Methodology

The nested-classes relationship between different labels are showed in Fig. (1). The general network structure shown in Fig.(2) is stemming from the Fabian Isensee Network[8] to process large 3D input blocks of 144x144x144 voxels. The original network is inspired by the U-net which allows the network to intrinsically recombine different scales throughout the entire network. This vertical depth is set as 5, which balance between the spatial resolution and feature representations. The context module is a pre-activation residual block, and is connected by 3x3x3 convolutions with input stride 2. The purpose of localization pathway is to extract features from lower levels of the network and change it into a high spatial resolution by employing the means of a simple upscale technology. The upsampled features and its corresponding level of the context aggregation feature are recombined via concatenation. Furthermore, the localization module, consisting of a 3x3x3 convolution followed by a 1x1x1 convolution, is designed to gather these features. The deep supervision is introduced in the localization pathway pathway by integrating segmentation layers at different levels of the network and combining them via elementwise summation to form the final network output. The output activation layer is multi-class activation layer substituting with Softmax layer converting the multi-classes problem to binary ones.

The multi-class-nested activation function shown in Fig.(3) is defined as

$$a(x) = \sum_{n=1}^m \sigma \left(\kappa \left[x + h \left(n - \frac{m+1}{2} \right) \right] \right) \quad (1)$$

Where σ is the sigmoid function, κ is the steepness and h is the spacing between consecutive sigmoids. For Brats 4-classes nested label case, assuming the $m+1=4$, h is 0.5 and steepness κ is 10. It is illustrated in the Fig.(1).

The corresponding loss function is defined as

$$L_{NCE} = -\frac{1}{N_{tot}} \sum_{pixels} \sum_{i \text{ classes } c} y_i^c w^c \log(Q^c[a(x_i)]) \quad (2)$$

w^c is the weight of corresponding label, N_{tot} is the sum number of pixels, and $y^c = 1$ for the ground-truth label c of pixel i and $y^c = 0$ otherwise.

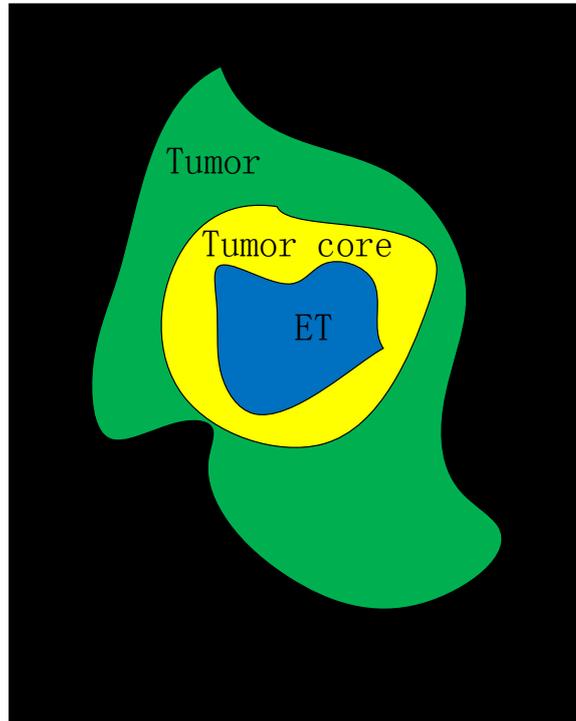


Fig.1 Nested-relationship sketch

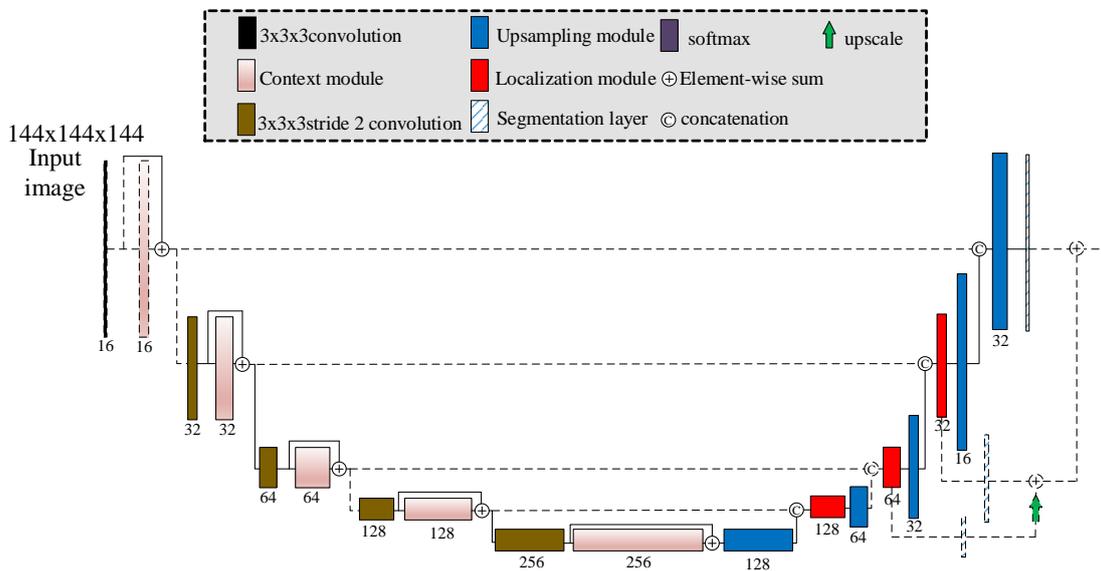


Fig.2 Network architecture: Context pathway(left) aggregates high level information; Localization pathway(right) localizes precisely.

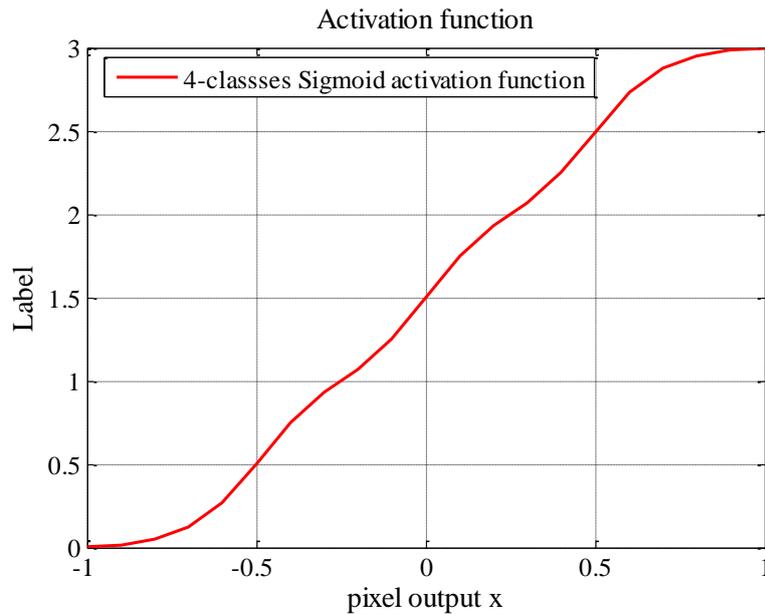


Fig.3 Multi-class activation function

3. Experiment results

As mentioned before, a hierarchically-nested multi-classes network based on the residual 3D Unet is proposed. Some slice of segmentation results containing the tumor, tumor core and enhancing core are shown in Fig.(4)(a,b,c,d). From the images, the topology geometry between different labels is constrained into the nested-classes relationship, and consequently avoiding the error from lack of prior in topology geometry. We train the networks using ADAM with a learning rate 0.0005, and to regularize the network, we use early stopping on the validation set and dropout in all residual block before the multi-classes sigmoid function. Overall, the approach, shown in Table 1 reached the result 84% for complete tumor, 76% for tumor core and 66% for enhancing core.

In BRATS dataset 2018 [9,10], there are 4 labels, Necrotic core, Edema, Non-enhancing core and Enhancing core. The dataset contains 4 different modalities for MRI, native(T1), post-contrast T1-weighted(T1Gd), T2-weighted(T2), and T2 Fluid Attenuated Inversion Recovery(FLAIR) .

The definition of three kinds of evaluation criteria, dice score, sensitivity and specificity are illustrated as:

$$Dice\ coefficient : DC = \frac{2|A \cap B|}{|A| + |B|} \quad (3)$$

And

$$Hausdorff\ Distance : H(A, B) = \max \{ \min \{ d(A, B) \} \} \quad (4)$$

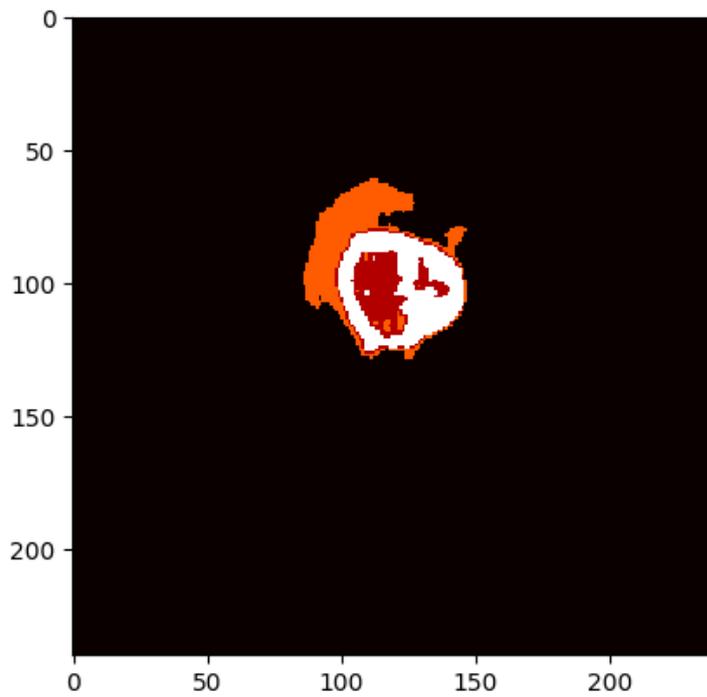
And

$$Sensitivity = TPR = \frac{TP}{TP + FN}$$

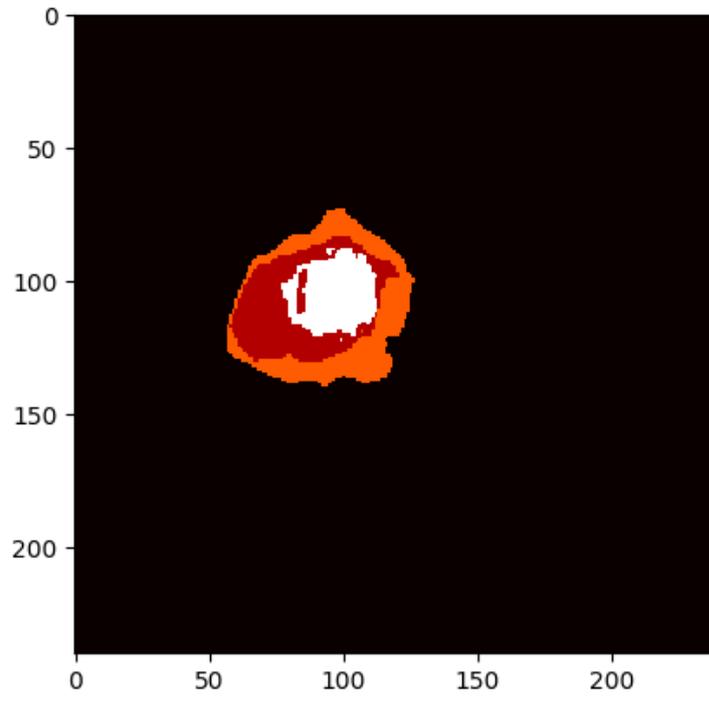
$$Specificity = TNR = \frac{TN}{TN + FP}$$
(5)

Table 1. Evaluation result of BraTS 2018

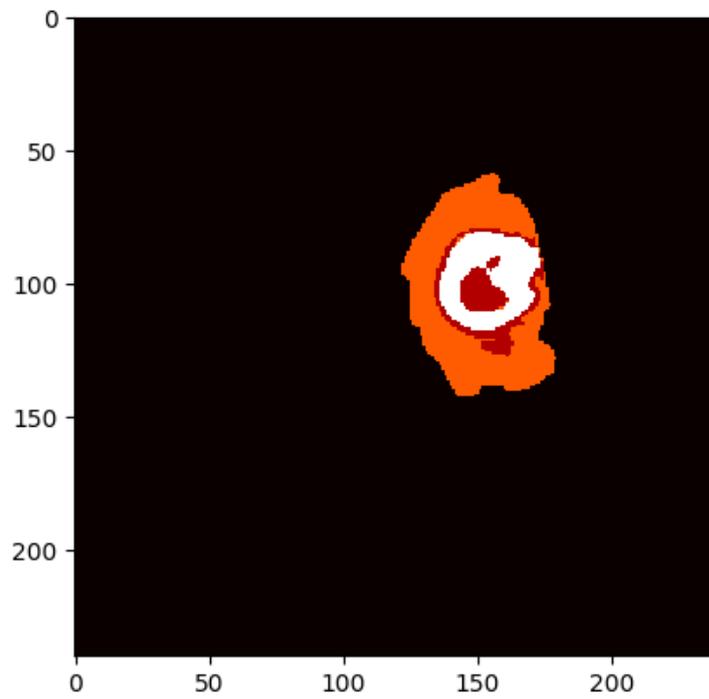
	Dice score			Sensitivity			Specificity		
	Enh	Whole	Core	Enh	Whole	Core	Enh	Whole	Core
Val	0.70	0.86	0.78	0.79	0.88	0.83	0.99	0.99	0.99



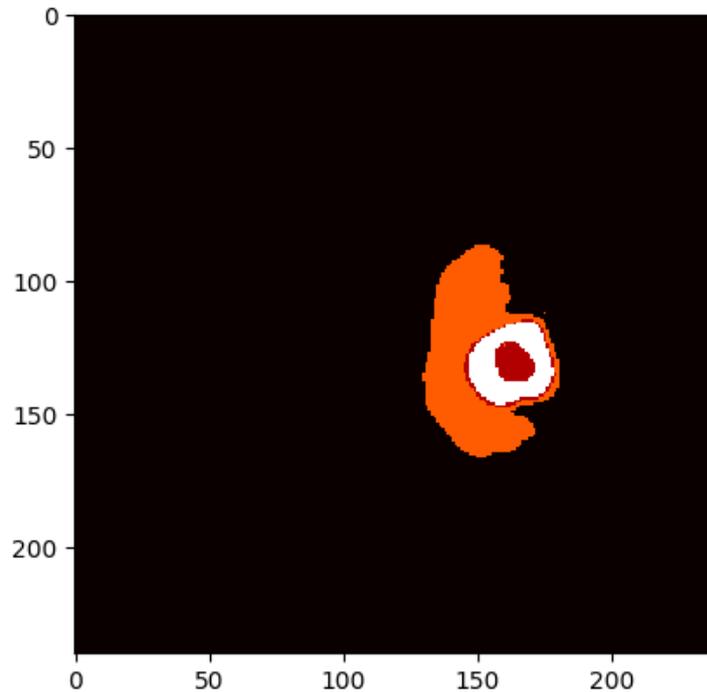
(a)



(b)



(c)



(d)

Fig.4. Segmentation result

4. Conclusions

In this paper we introduce the technique of multi-level activation for nested classes segmentation. The experiment results indicate that the multi-level activation function and its corresponding loss function are efficient compared with Softmax output layer based on the same network framework. Finally, the multi-level activation layer can be straightforwardly generalized to a deeper nesting hierarchy, and also easily applied into the complex and efficient deep learning network.

References

1. Nosrati, M.S., Hamarneh, G: Local Optimization Based Segmentation of SpatiallyRecurring, Multi-Region Objects With Part Configuration Constraints. *IEEE Transactions on Medical Imaging* 33(9) (sep 2014) 1845-1859
2. BenTaieb, A., Hamarneh, G: Topology Aware Fully Convolutional Networks for Histology Gland Segmentation. In Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W., eds.: *MICCAI 2016. Lecture Notes in Computer Science*, Cham, Springer International Publishing (2016)
3. Christ, P.F., Elshaer, M.E.A., Ettliger, F., et al.: Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D

Conditional Random Fields. In Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W., eds.: MICCAI. Volume 9900 of Lecture Notes in Computer Science., Springer International Publishing (2016)

4. Fidon, L., Li, W., Garcia-Peraza-Herrera, L.C., et al.: Generalised Wasserstein Dice Score for Imbalanced Multi-class Segmentation using Holistic Convolutional Networks. (jul 2017) 1-12

5. Bauer, S., Tessier, J., Krieter, O., et al.: Integrated spatio-temporal segmentation of longitudinal brain tumor imaging studies. In Menze, B., Langs, G., Montillo, A., Kelm, M., Müller, H., Tu, Z., eds.: Medical Computer Vision. Large Data in Medical Imaging, Cham, Springer International Publishing (2014) 74{83

6. Alberts, E., Charpiat, G., Tarabalka, Y., et al.: A Nonparametric Growth Model for Brain Tumor Segmentation in Longitudinal MR Sequences. In: MICCAI Brain Lesion Workshop. (2015) 69{79

7. Liu, Z., Li, X., Luo, P., et al.: Deep Learning Markov Random Field for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 8828(c) (2017) 1-1

8. Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, Klaus H. Maier-Hein, Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge, MICCAI BraTs Challenge(2017).

9. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp I, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694.

10. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117.

11. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q.

12. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF.

Brain Tumor Segmentation on Multimodal MRI Using Multi-Level Upsampling in Decoder

Yan Hu¹, Xiang Liu³, Xin Wen³, Chen Niu³ and Yong Xia^{1,2} (✉)

¹ Shaanxi Key Lab of Speech Image Information Processing (SAIIP),
School of Computer Science and Engineering,
Northwestern Polytechnical University,
Xi'an 710072, People's Republic of China

² Centre for Multidisciplinary Convergence Computing (CMCC),
School of Computer Science and Engineering,
Northwestern Polytechnical University,
Xi'an 710072, People's Republic of China

³ The First Affiliated Hospital of Xi'an Jiao Tong University
Xi'an 710072, People's Republic of China
yxia@nwpu.edu.cn

Abstract. Accurate brain tumor segmentation plays a pivotal role in clinical practice and research settings. In this paper, we propose the multi-level up-sampling network (MU-Net) for automated segmentation of brain tumors, including necrosis, edema, non-enhancing, and enhancing tumor, in multimodal magnetic resonance (MR) sequences. This model has an encoder-decoder structure. In this model, low level feature maps obtained by the encoder and high level feature maps obtained by the decoder are combined by using a newly designed global attention (GA) module. The proposed model has been evaluated on the BraTS 2018 Challenge validation dataset and achieved the Dice similarity coefficients of 0.87, 0.72 and 0.66 in the segmentation of whole tumor, core tumor and enhancing tumor, respectively. Our results indicate that the proposed model has a promising performance in automated brain tumor segmentation.

Keywords: Magnetic resonance imaging, brain tumor segmentation, encoder-decoder, multi-level upsampling, global attention

1 Introduction

Shape and localization of brain tumors are crucial for diagnosis, treatment planning and follow-up observation in clinical, while the manual segmentation of brain tumor in MRI requires a high degree of skills and concentration, and is time-consuming, expensive and prone to operator bias. Thus, a fully automated and reliable computer-aided segmentation algorithm is of great advance and significance.

As illustrated by Menze et al. [1], in the last decades, considerable research efforts have been devoted to this task. However, due to the variable shapes and locations, dif-

fusion and poor contrast of brain tumors, it remains a challenge to automatically segment brain tumors. In recent years, deep learning techniques have shown their advanced ability to feature presentation and classification, and hence the mainstream has changed from traditional machine learning algorithms to deep learning algorithms. Convolutional neural networks (CNNs) learn advanced presentation of images and segment the image simultaneously, have been widely used in brain tumor segmentation. One stream is patch-wise network. Zhao et al. [2] proposed a three-convolutional-pathway network, in which the input patches for three pathways have a size of 48×48 , 28×28 and 12×12 , respectively, and concatenated these three outputs for classification. Kamnitsas et al. [3] extended a 3-D CNN architecture, i.e. DeepMedic, with multi-scale, residual connections and fully connected conditional random field. Hu et al. [4] used a cascaded U-Net model and a patch-wise CNN to detect and segment brain tumors. Another stream is based on fully convolutional networks (FCNs). Pereira et al. [5] employed two U-Nets - one for the localization of tumors and the other for the segmentation of intra-tumor structures. Li et al. [6] used three parallel end-to-end networks for three views and generated the segmentation results using majority voting. Kamnitsas et al. [7] trained seven end-to-end networks and used ensemble learning to produce segmentation results.

In this work, we propose a FCN called the multi-level upsampling network (MU-Net) to segment brain tumor structures, including necrosis, edema and enhancing tumor from multimodality MRI. The main contributes are: (a) we designed a global attention module (GA) to combine the low level feature from encoder and high level feature from decoder ; (b) we designed a multi-level decoding architecture. The proposed algorithm has been evaluated on the BraTS 2018 Challenge validation dataset and achieved a promising result.

2 Dataset

The proposed MU-Net model was evaluated on the Brain Tumor Segmentation 2018 (BraTS 2018) Challenge dataset. There are 285 cases for training, including 210 HGG and 75 LGG cases [8-10]. Each case has four multimodal MR scans, including the T1, T1c, T2, and FLAIR. All these scans were co-registered to the same anatomical template, interpolated to the same dimension of $240 \times 240 \times 155$ and the same voxel size of $1.0 \times 1.0 \times 1.0 \text{ mm}^3$, and skull-stripped. Each case has been segmented manually, by one to four raters, following the same annotation protocol, and their annotations were approved by experienced neuro-radiologists. Annotations of tumor tissues comprise the enhancing tumor (ET-label 4), the peritumoral edema (ED-label 2), and the necrotic and non-enhancing tumor core (NCR/NET-label 1). The validation dataset consists of 66 cases, but their grade and ground truth are unseen.

3 Methods

3.1 Segmentation Model

The proposed MU-Net model adopts the encoder-decoder structure, consisting of five convolutional blocks, a spatial pyramid pooling (SPP) module [11], five global attention (GA) modules, and nine upsampling feature (UF) modules. The architecture of this model is shown in Fig. 1.

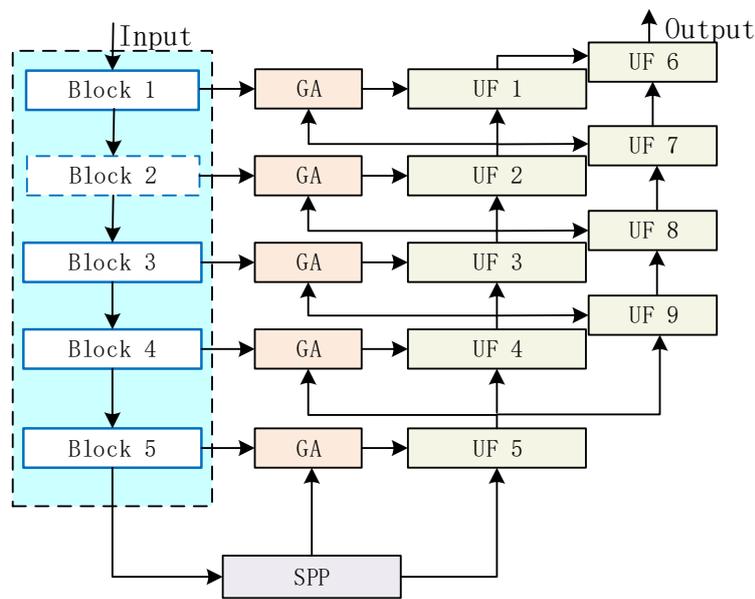


Fig. 1. Architecture of the proposed MU-Net model

The encoder branch is a variants of ResNet-101. The convolutional layer with $64 \times 7 \times 7$ kernels and a stride of 2 in the root block (i.e. Block 1) is replaced with five convolutional layers, each consisting 64 3×3 kernels. The stride of the third convolutional layer is 2, and the stride of other convolutional layers is 1. Other blocks in this branch is the same as those in ResNet-101 [12].

Between the encoder and decoder branches, we add a SPP module, in which there are five parallel operators, including three 3×3 dilated convolution with a dilation rate of 6, 12, and 18, respectively, a 1×1 convolution and a global pooling (see Fig. 2(a)). The input of the SPP module is processed by these five operators simultaneously, and feature maps generated by these operators are concatenated as the output of the SPP module.

The major part of the decoder branch contains five Decode modules (i.e. UF 1 – UF 5), which are designed to recover the size of feature maps. Usually, there are two 3×3 convolutions and a bilinear interpolation between them in each UF module (see Fig. 2(c)). However, since there is no down-sampling operation in the encoder block 3-5,

the interpolation operation is omitted in the UF 5, FU 4, and UF 1 modules such that the output feature maps have the same size as the input of the MU-Net model. Meanwhile, to combine low-level feature maps and high-level feature maps in the decoding process, we add five GA modules to the MU-Net model. Each GA module takes two groups of inputs - low-level feature maps from the corresponding encoder block and high-level feature maps from the UF module at the previous level. Two 3×3 convolutions are applied to low-level feature maps, respectively. High-level feature maps are also processed by two operations – one is the global average pooling followed by a 1×1 convolution as, and the other is a 3×3 convolution. The processed high-level feature maps are then used as the element-wise weighting mask of the processed low-level feature maps (see Fig. 2(b)). In addition, the output of each of UF 2 - UF 5 are fed simultaneously to the UF module at the next level. Eventually, the output of the last extra UF module and the output of UF 1 are concatenated and fed to another UF module to produce the segmentation results.

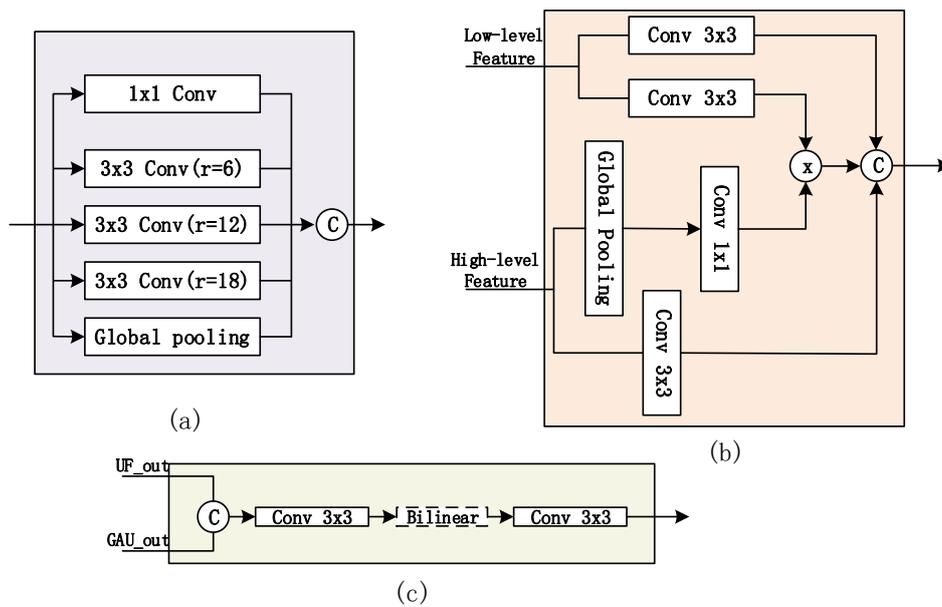


Fig. 2. Architecture of modules used in segmentation model. (a) shows the SPP module;(b) shows GA module; (c) shows the UF module.

3.2 Implementation

With the proposed MU-Net model, brain tumor segmentation can be performed on a slice-by-slice basis. Hence, the transvers slices in each training data were cropped to 224×224 , such that the pre-trained ResNet-101 can be used to initialize the encoding branch. We randomly selected positive slices (with tumor) and negative slices (without

tumor) at a rate of 5:1. We used the cross entropy as the loss function, and adopted the adaptive moment estimator (Adam) with an exponentially descending learning rate of 0.001 ~ 0.00001 as the optimizer.

4 Experiments and Results

4.1 Experiments Settings

Following the request of the challenge, four intra-tumor structures have been grouped into three mutually inclusive tumor regions: (a) whole tumor (WT) that consists of all tumor tissues, (b) tumor core (TC) that consists of the enhancing tumor and necrotic and non-enhancing tumor core, and (c) enhancing tumor (ET). The performance of segmenting each tumor region was quantitatively evaluated through an online system by using three metrics, including the average Dice similarity coefficient, sensitivity and Hausdorff distance.

4.2 Results

Preliminary results for the BraTS 2018 Training dataset have been obtained by hold-out using 75% of the data (228 cases) for training and the remaining 25% for validation (57 cases). Table 1 show the quantitative evaluation and Figure 3 shows example segmentation results on predicted cases from BraTS 2018 training data. It appears as though prediction of whole tumor is reasonable, while the intra structures segmentation are not good enough.

The evaluation of our algorithm's performance on BraTS 2018 validation set are presented in Table 2 and Figure 4. We can observe that performance on both the training and validation data are consistent, which indicates that this model generalizes well to unseen examples.

Table 1. Quantitative result of Validation on BraTS 2018 training Set.

	Dice			Sensitivity			Hausdorff-95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC
Mean	0.61	0.83	0.73	0.83	0.89	0.75	41.48	47.23	41.14
StdDev	0.27	0.11	0.17	0.17	0.09	0.21	37.49	23.49	28.90
Median	0.72	0.86	0.77	0.88	0.91	0.82	41.69	46.7	44.77
25quantile	0.50	0.78	0.64	0.80	0.85	0.62	5.12	30.36	12.37
75quantile	0.80	0.90	0.86	0.95	0.96	0.93	61.25	59.84	59.67

Fig.

3.

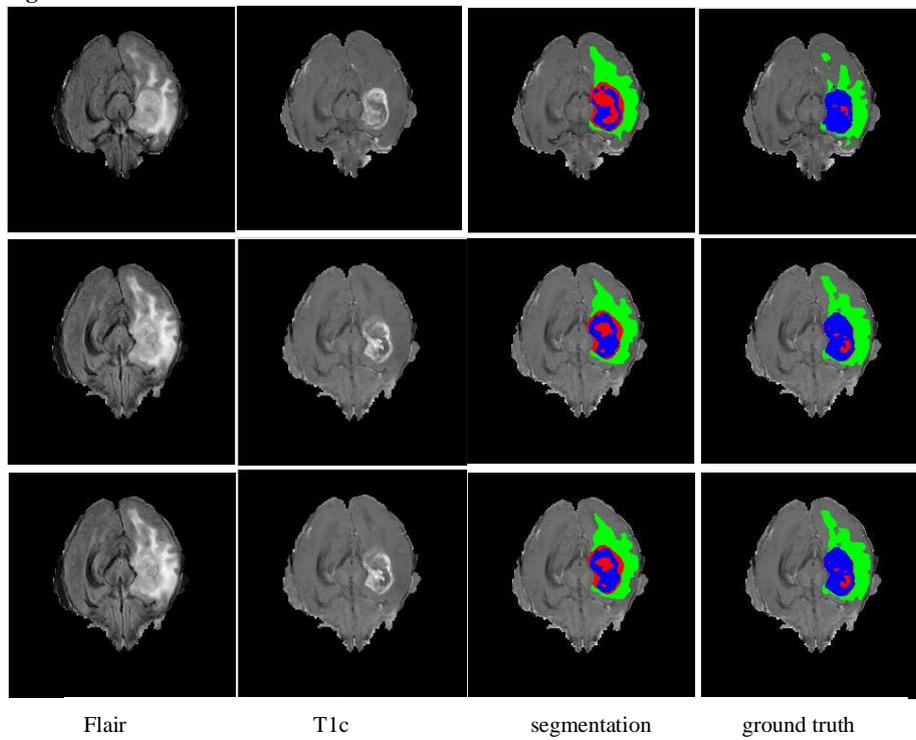


Fig. 4. Segmentation examples from the validation set. From top to bottom: the 55th, 57th, and 59th slices from the subject Brats18_TCIA01_147_1. *Red* - NCR&NET, *Blue* - ET, *Green* - ED.

Table 2. Quantitative result on BraTS 2018 validation set.

	Dice			Sensitivity			Hausdorf-95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC
Mean	0.66	0.87	0.72	0.63	0.84	0.73	7.56	6.73	15.74
StdDev	0.29	0.11	0.24	0.28	0.16	0.25	14.65	9.71	21.97
Median	0.79	0.90	0.80	0.72	0.89	0.82	2.24	3.61	8.75
25quantile	0.54	0.87	0.67	0.49	0.82	0.64	1.73	2.83	4.06
75quantile	0.86	0.92	0.90	0.85	0.93	0.90	6	6.95	16.74

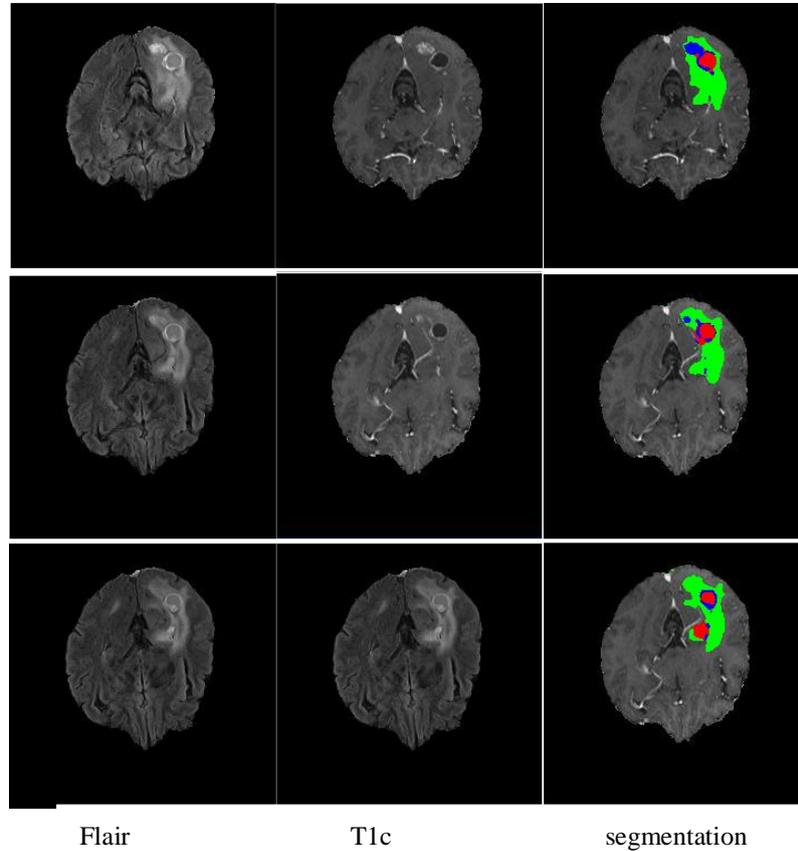


Fig. 4. Segmentation examples from the validation set. From top to bottom: the 54th, 57th, and 59th slices from the subject Brats18_CBICA_ANK_1. *Red* - NCR&NET, *Blue* - ET, *Green* - ED

5 Discussion

5.1 Multi-level Upsampling

To demonstrate the performance improvement resulted from using the GA module, we trained a similar network but without using multi-level upsampling on the BraTS 2018 training dataset and tested it on the validation dataset. Table 3 gives the performance of both models measured by the average Dice similarity coefficient, sensitivity, specificity and Hausdorff-95. It reveals that the dense upsampling connection is able to improve the performance.

Table 3. Comparison of model with Dense Upsampling and without Dense Upsampling on BraTS 2018 validation dataset

	Dice			Sensitivity			Hausdorf-95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC
With MU	0.66	0.87	0.72	0.63	0.84	0.73	7.56	6.73	15.74
Without MU	0.62	0.81	0.65	0.74	0.82	0.67	36.74	44.85	42.26

6 Conclusion

In this paper, we proposed a novel end-to-end segmentation deep model called MU-Net to address the task of segmenting brain tumor and its intra structures from multi-modal MRI scans. We designed a module named GA to combine high-level features and low-level features instead of concatenating them directly. The major part of the decoder branch contains five UF modules, the outputs of four of these five UF modules and corresponding GA are fed into the next level decoder branch. This algorithm has been evaluated on the BraTS 2018 Challenge validation dataset and achieved a Dice similarity coefficient of 0.87, 0.72 and 0.66 for the segmentation of whole tumor, core tumor and enhancing tumor, respectively.

References

1. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J. Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*34(10), 1993 (2015)
2. Zhao, L., Jia, K.: Multiscale cnns for brain tumor segmentation and diagnosis. *Computational and mathematical methods in medicine* 2016 (2016)
3. Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A.V., Criminisi, A., Rueckert, D., Glocker, B.: Deepmedic for brain tumor segmentation. In: *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. pp. 138-149. Springer (2016)
4. Hu, Y., Xia, Y.: 3d deep neural network-based brain tumor segmentation using multimodality magnetic resonance sequences. In: *International MICCAI Brainlesion Workshop*. pp. 423-434. Springer (2017)
5. Pereira, S., Oliveira, A., Alves, V., Silva, C.A.: On hierarchical brain tumor segmentation in mri using fully convolutional neural networks: a preliminary study. In: *Bioengineering (ENBENG), 2017 IEEE 5th Portuguese Meeting on*. pp. 1-4. IEEE (2017)
6. Li, Y., Shen, L.: Deep learning based multimodal brain tumor diagnosis. In: *International MICCAI Brainlesion Workshop*. pp. 149-158. Springer (2017)
7. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., et al.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: *International MICCAI Brainlesion Workshop*. pp. 450-462. Springer (2017)

8. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data* 4, 170117 (2017)
9. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive* (2017), <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>
10. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the preoperative scans of the tcga-lyg collection. *The Cancer Imaging Archive* (2017), <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>
11. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. *arXiv preprint arXiv:1802.02611* (2018)
12. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: *European conference on computer vision*. pp. 630-645. Springer (2016)

Multimodal Brain Tumor Segmentation using Cascaded V-Net

Rui Hua^{1,2}, Yu Zhang¹, Quan Huo¹, Yaozong Gao¹, Yu Sun², Feng Shi¹

¹United Imaging Intelligence, Shanghai, China

²School of Biomedical Engineering, Southeast University, Nanjing, China
220163910@seu.edu.cn

Abstract. In this report, we proposed a novel cascaded V-Net method to segment tumor substructures in multimodal brain magnetic resonance imaging (MRI). Although V-Net has been successfully used in many segmentation tasks, we demonstrate that its performance could be further enhanced by using a cascaded structure. Briefly, we use a V-Net consisting three levels with encoding and decoding paths and intra- and inter-path skip connections. Focal loss is used to improve performance on hard samples as well as balance the positive and negative samples. We first use it to roughly segment the tumor structure. Then the segmentation probability map is used together with the raw image as input for another round of segmentation. By repeating this 3 times as a cascaded strategy, the result is further improved. In another hand, we propose to segment the whole tumor first, and then divide it into the tumor core and enhancing tumor. Experimental results on BraTS 2018 online validation set achieved average Dice scores of 0.8647, 0.8006, 0.7213 for whole tumor, tumor core and enhancing tumor respectively.

Keywords: Deep Learning, Brain Tumor, Segmentation, V-Net

1 Introduction

Gliomas are the most common brain tumors and comprise about 30 percent of all brain tumors. They start in the glial cells of the brain or the spine [1]. Gliomas can be further categorized into low-grade gliomas (LGG) and high-grade gliomas (HGG) according to their pathologic evaluation. LGG are well-differentiated and tend to exhibit benign tendencies and portend a better prognosis for the patients. HGG are undifferentiated and tend to exhibit malignant and carry a worse prognosis. With the development of the Magnetic Resonance Imaging (MRI), multimodal MRI play an important role in disease diagnosis. Different modalities are sensitive to different tissues. For example, T2-weighted (T2) and T2 Fluid Attenuation Inversion Recovery (FLAIR) are sensitive to peritumoral edema, and post-contrast T1-weighted (T1Gd) is sensitive to necrotic core and enhancing tumor core. Thus, they can provide complementary information about gliomas.

Segmentation of brain tumor is an essential technique in disease diagnosis, surgical planning and prognosis [2]. Automatic segmentation provides quantitative information which is more accurate and has better reproducibility. Moreover, the automatic classification of brain tumor relies on the results of brain tumor segmentation. Automatic segmentation likes a powered engine and empower other intelligent medical application. However, segmentation of brain tumor in multimodal MRI scans is one of the most challenging tasks in medical imaging analysis due to their highly heterogeneous appearance, and variable localization, shape and size.

With the development of deep learning, state-of-the-art performance for brain tumor segmentation based on multimodal MRI have been achieved. For example, an end-to-end training using fully convolutional network (FCN) has a good performance in localization of the tumor, and patch-wise convolutional neural network (CNN) was used to segment the intra-tumor structure [3]. A cascaded anisotropic CNN was designed to segment three sub-regions with three Nets, and the segmentation result from last net was used as receptive field in the next net [4].

Inspired by the cascaded strategy and the good performance of V-Net in segmentation tasks, we proposed a cascaded V-Net to segment brain tumor into background and three sub-regions. The cascaded V-Net not only takes advantage of residual connection and using multi-resolution information but also uses the extra coarse localization to boost the performance.

2 Method

2.1 Dataset and Pre-processing

The data used in experiments come from BraTS 2018 training set and validation set [5, 6, 7, 8]. The training set include totally 210 HGG patients and 75 LGG patients. The validation set include 66 patients. Each patient has five image modalities including T1-weighted (T1), T2, T1Gd, FLAIR, and ground truth label. We used 80 percent of the training data as our training set, other 20 percent of the training data as our local test set. All data used in the experiments have been pre-processed with a standard procedure. First, images were bias corrected by using N4 [9]. Second, a histogram matching was applied to all images with respect to a template image. Finally, we rescaled the image intensity into the range of -1 to 1.

2.2 V-Net Architecture

V-Net was initially proposed to segment prostate by training an end-to-end CNN on MRI [10]. The architecture of our V-Net is shown in Fig. 1. The left side of V-Net reduces the size of the input by down-sampling, and the right side of V-Net recovers the semantic segmentation image which has the same size with input image by applying de-convolutions. The detailed parameters about V-Net is shown in Table 1. By means of introducing residual function and skip connection, V-Net has a better segmentation performance compared with classical CNN.

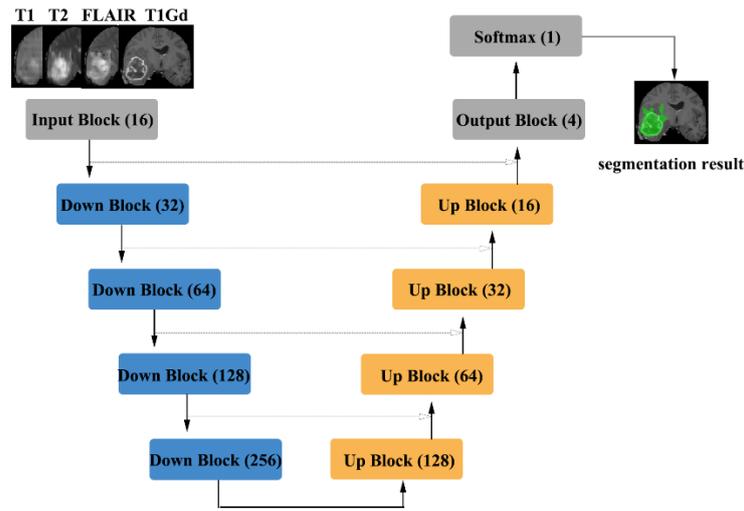


Fig. 1. The architecture of the V-Net.

Table 1. The detailed parameters about V-Net.

Blocks	Sub-blocks or layers	Input Dimensions	Output Dimensions
Input Block	Conv(k=3,p=1,s=1) + BN + ReLU	96*96*96*4	96*96*96*16
Down Block 1	Conv(k=2,p=0,s=2)+ BN + ReLU	96*96*96*16	48*48*48*32
	Residual Block Conv(k=3,p=1,s=1) + BN (input+output) + ReLU	48*48*48*32 48*48*48*32	48*48*48*32 48*48*48*32
Down Block 2	Conv(k=2,p=0,s=2) + BN + ReLU	48*48*48*32	24*24*24*64
	Residual Block Conv Block * 2 (input+output) + ReLU	24*24*24*64 24*24*24*64	24*24*24*64 24*24*24*64
Down Block 3	Conv(k=2,p=0,s=2) + BN + ReLU	24*24*24*64	12*12*12*128
	Residual Block Conv Block * 3 (input+output) + ReLU	12*12*12*128 12*12*12*128	12*12*12*128 12*12*12*128
Down Block4	Conv(k=2,p=0,s=2) + BN + ReLU	12*12*12*128	6*6*6*256
	Residual Block Conv Block * 3 (input+output) + ReLU	6*6*6*256 6*6*6*256	6*6*6*256 6*6*6*256
Up Block 1	Conv(k=2,p=0,s=2) + BN + ReLU	6*6*6*256	12*12*12*128
	Cat(output, skip) Residual Block Conv Block * 3 (input+output) + ReLU	12*12*12*128 12*12*12*256 12*12*12*256	12*12*12*256 12*12*12*256 12*12*12*256
Up Block 2	Conv(k=2,p=0,s=2) + BN + ReLU	12*12*12*256	24*24*24*64
	Cat(output+skip) Residual Block Conv Block * 3 (input+output) + ReLU	24*24*24*64 24*24*24*128 24*24*24*128	24*24*24*128 24*24*24*128 24*24*24*128

Up Block 3		Conv(k=2,p=0,s=2) + BN + ReLU	24*24*24*128	48*48*48*32	
	Residual Block		Cat(output+skip)	48*48*48*32	48*48*48*64
			Conv(k=3,p=1,s=1) + BN + ReLU	48*48*48*64	48*48*48*64
			Conv(k=3,p=1,s=1)+BN	48*48*48*64	48*48*48*64
			(input+output) + ReLU	48*48*48*64	48*48*48*64
Up Block 4		Conv(k=2,p=0,s=2) + BN + ReLU	48*48*48*64	96*96*96*16	
	Residual Block		Cat(output+skip)	96*96*96*16	96*96*96*32
			Conv(k=3,p=1,s=1) + BN	96*96*96*32	96*96*96*32
			(input+output) + ReLU	96*96*96*32	96*96*96*32
Out Block		Conv(k=1,p=0,s=1) + BN + ReLU	96*96*96*32	96*96*96*4	
		Softmax	96*96*96*4	96*96*96*1	

Note: Each Conv Block contains three convolution layers: Conv1(k=1, p=0, s=1), Conv2(k=3, p=1, s=1), and Conv3(k=1, p=0, s=1). k, kernel size; p, padding; s, stride.

2.3 Proposed Cascaded V-Net Framework

Although V-Net performs well in segmentation tasks, its performance may be further improved if we can provide extra information, such as coarse localization. Therefore, we propose a cascaded V-Net for tumor segmentation. Briefly, we 1) use a V-Net for the brain whole tumor segmentation; 2) another V-Net to further divide the tumor region into three substructures, e.g., tumor necrosis, edema, and enhancing tumor. Note that the coarse segmentation of whole tumor in the last V-Net is also used a part of input to boost the performance; and 3) the third V-Net to refine the segmentation result by using the image input together with the segmentation probability maps from the last step. Detailed steps are as follows.

The proposed framework is shown in Fig. 2. There are three networks to segment substructure of brain tumors sequentially. The first network (V-Net 1) is designed to segment the whole tumor. V-Net 1 uses four modality MR images as inputs, and outputs the mask of whole tumor (WT). The second network (V-Net 2) is designed to segment the brain tumor into three sub-regions: whole tumor (WT), tumor core (TC) and enhancing tumor (ET). V-Net 2 uses four modality MR images and the mask of WT as inputs, and outputs the segmented mask with three labels. The third network (V-Net 3) is designed to obtain more accurate results about the three sub-regions of brain tumor. V-Net 3 uses four modality MR images and the labeled mask as inputs and outputs the final accurate segmentation results. The cascaded V-Net take advantage of combining extra coarse localization information or segmentation information, and multimodal MR imaging to obtain more accurate segmentation results.

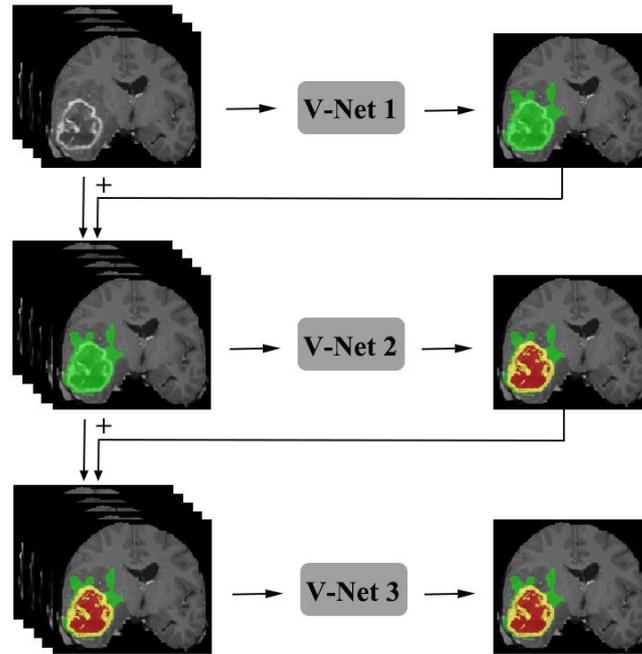


Fig. 2. The proposed cascaded framework for brain segmentation.

Network implementation.

Our cascaded V-Net was implemented in the deep learning framework PyTorch. In our network, we initialized weight with kaiming initialization [11], and used focal loss [12] as loss function. Adaptive Moment Estimation (Adam) [13] was used as optimizer with learning rate 0.001, batch size 8.

2.4 Post-processing

The predicted segmentation results were post-processed using connected component analysis. We consider that the isolated segmentation labels with small size are prone to artifacts and thus remove them. After the V-Net 1, the components with total voxel number below a threshold ($T=15000$) were discarded in the binary whole tumor map. After the V-Net 2, a mask of different labels was used in the connected component analysis. Moreover, if all the connected components were less than 15000 voxels, we will retain the largest connected component.

3 Experimental Results

3.1 Segmentation results on local test set

We used 20 percent of the training data as our local test set. The local test set include 42 HGG patients and 15 LGG patients. Some of the segmentation results are shown in Fig. 3. The segmentation of whole tumor achieved best results with average dice score 0.8505. And the segmentation of enhancing tumor seems not good enough. Moreover, in order to evaluate the preliminary experimental results, we calculated the average dice scores, sensitivity and specificity for whole tumor, tumor core and enhancing tumor respectively. The results are shown in Table 2.

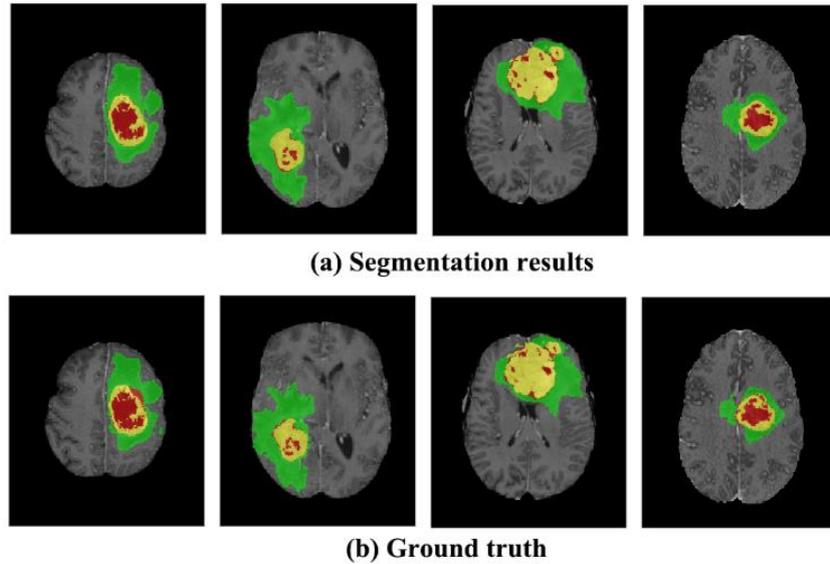


Fig. 3. The comparison of segmentation results and ground truth on four cases from local test set.

Table 2. Segmentation accuracy evaluation on local test set.

	Whole Tumor	Tumor Core	Enhancing Tumor
Dice mean \pm SD	0.8505 \pm 0.0972	0.7842 \pm 0.1919	0.7426 \pm 0.2080
Sensitivity mean \pm SD	0.9180 \pm 0.1091	0.7596 \pm 0.2199	0.7174 \pm 0.2337
Specificity mean \pm SD	0.9981 \pm 0.0012	0.9996 \pm 0.0008	0.9997 \pm 0.0003

3.2 Segmentation results on MICCAI BraTS 2018 validation set

The segmentation results on BraTS 2018 online validation set achieved average dice scores of 0.8647, 0.8006, 0.7213 for whole tumor, tumor core and enhancing tumor respectively. Interestingly, the average dice scores of whole tumor and tumor core are bigger than that in local test set. And they have the same tendency with that in local test set. The segmentation of whole tumor performed best while the segmentaion of enhancing tumor core performed worst. The details are shown in Table 3.

Table 3. Segmentation accuracy evaluation on BraTS 2018 validation set.

	Whole Tumor	Tumor Core	Enhancing Tumor
Dice mean \pm SD	0.8647 \pm 0.1238	0.8006 \pm 0.2231	0.7213 \pm 0.2633
Sensitivity mean \pm SD	0.9419 \pm 0.1301	0.7987 \pm 0.2363	0.7522 \pm 0.2316
Specificity mean \pm SD	0.9870 \pm 0.0091	0.9974 \pm 0.0029	0.9983 \pm 0.0026
Hausdorff95 mean \pm SD	8.3350 \pm 12.3272	7.4854 \pm 9.1365	4.4946 \pm 7.5252

4 Discussion

We found three interesting points in the preliminary experiment. Firstly, we used two kinds of pre-processing method to process the training data and the validation data and compared their segmentation results. Except the method proposed in the method part, we also normalize the images by subtracting the mean and dividing by the standard deviation of the brain region. As a result, there is almost no different in the two average dice scores for whole tumor, tumor core and enhancing tumor respectively.

Secondly, according to the evaluation results, our model performs better in online validation set than that in our local test set. The reason might be that we didn't do ten-fold cross validation. So the results are not generalizable enough.

Thirdly, the post-processing method is so important that it will make the average dices increase or decrease obviously. If the threshold is big, some of small cluster will be discarded improperly. If the threshold is small, some false positive results will be kept incorrectly. In order to have a better performance, we test lots of thresholds and choose the best one. However, this threshold may not be suitable for other models. All in all, the threshold used in component connected analysis is important to have a better performance.

5 Conclusions

In conclusion, we proposed a cascaded V-Net to segment brain tumor into background and three sub-regions of brain tumor. The experimental results on BraTS 2018 online validation set achieved average Dice scores of 0.8647, 0.8006, 0.7213 for whole tumor, tumor core and enhancing tumor respectively.

6 References

1. Mamelak, A.N., Jacoby, D.B. Targeted delivery of antitumoral therapy to glioma and other malignancies with synthetic chlorotoxin (TM-601). *Expert Opin Drug Deliv* 4, 175-186 (2007).
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci Data* 4, 170117, (2017).
3. Cui, S., Mao, L., Jiang, J., Liu, C., Xiong, S. Automatic Semantic Segmentation of Brain Gliomas from MRI Images Using a Deep Cascaded Neural Network. *J Healthc Eng* 2018, 4940593, (2018).
4. Bakas S., Kuijf H., Menze B., Reyes M. (eds) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. BrainLes. Lecture Notes in Computer Science, vol 10670. Springer, Cham, (2017)
5. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp F, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftikharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput

- K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
6. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
 7. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
 8. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
 9. N4BiasFieldCorrectionImageFilter Class Reference,
https://itk.org/SimpleITKDoxygen/html/classitk_1_1simple_1_1N4BiasFieldCorrectionImageFilter.html
 10. F. Milletari, N. Navab and S. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 2016, pp. 565-571. (2016)
 11. K. He, X. Zhang, S. Ren and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, pp. 1026-1034.(2015)
 12. T. Lin, P. Goyal, R. Girshick, K. He and P. Dollar, "Focal Loss for Dense Object Detection," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp. 2999-3007.(2018)
 13. Kingma, D., Ba, J. Adam: A Method for Stochastic Optimization. Computer Science.(2014)

Pre and Post Processing Techniques for Brain Tumor Segmentation

Vivek HV¹

Harvard University, Cambridge MA 02138, USA
vivekhv@mde.harvard.com

Abstract. Over the past few years, deep learning convolutional networks such as U-Net [1], ResNet [2] and Fully-Convolutional Networks (FCNs) [3] have been consistently used to obtain state-of-art performances for the task of semantic segmentation on MRI and CT images. Pre-processing and Post-processing methods, however, are often overlooked in favor of deeper or different deep learning network architectures. Pre-processing and post-processing methods can provide additional information and increased information density which can help improve the predictions of a deep learning network resulting in enhanced performance overall. This paper explores different methods of pre-processing and post-processing the data to augment a U-Net architecture. To train the algorithms we use the data sets of the Brain Tumor Segmentation challenge (BraTS) 2018 competition [4, 5, 9, 10].

Keywords: Pre-Processing · Post-Processing · Brain Tumor · Semantic Segmentation · U-Net · Deep Learning

1 Introduction

Gliomas are a type of tumor that can occur in the human brain and spinal cord. While they are the most common of brain malignancies - making their identification an important task - their heterogeneity in appearance and shape has made their identification one of the most challenging tasks in medical imaging.

Over the past several years, the literature on computational algorithms developed for these tasks has focused on using deep learning architectures. Networks such as U-Net [1] and DeepMedic [6] and ensembles thereof have shown consistent and robust performance on tasks like semantic segmentation. Due to the nature of Deep Learning, pre-processing and post-processing of data are often given lower priority over the development of newer or deeper model architectures.

This work presents different processing techniques used in conjunction with a U-Net model architecture trained for the task of semantic segmentation. The work uses data obtained from multi-institutional pre-operative MRI scans made available as part of BraTS 2018. Further, the work leverages BraTS evaluation tools to analyze the performance of the proposed methods.

2 Methods

2.1 Pre-Processing

Asymmetry Masks: As noted in the work by [7], symmetry in axial view is an important indicator for brain tumor segmentation as tumors usually break symmetric appearance of a healthy brain. Building on the work proposed in [8] and [7], asymmetry-based masks are generated from the FLAIR modalities. A Gaussian filter is used to remove noise from the masks.

To increase the probability of detecting tumors using this technique, the image is first transposed and added to the original image. To determine the correct threshold value, the asymmetry is then compared along the axial view. The masks generated were able to detect large tumors that cause asymmetry in axial or median views. The masks, however, perform poorly in cases where the tumor is very close to the center of the image modality since these do not contribute to asymmetry in the image. These asymmetry-masks are then encoded into the U-Net as an extra modality of the input data.

Threshold Masks: Using asymmetry-masks as the bounds of whole tumors, further masks are generated using thresholding on T1CE and T2 images. A Gaussian filter is used to remove noise from the masks. These masks can detect large tumor-cores and enhancing tumors in T1CE and T2 modalities respectively. Similar to the asymmetry-based masks, these are then fed into the U-Net as extra modalities.

Standardization and Normalization The images in all modalities are standardized with a mean of 0 and a standard deviation of 1. The image is then normalized to $[0, 1]$.

2.2 Pipeline

Patching The BraTS dataset offers four modalities: FLAIR, T1, T2 and T1CE. Each image in all modalities is of size 240x240 pixels. For the purpose of training in a limited memory environment, every slice is padded with zeros to 288x288 before being divided into 9 patches each of size 96x96 pixels.

Data Trimming From the dataset thus generated, all patches with low information density (empty FLAIR modality patches were used to determine information density) were removed. Further, data points without a tumor were randomly discarded with a probability of 0.5 to balance the dataset.

Data Augmentation The data is then augmented by transposing, flipping and rotating the images of each modality (8x).

Multi-Scale: The DeepMedic [6] algorithm used multi-scale patches to feed its model architecture. A similar approach is used in this work where patches of size 144x144 are re-scaled to 96x96 and added to the input pipeline.

2.3 Training

Data Split The BraTS 2018 dataset consisted of 285 cases divided into 210 cases of High-Grade Glioma and 75 cases of Low-Grade Glioma. The dataset was divided into 29 test cases and the remaining 256 cases were further divided into 90% training and 10% validation sets.

U-Net: A 2D U-Net is used for training in the semantic segmentation task. The U-Net architecture is 10 layers deep. The input is of shape (n x 96 x 96 x 11). The 11 channels are 96x96 patches of : FLAIR, FLAIR-Scaled, T1, T1 Scaled, T2, T2 Scaled, T1CE, T1CE Scaled, FLAIR Asymmetry Mask, T2-Threshold Mask and T1CE-Threshold Mask.

Training Parameters Three different networks are trained - one each for Whole Tumor (all labels), Tumor Core (labels 1 and label 4) and Enhancing Tumor (label 4 only). All networks use the same input dataset. The models are optimized using the Adam optimizer (starting learning rate = 0.0001) with a decreasing learning rate (factor 0.25 with patience of 5 epochs) for 50 epochs.

2.4 Post-Processing

The results from the three networks are combined to obtain final predicted masks of 240x240 pixels. For every patient, the following post-processing techniques are used to further obtain better results on the validation set:

- The slice with the largest tumor area is determined for each patient. Using this as a baseline, all other predicted tumors which have an area lesser than 50 pixels or which reside outside the contour determined by the largest tumor are removed.
- Using “Whole Tumor” predictions, the predicted “Tumor Core” regions bounded within the contour of the “Whole Tumor”. This removes artifacts from the predicted “Tumor Core” regions. The same technique is employed to bound “Enhancing Tumor” predictions within the “Tumor Core” regions.

3 Results

The BraTS 2018 evaluation too is used to evaluate the performance of the methods described in this work.

3.1 Local Test Set

The performance of the proposed pipeline on the local test set (n=29) is given in Table 1. Figure 1 provides an example of performance of the network on a slice in the local test set where dice score achieved was greater 0.9. Figure 2 an example of performance of the network on a slice in the local test set where dice score achieved was less than 0.7.

Table 1: Table captions should be placed above the tables.

	Whole Tumor	Tumor Core	Enhancing Tumor
Mean	0.8744	0.7182	0.6662
Standard Deviation	0.0721	0.1504	0.1661
Median	0.8933	0.7533	0.7185

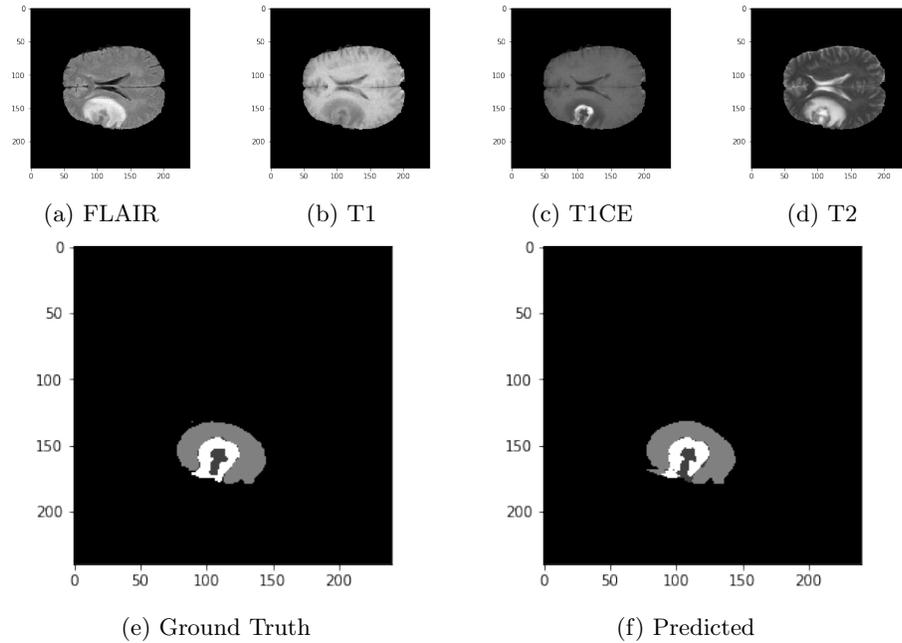


Fig. 1: Performance of the network on a slice in the Local test set (Dice Score > 0.9)

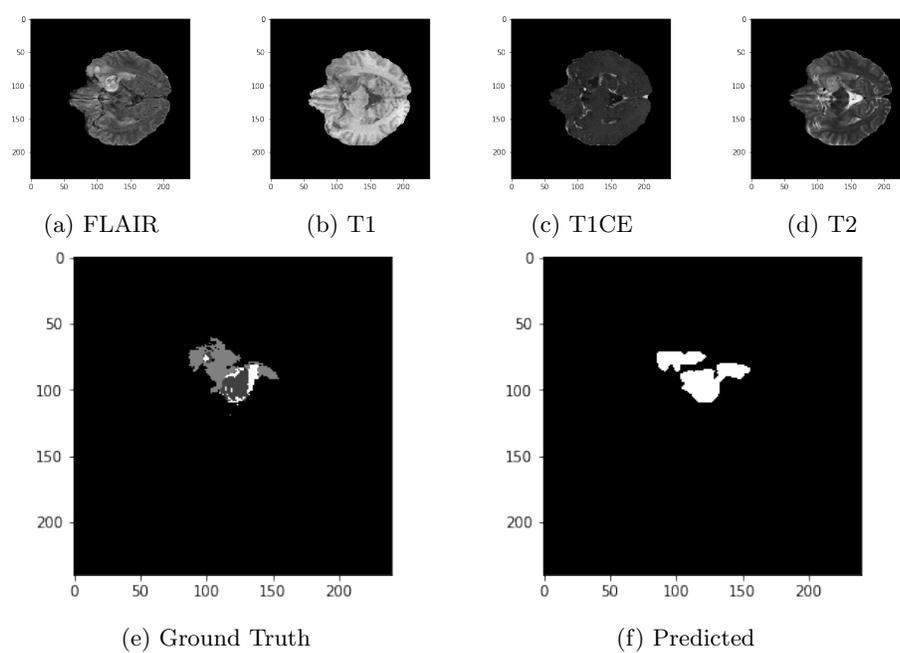


Fig. 2: Performance of the network on a slice in the Local test set (Dice Score<0.7)

3.2 BraTS Validation Set

The performance of the proposed pipeline on the BraTS 2018 validation set (n=66) is given in Table 2. Figure 3 provides an example of performance of the network on a slice in the BraTS validation set where dice score achieved was greater 0.9. Figure 4 an example of performance of the network on a slice in the BraTS validation set where dice score achieved was less than 0.7.

Table 2: Table captions should be placed above the tables.

	Whole Tumor	Tumor Core	Enhancing Tumor
Mean	0.8819	0.6485	0.5679
Standard Deviation	0.0720	0.2233	0.2747
Median	0.9001	0.7284	0.6297

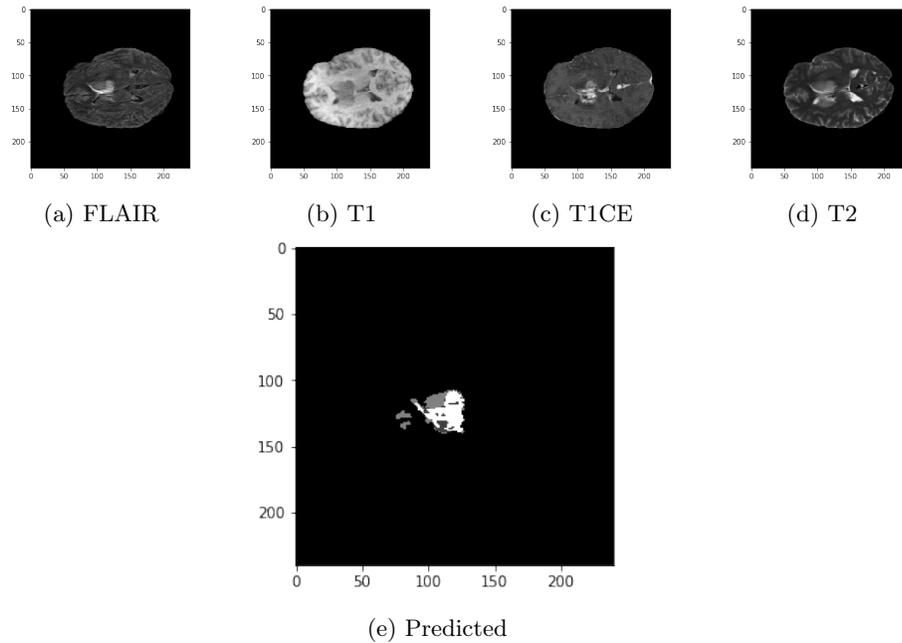


Fig. 3: Performance of the network on a slice in the BraTS validation set (Dice Score > 0.9)

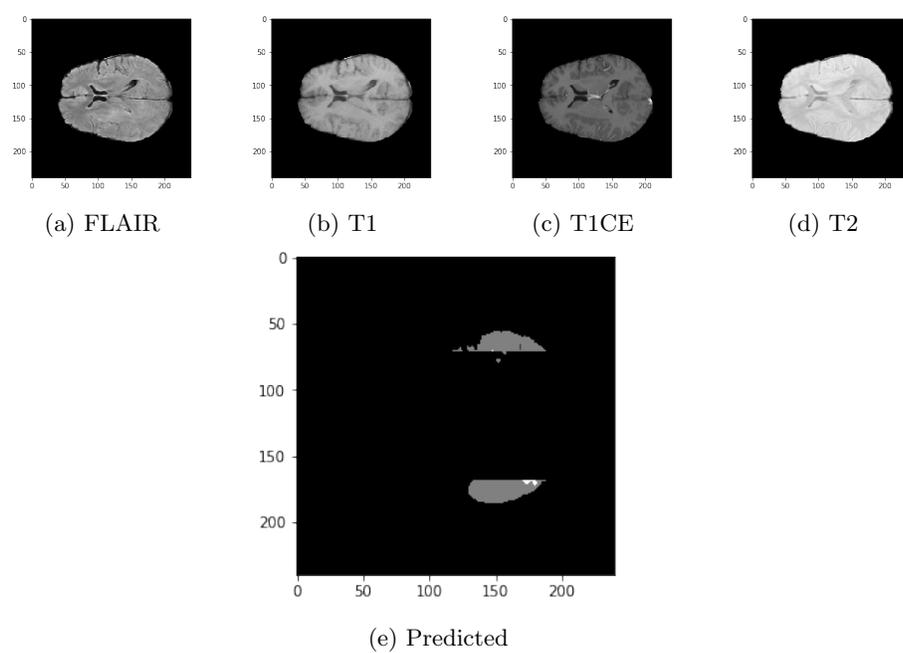


Fig. 4: Performance of the network on a slice in the BraTS validation set (Dice Score < 0.7)

4 Discussion

- It should also be noted that the poor performance on “Tumor Core” and “Enhancing Tumor” doesn’t align with the performance of “Whole Tumor” model. Future work will focus on improving the performance of the “Tumor Core” and “Enhancing Tumor” models.
- While the proposed methods in this paper show good results on “Whole Tumor”, further research is necessary to quantify the impact of each of the methods proposed for pre-processing and post processing.
- The training was first conducted using patches of size 64x64 pixels. The performance was seen to be lower than the models using 96x96 patches. In future work, the impact of patch size and the techniques proposed in this paper will be quantified.
- This work uses a 2D U-Net architecture. It has been seen that 3D U-Nets consistently perform better than 2D counterparts. Future experiments will expand on the work using 3D Deep Learning architectures.

5 Conclusion

In this paper, the author proposes different pre-processing and post-processing techniques that accompany Deep Learning models trained for semantic segmentation.

- Asymmetry-based techniques were used to generate region-of-interest masks
- Thresholding-based techniques were used to generate region-of-interest masks
- Multi-Scale images were generated to increase information for deep learning algorithms
- A U-Net was used for the deep learning part of the pipeline which was fed using the aforementioned techniques
- Simple size and location-based post-processing techniques were used to improve predictions generated by the U-Net

The pre-processing and post-processing techniques proposed in this paper can help augment the deep learning pipeline that has become the de-facto tool for semantic segmentation in medical imaging. However, further work is necessary to determine the impact of pre-processing and post-processing. The author plans to conduct future work in the following areas: 1. Improve performance of the “Tumor Core” and “Enhancing Tumor” regions. 2. Quantify the contribution of each of the pre-processing and post-processing steps and compare them to a baseline U-Net model 3. Use 3D deep learning architecture and compare the performance of the methods proposed in this paper

References

1. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention 2015 Oct 5 (pp. 234-241). Springer, Cham.
2. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conf Comput Vis Pattern Recognit [Internet]. IEEE; 2016. pp 7708.
3. Varghese A, Mohammed S, Ganapathy K. Brain Tumor Segmentation from Multi Modal MR images using Fully Convolutional Neural Network. Proceedings of International MICCAI BraTS Challenge 2017 (pp 1-8). Springer
4. Menze, B.H., Jakab, A., Bauer, S., et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BraTS). *Medical Imaging*, 34(10), 19932024 (2015)
5. Bakas, S., Akbari, H., Sotiras, A., et al. Advancing The Cancer Genome Atlas Glioma MRI Collections with Expert Segmentation Labels and Radiomic Features. *Nature Scientific Data*, (2017) [In Press]
6. Kamnitsas, K., Ledig, C., Newcombe, V.F., et al. Efficient Multi-scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation. *Medical Image Analysis*, 36, 61-78 (2017)
7. Prastawa M, Bullitt E, Sean H, et al. A brain tumor segmentation framework based on outlier detection *Medical Image Analysis*, 2004, Vol.8(3), pp.275-283
8. Haocheng S, Jianguo Z, Weishi Z. Efficient symmetry-driven fully convolutional network for multimodal brain tumor segmentation. 2017 IEEE International Conference on Image Processing, Sept. 2017, pp.3864-3868
9. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
10. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF

No New-Net

Fabian Isensee¹, Philipp Kickingereder², Wolfgang Wick³, Martin Bendszus²,
and Klaus H. Maier-Hein¹

¹ Division of Medical Image Computing, German Cancer Research Center (DKFZ),
Heidelberg, Germany

² Department of Neuroradiology, University of Heidelberg Medical Center,
Heidelberg, Germany

³ Neurology Clinic, University of Heidelberg Medical Center,
Heidelberg, Germany

Abstract. In this paper we demonstrate the effectiveness of a well trained U-Net in the context of the BraTS 2018 challenge. This endeavour is particularly interesting given that researchers are currently besting each other with architectural modifications that are intended to improve the segmentation performance. We instead focus on the training process, argue that a well trained U-Net is hard to beat and intend to back up this assumption with a strong participation in this years BraTS challenge. Our baseline U-Net, which has only minor modifications and is trained with a large patch size and a dice loss function already achieves competitive dice scores on the BraTS2018 validation data. By incorporating region based training, additional training data and a simple postprocessing technique, we obtain dice scores of 81.01, 90.83 and 85.44 and Hausdorff Distances (95th percentile) of 2.54, 4.97 and 7.04 for the enhancing tumor, whole tumor and tumor core, respectively.

Keywords: CNN, Brain Tumor, Glioblastoma, U-Net, dice loss

1 Introduction

Quantitative assessment of brain tumors provides valuable information and therefore constitutes an essential part of diagnostic procedures. Automatic segmentation is attractive in this context, as it allows for faster, more objective and potentially more accurate description of relevant tumor parameters, such as the volume of its subregions. Due to the irregular nature of tumors, however, the development of algorithms capable of automatic segmentation remains challenging.

The brain tumor segmentation challenge (BraTS) [1] aims at encouraging the development of state of the art methods for tumor segmentation by providing a large dataset of annotated low grade gliomas (LGG) and high grade glioblastomas (HGG). The BraTS 2018 training dataset, which consists of 210 HGG and 75 LGG cases, was annotated manually by one to four raters and all segmentations were approved by expert raters [2,3,4]. For each patient a T1 weighted, a post-contrast T1-weighted, a T2-weighted and a FLAIR MRI was provided.

The MRI originate from 19 institutions and were acquired with different protocols, magnetic field strengths and MRI scanners. Each tumor was segmented into edema, necrosis and non-enhancing tumor and active/enhancing tumor. The segmentation performance of participating algorithms is measured based on the DICE coefficient, sensitivity, specificity and 95th percentile of Hausdorff distance.

It is unchallenged by now that convolutional neural networks (CNNs) dictate the state of the art in biomedical image segmentation [5,6,7,8,9,10]. As a consequence, all winning contributions to recent BraTS challenges were exclusively build around CNNs. One of the first notably successful neural network for brain tumor segmentation was DeepMedic, a 3D CNN introduced by Kamnitsas et al. [5]. It comprises a low and a high resolution pathway that capture semantic information at different scales and recombines them to predict a segmentation based on precise local as well as global image information. Kamnitsas et al. later enhanced their architectures with residual connections for BraTS 2016 [11]. With the success of encoder-decoder architectures for semantic segmentation, such as FCN [12,13] and most notably the U-Net [14], it is unsurprising that these architectures are used in the context of brain tumor segmentation as well. In BraTS 2017, all winning contributions were at least partially based on encoder-decoder networks. Kamnitsas et al. [9], who were the clear winner of the challenge, created an ensemble by combining three different network architectures, namely 3D FCN [12], 3D U-Net [15,14] and DeepMedic [5], trained with different loss functions (dice loss [16,17] and crossentropy) and different normalization schemes. Wang et al. [10] used a FCN inspired architecture, enhanced with dilated convolutions [13] and residual connections [18]. Instead of directly learning to predict the regions of interest, they trained a cascade of networks that would first segment the whole tumor, then given the whole tumor the tumor core and finally given the tumor core the enhancing tumor. Isensee et al. [6] employed a U-Net inspired architecture that was trained on large input patches to allow the network to capture as much contextual information as possible. This architecture made use of residual connections [18] in the encoder only, while keeping the decoder part of the network as simple as possible. The network was trained with a multiclass dice loss and deep supervision to improve the gradient flow.

Recently, a growing number of architectural modifications to encoder-decoder networks have been proposed that are designed to improve the performance of the networks for their specific tasks [17,19,20,21,10,6,22,7]. Due to the sheer number of such variants, it becomes increasingly difficult for researchers to keep track of which modifications extend their usefulness over the few datasets they are typically demonstrated on. We have implemented a number of these variants and found that they provide no additional benefit if integrated into a well trained U-Net. In this context, our contribution to the BraTS 2018 challenge is intended to demonstrate that such a U-Net, without using significant architectural alterations, is capable of generating competitive state of the art segmentations.

2 Methods

In the following we present the network architecture and training schemes used for our submission. As hinted in the previous paragraph, we will use a 3D U-Net architecture that is very close to its original publication [15] and optimize the training procedure to maximize its performance on the BraTS 2018 training and validation data.

2.1 Preprocessing

With MRI intensity values being non standardized, normalization is critical to allow for data from different institutes, scanners and acquired with varying protocols to be processed by one single algorithm. This is particularly true for neural networks where imaging modalities are typically treated as color channels. Here we need to ensure that the value ranges match not only between patients but between the modalities as well in order to avoid initial biases of the network. We found the following workflow to work well. We normalize each modality of each patient independently by subtracting the mean and dividing by the standard deviation of the brain region. The region outside the brain is set to 0.

2.2 Network architecture

U-Net [14] is a successful encoder-decoder network that has received a lot of attention in the recent years. Its encoder part works similarly to a traditional classification CNN in that it successively aggregates semantic information at the expense of reduced spatial information. Since in segmentation, both semantic as well as spatial information are crucial for the success of a network, the missing spatial information must somehow be recovered. U-Net does this through the decoder, which receives semantic information from the bottom of the 'U' (refer to Fig. 1) and recombines it with higher resolution feature maps obtained directly from the encoder through skip connections. Unlike other segmentation networks, such as FCN [12] and previous iterations of DeepLab [13] this allows U-Net to segment fine structures particularly well.

Our network architecture is an instantiation of the 3D U-Net [15] with minor modifications. Following our successful participation in 2017 [6], we stick with our design choice to process patches of size 128x128x128 with a batch size of two. Due to the high memory consumption of 3D convolutions with large patch sizes, we implemented our network carefully to still allow for an adequate number of feature maps. By reducing the number of filters right before upsampling and by using inplace operations whenever possible, this results in a network with 30 feature channels at the highest resolution, which is nearly double the number we could train with in our previous model (on a 12 GB NVIDIA Titan X GPU). Due to our choice of loss function, traditional ReLU activation functions did not reliably produce the desired results, which is why we replaced them with leaky ReLUs (leakiness 10^{-2}) throughout the entire network. With a small batch size of 2, the exponential moving averages of mean and variance within a batch

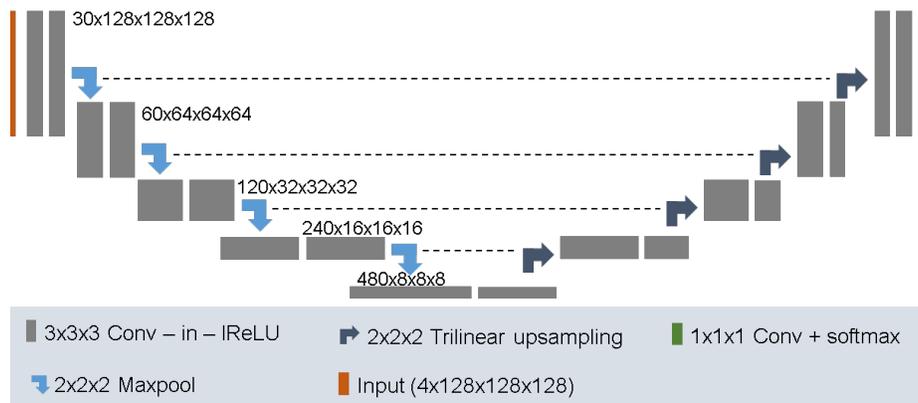


Fig. 1. We use a 3D U-Net architecture with minor modifications. It instance normalization [23] and leaky ReLU nonlinearities and reduces the number of feature maps before upsampling. Feature map dimensionality is noted next to the convolutional blocks, with the first number being the number of feature channels.

learned by batch normalization [24] are unstable and do not reflect the feature map activations at test time very well. We found instance normalization [23] to provide more consistent results and therefore used it to normalize all feature map activations (between convolution and nonlinearity). For an overview over our segmentation architecture, please refer to Fig. 1.

2.3 Training Procedure

Our network architecture is trained with randomly sampled patches of size 128x128x128 voxels and batch size 2. We refer to an epoch as an iteration over 250 batches and train for a maximum of 500 epochs. The training is terminated early if the exponential moving average of the validation loss ($\alpha = 0.95$) has not improved within the last 60 epochs. Training is done using the ADAM optimizer with an initial learning rate $lr_{init} = 1 \cdot 10^{-4}$, which is reduced by factor 5 whenever the above mentioned moving average of the validation loss has not improved in the last 30 epochs. We regularize with a l2 weight decay of 10^{-5} .

One of the main challenges with brain tumor segmentation is the class imbalance in the dataset. While networks will train with crossentropy loss function, the resulting segmentations may not be ideal in the sense of the dice score they obtain. Since the dice scores is one of the most important metrics based upon which contributions are ranked, it is imperative to optimize this metric. We achieve that by using a soft dice loss for the training of our network. While several formulations of the dice loss exist in the literature [25,16,17], we prefer to use a multi-class adaptation of [16] which has given us good results in segmentation

challenges in the past [8,6]. This multiclass Dice loss function is differentiable and can be easily integrated into deep learning frameworks:

$$\mathcal{L}_{\text{dc}} = -\frac{2}{|K|} \sum_{k \in K} \frac{\sum_i u_i^k v_i^k}{\sum_i u_i^k + \sum_i v_i^k} \quad (1)$$

where u is the softmax output of the network and v is a one hot encoding of the ground truth segmentation map. Both u and v have shape i by c with i being the number of pixels in the training patch and $k \in K$ being the classes.

When training large neural networks from limited training data, special care has to be taken to prevent overfitting. We address this problem by utilizing a large variety of data augmentation techniques. The following augmentation techniques were applied on the fly during training: random rotations, random scaling, random elastic deformations, gamma correction augmentation and mirroring. Data augmentation was done with our own in-house framework which is publically available at <https://github.com/MIC-DKFZ/batchgenerators>.

The fully convolutional nature of our network allows to process arbitrarily sized inputs. At test time we therefore segment an entire patient at once, alleviating problems that may arise when computing the segmentation in tiles with a network that has padded convolutions. We furthermore use test time data augmentation by mirroring the images and averaging the softmax outputs.

2.4 Region based prediction

Wang et al. [10] use a cascade of CNNs to segment first the whole tumor, then the tumor core and finally the enhancing tumor. We believe the cascade and their rather complicated network architecture to be of lesser importance, but the fact that they did not learn the labels (enhancing tumor, edema, necrosis) but instead optimized the regions that are finally evaluated in the challenge directly to be key to their good performance in last years challenge. For this reason we will also train a version of our model where we replace the final softmax with a sigmoid and optimize the three (overlapping) regions (whole tumor, tumor core and enhancing tumor) directly with the dice loss.

2.5 Cotraining

285 training cases is a lot for medical image segmentation, but may still not be enough to prevent overfitting entirely. We therefore also experiment with cotraining on institutional data. Since the label definitions between the BraTS dataset and our own differ slightly, we add a second segmentation layer (1x1x1 convolution) at the end, which acts as a supervised version of m heads [26]. During training, the BraTS segmentation layer only receives gradients from BraTS examples and the other segmentation layer is trained only on institutional data. The losses of both layers are averaged to obtain the total loss of a minibatch. The rest of the network weights are shared.

2.6 Postprocessing

One of the most challenging parts in the BraTS challenge data is distinguishing small blood vessels in the tumor core region (that must be labeled either as edema or as necrosis) from enhancing tumor. Since this is particularly detrimental for LGG patients that may have no enhancing tumor at all we replace all enhancing tumor voxels with necrosis if less than 500 enhancing tumor voxels are present in a patient.

3 Experiments and Results

We designed our training scheme by running a five fold cross-validation on the 285 training cases of BraTS 2018. Training set results are summarized in Table 3, validation set results can be found in table 2. Validation set results were obtained by using the five networks from the training cross-validation as an ensemble. For consistency with other publications, all reported values were computed by the online evaluation platform (<https://ipp.cbica.upenn.edu/>).

Due to the relatively small size of the validation set (66 cases vs 285 training cases) we base our main analysis on the cross-validation results. We are confident that conclusions drawn from the training set are more robust and will generalize well to the test set. The vast majority of observations we draw from the training set translate into the validation set as well.

	Dice			HD95		
	enh.	whole	core	enh.	whole	core
baseline	73.43	89.76	82.17	4.88	5.86	7.11
baseline + postprocess	77.11	89.76	82.17	3.99	5.86	7.11
baseline + regions	73.81	90.02	82.87	5.01	6.26	6.48
baseline + regions + postprocess	77.82	90.02	82.87	3.93	6.26	6.48
baseline + regions + cotraining	74.62	90.08	84.30	4.68	5.61	6.00
baseline + regions + cotraining + postprocess	77.17	90.08	84.30	3.68	5.61	6.00

Table 1. Results on BraTS 2018 training data (285 cases). All results were obtained by running a five fold cross-validation. Metrics were computed by the online evaluation platform.

We refer to our U-Net trained on the BraTS labels and training data as *baseline*. With Dice scores of 73.43/89.76/82.17 (enh/whole/core) on the training set and 79.59/90.80/84.32 (enh/whole/core) on the validation set this baseline model is by itself already strong. Adding region based training (*regions*) improved the dice scores of both the enhancing tumor as well as the tumor core. When training with our institutional data (*cotraining*), we gain an additional dice point on enhancing tumor and tumor core. Our postprocessing, which is targeted at fixing false positive enhancing tumor predictions in LGG patients has a substantial impact on enhancing tumor dice. On the training

	Dice			HD95		
	enh.	whole	core	enh.	whole	core
baseline	79.59	90.80	84.32	3.12	4.79	8.02
baseline + postprocess	80.36	90.80	84.32	2.55	4.79	8.02
baseline + regions	79.25	90.72	85.14	3.80	5.23	7.23
baseline + regions + postprocess	80.48	90.72	85.14	2.81	5.23	7.23
baseline + regions + cotraining	80.45	90.83	85.44	3.12	4.97	7.04
baseline + regions + cotraining + postprocess	81.01	90.83	85.44	2.54	4.97	7.04

Table 2. Results on BraTS2018 validation data (66 cases). Results were obtained by using the five models from the training set cross-validation as an ensemble. Metrics were computed by the online evaluation platform.

set it increases the mean dice by up to 4.01 points (baseline+regions vs baseline+regions+postprocess). However, its impact on the best performing model is less pronounced, resulting in only an increase of 2.55 dice points (baseline+regions+cotraining+postprocess on train set). This is most likely due to an increase in segmentation quality which partly removes the necessity for postprocessing. Especially the inclusion of institutional data seems to have resolved many false positives in LGG patients by itself. Comparing baseline+regions+postprocess vs baseline+regions+cotraining+postprocess we notice that training with additional training data only yields marginal improvements. In the training data, including institutional data reduced the enh. tumor dice scores by 0.65 but increased the tumor core dice by 1.43. On the validation set, the impact of our postprocessing is lower, yielding only a minor improvement on in dice score (0.56) for our best model. We believe this is caused by a discrepancy in data distribution between the training and validation set resulting in the validation set having relatively few patients with no enhancing tumor label. With enhancing tumor dice scores on the test set being generally lower than validation and training set scores for almost all participants in recent BraTS challenges, we believe the test set to contain many LGG cases without enhancing tumor label therefore our postprocessing to be a promising approach to maximize our dice scores.

Figure 2 shows a qualitative example generated from our best performing model. The patient shown is taken from the validation set (CBICA_AZA_1). As can be seen in the middle (t1ce), there are several blood vessels close to the enhancing tumor. Segmentation CNNs typically struggle to correctly differentiate between such vessels and actual enhancing tumor. This is most likely due to a) a difficulty in detecting tube-like structures b) few training cases where these vessels are an issue c) the use of dice loss functions that does not sufficiently penalize false segmentations of vessels due to their relatively small size. In the case shown here, our model correctly segmented the vessels as background.

Finally we compare our best model (baseline+regions+cotraining+ postprocess) with other entries in the validation leaderboard <https://www.cbica.upenn.edu/BraTS18/lboardValidation.html> (accessed on July 14th 2018, see Table 3).



Fig. 2. Qualitative results. The case shown here is patient CBICA_AZA_1 from the validation set. Left: flair, middle: t1ce, right: our segmentation. Enhancing tumor is shown in yellow, necrosis in turquoise and edema in violet.

	Dice			HD95		
	enh.	whole	core	enh.	whole	core
xuhuaren	83.70	90.76	87.40	3.35	5.20	5.55
SCAN	79.65	90.01	85.13	3.60	4.41	5.58
ETS_livia	79.26	90.29	85.42	3.32	5.38	6.59
NVDLMED	82.33	91.00	86.68	3.92	4.52	6.85
MIC-DKFZ	81.01	90.83	85.44	2.54	4.97	7.04

Table 3. Comparison with other teams (validation leaderboard).

4 Discussion

In this paper we demonstrated that a generic U-Net architecture that has only minor modifications can obtain very competitive segmentation, if trained correctly. Our baseline model is already strong on the BraTS 2018 validation data with dice scores of 79.59, 90.80 and 84.32 for enhancing tumor, whole tumor and tumor core, respectively. Using region based training improved results for the tumor core by about one dice point. Additional training data yielded some improvements for the enhancing tumor dice. Our simple postprocessing technique of removing enhancing tumor entirely from a patient if the total number of predicted enhancing tumor voxels was below 500 proved to be effective. Overall our current best model achieves dice scores 81.01, 90.83 and 85.44 and 95th percentiles of Hausdorff Distance of 2.54, 4.97 and 7.04 for enhancing tumor, whole tumor and tumor core, respectively. We will continue to improve our model until the release of the testing data and look forward to test our model on new, unseen patients.

References

1. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, “The multimodal brain tumor image segmentation benchmark (BRATS),” *IEEE TMI*, vol. 34, no. 10, pp. 1993–2024, 2015.
2. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, “Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features,” *Nature Scientific Data*, 2017 (In Press).
3. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, “Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection,” *TCIA*, 2017.
4. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, “Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection,” *TCIA*, 2017.
5. K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, “Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation,” *MIA*, vol. 36, pp. 61–78, 2017.
6. F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, “Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge,” in *International MICCAI Brainlesion Workshop*. Springer, 2017, pp. 287–297.
7. X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P. A. Heng, “H-denseunet: Hybrid densely connected unet for liver and liver tumor segmentation from ct volumes,” *arXiv preprint arXiv:1709.07330*, 2017.
8. F. Isensee, P. F. Jaeger, P. M. Full, I. Wolf, S. Engelhardt, and K. H. Maier-Hein, “Automatic cardiac disease assessment on cine-mri via time-series segmentation and domain specific features,” in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer, 2017, pp. 120–129.
9. K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, D. Rueckert *et al.*, “Ensembles of multiple models and architectures for robust brain tumour segmentation,” in *International MICCAI Brainlesion Workshop*. Springer, 2017, pp. 450–462.
10. G. Wang, W. Li, S. Ourselin, and T. Vercauteren, “Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks,” in *International MICCAI Brainlesion Workshop*. Springer, 2017, pp. 178–190.
11. K. Kamnitsas, E. Ferrante, S. Parisot, C. Ledig, A. V. Nori, A. Criminisi, D. Rueckert, and B. Glocker, “DeepMedic for brain tumor segmentation,” in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer, 2016, pp. 138–149.
12. J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
13. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2018.
14. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *MICCAI*. Springer, 2015, pp. 234–241.

15. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3d u-net: learning dense volumetric segmentation from sparse annotation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 424–432.
16. M. Drozdal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, “The importance of skip connections in biomedical image segmentation,” in *Deep Learning and Data Labeling for Medical Applications*. Springer, 2016, pp. 179–187.
17. F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *International Conference on 3D Vision*. IEEE, 2016, pp. 565–571.
18. K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” in *ECCV*. Springer, 2016, pp. 630–645.
19. S. Jégou, M. Drozdal, D. Vazquez, A. Romero, and Y. Bengio, “The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*. IEEE, 2017, pp. 1175–1183.
20. O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, “Attention u-net: Learning where to look for the pancreas,” *arXiv preprint arXiv:1804.03999*, 2018.
21. A. G. Roy, N. Navab, and C. Wachinger, “Concurrent spatial and channel squeeze & excitation in fully convolutional networks,” *arXiv preprint arXiv:1803.02579*, 2018.
22. B. Kayalibay, G. Jensen, and P. van der Smagt, “CNN-based segmentation of medical imaging data,” *arXiv preprint arXiv:1701.03056*, 2017.
23. D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Instance normalization: The missing ingredient for fast stylization,” *arXiv preprint arXiv:1607.08022*, 2016.
24. S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv preprint arXiv:1502.03167*, 2015.
25. C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, “Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, pp. 240–248.
26. S. Lee, S. Purushwalkam, M. Cogswell, D. Crandall, and D. Batra, “Why m heads are better than one: Training a diverse ensemble of deep networks,” *arXiv preprint arXiv:1511.06314*, 2015.

Batch Normalized PixelNet for Brain Tumor Segmentation and Survival Prediction

Mobarakol Islam^{1,2}(✉), V Jeya Maria Jose^{2,3}, and Hongliang Ren²(✉)

¹ NUS Graduate School for Integrative Sciences and Engineering (NGS), National University of Singapore, Singapore

² Dept. of Biomedical Engineering, National University of Singapore, Singapore

³ Dept. of Instrumentation and Control Engineering, National Institute of Technology, Tiruchirappalli, India

mobarakol@u.nus.edu, ren@nus.edu.sg

Abstract. Segmentation of Brain tumor from magnetic resonance imaging (MRI) is a vital process to improve diagnosis, treatment planning and to study the difference between subjects with tumor and healthy subjects. In this paper, we exploit a convolutional neural network (CNN) with hypercolumn technique to segment tumor from healthy brain tissue. Hypercolumn is the concatenation of a set of vectors which form by extracting convolutional features from multiple layers. Proposed model integrates batch normalization (BN) approach with hypercolumn. BN layers help to alleviate the internal covariate shift during stochastic gradient descent (SGD) training by zero-mean and unit variance of each mini-batch. Our model achieves a mean dice score of 88.50%, 80.16% and 75.68% of whole tumor, tumor core and enhancing tumor respectively in segmentation task and 67.9% in overall survival prediction task with the validation set of BraTS 2018 challenge.

Keywords: Brain Tumor Segmentation, Glioma, Convolutional Neural Network, Hypercolumn, PixelNet, Magnetic Resonance Imaging, Survival Prediction

1 Introduction

Gliomas are the most frequent brain tumor with the highest mortality rate which develops from glial cell [6]. Early detection, accurate segmentation and estimation of the relative volume are very crucial for overall survival (OS) prediction, treatment and surgical planning. In addition, manual segmentation of tumor tissue is tedious, time consuming and required strongly supervised by a human expert. It is also prone to inter and intra-rater variability. So it is highly necessary to develop automatic segmentation system to diagnose and estimate the volume, size, shape and location of the tumor. In recent years, the success of deep learning is huge as it shows state of art performance in the applications of segmentation, classification, regression and detection. Khan et al. [8]

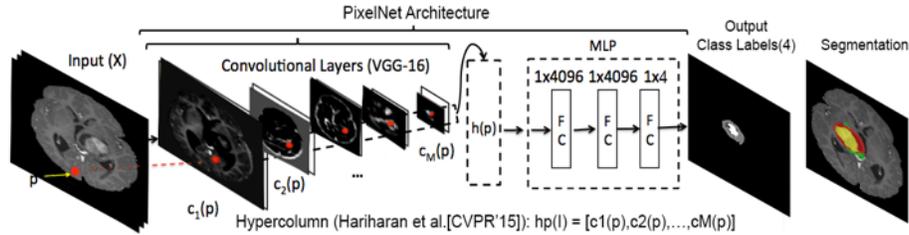


Fig. 1. Batch Normalized PixelNet Architecture.

exploits convolutional neural network (CNN) for glioma segmentation and extracts handcrafted features like histogram, co-occurrence matrix, neighbourhood gray tone difference, run length, volume and areas to predict OS using random forest regression model. Alain et al. [12] use the residual convolutional neural network with Bayesian dropout to segment tumor and calculates the geometric features (e.g. volume, heterogeneity, rim width, surface irregularity etc.) from the segmented tumor. Later, a simple artificial neural network (ANN) is utilized to predict the exact days of OS. 3D U-net and linear regression approaches are exploited to segment and predict OS [1].

In this paper, we propose a batch normalized CNN architecture with hypercolumn features inspired by multi-modal PixelNet [5, 9, 10] where a modest number of features are extracted from multiple convolution layers and trained a multi-layer perceptron (MLP) to predict segmentation class.

2 Methods

In this section, we represent our model and methodologies to process data for segmentation and overall survival prediction.

2.1 Dataset

BraTS 2018 (Brain Tumor Image Segmentation Benchmark) training database [2–4, 14] consists in total 285 cases of patients out of which the overall survival prediction data was provided for 163 cases. BraTS 2018 Validation dataset [2–4, 14] consists of 53 cases. All the above data are a multi-modal MRI scan of 210 high-grade glioma (HGG) and 75 low-grade glioma (LGG) and 4 different modalities including T1 (spin-lattice relaxation), T1c (T1-contrasted), T2 (spin-spin relaxation) and FLAIR (fluid attenuation inversion recovery). Each scan is a continuous 3D volume of 155 2D slices of size 240x240. The volume of the various modalities is already skull-stripped, aligned with T1c and interpolated to 1 mm voxel resolution. The provided ground truth with manual segmentation includes three labels: GD-enhancing tumor (ET — label 4), the peritumoral edema (ED

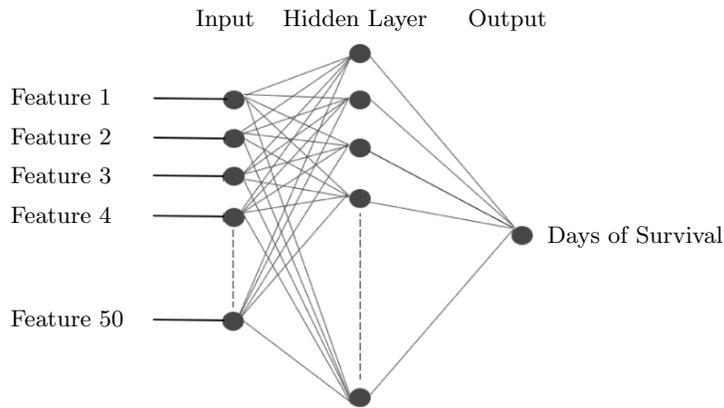


Fig. 2. Regression model using ANN.

— label 2), and the necrotic and non-enhancing tumor (NCR/NET — label 1). The predicted labels are evaluated by merging three regions: whole tumor (WT: all four labels), tumor core (TC: 1,2) and enhancing tumor (ET: 4).

2.2 Segmentation

Proposed Model Our proposed model is inspired by PixelNet [5] where we integrate batch normalization layer after each convolution layer. It consists of 15 pixel-blocks connected to a hypercolumn layer followed by a multi-layer perceptron (MLP) of 3 fully connected layers as in Fig.1. A pixel-block contains convolution, batch normalization (BN) and ReLu layers sequentially. Hypercolumn layer extracts the features from multiple convolution layers and concatenates them into a feature vector which propagates to the MLP for pixel-wise classification.

2.3 Survival Prediction

Following the extraction of the tumor from the MRI scans, the segmented tumor along with certain other parameters are used for survival prediction. The following paragraphs elucidate the features those are extracted along with the regression model that is built for predicting survival.

Feature Extraction The tumor geometry and its location hold a very important role in deciding the number of days of survival [16]. Figure 3 visually illustrates how the features such as location or centroid of the tumor, the size and shape of tumor affect the overall survival of the patient. It is evident that more the proximity of the tumor to the centre of the brain, the lower the overall survival of the patient. Also, lesser the size or smaller the shape of the tumor,

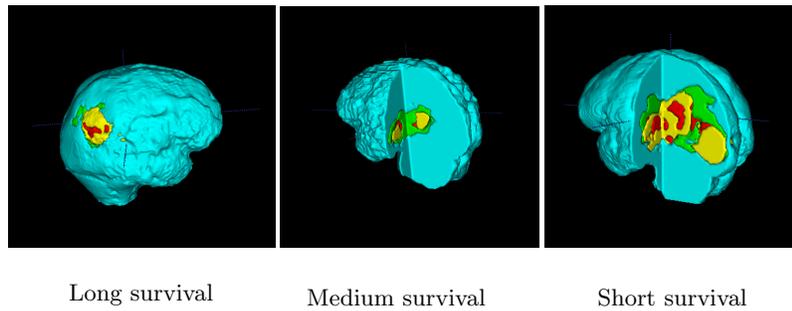


Fig. 3. Visualization of tumor with different survival rate.

higher the overall survival of the patient. So, we extract geometrical features which include First axis coordinates, Second axis coordinates, Third axis coordinates, Eigen Values, First axis length, Second axis length, Third axis Length, Centroid coordinates, meridional eccentricity and equatorial eccentricity for individual tumor type as well as whole tumor. Also, volume of the tumor and its ratio with respect to the total volume was extracted. Features of the image including mean, variance, standard deviation, entropy, skewness, Kurtosis, Entropy and histogram feature intensities of the tumor were extracted. Fractal dimension of the necrotic tumor has been found to play a pivotal role in the survival prediction, according to [17]. So, fractal dimension and fractal ratio were extracted for the necrotic part of the segmented tumor core. In addition, the age of the subject was also included.

Feature Selection Several experiments were performed using different combinations of extracted features and using the cross-validation errors, the most informative features are alone retained and the others are neglected. Features like eigenvalue and eccentricity, skewness, mean, variance were not found to have an important role in survival prediction. Also, geometric features of the GD-enhancing tumor (ET — label 4) was found to be only increasing the cross-validation errors and hence is removed. So, finally, a total of 50 features that were found to be the most informative are used in the regression model. These features are First axis coordinates, Second axis coordinates, Third axis coordinates, First axis length, Second axis length, Third axis Length, Centroid coordinates for part wise non-enhancing tumor core (NCR/NET — label 1), peritumoral edema (ED — label 2) as well as for the whole tumor without including GD-enhancing tumor (ET — label 4) in addition to Kurtosis, Entropy, Histogramic intensity, Fractal Dimension and age.

Regression Model With the selected features, we train a fully connected artificial neural network (ANN) with one hidden layer [18] and ReLu activation function. Figure 2 shows the ANN model that takes 50 features as input and

gives the number of days of survival as the output. We run an experiment to find which configuration of the hidden layers give the lowest MSE the best configuration was found to be 100 neurons in the hidden layer. Then, we find the best epoch by finding the epoch for which the mean squared error(MSE) is minimum by using a cross-validation data while training. However, by many experiments, we find that when the MSE is minimum, the accuracy is low and vice-versa. So, we find an epoch where the accuracy and the MSE of the model are balanced. Adam optimizer with MSE loss function has been used to conduct all experiments of ANN. We also tune the hyperparameters like learning rate, optimizer and batch size and the best result for BraTS2018 Validation Data was acquired for the 900th epoch with a batch size of 10.

3 Experiments and Results

3.1 Segmentation

We convert the 3D voxel 240x240x155 into 2D slices of 240x240 by ignoring blank slices of scan and ground-truth. We choose a sample of 2000 pixels per image and batch size of 5 in training time. We utilize Caffe framework with a single Nvidia GPU 1080Ti GPU to perform all the experiments. In testing phase, we feed all the pixels(240x240) into the model to predict the segmentation. Table 1 represents the Dice and Hausdorff performance of our model. It obtains dice accuracy of 88.50%, 80.16% and 75.68% of whole tumor (WT), tumor core (TC) and enhance tumor (ET) respectively.

Table 1. Dice and Hausdorff for BraTS 2018 validation dataset

	Dice			Hausdorff		
	WT	TC	ET	WT	TC	ET
Mean	88.50	80.16	75.68	5.57	7.70	4.26
StdDev	8.35	17.80	23.23	7.04	8.04	5.58
Median	90.51	86.87	83.15	3.08	4.53	2.23

3.2 Survival Prediction

Evaluation of survival prediction model is performed on the BraTS18 survival validation dataset(which is a subset (28 out of 53 subjects) of the segmentation validation dataset) [2-4,14]. Quantitative details for the ANN that gave the best accuracy of 67.9% has been listed in Table 2.

4 Discussion

For finding out the best regression model and the best features , several experiments are conducted. Experiments with different combination of the available

Table 2. Quantitative results of the survival prediction on the BraTS18 validation dataset using ANN

Cases	Accuracy	MSE	MedianSE	stdSE	SpearmanR
28	67.9%	96161.713	59473.481	117207.189	0.496

Table 3. Performance comparison with all the features in different machine learning models for BraTS2017 validation dataset.

Models	Accuracy	MSE	MedianSE	stdSE	SpearmanR
Linear Regression	50.5%	252353.061	95419.102	429879.191	0.263
SVM	33.3%	242147.277	62044.450	563941.194	0.142
Random Forest	42.4%	208660.63	33367.111	502312.762	0.213
Logistic Regression	39.4%	286470	40401	540201.363	0.479
ANN	54.5%	211967.681	53967.305	540221.112	0.206

Table 4. Performance comparison with the best 50 features in different machine learning models for BraTS2017 validation dataset.

Models	Accuracy	MSE	MedianSE	stdSE	SpearmanR
Linear Regression	50.5%	237148.501	78915.034	362705.604	0.221
SVM	42.4%	233367.604	127524.752	374037.279	0.218
Random Forest	39.4%	262224.703	42507.088	506754.007	0.324
Logistic Regression	36.4%	181509.182	58564	249213.006	0.13
ANN	60.6%	214207.487	60832.523	354332.371	0.293

features and regression models are conducted on the 33 subjects of BraTS17-Validation dataset [2, 4, 14] to find the best possible combination of features and regression model. The details of the experiments on the entire features and the best 50 features are listed in Table 3 and 3 respectively. When compared to ANN, other regression models such as Support Vector Machine(SVM) with Radial basis function(RBF) kernel [11], Random Forest [13], Linear Regression [15], Logistic Regression [7] are investigated but resulted in inferior performance. Also, using only the best 50 features showed improvement in accuracy and MSE for every models.

5 Conclusion

Batch Normalized pixelnet is found to give quality segmentations for the BraTS 2018 Validation dataset. The main advantage of the pixelnet is that it has freedom of sampling pixel during training phase. The background of the scan is removed during training and this helps the network to converge faster.

Acknowledgement

This work is supported by the Singapore Academic Research Fund under Grant R-397-000-227-112, NUSRI China Jiangsu Provincial Grant BK20150386 and

BE2016077 and NMRC Bedside & Bench under grant R-397-000-245-511 awarded to Dr. Hongliang Ren.

References

1. Amorim: 3d u-nets for brain tumor segmentation in miccai 2017 brats challenge. In: International MICCAI Brainlesion Workshop. pp. 9–14. Springer (2017)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive* 286 (2017)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *The Cancer Imaging Archive* (2017)
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data* 4, 170117 (2017)
5. Bansal, A., Chen, X., Russell, B., Gupta, A., Ramanan, D.: Pixelnet: Representation of the pixels, by the pixels, and for the pixels. arXiv preprint arXiv:1702.06506 (2017)
6. Holland, E.C.: Progenitor cells and glioma formation. *Current opinion in neurology* 14(6), 683–688 (2001)
7. Hosmer Jr, D.W., Lemeshow, S., Sturdivant, R.X.: Applied logistic regression, vol. 398. John Wiley & Sons (2013)
8. Iftexharuddin, K.M.: Glioblastoma and survival prediction. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers. vol. 10670, p. 358. Springer (2018)
9. Islam, M., Ren, H.: Multi-modal pixelnet for brain tumor segmentation. In: International MICCAI Brainlesion Workshop. pp. 298–308. Springer (2017)
10. Islam, M., Ren, H.: Class balanced pixelnet for neurological image segmentation. In: Proceedings of the 2018 6th International Conference on Bioinformatics and Computational Biology. pp. 83–87. ACM (2018)
11. Joachims, T.: Making large-scale svm learning practical. Tech. rep., Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund (1998)
12. Jungo, A., McKinley, R., Meier, R., Knecht, U., Vera, L., Pérez-Beteta, J., Molina-García, D., Pérez-García, V.M., Wiest, R., Reyes, M.: Towards uncertainty-assisted brain tumor segmentation and survival prediction. In: International MICCAI Brainlesion Workshop. pp. 474–485. Springer (2017)
13. Liaw, A., Wiener, M., et al.: Classification and regression by randomforest. *R news* 2(3), 18–22 (2002)
14. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging* 34(10), 1993 (2015)
15. Neter, J., Wasserman, W., Kutner, M.H.: Applied linear regression models (1989)

16. Pérez-Beteta, J., Martínez-González, A., Molina, D., Amo-Salas, M., Luque, B., Arregui, E., Calvo, M., Borrás, J.M., López, C., Claramonte, M., et al.: Glioblastoma: does the pre-treatment geometry matter? a postcontrast t1 mri-based study. *European radiology* 27(3), 1096–1104 (2017)
17. Reishofer, G., Studencnik, F., Koschutnig, K., Deutschmann, H., Ahammer, H., Wood, G.: Age is reflected in the fractal dimensionality of mri diffusion based tractography. *Scientific reports* 8(1), 5431 (2018)
18. Schalkoff, R.J.: *Artificial neural networks*, vol. 1. McGraw-Hill New York (1997)

Brain Tumor Segmentation and Tractographic Feature Extraction from Structural MR Images for Overall Survival Prediction

Po-Yu Kao¹, Thuyen Ngo¹, Angela Zhang¹,
Jefferson Chen², and B.S. Manjunath¹

¹ Vision Research Lab, University of California, Santa Barbara, CA, USA
poyu_kao@ece.ucsb.edu

² UC Irvine Health, University of California, Irvine, CA, USA

Abstract. This paper introduces a novel methodology to integrate human brain connectomics and parcellation for tumor segmentation and survival prediction. For segmentation, we utilize an existing brain parcellation atlas in the MNI152 1mm space and map the parcellation to each individual subject data. We leverage current state-of-the-art deep networks together with hard negative mining to achieve the final voxel level classification. This method achieves mean Dice scores of 0.904, 0.785 and 0.805 on segmentation of the whole tumor, tumor core and enhancing tumor, respectively. For survival prediction, we present a new method for combining features from connectomics data, brain parcellation, and the segmented brain tumor. Since the BraTS dataset does not include diffusion tensor imaging data for computing tractography, we use the average connectome information from the Human Connectome Project and map each subject brain volume onto this common connectome space. From this, we compute tractographic features that describe potential disruptions due to the brain tumor. These features are then used to predict the overall survival time. The proposed tractographic features achieve an average accuracy of 69% on training data and 50% on validation data.

Keywords: Brain Tumor Segmentation · Hard Negative Mining · Ensemble Modeling · Overall Survival Prediction · Tractographic Feature · Support Vector Machine

1 Introduction

Glioblastomas, or Gliomas, are one of the most common types of brain tumor. They have a highly heterogeneous appearance and shape and may happen at any location in the brain. High-Grade Glioma (HGG) is one of the most aggressive types of brain tumor with median survival of 15 months [15]. In [9], seven different 3D neural network models including two DeepMedics [10], three 3D FCNs [11] and two 3D U-Nets [4] with different parameters are combined, and the output probability maps from each model are averaged to obtain the final brain tumor mask. In [17], a hierarchical pipeline is designed to segment the

different types of tumor compartments using anisotropic convolutional neural networks. The network architecture in [6] is derived from a 3D U-Net with additional residual connections on context pathway and additional multi-scale aggregation on localization pathways, using the Dice loss in the training phase to circumvent class imbalance. For the brain tumor segmentation task, we propose a methodology to integrate multiple DeepMedics and 3D U-Nets in order to get a robust brain tumor segmentation in multimodal structural MR images. We also utilize the brain parcellation to bring the location information to DeepMedic. In order to increase the diversity of our ensemble, 3D U-Nets with dice loss and cross-entropy loss are included. A hard negative mining strategy is used while we train a 3D U-Net, with cross-entropy loss to mitigate class imbalance. The final segmentation mask of the brain tumor is calculated by taking the average of the output probability maps from each model in our ensemble.

In addition, we design a system to automatically extract important features from gliomas and predict the overall survival (OS) for glioma patients. In [14], they extract 40 features from the predicted brain tumor mask and uses a random forest regression to predict the glioma patient's OS. In [8], four features are extracted from each subject and a support vector machine (SVM) with Radial Basis Function (RBF) kernel is used to classify glioma patients into three different OS groups. In this paper, we propose a method to extract the tractographic features from the lesion regions on structural MR images via a average diffusion MR image which is averaged from a total of 1021 subjects (Q1-Q4, 2017). The volumetric features, spatial features, and morphological features are also extracted from these lesion regions. We use these lesion-related features along with age to predict the OS using a Support Vector Machine (SVM) classifier.

2 Task 1: Segmentation of Gliomas in Pre-operative MRI Scans

2.1 Materials

The Brain Tumor Segmentation (BraTS) 2018 dataset [1–3, 12] provides 285 training subjects with four different types of MR images (MR-T1, MR-T1ce, MR-T2 and MR-FLAIR) and expert-labeled ground truths of lesions, including necrosis, non-enhancing tumor, edema, and enhancing tumor. The dataset also offers 66 validation subjects with four different types of MR images and is distributed after pre-processing - i.e. co-registration to the same anatomical template, interpolated to the same resolution ($1mm^3$) and skull-stripped.

2.2 Data Pre-processing

Data normalization is a crucial step before training the deep neural network. We need to make sure the inputs have the same scale in order to speed up the training process and improve the network's performance. For each subject, we apply Z-score normalization within the brain regions for each type of MR image.

2.3 Network Architecture and Training

Our proposed ensemble combines eight 3D neural networks based on 3D U-Net [4] and DeepMedic [10]. In our ensemble, we have one original DeepMedic and four modified DeepMedics with different numbers of convolutional kernels and training parameters. The original DeepMedic does not consider the location information of each input patch inside the brain. We bring this location information to the network by creating 21 binary Harvard-Oxford subcortical brain parcellation masks [5] in the subject space as additional inputs to the neural network. We call this model DeepMedic-c25. We also train one DeepMedic with 24 additional brain parcellation channels and no data augmentation. We call this model DeepMedic-c25-n. A DeepMedic with double convolutional kernels, called DeepMedic-double and a DeepMedic with triple convolutional kernels, called DeepMedic-triple are included in our ensemble.

Our ensemble also has three 3D U-Nets, including one revised version of [6] with group normalization [18] and two 3D U-Nets, named 3D-UNet-ce-1 and 3D-UNet-ce-2 with cross-entropy loss. The 3D U-Net in [6] are trained using Dice loss in order to solve the class imbalance problem in the BraTS 2017 challenge. However, in order to increase the diversity of our ensemble, we also include two 3D U-Nets trained using cross-entropy loss. We utilize a hard negative mining strategy to solve the class imbalance problem while we train 3D U-Nets with the cross-entropy loss. In the end, we take the average of the output probability maps from each model and get the final brain tumor segmentation.

Hard Negative Mining We train 3D U-Nets with 128x128x128 patches randomly cropped from the original data. With such large dimensions, the majority of pixels are not classified as lesion and standard cross-entropy loss would encourage the model to favor the background class. To cope with this problem, we only select negative pixels with the largest losses (hard negative) to back-propagate the gradients. In our implementation the number of selected negative pixels is at most three times the number of positive pixels. Hard negative mining not only improves the performance of our classifier but also decreases its false positive rate.

2.4 Result

In this task, we compare the performance of our ensemble with each individual model. The quantitative results are shown in Table 1.

3 Task 2: Prediction of Patient Overall Survival from Pre-operative MRI Scans

3.1 Material

In addition to the data used in the brain segmentation challenge, the age (in years), survival (in days) and resection status are provided for 163 subjects in

Table 1. Quantitative results of the individual model and our ensemble on BraTS 2018 validation dataset (reported as mean \pm one standard deviation). Bold numbers highlight the best result for a given metric and tumor compartments (ET: enhancing tumor, WT: whole tumor, and TC: tumor core)

	Model	ET	WT	TC
<i>Dice</i>	DeepMedic	0.772 \pm 0.248	0.897 \pm 0.071	0.801 \pm 0.219
	DeepMedic-c25	0.752 \pm 0.280	0.895 \pm 0.064	0.805 \pm 0.230
	DeepMedic-c25-n	0.766 \pm 0.261	0.893 \pm 0.070	0.788 \pm 0.244
	DeepMedic-double	0.757 \pm 0.260	0.893 \pm 0.088	0.782 \pm 0.231
	DeepMedic-triple	0.777 \pm 0.236	0.895 \pm 0.085	0.792 \pm 0.230
	3D U-Net	0.770 \pm 0.251	0.892 \pm 0.060	0.770 \pm 0.224
	3D-UNet-ce-1	0.745 \pm 0.271	0.884 \pm 0.127	0.768 \pm 0.222
	3D-UNet-ce-2	0.745 \pm 0.269	0.891 \pm 0.110	0.760 \pm 0.244
	Our ensemble	0.785 \pm 0.248	0.904 \pm 0.073	0.805 \pm 0.218
<i>Sensitivity</i>	DeepMedic	0.816 \pm 0.217	0.919 \pm 0.089	0.802 \pm 0.217
	DeepMedic-c25	0.803 \pm 0.250	0.925 \pm 0.072	0.794 \pm 0.248
	DeepMedic-c25-n	0.806 \pm 0.244	0.918 \pm 0.092	0.792 \pm 0.251
	DeepMedic-double	0.794 \pm 0.235	0.913 \pm 0.113	0.801 \pm 0.225
	DeepMedic-triple	0.800 \pm 0.230	0.911 \pm 0.110	0.789 \pm 0.243
	3D U-Net	0.784 \pm 0.248	0.919 \pm 0.085	0.791 \pm 0.245
	3D-UNet-ce-1	0.785 \pm 0.249	0.896 \pm 0.160	0.791 \pm 0.230
	3D-UNet-ce-2	0.802 \pm 0.251	0.904 \pm 0.133	0.803 \pm 0.249
	Our ensemble	0.806 \pm 0.245	0.918 \pm 0.102	0.795 \pm 0.236
<i>Specificity</i>	DeepMedic	0.998 \pm 0.004	0.994 \pm 0.005	0.997 \pm 0.004
	DeepMedic-c25	0.998 \pm 0.003	0.993 \pm 0.006	0.998 \pm 0.003
	DeepMedic-c25-n	0.998 \pm 0.004	0.993 \pm 0.005	0.997 \pm 0.004
	DeepMedic-double	0.998 \pm 0.003	0.993 \pm 0.005	0.997 \pm 0.004
	DeepMedic-triple	0.998 \pm 0.004	0.993 \pm 0.005	0.997 \pm 0.003
	3D U-Net	0.998 \pm 0.003	0.993 \pm 0.005	0.997 \pm 0.003
	3D-UNet-ce-1	0.998 \pm 0.003	0.994 \pm 0.006	0.996 \pm 0.007
	3D-UNet-ce-2	0.998 \pm 0.003	0.994 \pm 0.005	0.996 \pm 0.004
	Our ensemble	0.998 \pm 0.003	0.994 \pm 0.005	0.998 \pm 0.003
<i>Hausdorff</i> ₉₅ (in mm)	DeepMedic	4.723 \pm 9.018	6.290 \pm 5.570	9.181 \pm 13.612
	DeepMedic-c25	3.376 \pm 4.918	10.992 \pm 18.590	9.902 \pm 18.277
	DeepMedic-c25-n	4.108 \pm 5.920	8.682 \pm 11.907	9.928 \pm 14.247
	DeepMedic-double	4.696 \pm 8.683	5.969 \pm 7.631	10.919 \pm 17.270
	DeepMedic-triple	5.826 \pm 12.333	6.583 \pm 9.590	9.787 \pm 12.743
	3D U-Net	4.180 \pm 7.920	5.657 \pm 7.353	10.641 \pm 14.207
	3D-UNet-ce-1	4.416 \pm 7.850	5.627 \pm 7.609	10.127 \pm 13.655
	3D-UNet-ce-2	5.956 \pm 11.858	4.726 \pm 4.435	10.882 \pm 15.407
	Our ensemble	3.744 \pm 7.832	3.946 \pm 3.166	7.833 \pm 13.076

the training dataset. 59 of them have resection status of Gross Total Resection (GTR). The validation dataset has 53 subjects with age (in years) and resection status. 28 of them have resection status GTR. For this task we predict the overall survival (OS) for glioma patients with resection status of GTR.

3.2 Methodology

Our proposed pipeline, shown in Figure 1, includes three stages: In the first stage, we use all structural MR images from each subject to obtain the predicted lesion mask for each subject. We then extract five different types of features from each subject. Four of them are from the predicted lesion mask, and the remaining one is the subject's age. In the second stage, we perform feature concatenation, feature normalization, and feature selection. In the final stage, we train a classifier using the features extracted from the training subjects and evaluate its performance on the validation subjects.

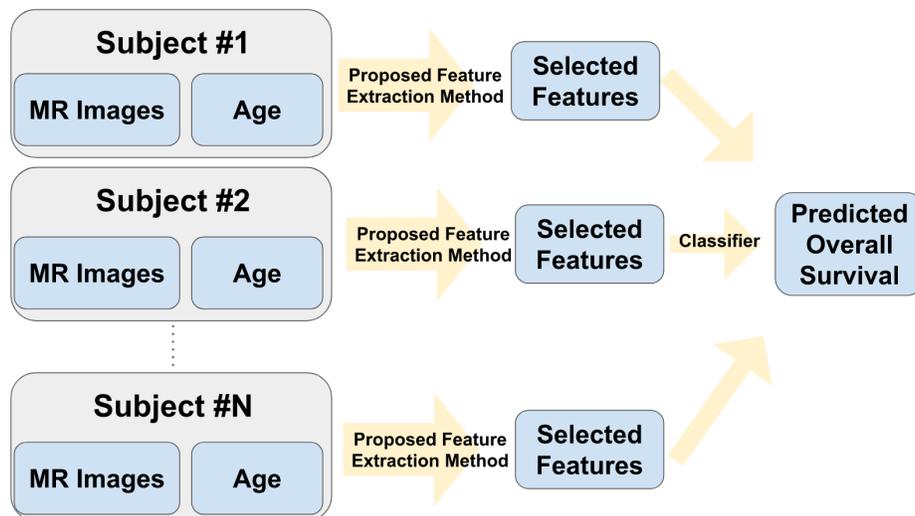


Fig. 1. Overall survival classification pipeline.

Glioma Segmentation: To segment the gliomas, we use the proposed method in the previous section to obtain the prediction of three different tissue types which includes necrosis, non-enhancing tumor, edema, and enhancing tumor.

Feature Extraction from the Glioma Segmentation: After we obtain the predicted lesion mask, we extract four different types of features from this mask for each subject. These features include volumetric features, spatial features,

morphological features, and tractographic features. For each subject, we extract approximately 800 features. The feature extraction pipeline for one subject is shown in Figure 2.

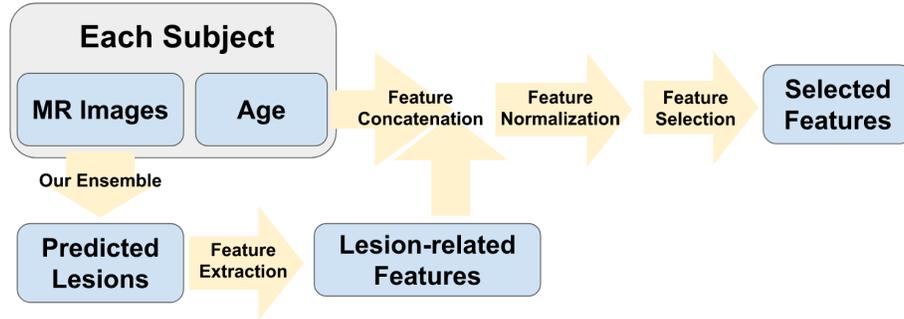


Fig. 2. Feature extraction pipeline for one subject.

Volumetric Features: The volumetric features include the volume and the ratio of brain to the different types of lesions, as well as the tumor compartments. We have 19 volumetric features from each subject.

Spatial Features: The spatial features describe the location of the tumor in the brain. We extract the centroids and volumes of different types of tumor tissue in each of the Harvard-Oxford subcortical structural atlas brain parcellation regions. FMRIB’s Linear Image Registration Tool (FLIRT) [7] is used to map the Harvard-Oxford subcortical structural atlas from the MNI152 1mm space to each individual’s subject space. For each subject, we extract around 90 spatial features.

Morphological Features: The morphological features include the length of the major axis of the lesion, the length of the minor axis of the lesion and the surface irregularity of the lesions. We extract 19 morphological features from each subject.

Tractographic Features: Tractographic features describe the potentially damaged parcellation regions impacted by the brain tumor through fiber tracking. Figure 3 shows the workflow for building a connectivity matrix for each subject. First, the predicted whole tumor mask and the average diffusion orientation distribution function from HCP-1021, created by QSDR [19], are obtained for each subject. FLIRT is used to map the whole tumor mask from subject space to MNI152 1mm space. Second, we use a deterministic diffusion fiber tracking method [20] to create approximately 1,000,000 tracts from the whole tumor

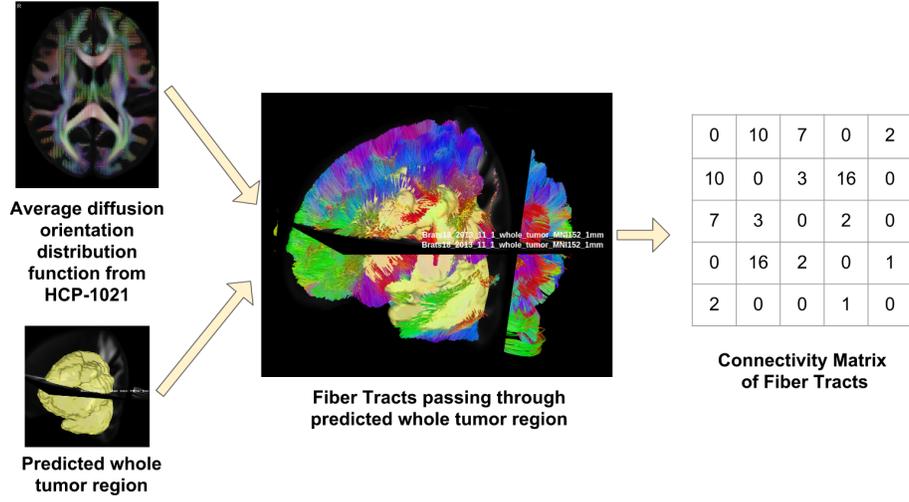


Fig. 3. Workflow for building a connectivity matrix for each subject. ITK-SNAP [21] is used for visualizing the 3D MR images and 3D labels.

region. Finally, a structural brain atlas is used to create a connectivity matrix W_{ori} for each subject. This matrix contains information about whether a fiber connecting one region to another passed through or ended at those regions, as shown:

W_{ori} is a $N \times N$ matrix, and N is the number of parcellation in a structural brain atlas.

$$W_{ori} = \begin{bmatrix} w_{ori,11} & w_{ori,12} & \cdots & w_{ori,1N} \\ w_{ori,21} & w_{ori,22} & \cdots & w_{ori,2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{ori,N1} & w_{ori,N2} & \cdots & w_{ori,NN} \end{bmatrix} \quad (1)$$

If w_{ij} is pass type, it shows the number of tracts passing through region j and region i . if w_{ij} is end type, it shows the number of tracts starting from a region i and ending in a region j . From the original connectivity matrix W_{ori} , we create a normalized version W_{nrm} and a binarized version W_{bin} .

$$W_{nrm} = W_{ori} / \max(W_{ori}) \quad (2)$$

$/$ is the element-wise division operator and $\max(W_{ori})$ is the maximum value of the original connectivity matrix W_{ori} .

$$W_{bin} = \begin{bmatrix} w_{bin,11} & w_{bin,12} & \cdots & w_{bin,1N} \\ w_{bin,21} & w_{bin,22} & \cdots & w_{bin,2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{bin,N1} & w_{bin,N2} & \cdots & w_{bin,NN} \end{bmatrix} \quad (3)$$

$w_{bin,ij} = 0$ if $w_{ori,ij} = 0$, and $w_{bin,ij} = 1$ if $w_{ori,ij} > 0$. Then, we sum up each column in a connectivity matrix to form a tractographic feature vector.

$$\mathbf{V} = \sum_{i=1}^N w_{ij} = [v_1, v_2, \dots, v_N] \quad (4)$$

Furthermore, we weight every element in the tractographic feature vector with respect to the ratio of the lesion in a brain parcellation region to the volume of this brain parcellation region.

$$\mathbf{V}_{wei} = \boldsymbol{\alpha} \odot \mathbf{V}, \boldsymbol{\alpha} = [t_1/b_1, t_2/b_2, \dots, t_N/b_N] \quad (5)$$

\odot is the element-wise multiplication operator, t_i is the volume of the whole brain tumor in the i -th brain parcellation, and b_i is the volume of the i -th brain parcellation. Automated Anatomical Labeling (AAL) [16] is used, so the dimension of the connectivity matrix \mathbf{W} is 116×116 . In the end, we extract six tractographic feature vectors from each subject which have 696 dimensions.

Feature Concatenation, Normalization and Selection: We concatenate all of the obtained features together with subject's age to form a final high dimensional feature vector for each subject. Then, we remove features that have low variance between subjects and apply z-score normalization on the remaining features. In the feature selection step, we combine recursive feature elimination with the repeated stratified K-fold cross validator. These feature processing steps are implemented by using scikit-learn [13].

Prediction of Patient Overall Survival: We first divide all 59 training subjects into three groups: long-survivors (e.g., >15 months), short-survivors (e.g., <10 months), and mid-survivors (e.g., between 10 and 15 months). Then, we train an SVM on all training subjects. Finally, we obtain predicted labels for the validation subjects and upload them to the evaluation website.

3.3 Results

In the overall survival prediction task, we examine the performance of each type of feature compared to combined features. The ground truth of lesions are used to extract four different types of features. Recursive feature elimination and cross-validated selection (RFECV) are used in the feature selection step, and a Support Vector Machine (SVM) is trained with 1000-times repeated stratified 5-folds cross-validator. The classification accuracy with one standard deviation for the training dataset is shown in Figure 4. By only using the tractographic features with RECV and a SVM classifier, the classification accuracy is 0.5 in the validation dataset. However, the classification accuracy drops to 0.429 when we use all combined features with RFECV and a SVM classifier in the validation dataset.

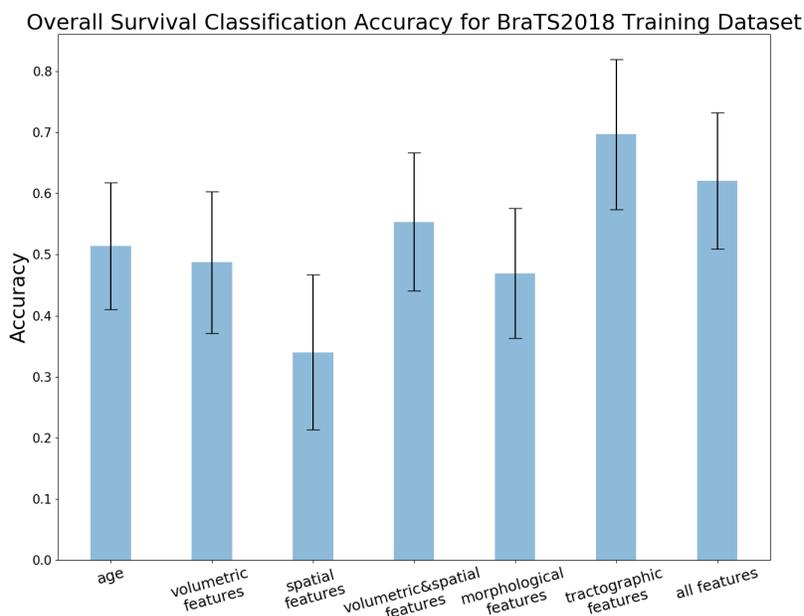


Fig. 4. Classification accuracy between each different types of features and all combined features. We consider the volume of different types of lesions in different brain parcellation regions as volumetric&spatial features.

4 Discussion

For the brain tumor segmentation task, our ensemble has better performance on all *Dice* scores for all tumor compartments compared to each individual model. It also has better performance on *Hausdorff* distances for the whole tumor and tumor core. *Specificity* (also called the true negative rate) are pretty similar between each individual model and our ensemble due to the class imbalance of the foreground and the background.

DeepMedic with 21 more binary brain parcellation channels has better performance on the *sensitivity* (also called the true positive rate) of whole tumor region and the *Hausdorff* distance of enhancing tumor region compared to other models and our ensemble. Original DeepMedic has a better performance on the *sensitivity* of enhancing tumor region, and the 3D U-Net which is train by cross-entropy loss has a better performance on the *sensitivity* of tumor core region compared to our ensemble and other models.

For the overall survival prediction task, the 12 selected tractographic features achieve 69 % classification accuracy which is better than other types of features and the combined features. These tractographic features are from the left precentral gyrus, right precentral gyrus, left insula, right insula, right mid-

cingulate area, left posterior cingulate gyrus, left superior occipital gyrus, right fusiform gyrus, right globus pallidus, right middle temporal pole, left inferior temporal gyrus and right inferior temporal gyrus. These regions are the key regions to affect the brain tumor patient's survival period.

The features we used to report the classification accuracy in Figure 4 are from the ground truth lesions. We repeat 5-fold cross-validation 1000 times to train a SVM classifier with different types of features and all combined features in order to find more reliable features. The tractographic features has the best performance among other different types of feature and the combined features.

However, when we evaluate the tractographic features which are extracted from the predicted tumor segmentation, the classification accuracy drops to 0.5. We believe this drop is due to the imperfection of our tumor segmentation tool.

5 Conclusions

Our proposed ensemble has the best performance on the *Dice* scores of the whole tumor, tumor core and enhancing tumor and the smallest *Hausdorff* distance in the whole tumor and tumor core compared to each individual model. It is shown that increasing the number of model and the diversity of model in our ensemble results in a more robust tumor segmentation.

Our proposed tractographic features are reliable features to predict the overall survival for the tumor patients. The following 12 regions are the pivotal: left precentral gyrus, right precentral gyrus, left insula, right insula, right midcingulate area, left posterior cingulate gyrus, left superior occipital gyrus, right fusiform gyrus, right globus pallidus, right middle temporal pole, left inferior temporal gyrus and right inferior temporal gyrus. The tumor spreading in these 12 regions has a greater impact on patient's overall survival compared to other regions inside the brain.

References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. The Cancer Imaging Archive. (2017). <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. The Cancer Imaging Archive. (2017). <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data* **4**, 170117 (2017)
4. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 424–432. Springer (2016)

5. Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al.: An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage* **31**(3), 968–980 (2006)
6. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In: *International MICCAI Brainlesion Workshop*. pp. 287–297. Springer (2017)
7. Jenkinson, M., Smith, S.: A global optimisation method for robust affine registration of brain images. *Medical image analysis* **5**(2), 143–156 (2001)
8. Jungo, A., McKinley, R., Meier, R., Knecht, U., Vera, L., Pérez-Beteta, J., Molina-García, D., Pérez-García, V.M., Wiest, R., Reyes, M.: Towards uncertainty-assisted brain tumor segmentation and survival prediction. In: *International MICCAI Brainlesion Workshop*. pp. 474–485. Springer (2017)
9. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., et al.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: *International MICCAI Brainlesion Workshop*. pp. 450–462. Springer (2017)
10. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis* **36**, 61–78 (2017)
11. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 3431–3440 (2015)
12. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging* **34**(10), 1993 (2015)
13. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011)
14. Shboul, Z.A., Vidyaratne, L., Alam, M., Iftekharruddin, K.M.: Glioblastoma and survival prediction. In: *International MICCAI Brainlesion Workshop*. pp. 358–368. Springer (2017)
15. Thakkar, J.P., Dolecek, T.A., Horbinski, C., Ostrom, Q.T., Lightner, D.D., Barnholtz-Sloan, J.S., Villano, J.L.: Epidemiologic and molecular prognostic review of glioblastoma. *Cancer Epidemiology and Prevention Biomarkers* (2014)
16. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage* **15**(1), 273–289 (2002)
17. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: *International MICCAI Brainlesion Workshop*. pp. 178–190. Springer (2017)
18. Wu, Y., He, K.: Group normalization. *arXiv preprint arXiv:1803.08494* (2018)
19. Yeh, F.C., Tseng, W.Y.I.: Ntu-90: a high angular resolution brain atlas constructed by q-space diffeomorphic reconstruction. *Neuroimage* **58**(1), 91–99 (2011)
20. Yeh, F.C., Verstynen, T.D., Wang, Y., Fernández-Miranda, J.C., Tseng, W.Y.I.: Deterministic diffusion fiber tracking improved by quantitative anisotropy. *PloS one* **8**(11), e80713 (2013)

21. Yushkevich, P.A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J.C., Gerig, G.: User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* **31**(3), 1116–1128 (2006)

Brain Tumor Segmentation in Multimodal 3D-MRI of BraTS'2018 Datasets using Deep Convolutional Neural Networks

Adel Kermi¹, Issam Mahmoudi¹ and Mohamed Tarek Khadir²

¹ LMCS Laboratory, National Higher School of Computer Sciences (ESI), BP.68M, Oued-Smar, 16309, El-Harrach, Algiers, Algeria

² LabGed Laboratory, Dept. of Computer Sciences, University Badji-Mokhtar of Annaba, BP.12, 23000, Annaba, Algeria

{a_kermi, di_mahmoudi}@esi.dz, khadir@labged.net

Abstract. This paper presents a fully automated and accurate brain tumor segmentation method based on 2D Deep Convolutional Neural Networks (DNNs) which automatically extracts whole tumor and intra-tumor regions, including enhancing tumor, edema and necrosis, from pre-operative multimodal MRI volumes of BraTS'2018 Datasets. Our network architecture was inspired by the U-net and has been modified to increase brain tumor segmentation performance. Among modifications, we employ Weighted Cross Entropy (WCE) and Generalized Dice Loss (GDL) as a loss function to address the class imbalance problem in the brain tumor data. The proposed segmentation system has been tested and evaluated on both, BraTS'2018 training and validation datasets. On the challenge validation dataset, our system achieved a mean enhancing tumor, whole tumor, and tumor core dice score of 0.783, 0.868 and 0.805 respectively. Other quantitative and qualitative evaluations are presented and discussed.

Keywords: Brain tumor segmentation, 3D-MRI, deep learning, convolutional neural networks, U-net, BraTS'2018.

1 Introduction

Brain tumor segmentation in multimodal Magnetic Resonance Imaging (MRI) is widely used as a vital process for surgical planning and simulation, treatment planning prior to radiation therapy, therapy evaluation [1-5], and intra-operative neuro navigation and image neurosurgery [6-8]. However, segmenting brain tumor manually is not only challenging task but also a time-consuming one, favoring therefore, the emergence of computerized approaches.

Despite considerable research works and encouraging results in the medical imaging domain, fast and precise 3D computerized brain tumors segmentation remains until now a challenging process and a very difficult task to achieve because brain tumors may appear in any size, shape, location and image intensity [2-5]. Many research recently

used deep learning methods [9], specifically Convolutional Neural Networks (CNNs) [9-12] that have shown their effectiveness and proved successful to automatically classify the normal and pathological brain MRI scans in the past few BraTS challenges as well as other semantic and medical segmentation problems.

This paper proposes an automated and accurate segmentation method of whole tumor and intra-tumor structures, including enhancing tumor, edema and necrosis, in multimodal 3D-MRI. It is based on 2D Deep Convolutional Neural Networks (DNNs) using a modified U-net architecture. The proposed DNN model is trained to segment both High Grade Glioma (HGG) and Lower Grade Glioma (LGG) volumes.

The rest of the paper is organized as follows. First, Section 2 presents an overview of the proposed segmentation method. Experimental results with their evaluations are given in Section 3. Finally, a conclusion and future work are presented in Section 4.

2 The proposed method

The proposed segmentation system is entirely automatic. The brain tumor segmentation process is based on deep learning more precisely on 2D Convolutional Neural Networks. It includes the main following steps: pre-processing of the 3D-MRI data, training using a U-net architecture, and brain tumoral structures prediction.

2.1 Data and their pre-processing

The BraTS'2018 challenge training dataset consists of 210 pre-operative multimodal MRI scans of subjects with HGG and 75 scans of subjects with LGG, and the BraTS'2018 challenge validation dataset includes 66 different multimodal 3D-MRI [10-13]. Images were acquired at 19 different centers using MR scanners from different vendors and with 3T field strength. They comprise co-registered native (T1) and contrast-enhanced T1-weighted (T1Gd) MRI, as well as T2-weighted (T2) and T2 Fluid Attenuated Inversion Recovery (FLAIR) MRI. All 3D-MRI of BraTS'2018 dataset have a volume dimension of $240 \times 240 \times 155$. They are distributed, co-registered to the same anatomical template and interpolated to the same resolution (1mm³). All MRI volumes have been segmented manually, by one to four raters, and their annotations were approved by experienced neuro-radiologists. Each tumor was segmented into edema, necrosis and non-enhancing tumor and active/enhancing tumor.

First, a minimal pre-processing of MRI data is applied. We employed the same pre-processing, as in [14]. We removed the 1% highest and lowest intensities, then each modality of MR images was normalized by subtracting the mean and dividing by the standard deviation of the intensities within the slice. To address the class imbalance problem in the data, we also employed data augmentation technique [15], which consists of adding new synthetic images by performing operations and transformations on data and the corresponding ground truth. The transformations comprise rotation, translation, and horizontal flipping and mirroring.

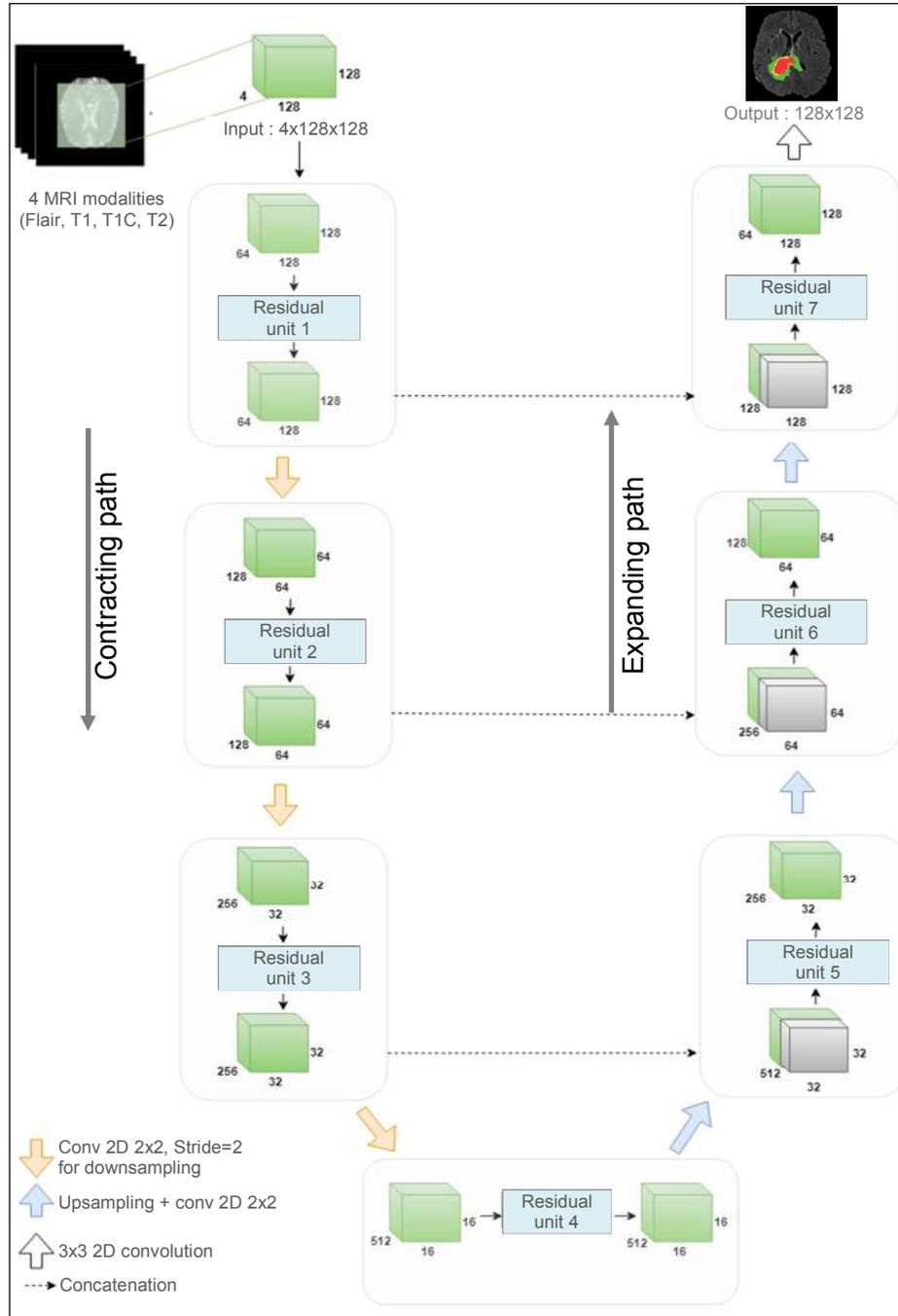


Fig. 1. Architecture of the proposed Deep Convolutional Neural Network.

2.2 Network architecture and Training

The CNN used in this study has a similar architecture as that of U-net [16]. Our network architecture can be seen in Fig. 1. It consists of a contracting path (left side) and an expanding path (right side). The contracting path consists of 3 pre-activated residual blocks, as in [17, 18], instead of plain blocks in the original U-net. Each block has two convolution units each of which comprises a Batch Normalization (BN) layer, an activation function, called Parametric Rectified Linear Unit (PReLU) [19], instead of ReLU function used in the original architecture [16], and a convolutional layer, like in [20], instead of using Maxpooling [16], with Padding = 2, Stride = 1 and a 3 x 3 size filter. For downsampling, a convolution layer with a 2 x 2 filter and a stride of 2 is applied. At each downsampling step we double the number of feature channels. The contracting path is followed by a fourth residual unit that acts as a bridge to connect both paths. In the same way, the expanding path is built using 3 residual blocks. Before each block, there is an upsampling operation which increases the feature map size by 2, followed by a 2 x 2 convolution and a concatenation with the feature maps corresponding to the contracting path. At the last layer of the expanding path, a 1 x 1 convolution with the Softmax activation function is used to map the multi-channel feature maps to the desired number of classes.

In total, the proposed network model contains 7 residual blocks, 25 convolution layers, 15 layers of BN and 10159748 parameters to optimize.

Our network was trained with axial slices extracted from training MRI set, including HGG and LGG cases, and the corresponding ground truth segmentations. The goal is to find the network parameters (weights and biases) that minimize a loss function. In this work, this can be achieved by using Stochastic Gradient Descent algorithm (SGD) [15], at each iteration SGD updates the parameters towards the opposite direction of the gradients. In our network model, we used a loss function that adds Weighted Cross Entropy (WCE) [15] and Generalized Dice (GDL) [21] to address the class imbalance problem present in brain tumor data. So, the two components of the loss function are defined as follows:

$$WCE = -\frac{1}{K} \sum_k \sum_i^L w_i g_{ik} \log(p_{ik}) \quad (1)$$

$$GDL = 1 - 2 \frac{\sum_i^L w_i \sum_k g_{ik} p_{ik}}{\sum_i^L w_i \sum_k (g_{ik} + p_{ik})} \quad (2)$$

where L is the total number of labels, K denotes the batch size. w_i represents the weight assigned to the i th label. p_{ik} and g_{ik} represent the value of the (i th, k th) pixel of the segmented binary image and of the binary ground truth image, respectively.

2.3 Brain tumoral structures prediction

After training our network, we use it to perform the prediction. This step consists to provide the network with the four MRI modalities of an unsegmented volume that it

has never processed or encountered before, and it must be able to return a segmented image.

3 Experimental results and discussion

In this study, we have tested and evaluated our segmentation system on pre-operative multimodal MRI scans of both the training/testing and the validation datasets of the BraTS'2018 challenge [1-4]. The results of automatically segmented tumors, denoted by A , can be compared with manually segmented tumors by human experts, denoted by B , which are considered as ground truth for evaluation.

3.1 Evaluation metrics

The BraTS competition has four metrics to assess the accuracy of segmentation results and to measure the similarity between the segmentations A and B . These measures can be defined with the following expressions [2, 22]:

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (3)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

$$HD(A, B) = \max(h(A, B), h(B, A)) \quad (6)$$

The *Dice* metric is computed as performance metric. It measures states the similarity between two volumes A and B , corresponding to the output segmentation of the model and clinical ground truth annotations, respectively. *Sensitivity* and *Specificity* are statistical measures employed to evaluate the behavior of the predictions and the proportions of True Positives (TP), False Negatives (FN), False Positives (FP) and True Negatives (TN) voxels. The *Sensitivity* metric, also known as true positive detection rate (TPR), measures the proportion of positive voxels of the real brain tumor that are correctly segmented as such, while *Specificity* metric, also called true negative detection rate (TNR), indicates how well the true negatives are predicted. Employing *Sensitivity* with *Specificity* can provide a good assessment of the segmentation result. $HD(A, B)$ is the *Hausdorff distance* between the two surfaces of A and B , where $h(A, B) = \max_{a \in A} \min_{b \in B} d(a, b)$, and $d(a, b)$ represents the Euclidean distance between a and b such as a and b are points of A and B , respectively. This metric indicates the segmentation quality at the border of the tumors by evaluating the greatest distance between the two segmentation surfaces, and is independent of the tumor size.

3.2 Performance on BraTS'2018 training dataset

Preliminary segmentation results for the 285 3D-MRI of the BraTS'2018 training dataset have been obtained using 80 % of this dataset (228 subjects) for training and the remaining 20 % (57 subjects) for the validation purposes. Results obtained by our automated system for 10 sample cases are shown in Fig. 2 and Fig. 3. Fig. 2 shows segmentation results from 5 multimodal MRI where HGG tumors are present and Fig. 3 shows other segmentation results from other 5 MRI with LGG tumors. In these figures, each row represents one clinical case. In the first four columns from left to right, images show one axial slice of MRI acquired in Flair, T1, T1C and T2 modality, respectively, used as input channels to our CNN model. In the fifth and the sixth columns, images show the ground truth (GT) and the prediction labels respectively, where we can distinguish intra-tumoral regions by color-code: enhancing tumor (yellow), peritumoral edema (green) and necrotic and non-enhancing tumor (red). As it can be seen, tumors in the brain MRI of the 10 cases vary in size, shape, position and intensity. By visual inspection, the proposed method usually generates segmentations (Prediction) enough alike to the ones obtain by the experts (GT).

Table 1. Quantitative evaluation of segmentation results on 20 % of BraTS'2018 training dataset (57 MRI scans) by using *Dice* and *Sensitivity* metrics.

	<i>Dice</i>			<i>Sensitivity</i>		
	<i>ET</i>	<i>WT</i>	<i>TC</i>	<i>ET</i>	<i>WT</i>	<i>TC</i>
<i>Mean</i>	0.717	0.867	0.798	0.778	0.907	0.84
<i>Std. Dev.</i>	0.275	0.078	0.226	0.3	0.107	0.223
<i>Median</i>	0.831	0.887	0.889	0.912	0.943	0.927
<i>25quantile</i>	0.726	0.839	0.808	0.777	0.896	0.847
<i>75quantile</i>	0.859	0.926	0.935	0.951	0.974	0.961

Table 2. Quantitative evaluation of segmentation results on 20 % of BraTS'2018 training dataset (57 MRI scans) by using *Specificity* and *Hausdorff distance* metrics.

	<i>Specificity</i>			<i>Hausdorff95</i>		
	<i>ET</i>	<i>WT</i>	<i>ET</i>	<i>WT</i>	<i>ET</i>	<i>WT</i>
<i>Mean</i>	0,999	0,998	0,999	4,742	8,706	6,4
<i>Std. Dev.</i>	0,001	0,001	0,002	2,079	2,822	3,685
<i>Median</i>	1	0,999	1	4,123	8,062	5,099
<i>25quantile</i>	0,999	0,998	0,999	3	6,442	3,871
<i>75quantile</i>	1	0,999	1	6,633	10,951	8,303

A quantitative evaluation of segmentation results of enhancing tumor (ET), whole tumor (WT) and tumor core (TC) using the four previously defined metrics is given in Table 1 and Table 2. Mean, standard deviation, median and 25th and 75th percentile are given for *Dice* and *Sensitivity* metrics in Table 1 and for *Specificity* and *Hausdorff distance* in Table 2. Values presented in Table 1 show high performance on the Dice metric for WT region, but lower performance for ET and TC regions.

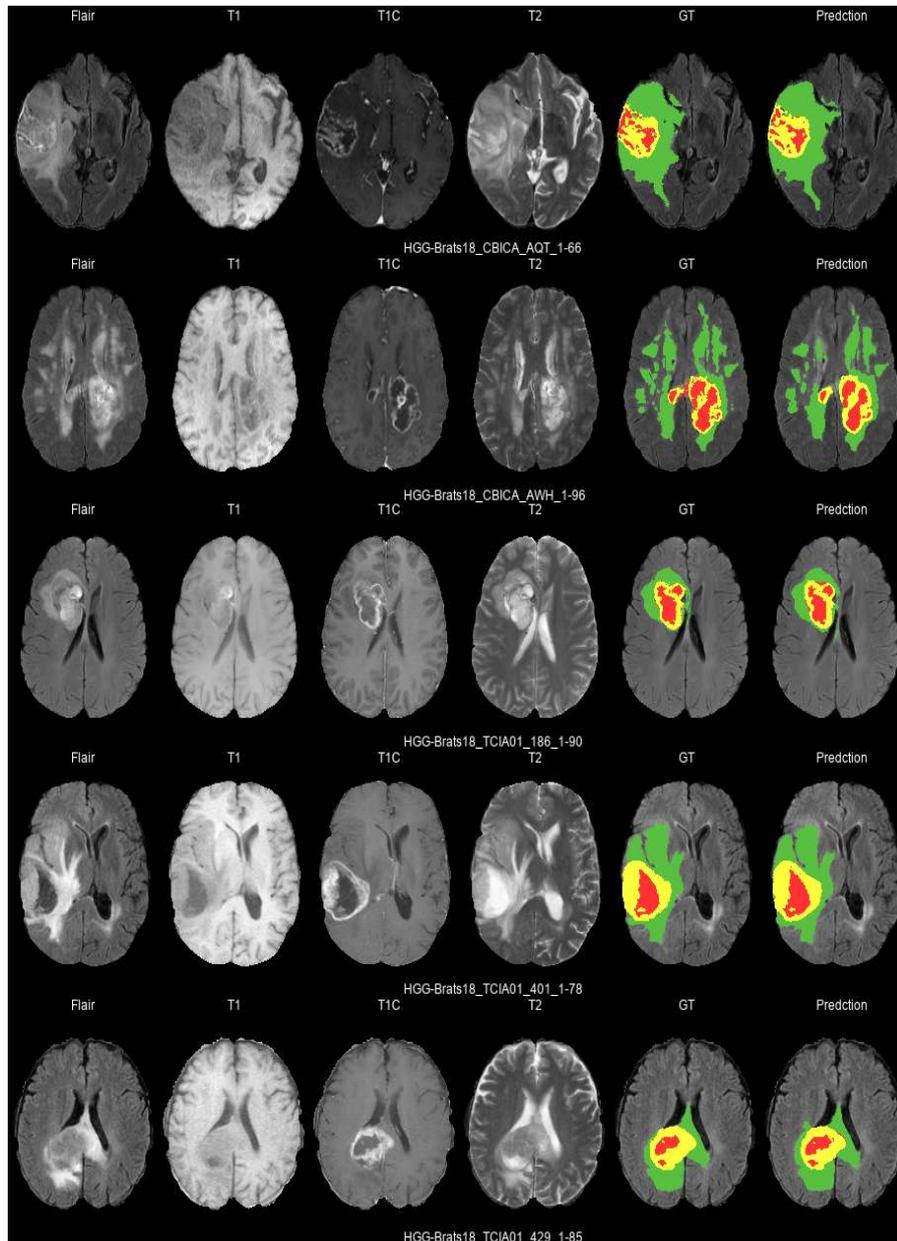


Fig. 2. Intra-tumoral structures segmentation results from 5 multimodal 3D-MRI with HGG of BraTS'2018 training dataset corresponding to 5 different subjects.

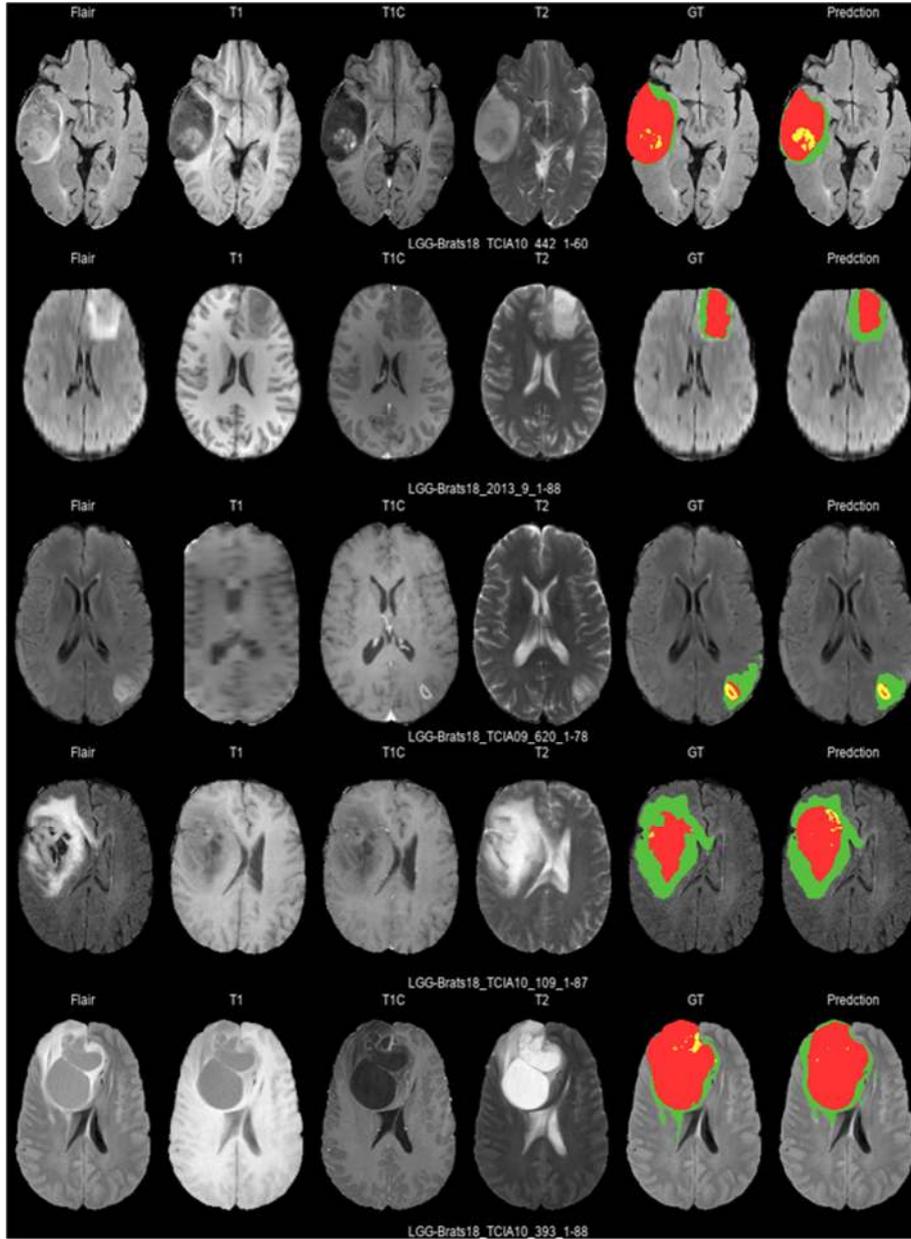


Fig. 3. Intra-tumoral structures segmentation results from 5 other multimodal 3D-MRI with LGG of BraTS'2018 training dataset corresponding to 5 different subjects.

3.3 Performance on BraTS'2018 validation dataset

For our participation to BraTS'2018 competition, we used 100 % of the training dataset (285 subjects) for training purpose. The performance on BraTS'2018 validation dataset, which is composed of 66 subjects, is diffused in the challenge leaderboard Web site¹ and presented with more statistics in Table 3, Table 4, Fig. 4 and Fig. 5. In this context, we can compare our segmentation results with those of other participants. Our method achieved a mean ET, WT, and TC dice score of 0.783, 0.868 and 0.805 respectively. These scores are close to those obtained by the top performing methods. Also, we obtained average *HD* scores of 3.728, 8.127 and 9.84 for ET, WT, and TC, respectively. These scores are not far from those obtained by the best performing methods.

Table 3. Quantitative evaluation of segmentation results on BraTS'2018 validation dataset (66 MRI scans) by using *Dice* and *Sensitivity* metrics.

	<i>Dice</i>			<i>Sensitivity</i>		
	<i>ET</i>	<i>WT</i>	<i>TC</i>	<i>ET</i>	<i>WT</i>	<i>TC</i>
<i>Mean</i>	0.783	0.868	0.805	0.826	0.895	0.807
<i>Std. Dev.</i>	0.216	0.101	0.199	0.241	0.149	0.222
<i>Median</i>	0.846	0.898	0.891	0.901	0.955	0.895
<i>25quantile</i>	0.769	0.855	0.756	0.82	0.901	0.71
<i>75quantile</i>	0.893	0.919	0.928	0.969	0.971	0.965

Table 4. Quantitative evaluation of segmentation results on BraTS'2018 validation dataset (66 MRI scans) by using *Specificity* and *Hausdorff distance* metrics.

	<i>Specificity</i>			<i>Hausdorff95</i>		
	<i>ET</i>	<i>WT</i>	<i>ET</i>	<i>WT</i>	<i>ET</i>	<i>WT</i>
<i>Mean</i>	0.997	0.991	0.997	3.728	8.127	9.84
<i>Std. Dev.</i>	0.004	0.007	0.003	4.471	10.426	15.385
<i>Median</i>	0.998	0.993	0.998	2.236	4.243	5.431
<i>25quantile</i>	0.997	0.988	0.997	1.637	3	2.871
<i>75quantile</i>	0.999	0.996	0.999	3.317	7.778	10.728

Boxplots showing the dispersion of Dice and Sensitivity scores are represented in Fig. 4 and boxplots of the dispersion of Specificity and HD scores are represented in Fig. 5. In these figures, boxplots show quartile ranges of the scores; whiskers and dots '●' indicate outliers; and 'x' indicates the mean score.

¹ <https://www.cbica.upenn.edu/BraTS18/lboardValidation.html>

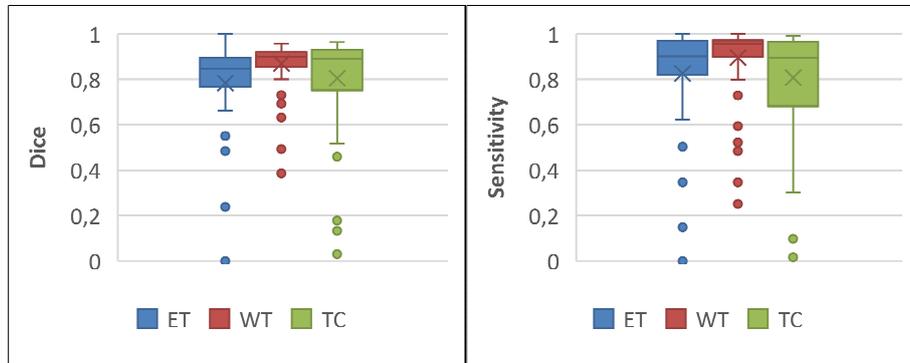


Fig. 4. Dispersion of *Dice* and *Sensitivity* scores for results segmentation of ET, WT, and TC in multimodal MRI scans of the 66 subjects of BraTS'2018 validation dataset.

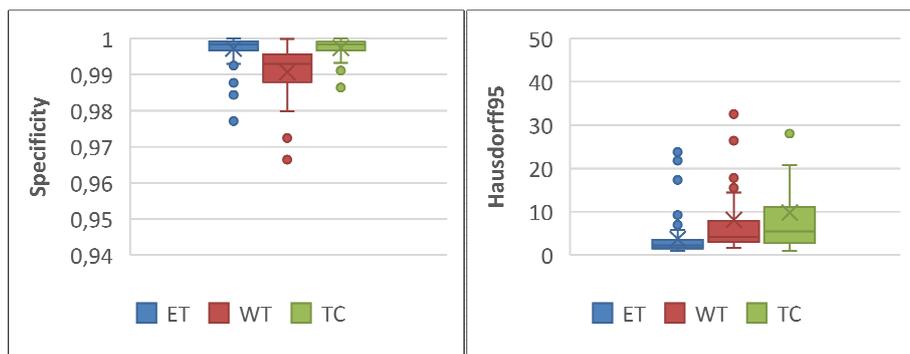


Fig. 5. Dispersion of *Specificity* and *Hausdorff* distance scores for results segmentation of ET, WT, and TC in multimodal MRI scans of the 66 subjects of BraTS'2018 validation dataset.

4 Conclusion and future work

In this paper, we have proposed a fully automatic and accurate method for segmentation of whole brain tumor and intra-tumoral regions using a 2D deep convolutional network based on a well-known architecture in medical imaging called "U-net". Our DNN model was trained to segment both HGG and LGG volumes.

The proposed method was tested and evaluated quantitatively on both BraTS'2018 training and challenge validation datasets. The total learning computation time of the 285 multimodal MRI volumes of BraTS'2018 training dataset is 185 hours on a Cluster machine with Intel Xeon E5-2650 CPU @ 2.00 GHz (64 Gb) and NVIDIA Quadro 4000 – 448 Core CUDA (2 GB) GPU. The average time to segment brain tumor and its components from a given MRI volume is about 62 seconds on the same GPU. The different tests showed that the segmentation results were very satisfactory, and the

evaluation measures confirm that our results are very similar to those obtained by the experts (i.e., growth truth), although the proposed method can be further improved.

As future work, we will use a more powerful GPU to further accelerate learning phase of DNN. Thus, we will be able to exploit more architectural variants of CNN and other data augmentation methods. Also, other interesting perspective consists to use ensemble learning methods, like Stacking and Blending, to improve segmentation performance in tumor core and active tumor regions. Another future work will focus on refining the segmentation results by reducing the false-positive rate using post-processing techniques, such as: applying a conditional random field (CRF) and removing small connected components.

Acknowledgments. This work was in part financially supported by an Algerian research project (CNEPRU) funded by the Ministry of Higher Education and Scientific Research (Project title and number: “PERFORM”, B*04120140014). We would like to thank Bakas, S., Ph.D and Postdoctoral researcher at SBIA of Perelman School of Medicine University of Pennsylvania – USA for providing the entire BraTS’2018 datasets employed in this study. We also gratefully acknowledge the support of CERIST, the Algerian Research Center for Scientific and Technical Information, for allowing us the use of “IBNBADIS” Cluster, without which we could not have performed these tests.

References

1. Dass, R., Priyanka, Devi, S.: Image Segmentation Techniques. International Journal of electronics & communication technology (IJECT) 3, 66-70 (2012)
2. Kermi, A., Andjough, K., Zidane, F.: A Fully Automated Brain Tumor Segmentation System in 3D-MRI using Symmetry Analysis of Brain and Level-Sets. IET Image Processing (2018)
3. Khotanlou, H., Colliot, O., Atif, J., Bloch, I.: 3D brain tumor segmentation in MRI using fuzzy classification, symmetry analysis and spatially constrained deformable models. Fuzzy Sets and Systems 160, 1457-1473 (2009)
4. Marcel Prastawaa, E.B., Guido Geriga: Simulation of Brain Tumors in MR Images for Evaluation of Segmentation Efficacy. Med Image Anal 13, 297–311 (2009)
5. Sharma, N., Aggarwal, L.M.: Automated medical image segmentation techniques. Journal of medical physics 35, 3–14 (2010)
6. Despotovic, I., Goossens, B., Philips, W.: MRI Segmentation of the Human Brain: Challenges, Methods, and Applications. Computational and Mathematical Methods in Medicine 2015, 23 (2015)
7. Wirtz, C.R., Albert, F.K., Schwaderer, M., Heuer, C., Staubert, A., Tronnier, V.M., Knauth, M., Kunze, S.: The benefit of neuronavigation for neurosurgery analyzed by its impact on glioblastoma surgery. Neurological Research 22, 354-360 (2000)
8. Yanyun, L., Zhijian, S.: Automated brain tumor segmentation in magnetic resonance imaging based on sliding-window technique and symmetry analysis. Chin Med Journal 127, 462-468 (2014)
9. Işın, A., Direkçođlu, C., Şah, M.: Review of MRI-based Brain Tumor Image Segmentation Using Deep Learning Methods. Procedia Computer Science 102, 317-324 (2016)
10. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific data 4, 170117 (2017)

11. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. *The Cancer Imaging Archive* (2017)
12. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. *The Cancer Imaging Archive* (2017)
13. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharrudin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans Med Imaging* 34, 1993-2024 (2015)
14. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., Larochelle, H.: Brain tumor segmentation with Deep Neural Networks. *Medical Image Analysis* 35, 18-31 (2017)
15. Goodfellow, I., Bengio, Y., Courville, A.: *Deep learning*. MIT press Cambridge (2016)
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234-241. Springer International Publishing, (Year)
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778 (2016)
18. He, K., Zhang, X., Ren, S., Sun, J.: Identity Mappings in Deep Residual Networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *Computer Vision – ECCV 2016*, pp. 630-645. Springer International Publishing, Cham (2016)
19. He, K., Zhang, X., Ren, S., Sun, J.: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1026-1034. IEEE Computer Society (2015)
20. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 Fourth International Conference on 3D Vision (3DV), pp. 565-571 (2016)
21. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In: Cardoso, M.J., Arbel, T., Carneiro, G., Syeda-Mahmood, T., Tavares, J.M.R.S., Moradi, M., Bradley, A., Greenspan, H., Papa, J.P., Madabhushi, A., Nascimento, J.C., Cardoso, J.S., Belagiannis, V., Lu, Z. (eds.) *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 240-248. Springer International Publishing, Cham (2017)
22. Castillo, L.S., Daza, L.A., Rivera, L.C., Arbeláez, P.: Volumetric Multimodality Neural Network For Brain Tumor Segmentation. *Proceedings of the 6th MICCAI BraTS Challenge* pp. 34-41, Quebec, Canada (2017)

3-D Densely Connected Fully Convolutional Neural Network for Brain Tumor Segmentation and Survival Analysis

Avinash Kori¹[0000-0002-5878-3584], Mehul Soni^[0000-0002-2968-6267], Pranjal B^[0000-0002-5195-4627], Mahendra Khened^[0000-0001-0841-6628], Varghese Alex^[0000-0001-5095-2358], and Ganapathy Krishnamurthi^[0000-0002-9262-7569]

Indian Institute of Technology Madras, Chennai 600036, India
gankrish@iitm.ac.in

Abstract. In this manuscript we utilize a 3-D fully convolutional neural network (CNN) for segmentation of the brain tumor and its constituents from multi modal Magnetic Resonance Images (MRI). The dense connectivity pattern used in the segmentation network enables effective reuse of features with lesser number of network parameters. From the segmentation, 22 features based on shape and intra-tumor volume fraction were extracted and fed to a multi-layered perceptron (MLP) based regressor for the task of survival prediction. On the BraTS validation data, the segmentation network achieved a whole tumor, tumor core and active tumor dice of 0.85 0.74, 0.71 respectively and the proposed survival prediction scheme achieved an accuracy of 53.6 %.

Keywords: Brain Tumor · MRI · CNN · 3-D · Survival prediction · Regressor.

1 Introduction

Manual tracing and detection of organs and tumor structure from medical images is considered as one of the preliminary step in diseases diagnosis and treatment planning. In a clinical setup this time consuming process is carried out by radiologists, however this approach becomes in-feasible when number of patients increase. This necessitates the scope of research in automated segmentation methods.

Segmentation of the gliomas from MR volumes is the preliminary step for treatment planning. Diffused boundaries of the lesion and partial volume effects in the MR images makes segmentation of gliomas from MR volumes a challenging task. In the recent years convolutional neural networks have produced state of the art results for the task of segmentation of gliomas from MR images. Typically, medical images are volumetric, organs being imaged are 3-D entities and henceforth we exploit the nature of 3-D CNN based architectures for segmentation task. Survival analysis is time prediction problem, which aids in better patient care and treatment planning.

2 Materials and Methods

The flowchart of the proposed the technique is illustrated in Figure 1

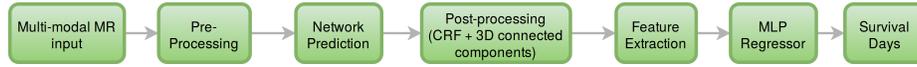


Fig. 1: Pipeline of architecture

2.1 Data

The training and validation data released as part of the BraTS 2018 challenge [4], [3], [1], [2] was used in this manuscript for segmentation and survival prediction tasks. The training dataset comprises of 210 high grade glioma volumes and 75 low grade gliomas. Each subject comprises of 4 MR sequences, namely FLAIR, T2, T1, T1 post contrast. Additionally on the training data, each subject was provided with pixel level segmentation. For the task of survival prediction, the training dataset included age, survival days and the resection status of 163 high grade glioma patients.

2.2 Data Pre-Processing and Patch Extraction

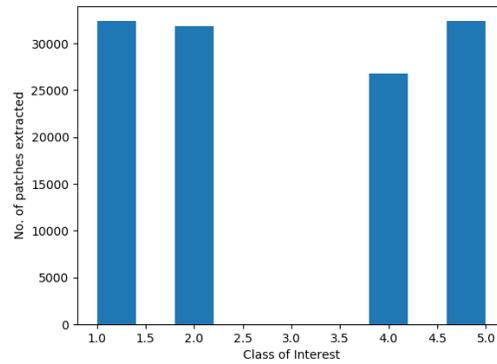


Fig. 2: Histogram of patches sampled surrounding certain class

As a part of pre-processing z-score normalization of individual MR sequence was done, wherein each sequence was subtracted by its the mean from the data

and divided by its standard deviation. Patches of size 64^3 were extracted from the brain. The class imbalance among the various classes in the data was circumvented by extracting relatively more number of patches extracted from lesser frequent cases such as necrosis (class 1) when compared to edema. Figure 2 illustrates the number of patches extracted around the pixel of each class.

2.3 Segmentation Network

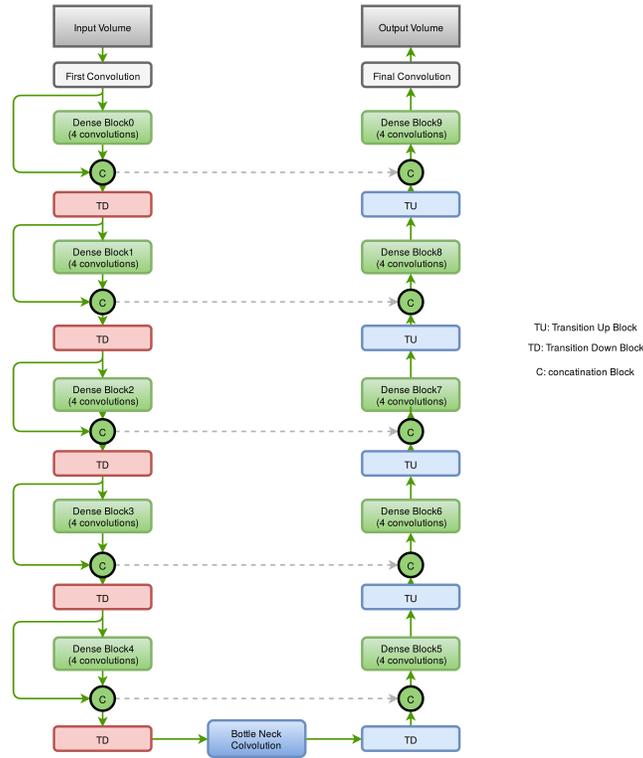


Fig. 3: Densely connected convolutional network used for segmentation task

The segmentation network is a 3-D fully convolutional semantic segmentation network. The network accepts an input of dimension 64^3 and predicts the class associated to all the voxels in the input. The network comprises of 57 layers. The dense connection between the various convolutional layers in the network aids in effective reuse of the features in the network. The presence of dense connections between layers lead to requiring more computational resources. This bottleneck is circumvented by keeping the number of convolutions to a small number say 4. Figure 3 shows the network architecture used in semantic segmentation task.

2.4 Training

Stratified sampling based on the grade of the gliomas was done so as to split the dataset into training, validation and testing in the ratio 70 :20 :10. The network was trained and validated on 182 and 63 respectively. To further address the issue of class imbalance in the network, the parameters of the network were trained by minimizing weighted cross entropy. The weight associated to each class was equivalent to the ratio of median of the class frequency to the frequency of the class of interest. The number of batches was set at 4, while the learning rate was initialized to 0.0001 and decayed by a factor of 10 % every-time the validation loss plateaued.

2.5 Testing

During inference, patches of dimension of 64^3 were extracted from the volume and fed to the network with the stride of 32. CNN's being a deterministic technique is bound to generate predict presence of lesion in physiological impossible place.

2.6 Post-processing

The false positives generated by the network were reduced by performing conditional random fields (CRF), these false positives were further reduced by performing 3-D connected component analysis.

2.7 Survival Prediction Task

Extraction of features for survival prediction: From the segmentation maps 4 features were handcrafted based on relative volumes of the various tumor regions to the whole tumor, 1 feature corresponded to the volume of tumor to brain volume, 3 features were based on relative location of the tumor core within the brain. Additionally, 13 shape based radiomic features ([5]) were extracted. Age of the patients were also taken into account. The extracted features were further standardized. The Table 1 lists the features extracted for survival analysis which is further elaborated below.

- Features were based on tumor and brain volumes (5 features):
 - ED to lesion: this feature was created by calculating the ratio of the volume of the edema region in the tumor to the volume of the whole tumor.
 - ET to lesion: ET refers to enhancing tumor. This is the ratio of the volume of enhancing tumor region to the whole tumor volume.
 - ED + ET to lesion: ED + ET indicates the volume of the edema added to the volume of the enhancing tumor and the sum is divided by the whole tumor volume. Hence, this is the ratio of edema and enhancing tumor region's volume to the whole tumor.

- NCR_NET to lesion: NCR stands for necrotic and NET refers to the non enhancing tumor region. These two regions form the tumor core. Hence, this feature gives the ratio of volume of the tumor core to the volume of the whole tumor.
 - Lesion to brain: the ratio of the tumor (lesion) volume to volume of the brain gives a measure of the size of the tumor with respect to the brain.
- Features based on location of the tumor (3 features): X location, Y location, Z location: these three values corresponding to the three coordinate values are the ratios of the centroid of the tumor core to the whole tumor center.
- Radiomic features based on shape (13 features): This group of 13 features includes descriptors of the three-dimensional size and shape of the ROI (region of interest). These features are independent from the gray level intensity distribution in the ROI and are therefore only calculated on the non-derived image and mask.

Table 1: Lists of features used for survival analysis

Type	Features
Based on tumor and brain volume	ED to lesion
	ET to lesion
	ED + ET to lesion
	NCR_NET to lesion
	Lesion to brain
Location of tumor core	X location
	Y location
	Z location
Shape based radiomic features	Elongation
	Flatness
	Least Axis
	Major Axis
	Maximum 2D Diameter Column
	Maximum 2D Diameter Row
	Maximum 2D Diameter Slice
	Maximum 3D Diameter
	Minor Axis
	Sphericity
	Surface Area
	Surface Volume
	Volume

Regression Model: For prediction of survival days, a multi-layer perceptron (MLP) comprising of two hidden layers and each layer with 3 neurons was employed for regressing on the training data. The MLP was trained for 200 epochs using Adam optimizer with a learning rate of 0.0001. For evaluation purpose, the predicted survival days were grouped into three classes depending on

the survival range, namely - short (< 300 days), mid (300 - 450 days) and long (> 450).

3 Results

Performance of the segmentation network The performance of the network was computed by inferring on 2 different dataset namely; a held out test data (n=40) and the BraTS data (n=66) is tabulated in Tables 2 & 3 respectively. Comparing the performance of network on the held out test data and the validation data, apart from generating good segmentation the network generalizes well on the unseen data. Figure 4 illustrates the segmentation generated by a trained network. Figure 5 illustrates the effect post-processing on segmentation.

Table 2: Performance of the segmentation on the held out test data (n=40)

	Whole Tumor	Tumor Core	Active Tumor
Mean	0.88	0.78	0.72
Std	0.10	0.26	0.31
Median	0.92	0.92	0.86

Table 3: Performance of the segmentation on the validation data (n=66)

	Whole Tumor	Tumor Core	Active Tumor
Mean	0.85	0.74	0.71
Std	0.11	0.22	0.28
Median	0.88	0.82	0.83

3.1 Performance of the model for survival prediction

The performance of the regression model for survival prediction was evaluated on the training (n=59) and validation (n=28) dataset and is tabulated in Table 4.

Table 4: Results of survival prediction on the training and validation data

Dataset	Accuracy	MSE	medianSE	stdSE	SpearmanR
Training	0.508	126665.607	29302.5	274604.067	0.296
Validation	0.536	976760.929	244116.5	1697543.43	0.172

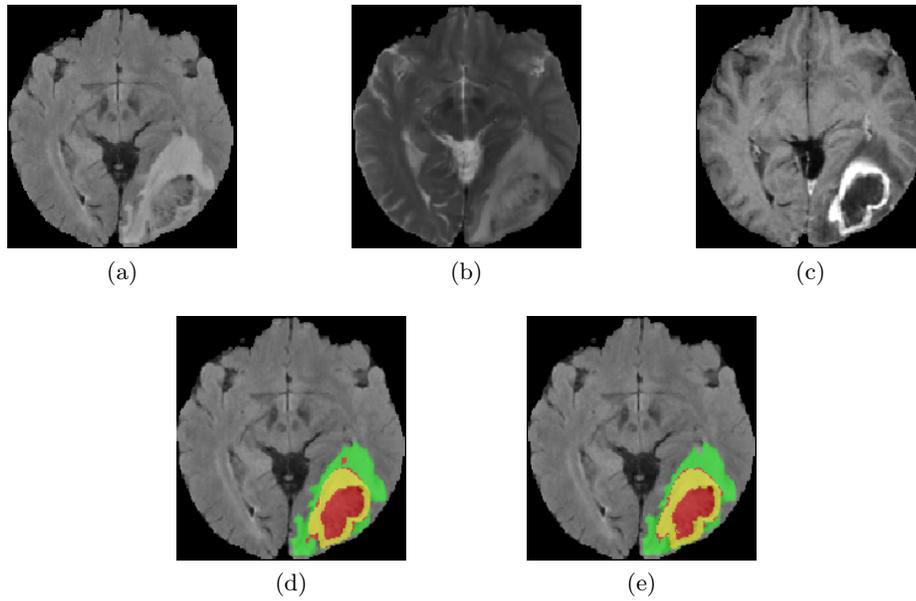


Fig. 4: (a) FLAIR, (b) T2, (c) T1c, (d) Prediction, (e) Segmentation. In images d and e, Green, Yellow & Red represent Edema, Enhancing Tumor and Necrosis present in the lesion.

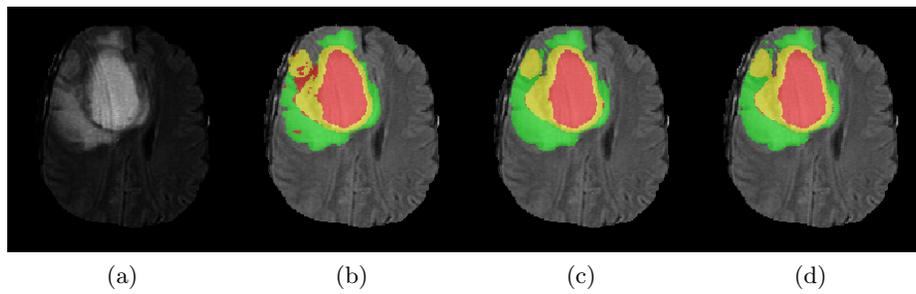


Fig. 5: (a) FLAIR, (b) Without Post-processing, (c) With Post-processing, (d) Segmentation. In images d and e, Green, Yellow & Red represent Edema, Enhancing Tumor and Necrosis present in the lesion.

4 Conclusion

In this manuscript we make use of 3-D fully convolutional semantic segmentation network brain tumor segmentation. The network accepts patches of size 64^3 as the input and predicts the class associated to all the voxels forming the input. The false positives generated by the network were minimized by using 3-D connected component analysis. On the BraTS 2018 validation data (n=66), the network achieved a competitive dice score of 0.85, 0.78 and 0.72 for whole tumor, tumor core and active tumor. From the segmentation maps features were extracted and a MLP based regression model was trained to predict the survival days. On the BraTS 2018 validation data (n=28), the proposed model achieved an accuracy score of 0.536.

References

1. S Bakas, H Akbari, A Sotiras, M Bilello, M Rozycki, J Kirby, J Freymann, K Farahani, and C Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive*, 286, 2017.
2. S Bakas, H Akbari, A Sotiras, M Bilello, M Rozycki, J Kirby, J Freymann, K Farahani, and C. Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *The Cancer Imaging Archive*, 2017.
3. Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4:170117, 2017.
4. Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993, 2015.
5. Joost JM van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina GH Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, and Hugo JW Aerts. Computational radiomics system to decode the radiographic phenotype. *Cancer research*, 77(21):e104–e107, 2017.

Glioma Segmentation with Cascaded Unet: Preliminary results

Dmitry Lachinov^{1,2}[0000-0002-2880-2887], Evgeny Vasiliev¹[0000-0002-7949-1919],
and Vadim Turlapov¹[0000-0001-8484-0565]

¹ Lobachevsky State University, Gagarina ave. 23, 603950 Nizhny Novgorod, Russian Federation

{dlachinov, vadim.turlapov}@gmail.com

² Intel, Nizhny Novgorod, Russian Federation dmitry.lachinov@intel.com

Abstract. Being powerful diagnostic tool, MRI analysis takes central position in brain tumor diagnostics and treatment. Thus it's precise evaluation is crucially important, however it requires time and is often performed using 2D projections that reduces analysis complexity but increases bias. On the other hand time consuming 3D evaluation like segmentation is able to provide precise estimation of a number of valuable spatial characteristics giving us understanding about the course of the disease.

Recent studies focusing at the segmentation task report superior performance of neural network based methods compared with classical computer vision algorithms, but still, it remains a challenging problem. In this paper we present two stage approach for automatic brain tumor segmentation. Alike recent methods in object detection our solution is based on neural networks; we employ localization network at the first stage of the pipeline and segmentation neural network at the second stage. In this paper we evaluate proposed method on BraTS 2018 validation dataset and report preliminary results.

Keywords: segmentation · BraTS · UNET · cascaded unet · deep learning

1 Introduction

Multimodal magnetic resonance imaging (MRI) is a powerful tool for studying human brain. Among it's different applications, it is mainly used for decease diagnostics and treatment planing. Accurate assessment of MRI results is critical throughout all these steps. Since MRI scans by themselves are the set of multiple three dimensional arrays, it's manual analysis and evaluation is non-trivial procedure and requires time, attention and qualification. However, lack of these resources can lead to unsatisfying results. Typically these scans are analyzed by clinical experts using two dimensional cut and projection planes. So it adds some bias to the resulting evaluation. On the other hand, accurate segmentation and 3D reconstruction is able to provide some insights on decease progression and

help a therapist to plan the treatment in the right way, but still it is not used due to unreasonable amount of time needed for manual labeling.

Denoting problem of automatic glioma segmentation Brain Tumor Segmentation (BRATS) challenge [1, 10] was created and became annual competition allowing participants to evaluate and compare their state of the art methods using unified framework. Participants are called to develop their algorithms and produce segmentation labels of the different glioma sub-regions: "enhancing tumor" (ET), the "tumor core" (TC) and the "whole tumor" (WT). The training data provided [2, 3] consists of 210 high grade and 75 low grade glioma MRIs manually labeled by experts in the field. Test dataset is split into two parts: validation data proposed methods can be evaluated with throughout the challenge and testing data. The performance of the methods is evaluated using Dice coefficient, Sensitivity, Specificity and Hausdorff distance.

Above-named challenge made a significant impact on the evolution of computational approaches for tumor segmentation. In the last few years variety of algorithms addressed to solve this problem were proposed. Compared with other methods convolutional neural networks have been showing the best state of the art performance for computer vision tasks in general and for biomedical image processing tasks in particular.

In this paper we present two-stage convolutional neural network based pipeline for brain tumor segmentation. At the first stage the tumor is localized, cropped and then segmented at the second stage. We compare multiple approaches for defining localization network and report their performance. Besides this, we compare 3d unet [5] with it's cascaded counterpart described later in this paper.

2 Method

In this study we propose two stage neural network for brain tumor segmentation. Figure 1 shows the workflow of the presented approach. This pipeline consists of two stages: tumor localization and segmentation of localized region. Before processing MRI scan is normalized, normalized data is fed to the localization network where it predicts tumor center and size. After the tumor is localized, it is cropped from the MRI and resampled to the predefined size. The segmentation part takes the resampled image as an input and returns corresponding annotation that is resampled to the original size and copied to the corresponding position inside MRI label at the final step.

2.1 Localization network

The purpose of localization network is to reduce amount of background in the segmentation input. This can be done by rejecting healthy voxels the model confident in. As a suitable description for possible tumor location we choose a bounding box. In our case we define it as a tumor size and center. Predicted

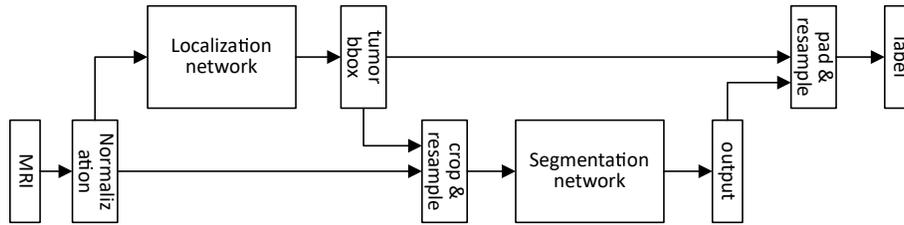


Fig. 1: Schematic representation of approach employed in this paper.

values should have high overlap ration with ground truth so the model doesn't reject the tumor voxels even before segmentation starts.

Generally speaking there are two possible way to predict a bound box: explicit estimation of bounding box center and size; or solution of another task like segmentation that can be used to extract information on bounding box. In this study we explore capabilities of both approaches using training dataset. Regression of the spatial characteristics is done using Resnet18 [6] model and Resnet10 model with Differentiable Spatial to Numeric Transform [11]. We use unet [5, 12] and proposed cascaded unet variant as a second type of algorithms that are able to implicitly estimate values described above.

Regression In order to predict the bounding box coordinates Resnet18 [6] was employed. The model was predicting tumor center and size - 6 values in total. Smooth l1 loss was used for training of this model.

DSNT As an alternative for heatmap regression we employed Diffirentiable Spatial to Numeric Transform layer [11]. We used this approach for tumor center prediction. For size prediction we employed two layer fully connected network on top of Resnet 18 [6]. Both networks were trained with smooth l1 loss.

U-net For the purpose of brain tumor localization we tried 3D variant of unet [5] described below, it's architecture is shown in figure 2. This network takes multimodal brain MRI as an input and produces binary segmentation. The tumor bounding box is constructed based on this output.

Cascaded U-net Cascaded unet was also applied for solving localization problem. It's detailed architecture is shown in figure 3.

2.2 Segmentation network

The segmentation part is intended to classify every voxel from given input into four classes: background, tumor core, peritumoral edema and enhancing tumor. The input volume comes from the previous step of the pipeline and consists of

tumor voxels and its surroundings. For this stage we've implemented 3d unet [5] with its cascaded counterpart. Both networks take $64 \times 64 \times 64$ volumes with localized tumor as an input.

U-net In this challenge we are using unet architecture extended to handle three dimensional input. The design of the network is illustrated in figure 2. It consists of encoding and decoding paths. On the encoding path number of filters is doubled after every downsampling step. Every convolution uses $3 \times 3 \times 3$ kernel and $1 \times 1 \times 1$ padding to keep the size of feature map the same. ReLU function is utilized as an activation function after every convolution. On the decoding path number of filters is reduced and feature map spacial size is increased by factor of two with each transpose convolution. Corresponding feature maps of encoding and decoding parts are linked using skip connections. As a final step, softmax layer is employed on the resulting feature map.

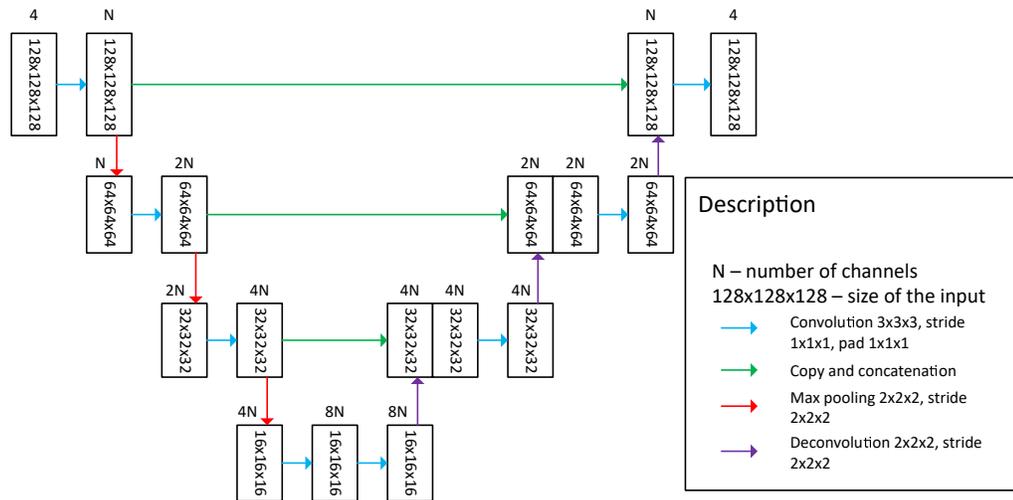


Fig. 2: Architecture of unet. ReLU is used as activation function after every convolution layer.

Cascaded U-net Proposed network is illustrated in figure 3 and consists of three basic blocks. Each block by itself is unet network [5, 12] with its own loss function in the end, the only difference is the feature maps size on the decoding path. Every next block takes downsampled volume as an input and produces segmentation of corresponding size. Alike DeepMedic [9] this architecture simultaneously processes the input image at multiple scales and extracts scale-specific features. The feature map before the last convolutional layer in every block is

concatenated to corresponding feature map of higher-scale block. It enables the context information flow between networks with different scales.

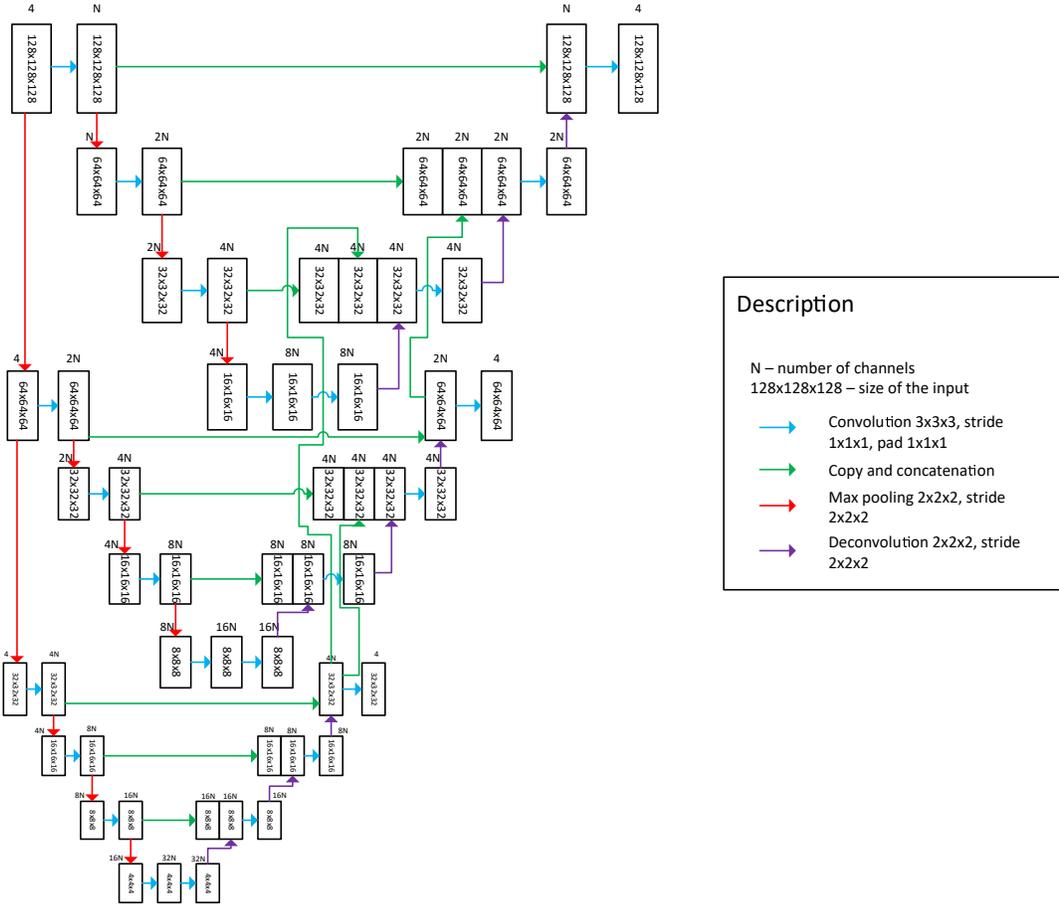


Fig. 3: Architecture of cascaded unet. ReLU is used as activation function after every convolution layer. Each block produces it’s own output.

2.3 Preprocessing & Data augmentation

We use data preprocessing described in [7]. First, we perform z-score normalization on non-zero (brain) voxels. After that we are eliminating outliers and noise by clamping all values to the range from -5 to 5. At the final step we shift brain voxels to the range [0;10] and assign zeros to background.

For data augmentation we artificially increase number of samples by employing

b-spline transformation to the original data. It has been done with ITK implementation [8].

As the last step localization and segmentation networks are trained together. Since the output of the localization network varies from one epoch to another, it serves as additional augmentation mechanism implicitly incorporated into the pipeline.

2.4 Training

The training has two steps. First, the localization network is trained so it can produce meaningful output. Output of bounding box is dilated for 10 voxels in every direction. Then both networks are trained together. For obtaining preliminary results we train these networks for 50 epochs and employ early stopping if the validation loss doesn't improve over time. Dice loss is used for localization network and multiclass dice loss for segmentation network. These CNNs were implemented in MXNet framework [4] and trained using two GTX 1080 with batch size 2 to enable data parallelism.

3 Preliminary results

In this section we report evaluation obtained with local validation first. In the scope of brain tumor segmentation bounding boxes built on segmentation output show better results. Here we are especially interested in intersection over union and overlap ratio since these two characteristics reflect number of rejected voxels that can possibly be tumor voxels. We've measured mean absolute error (MAE) for bounding box prediction as well as intersection over union (IoU) and overlap ration, results can be found in table 1.

Based on previously made evaluate we've built two convolution neural network

Table 1: Evaluation of localization networks performance using local validation

Method	MAE	IoU	Overlap ration
Resnet18	9.33	0.40	0.52
DSNT	7.03	0.54	0.66
unet	7.94	0.59	0.8
cascaded unet	4.61	0.75	0.81

pipelines based on unet and it's cascaded counterpart. These networks were evaluated using local validation. Results can be found in table 2.

Preliminary results for the BraTS 2018 validation dataset [2, 3] can be found in table 3. Method was evaluated on 66 MRI obtained with different scanners and from different institutions.

Table 2: Evaluation of glioma segmentation using local validation. Where dice_wt stands for dice score of whole tumor, et stands from enhancing tumor and tc stands for tumor core.

Method	dice wt	dice et	dice tc
unet	0.877	0.746	0.747
cascaded unet	0.893	0.756	0.768

Table 3: Evaluation of glioma segmentation using validation data provided by BraTS 2018 organizers. Where dice_wt stands for dice score of whole tumor, et stands from enhancing tumor and tc stands for tumor core.

Method	dice wt	dice et	dice tc
unet	0.882	0.754	0.754
cascaded unet	0.883	0.742	0.76

4 Conclusion

In this paper we presented automatic brain tumor segmentation algorithm. It consists of two stages: tumor localization and segmentation. We implemented and compared different localization methods and showed their performance for brain tumor localization. Moreover, we compared two segmentation networks as a part of presented pipeline: 3d unet and it's cascaded counterpart. Both methods achieved the same accuracy on validation data provided by organizers, however proposed method showed better performance on local validation.

Bibliography

- [1] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Sci Data* **4**, 170117 (Sep 2017). <https://doi.org/10.1038/sdata.2017.117>, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5685212/>, 28872634[pmid]
- [2] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive* (2017). <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>
- [3] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans

- of the tcga-egg collection. The Cancer Imaging Archive (2017). <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>
- [4] Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., Xiao, T., Xu, B., Zhang, C., Zhang, Z.: Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems. CoRR **abs/1512.01274** (2015), <http://arxiv.org/abs/1512.01274>
- [5] Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: Learning dense volumetric segmentation from sparse annotation. CoRR **abs/1606.06650** (2016), <http://arxiv.org/abs/1606.06650>
- [6] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CoRR **abs/1512.03385** (2015), <http://arxiv.org/abs/1512.03385>
- [7] Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: Contribution to the BRATS 2017 challenge. CoRR **abs/1802.10508** (2018), <http://arxiv.org/abs/1802.10508>
- [8] Johnson, H.J., McCormick, M., Ibáñez, L., Consortium, T.I.S.: The ITK Software Guide. Kitware, Inc., third edn. (2013), <http://www.itk.org/ItkSoftwareGuide.pdf>, *In press*
- [9] Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d CNN with fully connected CRF for accurate brain lesion segmentation. CoRR **abs/1603.05959** (2016), <http://arxiv.org/abs/1603.05959>
- [10] Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, ., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging **34**(10), 1993–2024 (Oct 2015). <https://doi.org/10.1109/TMI.2014.2377694>
- [11] Nibali, A., He, Z., Morgan, S., Prendergast, L.: Numerical coordinate regression with convolutional neural networks. CoRR **abs/1801.07372** (2018), <http://arxiv.org/abs/1801.07372>
- [12] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. CoRR **abs/1505.04597** (2015), <http://arxiv.org/abs/1505.04597>

Cascade of Random Forest Classifiers for Brain Tumor Segmentation ^{*}

Szidónia Lefkovits¹[0000-0002-7903-1111], László Szilágyi²[0000-0001-6722-2642],
and László Lefkovits²[0000-0002-4254-5880]

¹ Department of Computer Science
„Petru Maior” University, Tîrgu Mureş, Romania
`szidonia.lefkovits@science.upm.ro`

² Department of Electrical Engineering
Sapientia University of Transylvania, Tîrgu Mureş, Romania
`lefkolaci@ms.sapientia.ro`

Abstract. Brain tumor segmentation is a difficult task due to the strongly varying intensity and shape of gliomas. In this paper we propose a multi-stage discriminative framework for brain tumor segmentation based on BraTS 2018 dataset. The framework presented in this paper is an extended segmentation system of greater complexity based on our model described at BraTS 2016. This framework implements a multi-stage classifier based on the random forest (RF) algorithm. Our four-stage system attempts to follow the layered structure of tumor tissues provided in the annotation protocol. In each segmentation stage we dealt with four major difficulties: feature selection, determination of training database used, optimization of classifier performances and image post-processing. The framework was tested on the evaluation images from BraTS 2018. The results obtained are comparable to the best ones presented in previous BraTS Challenges.

Keywords: multi-stage classifier · random forest · feature selection · variable importance · MRI brain tumor segmentation

1 Introduction

Image processing is a powerful tool for computer-aided diagnosis especially in the medical field. The most important advantage of medical imaging is the fact that examination performed non-intrusively. MR imaging and diagnosis is increasingly being used for medical investigation. This article is restricted to MRI brain imaging and provides a framework for the automated brain tumor segmentation method proposed, delimiting different types of tumors from multi-modal MRI images. The automatic system based on machine learning overcomes the

^{*} The work of S. Lefkovits in this article was supported by the Communitas Foundation and the work of L. Lefkovits in this article was supported by Sapientia Foundation – Institute for Scientific Research (KPI).

laborious, lengthy work of segmentation done manually by experts. It is replicable and much faster than the segmentation performed by experts, which might be fairly different. The most important advantage of such a system is that it can lend assistance in determining the correct diagnosis, surgery or treatment plan and monitor the evolution of the disease.

The framework presented in this paper is an extended segmentation system of greater complexity based on our model presented at MICCAI-BraTS 2016 [14]. This model was built on a feature extraction algorithm [13] and single-staged random forest (RF) classifiers with optimized parameters. The random forest approach was used in few systems presented at BraTS 2017 [9, 18, 16, 17]. The segmentation results obtained showed that the tumor region is well detected, but the contours of the whole tumor and the interior tumor tissues are not well delimited. The source of the aforementioned errors could be the choice of training samples used, the unbalanced database provided, and its enormous size. These three factors cannot be counteracted by a single-stage RF classifier. Another deficiency in our previous model is that it considered almost any spatial relationship between the tumor tissues, according to the annotation protocol described in [11, 15].

In the current work we propose a multi-stage classifier based on the random forest algorithm. In our current experiments we attempt to circumvent the deficiencies of our old framework and improve the segmentation results.

The proposed framework is built around the model given in Fig. 1.

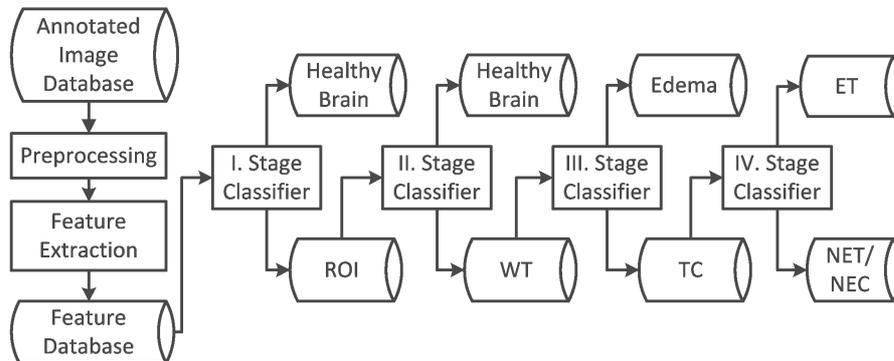


Fig. 1. Discriminative Model Proposed for Segmentation

The I. stage-classifier detects the tumorous zone from the entire 3D MRI image. This phase is tuned to have an extremely good Positive Predictive Value (PPV) detection index. It considers the tumor zone to be the goal of detection, and therefore this is the positive segmentation zone. Thus, it is able to delimitate the image ROI containing the tumor with a high probability, of approximately 0.99.

In the II. stage the WT (whole tumor)-classifier delimitates the whole tumor from healthy brain tissue. Afterwards the TC (tumor core)-classifier delimits the tumor from edema (III. stage). Finally, the last ET (enhanced tumor)-classifier (IV. stage) is able to segment the tissues inside the TC. The use of binary classifiers for all these classification decisions follows from the annotation protocol. It states that "the various tissue elements (edema, non-enhancing, enhancing, necrosis) usually follow an outside-inside sequence therefore one should start from the outside and delineate regions within the previous labels. Due to this «Mozart kugel» appearance it is enough to always delimitate what is outside and internal border should not be delimited" [11].

2 Method

The delimitation of the brain tumor from the healthy tissues can be achieved by a voxel-wise segmentation. To solve this task we propose a multi-stage discriminative model based mainly on the random forest algorithm and its facilities. Voxel-wise segmentation starts with the construction of the feature database obtained from the annotated image database. The feature database generation process is identical both for the segmentation (classification) and the training phases, as well. It consists of the following steps: preprocessing, local feature definition and extraction (Fig. 1 and 2).

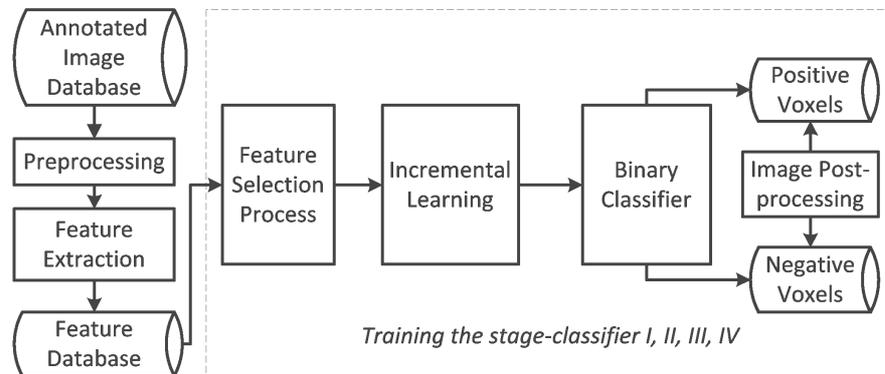


Fig. 2. Discriminative Model Proposed for Training each Stage

The database used in our segmentation made up of was the training and validation databases created for the BraTS 2018 Challenge [4]. The training set consists of 75 low-grade and 210 high-grade MRI brain images. The image data consists of 4 modalities T1, T1c, T2 and FLAIR, acquired from 19 different MRI scanners using different protocols [4, 6]. All the images had been segmented

manually by several experts, and the average annotation is in fact the ground truth given in the database. The modalities are co-registered, interpolated to the same resolution and skull-stripped. The annotated regions [7, 8] are labeled in 4 different classes: 0 for background and healthy tissue, 1 for NCR/NET (necrotic and/or non-enhancing tumor), label 2 for ED (the edema) and label 4 for ET (the enhancing tumor).

During preprocessing we handled three important artifacts: inhomogeneity correction, noise filtering and intensity standardization. For inhomogeneity reduction in MR images, we applied the N4 filter implemented in the ITK package [1]. The anisotropic filtering from the same package was used for noise reduction. Intensity normalization was done by histogram linear transformation in such a way that the first and third quartiles had predefined values.

In voxel-wise segmentation it is necessary to define a set of intensity- and local neighboring features. The following features were extracted: first order operators (mean, standard deviation, max, min, median, Sobel, gradient); higher order operators (laplacian, difference of gaussian, entropy, curvatures, kurtosis, skewness); texture features (Gabor filter); spatial context features (symmetry, projection, neighborhoods), – the same as in our previous work.

The segmentation workflow given in Fig. 1 requires four binary classifiers. Each classifier is trained and evaluated on its own feature database during its training process (Fig. 2). The global training consists of four training stages and each stage is composed of the following four steps:

1. feature selection based on variable importance [12] provided by the random forest;
2. incremental training of the RF stage-classifier;
3. optimization of the classification performance according to the task of the given tumor tissue segmentation;
4. image post-processing, with the role of reducing false detections and implementing the layered structure of tumor tissues.

The first step (1.), feature selection based on the variable importance provided by the RF algorithm, and the third step (3.), the performance optimization of the random forest classifier, were presented in our previous articles [13, 14]. These approaches were used to create our one-stage segmentation system presented at the previous BraTS Challenge in 2016 [14]. In our current work we use these algorithms in each of the four stages.

In the first step we defined 960 different features for each voxel. The RF classification algorithm is not able to deal with all the input image voxels and all 960 features previously defined, due to hardware and software limits. Therefore, this large amount of data was handled by taking advantage of the random forest variable importance evaluation. Our idea was to implement an iterative feature selection algorithm presented in [13]. The main idea of the algorithm is to evaluate the variable importance several times on a randomly chosen part of the feature database (Fig. 3). If the OOB error of the forest ensemble was below a certain threshold then the variable importance was taken into consideration and

cumulated. Averaging the variable importances in the iterations the algorithm was able to eliminate the most unimportant 20-40% of variables in each run.

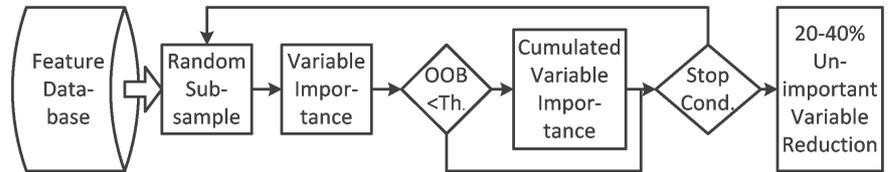


Fig. 3. Feature Selection Algorithm

In random forest approaches the training set is usually created out of the existing annotated images by random subsampling. In the case of BraTS2018 the annotated image set contains 285 MR images and each image is made up of about 1,500,000 voxels, which means about 450 million samples. This huge database is, in addition, extremely unbalanced. In consequence we must obtain a well-defined database for training our random forest classifier. The solution to this is the incremental learning procedure that consists of enlarging the current training set by incrementally adding incorrectly classified random subsamples. In the second step (2.), this incremental learning is repeated several times until the classification performances are adequate or the upper limit of hardware and/or software is reached. The flowchart of the incremental learning is given in Fig. 4.

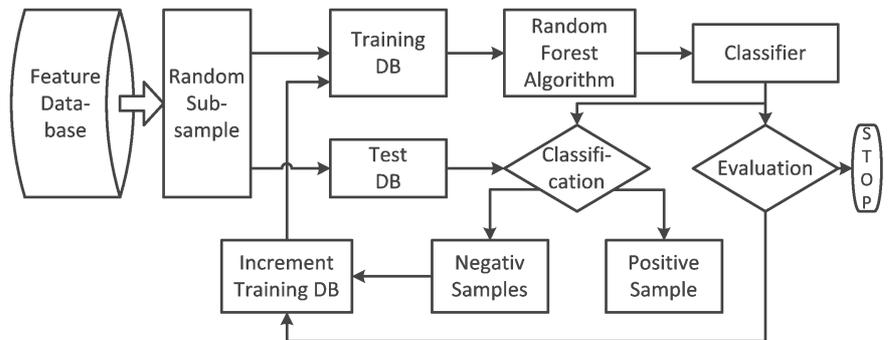


Fig. 4. Incremental Learning

The classifier performance optimization (step 3.) is in strong correlation with the segmentation task set. This assumes the correct choice of training param-

eters. The random forest classification performance can be tuned via three important parameters: m_{tires} – the number of randomly chosen features used as a splitting criterion in each node of the trees; the n_{trees} – the number of trees in the forest; n_{nodes} – the maximum number of nodes in each tree. These parameters determine the size of the random forest ensemble, the segmentation performances, training time and system complexity, as well. In our experiments [14] these durations can be drastically reduced without any loss in segmentation accuracy.

The last step (4.), after the training of each stage-classifier (Fig. 2), is an image post-processing step to do with the goal of the current stage. Here we managed to incorporate some knowledge about the tumor, such as the number of distinct tumors in a brain, one tumor is a connected zone within the healthy brain tissue, the tumor core is inside the edema, the enhanced tumor is a connected zone inside the whole tumor, etc. By applying this post-processing step we succeeded to eliminate the most of the false detections and improve the quality of segmentation.

3 Preliminary Results

The proposed discriminative model is quite laborious and the four proposed classifiers have to be tuned separately (Fig. 1). In this phase 70 HGG images are used for training the RF classifier and remaining of 140 HGG images are used for tuning and testing segmentation performances.

The I. stage-classifier determines the ROI (region of interest) that contains the tumor region with high probability. This binary classifier was trained on the whole brain in order to delimit the healthy region from the tumoral region. The tumor zone was determined by this first classifier obtaining a PPV (Positive Predictive Value) index of around 90%. So as to improve these values a region dilation of 3 voxels (with a ball-structuring element) was applied, taking into consideration only the two most important connected regions. In this way the ROI obtained is about twice as large as the whole tumor, but the PPV index reached on average 0.99 on the 70 training images and 0.97 on the remaining 140 images used for testing and evaluation. The correct determination of this ROI has a crucial role in the subsequent stages. Table 1 shows the average

Table 1. First Stage Classifier Results

		Training-70 MRI		Testing-140 MRI	
		PPV	Sensibility	PPV	Sensibility
Binary classif.	Mean	0.975	0.724	0.913	0.656
	StdDev	0.082	0.164	0.146	0.185
Post-processing	Mean	0.996	0.518	0.981	0.531
	StdDev	0.012	0.154	0.041	0.163

and standard deviation of the PPV index after the binary segmentation and

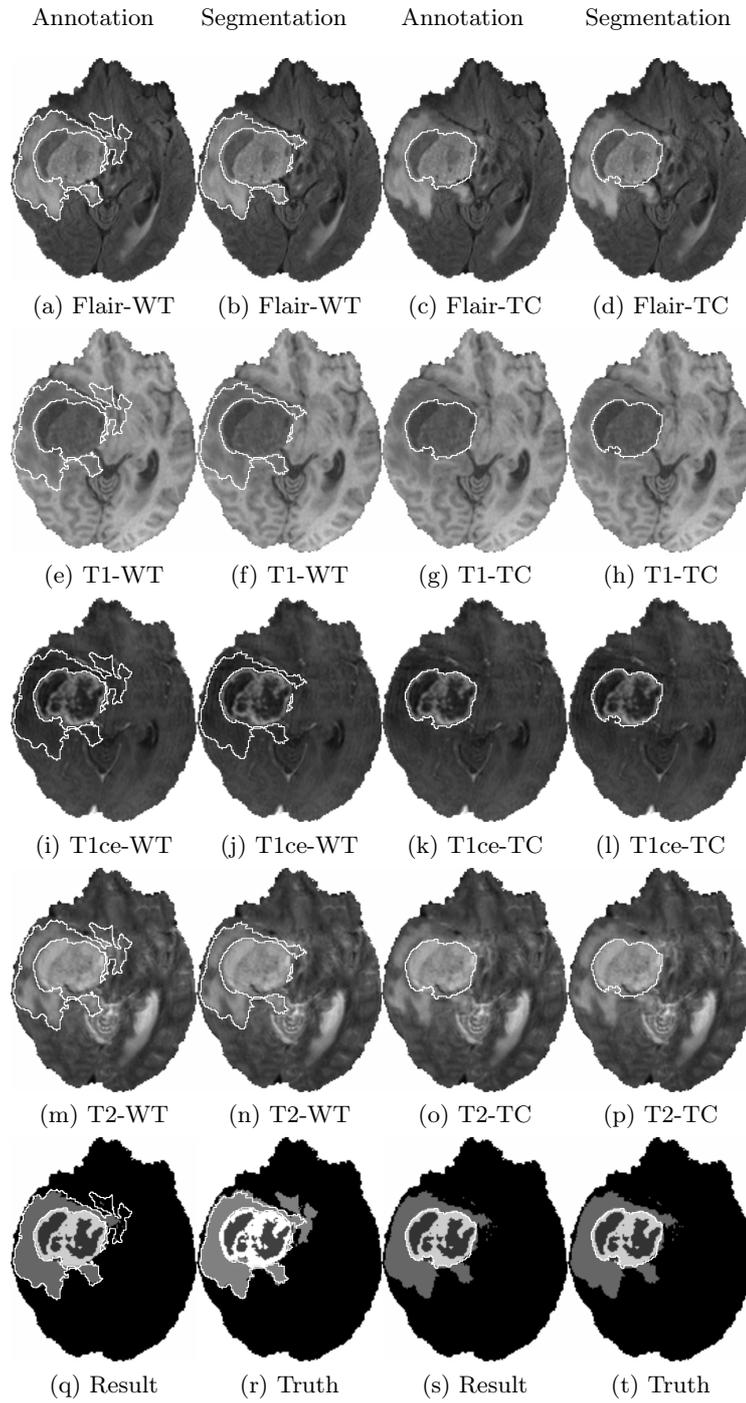


Fig. 5. Segmentation Example

post-processing in the I. stage. The ROI obtained reduces the image region and, implicitly, the feature database by about 8 times. This allows us to create a more precise classifier in the next stages.

The II. stage-classifier is applied only on the ROI. Its segmentation task is to delimit the remainder of healthy tissue from the WT. In this stage the segmentation with post-processing creates two disjunct regions, considering the tumor zone a connected region inside the healthy tissues.

The III. stage-classifier is applied only inside the WT region and its task is to delimit the edema region from the TC region. In post-processing we consider that the edema forms a connected region that frames the TC and is the margin of the WT. By obtaining the edema from the WT, the TC region is found at the same time.

The IV. stage-classifier is applied only inside the TC region (obtained in the previous stage) in order to segment the ET (enhanced tumor) and the NET (non-enhanced tumor)/NEC (necrotic tumor).

The results obtained after the four stages were uploaded to the leaderboard [5] are presented in Tables 2, 3 and Fig. 5.

Table 2. Segmentation Results for the Training Database

Label	Dice			Sensitivity			Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Mean	0.819	0.887	0.823	0.842	0.852	0.799	0.997	0.996	0.998	5.457	5.722	8.833
StdDev	0.116	0.092	0.155	0.146	0.140	0.196	0.004	0.004	0.003	7.924	6.411	8.705
Median	0.845	0.921	0.879	0.892	0.899	0.878	0.998	0.998	0.999	2.236	3.535	7.071
25quantile	0.779	0.864	0.779	0.777	0.809	0.734	0.997	0.996	0.997	1.414	2.000	3.162
75quantile	0.898	0.941	0.924	0.941	0.946	0.934	0.999	0.999	1.000	5.171	7.071	10.724

Table 3. Segmentation Results for the Validation Database

Label	Dice			Sensitivity			Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Mean	0.719	0.873	0.689	0.764	0.852	0.656	0.998	0.995	0.998	7.304	7.068	12.662
StdDev	0.273	0.109	0.281	0.252	0.154	0.288	0.003	0.005	0.004	11.080	11.309	13.429
Median	0.819	0.901	0.797	0.850	0.913	0.738	0.999	0.997	0.999	3.158	4.000	8.305
25quartile	0.724	0.851	0.647	0.708	0.775	0.552	0.998	0.994	0.998	2.000	2.449	6.335
75quartile	0.886	0.933	0.892	0.928	0.954	0.856	0.999	0.998	0.999	6.708	6.083	12.400

4 Conclusion

In this paper we developed a four-stage discriminative model for brain tumor segmentation based on multi-modal MRI data. Our four-stage model attempts to

implement the layered tissue structure by adequate training of binary classifier and image post-processing in each segmentation stage. In each stage we solved the four important issues concerning discriminative models. Our results show that binary classifiers are very efficient for the layered segmentation task. One of the most important results is the determination of a ROI that has to enclose the whole tumor with a very high probability. In the first segmentation stage, a PPV of about 0.97 has been reached by decreasing the sensibility to 0.6. This ROI reduces the size of the feature database by about 8 times and provides a reliable ROI for the next segmentation stages. The preliminary results we presented were obtained by the four-stage segmentation framework we developed, trained on 70 HGG and tested on the rest of 140 HGG image sets. In the future we propose training our system with all HGG and LGG images provided for training in order to reach similar performances on the BraTS2018 validation dataset. The system developed is a complex implementation using a large variety of software packages and modules such as ITK in C++ [1], Java, ImageJ and Fiji with Trainable Weka Segmentation [2], the random forest package from R [3], Matlab for evaluation and conversion into the desired image type.

References

1. ITK. <http://www.itk.org/>
2. Trainable segmentation. http://imagej.net/Scripting_the_Trainable_Segmentation
3. The R project for statistical computing. <https://www.r-project.org/>
4. Multimodal Brain Tumor Segmentation Challenge 2018. <https://www.med.upenn.edu/sbia/brats2018/>
5. Multimodal Brain Tumor Segmentation Challenge Evaluation Website. <https://www.cbica.upenn.edu/BraTS18/>
6. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., S., Freymann, J., B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features, *Scientific data*, Nature Publishing Group, 4:170117, (2017)
7. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., S., Freymann, J., B., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection, *The Cancer Imaging Archive*, DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q,(2017)
8. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., S., Freymann, J., B., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection, *The Cancer Imaging Archive*, DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF,(2017)
9. Bharath, H.N., Coleman, S., Sima D.M., Huffel Van S.: Tumor segmentation from multi-parametric MRI using random forest with superpixel and tensor based feature extraction. In *Proceedings of MICCAI-BRATS Challenge*, (2017)
10. Breiman, L.: Random forests. *Machine learning* 45(1), 5–32 (2001)
11. Jakab, A.: Segmenting Brain Tumors with the Slicer 3D Software Manual for providing expert segmentations for the BRATS Tumor Segmentation Challenge. University of Debrecen / ETH Zürich

12. Goldstein, B. A., Polley, E. C., Briggs, F.: Random forests for genetic association studies. In: Statistical applications in genetics and molecular biology, 10(1),1544–6115, (2011).
13. Lefkovits, L., Lefkovits, S., Emerich, S., Vaida M.F.: Random forest feature selection approach for image segmentation. In: Proceedings of the 9th International Conference on Machine Vision (ICMV), 10341:1034117, (2016)
14. Lefkovits, L., Lefkovits, S., Szilágyi, L.: Brain tumor segmentation with optimized random forest. In Proceedings of MICCAI-BRATS Challenge, 80-99, (2016).
15. Menze, B.H., Jakab, A., Bauer, S. et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Tr.on Medical Imaging 34(10),1993–2024, 2015
16. Phophalia, A., Maji, P.: Multimodal Brain Tumor Segmentation Using Ensemble of Forest Method, In Proceedings of MICCAI-BRATS, (2017)
17. Revanuru, K., Shah N.: Fully Automatic Brain Tumour Segmentation using Random Forests and Patient survival prediction using XGBoost. In Proceedings of MICCAI-BRATS, (2017)
18. Soltaninejad, M., Zhang, L., Lambrou, T., Yang, G., Allinson, N., Ye, X.: MRI Brain Tumor Segmentation using Random Forests and Fully Convolutional Networks. In Proceedings of MICCAI-BRATS, (2017)

Fused U-Net for Brain Tumor Segmentation based on Multimodal MR Images

Xiaochuan Li

Tianjin University, Tianjin, China

Abstract. Multi-modality MR images are important diagnosis tools for brain tumor evaluation and treatment. However, it is a non-trivial task to segment the whole tumor from MR images due to the complicated tumor micro-environments and different gradings of brain tumors. In this paper, we have proposed a multi-stage segmentation method for accurate extraction of the whole tumor, enhancing tumor and tumor core based on a cascaded modified U-Net. The modified U-Net of each stage is implemented as three separate 2D modified U-net for the axial, sagittal and coronal views, followed by fusion operation to include tumor features from all views. The network of the subsequent stages will use results from previous stages as an additional input. On the BraTS Challenge 2018, our method has achieved an average dice score of 0.895/0.810/0.863 over training dataset and 0.883/0.720/0.788 over validation dataset for whole tumor, enhancing tumor and tumor core, separately.

Keywords: Brain Tumor, Fully Convolutional Neural Network, MRI.

1 Introduction

Gliomas, including glioblastoma and lower grade glioma, are the most common brain malignancies with different degrees of aggressiveness and variable prognosis. With different histological structure, gliomas can be divided into various heterogeneous sub-regions such as edema, necrotic core, enhancing and non-enhancing tumor core. Accurate segmentation of the tumor sub-region is important for the precise diagnosis and treatment, such as radiotherapy[1, 2, 3]. Meanwhile, segmenting tumor structure from MR images is complicated due to the heterogeneous tumor micro-environments, and different appearance of various gradings of the brain tumors.

In this paper, we propose a three-stage U-Net[4, 5, 6] for the task of segmentation of brain tumors, including whole tumor, tumor core and enhancing tumor. U-Net has been successfully applied to many biomedical image tasks, and showed good performance on semantic segmentations[7, 8, 9]. We designed a three-stage segmentation system using both 2D and 3D multimodal U-Nets based on MRI scans for automated segmentation of brain tumor and intra-tumor sub-regions, including edema, enhancing tumor and necrotic/non-enhancing tumor.

2 Methods

2.1 Data

The BraTS 2018 Challenge has provided 210 glioblastoma (GBM/HGG) and 75 (lower grade glioma) LGG brain MR cases [10, 11, 12, 13]. Each case is composed of four MR sequences, including FLAIR, T1, T1CE and T2. Meanwhile, each case is also given a ground truth annotation image, consisting of GD-enhancing tumor (ET, label 4), peritumoral edema (ED, label 2), necrotic and non-enhancing tumor (CNR/NET, label 1). The data is co-registered to the same anatomical template and interpolated to the sample resolution. For the prediction of patient overall survival, the data is defined in days, storing in the csv file corresponding to each case name. The validation set consists of 66 cases with no distinction between HGG or LGG.

2.2 Data Preprocessing

In our approach, before feeding the data to the deep learning network, each case is normalized by its own mean and standard deviation for both training and validating phases. In training, we have added data augmentations including horizontal flip, vertical flip, scale and rotate, randomly. Also, we have added the shear and elastic transformation. Gaussian noise is also added in all input MR modalities. Those data augmentation procedures have greatly increase our model generalization, and effectively avoid overfitting of the training model.

2.3 Multimodality 2D Slice Neural Network

In our approach, the whole process pipeline is described as Fig.1. We have adopted a three-stage approach to extract the tumor structures, which consists of whole tumor extraction, tumor core extraction and enhancing tumor extraction.

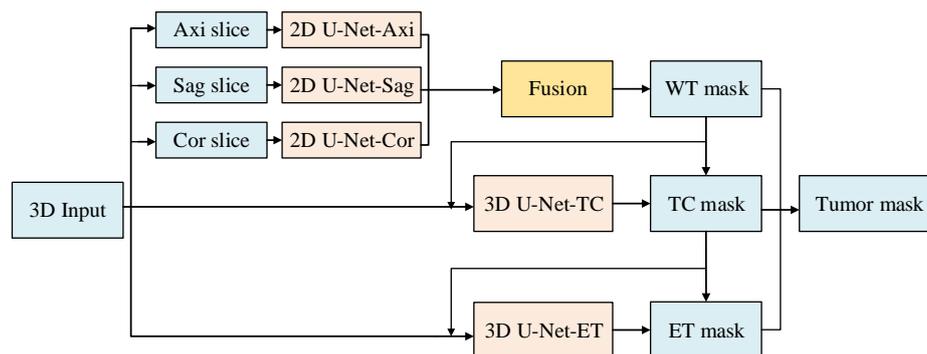


Figure. 1 The pipeline of the proposed method

In whole tumor extraction, we have applied three standard U-Net structures (U-Net-Axi, U-Net-Sag, U-Net-Cor), with each corresponds to one view direction of the

original MR images. The U-Net is implemented with feature numbers of [64, 128, 256, 512, 1024], separately. During downsampling and upsampling of each layer, we adopt the kernel size of 3 in convolutional layer and kernel size of 2 in max pooling layer. The output of the network has 2 channels, each represents the background and foreground (whole tumor). During training and validating stage, each U-Net for each view is trained and validated separately, using the log-softmax of output channels with the NLL-Loss. While in testing stage, testing cases will be feed forwarded through each U-Net-Axi, U-Net-Sag and U-Net-Cor, and the results of each view are ensemble with equal weights, followed by a post-processing operation. During post-processing, the final binary labels are fused as the maximum indexes of log-softmax values from the 2 output channels.

2.4 Multimodality 3D U-Nets for tumor core segmentation

Segmentation of tumor core regions was performed by a 3D U-Net. The results of the whole tumor segmented is used as an additional input, together with the 4 input MR channels of T1, T1CE, T2 and flair to generate a 5-channel concatenated input. The 3D U-Net is implemented with feature numbers of [16, 32, 64, 128]. Besides, the kernel size was set as 3 in convolutional layers and as 2 in pooling layers, the same as in the 2D U-Net. The output of the net is a 3-channel tensor, each channel refers to the probability map of background, tumor core and edema, respectively.

Since the region of brain tumor normally correspond to 20% or less of the overall brain tissue, dice coefficient[14, 15, 16] is used as a loss function to train the net, which can be written as:

$$D = \frac{\sum_{i \in V} p_i \cdot l_i}{\sum_{i \in V} p_i + \sum_{i \in V} l_i} \quad (1)$$

where V is the set of voxels in the 3D image, p_i is the softmax value of the voxel. Besides, we found that the predicted mask of tumor core is usually larger than the ground-truth, which causes the decline of segmentation accuracy. To solve the problem, we design a united dice loss function which also include dice score of edema to revise boundary:

$$D = D_{TC} + \alpha \cdot D_{ED} \quad (2)$$

where α is the coefficient of the dice loss of edema region. During training, we set $\alpha = 0.3$.

In tumor core segmentation, the input is a 5-channel 128*128*128 cubic patch in training, which is in the center of raw multimodality MRIs for simplicity, and the output is a 3-channel (background, tumor core and edema) patch of the same size. Subsequently, the 3-channel output of the network was multiplied by the previous whole tumor prediction probability map to remove outside false positive voxels. The complete architecture of the tumor core extraction network is shown in Fig. 2.

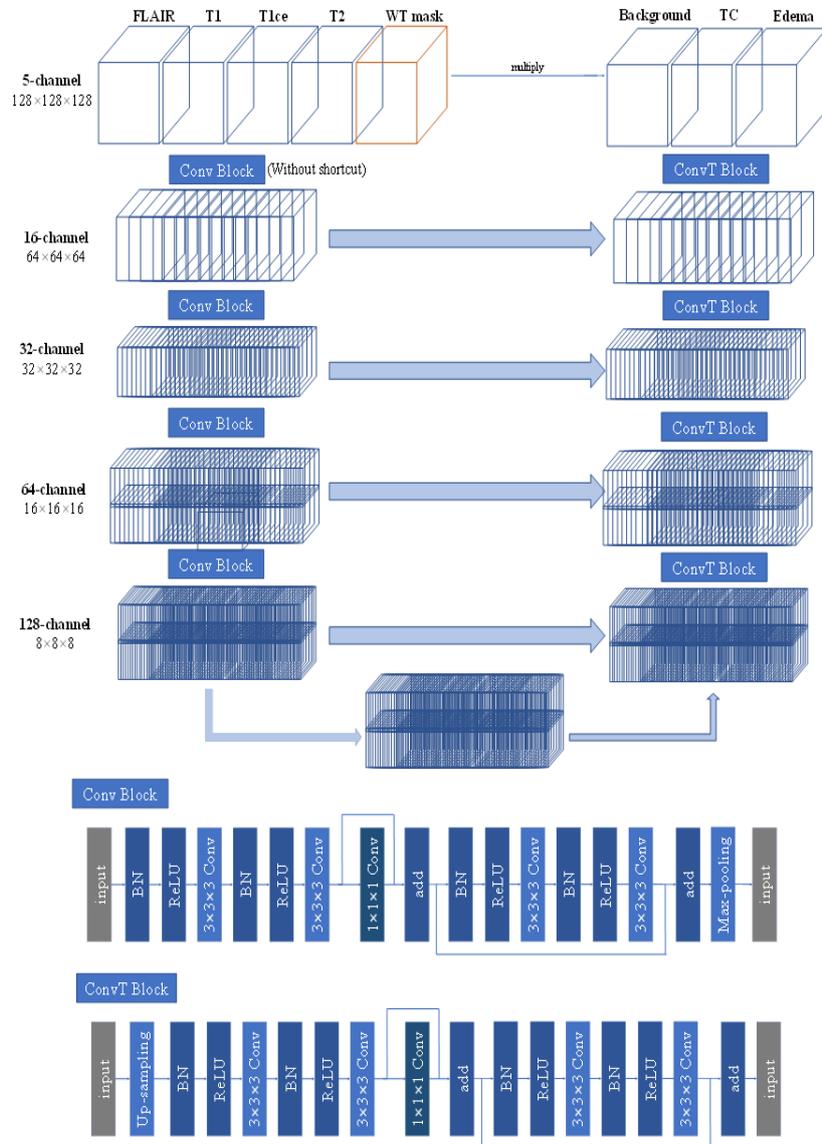


Figure. 2 Architecture of the 3D tumor core extractor

2.5 Multimodality 3D U-Nets for enhancing tumor segmentation

In enhancing tumor segmentation, a 3D U-Net with similar structure as in tumor core segmentation was used. A 5-channel 128*128*128 cubic patch input, which is the concatenation of raw MRIs and the mask of tumor core, is fed into the network. The output is a 3 channel patch of the same size of input, and each channel of output tensor refers to the probability map of background, necrotic/non-enhancing tumor and enhancing

tumor, respectively. The output is multiplied by the tumor core mask as a post-processing step. The loss function can be written as:

$$D = D_{ET} + \alpha \cdot D_{NC} \quad (3)$$

where α is set as 0.4 here.

Finally, we add three output prediction to generate the final mask, including ET-label 4, NCT/NET-label 1 and Edema-label 2.

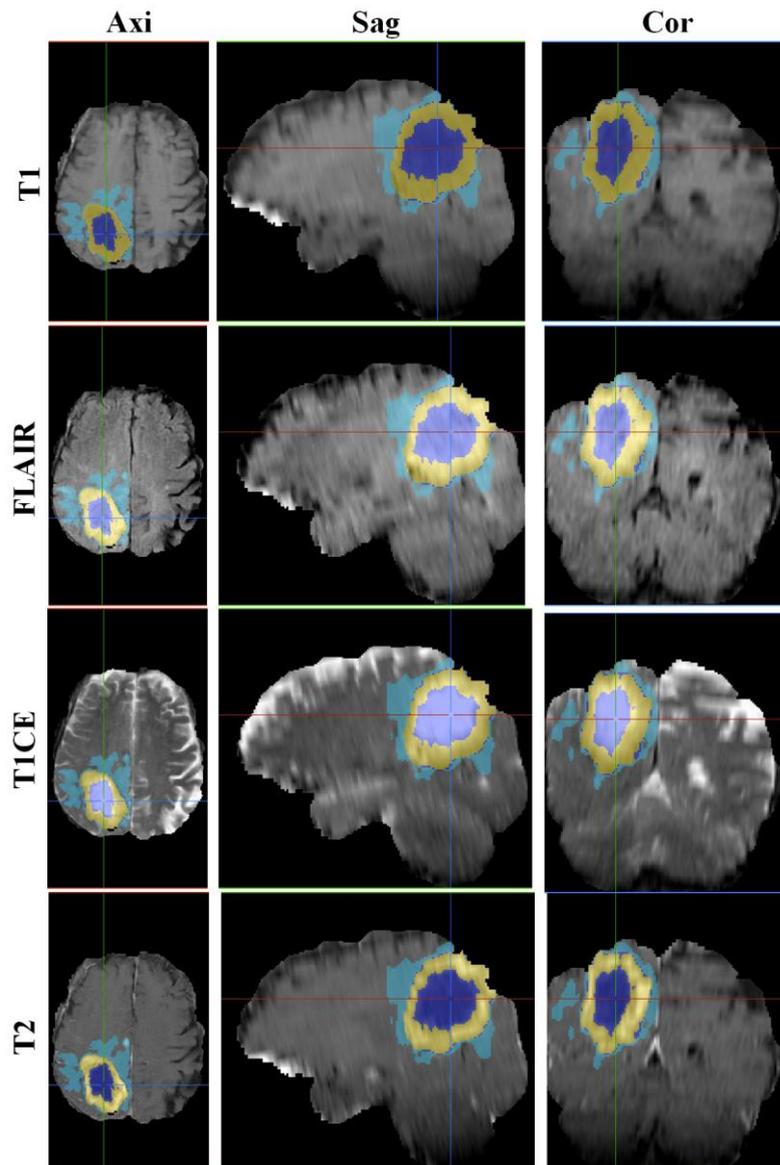


Figure. 3 Segmentation results for typical case

3 Results

We have evaluated our method for both training and validating phases. In training phase, we have applied cross-validation method. The whole data is split into 15 subsets equally, with each cross-validation procedure includes 13 training subsets, 1 test subset and 1 validation subset. The cross-validation procedure is repeatedly executed 15 times and the dice score of the tumor parts for each procedure is recorded. The final dice is calculated as the average in all procedures, and an average dice score of 0.891, 0.863 and 0.810 for whole tumor, tumor core and enhancing tumor is achieved. In validating phase, we have uploaded the output of our proposed network to the organizers' website to obtain the corresponding dice scores of tumor parts. The final results for validation on leaderboards are described in Table [1], with an average dice score of 0.8828, 0.7882 and 0.7197, an average Hausdorff score of 29.21969, 11.06356 and 7.93296 for whole tumor, tumor core and enhancing tumor, respectively. The typical results for segmentation is presented in Fig. 3.

4 Discussions and Conclusions

In this paper, we have proposed a cascaded U-Net architecture that progressively segment the whole tumor, tumor core and enhancing tumor from the whole brain MR images. The 2D and 3D modified U-net have achieved good segmentation results, using a balanced dice score, and model ensembles. During training and validating, we have applied data preprocessing and augmentation techniques to increase the segment accuracy. The cross validation on training dataset and official validation results show that our method has achieved promising results. Currently, segmentation errors mainly lie in the boundary parts of the tumors and several cases have shown lower dice score because of the poor image qualities. Future improvements will be on model tuning and refinement on tumor boundaries, and to increase the robustness of the proposed method towards low quality MR images. We are also going to predict patient overall survival using the refined segmentation results and a feature based learning method.

Table 1 Performance of the proposed method on validation dataset

	Mean	StdDev	Median	25 quantile	75 quantile
Dice ET	0.71965	0.26682	0.8171	0.72289	0.86515
Dice WT	0.8828	0.06948	0.90126	0.87261	0.92758
Dice TC	0.78823	0.20382	0.84841	0.75922	0.90723
Sensitivity ET	0.80544	0.24675	0.88398	0.80673	0.91933
Sensitivity WT	0.92736	0.08467	0.95204	0.92418	0.97418
Sensitivity TC	0.86726	0.17046	0.94171	0.82519	0.96627
Hausdorff95 ET	7.93296	14.35554	2.44949	2	4.12311
Hausdorff95 WT	29.21969	31.2731	10.58258	3.80624	57.55182
Hausdorff95 TC	11.06356	13.41404	6.32456	3	11.35782

References

1. Prastawa M, Bullitt E, Ho S, et al. A brain tumor segmentation framework based on outlier detection[J]. *Medical image analysis*, 2004, 8(3): 275-283.
2. Havaei M, Davy A, Warde-Farley D, et al. Brain tumor segmentation with deep neural networks[J]. *Medical image analysis*, 2017, 35: 18-31.
3. Corso J J, Sharon E, Dube S, et al. Efficient multilevel brain tumor segmentation with integrated bayesian model classification[J]. *IEEE transactions on medical imaging*, 2008, 27(5): 629-640.
4. Ronneberger O, Fischer P, Brox T. "U-net: Convolutional networks for biomedical image segmentation". In *International Conference on Medical Image Computing and Computer-Assisted Intervention 2015 Oct 5* (pp. 234-241). Springer, Cham.
5. Çiçek Ö, Abdulkadir A, Lienkamp S S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//*International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2016: 424-432.
6. F. Milletari, N. Navab, and S. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," *CoRR*, vol. abs/1606.04797, 2016. [Online]. Available: <http://arxiv.org/abs/1606.04797>.
7. Christ P F, Elshaer M E A, Ettliger F, et al. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields[C]//*International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2016: 415-423.
8. Sudre C H, Li W, Vercauteren T, et al. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations[M]//*Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, Cham, 2017: 240-248.
9. Li X, Chen H, Qi X, et al. H-DenseUNet: Hybrid densely connected UNet for liver and liver tumor segmentation from CT volumes[J]. *arXiv preprint arXiv:1709.07330*, 2017.
10. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp F, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694.
11. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117.
12. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
13. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of

- the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
14. Milletari F, Navab N, Ahmadi S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//3D Vision (3DV), 2016 Fourth International Conference on. IEEE, 2016: 565-571.
 15. Drozdal M, Vorontsov E, Chartrand G, et al. The importance of skip connections in biomedical image segmentation[M]//Deep Learning and Data Labeling for Medical Applications. Springer, Cham, 2016: 179-187.
 16. Sudre C H, Li W, Vercauteren T, et al. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations[M]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham, 2017: 240-248.

Coarse-to-Fine Deep Convolutional Neural Networks for Multi-Modality Brain Tumor Semantic Segmentation

Mingyuan Liu¹

¹ Department of biological science and medical engineering,
Beihang University, Beijing, China.
Email: liumingyuan95@buaa.edu.cn

Abstract. Semantic segmentation is a difficult problem, for it requires algorithms include both multi-level reasoning and the pixel-level accuracy. In recent years, deep convolutional neural networks have achieved many result-breaking results in dense prediction and scientists are still working on proposing new methods to produce more precise boundary. In this paper, a two-stage coarse-to-fine model is put forward, of which the first phase is designed to obtain the overall shape of target region and the second one to identify pixel-level details. The method has been tested on the training data and validation data of BrainTS18 challenge by values of dice, sensitivity, specificity and hausdorff95. The final results on testing data will be delivered soon.

Keywords: Tumor segmentation, Two-stage learning, Deep learning.

1 Introduction

Deep convolutional neural networks have achieved lots of result-breaking results in the field of image processing, including classification [11,12], object detection [13] and of course image segmentation [9,10]. In the field of image segmentation, the most challenging one is a semantic segmentation problem. It requires algorithms not only segment the object out, but also categorize pixels into different classes.

Up to now, most of the models proposed for dense prediction problems are adapted from neural networks proposed for natural scene [9]. Those structures, however, are not sensitive enough to depict the boundary of target objects clearly. For example, in an indoor dense prediction problem, when algorithms are required to segment a fork out, there is no need to differentiate each tooth of the fork. However, if the shape of an organ or tissue are multitooth, it may indicate cancerization. As a result, designing a model merely adapt existing models to medical field may not be the best solution.

Multiple stages training is not proposed for the first time[7,8]. Those models are reported to performs better in solving image segmentation problems. Based on those research mentioned above, the second or later neural networks will improve the performance and that improvement could be attributed to better segmentation of details. However, the widespread practice is simply concatenate two same CNNs together and

push the data through them one by one. This method raises questions that whether the second or later neural networks should as complex as the first one. If the first stage is designed to achieve a coarse result and the second stage are designed to polish the output of it, would the structure become more brief and efficient?

In this paper, a two-stage coarse-to-fine model are proposed for semantic segmentation problem. The convolutional neural network of the first stage is designed with pooling operations, so that a large receptive field could be obtained for image reasoning. The second one is proposed aim at dealing with details in a limited range and polish the coarse result to a better one.

To evaluate the result of our method, we participate the challenge called Multi-modal Brain Tumor Segmentation Challenge 2018. There are 285 MRI scans in the training dataset. All of them are 3T multimodal MRI images, including T1-weighted images, post-contrast T1-weighted images, T2-weighted images and T2 Fluid Attenuated Inversion Recovery (FLAIR) images. Labels are annotated by neuroradiologists [1,2,3,4].

In the method part of this paper, we will introduce the structure of our model and some details of it. After that, two stages will be introduced one by one. Then the method to evaluate the result will be put forward. In the result part, the result tested on the validation set will be published.

2 Method

Deep convolutional neural networks are capable of integrating data of multiple modality together and provide state-of-the-art dense prediction result. **错误!未找到引用源。** illustrates the overview of the fully automatic coarse-to-fine model that we proposed.

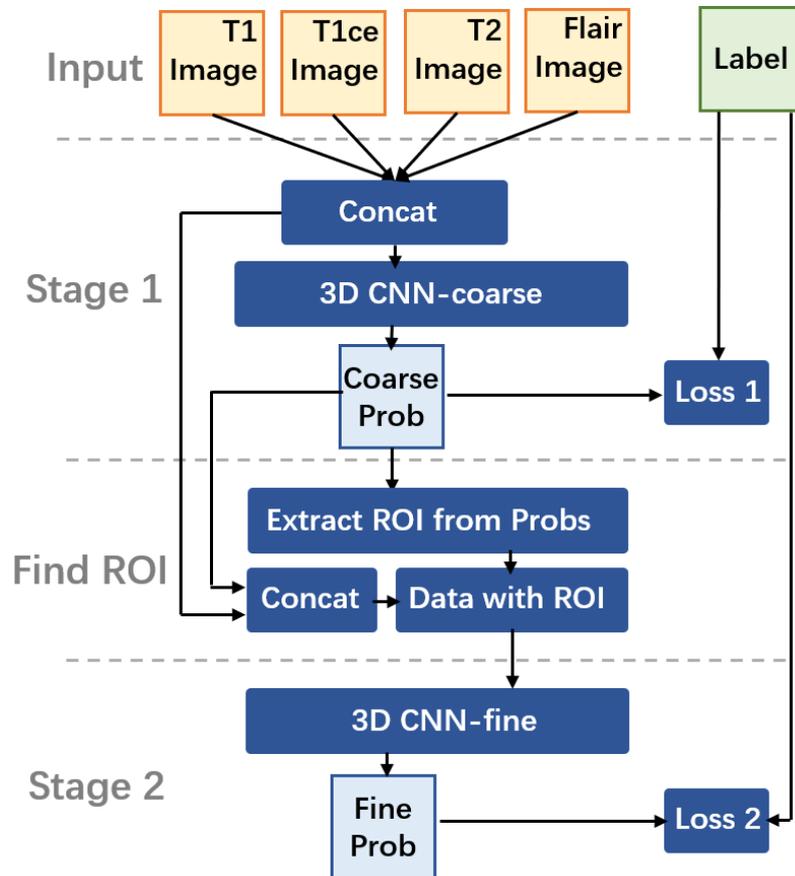


Fig. 1. This figure is the flow chat of the model that we proposed

2.1 Input

In the input stage, four modalities of images, with shape $155 \times 240 \times 240$, are directly concatenated together as the input of the model. This practice is trying to prompt neural networks learn key features itself, without manual intervention, driven by data.

In this experiment, all the images used for training is provided by the BrainTS18 challenge. Data form other sources is not used.

2.2 Training

In the first stage of the whole model, a U-Net[5,6] like, 3-dimensional convolutional neural network are leveraged to provide the coarse segmentation result. The first model is consisted of pooling, convolution and skip layers. Although pooling operations could enlarge the receptive fields of the first model, vital details are also lost. To make up for the information loss in the first stage, the second convolutional neural

network is proposed to recalculate the boundary of target object. In the second stage, there is no pooling operations, which is designed specifically for improving point wise accuracy.

The input of the second phase is the combination of original data including four different modalities and the probability map provided by the first stage. This practice, in a sense, is mimicking the learning progress of human beings. Take the math learning as a case, people usually learn elementary mathematics at first and then learn the advanced math based on previous knowledge, so it is with this two-stage model. The first stage is designed to obtain high-level features. It is easy to achieve and most of the deep learning algorithms could do it well. The second phrase mainly focus on details based on the knowledge learned from the first stage.

The whole model is not an end to end one, for gradients could not pass from the second loss directly to the input. So, two convolutional neural networks should be trained respectively.

2.3 Loss function

There are tow losses in the model. The first one is a balanced dice loss function. In the field of medical image segmentation, the area of target object is usually much smaller than the background. So that the loss will be very small even if the model predicts all the pixels as background, and the ratio of foreground and background is much higher in three-dimensional case. As a result, we multiply the loss of target areas with a large coefficient and the background with a small area to balance the difference of quality between target areas and background. The second one is cross entropy loss.

3 Results

3.1 Dataset

Data engaged in both training and validation phrase are acquired from BrainTS18 challenge [1,2,3,4]. In the training phrase, there are 285 objects, each of them is consists of four modalities, T1-weighted images, post-contrast T1-weighted images, T2-weighted images and T2 Fluid Attenuated Inversion Recovery (FLAIR) images. Those images are from 19 different facilities and are annotated followed the same protocol. There are 66 images without annotation in the validation dataset, results will be reported later.

3.2 Evaluation

For reasons of time, up to the first draft is submitted (14th July), we only finished the training phrase of the first stage of the model. The training of the second one is still in progress, for there is too much large-scale training data.

Table 1. shows the result that our model has achieved on the validation set. The result of three distinct parts are validated including Enhancing Tumor (ET), the whole tumor (WT) and tumor core (TC). All of them are listed in the horizontal direction of the table.

The results are evaluated by four indexes, dice coefficient value, sensitivity value, specificity value and Hausdorff95 distance. For each one of them, we reported the mean value, standard deviation and the median value.

Table 1. Results of the validation dataset

	ET	WT	TC
Dice value mean	0.7639	0.8958	0.7905
Dice value std	0.2529	0.0607	0.2206
Dice value median	0.8565	0.9112	0.8787
Sensitivity mean	0.8165	0.9119	0.8031
Sensitivity std	0.2224	0.0846	0.2256
Sensitivity median	0.8924	0.9340	0.8971
Specificity mean	0.9975	0.9934	0.9972
Specificity std	0.0038	0.0047	0.0032
Specificity median	0.9986	0.9953	0.9981
Hausdorff95 mean	4.0714	4.4924	8.1971
Hausdorff95 std	7.1522	3.1958	9.6222
Hausdorff95 median	2.0000	3.5348	4.4721

To achieve a subject understanding of the results we also visualized some examples. People could find out the difference between our result and the ground truth. Fig 2 illustrates the output of our model. From left to right is the origin T2-weighted images, the middle is the ground truth and the right one is the output of the model. By comparing the difference between the ground truth and the output, we could see that the first stage has already learnt the coarse structure of the images.

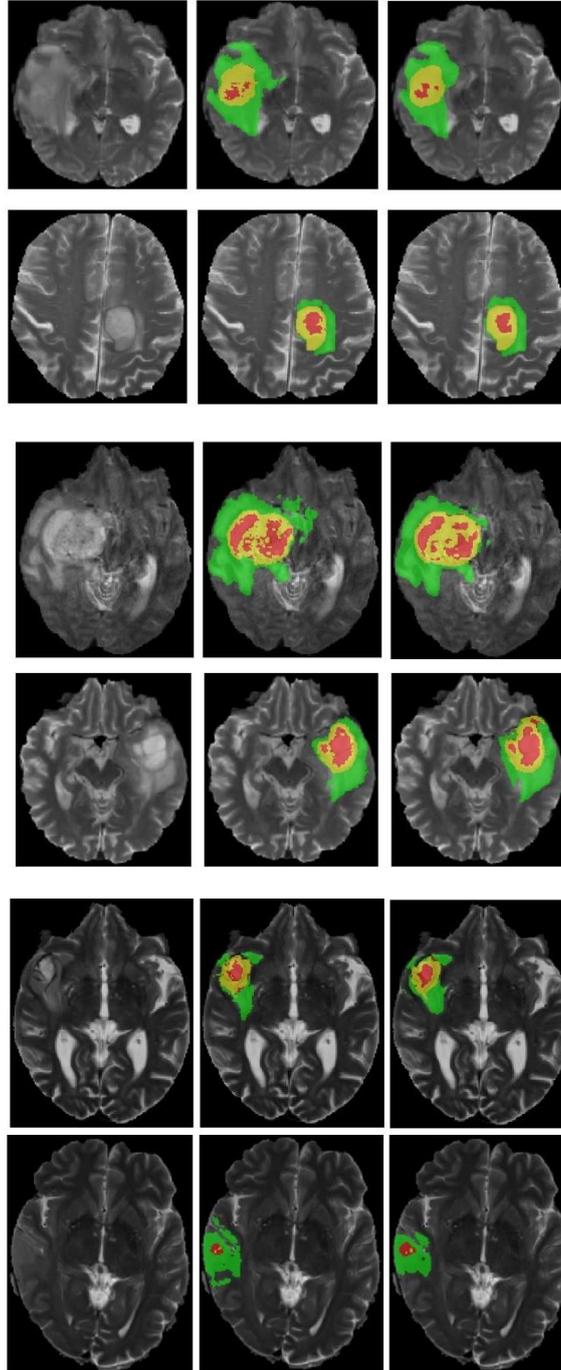


Fig. 2. From left to right is the origin image, origin images with ground truth and the prediction of the model.

4 Conclusion and discussion

In this paper, we described a two-stage framework for semantic segmentation. Up to now our method has only been tested on the validation dataset and the result on the testing data will be reported later. To further understand whether our method works better, a larger dataset is required.

References

1. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Γ , Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharruddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EFLNCS Homepage, <http://www.springer.com/lncs>, last accessed 2016/11/21.
5. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer (2015) 234–241
6. Çiçek, O., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer (2016) 424–432
7. Wang C, MacGillivray T, Macnaught G, et al. A two-stage 3D U-net framework for multi-class segmentation on full resolution image[J]. *arXiv preprint arXiv:1804.04341*, 2018.

8. Roth H R, Oda H, Zhou X, et al. An application of cascaded 3D fully convolutional networks for medical image segmentation[J]. *Computerized Medical Imaging and Graphics*, 2018, 66: 90-99.
9. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015: 3431-3440.
10. Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2018, 40(4): 834-848.
11. Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//*AAAI*. 2017, 4: 12.
12. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: 770-778.
13. Girshick R. Fast r-cnn[C]//*Proceedings of the IEEE international conference on computer vision*. 2015: 1440-1448.

Automatic Brain Tumor Segmentation by Cascaded Lightweight CNN Architecture

Jun Ma¹

Department of Mathematics, Nanjing University of Science and Technology, Nanjing,
210094, China
junma@njust.edu.cn

Abstract. This is brief draft of our developed automatic brain tumor sub-regions segmentation approach. We design a two-stage cascaded framework and propose a lightweight CNN architecture with only ten layers. Experiments with BraTS 2018 validation set show that the proposed method achieved average Dice scores of 0.743, 0.881, 0.773 for enhancing tumor core, whole tumor and tumor core, respectively.

Keywords: Brain tumor, convolutional neural network, dilated kernel, segmentation

1 Introduction

Globally, brain tumors are one of the leading reasons for cancer-related death in human beings. Medical imaging technologies play an important role in the diagnosis of cancer, preoperative planning, intraoperative navigation and post-operative evaluation. Magnetic Resonance Imaging (MRI) is the most frequently used imaging method in the diagnosis of brain tumors, because it is no invasive and does not have radiation.

Brain tumor segmentation in MRI sequences is crucial for quantitative analysis in clinic. However, it is time-consuming and labor-intensive for radiologists to manually delineation the brain tumor. Automatic segmentation of brain tumor has a potential to provide an objective and accurate analysis of the tumors.

In this paper, we design a two-stage cascaded framework to extract brain tumor subregions. In addition, we propose a novel lightweight Convolutional Neural Network (CNN) architecture which is iteratively employed in each segmentation sub-task. Besides, instead taking all the modalities as inputs, we analysis the characters of the brain tumors in each modality and make good use of the complementary information between different modalities.

2 Methods

2.1 Modality Selection

From [10], we find that different brain tumor structures are annotating by different strategies in clinic. Specifically, the edema (belongs to the whole tumor)

is segmented primarily from T2 images. FLAIR is used to cross-check the extension of the edema. The enhancing tumor and the tumor core are identified from T1ce. Motivated by this annotation protocol, we select different modalities for the segmentation of different brain structures. Table 1 contains an overview of the used modalities for different sub-regions segmentation. In other words, we only use FLAIR and T2 sequences to segment the whole tumor and use T1ce to segment tumor core and enhancing tumor.

Table 1. Overview of the used modalities for the segmentation of different sub-regions.

Sub-regions	Used Modality
whole tumor	FLAIR and T2
tumor core	T1ce
enhancing tumor	T1ce

2.2 Two-stage Cascaded Framework

The proposed two-stage cascaded framework is shown in Fig 1. We iteratively use one lightweight CNN architecture to sequentially segment the substructures of brain tumor. Firstly, we segment the whole tumor from FLAIR and T2 MR sequences. After that, we merge the segmentation results of FLAIR and T2 by simply making an union. Thus, a whole tumor ROI can be obtained. To ensure the ROI contains the fully tumor, we make an extension with 5 pixels for the bounding box of the tumor ROI. Following, we crop the FLAIR and T1 sequences by the bounding boxes and train one lightweight CNN architecture to make a binary prediction for whole tumor segmentation, the other same architecture to make a triple prediction for tumor core and enhancing tumor segmentation.

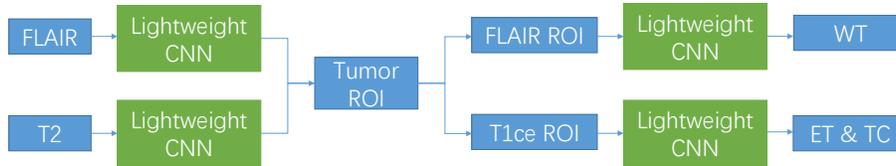


Fig. 1. The proposed two-stage cascaded framework for brain tumor sub-regions segmentation.

2.3 Proposed Lightweight CNN Architecture

For 3D volume segmentation, traditional architectures, such as Unet and FCN, have high memory consumption in the training phase. However, changing the inputs as 2D slices, the context information would be lost. As a trade-off between memory consumption and context information, we present a lightweight CNN architecture (Table 2) for 3D segmentation which integrate dilated convolution with increased and decreased dilated rates, residual connections and batch normalize.

Table 2. The configurations of the proposed ten layers lightweight architecture.

Layers	Configurations (kernel size, channel number)
Conv	$(3 * 3 * 3)$, 8
Dilated Conv	$(3 * 3 * 3)$, 16, dilated factor = 1
Dilated Conv	$(3 * 3 * 3)$, 32, dilated factor = 2
Dilated Conv	$(3 * 3 * 3)$, 64, dilated factor = 4
Dilated Conv	$(3 * 3 * 3)$, 64, dilated factor = 4
Dilated Conv	$(3 * 3 * 3)$, 64, dilated factor = 2
Dilated Conv	$(3 * 3 * 3)$, 64, dilated factor = 1
Conv	$(3 * 3 * 3)$, 64
Conv	$(1 * 1 * 1)$, 64
Conv	$(1 * 1 * 1)$, 2 or 3 for binary/triple segmentation, respectively

Dilated Convolution with increased and decreased dilated rates. Dilated convolutions have been verified as a very effective structure in deep neural networks [12], [4]. The main idea of dilated convolution is to insert "holes" between pixels in traditional convolutional kernels to enlarge the respective field. To obtain multi-scale semantic information, we employ different dilation factors in the proposed architecture. The dilation factors are set to 1, 2, 4 with increased and decreased sequences.

Residual Connections and Batch Normalization. To training deep CNNs more effective, residual connections were firstly introduced by He et al [7]. The main idea of residual connections is to learn residual functions through the use of identity-based skip connections which ease the flow of information across units. Our proposed lightweight architecture adds residual connections to each layer. In addition, each convolutional layer is associated with an element-wise parametric rectified linear Unit (prelu) layer [6] and a batch normalization layer [8]. These components greatly speed the convergence of the training process.

3 Experiments and Results

3.1 Preprocessing.

We use the BraTS 2018 dataset ([3] [1], [2]) for experiments. To enforce the MRI data more uniform, we conduct following preprocessing (Table 3) for the used modalities. It can be found that FLAIR data is added a histogram equalization compared to T2 and T1ce. This is because the intensity distribution of FLAIR varies considerably across different cases. We present two examples of FLAIR images (Fig. 2) in training data set. It can be seen the intensity distribution of the two cases differs remarkably even imposed z-score normalization. Therefore, we further make a histogram equalization to make them share similar intensity distribution. For T2 and T1ce sequences, however, there are no such significant intensity differences between different cases. So a simple z-score preprocessing is enough.

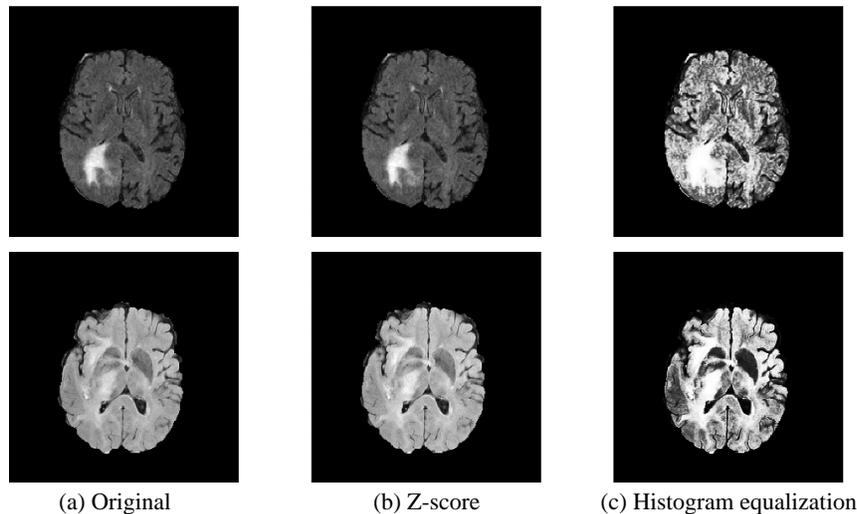


Fig. 2. Preprocessing of two FLAIR images. The first row is the case named "Brats18_TCIA02_135_1 (78th slice)" and the second row is the case named "Brats18_TCIA02_283_1 (78th slice)". After z-score normalization, there is still a great difference between the two images (the 2nd column). Further, a histogram equalization is employed to make them share similar intensity distribution (the 3rd column).

3.2 Implementation Details.

The BraTS 2018 training dataset is randomly divided to training data (80%), validation data (10%) and test data (10%). Our network is implemented with

Table 3. Data preprocessing.

Modality	Preprocessing
FLAIR	z-score, scale to [0, 1] and histogram equalization
T2	z-score and scale to [0, 1]
T1ce	z-score and scale to [0, 1]

NiftyNet ([9], [5]) and trained with resized volume of size $64 * 64 * 64$. The batch size is set as 2. The optimizer is adam with an initial learning rate 0.001 and a L2 weight decay of 10^{-5} . The loss function is Dice coefficient [11] which can deal with the data imbalance. No external data is used and data augmentation during training includes random rotation, random spatial scaling and random flipping. The whole training phase costs 30 hours in a NIVADA 1080Ti GPU.

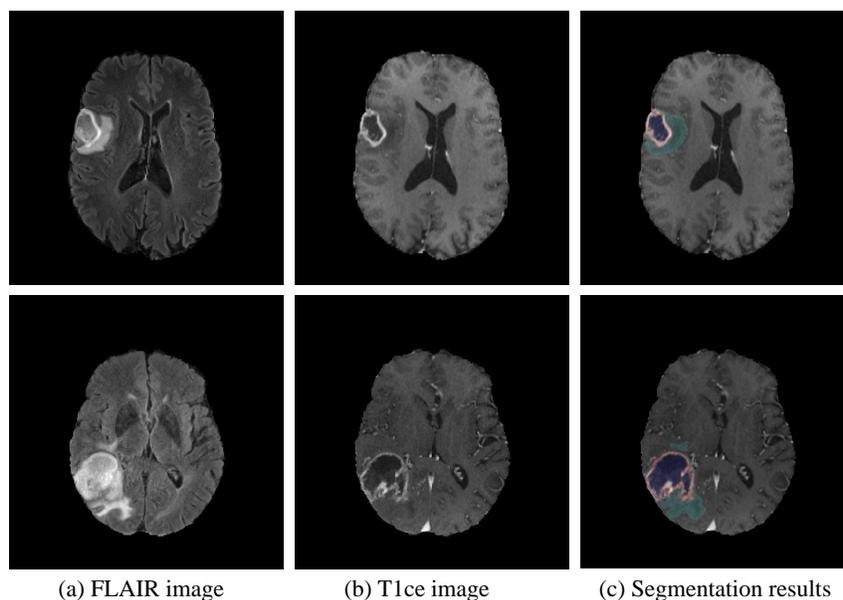


Fig. 3. Segmentation results of the brain tumor sub-regions from two validation cases (named "Brats18_CBICA_ABT_1" (1st row) and "Brats18_CBICA_AMF_1" (2nd row)). Green: edema; Blue: non-enhancing tumor core; Pink: enhancing tumor core.

3.3 Segmentation Results.

We test our framework on the BraTS 2018 validation dataset which consists of 66 new cases. Fig. 3 shows two examples for tumor sub-regions segmentation from two validation cases. The green, blue and pink colors show the edema, non-enhancing and enhancing tumor core, respectively. The quantitative results of training dataset (testing data partition) and validation dataset are shown in Table 4.

Table 4. Quantitative segmentation results of training and validation dataset.

Dataset	Dice_ET	Dice_WT	Dice_TC
Training dataset (testing data)	0.585	0.885	0.557
Validation dataset	0.743	0.881	0.773

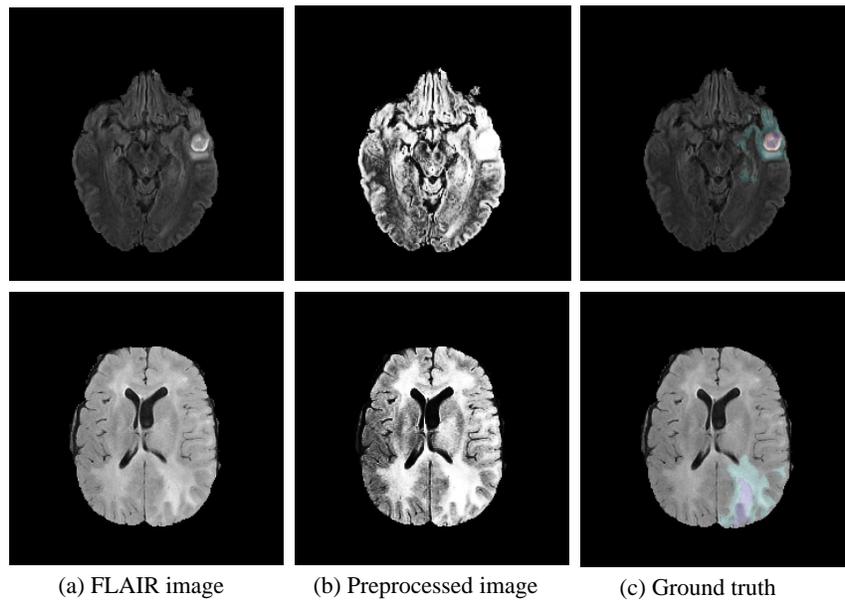


Fig. 4. Two failed cases in the training dataset (named "Brats18_CBICA_ASH_1_seg" (1st row) and "Brats18_TCIA10.413.1" (2nd row)).

3.4 Failed Cases

It is strange that the performance in training dataset (testing data partition) is significantly lower than the performance in validation dataset. Fig. 4 shows some failed cases with almost 0 dice score in training dataset. In the first stage, for these cases, the whole tumor is under-segmented and an incomplete tumor bounding box is generated. It further degenerates the tumor core segmentation in the second stage. Compared to the other well-segmented cases, these cases have low contrast in the tumors' boundaries. After preprocessing, the detailed information are lost (2nd column). From this perspective, the histogram equalization may not be the best strategy to normalize the intensity distribution.

4 Discussion and Conclusion

We propose a lightweight CNN architecture for brain tumor sub-regions segmentation. The architecture only has 10 layers and 0.8M parameters. Experiments with BraTS 2018 validation set show that the proposed method achieved average Dice scores of 0.743, 0.881, 0.773 for enhancing tumor core, whole tumor and tumor core, respectively. We also analysis some failed cases and refining the preprocessing method would be a solution to improve the performance.

References

1. Bakas S, Akbari H, S.A.B.M.R.M.K.J.F.J.F.K.D.C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. The Cancer Imaging Archive (2017)
2. Bakas S, Akbari H, S.A.B.M.R.M.K.J.F.J.F.K.D.C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. The Cancer Imaging Archive (2017)
3. Bakas S, Akbari H, S.A.B.M.R.M.K.J.F.J.F.K.D.C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Nature Scientific Data* **4**(170117) (2017)
4. Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 834–848 (2018). <https://doi.org/10.1109/TPAMI.2017.2699184>
5. Gibson, E., Li, W., Sudre, C.H., Fidon, L., Shakir, D.I., Wang, G., Eaton-Rosen, Z., Gray, R., Doel, T., Hu, Y., Whyntie, T., Nachev, P., Modat, M., Barratt, D.C., Ourselin, S., Cardoso, M.J., Vercauteren, T.: Niftynet: a deep-learning platform for medical imaging. *Computer Methods and Programs in Biomedicine* **158**, 113–122 (2018)
6. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015. pp. 1026–1034 (2015)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)

8. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks pp. 630–645 (2016)
9. Li, W., Wang, G., Fidon, L., Ourselin, S., Cardoso, M.J., Vercauteren, T.: On the compactness, efficiency, and representation of 3d convolutional networks: Brain parcellation as a pretext task. In: Information Processing in Medical Imaging - 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings. pp. 348–360 (2017)
10. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, ., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging* **34**(10), 1993–2024 (2015)
11. Milletari, F., Navab, N., Ahmadi, S.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: Fourth International Conference on 3D Vision, 3DV 2016, Stanford, CA, USA, October 25-28, 2016. pp. 565–571 (2016)
12. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. *CoRR abs/1511.07122* (2015)

Automatic Brain Tumor Segmentation Using a Two-Stage Multi-Modal FCNN*

Michał Marcinkiewicz¹, Jakub Nalepa^{1,2}, Pablo Ribalta Lorenzo^{1,2}, Wojciech Dudzik^{1,2}, and Grzegorz Mrukwa^{1,2}

¹ Future Processing, Gliwice, Poland

{mmarcinkiewicz, jnalepa}@future-processing.com

² Silesian University of Technology, Gliwice, Poland

Abstract. In this paper, we present an approach for automatic brain tumor segmentation by exploiting multiple MRI modalities and processing them in two stages using a fully-convolutional neural network (FCNN). The first stage detects regions of interests, whereas the second stage performs the multi-class classification. The total time required to process one full volume of our algorithm amounts to around 15 seconds. The experiments showed that the preliminary DICE scores over the BraTS'18 validation set are 0.7229, 0.8623, and 0.7675 for enhancing tumor, whole tumor, and tumor core, respectively.

Keywords: Brain Tumor · Segmentation · Deep Learning · CNN.

1 Introduction

Brain tumor segmentation is a vital research topic in the field of medical image analysis and processing due to its wide practical applicability, especially in the context of brain tumor prognosis and staging. Approaches for this task can be divided into *atlas-based*, *unsupervised*, *supervised*, and *hybrid* techniques. In the *atlas-based* algorithms, manually segmented images (referred to as *atlases*) are used to segment incoming (previously unseen) scans [21]. These atlases model the anatomical variability of the brain tissue [19]. Atlas images are extrapolated to new frames by warping and applying non-rigid registration techniques. An important drawback of such techniques is the necessity of creating large (and representative) annotated reference sets. It is time-consuming and error prone in practice, and may lead to atlases which cannot be applied to other tumors because they do not encompass certain types of brain tumors [5, 1].

Unsupervised algorithms search for hidden structures within unlabeled data [7, 16]. In various meta-heuristic approaches, e.g., in evolutionary algorithms [27], brain segmentation is understood as an optimization problem, in which pixels (or voxels) of similar characteristics are searched. It is tackled in a biologically-inspired manner, in which a population of candidate solutions (being the pixel

* This work was supported by the Polish National Centre for Research and Development (POIR.01.02.00-00-0030/15). We applied the *sequence-determines-credit* approach for the sequence of authors.

or voxel labels) evolves in time [6]. Other unsupervised algorithms encompass clustering-based techniques [24, 28, 12], and Gaussian modeling [25]. In *supervised* techniques, manually segmented image sets are utilized to train a model. Such algorithms include, among others, decision forests [8, 32], conditional random fields [29], support vector machines [15], and extremely randomized trees [20].

Deep neural networks, which established the state of the art in a plethora of image-processing and image-recognition tasks, have been successful in segmentation of different kinds of brain tissue as well [18, 14, 10] (they very often require computationally intensive data pre-processing). Holistically nested neural nets for MRI were introduced in [31]. White matter was segmented in [9]. Interestingly, the winning BraTS’17 algorithm used deep neural nets ensembles [13]. However, the authors reported neither training nor inference times of their algorithm which may prevent from using it in clinical practice. *Hybrid* algorithms couple together methods from other categories [26, 22, 30].

In this work, we address the aforementioned issues and propose a deep learning algorithm for automated brain tumor segmentation which exploits a new multi-modal fully-convolutional neural network (Section 3). The experimental evidence (Section 4) elaborated over the newest release of the Brain Tumor Segmentation dataset (Section 2) shows that it can effectively deal with multi-class classification, and deliver high-quality tumor segmentation.

2 Data

The Brain Tumor Segmentation (BraTS) dataset [17, 2–4] encompasses MRI-DCE data of 285 patients with diagnosed gliomas—210 high-grade glioblastomas (HGG), and 75 low-grade gliomas (LGG). Each study case was manually annotated by one to four expert readers. The data comes in four co-registered modalities: native pre-contrast (T1), post-contrast T1-weighted (T1c), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR). All the pixels have one of four labels attached: healthy tissue, Gd-enhancing tumor (ET), peritumoral edema (ED), the necrotic and non-enhancing tumor core (NCR/NET). The studies were interpolated to the same shape ($155 \times 240 \times 240$ with voxel size 1 mm^3) and they were skull-stripped.

3 Methods

Our approach is driven by an assumption that the most salient features of a lesion are not exposed in a single image modality. There are multiple ways to exploit all the modalities in deep learning-based engines. One way is to store three (four) modalities as channels of a single image, like RGB (RGBA), and process it as a standard color image. This approach has a significant downside—only the first layer (which extracts the most basic features) has access to the modalities as separate inputs. Consecutive layers in the network process the outputs of the previous layers—a mix of features from all the modalities. Hu and Xia [11]

utilized a slightly different approach, in which they processed each modality separately, and merged them at the very end to produce the segmentation.

3.1 The Proposed Two-Stage Multi-Modal FCNN

In this work, we combine both techniques—our U-net separates processing pathways and merges them at the very bottom of the network, where the feature space is compacted the most, and at each bridged connection. By doing that, we assure that the low- and high-level features are extracted separately for all modalities in the contracting path. Those features can “interact” with each other in the expanding path, producing high-quality segmentations. Our preliminary experiments showed that the pre-contrast T1 modality carries the smallest amount of information, therefore in order to reduce the amount of segmentation time and resources (thus to make our method easily applicable in a real-life clinical setting), we did not use that modality in our pipeline. However, the proposed U-net-based architecture is fairly flexible and allows for using any number of input modalities. The first stage of image analysis involves taking the whole image as an input, and producing binary mask of the region of interest (therefore, it performs detection of a tumor). This binary mask is used to select the voxels of all modalities from the original images (rendering remaining pixels as background). This region is passed to the segmentation unit by the U-net in the second stage for the final multi-class segmentation.

Our models (Figure 1) are based on a well-known U-net [23], with considerable changes to the architecture. First, there are separate pathways for each modality, effectively making 3 contracting paths. In the original architecture the number of filters was doubled at each down-block, whereas in our model it is constant everywhere, except in the very bottom part of the network (where the concatenation and merging of the paths takes place) where it is doubled. The down-block in our model consists of three convolutional layers (48 filters of the size 3×3 each, with stride 1). The second alteration to the original U-net are the bridged connections, which join (concatenate) activations from each pathway of the contracting paths with their corresponding activations from the expanding path, where they become merged. This procedure allows the network to extract high-level features while preserving the context stored earlier. The expanding path is standard—each up-block doubles the size of an activation map by the upsampling procedure, which is followed by two convolutional layers (48 filters of the size 3×3 each, with stride 1). At the very last layer, there is a 1×1 convolution with 1 filter in the detection (the first stage) or 3 filters in the multi-class classification (the second stage).

3.2 Pre-Processing

The data was acquired with different clinical protocols, various scanners, and at 19 institutions, therefore the pixel intensity distribution may vary significantly. In order to segment all the images (of different modalities) by one algorithm, the data was pre-processed. Our pre-processing included normalization of each

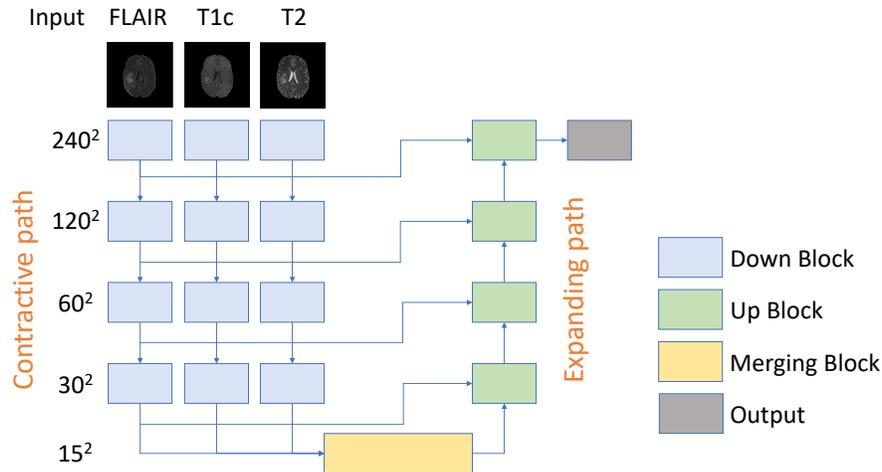


Fig. 1. The proposed network architecture. Three separate pathways (for FLAIR, T1c, and T2) are shown as a part of the contractive path. At each level (each set of down blocks) the output is concatenated and sent to a corresponding up block. At the very bottom, there is a merging block, where all the features are merged before entering the expanding path. The output layer is a 1×1 convolution with one filter for the first stage, and three filters for the second stage.

modality (as a volume) by extracting the brain volume by a simple thresholding routine. Then the mean value of the pixel intensity in the brain region was subtracted and divided by its standard deviation. The value of the pixels outside the brain region was put to zero.

3.3 Post-Processing

After the first stage, the activation is binarized using a threshold of 0.5. The binary mask is post-processed using the 3D connected components analysis—two connected components with the largest volumes remain, and are used to produce input to the second stage. The output of the second stage is an activation map of the size $240 \times 240 \times 3$, where the last dimension represents the number of classes. The activation is then passed through the softmax operation, which performs the final classification.

4 Results

4.1 Experimental setup

The DNN models were implemented using Python3 with the Keras library over CUDA 9.0 and CuDNN 5.1. The experiments were run on a machine equipped

with an Intel i7-6850K (15 MB Cache, 3.80 GHz) CPU with 32 GB RAM and NVIDIA GTX Titan X GPU with 12 GB VRAM.

4.2 Training and Inference

The metric for training was the DICE score for both stages. The optimizer was Nadam (Adam with Nesterov momentum) with the initial learning rate 10^{-5} , and the optimizer parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$. The training ran until DICE score over the validation set did not increase by at least 0.002 in 10 epochs.

Both networks are relatively small, which directly translates to the low computational requirements during inference—one complete volume can be processed and classified end-to-end within around 15 seconds. The training time for one epoch is around 10 minutes (similar for both stages), and the networks converges in around 30–40 epochs (the complete training takes 10–14 hours).

4.3 The Results

The preliminary results obtained over the BraTS'18 training and validation datasets are presented in Table 1 (we present the results for the whole tumor, WT, tumor core, TC, and the enhancing tumor, ET, classes). The whole tumor class is the performance of the first stage of our classification system (evaluated on all the classes merged into one—exactly as the first stage model is trained).

Table 1. The segmentation performance over the BraTS'18 validation dataset obtained using our two-stage FCNN trained with T1c, T2, and FLAIR modalities.

Dataset	Label	DICE	Sensitivity	Specificity
Training	ET	0.6383	0.7416	0.9973
	WT	0.8299	0.8191	0.9920
	TC	0.7437	0.7640	0.9947
Validation	ET	0.7229	0.7773	0.9976
	WT	0.8623	0.8553	0.9931
	TC	0.7675	0.7564	0.9967

5 Discussion

The results show that our two-stage segmentation is able to correctly classify most of the pixels. We had already tested the first stage of the segmentation (binary classification—all the classes were merged into one) on the validation data of the BraTS'17 challenge and obtained fairly similar results (Table 2).

Table 2. The segmentation performance (the whole tumor class only) over 46 BraTS'17 validation patients obtained using our FCNN (first stage only).

	DICE	Sensitivity	Specificity
Mean	0.8301	0.8391	0.9914
Std. Dev.	0.1176	0.1390	0.0068
Median	0.8676	0.8752	0.9931

6 Conclusions

In this paper, we presented our approach to the BraTS'18 segmentation challenge. We developed two fully-convolutional neural network models, which together form a two-stage system, and we obtained high-quality segmentation results over both BraTS'17 and BraTS'18 validation datasets. Training time for both models was around 10 hours, while the inference time for a whole volume is as low as 15 seconds. Due to time restrictions and limited computational environment, we were limited in the number of architectural models and training schemes we were able to investigate. We expect to significantly improve our pipeline for the main challenge.

References

1. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage* **46**(3), 726 – 738 (2009)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Nature Scientific data* **4**, 1–13 (9 2017). <https://doi.org/10.1038/sdata.2017.117>
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection (2017), the Cancer Imaging Archive. <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection (2017), the Cancer Imaging Archive. <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>
5. Bauer, S., Seiler, C., Bardyn, T., Buechler, P., Reyes, M.: Atlas-based segmentation of brain tumor images using a Markov Random Field-based tumor growth model and non-rigid registration. In: *Proc. IEEE EMBC*. pp. 4080–4083 (2010). <https://doi.org/10.1109/IEMBS.2010.5627302>
6. Chander, A., Chatterjee, A., Siarry, P.: A new social and momentum component adaptive PSO algorithm for image segmentation. *Expert Systems with Applications* **38**(5), 4998 – 5004 (2011)
7. Fan, X., Yang, J., Zheng, Y., Cheng, L., Zhu, Y.: A novel unsupervised segmentation method for MR brain images based on fuzzy methods. In: Liu, Y., Jiang, T., Zhang, C. (eds.) *Proc. CVBIA*. pp. 160–169. Springer, Berlin (2005)

8. Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. *NeuroImage* **57**(2), 378 – 390 (2011)
9. Ghafoorian, M., et al.: Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities. *CoRR* **abs/1610.04834** (2016)
10. Ghafoorian, M., et al.: Transfer learning for domain adaptation in MRI: application in brain lesion segmentation. In: *Proc. MICCAI*. pp. 516–524 (2017)
11. Hu, Y., Yong, X.: Automatic brain tumor segmentation using a 3d deep detection-classification model. *Proceedings of the 6th MICCAI BraTS Challenge* (2017)
12. Ji, S., Wei, B., Yu, Z., Yang, G., Yin, Y.: A new multistage medical segmentation method based on superpixel and fuzzy clustering. *Comp. and Math. Meth. in Med.* **2014**, 747549:1–747549:13 (2014)
13. Kamnitsas, K., et al.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Inj.* pp. 450–462. Springer (2018)
14. Korfiatis, P., Kline, T.L., Erickson, B.J.: Automated segmentation of hyperintense regions in FLAIR MRI using deep learning. *Tomography: a journal for imaging research* **2**(4), 334340 (2016). <https://doi.org/10.18383/j.tom.2016.00166>
15. Ladgham, A., Torkhani, G., Sakly, A., Mtibaa, A.: Modified support vector machines for MR brain images recognition. In: *Proc. CoDIT*. pp. 032–035 (2013). <https://doi.org/10.1109/CoDIT.2013.6689515>
16. Mei, P.A., de Carvalho Carneiro, C., Fraser, S.J., Min, L.L., Reis, F.: Analysis of neoplastic lesions in magnetic resonance imaging using self-organizing maps. *Journal of the Neurological Sciences* **359**(1-2), 78–83 (2015)
17. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, ., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftexharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging* **34**(10), 1993–2024 (Oct 2015). <https://doi.org/10.1109/TMI.2014.2377694>
18. Moeskops, P., Viergever, M.A., Mendrik, A.M., de Vries, L.S., Benders, M.J.N.L., Isgum, I.: Automatic segmentation of MR brain images with a convolutional neural network. *IEEE Transactions on Medical Imaging* **35**(5), 1252–1261 (2016). <https://doi.org/10.1109/TMI.2016.2548501>
19. Park, M.T.M., Pipitone, J., Baer, L.H., Winterburn, J.L., Shah, Y., Chavez, S., Schira, M.M., Lobaugh, N.J., Lerch, J.P., Voineskos, A.N., Chakravarty, M.M.: Derivation of high-resolution MRI atlases of the human cerebellum at 3T and segmentation using multiple automatically generated templates. *NeuroImage* **95**, 217 – 231 (2014)
20. Pinto, A., Pereira, S., Correia, H., Oliveira, J., Rasteiro, D.M.L.D., Silva, C.A.: Brain tumour segmentation based on extremely rand. forest with high-level features. In: *Proc. IEEE EMBC*. pp. 3037–3040 (2015). <https://doi.org/10.1109/EMBC.2015.7319032>

21. Pipitone, J., et al.: Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates. *NeuroImage* **101**, 494 – 512 (2014)
22. Rajendran, A., Dhanasekaran, R.: Fuzzy clustering and deformable model for tumor segmentation on MRI brain image: A combined approach. *Procedia Engineering* **30**, 327 – 333 (2012). <https://doi.org/https://doi.org/10.1016/j.proeng.2012.01.868>
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. *CoRR* **abs/1505.04597** (2015)
24. Saha, S., Bandyopadhyay, S.: MRI brain image segmentation by fuzzy symmetry based genetic clustering technique. In: *Proc. IEEE CEC*. pp. 4417–4424 (2007). <https://doi.org/10.1109/CEC.2007.4425049>
25. Simi, V., Joseph, J.: Segmentation of glioblastoma multiforme from MR images - a comprehensive review. *The Egyptian Journal of Radiology and Nuclear Medicine* **46**(4), 1105–1110 (2015)
26. Soltaninejad, M., Yang, G., Lambrou, T., Allinson, N., Jones, T.L., Barrick, T.R., Howe, F.A., Ye, X.: Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI. *Int. J. of Comp. Assist. Radiol. and Surgery* **12**(2), 183–203 (2017)
27. Taherdangkoo, M., Bagheri, M.H., Yazdi, M., Andriole, K.P.: An effective method for segmentation of MR brain images using the ant colony optimization algorithm. *Journal of Digital Imaging* **26**(6), 1116–1123 (2013)
28. Verma, N., Cowperthwaite, M.C., Markey, M.K.: Superpixels in brain MR image analysis. In: *Proc. IEEE EMBC*. pp. 1077–1080 (2013). <https://doi.org/10.1109/EMBC.2013.6609691>
29. Wu, W., Chen, A.Y.C., Zhao, L., Corso, J.J.: Brain tumor detection and segmentation in a CRF (conditional random fields) framework with pixel-pairwise affinity and superpixel-level features. *Int. J. of Comp. Assist. Radiol. and Surgery* **9**(2), 241–253 (2014)
30. Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., Fan, Y.: A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *CoRR* **abs/1702.04528** (2017)
31. Zhuge, Y., Krauze, A.V., Ning, H., Cheng, J.Y., Arora, B.C., Camphausen, K., Miller, R.W.: Brain tumor segmentation using holistically nested neural networks in MRI images. *Med. Phys.* pp. 1–10 (2017). <https://doi.org/10.1002/mp.12481>
32. Zikic, D., Glocker, B., Konukoglu, E., Criminisi, A., Demiralp, C., Shotton, J., Thomas, O.M., Das, T., Jena, R., Price, S.J.: Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR. In: *Proc. MICCAI*. pp. 369–376. Springer (2012)

Ensembles of densely-connected CNNs with label-uncertainty for Brain tumor segmentation.

Richard McKinley, Raphael Meier, Roland Wiest

Support Centre for Advanced Neuroimaging Inselspital, University Hospital, University of Bern, Switzerland

Abstract. We introduce a new family of classifiers based on our previous DeepSCAN architecture, in which densely connected blocks of dilated convolutions are embedded in a shallow U-net-style structure of down/upsampling and skip connections. These networks are trained using a newly designed loss function which models label noise and uncertainty. We present results on the validation dataset of the Multimodal Brain Tumor Segmentation Challenge 2018.

Introduction

We present a network architecture for semantic segmentation, heavily inspired by the recent Densenet architecture for image classification [6], in which pooling layers are replaced by heavy use of dilated convolutions [15]. Densenet employs dense blocks, in which the output of each layer is concatenated with its input before passing to the next layer. A typical Densenet architecture consists of a number of dense blocks separated by transition layers: the transition layers contain a pooling operation, which allows some degree of translation invariance and downsamples the feature maps. A Densenet architecture adapted for semantic segmentation was presented in [7], which adopted the now standard approach of U-net [14]: a downsampling path, followed by an upsampling path, with skip connections passing feature maps of the same spatial dimension from the downsampling path to the upsampling path.

In a previous paper [11], we described an alternative architecture adapting Densenet for semantic segmentation: in this architecture, which we called DeepSCAN, there are no transition layers and no pooling operations. Instead, dilated convolutions are used to increase the receptive field of the classifier. The absence of transition layers means that the whole network can be seen as a single dense block, enabling gradients to pass easily to the deepest layers. While we believe that this approach offers many advantages over U-net, by avoiding pooling and upscaling, this comes at the price of very high memory consumption, since all feature maps are present at the resolution of the final segmentation image. This restricts the possible depth, batch size, and input patch size of the network.

In this paper we describe a family of CNN models for segmentation which represent a continuum from our previously described DeepSCAN models to U-net-like models, in which a pooling-free dense net is embedded inside a U-net style network. This allows the dense part of the network to operate at a lower resolution, improving memory efficiency while maintaining many good properties of the original DeepSCAN architecture.

We describe the general architecture of the family of DeepSCAN models, plus the particular features of the network as applied to brain tumor segmentation, including pre-processing, data augmentation, and a new uncertainty-motivated loss function. We report preliminary results on the validation portion of the BRATS 2018 dataset.

The DeepSCAN family of models

We describe here the constituent parts of the DeepSCAN family of models.

Densely connected layers and Densenet

Densenet [6] is a recently introduced architecture for image classification. The fundamental unit of a densenet architecture is the densely connected block, or dense block. Such a block consists of a number of consecutive dense units, as pictured in Figure 1. In such a unit, the output of each convolutional layer (where a layer here means some combination of convolutional filters, nonlinearities and batch normalization) is concatenated to its input before passing to the next layer. The goal behind Densenet is to build an architecture which supports the training of very deep networks: the skip connections implicit in the concatenation of filter maps between layers allows the flow of gradients directly to those layers, providing an implicit deep supervision of those layers.

In the original Densenet architecture, which has state-of-the-art performance on the CIFAR image recognition task, dense blocks are combined with transition blocks: non-densely connected convolutional layers, followed by a maxpooling layer. This helps to control parameter explosion (by limiting the size of the input to each dense block) and limit redundancy between features, but also means that the deep supervision is not direct, at the lowest layers of the network. This Dense-plus-transition architecture was also adopted by Jegou et al. [7], whose Tiramisu network is a U-net-style variation of the Densenet architecture designed for semantic segmentation.

In our previous paper [11], we dispensed with the transition layers: this means, in effect that the whole network (except for the final one by one convolutions) is a single dense block. This led to networks which were highly parameter efficient, but which had a very large memory footprint. In the current paper we hybridize this approach with the down/up-sampling approach of U-net [14].

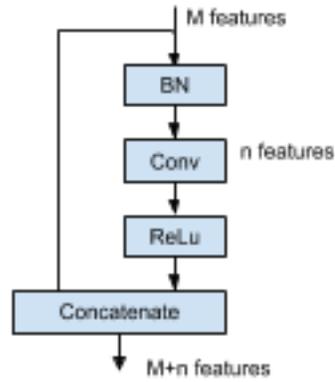


Fig. 1. A Dense unit, as used in the DeepSCAN architecture

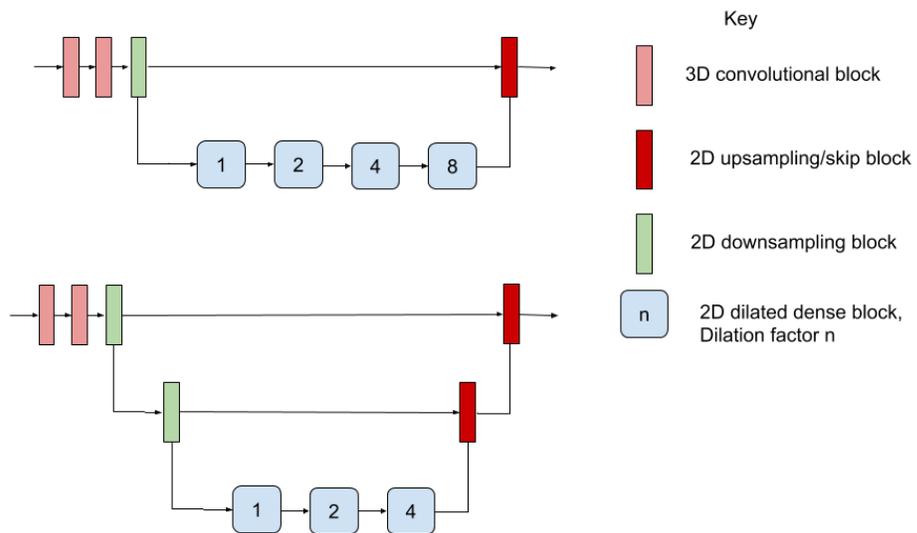


Fig. 2. Two DeepSCAN architectures, as applied to brain tumor segmentation

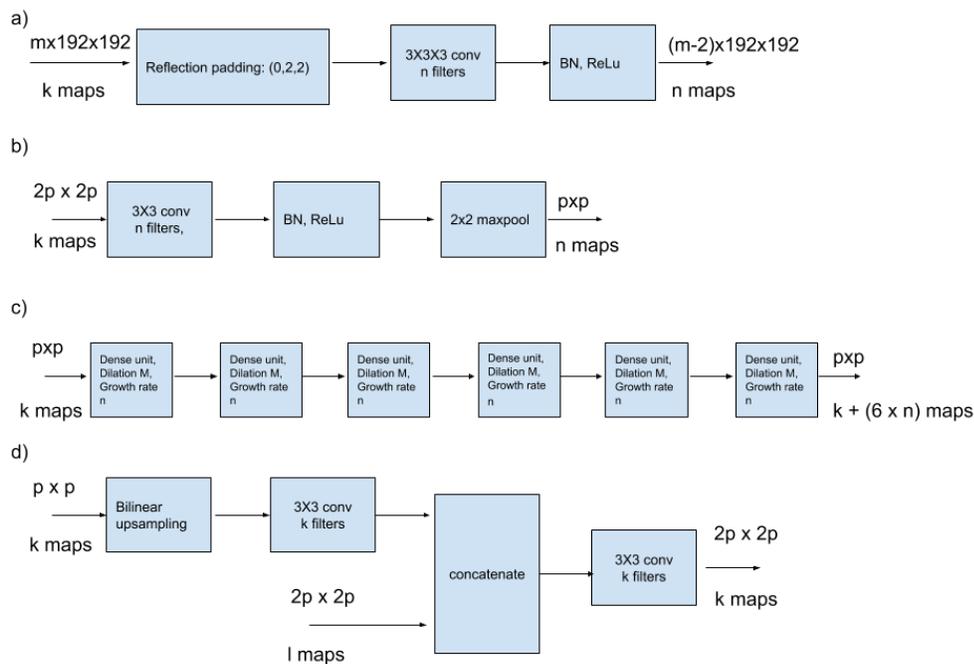


Fig. 3. Units of the DeepSCAN architecture: a) 3D convolutional blocks, b) Downsampling block, c) Dense block, with dilation M , d) upsampling block. Except in the 3D block, all convolutions are preceded by 2 by 2 reflection padding.

Dilated convolutions

The role of pooling layers in CNNs is twofold: to efficiently increase the receptive field and to allow some translation invariance. Translation invariance is of course undesirable in semantic segmentation problems, where what is needed is instead translation equivariance: a translated input corresponding to a translated output. To that end, we use layers with dilated convolutions to aggregate features at multiple scales. Dilated convolutions, sometimes called *trous convolutions*, can be best visualised as convolutional layers “with holes”: a 3 by 3 convolutional layer with dilation 2 is a 5 by 5 convolution, in which only the centre and corner values of the filter are nonzero, as illustrated in Figure 4. Dilated convolutions are a simple way to increase the receptive field of a classifier without losing spatial information.

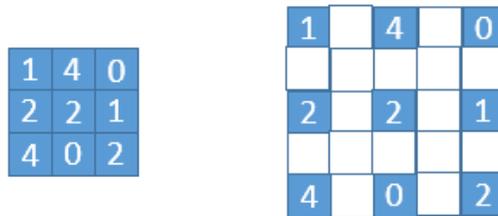


Fig. 4. Left, a 3 by 3 kernel. Right, a 3 by 3 kernel with dilation 2, visualised as a 5 by 5 kernel

Label-uncertainty loss

We introduce a new loss function, which we call label-uncertainty loss, inspired by the recent trend in networks able to quantify their own uncertainty. In brief, for each voxel, and each segmentation subtask (whole tumor, tumor core, and enhancing tumor) the network outputs two probabilities: the probability p that the label is positive, and the probability q that the label predicted does not correspond to the label in the ground-truth annotation (i.e., the probability of a ‘label flip’). If BCE stands for the standard binary cross-entropy loss, and y is the target label, then the loss function we minimize is:

$$BCE(p, (1 - x) * q + x * (1 - q)) + BCE(q, z) \quad (1)$$

where

$$z = (p > 0.5) * (1 - x) + (p < 0.5) * x \quad (2)$$

If q is close to zero, and the label is correct, the first term is approximately the ordinary BCE loss: if q is close to 0.5 (representing total uncertainty as to the correct label) the first term tends to zero. This loss therefore attenuates loss in areas of high uncertainty, in a similar fashion to the heteroscedastic loss of [9]. However, in [9] the uncertainty in the classification is modeled by assuming that logits have a Gaussian distribution, and estimating the variance of that Gaussian: this cannot be performed directly by gradient descent, instead requiring Monte Carlo sampling of the Gaussian distribution to perturb the output of the network. By contrast, label-uncertainty can be incorporated directly into the loss-function of the network. In fact, the label-uncertainty q can also be viewed as a variance: if we assume that the logit of p follows not a Gaussian but a logistic distribution (as is the standard assumption in classical statistical learning) with mean $\text{logit}(p)$, then if the probability that a sample from that distribution is below zero is q , the variance of the logistic distribution is $\text{abs}(\text{logit}(p)/\text{logit}(q))$

Since the label-uncertainty loss incorporates the current prediction in evaluating the probability of a label flip, it is important to apply the loss to a network which has already been pre-trained with ordinary BCE loss: for each of our networks we trained to converge with ordinary BCE loss (typically 10-20 epochs) then switched to using label uncertainty loss. We observed more stability when using both ordinary BCE and label uncertainty. Further, to counter the effects of label imbalance, we adopt the technique of *focal loss* from [10]: therefore, the final loss function used was

$$(1 - p_x)^\gamma (BCE(p, x) + BCE(p, (1 - x) * q + x * (1 - q))) + BCE(q, z) \quad (3)$$

where p_x is p if x is 1 and $(1-p)$ otherwise.

The DeepSCAN architecture

The design principles of the DeepSCAN models are i) non-isotropic input volumes, with one dimension being rather small (in this case, 5 by 192 by 192) ii) initial application of enough 3D convolutions to reduce the short dimension to length 1, and iii) a subsequent hybrid of 2D U-net and 2D Densenet, in which one or steps of convolution and maxpooling are followed by a number of densely connected blocks of dilated convolutions, with the dilation factor increasing with increasing depth, and then finally U-net-style upsampling blocks with skip connections from the previous downward path. The building blocks of these networks are shown in Figure 3, and two architectures built from these blocks are shown in Figure 2.

Initial application to Brain Tumor Segmentation

Brain Tumor segmentation has become a benchmark problem in medical image segmentation, due to the existence since 2012 of a long-running competition, BRATS, together with a large curated dataset [12,1,2,3] of annotated images. Both fully-automated and semi-automatic approaches to brain-tumor segmentation are accepted to the challenge, with supervised learning approaches dominating the fully-automated part of the challenge. A good survey of approaches which dominated BRATS up to 2013 can be found here [4]. More recently, CNN-based approaches have dominated the fully-automated approaches to the problem [5,8,13]

We trained two networks, as pictured in Figure 2. The networks were built using Pytorch, and trained using the Adam optimizer. Rather than using a softmax layer to classify the three labels (edema, enhancing, other tumor) we employ a multi-task approach to hierarchically segment the tumor into the three overlapping targets: whole tumor, tumor core and enhancing: thus the output of the network is three sigmoid units, one for each target.

Data preparation and homogenization

The raw values of MRI sequences cannot be compared across scanners and sequences, and therefore a homogenization is necessary across the training examples. In addition, learning in CNNs proceeds best when the inputs are standardized (i.e. mean zero, and unit variance). To this end, the nonzero intensities in the training, validation and testing sets were standardised, this being done across individual volumes rather than across the training set. This achieves both standardisation and homogenisation.

Data Augmentation

During training, we applied the following data augmentation: randomly flipping along the midline, random rotations in a randomly chosen principal axis, and random shifting and scaling of the standardised intensity values.

Training and results

The network segments the volume slice-by slice: the input data is five consecutive slices from all four modalities, Ground truth for such a set of slices is the lesion mask of the central slice.

Slices from all three directions (sagittal, axial, coronal) were fed to the classifier for training, and in testing the results of these three directions were ensemble

by averaging, for both versions of the classifier (meaning that segmentations were ensembled over six separate classifications). When applied to the BRATS 2018 validation dataset, the mean Dice scores for Enhancing Tumor, Whole core and Tumor Core were 0.79468, 0.90084 and 0.85396 respectively.

References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. In: in press, *Nature Scientific Data* (2017)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. In: *The Cancer Imaging Archive* (2017)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lyg collection. In: *The Cancer Imaging Archive* (2017)
4. Bauer, S., Wiest, R., Nolte, L.L., Reyes, M.: A survey of mri-based medical image analysis for brain tumor studies. *Physics in medicine and biology* 58 13, R97–129 (2013)
5. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A.C., Bengio, Y., Pal, C.J., Jodoin, P.M., Larochelle, H.: Brain tumor segmentation with deep neural networks. *Medical image analysis* 35, 18–31 (2017)
6. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017)
7. Jégou, S., Drozdal, M., Vázquez, D., Romero, A., Bengio, Y.: The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. vol. abs/1611.09326 (2016), <http://arxiv.org/abs/1611.09326>
8. Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis* 36, 61–78 (2017)
9. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? In: *NIPS* (2017)
10. Lin, T.Y., Goyal, P., Girshick, R.B., He, K., Dollár, P.: Focal loss for dense object detection. *2017 IEEE International Conference on Computer Vision (ICCV)* pp. 2999–3007 (2017)
11. McKinley, R., Jungo, A., Wiest, R., Reyes, M.: Pooling-free fully convolutional networks with dense skip connections for semantic segmentation, with application to brain tumor segmentation. In: Crimi, A., Bakas, S., Kuijff, H., Menze, B., Reyes, M. (eds.) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. pp. 169–177. Springer International Publishing, Cham (2018)
12. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C.R., Dojat,

- M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging* 34(10), 1993–2024 (Oct 2015)
13. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in mri images. *IEEE Transactions on Medical Imaging* 35(5), 1240–1251 (May 2016)
 14. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* pp. 234–241 (2015), <http://arxiv.org/abs/1505.04597>
http://link.springer.com/10.1007/978-3-319-24574-4_28
 15. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. In: *Proceedings of International Conference on Learning Representations (ICLR 2017)* (2017)

3D U-net for Brain Tumour Segmentation

Raghav Mehta and Tal Arbel

McGill University
{raghav,arbel}@cim.mcgill.ca

Abstract. In this work, we present a 3D Convolutional Neural Network (CNN) for brain tumor segmentation from Multimodal brain MR volumes. The network is a modified version of popular 3D U-net [12] architecture, that takes as input multi-modal brain MR volumes, processes them at multiple scales, and generates same resolution multi-class tumour segmentation as output. The network is trained end-to-end on BraTS [1–4] 2018 Training dataset using weighted Categorical Cross Entropy (CCE) loss function. A curriculum on class weights is employed to address class imbalance. We achieve competitive segmentation results on BraTS [1–4] 2018 Validation dataset with Dice scores of 0.7882, 0.9091, and 0.8250 for enhancing tumour, whole tumour, and tumour core, respectively.

Keywords: Tumour Segmentation · Deep Learning · Brain MRI

1 Introduction

Multimodal Brain Tumor Image Segmentation Benchmark (BraTS) challenge, since its introduction in 2012, provides good benchmark datasets [1–4] of MR scans of low- and high-grade glioma patients with repeat manual tumour delineations by several human expert. The challenge datasets assist in better evaluation of various automatic tumour segmentation techniques [5–8]. These automatic tumour segmentation techniques are necessary as they allow faster tumour delineation and don't require any human input. This is in contrast to manual delineation method which is a long arduous process and prone to human errors. Automatic quantitative analysis of brain tumours can assist in better and faster diagnosis procedure and surgical planning. This makes automatic tumour segmentation techniques highly impactful in a clinical setting.

Development of good and reliable tumour segmentation techniques remains a challenging task due to many sources of variabilities like tumour types, difference in shape and size of tumour, intensity and contrast difference in MR images, etc. Classical approaches based on Multi Atlas segmentation, probabilistic graphical models like Markov Random Field (MRF) [5] and Conditional Random Field (CRF), Random Forest (RF) [6], etc. have been successfully used for the task of tumour segmentation. They model the statistical distributions of image intensity and image context in different manners. For example, in [5], an iterative multilevel MRF was developed to leverage context and voxel information for tumour segmentation. In addition to these approaches, method based on generative models have also been explored [7] for tumour segmentation.

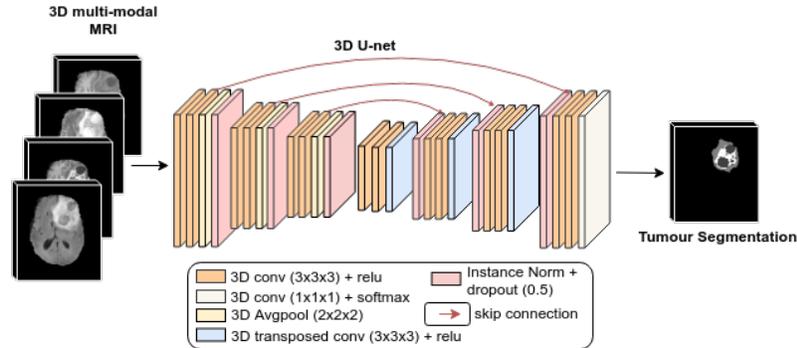


Fig. 1. 3D U-net CNN architecture takes as input four full 3D MR image sequences, and generates the multi-class segmentation of the tumour into sub-types.

Inspired by the success of deep learning in many tasks related to natural images like semantic segmentation [9], object detection [10], and classification [11], many deep learning based approaches have been proposed for various tasks in medical images like segmentation [12], synthesis [13] and classification [14]. Various CNN architectures have been explored for brain tumour segmentation which either explicitly [15, 8] or implicitly [16, 17] models necessary global and local image context. These architectures either takes MR images at multiple resolution as input [15, 8] or processes single resolution MR images at multiple scales [16, 17]. One of the advantages of deep learning based approaches over classical segmentation methods like MRF, RF etc. is that they don't require any hand-crafted features because the networks are trained in end-to-end manner with appropriate loss functions. In recent BraTS challenges [1], deep learning based approaches have these classical methods.

In this work, we apply a modified version of the popular 3D U-net [12] architecture for brain tumour segmentation task on BraTS 2018 datasets. U-net architecture has been successfully used for many medical imaging segmentation tasks like liver and lesion segmentation [18], retinal layer segmentation [19], organ segmentation [20] etc. The network is trained using Categorical Cross Entropy (CCE) loss function on BraTS 2018 training dataset and a curriculum on class weights is employed to address class imbalance. We achieved competitive results on BraTS 2018 validation dataset with Dice scores of 0.7882, 0.9091, and 0.8250 for enhancing tumour, whole tumour, and tumour core, respectively.

2 Method

A flowchart of the 3D U-net architecture can be seen in Figure 1. The network takes as input full 3D volumes of all available sequences of a patient and generates

multi-class segmentation of tumours into sub-types, at the same resolution. The 3D U-net is similar to the one proposed in [12], with some modifications. The U-net consists of 4 resolution steps for both encoder and decoder paths. At the start, we use 2 consecutive 3D convolutions of size 3x3x3 with k filters, where k denotes the user-defined initial number of convolution filters (10). Each step in the encoder path consists of 2 3D convolutions of size 3x3x3 with $k * 2^n$ filters, where n denotes the U-net resolution step. This is followed by average pooling of size 2x2x2. At the end of each encoder step, instance normalization [21] is applied, followed by dropout [22] with 0.05 probability. In the decoder path at each step, 3D transposed convolution of size 3x3x3 is applied, with 2x2x2 stride and $k * 2^n$ filters for the upsampling task. The output of the transposed convolution is concatenated with the corresponding output of the encoder path. This is, once again, followed by instance normalization and Dropout with 0.1 probability. Finally, 2 3D convolution of size 3x3x3 with $k * 2^n$ filters are applied. Rectified linear unit is chosen as a non-linearity function for every convolution layer. The last layer has C filters, where C denotes the total number of classes. This is followed by SoftMax non-linearity.

2.1 Loss function

We optimize weighted Categorical Cross Entropy (CCE) loss function during training. The equation for the same is given below.

$$CCE^i = - \sum_n w_n^i \sum_l t_{n,l}^i \log p_{n,l}^i \quad (1)$$

$$w_n^i = w_l * y_n^i \quad \text{where, } w_l = \left(\frac{\sum_{k=0}^{k=C} m_k}{m_l} \right) * r^{ep} + 1, \quad (2)$$

where, w_n^i and w_l denote the weight for voxel n of volume i and the weight of class l . m_l is total number of voxels of l^{th} class in the training dataset. w_l are decayed over each epoch ep with a rate of $r \in [0, 1]$. It should be noted that w_l converges to 1 as ep becomes large ensuring that all sample receive an equal weight at the later training stages.

3 Experiments and Results

3.1 Data

BraTS 2018 Training Set: The BraTS 2018 training dataset is comprised of 210 high-grade and 75 low-grade glioma patient datasets. Each dataset contains T1, T1 post contrast (T1c), T2, and Fluid Attenuated Inverse Recovery (FLAIR) MR volumes, along with an expert tumour segmentation. Each brain tumour is manually delineated into 3 classes: edema, necrotic/non-enhancing core, and enhancing tumour core.

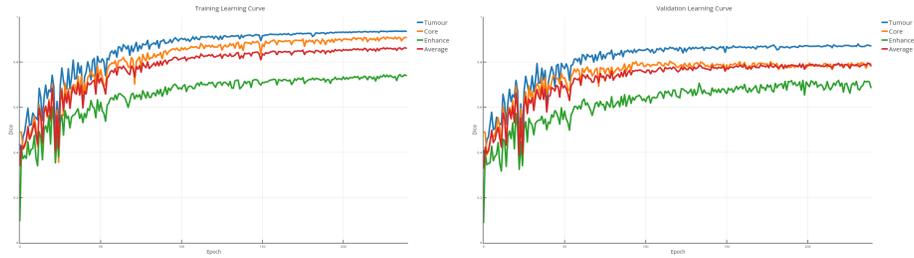


Fig. 2. Training (Left) and Validation (Right) Dice Scores as a function of number of epochs for one of the five cross-validation fold.

BraTS 2018 Validation Set: The BraTS 2018 validation dataset is comprised of 66 patient datasets. Each dataset contains T1, T1 post contrast (T1c), T2, and Fluid Attenuated Inverse Recovery (FLAIR) MR volumes. No expert tumour segmentation masks are provided and the grade of each glioma is not specified [1–4]

3.2 Pre-processing

The BraTS challenge provides isotropic, skull-stripped, and co-registered MR volumes. We follow this up with minimum pre-processing steps. The intensity of volumes were re-scaled using mean subtraction, divided by the standard deviation, and re-scaled from 0 to 1 and were cropped to 184 x 200 x 152.

3.3 5-Fold Cross Validation

We do 5-fold cross validation on the training dataset. The BraTS 2018 training dataset is randomly split into five folds with 57 patient dataset each such that each fold contains 42 high-grade patients and 15 low-grade patients. We train our network 5 times such that 4 folds are used to train the network and the remaining fold is used to validate the network.

Parameters We optimize the loss function in equation 1 using Adam [23] with a learning rate of 0.001 and batch size of 1. The network is trained for total 240 epochs. Learning rate is decayed by the factor of 0.75 after every 50 epochs. The decay rate r in equation 2 is set to 0.95. The initial class weights w_l in equation 2 are set to [1, 350, 160, 465] for background, tumor core, edema, and enhancing tumor, respectively. We regularize the model using data augmentation, where at each training iteration a random affine transformation is applied to the MR volumes and the corresponding segmentation mask. Random translation, rotation, scaling and shear transformations are applied, where the range of transformations is sampled from a uniform distribution of $[-5, 5]$, $[-3^\circ, 3^\circ]$, $[-0.1, 0.1]$, and $[-0.1, 0.1]$, respectively. Volumes are also randomly flipped left to right.

	Dice			Sensitivity			Specificity			Hausdorff-95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Mean	0.690	0.888	0.793	0.774	0.880	0.802	0.998	0.995	0.996	7.251	6.600	7.941
StdDev	0.294	0.094	0.206	0.245	0.118	0.210	0.004	0.006	0.007	13.318	11.215	11.805
Median	0.817	0.918	0.876	0.861	0.913	0.879	0.999	0.996	0.999	2.237	3.606	4.062
25quantile	0.641	0.878	0.748	0.709	0.850	0.723	0.997	0.994	0.996	1.414	2.236	2.236
75quantile	0.878	0.941	0.926	0.935	0.958	0.942	0.999	0.998	0.999	5.385	6.557	9.327

Table 1. Various Evaluation metric statistics for 5-fold cross validation on BraTS 2018 training dataset for enhancing tumour (ET), whole tumour (WT), and tumour core(TC).

	Dice			Sensitivity			Specificity			Hausdorff-95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Mean	0.788	0.909	0.825	0.824	0.911	0.811	0.998	0.995	0.998	3.520	4.923	8.316
StdDev	0.233	0.059	0.179	0.222	0.082	0.212	0.004	0.004	0.002	4.992	8.154	13.521
Median	0.869	0.921	0.902	0.893	0.933	0.901	0.999	0.996	0.999	1.732	2.914	3.240
25quantile	0.809	0.894	0.773	0.824	0.880	0.711	0.998	0.994	0.998	1.414	2.000	2.000
75quantile	0.911	0.951	0.945	0.942	0.964	0.958	0.999	0.998	0.999	3.000	4.970	8.658

Table 2. Various Evaluation metric statistics for BraTS 2018 validation dataset for enhancing tumour (ET), whole tumour (WT), and tumour core(TC).

Learning Curves Figure 2 shows an example of evolution of various dice scores (Tumour, Enhance, Core, and Average) as a function of number of epochs for one of the 5 cross-validation fold.

3.4 Quantitative Results

Tables 1 and 2 show different evaluation metrics statistics provided by the challenge organizers on 5-folded BraTS 2018 training dataset and BraTS 2018 validation dataset. From these tables, we can observe that we get the least performance for enhancing tumour. This was expected as only T1c MR modality helps in identifying enhancing tumour, while for other tumour sub-types all modalities assist in better tumour segmentation.

Please note that we use total five networks, obtained by the corresponding cross-validation, as an ensemble to predict segmentation for BraTS 2018 validation dataset.

3.5 Qualitative Results

Figures 3 and 4 show example segmentation slices for high-grade gliome and low-grade glioma patients from BraTS 2018 training dataset part of 5 fold cross validation. We can observe that the network makes more mistakes for low-grade glioma cases compare to high-grade glioma cases. Predicted segmentation slices for some of the BraTS 2018 validation dataset patients can be seen in Figures 5 and 6.

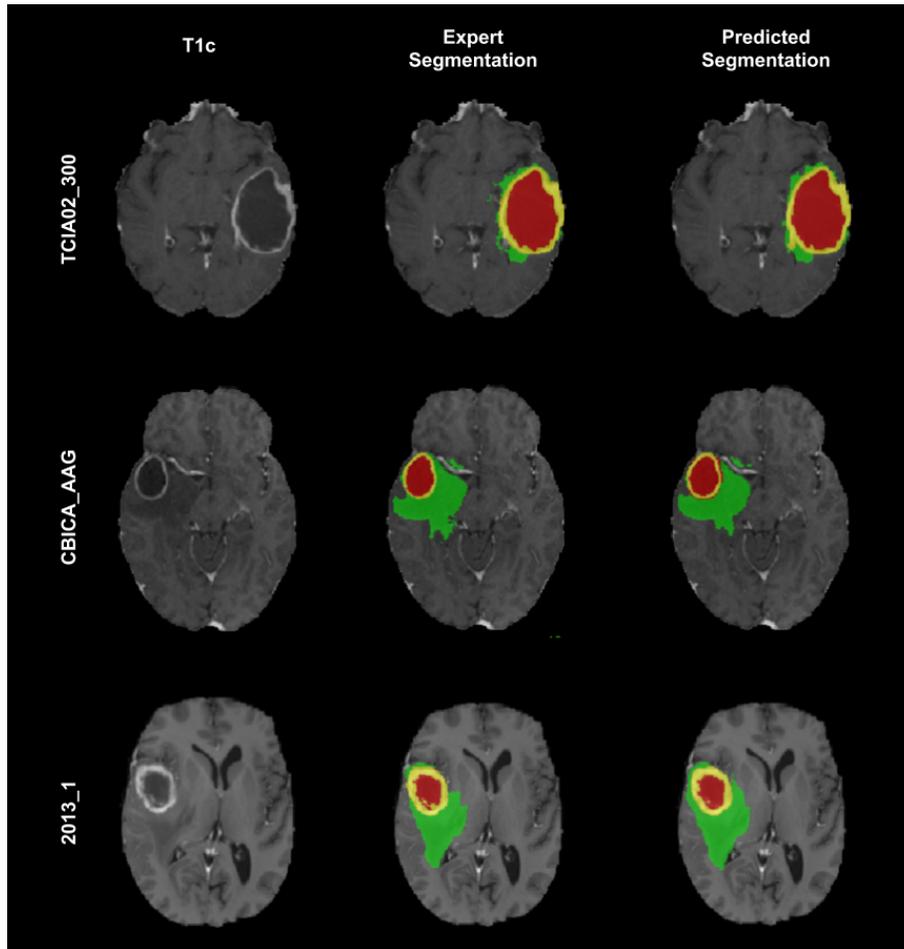


Fig. 3. Examples of high-grade glioma segmentation results for BraTS 2018 training datasets. On T1c MR volume (Column 1), Expert Segmentation (Column 2) and Predicted Segmentation (Column 3) are overlaid. The green label is edema, the red label is non-enhancing or necrotic tumour code, and the yellow label is enhancing tumour core.

4 Conclusion

In this work, we showed how a simple CNN network like 3D U-net [12] can be successfully applied for the task of tumour segmentation. U-net process the input multi-modal MR images at multiple scales, which allows it to learn local and global context necessary for tumour segmentation. The network was trained using a curriculum on class weights to address class imbalance, showing competitive results for brain tumour segmentation on BraTS 2018 validation dataset.

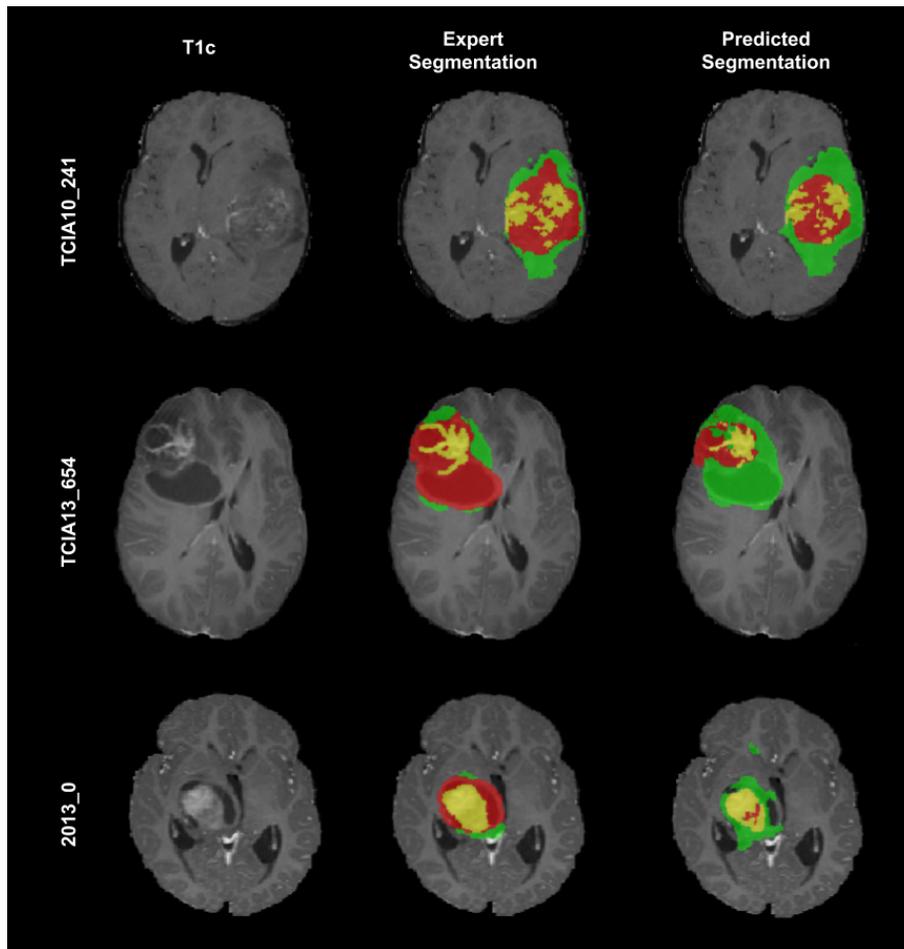


Fig. 4. Examples of low-grade glioma segmentation results for BraTS 2018 training datasets. On T1c MR volume (Column 1), Expert Segmentation (Column 2) and Predicted Segmentation (Column 3) are overlaid. The green label is edema, the red label is non-enhancing or necrotic tumour code, and the yellow label is enhancing tumour core.

References

1. Menze et al. "The multimodal brain tumor image segmentation benchmark (BRATS)." *IEEE TMI* 34, no. 10 (2015): 1993.
2. Bakas et al. "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features." *Scientific data* 4 (2017): 170117.
3. Bakas et al. "Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection." *The Cancer Imaging Archive* 286 (2017).
4. Bakas et al. "Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection." *The Cancer Imaging Archive* (2017).

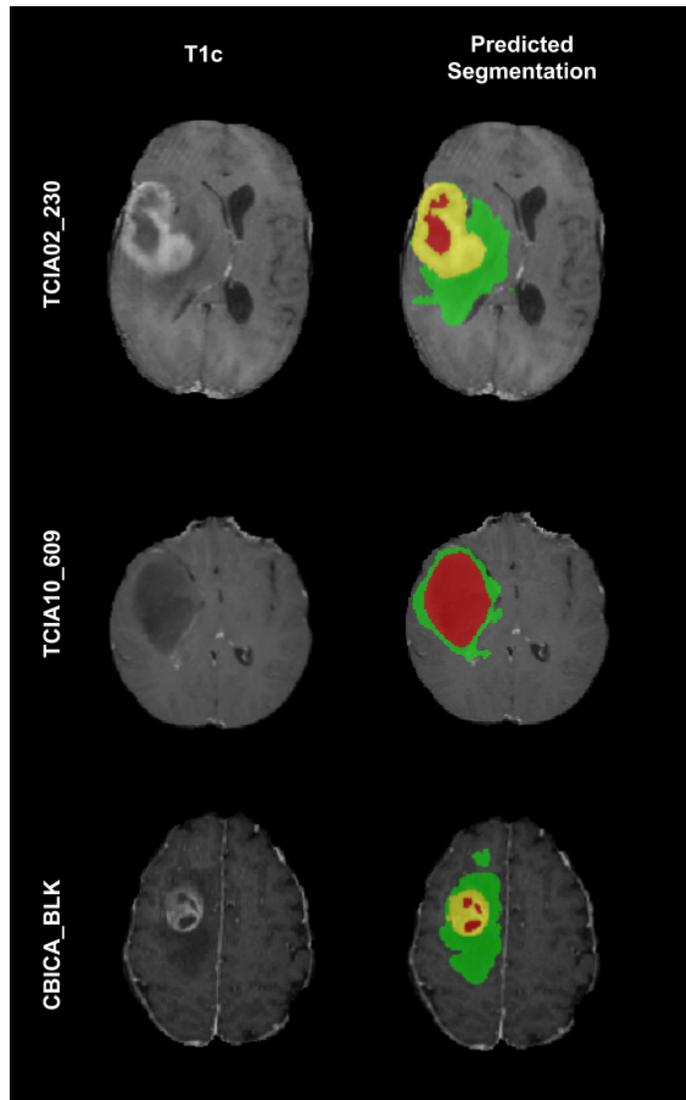


Fig. 5. Examples of segmentation results for BraTS 2018 validation datasets. On T1c MR volume (Column 1) predicted segmentation (Column 2) is overlaid. The green label is edema, the red label is non-enhancing or necrotic tumour core, and the yellow label is enhancing tumour core.

5. Subbanna et al. "Iterative multilevel MRF leveraging context and voxel information for brain tumour segmentation in MRI." In Proceedings of the IEEE CVPR, pp. 400-405. 2014.
6. Zikic et al. "Context-sensitive classification forests for segmentation of brain tumor tissues." Proc MICCAI-BraTS (2012): 1-9.

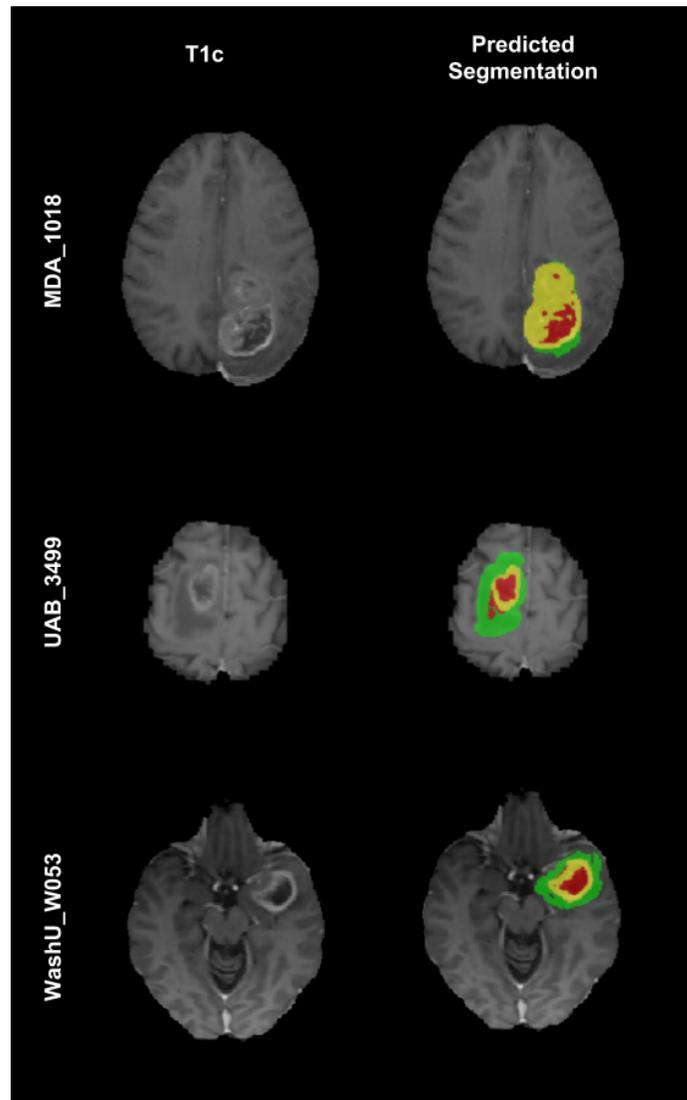


Fig. 6. Examples of segmentation results for BraTS 2018 validation datasets. On T1c MR volume (Column 1) predicted segmentation (Column 2) is overlaid. The green label is edema, the red label is non-enhancing or necrotic tumour core, and the yellow label is enhancing tumour core.

7. Menze et al. "A generative model for brain tumor segmentation in multi-modal images." In MICCAI, pp. 151-159, 2010.
8. Kamnitsas et al. "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation." *Medical image analysis* 36 (2017): 61-78.

9. Long et al. "Fully convolutional networks for semantic segmentation." In Proceedings of the IEEE CVPR, pp. 3431-3440. 2015.
10. Ren et al. "Faster R-CNN: towards real-time object detection with region proposal networks." IEEE Transactions on Pattern Analysis and Machine Intelligence 6 (2017): 1137-1149.
11. Krizhevsky et al. "Imagenet classification with deep convolutional neural networks." In NIPS, pp. 1097-1105. 2012.
12. Cicek et al. "3D U-Net: learning dense volumetric segmentation from sparse annotation." In MICCAI, pp. 424-432. Springer, Cham, 2016.
13. Chatsias et al. "Multimodal mr synthesis via modality-invariant latent representation." IEEE TMI 37, no. 3 (2018): 803-814.
14. Mazurowski et al. "Deep learning in radiology: an overview of the concepts and a survey of the state of the art." arXiv preprint arXiv:1802.08717 (2018).
15. Havaei et al. "Brain tumor segmentation with deep neural networks." Medical image analysis 35 (2017): 18-31.
16. Havaei et al. "HeMIS: Hetero-modal image segmentation." In MICCAI, pp. 469-477. Springer, Cham, 2016.
17. Shen et al. "Boundary-aware fully convolutional network for brain tumor segmentation." In MICCAI, pp. 433-441. Springer, Cham, 2017.
18. Christ et al. "Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields." In MICCAI, pp. 415-423. Springer, Cham, 2016.
19. Roy et al. "ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks." Biomedical optics express 8, no. 8 (2017): 3627-3642.
20. Roth et al. "Hierarchical 3D fully convolutional networks for multi-organ segmentation." arXiv preprint arXiv:1704.06382 (2017).
21. Ulyanov et al., "Instance normalization: The missing ingredient for fast stylization." arXiv preprint arXiv:1607.08022.
22. Srivastava et al. "Dropout: A simple way to prevent neural networks from overfitting." JMLR. 2014, 15(1):1929-58.
23. Kingma, D.P. and Ba, J., "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980.

Ensemble of Fully Convolutional Neural Networks for Brain Tumour Semantic Segmentation.

Miguel Monteiro^{1,2}[0000–0003–3425–5645]
and Arlindo L. Oliveira^{2,3}[0000–0001–8638–5594]

¹ mab.mtr@gmail.com

² INESC-ID, 1000-029 Lisboa, Portugal

³ Instituto Superior Técnico, 1049-001 Lisboa, Portugal

Abstract. In this paper, we present the methods and results of our entry in the Multimodal Brain Tumor Segmentation Challenge 2018. For the semantic segmentation task, we used an ensemble of Fully-Convolutional Neural Networks with a custom loss function which combines the cross-entropy loss with the generalized Dice loss. Our ensemble consisted of a V-Net, a modified V-Net and another modified V-Net with a Conditional Random Field as Recurrent Neural Network layer on top. On the validation set, we obtained an average Dice coefficient of 0.724 for the enhancing tumour core, 0.871 for the whole tumour and 0.803 for the tumour core. For the survival prediction task, we concluded that the best approach was to use a simple one-variable linear regression with age. This method proved to be the best at preventing overfitting given the lack of explanatory variables other than age. On the validation set, we obtained 0.536% accuracy on the main performance metric.

Keywords: Semantic Segmentation · Brain Tumour · Deep Learning · Ensemble · V-Net · CRF as RNN.

1 Introduction

The Multimodal Brain Tumor Segmentation Challenge 2018 (BraTS 2018) [3, 5] proposes its participants to submit methods that are capable performing automatic semantic segmentation of brain tumours given multi-modal three-dimensional images obtained via Magnetic Resonance Imaging (MRI). The data contains pre-operative multimodal MRI scans of higher grade glioma/glioblastoma (HGG) [1] and lower grade glioma (LGG) [2].

There are four images modalities provided for each patient: native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2) and T2 Fluid Attenuated Inversion Recovery (FLAIR). The images have already been co-registered to the same anatomical template, interpolated to the same resolution (1 mm³) and skull-stripped. The scans were provided by multiple institutions using different scanners and acquired under different clinical protocols, resulting in a large amount of variability in the data.

All of the images have been manually segmented by four raters following the same annotation protocol. Their annotations were subsequently approved by experienced neuro-radiologists. There are four different possible labels in the annotations:

- The background (label 0);
- The necrotic and non-enhancing tumour core (NCR/NET) (label 1);
- The peritumoral edema (ED) (label 2);
- The enhancing tumour core (ET) (label 4).

Figure 1 shows example slices for all the available modalities and the respective ground-truth segmentation for one case in the data.

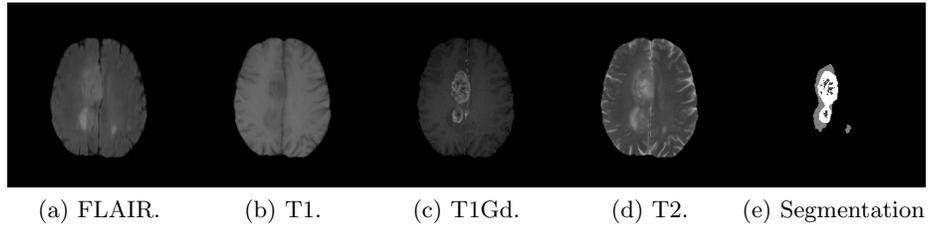


Fig. 1: Example slices for different modalities and expert segmentation.

The performance in the challenge will be measured using the Dice Coefficient (DC) of the whole tumour (ED + NCR/NET + ET), the tumour core (NCR/NET + ET) and the enhancing tumour core (ET). In addition, as secondary metrics, the sensitivity, specificity and Hausdorff distance metrics will also be used to assess performance.

Along with the segmentation task, BraTS 2018, also includes a regression task. This task consists in trying to predict the survival time (in days) after the patient was operated. To do this, along with the image data, the survival times for each patient and the patient’s age is provided. Even though this is a regression task, it will be evaluated based on the classification accuracy of subjects as long-survivors (*e.g.* >15 months), short-survivors (*e.g.* <10 months), and mid-survivors (*e.g.* between 10 and 15 months). After the challenge, mean and median squared errors of the survival time will also be compared.

2 Semantic Segmentation Task

The goal of this task is to perform semantic segmentation of brain tumours given multi-model MRI data.

2.1 Methods

We began by pre-processing the data as follows:

- We trimmed 20 voxels on each side of the x and y axes and 5 voxels at the top of the z axis. We observed that these zones were always empty and would add unnecessary overhead to the training of our algorithm.
- We scaled each channel of the image individually to have zero mean and unit variance. We also tried normalizing via histogram matching with a reference image but we observed no benefit in doing so.
- We stacked the four normalized image modalities into a tensor of shape $(200, 200, 150, 4)$;
- We one-hot encoded the labels into a binary tensor with shape $(200, 200, 150, 4)$, where 4 here is the number of classes.

To perform the segmentation we used an ensemble of Fully-Convolutional Neural Networks (FCNNs) with an encoder-decoder type architectures [4]. Our ensemble consisted of three networks and the final segmentation was obtained by majority voting. For regularization, all the networks in the ensemble used dropout [8] with a keep probability was set to 0.8.

The first network in the ensemble was the V-Net [6] as described in the original paper. The V-Net architecture uses fine-grained feature forwarding and large convolution kernels, $5 \times 5 \times 5$, to ensure that the features used in the final segmentation layer have high-resolution and broad context. The second network in the ensemble was a variation of the V-Net which used $3 \times 3 \times 3$ kernels and skip connection between every convolution. The smaller kernel helps to reduce overfitting and in conjunction with the skip connections also speeds up training. Lastly, we used the same modified V-Net architecture but with a Conditional Random Field (CRF) as Recurrent Neural Network (RNN) layer [7,10] on top of the network. This is meant to take advantage of the spatial and color correlations between all voxels in the image regardless of the distance between them.

As our loss function, we choose the combination of the cross-entropy loss and the generalized Dice loss. The cross-entropy part of the loss serves as a form of stabilizing training by ensuring the background (which comprises of 99% of the voxels most of the time) is segmented as such. The generalized Dice loss pushes training faster towards the objective by directly maximizing the DC on the cross-entropy loss reaches a value close to zero.

The cross-entropy loss function for multi-class classification is given by:

$$Loss_{cross-entropy} = - \sum_{i \in N} \sum_{l \in \mathcal{L}} y_i^{(l)} \log \hat{y}_i^{(l)}, \quad (1)$$

where, given the set of all examples N and the set of all labels \mathcal{L} , $y_i^{(l)}$ is the one-hot encoding (0 or 1) for example i and label l and $\hat{y}_i^{(l)}$ is the predicted probability for the same example/label pair.

The generalized Dice loss [9] is given by:

$$Loss_{GDL} = 1 - 2 \frac{\sum_{l \in \mathcal{L}} \left(w_l \sum_{i \in N} y_i^{(l)} \hat{y}_i^{(l)} \right) + \epsilon}{\sum_{l \in \mathcal{L}} \left(w_l \sum_{i \in N} y_i^{(l)} + \hat{y}_i^{(l)} \right) + \epsilon}, \quad (2)$$

where ϵ is a small constant to avoid dividing by 0 and w_l is given by:

$$w_l = \frac{1}{\left(\sum_{i \in N} y_i^{(l)} \right)^2}. \quad (3)$$

If the denominator of equation 3 is zero we set w_l to zero for that particular label (this is something that happens for some examples in the LGG dataset). The generalized Dice loss directly tries to maximize the DC of each class whilst weighing each class inversely proportional to the square of their frequency. This ensures that even in highly unbalanced semantic segmentation tasks all the classes are taken into account equally.

The final loss function is therefore given by:

$$Loss = Loss_{cross-entropy} + Loss_{GDL}. \quad (4)$$

We trained the network for 100 epochs using both the HGG and LGG datasets simultaneously. We kept a randomly sampled validation set consisting of 12% of the data (30 examples) to monitor overfitting. The initial learning rate was 1e-3 with an exponential learning rate decay by 0.98 every 200 steps:

$$lr := lr \times 0.98^{\frac{\text{current step}}{200}}. \quad (5)$$

We implemented our networks in TensorFlow. The code for the V-Net architecture can be found at <https://github.com/MiguelMonteiro/VNet-Tensorflow> and the code for the CRF as RNN layer can be found at <https://github.com/MiguelMonteiro/CRFasRNNLayer>.

2.2 Results and Discussion

Our results for the mean of the main evaluation metrics on the validation set are shown in Table 1.

Table 1: Validation set results.

DC Enhancing Tumour	DC Whole Tumour	DC Tumour Core
0.724	0.871	0.803

As expected, we obtain a higher DC for the whole tumour and a lower DC for the tumour core and the enhancing tumour core. This is explained by the fact

that segmenting the whole tumour is easier than distinguishing between different intra-tumour regions. Our results are still lacking when compared with the top results from other teams in the validation set, nonetheless, they are in concordance with average results submitted. We believe that using data augmentation and more complex pre-processing such as histogram matching to a histogram derived from multiple images could possibly improve the results.

3 Survival Prediction Task

In this task, the goal is to predict the survival times of patients after the operation given the MRI images and the patient's age.

3.1 Methods

We started from the assumption that we had perfectly solved the segmentation task, hence, we began by extracting human-engineered features from the ground-truth segmentations as if these were always available to us. The underlying reasoning for doing this is that if the expert segmentation does not help predict survival times, then there is little hope that an imperfect segmentation will do so.

The human-engineered features we extracted were for each of the labels (NCR/NET, ED, ET) and for the label combinations (tumour core and whole tumour):

- The volume of the labelled region;
- The distance between the centre of the image and the centre of mass of the labelled region. These are actually three features for each of the Cartesian coordinates (x , y and z);
- The surface area of the labelled region, as defined by the number of voxel faces of the region that are in contact with the exterior of the region;
- The ratio between the volume of the labelled region and the volume of the whole brain;
- The ratio between the surface area of the labelled region and its volume.

Having extracted these features we concatenated them with the patient's age to obtain our full feature set and we removed the patients which do not have the status Gross Total Resection (GTR).

We tested the following out of the box regression algorithms: linear regression, support vector machines, random forest, gradient boosted trees. We performed 10-fold cross-validation and look at the mean and standard deviation of the Mean Absolute Error (MAE) to determine which of the classifiers is performing best.

Table 2: MAE in days.

	All features (days)	Age only (days)
Linear Regression	468.1 \pm 226.4	210.7 \pm 86.5
Gradient Boosting	221.8 \pm 96.8	221.7 \pm 97.0
Random Forest	244.9 \pm 92.7	232.4 \pm 99.0
SVM (linear kernel)	221.1 \pm 97.2	211.8 \pm 85.6

3.2 Results and Discussion

The results for all of the regression algorithms under analysis using all of the extracted features and using only the age feature are shown in Table 2.

As can be seen from Table 2 the best performing method is the simple one-variable linear regression. This indicates that all of the other methods are over-fitting to the more complex features we extracted. We can further explore this hypothesis by looking at the correlation coefficients between features and the target variable (survival in days). In Table 3 we show the top 5 correlation coefficients between features and the target variable and in Table 4 we show the top 5 partially correlation coefficients between the features and the target variable accounting for the effect of the age.

Table 3: Correlation coefficients.

rank	Feature Name	Correlation Coefficient
1	Age	-0.52
2	NCR/NET center of mass z	0.24
3	Whole tumour center of mass x	-0.20
4	Tumour core center of mass z	0.19
5	ED center of mass x	-0.19

Table 4: Correlation coefficients controlling for Age.

rank	Feature Name	Correlation Coefficient
1	NCR/NET center of mass z	0.18
2	Tumour core center of mass z	0.16
3	ET center of mass z	0.14
4	ED volume	0.14
5	ED volume to brain volume ratio	0.13

As we can see from Tables 3 the age feature has more than twice the amount of correlation with the target than the second most correlated feature. In addition, by looking at Table 4 we can see that we control of the variable age, the most correlated feature with the target has only 18% correlation. Hence, we conclude that the age of the patient is still the best predictor of survival time given that all other features we extracted are much less correlated with the target and some of their correlation is already explained by the age variable. Since age is related with less ability to recover and other comorbidities, until better features can be found, a simple one-variable linear regression with age is the good low variance predictor of the survival time.

Training a one-variable linear regression with age on the training set and predicting the survival times for the validation set we obtained 0.536 classification accuracy.

4 Conclusion

For the semantic segmentation task, we used a FCNN with a well regularized V-Net architecture. On the validation set, we obtained an average Dice coefficient of 0.724 for the enhancing tumour core, 0.871 for the whole tumour and 0.803 for the tumour core.

For the survival prediction task, we concluded that the simple one-variable linear regression with age is the best approach given the lack of other explanatory features in the data. We obtained an accuracy of 0.536 on the validation set. We observed that the correlation between all of the features we extracted from the data and the survival time was mostly already explained by the age of the patient. This is likely to age being related to recovery ability, other comorbidities and severity of the disease.

Acknowledgements

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) under projects PRECISE - Accelerating progress toward the new era of precision medicine (LISBOA-01-0145-FEDER-016394), and UID/CEC/50021/2013.

References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive* **286** (2017). <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *The Cancer Imaging Archive* **286** (2017). <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>

3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data* **4**, 170117 (2017)
4. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 3431–3440 (2015)
5. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Ç. Demiralp, Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftexharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging* **34**(10), 1993–2024 (Oct 2015). <https://doi.org/10.1109/TMI.2014.2377694>
6. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *3D Vision (3DV), 2016 Fourth International Conference on*. pp. 565–571. IEEE (2016)
7. Monteiro, M., Figueiredo, M.A.T., Oliveira, A.L.: Conditional Random Fields as Recurrent Neural Networks for 3D Medical Imaging Segmentation. *ArXiv e-prints* (Jul 2018)
8. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research* **15**(1), 1929–1958 (2014)
9. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 240–248. Springer (2017)
10. Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.S.: Conditional random fields as recurrent neural networks. In: *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*. pp. 1529–1537. ICCV '15, IEEE Computer Society, Washington, DC, USA (2015). <https://doi.org/10.1109/ICCV.2015.179>, <http://dx.doi.org/10.1109/ICCV.2015.179>

3D MRI brain tumor segmentation using autoencoder regularization

Andriy Myronenko

Nvidia, Santa Clara, CA
amyronenko@nvidia.com

Abstract. Automated segmentation of brain tumors from 3D magnetic resonance images (MRIs) is necessary for diagnosis, monitoring, and treatment planning of the disease. Manual delineation practices not only require anatomical knowledge, but are also expensive, time consuming and can be inaccurate due to human error. Here, we describe a semantic segmentation network for tumor subregion segmentation from 3D MRIs based on encoder-decoder architecture. Due to a limited training dataset size, a variation auto-encoder branch is added to reconstruct the input image itself in order to regularize the shared decoder and impose additional constraints on its layers. The results of the proposed approach on BRATS 2018 online validation set are 0.8233, 0.9100, 0.8668 average Dice for enhanced tumor core, whole tumor, and tumor core respectively.

1 Introduction

Brain tumors are categorized into primary and secondary tumor types. Primary brain tumors originate from brain cells, whereas secondary tumors metastasize into the brain from other organs. The most common type of primary brain tumors are Gliomas, which arise from brain glial cells. Gliomas can be of low-grade (LGG) and high-grade (HGG) subtypes. High grade gliomas are an aggressive type of malignant brain tumor that grows rapidly, usually require surgery and radiotherapy and have poor survival prognosis. Magnetic Resonance Imaging (MRI) became a key diagnostic tool for brain tumor analysis, monitoring and surgery planning. Usually, several complimentary 3D MRI modalities are acquired, such as T1, T1 with contrast agent (T1c), T2 and Fluid Attenuation Inversion Recover (FLAIR) to emphasize different tissue properties and areas of tumor spread. For example the contrast agent, usually gadolinium, emphasizes hyperactive tumor subregions in T1c MRI modality.

Automated segmentation of 3D brain tumors can save physicians time and provide accurate reproducible solution for further tumor analysis and monitoring. Recently, deep learning based segmentation techniques overcame traditional computer vision methods for dense semantic segmentation. Convolutional neural networks (CNN) are able to learn from examples, and demonstrated state-of-the-art segmentation accuracy both in 2D natural images [5] and in 3D medical image modalities [13].

Semantic segmentation CNNs are usually encoder-decoder based, where the encoder part of the network extracts deep image features, and the decoder part reconstructs a dense segmentation mask [5, 6, 8, 14, 13]. Here, we describe our 3D segmentation network design for brain tumor segmentation task and evaluate it on BRATS 2018 challenge data [12, 4, 2, 3]. Last year, the BRATS 2017 winning challenge submissions included Kamnitsas et al. [10] who proposed to ensemble several models for robust segmentation, and Wang et al. [15] who proposed to segment tumor subregions in cascade using anisotropic convolutions. In this work, we add the variational autoencoder branch to the network to reconstruct the input images jointly with segmentation in order to regularize the shared encoder. At inference time only the main segmentation encode-decoder part is used. We describe our network architecture in the following section.

2 Methods

Our segmentation approach follows encoder-decoder based CNN architecture with an asymmetrically larger encoder part to extract image features and a smaller decoder to reconstruct the segmentation mask. We add an additional branch to the encoder endpoint to reconstruct the original image, similar to auto encoder architecture. The motivation for using the auto-encoder branch is to add additional guidance and regularization to the encoder part, since the training dataset size is limited. We follow the variational auto-encoder (VAE) approach to better cluster/group the features of the encoder endpoint. We describe the building parts of our networks in the next subsections.

2.1 Encoder part

The encoder part uses ResNet [9] blocks, where each block consists of two convolutions with normalization and ReLU, followed by additive identity skip connection. For normalization, we use Group Normalization (GN) [16], which shows better than BatchNorm performance when batch size is small (batch size of 1 in our case). We follow a common CNN approach to progressively downsize image dimensions by 2 and simultaneously increase feature size by 2. For downsizing we use strided convolutions.

All convolutions are 3x3x3 with initial number of filters equal to 32. The encoder part structure is shown in Table 1. The encoder endpoint has 256 features, and is 8 times spatially smaller than the input image. We decided against further downsizing to preserve more spatial content.

2.2 Decoder part

The decoder structure is similar to the encoder one, but with a single block per each spatial level. Each decoder level begins with upsizing: reducing the number of features by a factor of 2 (using 1x1x1 convolutions) and doubling the spatial dimension (using 3D bilinear upsampling), followed by an addition of encoder

Table 1. Encoder structure, where GN stands for group normalization (with group size of 8), Conv - 3x3x3 convolution, AddId - addition of identity/skip connection. Repeat column shows the number of repetitions of the block. We refer to the final output of the encoder, as the encoder endpoint

Name	Ops	Repeat	Output size
Input			4x160x192x128
InitConv	Conv	1	32x160x192x128
EncoderBlock0	GN,ReLU,Conv,GN,ReLU,Conv, AddId	1	32x160x192x128
EncoderDown1	Conv stride 2	1	64x80x96x64
EncoderBlock1	GN,ReLU,Conv,GN,ReLU,Conv, AddId	2	64x80x96x64
EncoderDown2	Conv stride 2	1	128x40x48x32
EncoderBlock2	GN,ReLU,Conv,GN,ReLU,Conv, AddId	2	128x40x48x32
EncoderDown3	Conv stride 2	1	256x20x24x16
EncoderBlock3	GN,ReLU,Conv,GN,ReLU,Conv, AddId	4	256x20x24x16

output of the equivalent spatial level. The end of the decoder has the same spatial size as the original image, and the number of features equal to the initial input feature size, followed by 1x1x1 convolution into 1 channel and sigmoid function. The decoder structure is shown in Table 2.

Table 2. Decoder structure, where GN stands for group normalization (with group size of 8), Conv - 3x3x3 convolution, Conv1 - 1x1x1 convolution, AddId - addition of identity/skip connection, UpLinear - 3D linear spatial upsampling

Name	Ops	Repeat	Output size
DecoderUp2	Conv1, UpLinear, +EncoderBlock2	1	128x40x48x32
DecoderBlock2	GN,ReLU,Conv,GN,ReLU,Conv, AddId	1	128x40x48x32
DecoderUp1	Conv1, UpLinear, +EncoderBlock1	1	64x80x96x64
DecoderBlock1	GN,ReLU,Conv,GN,ReLU,Conv, AddId	1	64x80x96x64
DecoderUp0	Conv1, UpLinear, +EncoderBlock0	1	32x160x192x128
DecoderBlock0	GN,ReLU,Conv,GN,ReLU,Conv, AddId	1	32x160x192x128
DecoderEnd	Conv1, Sigmoid	1	1x160x192x144

2.3 VAE part

Starting from the encoder endpoint output, we first reduce the input to a low dimensional space of 256 (128 to represent mean, and 128 to represent std).

Then, a sample is drawn from the Gaussian distribution with the given mean and std, and reconstructed into the input image dimensions following the same architecture as the decoder, except we don't use the inter-level skip connections from the encoder here. The VAE part structure is shown in Table 3.

Table 3. VAE decoder branch structure, where GN stands for group normalization (with group size of 8), Conv - 3x3x3 convolution, Conv1 - 1x1x1 convolution, AddId - addition of identity/skip connection, UpLinear - 3D linear spatial upsampling, Dense - fully connected layer

Name	Ops	Repeat	Output size
VD	GN, ReLU, Conv (16) stride 2, Dense (256)	1	256x1
VDraw	sample $\sim \mathcal{N}(\mu(128), \sigma^2(128))$	1	128x1
VU	Dense, ReLU, Conv1, UpLinear	1	256x20x24x16
VUp2	Conv1, UpLinear	1	128x40x48x32
VBlock2	GN,ReLU,Conv,GN,ReLU,Conv, AddId	1	128x40x48x32
VUp1	Conv1, UpLinear,	1	64x80x96x64
VBlock1	GN,ReLU,Conv,GN,ReLU,Conv, AddId	1	64x80x96x64
VUp0	Conv1, UpLinear,	1	32x160x192x128
VBlock0	GN,ReLU,Conv,GN,ReLU,Conv, AddId	1	32x160x192x128
Vend	Conv1	1	4x160x192x128

2.4 Loss

Our loss function consists of 3 terms:

$$\mathbf{L} = \mathbf{L}_{dice} + 0.1 * \mathbf{L}_{L2} + 0.1 * \mathbf{L}_{KL} \quad (1)$$

\mathbf{L}_{dice} is a soft dice loss [13] applied to the decoder output p_{pred} to match the segmentation mask p_{true} :

$$\mathbf{L}_{dice} = \frac{2 * \sum p_{true} * p_{pred}}{\sum p_{true}^2 + \sum p_{pred}^2 + \epsilon} \quad (2)$$

where summation is voxel-wise, and ϵ is a small constant to avoid zero division.

\mathbf{L}_{L2} is an L2 loss on the VAE branch output I_{pred} to match the input image I_{input} :

$$\mathbf{L}_{L2} = \|I_{input} - I_{pred}\|_2^2 \quad (3)$$

\mathbf{L}_{KL} is standard VAE penalty term [11, 7], a KL divergence between the estimated normal distribution $\mathcal{N}(\mu, \sigma^2)$ and a prior distribution $\mathcal{N}(0, 1)$, which

has a closed form representation:

$$\mathbf{L}_{\text{KL}} = \frac{1}{N} \sum \mu^2 + \sigma^2 - \log \sigma^2 - 1 \quad (4)$$

where N is total number of image voxels. We empirically found a hyper-parameter weight of 0.1 to provide a good balance between dice and VAE loss terms in Equation 1.

2.5 Optimization

We use Adam optimizer with initial learning rate of $\alpha_0 = 1e-4$ and progressively decrease it according to:

$$\alpha = \alpha_0 * \left(1 - \frac{e}{N_e}\right)^{0.9} \quad (5)$$

where e is an epoch counter, and N_e is a total number of epochs (300 in our case). We use batch size of 1, and draw input images in random order (ensuring that each training image is drawn once per epoch).

2.6 Regularization

We use L2 norm regularization on the convolutional kernel parameters with a weight of $1e-5$. We also use the spatial dropout with a rate of 0.2 after the initial encoder convolution. We have experimented with other placements of the dropout (including placing dropout layer after each convolution), but didn't find any additional accuracy improvements.

2.7 Data preprocessing and augmentation

We normalize all input images to zero mean and unit std (based on non-zero voxels only). We apply a random (per channel) intensity shift ($-0.1..0.1$ of image std) and scale ($0.9..1.1$) on input image channels. We also apply a random axis mirror flip (for all 3 axes) with a probability 0.5.

3 Results

We report the preliminary results using the data provided by BRATS 2018¹. The training dataset contains 285 cases (210 HGG and 75 LGG). Each case contains four 3D MRI modalities (T1, T1c, T2, FAIR), which are rigidly aligned, skull stripped, MRI bias corrected, and resampled to 1x1x1mm isotropic resolution (with image sizes of 240x240x155 voxels). The ground truth segmentation mask is provided for 3 classes: enhanced tumor core (ET), whole tumor (WT) and tumor core (TC). Class labels are nested such that the largest area class WT

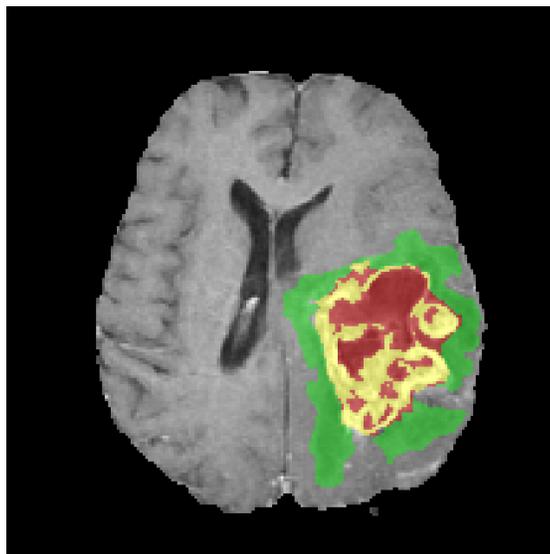


Fig. 1. An example of 2D axial slice of T1c MRI. The whole tumor (WT) class includes all visible labels (a union of green, yellow and red labels), the tumor core (TC) class is a union of red and yellow, and the enhancing tumor core (ET) class is shown in yellow (a hyperactive tumor part) .

also includes TC and ET as its subsets, and TC includes ET as its subset (see Figure 1)

The validation dataset contains 66 cases with unknown glioma grade and unknown segmentation. We uploaded our segmentation results to the BRATS 2018 server for evaluation of per class Dice, sensitivity, specificity and Hausdorff distances.

We implemented our network in Tensorflow [1] and trained it on NVIDIA Tesla V100 GPU. During training we use random crop of size 160x192x128, which ensures that most image content remains within the crop area. We concatenate 4 available 3D MRI modalities into the 4 channel image as an input. We train a separate network for binary segmentation of each of the 3 label classes separately (using the same hyper-parameters).

Table 4 shows the results of our model on the BRATS 2018 validation dataset. Aside from evaluating a single model, we also applied test time augmentation (TTA) by mirror flipping the input 3D image axes, and average the output of the resulting 8 flipped segmentation probability maps; this slightly improved the dice accuracy. Finally, we ensemble a set of 10 models (trained from scratch), to further improve the performance. At the time of the paper submission (Jul 13, 2018), our dice accuracy performance was second best (team name NVDLMED²)

¹ <https://www.med.upenn.edu/sbia/brats2018.html>

² <https://www.cbica.upenn.edu/BraTS18/lboardValidation.html>

for all of the 3 segmentation labels (ET, WT, TC), given 390 teams registered for the BRATS 2018 challenge.

Table 4. Mean Dice and Hausdorff measurements of the proposed segmentation method on BraTS 2018 validation set. EN - enhancing tumor core, WT - whole tumor, TC - tumor core.

	Dice			Hausdorff (mm)		
	ET	WT	TC	ET	WT	TC
Single Model	0.8145	0.9042	0.8596	3.8048	4.4834	8.2777
Single Model + TTA	0.8173	0.9068	0.8602	3.8241	4.4117	6.8413
Ensemble of 10 models	0.8233	0.9100	0.8668	3.9257	4.5160	6.8545

4 Conclusion

We described a segmentation network for brain tumor segmentation from multimodal 3D MRIs, where variational autoencoder branch is added to further regularize the decoder in the presence of limited training data. Preliminary results on BRATS 2018 validation datasets are 0.8233, 0.9100, 0.8668 average Dice for enhanced tumor core, whole tumor and tumor core respectively.

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), <https://www.tensorflow.org/>, software available from tensorflow.org
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., John Freymann, K.F., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. The Cancer Imaging Archive (2017), <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., John Freymann, K.F., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. The Cancer Imaging Archive (2017), <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Scientific data 4 (9 2017)

5. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. arXiv:1802.02611 (2018)
6. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1800–1807 (2017)
7. Doersch, C.: Tutorial on variational autoencoders (2016), <http://arxiv.org/abs/1606.05908>
8. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: Proceedings of the International Conference on Computer Vision (ICCV) (2017)
9. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: European Conference on Computer Vision (ECCV) (2016)
10. Kamnitsas, K., W. Bai, E.F., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., Glocker, B.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: International Conf. on Medical Image Computing and Computer Assisted Intervention. Multimodal Brain Tumor Segmentation Challenge (MICCAI, 2017). LNCS
11. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: The International Conference on Learning Representations (ICLR) (2014)
12. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E.R., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M.T., Sarikaya, D., Schwartz, L.H., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekeley, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G.B., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans. Med. Imaging* **34**(10), 1993–2024 (2015)
13. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: Fourth International Conference on 3D Vision (3DV) (2016)
14. Ronneberger, O., P.Fischer, Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). LNCS, vol. 9351, pp. 234–241. Springer (2015)
15. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: International Conf. on Medical Image Computing and Computer Assisted Intervention. Multimodal Brain Tumor Segmentation Challenge (MICCAI, 2017). LNCS
16. Wu, Y., He, K.: Group normalization. In: European Conference on Computer Vision (ECCV) (2018)

ESPNet for Glioma Segmentation

Nicholas Nuechterlein and Sachin Mehta

The University of Washington, Seattle, WA 98195, USA
nknuecht@cs.washington.edu, sacmehta@uw.edu

Abstract. Accurate segmentation of brain tumors is clinically relevant in diagnoses, prognoses, and treatment of glioma [15]. Early machine learning brain tumor segmentation efforts relied heavily on random forests [1, 2]. More recently, 2D and 3D CNNs trained on image patches sampled from 2D slices of magnetic resonance (MR) volumes have become the model of choice [14]. In particular, variants of Ronneberger et al.’s U-Net have performed well on this task [7]. ESPNet is a fast and efficient network for semantic segmentation, benchmarked on the Cityscape and Mapillary datasets, which loosely inherits Ronneberger et al.’s U-Net network structure with convolutional blocks swapped out for ESP (efficient spatial pyramid) modules [17]. We extend ESPNet to 3D medical imaging data in this report. We train in two stages at full resolution on four-channel, multimodal MR volumes to predict a voxelwise four-class segmentation of glioma tumors. We first find the tumor within the brain, then the tumor core within the tumor, and then the necrotic core within the tumor core. Our efforts follow an in-depth investigation of the underlying glioma MR data. Our preliminary results are encouraging (whole tumor dice score 0.889), but work remains to be done on the individual classes.

Keywords: Glioma · GBM · ESPNet · Image reinforcement · Semantic segmentation.

1 Introduction

Glioblastoma (GBM), a grade IV glioma, is the most common and most aggressive primary brain tumor. The average GBM patient lives only 12-15 months post diagnosis, a number which has only grown by three months in the last forty years [15]. Due to GBMs highly heterogeneous appearance, extent, and shape, segmentation of brain tumors in MR volumes is one of the most challenging tasks in neuroradiology [15]. This is compounded by the sparsity of data and the variation incurred by differing scanner models and manufacturers, imaging sites, variation in clinical standards and protocols, and the noise introduced by the movement of patients’ heads during scans. At every clinical visit, GBM patients generally receive standard of care FLAIR, post-contrast T1-weighted (T1ce), T2, and T1 MR sequences, each of which is described by a distinct volume. These sequences give complementing information about the tumor extent and composition.

Automated brain tumor segmentation also ranks among the most difficult problems in medical image analysis. The notion that successful training of deep networks requires massive amounts of data is widely held. Not only are MR scans scarce, they are high dimensional (e.g. $240 \times 240 \times 155 \times 4$) and contain high class imbalances (e.g. $\geq 95\%$ background class). Thus, naive models are predisposed to exhibit extreme background bias. Segmentation of brain tumors is also difficult due to heterogeneities in tumor intensity, which we attempt to correct during preprocessing.

In similar biomedical domains, patchwise approaches have helped address problems of data shortages and dimensionality. Ciresan et al. proposed a sliding-window method to segment electron microscopic images of the brain, which both localized the problem and exaggerated the dataset [7, 14]. Ronneberger et al.’s 2D encoder-decoder network, U-Net, outperformed Ciresans method. U-Net is a fully convolutional network where the traditional pooling operations in the contracting (encoding) path are replaced by upsampling operations in the symmetric expanding (decoding) path. Skip connections are passed from encoding blocks on the contracting path to same-level decoding blocks in the expanding path [7].

While some success has been reached using 2D fully convolutional networks (FCNs) like U-Net, these models ignore crucial 3D spatial context, which is undesirable given that most clinical imaging data is volumetric. However, even among 3D FCNs such as DeepMedic, a previous winner of the BraTS competition, fine spatial information is discarded in pooling [9]. This motivates our interest in U-Net’s skip connections and, in particular, the architecture of Milletari et al.’s 3D extension of U-Net, V-Net. V-Net benchmarked well on the “PROMISE2012 challenge, where it gave impressive segmentations of MR prostate scans after training on only 50 examples [8].

ESPNet is a faster, more efficient take on U-Net’s encoder-decoder architecture [17]. We seek to extend and benchmark ESPNet on 3D medical imaging data. We conduct extensive studies on the BraTS 2018 data, locating most of the segmentation signal in the FLAIR and T1ce sequences [1–4]. We train three two-class segmentation networks. The first distinguishes between tumor and background, the second between tumor and tumor core, and the third between tumor core and necrotic tissue. The difference between the entire tumor and the tumor core is assumed to be edema, and the difference between the tumor core and the necrotic core is assumed to be the enhancing tissue class.

2 Methods

2.1 Preprocessing

We trained the first stage of our network using only min-max normalization. In the second and third stage we perform histogram equalization. We found histogram equalization particularly useful in low grade gliomas (LGG) where the clearly defined structure in enhancing and necrotic tissue found in high grade

gliomas (HGG) is generally absent. Before stage one, we removed any border slices that contained no tumor in any training samples. This helped greatly in dimensionality reduction and allowed us to increase batch size during training.

2.2 Network Structure

The primary distinction between the encoder portion of ESPNet and a traditional convolutional encoder is that we use ESP (efficient spatial pyramid) blocks in the place of convolutional blocks after the first layer [17]. Using ESP blocks reduces the number of network parameters and allows us to deepen the network. We also replace pooling operations with strided ESP convolutional layers to learn non-linear downsampling operations. Additionally, we send feature maps forward via long-range concatenation connections from the first ESP module in an ESP block to the last ESP module in the block; this improves information flow within the network. These connections are shown as solid black arrows in the encoder in Figure 1.

When we attach a light-weight decoder, we share compressed representations of the feature maps in the encoder with same-level feature maps in the decoder via skip-connection concatenation. Compression is achieved by applying $1 \times 1 \times 1$ convolutions to learn a lower dimensional representation of the feature maps to be passed to the decoder. The skip connections and the $1 \times 1 \times 1$ convolutions are shown with dotted green arrows in Figure 1. Skip-connections allow fine details lost in downsampling in the encoder to be recovered in the decoder, which gives the segmentation maps a granularity simple interpolation cannot achieve. The decoder uses $2 \times 2 \times 2$ deconvolution kernels to upsample the encoder output.

After initially downsampling with a standard strided convolution, we down-sample again with a strided ESP module. Such dramatic downsampling is done to reduce memory overhead. We add two ESP modules and downsample again with a strided ESP module. We add five more ESP modules before we reduce the outputted feature maps into two channels, corresponding to the tumor and the background class, using two pointwise convolutions; this is shown by the thick black arrow at the end of the encoder in Figure 1. We then upsample and pass the output through a softmax to produce our segmentation mask. We use the same architecture for all network stages.

2.3 Training

We randomly partitioned the dataset into a training set and a validation set using an 80:20 split (228:57). We trained the encoder and decoder in separate phases. To investigate the contributions of each modality during data exploration, we trained encoders on each of the four modalities. The encoder trained on the FLAIR sequences performed best on the whole tumor task, and the encoder trained on the T1ce sequences performed best on the enhancing and necrotic tissue classes.

We used mIOU for our loss function instead of cross entropy for empirical reasons as we and others have observed [10]. We weight our mIOU loss to address

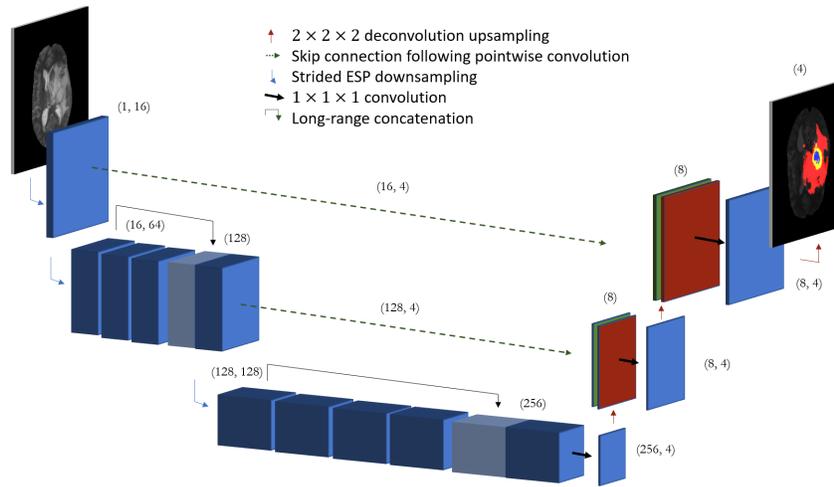


Fig. 1. ESPNet. The encoder is shown on the left; the light-weight decoder is shown on the right. Parentheses give the channel dimensions of incoming and outgoing feature maps. Light-blue feature maps indicated concatenation by long-range connections. Arrows are defined in the legend. The above is a single instances of ESPNet; we train three instances of ESPNet, one for each stage in our pipeline.

severe class imbalance. Random crops and random flips were used heavily for data augmentation.

We implemented our models in PyTorch. When training at full resolution over all modalities, we used a batch size of four on an NVIDIA Titan X. We trained for 300 epochs. We found that the optimizer Adam outperformed SGD with momentum [18]. We experimented with learning decay and settled on a learning rate of $10e^{-4}$, which we decreased to $10e^{-5}$ after 200 epochs. Code for this adaptation of ESPNet will be available at <https://github.com/nknuecht/GBMNet>.

3 Results

We train on the Multimodal Brain Tumor Segmentation Challenge (BraTS) 2018, which provides 285 multi-institutional pre-operative multimodal MR tumor scans, each consisting of T1, post-contrast T1-weighted (T1ce), T2, and FLAIR volumes [1–4]. Each case is annotated with the following voxel labels: enhancing tumor, peritumoral edema, background, and necrotic core and non-enhancing tumor [6]. Necrotic core and non-enhancing tumor share a single label. We withheld a random validation set of 57 volumes. These data are co-registered to the standard MNI anatomical template, interpolated to the same resolution, and skull-stripped. Ground-truth segmentations are manually drawn and approved by expert neuroradiologists [6].

Our preliminary experiments show that our three-stage application of ESPNet shows competitive results for whole tumor segmentation. Our exploratory experiments show that networks trained on the FLAIR modality consistently excel on the edema compartment and the merged whole tumor class when compared to networks trained on other modalities. However, networks trained on all modalities have the overall performance, despite incurring more memory overhead that constrains batch size.

ESPNet	Dice Score	Sensitivity	Specificity
Whole Tumor	0.889	0.9457	0.9971
Tumor Core	0.590	0.7583	0.9989
Enhancing Tumor	0.533	0.6346	0.9985
Edema	0.638	0.9457	0.9971

Table 2. Results obtained on our partitioned validation set.

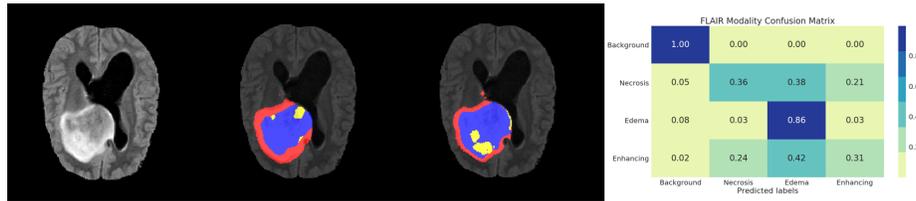


Fig. 2. Overlay of segmentation mask on a FLAIR image. FLAIR shows edema compartment (red) best. Water content outside of the cerebrospinal fluid (CSF) is bright which prevents spurious edema predictions in the ventricles. Left: input; Center: prediction; Right: ground truth.

4 Discussion

Our data exploration findings are mostly in line with clinical expectations. Networks trained on FLAIR volumes likely perform best because FLAIR sequences are the modality most sensitive to water, which is disproportionately present in edema. FLAIR sequences are also based on T2 magnetic gradients, which may explain the heightened performance achieved by T2-trained networks on the edema compartment. Edema surrounds the tumor and thus the edema compartment and the entire tumor share the same outer boundary: this may explain why FLAIR-trained networks perform better on whole tumor segmentation than the

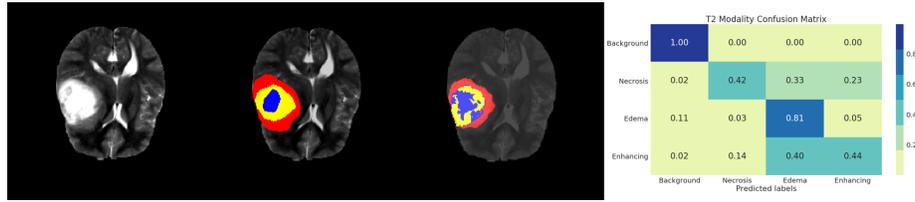


Fig. 3. Overlay of segmentation mask on a T2 image. Edema is bright on T2 scans as is the CSF fluid in the ventricles. T2 captures edema well, but edema can also wash out necrotic areas and sometimes be confused for CSF. Left: input; Center: prediction; Right: ground truth.

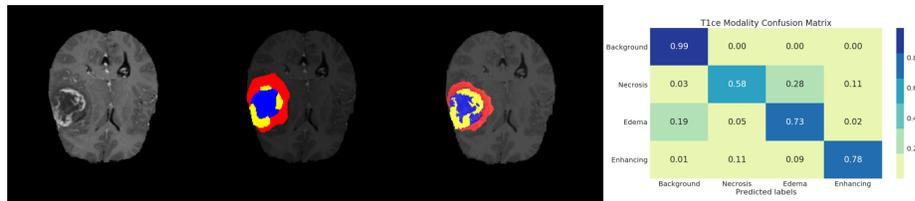


Fig. 4. Overlay of segmentation mask on a T1ce image. T1ce shows necrosis compartment (blue) best as compared to water-sensitive scans such as FLAIR and T2 (shown above). T1 lacks the contrast to pick up fine details in enhancing tissue, which is illuminated by the gadolinium contrast agent injected during the T1ce sequence. Left: input; Center: prediction; Right: ground truth.

networks trained on other modalities. The T1ce sequence is the modality most often consulted by neuro-oncologists because it enhances the contrast agent leakage around the malignant tumor.

A possible explanation for our poor performance on individual tumor compartments is that it is not always the case that the necrotic core is surrounded by enhancing tumor or that the tumor core (necrotic and enhancing tissue) is completely surrounded by edema, which we assume when using our staged model. We also find LGG tumors more difficult to segment, likely because they are more diffuse than HGG tumors and make up a smaller portion of the dataset.

5 Conclusion

Segmentation of glioma tumors is relevant to clinical diagnoses, prognoses, and treatment in neuro-oncology [15]. This paper extends ESPNet, a fast and efficient network designed for vanilla 2D semantic segmentation, to 3D data in the medical imaging domain. We followed a three stage training scheme to build a four-class voxelwise tumor compartment segmentation model. We achieved a dice score of 0.889 on the whole tumor class, but work remains to be done on the individual classes. Adding a CRF on top of our final output should improve our results as it for DeepMedic [9]. Better data normalization and hyperparameter tuning

may also improve our performance. Training on the entire dataset will also be helpful.

References

1. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017. DOI:10.7937/K9/TCIA.2017.KLXWJJ1Q
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection"
5. Bauer, S., Fejes, T., Slotboom, J., Wiest, R., Nolte, L.P., Reyes, M.: Segmentation of brain tumor images based on integrated hierarchical classification and regularization. *Proceedings MICCAI-BRATS* (2012)
6. Menze, B.H., Geremia, E., Ayache, N., Szekely, G.: Segmenting glioma in multi-modal images using a generative-discriminative model for brain lesion segmentation. *Proceedings MICCAI-BRATS* (2012)
7. Ronneberger O, Fischer P, Brox T.: U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* 2015 Oct 5 (pp. 234-241). Springer, Cham.
8. Milletari, F., Navab, N., Ahmadi, S.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. *CoRR* abs/1606.04797 (2016)
9. Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A., Criminisi, A., Rueckert, D., Glocker, B.: Deepmedic on brain tumor segmentation. *Athens, Greece Proc. BRATS-MICCAI* (2016)
10. K. Kamnitsas, W. Bai, E.Ferrante, S.G. McDonagh, M. Sinclair, N.Pawlowski, M. Rajchl, M. C. H. Lee, B. Kainz, D. Rueckert, B. Glocker, Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation, *CoRR* vol. abs/1711.01468, 2017. [Online]. Available: <http://arxiv.org/abs/1711.01468>
11. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection"

12. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection"
13. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection"
14. Cireşan, D.C., Gambardella, L.M., Giusti, A., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: NIPS. pp. 2852-2860 (2012)
15. [14] Fink JR, Muzi M, Peck M, Krohn KA. Multimodality brain tumor imaging: MR imaging, PET, and PET/MR imaging. *J Nucl Med.* 2015;56:1554-1561. doi: 10.2967/jnumed.113.131516
16. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC. "N4ITK: Improved N3 Bias Correction", *IEEE Trans Med Imaging.* 29(6), 1310-1320 (2010)
17. Sachin Mehta, Mohammad Rastegari, Anat Caspi, Linda Shapiro, and Hannaneh Hajishirzi (2018). ESPNet: Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation. arXiv preprint arXiv:1803.06815.
18. Diederik P. Kingma and (2014). Adam: A Method for Stochastic Optimization. CoRR, abs/1412.6980.

Ensemble Learning Models for Accurate Segmentation of Brain Tumor and Prediction of Patient Treatment Outcome: BraTS'2018 Challenge

Alexander F. I. Osman^[0000-0002-1286-475X]

Medical Physics Department, Al-Neelain University, Khartoum 11121, Sudan
alexanderfadul@yahoo.com

Abstract. Machine learning approaches for tumor segmentation and prediction of treatment outcome are currently representing a trend research in the radiation oncology field. The purpose of this study is to propose ensemble machine learning methods for glioma tumor segmentation and prediction of patient's survival. BRATS'2018 multimodal magnetic resonance imaging (MRI) scans and clinical data of patients pathologically confirmed with glioma tumor were used for this purpose. Ensemble model of multiple algorithms was developed for full tumor structures segmentation and predicting the patient-specific expected survival. The ensemble models were partially trained BRATS'2018 training dataset. The segmentation model performance was and qualitatively evaluated, the survival prediction model was evaluated with the accuracy and area under the curve (AUC) metrics. The preliminary segmentation results during the training phase qualitatively resemble the ground truth labels. For survival prediction, AUC of 0.72, 0.58, and 0.63 was achieved in predicting short, mid, and long-survivors groups with over all accuracy of 55%. Both models are still under developing but they look promising to provide more accurate segmentation and patient's treatment outcome predictions.

Keywords: Glioma Tumors, Image Segmentation, Ensemble Learning, Treatment Outcome.

1 Introduction

Gliomas are the most common primary brain malignancies [1]. Glioma tumors are categorized clinically into high-grade and low-grade. High-grade glioma (HGG) tumors are invasive and aggressively grow rapidly leading to the patient's death i.e. glioblastoma multiform (GBM) is the most common one with a median survival rate of two years or less [2-3]. On the other hand, low-grade glioma (LGG) tumors are slow-growing, with a life expectancy of several years. Treatment protocols for gliomas generally start with patient undergo through surgical procedure to remove the tumor then followed by irradiation of the tumor bed [4]. The goal of three-dimensional (3D) conformal radiation therapy is to irradiate the tumor volume while limiting damage to the surrounding normal tissues. Achieving this goal requires accu-

rate determination of 3D treatment volumes. Radiation oncologists traditionally model the brain treatment target through a time-intensive manual procedure involving the outlining of the gross tumor volume (GTV) on numerous 2D imaging “slices” using either computed tomography (CT) or magnetic resonance imaging (MRI) [5]. Recently, the search for improvements in target volume definition methodology has concentrated on improved imaging modalities [6, 7]. It has been demonstrated that MRI is more sensitive than CT in both lesion detection and in the margin delineation of gliomas [8]. However, limitations remain in the delineation of tumor volumes and in the ability of different radiation oncologists to reproduce consistent results [8]. Although the technology for conformal radiation treatment planning has developed to a high level of accuracy, the definition of the tumor GTV is still based on time-intensive, highly subjective “inter-observer variations” manual outlining [9-10]. Manual outlining is the type of process that should be an excellent candidate for automation through the development of a computerized segmentation system.

Automatic segmentation of MR images offers the potential to accurately define complex treatment volumes, to speed the contouring process in radiation therapy treatment planning, and to provide a standardized reproducible measurement protocol that can be employed by geographically diverse facilities and physicians in treating brain tumors. Several algorithms [11-17] have been developed using different machine learning (ML) methods for computational segmentation of brain tumors in MRIs. However, it still remains challenging for reasons such as availability of big data of multimodal MR images with their ground truth data, etc. Multimodal Brain Tumor Segmentation (BRATS) challenges [17-18] have focused on brain tumor segmentation algorithms and established a framework for benchmarking and evaluating different segmentation algorithms. BRATS is making repository of a large “standard” dataset with accompanying delineations of the relevant tumor sub-regions, and patient overall survival data. Over the past few years, discriminative probabilistic methods [13-14] that rely on a random forest classifier and convolutional neural networks [10,11,15] have shown as the most successful methods for segmentation of brain tumors. Last year, a ML method based on ensembles of multiple models for brain tumor segmentation won the first place in the Brain Tumor Segmentation (BRATS) 2017 competition. Fusing segmentations from different algorithms have shown indications for better performance than the best individual algorithm applied to the same task [19]. Ensemble learning [20] helps improve machine learning results by combining several models. This approach allows the production of better predictive performance compared to a single model. The main purpose of an ensemble is maximizing individual accuracy and diversity.

Machine learning algorithms for predicting radiotherapy outcomes (e.g., survival, treatment failure, toxicity) are receiving much attention in literature, for example, in decision support systems for precision medicine [24,25]. Patient-specific overall survival (OS) prediction using the patient clinical data and tumor radiomic/imaging features could serve as a survival-predictor and hence accurate prognosis for glioma patients would provide essential guidelines for their treatment planning. Prediction of GBM patient overall survival (OS) via integrative analyses of radiomic features could provide reliable information of patient’s status. The central hypothesis of radiomics is

that quantitative features measured from routine medical images are related to the tumor phenotype and provide important information for personalized medicine. When using quantitative image features for radiomics, the initial workflow involves segmenting an image, applying image processing, fine-tuning parameters for the radiomics features needed to be measured, and extracting them. Each step is dependent on image modality and research goals. A variety of statistical tools are prevalent and useful for understanding the dynamics of radiomics features and relating them to clinical endpoints in medical research. When the primary outcomes are time-to-events, such as progression-free survival and overall survival, the conventional survival analysis and relevant methods can be applied as a benchmark to evaluate the goodness of fit of machine learning techniques. This paper presents developing an ensemble model of multiple ML algorithms for fully automated glioma tumor segmentation process in magnetic resonance imaging (MRI) scans and patient-specific survival expected period. The survival model is estimating the glioblastoma multiforme (GBM) tumors treatment outcome using radiomic signatures determined via machine learning methods and multi-modal MRIs image processing.

2 Materials and Methods

2.1 Data

Multimodal MR images and patient's clinical data were provided by BRATS'2018 challenge [19, 21-23]. The data were for patient clinically diagnosed with glioma tumors and pathologically confirmed. The provided MR images were clinically-acquired in a routine manner in multi-institutions using different scanners and various protocols.

For the segmentation task, MRI dataset of 285 patients were provided as training dataset 66 patients as validation dataset. MR images consisting the training dataset were provided with their ground truth segmentation labels of various glioma sub-regions. The manually segmented labels were performed by experts following the same annotation protocol, and their annotations were revised and approved by board-certified neuroradiologists. Annotation labels included were the GD-enhancing tumor (ET), the peritumoral edema (ED), and the necrotic and non-enhancing tumor (NCR/NET). Each patient multimodal imaging data was consisted of four MRI modalities precisely; 1) T1-weighted (highlights fat locations), 2) post-contrast/gadolinium T1c-weighted (taken after the injection of the contrast agent gadolinium), 3) T2-weighted (highlights water locations), and 4) T2-Flair (Fluid Attenuated Inversion Recovery, an MR imaging technique that produces images similar to T2-weighted images, but with free water suppressed). They were co-registered with aligning the volumes of the four MRI modalities to the same anatomical template, and interpolated to the same resolution of $1 \times 1 \times 1 \text{ mm}^3$ and skull-stripped to correlate all four images for each patient.

The provided clinical OS data and patient age were included with correspondences to the pseudo-identifiers of the GBM/HGG imaging data. For the survival prediction

task, clinical data of 163 (accompanied with the ground truth data) and 53 patients were provided for training and validating the algorithm, respectively.

2.2 The model

Theory of ensemble method. An ensemble is itself a supervised learning algorithm, because it can be trained and then used to make predictions. The trained ensemble, therefore, represents a single hypothesis. This hypothesis, however, is not necessarily contained within the hypothesis space of the models from which it is built. Thus, ensembles can be shown to have more flexibility in the functions they can represent. This flexibility can, in theory, enable them to over-fit the training data more than a single model would, but in practice, some ensemble techniques (especially bagging) tend to reduce problems related to over-fitting of the training data. Ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone [20]. Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model. There are many types of ensemble methods. They include *BAGGing*, or *Bootstrap AGGregating* and *Random Forest*. *BAGGing* gets its name because it combines *Bootstrapping* and *Aggregation* to form one ensemble model. Given a sample of data, multiple bootstrapped subsamples are pulled. A Decision Tree is formed on each of the bootstrapped subsamples. After each subsample Decision Tree has been formed, an algorithm is used to aggregate over the Decision Trees to form the most efficient predictor. Random Forest Models can be thought of as *BAGGing*, with a slight tweak. When deciding where to split and how to make decisions, *BAGGed* Decision Trees have the full disposal of features to choose from. Therefore, although the bootstrapped samples may be slightly different, the data is largely going to break off at the same features throughout each model. In contrary, Random Forest models decide where to split based on a random selection of features. Rather than splitting at similar features at each node throughout, Random Forest models implement a level of differentiation because each tree will split based on different features. This level of differentiation provides a greater ensemble to aggregate over, ergo producing a more accurate predictor.

Ensemble of different models. We used different methods of assembling to develop our segmentation and survival prediction models.

2.3 Algorithm Evaluation

The performance of each participating algorithm has to be evaluated on the online evaluation platform (CBICA's IPP) at University of Pennsylvania (<https://ipp.cbica.upenn.edu/>) for its performance in segmentation and survival prediction tasks. At the moment, our model's performance was evaluated qualitatively. BRATS'2018 requires that the segmentation model to be evaluated with "Dice score" and the "Hausdorff distance" evaluation metrics. In addition, the "Sensitivity" and "Specificity" metrics have also included. For survival prediction task, accuracy is

used for classification of subjects as long-survivors (e.g., >15 months), short-survivors (e.g., <10 months), and mid-survivors (e.g. between 10 and 15 months) for predicted survival status for subjects with resection status of gross total resection). For regression the model is evaluated with the mean and median square error of survival time predictions.

3 Results

The predicted results of glioma segmentation are presented here. The qualitative results represent the whole tumor are shown below (Fig. 1) on the training dataset.

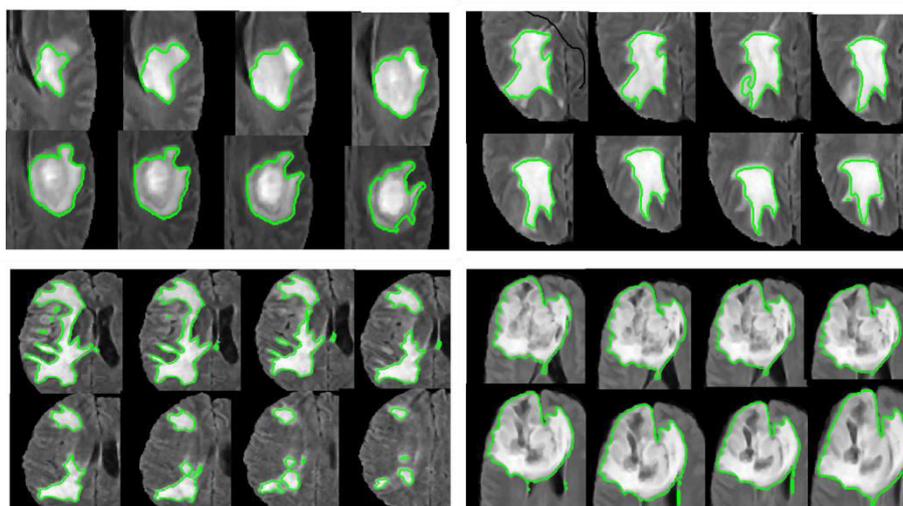


Fig. 1. The predicted segmentation labels of glioma on MR images on the training dataset (four patients MRIs).

We could not report the quantitative results here as well as the segmentation results of the other glioma sub-structures.

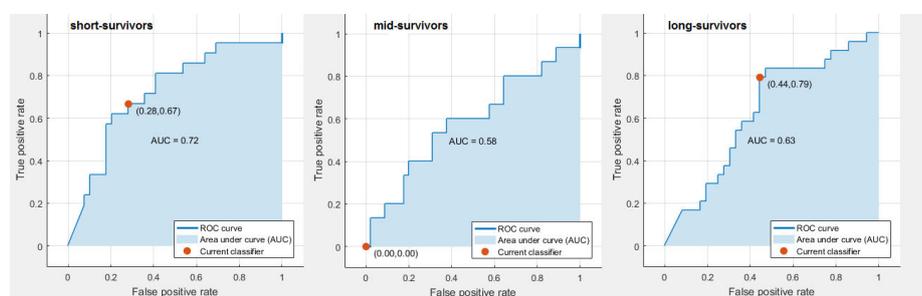


Fig. 2. The receiver operating characteristic (ROC) curve for SVM classification of survival class prediction in the training phase.

The above figure (Fig. 2) show the ROC curve for short, med, and long-survivors class predictions. The curve shows true positive rate versus false positive rate for the trained classifier. The overall accuracy is 55% for a SVM linear model.

4 Discussion

The goal of any machine learning problem is to find a single model that will best predict the required outcome. Rather than making one model and hoping this model is the best/most accurate predictor we can make, ensemble methods combine multiple different models and average those models to produce one final model. Therefore, for this reason we changed our mind to develop an ensemble model for BRATS'2018 competition rather than enhancing our last year (2017) contribution which was a single model based-on support vector machine approach. An ensemble model of multiple algorithms architecture was developed for glioma segmentation and patient-specific survival prediction. The model is successfully built and currently in validation phase.

The segmentation model is an ensemble of different classifiers. The preliminary segmentation results look very good qualitatively. However, it's still under developing and Dice and the results of the sub-structures are not reported here. With recently introduced MR-guided radiation therapy (MR-guided RT), automated glioma tumor segmentation may have an effective role in Adaptive RT and expediting the clinic work flow.

When using quantitative image features for radiomics, the initial workflow involves segmenting an image, applying image processing, fine-tuning parameters for the radiomics features you wish to measure, and extracting them. A variety of statistical tools are prevalent and useful for understanding the dynamics of radiomics features and relating them to clinical endpoints in medical research. When the primary outcomes are time-to-events, such as progression-free survival and overall survival, the conventional survival analysis and relevant methods can be applied as a benchmark to evaluate the goodness of fit of machine learning techniques. The preliminary results of survival prediction during the training phase is very good also but we could not include the results here. We are refining the model to obtain more accurate survival prediction. The SVM linear classification model was trained on the BRATS'2018 survival data and validated with k-fold cross-validation (k=5).

5 Conclusions

An ensemble machine learning algorithm is proposed for brain tumor segmentation and patient's survival. BRATS'2018 multimodal MR images and clinical data of patients diagnosed with glioma tumor were used for this purpose. An ensemble model of multiple algorithms was developed for full tumor structures segmentation and predicting the patient-specific expected survival. The ensemble model was trained on part of

the BRATS'2018 training dataset. Its performance was quantitatively and qualitatively evaluated. The preliminary segmentation results during the training phase look qualitatively resemble. The model is still under developing but it looks promising to provide more accurate segmentation and patient's survival predictions.

References

1. Holland, E.C.: Progenitor cells and glioma formation. *Current Opinion in Neurology* 14(6), 683–688 (2001).
2. Ohgaki, H., Kleihues, P.: Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. *Journal of Neuropathology & Experimental Neurology* 64(6), 479–489 (2005).
3. Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavanee, W.K.: WHO classification of tumours of the central nervous system. 4th edn. WHO/IARC, Lyon, France (2007).
4. Mazzara, G.P., Velthuizen, R.P., Pearlman, J.L., Greenberg, H.M., Wagner, H.: Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation. *International Journal of Radiation Oncology Biology Physics* 59(1), 300–312 (2004).
5. Morris, D.E., Bourland, J.D., Rosenman, J.G., et al.: Three-dimensional conformal radiation treatment planning and delivery for low- and intermediate-grade gliomas. *Seminars in Radiation Oncology* 11(2), 124–137 (2001).
6. Henkelman, R.M.: New imaging technologies: Prospects for target definition. *International Journal of Radiation Oncology Biology Physics* 22(2), 251–257 (1992).
7. Jansen, E.P., Dewit, L.G., Van Herk, M., et al.: Target volumes in radiotherapy for high-grade malignant glioma of the brain. *International Journal of Radiation Oncology Biology Physics* 56(2), 151–156 (2000).
8. Caudrelier, J.M., Vial, S., Gibon, D., et al.: MRI definition of target volumes using fuzzy logic method for three-dimensional conformal radiation therapy. *International Journal of Radiation Oncology Biology Physics* 55(1), 223–233 (2003).
9. Yamamoto, M., Nagata, Y., Okajima, K.: Differences in target outline from CT scans of brain tumours using different methods and different observers. *Radiotherapy & Oncology* 50(2), 151–156 (1999).
10. Pitkanen, M.A., Holli, K.A., Ojala, A.T., et al.: Quality assurance in radiotherapy of breast cancer—variability in planning target volume delineation. *Acta Oncology* 40(1), 50–55 (2001).
11. Kamnitsas, K., Ledig, C., Newcombe, V.F., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis* 36, 61–78 (2017).
12. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., et al.: 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic resonance imaging* 30(9), 1323–1341 (2012).
13. Zikic, D., Glocker, B., et al.: Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7512, pp. 369–376. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33454-2_46
14. Le Folgoc, L., Nori, A.V., Ancha, S., Criminisi, A.: Lifted auto-context forests for brain tumour segmentation. In: Alessandro, C., et al. (eds.) BrainLes 2016. LNCS, vol. 10154, pp. 171–183. Springer, Heidelberg (2016).

15. Urban, G., Bendszus, M., Hamprecht, F., Kleesiek, J.: Multi-modal brain tumor segmentation using deep convolutional neural networks. In: MICCAI-BraTS 2014, pp. 31–35 (2014).
16. Osman, A. F. I.: Automated Brain Tumor Segmentation on Magnetic Resonance Images (MRIs) and Patient Overall Survival Prediction using Support Vector Machines. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 435–449. Springer, Cham (2018).
17. Kamnitsas K., Bai, W., Ferrante, E., et al.: Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds) BrainLes 2017. LNCS, vol. 10670, pp. 450–462. Springer, Cham (2018).
18. Archip, N., Jolesz, F.A., Warfield, S.K.: A validation framework for brain tumor segmentation. *Academic Radiology* 14(10), 1242–1251 (2007).
19. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, G., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging* 34(10), 1993–2024 (2015). DOI: 10.1109/TMI.2014.2377694
20. Dietterich, T.G. Ensemble Methods in Machine Learning. In: Kittler, J., and Roli, F. (eds.) First International Workshop on Multiple Classifier Systems, LNCS, pp. 1–15. Springer, Verlag (2000).
21. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
22. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
23. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
24. Lambin, P., van Stiphout, R.G., Starmans, M.H., et al.: Predicting outcomes in radiation oncology—multifactorial decision support systems. *Nature Reviews Clinical Oncology* 10(1), 27–40 (2013).
25. Lambin, P., Roelofs, E., Reymen, B., et al.: Rapid learning health care in oncology’ – An approach towards decision support systems enabling customised radiotherapy. *Radiotherapy & Oncology* 109(1), 159–164 (2013).

Automatic Brain Tumor Segmentation using U-Net based 3D Fully Convolutional Network

Anmol Popli, Manu Agarwal, and G.N. Pillai

Indian Institute of Technology Roorkee, India
apopli@ee.iitr.ac.in

Abstract. Gliomas are the most common primary brain malignancies, with different degrees of aggressiveness, variable prognosis and various heterogeneous histological sub-regions. Due to this highly heterogeneous appearance and shape, segmentation of brain tumors in multimodal MRI scans is one of the most challenging tasks in medical image analysis. In this paper, we propose a U-Net based 3D Fully Convolutional Neural Network to segment tumor subregions and background from multimodal Magnetic Resonance Imaging (MRI) scans of the brain. The tumor subregions are three hierarchical regions: whole tumor, tumor core and enhancing tumor core. Our 3D FCN architecture has been derived from the U-Net architecture and optimized in order to maximize performance on three-dimensional multimodal MRI data. In addition, we employ N4ITK bias field correction as part of our preprocessing stage to rectify bias and intensity inhomogeneities in MRI images. Our proposed method has achieved very good results on BraTS 2018 training and validation datasets. The Dice scores on training set are 0.9069, 0.8683 and 0.7265 for whole tumor, tumor core and enhancing tumor core respectively. The corresponding scores on validation set are 0.8941, 0.8248 and 0.7439.

Keywords: Brain tumor · Deep Learning · U-Net.

1 Introduction

Automatic segmentation of brain tumors and substructures is an alluring concept. It has the potential to guide for better diagnosis, surgical planning and treatment assessment of brain tumors by providing us with more accurate information such as relevant tumor states and volume of its subregions. As the nature of tumour is irregular, development of automated methods is still a challenging process.

If done using manual work, segmentation is a highly time consuming and subjective task. In this paper we present our robust segmentation algorithm based on 3D convolutional neural networks inspired from U-Net architecture. In this segmentation task there is a considerable challenge because of variable size, shape and localization of brain tumor among patients. The brain tumor segmentation challenge (BraTS) aims to propagate the development of state of the art

methods for tumor segmentation by providing a large dataset of annotated low grade gliomas (LGG) and high grade glioblastomas(HGG).

In contrast to traditional methods, deep learning doesn't depend on the creation of handmade filters or features to differentiate tumor from normal brain anatomy. Instead of that raw image intensities are provided as input to neural network architecture to calculate output signal. The involvement of non linearities and multiple degrees of freedom allow our algorithms to learn complex level patterns and abstract high level information. Many algorithms are using 2D CNNs till now which do not utilize the volumetric information. With increasing awareness, 3D CNNs are proving to be effective on these tasks with a little increase in computational complexity.

The U-Net architecture [8] has proved to be quite effective at segmentation tasks. We propose a U-Net based 3D Fully Convolutional Neural Network for automatic segmentation of tumor subregions from brain MR images. We have adapted the U-Net architecture to our 3D MRI data by making certain design choices in accordance with our data. Whereas other kinds of FCNs manually combine feature maps at different levels of the network, the U-Net, owing to its configuration, has the advantage of intrinsically recombining different scales throughout the network. We also implement bias correction and normalization to make the MRI data suitable for being fed to our network.

2 Methods

2.1 Data Preprocessing

Data We have trained and tested our network on the BraTS 2018 [9, 1–3] dataset. The training set comprises of images from 285 patients (210 HGG and 75 LGG). The BraTS 2018 validation set contains images from 66 patients with brain tumors of unknown grade. Corresponding to each patient, four modalities of MR images were generated: T1, T1c, T2 and FLAIR. The provided data have been co-registered to the same anatomical template, interpolated to the same resolution (1 mm³) and skull-stripped. All the imaging datasets have been segmented manually, by one to four raters, following the same annotation protocol, and their annotations were approved by experienced neuro-radiologists. Annotations comprise the GD-enhancing tumor (ET label 4), the peritumoral edema (ED label 2), and the necrotic and non-enhancing tumor core (NCR/NET label 1)

Bias Correction As the first step in data pre-processing, we perform N4ITK bias field correction [10] on the MRI images. MR scans often display intensity non-uniformities due to variations in the magnetic field. So, one part of an image might appear lighter or darker when visualized, solely because of variations in the magnetic field. The map of these variations is called the bias field. The bias field can cause problems for a classifier as the variations in signal intensity are not due to any anatomical differences. N4ITK Bias Field Correction attempts to correct the bias field by extracting it from the image.

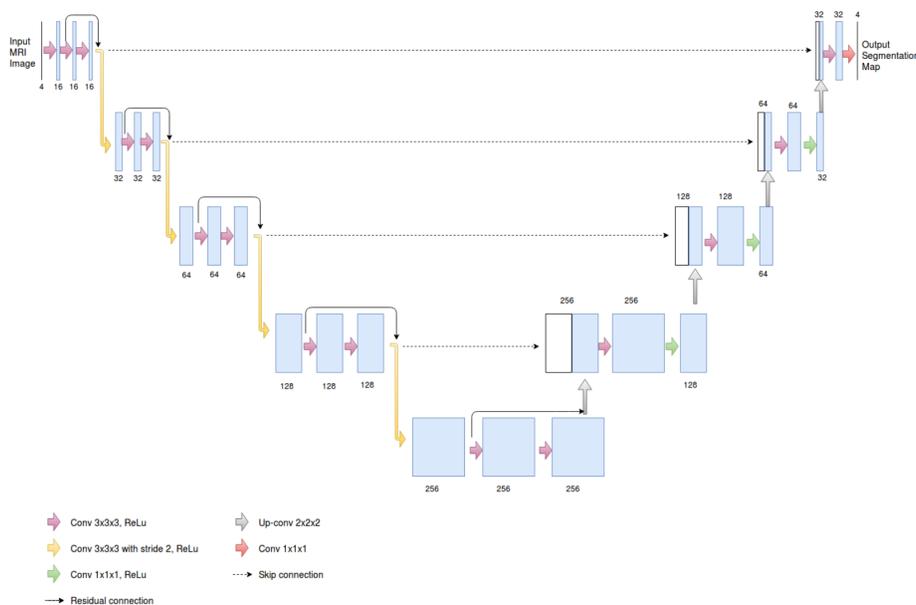


Fig. 1. Network architecture. Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top or bottom of the box. White boxes represent copied feature maps. The arrows denote the different operations.

Normalization Since the MRI intensity values are non-standardized, we perform normalization so that the volumes have zero mean and unit variance. We perform this step independently for each modality of every patient. This involves subtracting the mean and dividing by the standard deviation of only the brain region.

2.2 U-Net based 3D Fully Convolutional Network

Our network is based on the popular U-Net architecture [8]. Our network architecture is illustrated in Fig. 1. Just like the U-Net, it consists of a contracting path (contextual pathway) and an expansive path (localization pathway). Since the U-Net is a fully connected network, our model allows end-to-end training and testing for MR image segmentation. There have been earlier approaches that train 2D FCNs on individual 2D slices. But such approaches result in loss of 3D contextual information. We train a 3D FCN that captures the 3D contextual information of MRI volumes effectively. A similar approach of U-Net inspired 3D FCN has been followed by [4].

Network Architecture As illustrated in Fig. 1, the contracting path consists of repeated application of two 3x3x3 convolutions each followed by a rectified linear unit (ReLU) and then a 3x3x3 convolution with stride 2 followed by ReLU.

The stride 2 convolution downsamples the data along with doubling the number of feature channels. Every step in the expansive path consists of an upsampling of the feature map followed by a $2 \times 2 \times 2$ convolution ("up-convolution") that halves the number of feature channels, a concatenation with the corresponding feature map from the contracting path, a $3 \times 3 \times 3$ convolution followed by a ReLU, and a $1 \times 1 \times 1$ convolution followed by ReLU. At the final layer a $1 \times 1 \times 1$ convolution is used to map each 32-component feature vector to the 4 desired classes. As proved in [6], residual learning frameworks ease the training of deeper neural networks. Hence, we implement residual connections across the two $3 \times 3 \times 3$ convolutions followed by ReLUs in every step of the contextual pathway. This enables us to train a deeper FCN.

Network Optimization Developing and training a network for 3D images involves a trade-off among receptive field, model complexity and memory consumption. While a model with a small receptive field learns only local features, a larger receptive field enables a model to draw contextual information from more global features. Since the memory requirement of 2D networks is low, they are able to use a very large receptive field, thus capturing global features from the entire image. The 2D U-Net is one such network. In case of large receptive fields in 3D, the resulting large 3D patches consume large amount of memory and therefore restrict the model complexity. We have therefore optimized the U-Net architecture in such a way so as to be able to train on a sufficiently large 3D patch while maintaining considerable model complexity. We start off with only 16 feature channels in the first layer compared to 64 in the original U-Net. We use $1 \times 1 \times 1$ convolutions in the localization pathway as they reduce the memory consumption by filter weights without any negative impact on performance. The $1 \times 1 \times 1$ convolution also further reduces the number of feature maps which is critical for reducing memory consumption. With the aforementioned design, we are able to train the network on large 3D input patches of $112 \times 112 \times 112$ voxels.

Using residual connections in our network enables us to effectively train a deep network of depth 5 in both the contextual and localization pathways. Consequently, we can start off with fewer number of feature channels and a large patch size of $112 \times 112 \times 112$ and still be able to capture a sufficiently large receptive field. We further enhance this capability to train a deeper network by implementing deep supervision [5] in the later layers of our network. Through these design choices, we are able to achieve a perfect balance of receptive field, model complexity and memory consumption.

3 Experiments and Results

3.1 Implementation Details

We trained our model on the entire BraTS 2018 training set by utilizing the HGG and LGG cases together as a whole. Post training, we obtained segmentation results by applying our algorithm on BraTS 2018 training and validation sets. We

uploaded these results to the evaluation applications in CBICA's IPP, and the portal provided us with quantitative evaluations including Dice score, sensitivity, specificity and Hausdorff distance with respect to the ground truth for each of whole tumor, tumor core and enhancing tumor core.

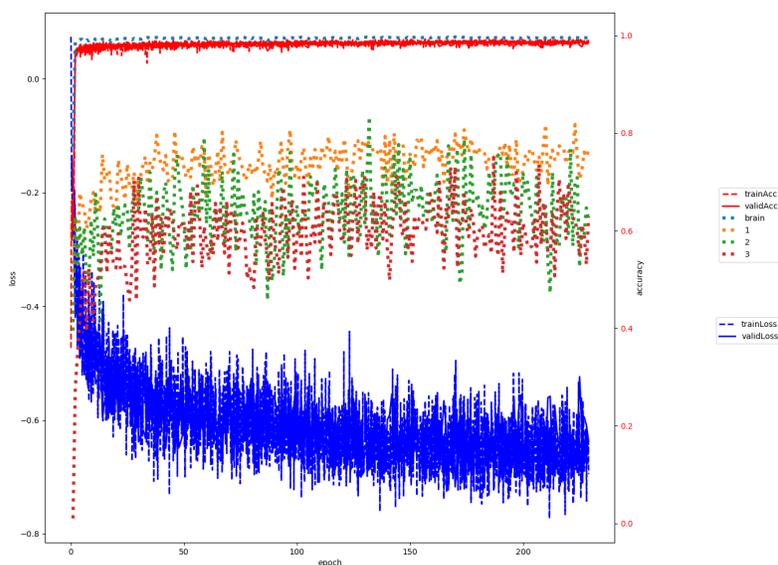


Fig. 2. Plot of accuracies and losses as a function of number of epochs trained. The numbers 1, 2 and 3 denote accuracies corresponding to edema, enhancing tumor and necrosis respectively.

We train our FCN on randomly sampled patches of size $112 \times 112 \times 112$ voxels, with 2 patches per batch. We train by Adam Optimizer with an initial learning rate of $5e-4$ and decay by a factor of 0.985 per epoch, with exponential decay rates for 1st and 2nd moment estimates as 0.9 and 0.999 respectively. We also implement l2 regularization in our network. Due to the high degree of class imbalance in brain MRI data, we use Soft Dice Loss [7] as our loss function. Soft dice loss takes into account the class imbalance by using one hot encoding of the ground truth of the image being processed. Because of the limited amount of training data (only 285 patients), we employ data augmentation while training our network in order to avoid overfitting. The data augmentation scheme comprises of scaling, rotation, mirroring, brightness variation and elastic deformations.

Our model was trained on an NVIDIA GTX 1080 Ti GPU for 230 epochs (see Fig. 2), which took around 83 hours. Since the memory consumption when running a forward pass during test time is much less compared to the consumption during training, we are able to process an entire MR image of a patient at once during test time. Thus, there is no issue of consolidating segmented patches which prevailed in some earlier proposed methods.

3.2 Segmentation Results

Qualitative results of segmentation are depicted in Fig. 3 and Fig. 4 for HGG and LGG images respectively. These results show slices from the three orthogonal views - axial, sagittal and coronal. For simplicity of visualization, we only show FLAIR and T1c modalities. Ground truth and segmentation masks are overlaid on T1c images. The orange, red and yellow colors show the edema, non-enhancing and enhancing tumor cores, respectively. As is apparent from these segmentations, our network is capable of accurately segmenting the tumor sub-regions from HGG as well as LGG images.

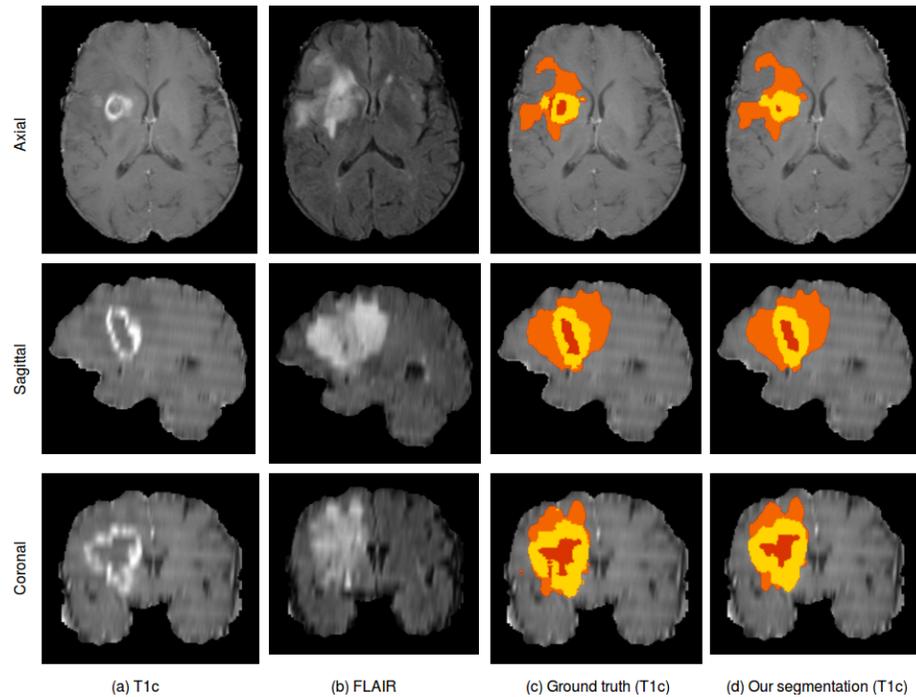


Fig. 3. Segmentation result of the brain tumor (HGG) from a training image. Orange: edema; Red: non-enhancing tumor core; Yellow: enhancing tumor core. For simplicity of visualization, the results are shown on 3 slices from 3 orthogonal views.

Table 1 presents quantitative evaluations on the BraTS 2018 training set. It shows that our proposed algorithm achieves average Dice scores of 0.72651, 0.90686 and 0.86827 for enhancing tumor core, whole tumor and tumor core, respectively. Table 2 presents quantitative evaluations with the BraTS 2018 validation set. It shows that our proposed algorithm achieves average Dice scores of 0.7439, 0.89411 and 0.82481 for enhancing tumor core, whole tumor and tumor core, respectively. On comparison with the Dice scores achieved on the training set (0.72651, 0.90686, 0.86827) it can be observed that the values are almost similar and hence our model does not overfit the training data. In case of both training set and validation set, the median scores are higher than the mean scores which shows that good segmentation results are achieved for majority of the images, while some outliers contribute to the lower mean scores.

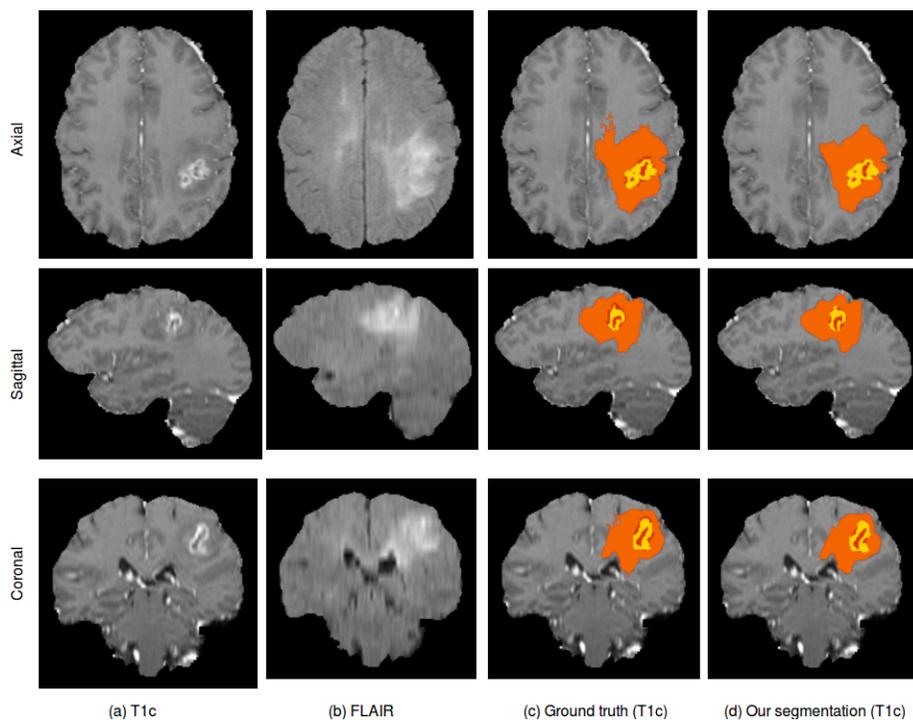


Fig. 4. Segmentation result of the brain tumor (LGG) from a training image. Orange: edema; Red: non-enhancing tumor core; Yellow: enhancing tumor core. For simplicity of visualization, the results are shown on 3 slices from 3 orthogonal views.

Table 1. Dice and Hausdorff measurements of the proposed method on BraTS 2018 training set. EN, WT, TC denote enhancing tumor core, whole tumor and tumor core, respectively.

	Dice			Hausdorff(mm)		
	ET	WT	TC	ET	WT	TC
Mean	0.72651	0.90686	0.86827	4.81241	5.34302	5.64333
StdDev	0.27467	0.05758	0.11079	9.36169	8.84861	9.9624
Median	0.83253	0.92442	0.90206	2	3	2.82843
25quantile	0.7156	0.88674	0.84177	1.41421	2.23607	1.73205
75quantile	0.885	0.94398	0.93722	3.93541	5	5.47723

Table 2. Dice and Hausdorff measurements of the proposed method on BraTS 2018 validation set. EN, WT, TC denote enhancing tumor core, whole tumor and tumor core, respectively.

	Dice			Hausdorff(mm)		
	ET	WT	TC	ET	WT	TC
Mean	0.7439	0.89411	0.82481	3.94542	6.27732	10.45925
StdDev	0.26204	0.10444	0.19492	7.63795	15.59158	22.51359
Median	0.85073	0.91811	0.90277	2	3	3.31662
25quantile	0.76936	0.89231	0.79973	1.41421	2	1.73205
75quantile	0.89215	0.94282	0.94582	3.60555	4.96991	5.71951

4 Discussion and Conclusion

The scores show that our developed method has achieved competitive performance on BraTS 2018 dataset for automatic brain tumor segmentation. We have also observed that the N4ITK bias field correction improves the segmentation results. Thus the bias correction step is effectively smoothening the multiplicative field and removing bias and intensity inhomogeneities.

In conclusion, we have developed an algorithm to segment glioma subregions from multimodal brain MRI scans. Our model makes efficient use of 3D contextual information and effectively maintains a balance between input patch size and model complexity while being limited by 11 GB of GPU RAM. Experimental results on BraTS 2018 validation set show that our proposed method achieved mean Dice scores of 0.7439, 0.89411 and 0.82481 for enhancing tumor core, whole tumor and tumor core, respectively. The corresponding values for BraTS 2018 training set were 0.72651, 0.90686 and 0.86827, respectively.

References

1. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI

- collections with expert segmentation labels and radiomic features”, *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. ”Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection”, *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
 3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. ”Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection”, *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
 4. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. *MICCAI BraTS Challenge (2017)*
 5. B. Kayalibay, G. Jensen, and P. van der Smagt, CNN-based segmentation of medical imaging data, *arXiv preprint arXiv:1701.03056*, 2017
 6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *CVPR*. pp. 770778 (2016)
 7. F. Milletari, N. Navab, and S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in *International Conference on 3D Vision*. IEEE, 2016, pp. 565571
 8. O. Ronneberger, P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234-241
 9. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp , Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. ”The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)”, *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
 10. N. J. Tustison and J. C. Gee, J.c.: N4itk: Nicks n3 itk implementation for mri bias field correction, *Insight Journal*, 2009

Brain Tumor Segmentation on multi-modal MR images with Deep Architectures

Santi Puch¹, Irina Sánchez¹, Aura Hernández², Gemma Piella³, and Vesna Prčkovska¹

¹ QMENTA Imaging Inc, Barcelona, Spain
{santi, irina, vesna}@qmenta.com

² Computer Vision Center, Universitat Autònoma de Barcelona, Barcelona, Spain
aura@cvc.uab.es

³ SIMBIOsys, Universitat Pompeu Fabra, Barcelona, Spain
gemma.piella@upf.edu

Abstract. In this work, we present a study of several 3D fully-convolutional neural networks trained on the task of brain tumor segmentation. We focus on context-aware and efficient architectures: UNet, a well-established segmentation network; ResUNet, a residual version of UNet; and a novel architecture, ContextNet, an extension of ResUNet that incorporates ampler context via planar filters with large kernels. We train these architectures on the BraTS 2018 dataset with variations of the available input modalities for the network, achieving 0.895 of whole tumor DICE, 0.758 of enhancing tumor DICE and 0.774 of tumor core DICE on the BraTS 2018 validation data. We also leverage the learned representations of the network, along with basic measures of size of the different tumoral structures obtained from the model's prediction, to predict the patient overall survival, obtaining an accuracy of 0.536 when classifying subjects as long, short and mid-survivors.

Keywords: Brain Tumors · 3D fully-convolutional CNN · Magnetic Resonance Imaging · Global Planar Convolution · Genetic Programming

1 Introduction

It is estimated that, as of today, 700.000 people in the United States are living with a primary brain tumor, from which 80% are benign and 20% are malignant tumors[13]. Of all malignant brain tumors, 81% are gliomas, which are tumors that originate in glial cells[16]. Glioblastomas are the most common type of glioma, representing 45% of all gliomas; they are one of the most aggressive types of brain tumors, having an estimated 5-year relative survival rate of approximately 5%, which means that only 5% of people diagnosed with a glioblastoma will still be alive 5 years after being diagnosed [16].

It is clear that such dismal prognosis requires proper treatment planning and follow-up, which can be greatly improved if proper in-vivo, non-invasive delineation and identification of glioma structures is in place. However, this

poses a significant burden on the radiologist: multiple imaging modalities have to be assessed in parallel, as each highlights different regions of the tumor, and the process of delineation in a 3D acquisition is tedious and prone to errors. As a consequence, inter-observer variability has been reported to be a major — if not the largest — factor of inaccuracy in radiation therapy, constituting the weakest link in the radiotherapy chain that goes from diagnosis and consultation, going through 3D imaging and target volume delineation, to treatment delivery [20].

Therefore, automating the delineation and identification process on MR images would accelerate treatment planning and improve treatment follow-up. However, the problem of tumor segmentation poses several challenges, such as blurry or smoothed boundaries, variability of shape, location and extension or heterogeneity of appearance of brain tumors on MR images.

The research community has concentrated efforts in order to address the brain tumor segmentation task, and to this end initiatives like the Brain Tumor Segmentation (BraTS) challenge[11] have made the problem accessible to a larger audience. As a result, a large variety of computational methods have been proposed to automate the delineation of brain tumors. These methods can be broadly categorized in two groups: generative models, which rely on prior knowledge about tissue appearance and distribution; and discriminative models, which directly learn the relationship between the image features and the segmentation labels. Deep Learning approaches, especially Convolutional Neural Networks (CNNs), cover a large portion of the recently proposed discriminative methods. The majority of these works have based their methods in well-known semantic segmentation networks, either using a 2D variant on one or more planes of the brain, or implementing a 3D architecture that takes spatial information in all directions into account. In all these works, several training strategies are leveraged, such as dense training with patches combined with patch sampling schemes, and a distinction between local, refined features and global, coarse features is accomplished via multiresolution approaches or skip connections.

In this project, we explore different 3D fully-convolutional architectures in order to solve the tumor segmentation problem, focusing on spatial context awareness, precision and efficiency. We start with a baseline architecture, a 3D variation of UNet [18], and we introduce incremental variations to the architecture to improve its performance. Concretely, we present in this work ResUNet, a variation of UNet that incorporates residual elements introduced by [6]. We further refine this architecture by adding a set of planar convolutional layers with large kernels, with the objective of enlarging the context perceived by the network; we call this architecture ContextNet. We train these architectures in the 2018 BraTS Challenge dataset, that consists of 210 High Grade Glioma (HGG) cases and 75 Low Grade Glioma (LGG) cases with manual annotations of the distinct intra-tumoral structures of interest [11] [3] [1] [2]. In addition, we leverage the representations learned by the network and the resulting segmentation, as well as the age, to predict the patient overall survival in days with a machine learning model found via a genetic programming algorithm.

2 Methods

2.1 Data

The data used in this project originates in the 2018 version of the yearly Multimodal Brain Tumor Segmentation Challenge dataset. This dataset consists of 285 multi-institutional clinically-acquired pre-operative scans. Each multimodal scan is formed by T1, T1-Gd, T2 and T2-FLAIR volumes acquired with various scanners from 19 institutions. All the scans have been segmented manually by one to four raters, and approved by experienced neuroradiologists. The ground-truth labels comprise the enhancing tumor, the peritumoral edema and the necrotic and non-enhancing tumor when present. Each multi-modal scan in the BraTS challenge dataset is co-registered to the same anatomical template, skull-stripped and resampled to 1mm^3 isotropic resolution. Therefore, no further preprocessing is needed. Overall survival data, defined in days, is also provided for 164 of the HGG subjects, as well as the age in years and the resection status.

2.2 CNN architectures

All the Convolutional Architectures implemented and trained are 3D and fully-convolutional by design, meaning that they can be trained using 3D patches of data, and then they can be used for inference with whole brain volumes.

UNet This reference architecture proposed initially in [18] is based on [10], with a similar contracting path, but in order to improve the localization capabilities of the network a supplementary expanding path is introduced, in which high-resolution features from the contracting path are combined with upsampled feature maps. This architecture is extended to 3D by replacing all the convolutions, transposed convolutions and max pooling by their 3D alternatives, similarly to [4].

In our experiments we use Rectified Linear Unit (ReLU) as activation function and Batch Normalization [7] after convolutions and before activations.

ResUNet The work of [6] introduced the concept of deep residual learning: instead of stacking a series of layers and letting them learn the desired underlying mapping, these layers can be set to explicitly fit a residual mapping. This residual learning framework not only improved the performance on the image classification task by a wide margin, but also alleviated the degradation problem found when an excessive amount of layers was used.

The motivation of introducing residual layers originates from the fact that the contracting path has a twofold purpose: on one hand it learns increasingly abstract features that encode contextual information necessary to decide *what*, and on the other hand it connects feature maps from lower-level representations with the expanding path to decide *where*. Residual elements allow for an increased number of layers, which has been shown empirically to increase the

representational power of the network [19] [6], thus helping with the first task (*identification*). They also facilitate learning of identity mappings, which enables the possibility of passing low-level representations throughout the network, thus easing the second task (*delineation*).

The architecture of ResUNet is essentially an extension of UNet with residual elements. The convolutional layers in UNet are replaced by residual layers; concretely, 2 residual layers are used at each resolution level.

Again, as in UNet, we use Rectified Linear Unit (ReLU) as activation function and Batch Normalization after convolutions and before activations.

ContextNet ContextNet is a novel architecture introduced in this work that aims to enhance the context-awareness capabilities of 3D imaging segmentation architectures. It is built upon the aforementioned ResUNet architecture and inspired by [17]. An overview of the architecture is shown in fig. 1.

The localization aspect of semantic segmentation networks is addressed by skip-connections and residual elements, as these components let low level representations pass through the network and inform the latest layers about fine-grained spatial details. However, the classification aspect of semantic segmentation networks, that deals with proper identification of the delineated structures, is hindered by the fact that these networks are focused on proper boundary alignment. State-of-the-art classification architectures rely on layers that are globally connected, which in the most extreme case (all the nodes are connected with each other) corresponds to a fully-connected layer. It is clear that such type of operation is not feasible in a fully-convolutional architecture, however we can approximate global connectivity by increasing kernel size in convolutions: in the limit, the kernel is as big as the input feature map, which can be interpreted again as a fully-connected layer. The problem with increasing the kernel size is the computational and memory requirements associated with it, and it is not feasible in the case of 3D CNNs with current accelerated computing hardware.

However, global connectivity can be approximated by constraining the kernel parameters' subspace. Specifically, we can constrain the convolutional kernels to have one dimension less than they would normally have, which in practice is implemented by having kernels with size 1 in one of the dimensions. This reduction of parameters in one of the three dimensions allows the growth of kernel sizes in the other dimensions, thus providing improved global connectivity.

On the basis of this reasoning, we introduce a new module in this architecture named *Global Planar Convolution*, abbreviated GPC. A GPC convolves planar filters (i.e. filters in which one of the three dimensions has size 1) in each of the three orthogonal directions, and then combines the resulting planar feature maps via summation. We introduce these modules in between skip-connections, similarly to bottleneck modules in [6]. We further improve the localization capabilities by including an extra residual layer after each GPC module, however these residual layers do not include an activation in the end. The resulting feature maps from each of this altered skip-connections is then summed with the feature maps outputted by the upsampling layers in the expanding path, and

the summed feature maps are then passed through an Exponential Linear Unit (ELU) [5]. In our experiments we are able to set the filter size of GPC modules to 15.

2.3 Patient overall survival prediction

Several features are extracted from the network for the task of survival prediction of brain tumor patients. On one hand, the final segmentation map produced by the trained networks is used to compute the ratio between the extent of each tumoral structure and the whole tumor, as well as the ratio between lesion volume and whole-brain volume. This results in 3 features for the tumoral structure ratios, plus 1 feature for the lesion to brain volume ratio. On the other hand, the feature maps from the deepest convolutional layer in ContextNet (section 2.2) are extracted and their mean, standard deviation and max values are used as features for the prediction task. This adds $15 \cdot 3 = 45$ features, as there are 15 feature maps (one per filter) and we obtain 3 measures from each of them. Finally, we also include the age of the patient as a predictor. In the end, the resulting feature vector produced for each brain tumor patient consists of $4 + 45 + 1 = 50$ features.

A genetic programming algorithm is used to search for the best model, according to a specific optimization score, in the space of all possible feature extraction, construction and preprocessing operations, machine learning algorithms and their possible combinations [14] [15]. We let the evolutionary algorithm run for 200 generations, each creating an offspring of 200 elements and retaining only 100 of them at the end of each generation. The offspring is produced as a mix of mutation (90%) and cross-over (10%), and the evolutionary algorithm evaluates the fitness of each model in a generation in terms of the coefficient of determination R^2 .

2.4 Experimental design

Local dataset split In order to evaluate the segmentation performance of the implemented models, the dataset is split into train and validation sets. We use a split ratio of 70%-30% for training and validation, respectively. This results in 199 subjects for training and 86 subjects for evaluation on the tumor segmentation task. A 6-fold cross-validation scheme is used to train and evaluate the performance of the machine learning models on the task of overall survival prediction.

CNN training details Categorical cross-entropy was selected as the loss function to be minimized during training. The complete loss function included L1 and L2 penalization of the weights (for regularization purposes), with penalization ratios of $1E-6$ and $1E-4$, respectively.

All the models were trained using the ADAM optimizer [9]. The initial learning was set to $1E-3$ in all experiments, and a learning rate decay policy was

integrated in order to stabilize training as the training procedure progressed. Concretely, we used an exponential decay of the learning rate every 1000 training steps with a decay rate of 0.9. The number of training steps was 35000. The training procedure alternated 1000 training steps with 1 complete evaluation of the model. Batch size was set to 6, which maximized the memory consumption in the most memory demanding architectures.

During training, the data ingestion pipeline was configured to extract patches of size $80 \times 80 \times 80$ with 50% probability of being centered on a background voxel and 50% on a tumor voxel (50% background, 20% edema, 15% enhancing tumor and 15% necrosis and non-enhancing tumor). Whole brain volumes were used during evaluation in order to provide a realistic value of performance in a real-world scenario. The CNNs were trained on two different hardware configurations, depending on the availability of computing resources: 1) AWS p2.xlarge instance with a single NVIDIA K80 with 12 GiB of GPU memory; 2) on-premises server with two NVIDIA GeForce GTX 1080 Ti with 11GiB of GPU memory.

Restriction of availability of input modalities We perform data ablation experiments by restricting the available input modalities at training time, but always maintaining the minimum required modalities to properly identify all structures, namely T1-Gd and FLAIR. The motivation for such experiments is twofold. First, we want to assess the relative contribution of each modality to the overall segmentation, and inspect if some modalities are redundant or indeed provide useful information. Second, it is convenient and even necessary to have models that can work with a restricted number of modalities (as in some clinical cases not all MR sequences are included in the protocol) even if such models with restricted input information do not perform as well as models trained without data restrictions.

2.5 Evaluation

The primary evaluation score for the segmentation task is the Sørensen-Dice coefficient, usually abbreviated as DICE. In this context, the DICE coefficient compares the similarity between the set of true examples and the set of positive examples:

$$DICE = \frac{2TP}{2TP + FP + FN} \quad (1)$$

The Hausdorff distance is used to evaluate the distance between segmentation boundaries. Results are reported using the 95% quantile of the maximal surface distance between the ground truth P_1 and the predicted segmentation T_1 [12]:

$$Haus(P, T) = \max\left(\sup_{p \in \partial P_1} \inf_{t \in \partial T_1} d(p, t), \sup_{t \in \partial T_1} \inf_{p \in \partial P_1} d(t, p)\right) \quad (2)$$

The survival prediction task is evaluated in terms of classification accuracy of patients as short, mid or long-survivors, and the mean squared error of the predictions.

3 Results

Table 1 shows the DICE coefficients on the local validation set for all target structures of all the architectures, trained with different data configurations. The best model for whole tumor segmentation is ContextNet trained with all modalities (0.897 DICE score), while the best model for enhancing tumor and tumor core segmentation is ResUNet trained only with T1-Gd, FLAIR and T1, achieving 0.752 and 0.799 DICE scores, respectively.

	Enhancing Tumor	Whole Tumor	Tumor Core
UNet - all modalities	0.698 ± 0.229	0.847 ± 0.095	0.694 ± 0.235
ResUNet - all modalities	0.739 ± 0.207	0.892 ± 0.064	0.785 ± 0.200
ResUNet - T1-Gd, FLAIR, T1	0.752 ± 0.193	0.882 ± 0.080	0.799 ± 0.171
ResUNet - T1-Gd, FLAIR	0.723 ± 0.218	0.884 ± 0.070	0.790 ± 0.184
ContextNet - all	0.752 ± 0.207	0.897 ± 0.059	0.797 ± 0.195
ContextNet - T1-Gd, FLAIR, T1	0.743 ± 0.216	0.881 ± 0.071	0.770 ± 0.211
ContextNet - T1-Gd, FLAIR	0.734 ± 0.231	0.878 ± 0.080	0.770 ± 0.216

Table 1: DICE of target structures on the local validation set.

On the BraTS 2018 validation data we reach a similar DICE score in whole tumor segmentation with the same model (ContextNet trained with all modalities) as reported in table 2. We also provide the evaluation results of ensembling several models to give more robust segmentation maps, as in [8]. This model ensembling approach allows us to obtain 0.895, 0.758 and 0.774 of whole tumor, enhancing tumor and tumor core DICE scores, respectively.

	Dice			Hausdorff95		
	ET	WT	TC	ET	WT	TC
UNet						
All modalities	0.711 ± 0.293	0.871 ± 0.096	0.706 ± 0.284	11.676 ± 25.6781	18.536 ± 28.886	22.752 ± 29.088
T1-Gd, FLAIR, T1	0.704 ± 0.289	0.856 ± 0.103	0.73043 ± 0.255	9.921 ± 20.218	13.723 ± 24.201	14.758 ± 23.668
T1-Gd, FLAIR	0.711 ± 0.282	0.862 ± 0.092	0.734 ± 0.243	10.743 ± 22.139	22.677 ± 29.220	17.023 ± 24.145
ResUNet						
All modalities	0.729 ± 0.279	0.882 ± 0.071	0.741 ± 0.256	5.578 ± 11.249	9.896 ± 16.803	9.532 ± 12.407
T1-Gd, FLAIR, T1	0.700 ± 0.296	0.867 ± 0.088	0.725 ± 0.27	12.197 ± 25.005	23.835 ± 29.099	23.203 ± 29.459
T1-Gd, FLAIR	0.738 ± 0.271	0.877 ± 0.077	0.750 ± 0.256	7.375 ± 16.942	16.687 ± 23.720	19.632 ± 29.278
ContextNet						
All modalities	0.735 ± 0.281	0.883 ± 0.112	0.753 ± 0.269	7.004 ± 13.944	7.594 ± 12.453	9.505 ± 11.557
T1-Gd, FLAIR, T1	0.736 ± 0.271	0.857 ± 0.086	0.746 ± 0.26	10.178 ± 20.717	23.062 ± 29.066	14.584 ± 20.348
T1-Gd, FLAIR	0.719 ± 0.274	0.872 ± 0.113	0.745 ± 0.226	10.411 ± 20.932	13.131 ± 20.018	15.867 ± 20.405
Ensemble of models						
All models	0.755 ± 0.266	0.894 ± 0.071	0.769 ± 0.258	5.564 ± 10.813	9.454 ± 17.74	7.895 ± 8.567
ContextNet + ResUNet	0.758 ± 0.264	0.895 ± 0.07	0.774 ± 0.253	4.502 ± 8.227	10.656 ± 19.286	7.103 ± 7.084

Table 2: Evaluation scores obtained on the BraTS 2018 validation set.

We also show some example segmentations in fig. 2 that illustrate cases in which the trained models obtain a good segmentation, confound some tumoral structures or deal with challenging examples.

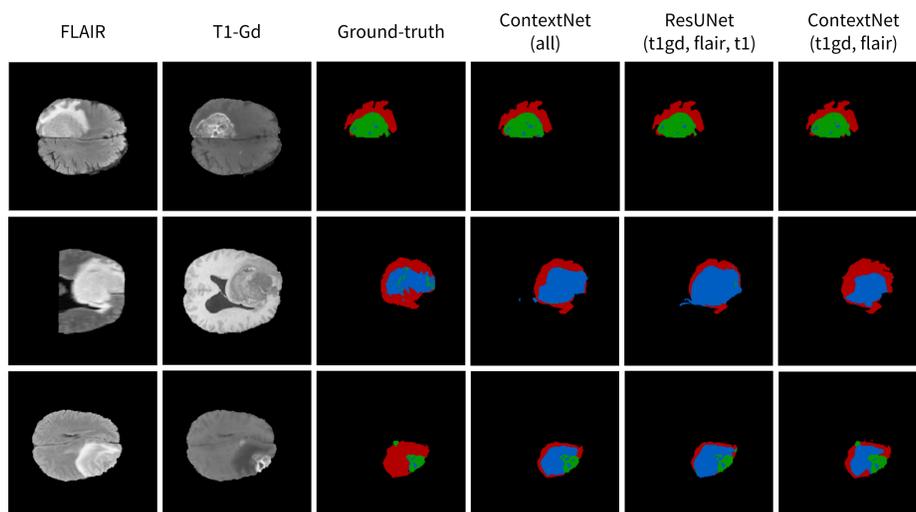


Fig. 2: FLAIR and T1-Gd MR modalities, ground-truth labels and segmentations produced by ContextNet trained with all modalities, ResUNet trained without T2, and ContextNet trained only with T1-Gd and FLAIR of 3 subjects from the local validation split of the BraTS 2018 dataset.

Finally, using the feature extraction and evolutionary algorithm approaches explained in section 2.3, we reach a patient classification accuracy of 0.536 and a mean squared error of the predicted survival values of 107493.9 on the validation data provided by the BraTS 2018 organizers.

4 Conclusion

In this work we present several 3D fully-convolutional CNNs to address the task of automatic tumor segmentation from magnetic resonance images, with the objective of accelerating and improving radiotherapy planning and monitoring of patients with gliomas of varied grades. We start with a baseline architecture, UNet, and gradually improve its performance by adding residual elements (ResUNet) and enlarging the receptive field of its components via large planar filters (ContextNet). The representational power of CNNs in images is leveraged to extract features for the task of patient overall survival prediction, that we address via the search of an optimal machine learning pipeline with genetic programming. We demonstrate competitive performance on the BraTS 2018 validation data in both tasks, achieving 0.895 DICE score for whole tumor segmentation and 4.502 Hausdorff distance in enhancing tumor, and 0.536 classification accuracy on the task of survival prediction. Future work includes improving the performance of individual models by means of hyper-parameter optimization, and in-depth investigation of intermediate representations and behavior of the models for improved interpretability.

References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive* **286** (2017)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lyg collection. *The Cancer Imaging Archive* **286** (2017)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci Data* **4**, 170117 (09 2017)
4. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *CoRR* **abs/1606.06650** (2016), <http://arxiv.org/abs/1606.06650>
5. Clevert, D., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (elus). *CoRR* **abs/1511.07289** (2015), <http://arxiv.org/abs/1511.07289>
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. *Arxiv.Org* **7**(3), 171–180 (2015). <https://doi.org/10.3389/fpsyg.2013.00124>, <http://arxiv.org/pdf/1512.03385v1.pdf>
7. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR* **abs/1502.03167** (2015), <http://arxiv.org/abs/1502.03167>
8. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S.G., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M.C.H., Kainz, B., Rueckert, D., Glocker, B.: Ensembles of multiple models and architectures for robust brain tumour segmentation. *CoRR* **abs/1711.01468** (2017), <http://arxiv.org/abs/1711.01468>
9. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *CoRR* **abs/1412.6980** (2014), <http://arxiv.org/abs/1412.6980>
10. Long, J., Shelhamer, E., Darrell, T.: Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* pp. 3431–3440 (2015). <https://doi.org/10.1109/CVPR.2015.7298965>
11. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftexharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging* **34**(10), 1993–2024 (oct 2015). <https://doi.org/10.1109/TMI.2014.2377694>, <http://ieeexplore.ieee.org/document/6975210/>

12. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging* **34**(10), 1993 (2015)
13. National Brain Tumor Society: Quick brain tumor facts (2018), <http://braintumor.org/brain-tumor-information/brain-tumor-facts/>, [Online; accessed 12-June-2018]
14. Olson, R.S., Bartley, N., Urbanowicz, R.J., Moore, J.H.: Evaluation of a tree-based pipeline optimization tool for automating data science. In: *Proceedings of the Genetic and Evolutionary Computation Conference 2016*. pp. 485–492. GECCO '16, ACM, New York, NY, USA (2016), <http://doi.acm.org/10.1145/2908812.2908918>
15. Olson, R.S., Urbanowicz, R.J., Andrews, P.C., Lavender, N.A., Kidd, L.C., Moore, J.H.: Applications of Evolutionary Computation: 19th European Conference, EvoApplications 2016, Porto, Portugal, March 30 – April 1, 2016, *Proceedings, Part I*, chap. Automating Biomedical Data Science Through Tree-Based Pipeline Optimization, pp. 123–137. Springer International Publishing (2016), http://dx.doi.org/10.1007/978-3-319-31204-0_9
16. Ostrom, Q.T., Bauchet, L., Davis, F.G., Deltour, I., Fisher, J.L., Langer, C.E., Pekmezci, M., Schwartzbaum, J.A., Turner, M.C., Walsh, K.M., Wrensch, M.R., Barnholtz-Sloan, J.S.: The epidemiology of glioma in adults: a "state of the science" review. *Neuro-oncology* **16**(7), 896–913 (Jul 2014)
17. Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J.: Large kernel matters - improve semantic segmentation by global convolutional network. *CoRR* **abs/1703.02719** (2017), <http://arxiv.org/abs/1703.02719>
18. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation (may 2015), <http://arxiv.org/abs/1505.04597>
19. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 1–9 (June 2015). <https://doi.org/10.1109/CVPR.2015.7298594>
20. Weiss, E., Hess, C.F.: The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical aspects and practical experiences. *Strahlenther Onkol* **179**(1), 21–30 (Jan 2003)

Glioma Segmentation In a Few Seconds Using Fully Convolutional Network and Transfer Learning

Elodie Puybareau, Joseph Chazalon and Guillaume Tochon

EPITA Research and Development Laboratory (LRDE), France
elodie.puybareau@lrde.epita.fr

Abstract. In this paper, we propose a fast automatic method that segments glioma without any manual assistance, using a fully convolutional network (FCN) and transfer learning. This FCN is the base network of VGG-16, pre-trained on ImageNet for natural image classification, and fine tuned with the training dataset of the MICCAI 2018 BraTS Challenge. It relies on the "pseudo-3D" method published at ICIP 2017, which allows for segmenting objects from 2D color images which contain 3D information of MRI volumes. For each n^{th} slice of the volume to segment, we consider three images, corresponding to the $(n - 1)^{\text{th}}$, n^{th} , and $(n + 1)^{\text{th}}$ slices of the original volume. These three gray-level 2D images are assembled to form a 2D RGB color image (one image per channel). This image is the input of the FCN to obtain a 2D segmentation of the n^{th} slice. We process all slices, then stack the results to form the 3D output segmentation. With such a technique, the segmentation of a 3D volume takes only a few seconds.

Keywords: glioma · tumor segmentation · fully convolutional network.

1 Introduction

1.1 Motivation

Gliomas are the most common brain tumors in adults, growing from glial cells and invading the surrounding tissues [8]. Two classes of tumors are observed. The patients with the more aggressive ones, classified as high-grade gliomas (HGG), have a median overall survival of two years or less and imply immediate treatment [12, 15]. The less aggressive ones, the low-grade gliomas (LGG), allow an overall survival of several years, with no need of immediate treatment. Multimodal magnetic resonance imaging (MRI) helps practitioners to evaluate the degree of the disease, its evolution and the response to treatment. Images are analyzed based on qualitative or quantitative measures of the lesion [7, 18]. Developing automated brain tumor segmentation techniques that are able to analyze these tumors is challenging, because of the highly heterogeneous appearance and shapes of these lesions. Manual segmentations by experts can also be a challenging task, as they show significant variations in some cases.

During the past 20 years, different algorithms for segmentation of tumor structures has been developed and reviewed [1, 5, 6]. However, a fair comparison of algorithms implies a benchmark based on the same dataset, as it has been proposed during MICCAI BraTS Challenges [14].

1.2 Context

This work has been done in the context of the MICCAI 2018 Multimodal Brain Tumor Segmentation Challenge (BraTS)¹. The aim was to provide a fully automated pipeline for the segmentation of the glioma from multi modal MRI scans without any manual assistance and to predict the patient overall survival.

Despite the relevance of glioma segmentation, this segmentation is challenging due to the high heterogeneity of tumors. The development of an algorithm that can perform fully automatic glioma segmentation and overall prediction of survival would be an important improvement for patients and practitioners.

We received data of 286 patients, with associated masks to develop our method. The data, available online, have been annotated and preprocessed [2–4]. The volumes given are T1, T1ce, T2 and FLAIR. Our method is then evaluated on new volumes: a validation set released by the organizers without the manual segmented masks (these masks will not be released), to obtain preliminary results. For the challenge evaluation, a new set will be released and we are asked to send the segmentations within 48h after receiving the data. This challenge will establish a fair comparison to state-of-the-art methods, and a release of a large annotated dataset.

1.3 Related Works

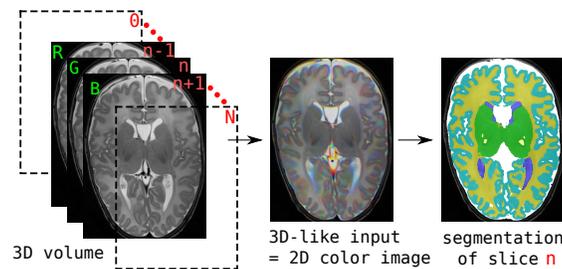


Fig. 1. Illustration of 3D-like color image and associated segmentation used in [19].

As machine learning really improved the results of some segmentation tasks, the use of such strategy seems meaningful in the context of medical image segmentation.

¹ <https://www.med.upenn.edu/sbia/brats2018.html>

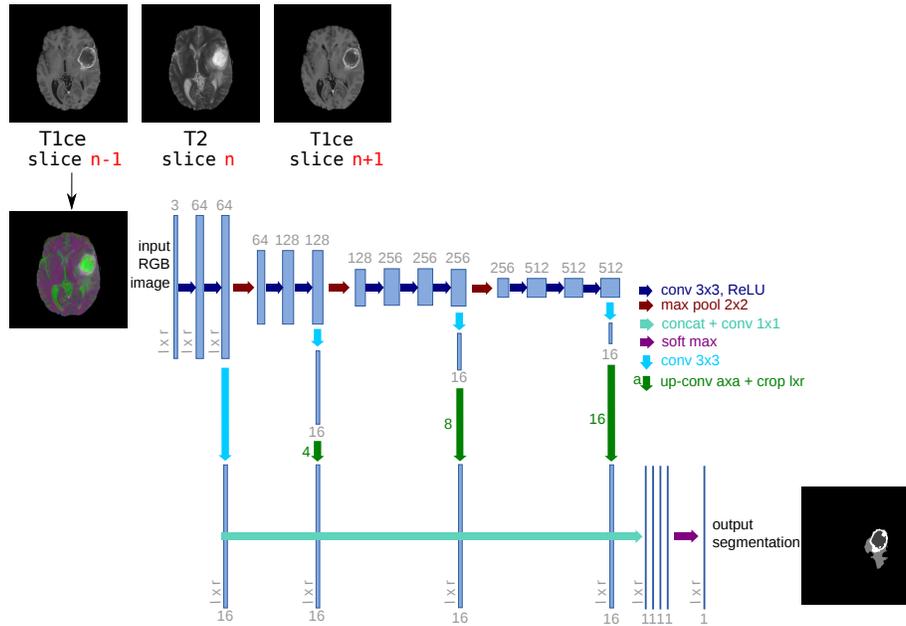


Fig. 2. Architecture of the proposed network. We fine tune it and combine linearly fine to coarse feature maps of the pre-trained VGG network [17]. Note that each input color image is built from the slice n and its neighbouring slices $n - 1$ and $n + 1$.

In a work published in the IEEE Intl. Conf. on Image Processing (ICIP) in 2017 [19], 3D brain MR volumes are segmented using fully convolutional network (FCN) and transfer learning. The network used for transfer learning is VGG (Visual Geometry Group) [17], pre-trained on the ImageNet dataset. It takes as input a 2D color image that is here a 3D-like image, or *3D-like* image, composed of 3 consecutive slices of the 3D volume (see Fig. 1). This method uses only one modality, and reaches good results for brain segmentation.

For this challenge, we rely on the same ideas than the work on brain MRI segmentation [19], leveraging the power of a fully convolutional network pre-trained on a large dataset and later fine-tuned on the training set of the challenge to segment tumors.

2 Method

An overview of the proposed method is given in Fig. 2. The method is fully automatic, and takes pseudo-3D images as input. It is really fast as about 10 seconds are needed to process a complete volume.

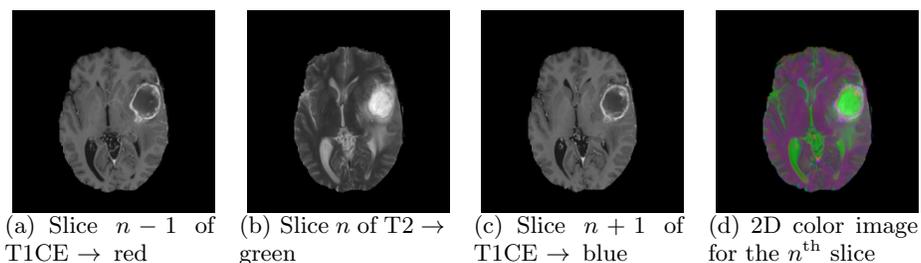


Fig. 3. Three successive slices (a-c) are used to build a 2D color image (d) from for example T1CE and T2 images.

2.1 Pre-Processing

Let n, m be respectively the minimum non-null and maximum gray-level value of the histogram. We quantize all voxel values using a linear function so that the gray-level range $[n, m]$ is mapped to $[-127, 127]$. For our application, the question amounts to how to prepare appropriate inputs (RGB input images) given that a brain MR image is a 3D volume. To that aim, we propose to stack successive 2D slices. Precisely, to form an input artificial color image for the pre-trained network to segment the n^{th} slice, we use the slice n of the volume, its predecessor, the $(n - 1)^{\text{th}}$ slice, and its successor, the $(n + 1)^{\text{th}}$ slice, as respectively the green, red and blue channels (note that the order is not relevant). This process is depicted in Fig. 2 (left). Each 2D color image thus forms a representation of a part (a small volume) of the MR volume.

A variant of this procedure is to combine modalities. The n^{th} slice is from one modality and its $(n - 1)^{\text{th}}$ and $(n + 1)^{\text{th}}$ from a second modality. This combination brings not only 3D information but also multi-modality information. Fig. 3 illustrates this variant.

2.2 Deep FCN for Tumor Segmentation

Fully convolutional network (FCN) and transfer learning has proved their efficiency for natural image segmentation [11]. The paper [19] proposed to rely on a FCN and transfer learning to segment 3D brain MR images, although those images are very different from natural images. As it was a success, we adapted it to glioma segmentation. We rely on the 16-layer VGG network [17], which was pre-trained on millions of natural images of ImageNet for image classification [10]. For our application, we keep only the 4 stages of convolutional parts called “*base network*”, and we discard the fully connected layers at the end of VGG network. This base network is mainly composed of convolutional layers: $z_i = w_i \times x + b_i$, Rectified Linear Unit (ReLU) layers for non-linear activation function: $f(z_i) = \max(0, z_i)$, and max-pooling layers between two successive stages, where x is the input of each convolutional layer, w_i is the convolution parameter, and b_i is the bias term. The three max-pooling layers divide the base network into four stages of fine to coarse feature maps. Inspired by the work

in [11,13], we add specialized convolutional layers (with a 3×3 kernel size) with K (e.g. $K = 16$) feature maps after the convolutional layers at the end of each stage. All the specialized layers are then rescaled to the original image size, and concatenated together. We add a last convolutional layer with kernel size 1×1 at the end. This last layer combine linearly the fine to coarse feature maps in the concatenated specialized layers, and provide the final segmentation result. The proposed network architecture is schematized in Fig. 2.

The architecture described above is very similar with the one used in [13] for retinal image analysis, where the retinal images are already 2D color images. Using such a 2D representation avoids the expensive computational and memory requirements of fully 3D FCN.

For the training phase, we use the multinomial logistic loss function for a one-of-many classification task, passing real-valued predictions through a softmax to get a probability distribution over classes. During training, we use the classical data augmentation strategy by scaling and rotating. We rely on the ADAM optimization procedure [9] (AMSGrad variant [16]) to minimize the loss of the network. The relevant parameters of the methods are the following: the learning rate is set to 0.002 (we did not use learning rate decay), the `beta_1` and `beta_2` are respectively set to 0.9 and 0.999, and we use a fuzz factor (epsilon) of 0.001.

At test time, after having pre-processed the 3D volume (requantization), we prepare the set of 2D color images and pass every image through the network.

We run the train and test phase on an NVIDIA GPU card. The testing one lasts less than 10 seconds.

The output of the network for one slice during the inference phase is a 2D segmented slice. After treating all the slices of the volume, all the segmented slices are stacked to recover a 3D volume with the same shape as the initial volume, and containing only the segmented lesions.

3 Experiments and Results

This section presents the experiments and results obtained, during the development of our method (using the training dataset) using the Dice coefficient

3.1 Experiments

In this part, we used the training scans provided by the challenge. *Note: we here present the experiments and hypothesis we are currently working on. We do not have yet the results, this part will hence be rewritten for the final version.*

Training and testing. We trained our model on randomly chosen scans. The model was trained using the parameters described in the previous section. We tested on the remaining scans and cross-validated.

Comparison of pseudo-3D ranks. In the pseudo-3D procedure, we associate to the n^{th} slice the $(n - x)^{\text{th}}$ and $(n + x)^{\text{th}}$ slices, with x being a small integer. As pseudo-3D is intended to provide 3D information, we can wonder what would be the best x in our case. We make x vary in that aim. Results will be shown later.

Axis and combination Our method deals with 2D RGB images that are pseudo-3D. A clue to improve segmentation results and to get a better contours detection would be to associate three networks, each network being trained on a particular axis (axial, sagittal and coronal), and to combine their results to obtain the final segmentation. We study the benefits of such approach.

3.2 Results

In this part, we will present the results we will obtain at the end of our development phase.

Note As we are currently doing our tests, we do not have the results yet. The preliminary results before post-processing is a dice of 0.87 for the whole tumor, computed on a fixed test set (10% of training set). The discussion part will be added for the final submission, as well as the part on overall survival prediction as we are currently working on. The code will be available on <https://www.lrde.epita.fr/wiki/NeoBrainSeg>

4 Conclusion

We propose in this article a method to segment accurately glioma in few seconds based on transfer learning from VGG-16, a pre-trained network used to classify natural images. This method takes the advantage of keeping 3D information of the MRI volume and the speed of processing only 2D images, thanks to the pseudo-3D concept.

This method can also deal with multi-modality, and can be applied to other segmentation problems, such as in [21], where a similar method is proposed to segment white matter hyperintensities, but pseudo-3D has been replaced by an association of multimodality and mathematical morphology pre-processing to improve the detection of small lesions. Hence, we might also try to modify our inputs thanks to some highly non-linear filtering to help the network segment tumors, precisely some mathematical morphology operators [20].

The strength of this method is its modularity and its simplicity. It is easy to implement, fast, and does not need a huge amount of annotated data for training (in the work on brain segmentation [19], there is only 2 images for training for some cases).

Acknowledgments The authors want to thank the organizers of the BraTS 2018 Challenge and the MICCAI Brainles Workshop. The GPU card “Quadro P6000” used for the work presented in this paper was donated by NVIDIA Corporation.

References

1. Angelini, E.D., Clatz, O., Mandonnet, E., Konukoglu, E., Capelle, L., Duffau, H.: Glioma dynamics and computational models: a review of segmentation, registration, and in silico growth algorithms and their clinical applications. *Current Medical Imaging Reviews* **3**(4), 262–276 (2007)

2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive* **286** (2017)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *The Cancer Imaging Archive* **286** (2017)
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data* **4**, 170117 (2017)
5. Bauer, S., Wiest, R., Nolte, L.P., Reyes, M.: A survey of mri-based medical image analysis for brain tumor studies. *Physics in Medicine & Biology* **58**(13), R97 (2013)
6. Bonnín Rosselló, C.: Brain lesion segmentation using Convolutional Neuronal Networks. B.S. thesis, Universitat Politècnica de Catalunya (2018)
7. Eisenhauer, E.A., Therasse, P., Bogaerts, J., Schwartz, L.H., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M., et al.: New response evaluation criteria in solid tumours: revised recist guideline (version 1.1). *European journal of cancer* **45**(2), 228–247 (2009)
8. Holland, E.C.: Progenitor cells and glioma formation. *Current opinion in neurology* **14**(6), 683–688 (2001)
9. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *CoRR abs/1412.6980* (2014)
10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. pp. 1097–1105 (2012)
11. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. pp. 3431–3440 (2015)
12. Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., Burger, P.C., Jouvet, A., Scheithauer, B.W., Kleihues, P.: The 2007 who classification of tumours of the central nervous system. *Acta neuropathologica* **114**(2), 97–109 (2007)
13. Maninis, K.K., Pont-Tuset, J., Arbeláez, P., Gool, L.V.: Deep retinal image understanding. In: *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Part II. Lecture Notes in Computer Science*, vol. 9901, pp. 140–148. Springer (2016)
14. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging* **34**(10), 1993 (2015)
15. Ohgaki, H., Kleihues, P.: Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. *Journal of Neuropathology & Experimental Neurology* **64**(6), 479–489 (2005)
16. Reddi, S.J., Kale, S., Kumar, S.: On the convergence of adam and beyond. In: *International Conference on Learning Representations* (2018)
17. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *CoRR abs/1409.1556* (2014)
18. Wen, P.Y., Macdonald, D.R., Reardon, D.A., Cloughesy, T.F., Sorensen, A.G., Galanis, E., DeGroot, J., Wick, W., Gilbert, M.R., Lassman, A.B., et al.: Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *Journal of clinical oncology* **28**(11), 1963–1972 (2010)

19. Xu, Y., Géraud, T., Bloch, I.: From neonatal to adult brain MR image segmentation in a few seconds using 3D-like fully convolutional network and transfer learning. In: Proceedings of the 23rd IEEE International Conference on Image Processing (ICIP). pp. 4417–4421. Beijing, China (September 2017)
20. Xu, Y., Géraud, T., Najman, L.: Connected filtering on tree-based shape-spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **38**(6), 1126–1140 (June 2016)
21. Xu, Y., Géraud, T., Puybareau, E., Bloch, I., Chazalon, J.: White matter hyperintensities segmentation in a few seconds using fully convolutional network and transfer learning. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries— 3rd International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Revised Selected Papers, Lecture Notes in Computer Science, vol. 10670, pp. 501–514. Springer (2018)

Ensembles of Multiple Scales, Losses and Models for Brain Tumor Segmentation and Overall Survival Time Prediction Task

Xuhua Ren¹, Lichi Zhang¹, Dinggang Shen² and Qian Wang¹

¹ Institute for Medical Imaging Technology, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China

² Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599
renxuhua@sjtu.edu.cn

Abstract. Deep learning approaches such as convolutional neural nets have consistently outperformed traditional methods such as instance, semantic segmentation. However, the various proposed networks perform differently, with largely influenced by architectural hyper-parameters and training settings. This paper explores ensembles of multiple scales, losses and models for robust performance through of predictions from numerous models. Moreover, we proposed an automated survival prediction method by utilizing different types of radiomics and hand-crafted features. The method tests feature significance and correlation values, and then utilizes the most significant features with a random forest auto-context regression model to perform survival prediction. Our framework wins the **1st** rank in the BRATS 2018 segmentation task and **3rd** rank in OS prediction task in validation set.

Keywords: Deep learning, Image segmentation, Survival prediction.

1 Introduction

Brain tumors are the most risky types of cancer [1]. These tumors that originally develop in the brain. They arise from glioma cells and according to their aggressiveness, these tumors are broadly categorized into high and low-grade gliomas [2]. High grade gliomas (HGG) develop rapidly and aggressively, forming abnormal vessels and often a necrotic core, accompanied by surrounding edema and swelling [3]. They are malignant, with high mortality and average survival rate of less than two years even after treatment. Low grade gliomas (LGG) can be benign or malignant, grow slower, but they may recur and evolve to HGG, thus their treatment is warranted. Magnetic resonance imaging (MRI) is widely used in both clinical routine and research studies. It facilitates tumor analysis by allowing estimation of extent, location and investigation of its subcomponents. This however requires accurate delineation of the tumor, which proves challenging due to its complex structure and appearance.

These factors make manual delineation time-consuming and subject to inter- and intra-rater variability [4].

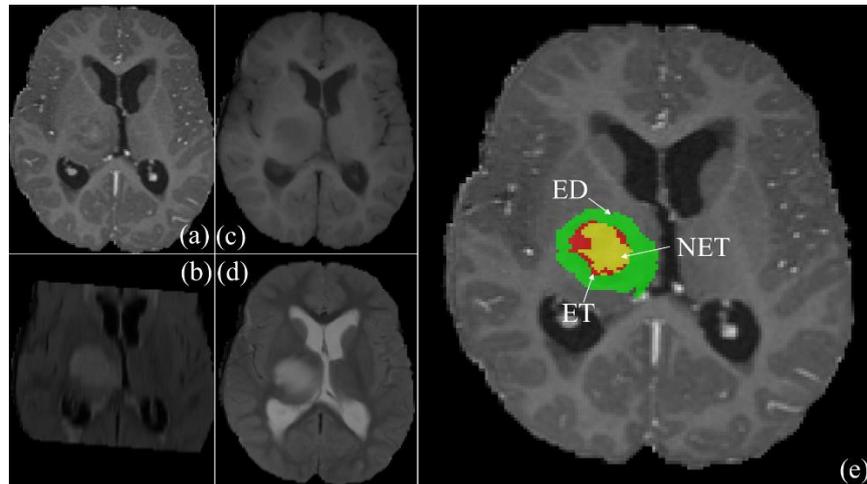


Fig. 1. Brain tumor segmentation via BRATS18 by using input (a) T1CE, (b) FALIR, (c) T1, and (d) T2 sequences. The segmented peritumoral edema/tumor infiltration, enhancing tumor, and non-enhancing tumor core are seen overlaid on (e).

Automatic segmentation systems aim at providing an objective solution. The early work for this task are the atlas-based outlier detection method [5]. The past few years saw rapid developments of machine learning methods, such as Random Forests being among the most successful [6]. More recently, convolutional neural networks (CNN) have gained popularity by exhibiting very promising results for segmentation of brain tumors [7]. A variety of CNN architectures have been proposed, each presenting different strengths and weaknesses. Additionally, networks have a vast number of hyper-parameters. The multiple configuration choices are very important for performance.

Several studies have proposed different methods for predicting the survivability of patients with brain tumors. In [8], the authors use the different subtype tumor volumes, the extent of resection, location, size and other imaging features in order to evaluate the capability of these features in predicting survival. While, most of the survival prediction studies utilize traditional regression models without machine learning ensemble algorithms, e.g., random forest.

In this work we push towards constructing a more reliable and deep learning model. We bring together a variety of CNN architectures, input sizes and losses in diverse ways to introduce high variance between them. We construct a framework with the aim of averaging bias and variance by ensemble models. Our approach leads to: (1) a system which is robust to samples of different distributions, (2) enables hard-mining for minority samples. Our method won the **first** position for segmentation task in validation set. Moreover, we used the proposed OS prediction pipeline to participate

in the BRATS 2018 of Survival Prediction and the method ranked **third** in validation set.

2 Tumor segmentation task

A variety of CNN architectures has shown good results in recent literature. Such architectural choices are very important influence for results. In this work, we utilized three types of CNN architectures: DeepMedic [9], Cascaded Anisotropic CNN [10] and Cascaded Anisotropic Unet [11].

2.1 DeepMedic

Model description: The first architecture we employ is DeepMedic, it is the 11-layers deep, multi-scale 3D CNN for medical image segmentation. The architecture consists of two parallel convolutional pathways that process the input at multiple scales to achieve a large receptive field for the final classification while keeping the computational cost low. Inspired by Residual, the use of residual connection is adopted. We only utilized the default parameter for DeepMedic version during these Challenge.

Training details: The wider variant is trained on larger inputs of width 34 and 22 for the two scales respectively. They are trained with cross-entropy loss, with all hyper-parameters adopted from the original configuration.

2.2 Cascaded Anisotropic Convolutional Neural Networks

Model description: They use three networks to sequentially segment sub-region of brain tumor, and each of these networks deals with a binary segmentation problem. The first network (WNet) segments the whole tumor from four modalities 3D image patches. Then a bounding box of the whole tumor is obtained. The cropped region would be used in the second network (TNet) for enhancing tumor (ET) and non-enhancing tumor (NET) which has the same architecture with WNet. Finally, the third network called ENet would separate ET and NET. Rely on the three networks, they could solve the BRATS segmentation task. Binary segmentation problems and take advantage of the hierarchical structure of tumor sub regions to reduce false positives. And, it uses dilated convolution, residual connection and multi-scale prediction to improve segmentation performance. They also propose to fuse the output of CNNs in three orthogonal views (i.e., WNet, TNet and ENet for three views: axial, sagittal, and coronal) for more robust segmentation of brain tumor.

Training details: Our networks were implemented in Tensorflow [12] using NiftyNet [13]. We used Adaptive Moment Estimation (Adam) for training. Batch size is 5, and maximal iteration 20k.

2.3 U-Net

Model description: We employ one versions of the U-Net architecture in our ensemble models. In this version we use residual block to increase model complexity, where residual skip connections [14] are implemented via summations of the signals in the up-sampling part of the network, all layers use batch normalization [15], RELU [16] and zero-padding. We also use three networks to hierarchically and sequentially segment substructures of brain tumor, and each of these networks deals with a binary segmentation problem, as same as Cascaded Anisotropic Convolutional Neural Networks.

Training details: Our networks were implemented in Tensorflow using NiftyNet. We used Adaptive Moment Estimation (Adam) for training. Batch size is 5, and maximal iteration 20k.

2.4 Hybrid Loss

The **Dice** similarity coefficient (DSC) measures the amount of agreement between two image regions. It is widely used as a metric to evaluate the segmentation performance with the given ground truth in medical images. The DSC is defined in (1), we utilize $||$ to indicate the number of foreground voxels in the ground truth and segmentation images.

$$\text{DSC} = \frac{2|S \cap R|}{|S| + |R|}$$

where S is the segmentation result and R is the corresponding ground truth label. This function however is not differentiable and hence cannot directly be used as a loss function in deep learning. Hence, continuous versions of the Dice score have been proposed that allow differentiation and can be used as loss in optimization schemes based on stochastic gradient descent:

$$LD = -\frac{2 \sum_i^N s_i r_i}{\sum_i^N s_i + \sum_i^N r_i}$$

We also address to add a modulating factor $(1 - p_t)^\gamma$ to the cross-entropy loss, with tunable focusing parameter $\gamma \geq 0$. The focal loss [17] definition as:

$$FL = -1(1 - p_t)^\gamma \log(p_t)$$

Finally, we propose a **hybrid loss** HL which add dice loss and focal loss together.

$$HL = LD + FL$$

2.5 Ensemble

The above models are all trained completely separately. At testing time, each model segments individually an unseen image and outputs its class-confidence maps. We utilize three level ensemble strategies: input image sizes, CNN models and losses. So, we have 16 number of models in the finally. Considering each of Cascade CNN has three direction CNNs, we have 49 single model CNNs for the three-sub-region segmentation task.

Table 1. Ensemble strategies

Item	DeepMedic	Cascade CNN			Cascade Unet		
		WNET	TNET	ENET	WNET	TNET	ENET
Input	34x34x34 & 22x22x22	144x144x19	72x72x19	64x64x19	144x144x19	72x72x19	64x64x19
		144x144x29	72x72x29	64x64x29	144x144x29	72x72x29	64x64x29
		72x72x19	64x64x19	32x32x19	72x72x19	64x64x19	32x32x19
		72x72x29	64x64x29	32x32x29	72x72x29	64x64x29	32x32x29
Loss	Entropy	Dice		Hybrid	Dice		Hybrid

2.6 Postprocess

Finally, the above models are all trained completely separately. At testing time, each model segments individually an unseen image and outputs its class-confidence maps. Then we ensemble all the models results into one result with some postprocessing: get rid of tiny sub regions; get rid of sub regions which has large eccentricity.

3 OS time prediction task

3.1 Feature Extraction

We extract many features from the raw MRI modality (T1C, T2, T1 and Flair). The features are described by the First Order Features, the Shape Features, the Gray Level Co-occurrence Matrix Features, the Gray Level Size Zone Matrix Features, the Gray Level Run Length Matrix Features, the Gray Level Dependence Matrix Features and Neighbouring Gray Tone Difference Matrix Features [8], all the features are extracted from Pyradiomics [18]. In addition, we extract some hand-crafted features including volumetric features, histogram-based statistics (mean, variance, skewness, kurtosis, energy, and entropy) features are extracted from the tumor and the different tumor sub-regions (ED, ET, and NCR/NET). Further, the tumor locations and the spread of the tumor in the brain are also considered. These features represent the frequency at different intensity bins (number of bins = 11, 23, 46). Finally, we have 2380 number of features in the feature space.

3.2 Feature Extraction

Feature selection is performed in three steps; first, we utilize Pearson Correlation Coefficient to reduce the redundancy of feature space. Then, the significant features are selected by VarianceThreshold, Univariate feature selection, Recursive feature elimination and

Feature selection using SelectFromModel algorithms. We come true all the feature selection algorithms using scikit-learn [19] and we can obtain 35 number of features after the correlation reducing and feature selection. Finally, the 28 features are reduced from 35 features using a recursive feature selection algorithm.

3.3 Random Forest Regression

The 28 features are used with Random Forest [20] regression model for survival prediction. We also utilize a strategy called auto-context [21, 22] to increase the robustness of model.

4 Evaluation

4.1 Material

Our system was evaluated on the data from the Brain Tumor Segmentation Challenge 2018 (BRATS) [4, 23, 24, 25]. The training set consists of 210 cases with high grade glioma (HGG) and 75 cases with low grade glioma (LGG), for which manual segmentations are provided. The segmentations include the following tumor tissue labels: 1) necrotic core and non-enhancing tumor, 2) edema, 4) enhancing core. Label 3 is not used. The validation set consists of 66 cases, both HGG and LGG but the grade is not revealed. Reference segmentations for the validation set are hidden and evaluation is carried out via an online system that allows multiple submissions. For evaluation, the 3 predicted labels are merged into different sets of whole tumors (all labels), the core (labels 1,4) and the enhancing tumor (label 4). For each subject, four MRI sequences are available, FLAIR, T1, T1 contrast enhanced (T1ce) and T2. The datasets are pre-processed by the organizers and provided as skull-stripped, registered to a common space and resampled to isotropic $1mm^3$ resolution. Dimensions of each volume are $240 \times 240 \times 155$.

4.2 Results

We provide the segmentation results that ensemble results achieved on the validation set of the BRATS'18 challenge on Table 2. We show results achieved on the validation set and two teams that ranked in the next two positions at the validation stage.

Table 2. Performance of our method on the validation set of BRATS 2018

	DSC			Hausdorff 95		
	Enh	Whole	Core	Enh	Whole	Core
Our (1st)	0.84	0.91	0.87	3.4	8.2	6.7
2nd	0.79	0.91	0.85	3.8	5.3	7.2
3rd	0.80	0.91	0.82	2.8	4.5	7.0

Whereas with the BRATS18 validation dataset, survival prediction features are extracted from the fused segmented tumor outcome. Table 3 shows the online evaluation results we achieve 3rd place with the BRATS 18 validation dataset.

Table 3. Performance of our method on the validation and set of BRATS 2018

	Acc	MSE	MedianSE	StdSE
Our (3rd)	0.429	104086	51894	138111
1st	0.536	101012	51006	140511
2nd	0.464	192116	115442	198286

Conclusion

Neural networks have been proven very impotent, often making unpredictable errors. For this reason, we first concentrate on improving robustness and using three levels ensemble strategies including input image sizes, CNN models and losses. Our method won the **1st rank** in the validation stage of BRATS 2018 competition, indicating strong generalization. Finally, the segmented tumor volumes and its' sub regions along with other structural, texture, volumetric, and histogram features derived from these are used in an RF regression model for survival prediction. We use the proposed pipeline to participate in the Task of Survival Prediction and obtain the **3rd rank**.

References

1. DeAngelis, L.M.: Brain tumors. *New England Journal of Medicine* 344, 114-123 (2001)
2. Bauer, S., Wiest, R., Nolte, L.-P., Reyes, M.: A survey of MRI-based medical image analysis for brain tumor studies. *Physics in Medicine & Biology* 58, R97 (2013)
3. Louis, D.N., Perry, A., Reifenberger, G., Von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., Ellison, D.W.: The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta neuropathologica* 131, 803-820 (2016)
4. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging* 34, 1993 (2015)
5. Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. *Medical image analysis* 8, 275-283 (2004)
6. Zikic, D., Glocker, B., Konukoglu, E., Criminisi, A., Demiralp, C., Shotton, J., Thomas, O.M., Das, T., Jena, R., Price, S.J.: Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 369-376. Springer, (Year)
7. Kleesiek, J., Biller, A., Urban, G., Kothe, U., Bendszus, M., Hamprecht, F.: Ilastik for multi-modal brain tumor segmentation. *Proceedings MICCAI BraTS (Brain Tumor Segmentation Challenge)* 12-17 (2014)
8. Aerts, H.J., Velazquez, E.R., Leijenaar, R.T., Parmar, C., Grossmann, P., Carvalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D.: Decoding tumour phenotype by

- noninvasive imaging using a quantitative radiomics approach. *Nature communications* 5, 4006 (2014)
9. Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A.V., Criminisi, A., Rueckert, D., Glocker, B.: DeepMedic for brain tumor segmentation. In: *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 138-149. Springer, (Year)
 10. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: *International MICCAI Brainlesion Workshop*, pp. 178-190. Springer, (Year)
 11. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*, pp. 234-241. Springer, (Year)
 12. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M.: Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467* (2016)
 13. Gibson, E., Li, W., Sudre, C., Fidon, L., Shaker, D.I., Wang, G., Eaton-Rosen, Z., Gray, R., Doel, T., Hu, Y.: NiftyNet: a deep-learning platform for medical imaging. *Computer methods and programs in biomedicine* 158, 113-122 (2018)
 14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. (Year)
 15. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015)
 16. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807-814. (Year)
 17. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. *arXiv preprint arXiv:1708.02002* (2017)
 18. van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G., Fillion-Robin, J.-C., Pieper, S., Aerts, H.J.: Computational radiomics system to decode the radiographic phenotype. *Cancer research* 77, e104-e107 (2017)
 19. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V.: Scikit-learn: Machine learning in Python. *Journal of machine learning research* 12, 2825-2830 (2011)
 20. Breiman, L.: Random forests. *Machine learning* 45, 5-32 (2001)
 21. Tu, Z.: Auto-context and its application to high-level vision tasks. In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1-8. IEEE, (Year)
 22. Zhang, L., Wang, Q., Gao, Y., Wu, G., Shen, D.: Automatic labeling of MR brain images by hierarchical learning of atlas forests. *Medical physics* 43, 1175-1186 (2016)
 23. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific data* 4, 170117 (2017)
 24. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. *The Cancer Imaging Archive*, (2017)

25. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. The Cancer Imaging Archive, (2017)

Adversarial Framework for Learning Multiple Clinical Tasks

Mina Rezaei^[0000–0001–6994–6345] and Christoph Meinel

Hasso Plattner Institute, Berlin, Germany
mina.rezaei@hpi.de, christoph.meinel@hpi.de

Abstract. Inspired by the recent success of generative adversarial networks (GANs), we propose a new adversarial network called voxel-GAN for semantic segmentation of brain tumour. We introduce a conditional generative adversarial network (cGAN) for brain tumour semantic segmentation and prediction of the patient overall survival. The proposed network comprises three components a segmentor, a discriminator, and a survival predictor. The segmentor is trained on 3D brain magnetic resonance images (MRI) to learn the segmentation label's in voxel level, while the discriminator is trained to discriminate a segmentor output, coming from the ground truth or generated artificially. The segmentor network and the discriminator network simultaneously train via adversarial loss, with two-player mini-max games. On the other hand, the survival predictor network is trained on statistical features of 3D tumours region predicted by the segmentor, and clinical data to predict the time of survival. We measure survival time in days, based on deep neural network of Cox Proportional Hazard model (CoxPH).

Keywords: 3DGAN, Deep CoxPH Model, Brain Tumour Segmentation, Prediction of Patient Overall Survival

1 Introduction

Automating brain tumour segmentation is a challenging task due to the high diversity in the appearance of tumour tissues among different patients, and in many cases, the similarity with the normal tissues. An accurate brain tumour segmentation algorithm might be able to improve the prediction accuracy and efficiency.

As mentioned by Menze et al. [21], the number of clinical study for automatic brain tumour segmentation has grown significantly in the last decades. We can roughly divide the current automated algorithms into two categories: those based on generative models and those based on discriminative models. Generative probabilistic approaches build the statistical model based on natural features of dataset such as categories and dimensions. Eventually, generative models discover the complex regularities of the natural data. Traditionally, these models have been applied where simple conditionally independent Gaussian models [9] or Bayesian learning [25] is used for tissue appearance. On the

other hand, discriminative models directly learn the relationship between the local features of images and segmentation labels. Traditional discriminative approaches such as SVMs [5], random forests [8] have been used in medical image segmentation. Deep neural networks (DNNs) are one of the most popular discriminative approaches, where the machine learns the hierarchical representation of features without any handcrafted features [19]. In the field of medical image segmentation, Ronneberger et al. [26] presented a fully convolutional neural network, named UNet, for segmenting neuronal structures in electron microscopic stacks.

One of the major challenges in medical image segmentation is the existence of unbalanced data where non-healthy class pixels are often fewer than healthy pixels. A trained network of unbalanced data, makes predictions with high precision and low recall which is undesired in medical routine applications. Among the extensive usage of UNet for automated medical image segmentation, there are methods that try to mitigate the imbalanced data issue by cascade training [13], equal selection of training samples [12], samples re-weighting [24], or similarity loss functions [26].

Recently, in the machine learning community GANs [11] have gained a lot of momentum. Classical GANs consist of two models: a generative model and a discriminative model. The generator model takes random input values and transforms them into realistic images. The discriminator model is a classifier that determines whether a given image from generator output looks like a real image from the dataset or like an artificially created image. Mirza et al. [22] extended the GANs to the conditional setting by adding the generator network class-conditional. Conditional GANs (cGANs) have the advantage of being able to provide better representations for data generation since there is control on the modes of the data being generated. This makes cGANs suitable for image segmentation tasks.

Xue et al. [27] presented a cGAN, called SegAN, for medical image segmentation using a multi-scale loss function for the generative network. SegAN utilizes UNet as the generator and FCN as the discriminator. Moeskops et al. [23] used dilated convolution as the generator and FCN as the discriminator through adversarial training for 2D brain MRI segmentation. Adversarial networks achieved good results on small datasets for prostate cancer detection [18] and mammography mass segmentation [28]. Isola et al. [14] introduced cGAN named pix-to-pix as a general setting for image to image translation. They used UNet as the generator and FCN as the discriminator. However, pix-to-pix achieved good results in semantic segmentation when the data is balanced such as ImageNet [14].

Unlike previous works [14, 27, 23], we introduce an adversarial framework that takes 3D medical images and outputs 3D semantic segmented tumours region. In our proposed framework, the segmentor network captures tumours structure implicitly and synthesize high-quality 3D tumorous region. To mitigate the problem of imbalanced data in semantic segmentation of brain tumour(s) and achieve a much better trade-off between precision and recall, our segmentor loss

is squarely weighted and mixed with traditional with ℓ_1 term. The discriminator distinguish a synthesized voxel created by segmentor is real or fake.

Moreover, in this work we address the patient overall survival. Survival analysis is a branch of statistics for analyzing and modeling method to estimate the time until an event of interest will occur, such as death in biological organisms or failure in mechanical systems. The most common survival analysis modeling techniques are based on the Kaplan-Meier (KM) model [15], and Cox Proportional Hazard (CoxPH) model [6]. Whereas the Kaplan-Meier method provides a survival based on log-rank and compute the survival curves in two or more groups without considering the feature vectors, therefore those models are not able to predict the survival function of a specific unit. In contrast, the CoxPH allows analyzing the effect of several risk factors on survival and provides prediction on the survival and hazard functions of an individual based on its feature vector.

Our proposed survivors predictor network is trainable multi-layers preceptron with some hidden layers and takes statistical feature vector from 3D segmented tumours predicted by segmentor, and clinical data such as the age of patients, and the resection status. The output of survivors network is patient overall survival days based on negative log partial likelihood; similar to the DeepSurv model [16].

2 Method

2.1 Semantic Segmentation using Voxel-GAN

In a conventional generative adversarial network, generative model G tries to learn a mapping from random noise vector z to output image y ; $G : z \rightarrow y$. Meanwhile, a discriminative model D estimates the probability of a sample coming from the training data x_{real} rather than the generator x_{fake} . The GAN objective function is a two-player mini-max game like Eq.(1).

$$\min_G \max_D V(D, G) = E_y[\log D(y)] + E_{x,z}[\log(1 - D(G(x, z)))] \quad (1)$$

Similar conditional GAN [22]; in our proposed method, segmentor network learns mapping from a given 3D multimodal MR images x and Gaussian vector z to a 3D semantic segmentation y_{seg} ; $S : \{x, z\} \rightarrow \{y_{seg}\}$. The training procedure for the segmentation task is similar to two-player mini-max game as shown in Eq.(2). While the segmentor generated segmentation voxels label, the discriminator takes the ground truth annotated by an expert, and the segmentor's output to distinguish segmentor output is real or fake.

$$\mathcal{L}_{adv} \leftarrow \min_S \max_D V(D, S) = E_{x,y_{seg}}[\log D(x, y_{seg})] + E_{x,z}[\log(1 - D(x, S(x, z)))] \quad (2)$$

To mitigate the problem of unbalanced data in medical image segmentation and achieve a much better trade-off between precision and recall, the segmentor loss is weighted by considering number of classes and mixed with ℓ_1 term.

$$\mathcal{L}_{L1}(S) = E_{x,z} \| y_{seg} - S((x)^w, z) \| \quad (3)$$

We investigate square weighting on class voxel frequencies for the whole training dataset as follows:

$$W = \frac{N}{C |v_c|^2 + \epsilon} \quad (4)$$

where N indicates the number of training volumes, C is the number of semantic classes (e.g in BraTS $C = 4$), and $|v_c|$ is the number voxels in class. We consider $\epsilon = 1$ in order to prevent division by zero.

The segmentor loss Eq.(4) is mixed with ℓ_1 term to minimize the absolute difference between the predicted value and the existing largest value. Previous studies [14, 27] on cGANs have shown the success of mixing the cGANs objective with ℓ_1 distance. Hence, the ℓ_1 objective function takes into account CNN feature differences between the predicted segmentation and the ground truth segmentation and resulting in fewer noises and smoother boundaries.

The final objective function for semantic segmentation of brain tumours \mathcal{L}_{seg} calculated by adversarial loss and additional segmentor ℓ_1 loss as follows:

$$\mathcal{L}_{seg}(D, S) = \mathcal{L}_{adv}(D, S) + \mathcal{L}_{L1}(S) \quad (5)$$

2.1.1. Segmentor Network In our proposed architecture, the segmentor is a 3D fully convolutional encoder-decoder network that generates a label for each voxel. The encoder takes $64 \in 64 \in 64$ of multi-modal multisite MRI/CT images and decoder output size is $64 \in 64 \in 64$. Similar to UNet [26], we added the skip connections between each layer i and layer $n - i$, where n is the total number of layers. Each skip connection simply concatenates all channels at layer i with those at layer $n - i$.

2.1.2. Discriminator Networks Similar to the segmentor, the discriminator is 3D fully convolutional encoder network which classifies whether a generated voxel's label belongs to right class.

2.2 Survival Analysis

2.2.1. Background In statistical modeling of survival analysis [6, 16, 10], a survival data is comprised to three elements of a patients baseline data x , a failure event time T , and an event indicator E . The $E = 1$, if an event such as death observed, the time interval T corresponds to the time elapsed between the time in which the baseline data was collected. Here in BraTS [4, 2, 3] survival prediction T is a number of days, patient data includes an ages, a resection

status, and a 3D MRI preoperative images. The event indicator is equal zero ($E = 0$), when in middle of study the patient quite and study get ended which in current experiment we don't have such cases.

In survival studies based on CoXPH model, the survival function Eq.(6), and hazard function Eq.(7) are the two fundamental functions.

$$S(t) = Pr(T > t) \quad (6)$$

The survival function shows the probability that an individual has survived beyond time t . The hazard function (or conditional failure rate, conditional mortality rate) approximates probability that an event occurs in the small time interval $[t; t + \Delta t]$, under the condition that an individual would remain event-free up to time t :

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{Pr(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \quad (7)$$

2.2.2. Survivors Predictor Network We design a multi-layers preceptron to address patient survival days. Our architecture takes statistical feature vector of predicted 3D voxel of tumours region by segmentor network, patient ages, and resection status as an input and estimate the risk function $h_\theta(x)$ parameterized by the weights of the network θ . Similar to the previous network [10, 16], we set the loss function to be the negative log partial likelihood and predict survival days as follows:

$$\mathcal{L}_{sv}(\theta) = - \sum_{i:E_i=1} h_\theta(x_i) - \log \sum_{j \in R(T_i)} exp^{h_\theta(x_j)} \quad (8)$$

3 Experiments

In order to evaluate the performance of the proposed network on task of segmentation, we applied the ISLES-2018 [20, 17], and the BraTS-2018 [4, 21, 3, 2] challenges datasets. Moreover, we test the performance of the proposed survival prediction algorithm on the BraTS-2018, for predicting overall survival days for patient with brain cancer.

3.1 Implementation Details

Our proposed Voxel-GAN, and survival predictor are implemented based on a Keras [7] library with backend of Tensorflow [1] supporting 3D convolutional network and is publicly available ¹. All training and experiments were conducted on a workstation equipped with a multiple NVIDIA TITAN X GPUs. The learning rate was initially set to 0.001. The Adadelta optimizer is used in both the segmentor and the discriminator. The model is trained for up to 80 epochs.

¹ <https://github.com/HPI-DeepLearning/VoxelGAN>

3.2 Data Preprocessing and Data Augmentation

The proposed Voxel-GAN takes a set of 3D images acquired from 4 different MRI modalities; Flair, T1, T1ce, and T2 as different channels. To reduce the effect of the absolute pixel intensities to the model, we applied intensity normalization on each MRI modality from individual patient by subtracting the mean and dividing by the standard deviation of the brain region.

To prevent over fitting issue, we applied randomly cropped, re-size, scaling, rotation between -10 and 10 degree, and Gaussian noise as data augmentation.

3.3 Evaluation

We followed the evaluation criteria introduced by the BraTS ², the ISLES ³ challenge organizer.

3.2.1. Brain Tumour Segmentation The segmentation of the brain tumor from medical images is highly interesting in surgical planning and treatment monitoring; The goal of segmentation is to delineate different tumor structures such as active tumorous core, enhanced tumorous, and whole tumorous. From Table (1), the proposed Voxel-GAN achieved better results in terms of Dice compared to 2D-cGAN. One likely explanation is that the Voxel-GAN architecture is trained on 3D convolutional features and the segmentor loss is weighted for imbalanced data. It worth to mention the reported results on Table (1), is based trained model with 80 epochs and 5 fold cross validation in terms of Dice, Hausdorff distance, sensitivity, and specificity.

Table 1. The achieved accuracy for semantic segmentation by Voxel-GAN in terms of Dice, sensitivity, specificity, and Hausdorff distance on five fold cross validation after 80 epochs while the reported results in second and third rows are after 120 epochs.

Methods	Dice			Hdff			Sen			Spec		
	WT	ET	CT	WT	ET	CT	WT	ET	CT	WT	ET	CT
Voxel-GAN	0.84	0.63	0.79	6.41	7.1	10.38	0.86	0.74	0.78	0.99	0.99	0.99
cGAN	0.81	0.61	0.64	7.30	9.22	12.04	0.75	0.61	0.55	0.99	0.99	0.99

3.2.2. Prediction of Patient Overall Survival The models input features include patients age, resection status as well as morphological (volume), ratio of number of voxels of whole tumours region to number of voxels of lesion, ratio of number of voxels of tumour core to number of voxels of lesion, ratio of number of

² <http://www.med.upenn.edu/sbia/brats2018/evaluation.html>

³ <https://www.smir.ch/ISLES/Start2018>

Table 2. The achieved accuracy for semantic segmentation on ISLES dataset by Voxel-GAN in terms of Dice, Hausdorff distance, Precision, and Recall on five fold cross validation after 140 epochs.

	Dice	Hausdorff	Precision	Recall
Voxel-GAN	0.83	10.3	0.81	0.78

voxels of enhanced tumour to number of voxels of lesion, etc. were fed as input to multi-layers perceptron with 5 different hidden layers. The achieved accuracy for local test data in term of concordance index or C-index is 0.58.

4 Conclusion and Future Work

In this paper, we introduced a new 3D generative adversarial network to mitigate the issue of unbalanced data in brain tumour segmentation task. To this end, we proposed a segmentor network and couple discriminator networks where a segmentor generates voxel's label, and discriminator classifies whether segmentor output is real or fake. Whilst, another discriminator called survival predictor network, trained on statistical feature produced by segmentor and clinical data such as patient age, resection status and learned to predict survival days using negative log of likelihood. The proposed segmentor loss function squarely weighted on imbalanced class labels and shown good trade-off between achieved precision and recall on the ISLES dataset. Moreover, We applied preprocessing and provided data augmentation to prevent overfitting issue. Experimental evaluation of the BraTS-2018 and ISLES-2018 benchmarks show that the proposed model achieved acceptable results for segmentation of brain tumours.

References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al.: Tensorflow: a system for large-scale machine learning. In: OSDI. vol. 16, pp. 265–283 (2016)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017). <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive (2017). <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data (2017)

5. Bauer, S., Nolte, L.P., Reyes, M.: Fully Automatic Segmentation of Brain Tumor Images Using Support Vector Machine Classification in Combination with Hierarchical Conditional Random Field Regularization, pp. 354–361. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
6. Breslow, N.: Covariance analysis of censored survival data. *Biometrics* pp. 89–99 (1974)
7. Chollet, F., et al.: Keras (2015)
8. Criminisi, A., Shotton, J., Konukoglu, E., et al.: Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends in Computer Graphics and Vision* **7**(2–3), 81–227 (2012)
9. Fischl, B., Salat, D.H., van der Kouwe, A.J., Makris, N., Ségonne, F., Quinn, B.T., Dale, A.M.: Sequence-independent segmentation of magnetic resonance images. *Neuroimage* **23**, S69–S84 (2004)
10. Fotso, S.: Deep neural networks for survival analysis based on a multi-task framework. arXiv preprint arXiv:1801.05512 (2018)
11. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Networks. ArXiv e-prints (2014)
12. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H.: Brain tumor segmentation with deep neural networks. *Medical image analysis* **35**, 18–31 (2017)
13. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. 2017 International MICCAI BraTS Challenge (2017)
14. Isola, P., Zhu, J., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. *CoRR* **abs/1611.07004** (2016)
15. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. *Journal of the American statistical association* **53**(282), 457–481 (1958)
16. Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Y.: Deep-surv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology* **18**(1), 24 (2018)
17. Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., Büchler, P.: The virtual skeleton database: an open access repository for biomedical research and collaboration. *Journal of medical Internet research* **15**(11) (2013)
18. Kohl, S., Bonekamp, D., Schlemmer, H., Yaqubi, K., Hohenfellner, M., Hadaschik, B., Radtke, J., Maier-Hein, K.H.: Adversarial networks for the detection of aggressive prostate cancer. *CoRR* **abs/1702.08014** (2017)
19. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
20. Maier, O., Menze, B.H., von der Gabelntz, J., Häni, L., Heinrich, M.P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., et al.: Isles 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral mri. *Medical image analysis* **35**, 250–269 (2017)
21. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging* **34**(10), 1993–2024 (2015)
22. Mirza, M., Osindero, S.: Conditional generative adversarial nets. *CoRR* **abs/1411.1784** (2014)
23. Moeskops, P., et al.: Adversarial Training and Dilated Convolutions for Brain MRI Segmentation, pp. 56–64. Springer International Publishing, Cham (2017)

24. Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., Glocker, B.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers. vol. 10670, p. 450. Springer (2018)
25. Pohl, K.M., Fisher, J., Grimson, W.E.L., Kikinis, R., Wells, W.M.: A bayesian model for joint segmentation and registration. *NeuroImage* **31**(1), 228–239 (2006)
26. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer International Publishing (2015)
27. Xue, Y., Xu, T., Zhang, H., Long, L.R., Huang, X.: Segan: Adversarial network with multi-scale l1 loss for medical image segmentation. *CoRR* **abs/1706.01805** (2017)
28. Zhu, W., Xie, X.: Adversarial deep structural networks for mammographic mass segmentation. *CoRR* **abs/1612.05970** (2016), <http://arxiv.org/abs/1612.05970>

Brain Tumour Segmentation Method based on Sparse Feature Vectors

J.P. Serrano-Rubio¹, Richard Everson², and Hugo Hutt²

¹ Technological Institute of Irapuato,
Information Technologies Laboratory, Irapuato, Guanajuato, Mexico.
juserrano@itesi.edu.mx

² College of Engineering, Mathematics and Physical Sciences,
University of Exeter, United Kingdom.
{R.M.Everson,hwh202}@exeter.ac.uk

Abstract. This paper presents an automatic method for brain tumour segmentation from magnetic resonance images. The method uses the feature vectors obtained by an efficient feature encoding approach which combines the advantages of the supervoxels and sparse coding techniques. Extremely Randomized Trees (ERT) is trained using these feature vectors to detect the whole tumour and for multi-label classification of abnormal tissues. The obtained predictions of the ERT are used to estimate probability maps. A Conditional Random Field (CRF) algorithm is implemented to complement the segmentation task. The probability maps of the images and the Euclidean distance between the feature vectors of neighbour supervoxels define the conditional random field energy function. The minimization of the energy function is performed via graph cuts. The proposed methods are evaluated on real patient data obtained from BraTS 2018 challenge.

Keywords: Brain Tumour Segmentation, Sparse Coding Techniques, Conditional Random Fields

1 Introduction

Gliomas is a type of brain tumours with the highest mortality rate in adults. The gliomas are classified according to their histopathological appearances into Low Grade Gliomas (LGG) and High Grade Gliomas (HGG) for determining the best treatment for the patient [1]. Furthermore the treatment of gliomas depend on its size and location within the brain, since the tumours can grow and infiltrating over the surrounding normal brain tissue what sometimes it complicates its surgical removal.

One of the most important and critical procedures for the diagnosis of brain tumours is the brain segmentation. Brain segmentation has been useful for the analysis and visualization of brain structures [3, 2] with the purpose of monitoring, guidance and planing the surgeries [4]. The goal of the brain tumour segmentation is to delineate the pathological regions such as the peritumoral

edema, non-enhancing tumour and enhancing tumour. Nowadays brain tumour segmentation methods are divided into two categories: a) semi-automatic and b) automatic methods. Semi-automatic methods are extremely costly due to that they requires an expert to detect the tumour. Automatic methods automatically detect and segment the brain tumours by using machine learning algorithms which can assign each tissue to its respectively class. Brain tumour segmentation for automatic methods can be a challenge task, because the tumours can appear anywhere in the brain and they can present a wide number of shapes and sizes.

In this paper we propose a fully automatic method for segmentation of brain tumours. This method incorporates an efficient feature encoding approach which is based on three algorithms: 1) sparse coding technique, 2) a Gaussian pyramid and 3) supervoxels. 3D image patches of images are projected into sparse dictionary spaces in order to obtain feature vectors which are named as sparse feature vectors. For classification purposes each feature vector is assigned to only one supervoxel. An Extremely Randomly Trees (ERT) algorithm is trained using the sparse feature vectors to detect the whole tumour and for multi-label classification of abnormal tissues. The tumour segmentation method is performed in two phases:

- The whole tumour is detected for identifying the region where is located the tumour.
- A multi-label classification is performed on the delimited region where is located the tumour in order to recognize the enhancing tumour, peritumoral edema and enhancing tumour.

A Conditional Random Field (CRF) is implemented to complement the segmentation task. The obtained predictions of ERT are used to estimate probability maps for each MRI. The energy function of the CRF is defined by the probability maps and the euclidean distance between the sparse feature vectors of neighbour supervoxels. The minimization of the energy function is performed via graph cuts.

For convenience of the reader, this paper is organized as follows: section 2 presents the details of the MRI database which is employed to evaluated the proposal method and Section 3 presents the feature learning approach to obtain the sparse feature vectors for normal and abnormal tissue and the brain tumour segmentation model. Section 4 presents the experimental details and results. Finally section 5 presents the conclusions.

2 BraTS 2018 Database

The database of magnetic resonance images is adopted from BraTS 2018 challenge as part of the International Conference on Medical Image Computing and Computer-Assisted Interventions (MICCAI) conference [3, 5–7]. The magnetic resonance images are 3D volumes whose dimension is $240 \times 240 \times 155$. For each brain, four MRI modalities are available, Flair, T1, T1 contrast enhanced

(T1C), and T2. The training dataset consists of 210 brains with high grade glioma (HGG) and 75 brains with low-grade glioma (LGG). Each brain has a manual segmentation which includes the following tissue labels: for normal tissue and background regions (label 0), non-enhancing tumour (label 1), peritumoral edema (label 2) and enhancing tumour (label 4). The validation dataset consists of 66 brains whose grade of the magnetic resonance images is not described.

3 Method

Each MRI is normalized by setting the mean to 0 and standard deviation to 1 of the voxel intensities. Our method is based on the following steps:

1. Feature Learning Approach.
 - Calculate sparse dictionaries for each MRI modality.
 - Calculate supervoxels.
 - Obtain the sparse feature vectors.
2. Gliomas Segmentation
 - Sparse feature vectors are classified using an ERT.
 - Probability maps are estimated using the obtained predictions of the ERT model.
 - Complement the segmentation task using a CRF.

This method is repeated two times. Firstly we perform this method to delimit the region where is located the tumour. Secondly we classify the abnormal tissue on the delimited region in order to obtain the final tumour segmentation.

3.1 Feature Learning Approach

Sparse dictionaries The feature learning approach incorporates a sparse coding technique and a Gaussian pyramid to learn a set of sparse dictionaries over different scales of the BraTS 2018 database. The sparse coding captures structures and patterns of the data as well as defining an efficient representation (sparse dictionaries) of the input images. A Gaussian pyramid is implemented to generate a sequence of images for each MRI modality over three different levels $l = 1, 2, 3$ where the level $l = 1$ represent the original image. Our approach uses the four MRI modalities of each patient, therefore we obtain 4 sparse dictionaries for each level of the Gaussian pyramid. The sparse coding technique finds a set of basis vector (sparse dictionary) such that an input vector can be represent as a linear combination of these basis vectors. Let \mathbf{I} be a set of images at level l of the Gaussian pyramid. 3D image patches $p_1, p_2, \dots, p_{\psi-1}, p_{\psi} \in \mathbb{R}^{k \times k \times k}$ are randomly selected from \mathbf{I} . All 3D image patches are reshaped into the vector $w_i \in \mathbb{R}^{k^3}$. The sparse dictionary $\mathbf{B} \in \mathbb{R}^{k^3 \times n}$ is obtained by using the sparse coding technique over n linear filters as follows:

$$\underset{\{\mathbf{B}, a_i\}}{\text{minimize}} \sum_{i=1}^{\psi} \underbrace{\|w_i - \mathbf{B}a_i\|_2^2}_{\text{residue function}} + \underbrace{\beta \|a_i\|_1}_{\text{sparsity function}}$$

subject to $\| \mathbf{B}_{(:,j)} \|_2 = 1 \forall_j$

where a_i is a coefficient vector, $\mathbf{B}_{(:,j)}$ is the j -th column of the sparse dictionary and β is the parameter which controls the sparsity of the solution [8]. Figure 1 illustrates the linear combination for representing the image patch p_i and one component of the sparse dictionary.

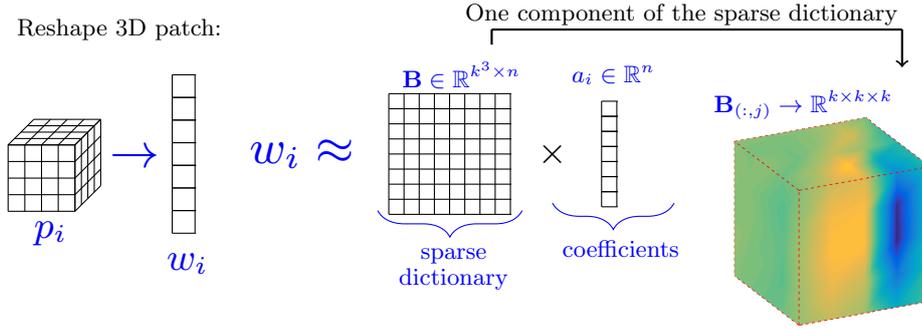


Fig. 1. Linear combination for representing the image patch p_i as a linear combination of a basis vectors (sparse dictionary).

Supervoxels and Sparse Feature Vectors The MRIs are partitioned into regions of perceptually similar voxels called supervoxels [9] in order to decrease the computational complexity of the images and to obtain meaningful structure of brain regions due to that they present an irregular shape which provides a better alignment with the tissue boundaries [10]. We incorporate the four MRI modalities to calculate the supervoxels using an extension of the Superpixels Extracted via Energy-Driven Sampling (SEEDS) algorithm [11]. The set of supervoxels is denoted as $\mathbf{S} = \{s_1, s_2, \dots, s_{\eta-1}, s_{\eta}\}$. The label c for each supervoxel is assigned by majority vote. Connected-Component Label algorithm is implemented to valid that all supervoxels are spatially connected as an only set. We use the supervoxels as atomic units to extracting features for normal and abnormal brain tissue. Figure 2 shows the procedure to obtain a feature vector d_{s_i} for the supervoxel s_i using only one sparse dictionary. Let $\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_{\kappa-1}, \mathbf{Q}_{\kappa} \in \mathbb{R}^{k \times k \times k}$ a set of 3D image patches which are located into the regions delimited by a supervoxel. The 3D image patches are reshaped to vectors $q_1, q_2, \dots, q_{\kappa-1}, q_{\kappa} \in \mathbb{R}^{k^3}$ and projected into the sparse dictionary space. The feature vector d_{s_i} is obtained with a maxpooling technique which uses the highest value of the positive and negative components of the sparse vectors $u_1, u_2, \dots, u_{\kappa}$. Note that for each sparse dictionary we calculate a feature vector for each supervoxel, therefore 12 feature vectors describe each supervoxel. These feature vectors are concatenated to assign only one feature vector \mathbf{D}_i to each supervoxel.

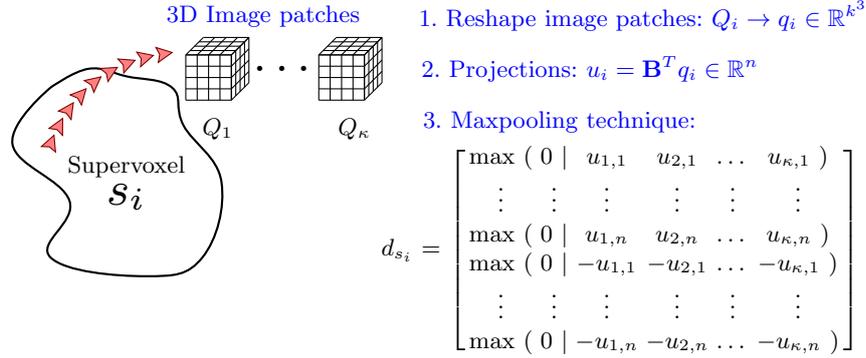


Fig. 2. Features determination for only one supervoxel using one sparse dictionary.

3.2 Segmentation of Gliomas

Figure 3 shows a flowchart to illustrate the automatic brain tumour segmentation method. For each patient a subset of 50 feature vectors of each tissue class are randomly selected in order to obtain a training set. An ERT classifier is trained to automatically assign each supervoxel to its respective class. ERT combines the predictions of several classifiers using random splits to generate different trees and then to calculate an output final classification [12]. Each tree learns a weak predictor for each supervoxel $\theta(D_i)$. The class c which corresponds to supervoxel s_i is calculated as the most frequent estimated class:

$$c = \text{mode}[\theta_1(\mathbf{D}_i), \theta_2(\mathbf{D}_i), \dots, \theta_T(\mathbf{D}_i)] \quad (1)$$

where T is the number of trees. Equation 2 is used to calculate the probability map for the images of each patient. The probability map assigns all supervoxels probability values of belonging to each tissue class c .

$$P(c|\theta(\mathbf{D}_i)) = \frac{1}{T} \sum_{t=1}^T \mathbf{1}_A(\theta_t(\mathbf{D}_i)), \text{ where: } \mathbf{1}_A(\theta_t(\mathbf{D}_i)) = \begin{cases} 1, & \theta_t(\mathbf{D}_i) = c \\ 0, & \theta_t(\mathbf{D}_i) \neq c \end{cases} \quad (2)$$

To complement the segmentation task a Conditional Random Field (CRF) operates over the supervoxels of the images. The purpose of the CRF is to obtain the labelling of the supervoxels such that minimize the following equation:

$$E(\mathbf{X}, \mathbf{D}) = \underbrace{\sum_{i \in S} f_i(\mathbf{X}_i | \mathbf{D}_i)}_{\text{Unary potential}} + \beta \underbrace{\sum_{i \in S} \sum_{j \in N_i} f_{i,j}(\mathbf{X}_i, \mathbf{X}_j | \mathbf{D}_i, \mathbf{D}_j)}_{\text{Pairwise potential}} \quad (3)$$

where S denotes the set of supervoxels, N_i is the set of neighbours of supervoxel i , β controls the relative importance of the smoothing term. The energy function of the conditional random field defines a posterior probability distribution

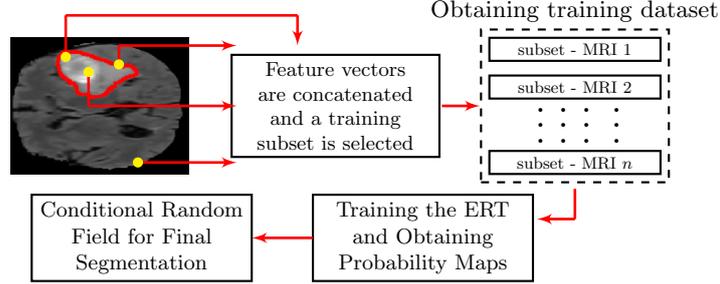


Fig. 3. Automatic Segmentation Method.

$P(\mathbf{X} | \mathbf{D})$ given a set of class label \mathbf{X} and a set of features for each supervoxel \mathbf{D} . The tumour segmentation problem is described as the partitioning of vertices of a graph into disjoint subsets. Graph cuts is performed to partition the graph and obtain the maximum a posteriori inference of the labels \mathbf{X} [13]. Equation 4 presents the unary potential which is defined by the probability map of the supervoxels. Equation 5 presents the pairwise potential which models the relationship among neighbouring of supervoxels.

$$f_i(\mathbf{X}_i | \mathbf{D}_i) = -\log(P(\mathbf{X}_i | \mathbf{D}_i)) \quad (4)$$

$$f_{i,j}(\mathbf{X}_i, \mathbf{X}_j | \mathbf{D}_i, \mathbf{D}_j) = \begin{cases} \exp(-\|\mathbf{D}_i - \mathbf{D}_j\|), & \text{if } \mathbf{X}_i \neq \mathbf{X}_j \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

4 Experiments and Results

We evaluate our method using real patient data obtained from BraTS 2018 challenge. The experiment details are explained as follows. The experimental setup for gliomas segmentation is given as follows: a) the sparse dictionaries are estimated separately over 128 filters from 500000 randomly sampled patches of dimension $3 \times 3 \times 3$. The supervoxels are calculated using an initial dimension of $4 \times 4 \times 4$ and $3 \times 3 \times 3$ to detect the tumour region and multi-label classification of abnormal tissue respectively, c) Number of trees $T = 100$ for ERT model.

Tables 1 and 2 presents the results for training phase and validation phase respectively. The Mean, StdDev, Median, 25q, and 50q indicate the mean, median, standard deviation, 25th percentile and 75th percentile respectively, for the evaluated metrics. The evaluated metrics are Dice, sensitivity, specificity and 95th percentile Hausdorff for enhancing tumour (ET), whole tumour (WT) and tumour core (TC).

Acknowledgements

The first author's research was partially supported by a CONACYT postdoctoral grant and CONACyT Research Grant 256126. The first author would also like to thank the University of Exeter for its hospitality.

Table 1. Results using the training set. Training set is composed by 285 brains.

	Training Phase											
	Dice			Sensitivity			Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Mean	0.5948	0.8398	0.7333	0.7837	0.9185	0.8451	0.9946	0.9835	0.9916	11.69	8.9686	14.6677
StdDev	0.2520	0.0953	0.1499	0.1673	0.0998	0.1671	0.0044	0.0132	0.0067	12.7416	9.0691	11.7043
Median	0.6751	0.8659	0.7730	0.8284	0.9473	0.9003	0.9958	0.9866	0.9933	5.6568	5.3851	11.3578
25q	0.5213	0.8112	0.6740	0.7309	0.8932	0.8057	0.9925	0.9781	0.9893	3	3.3166	6.4031
50q	0.7665	0.898	0.8425	0.8897	0.9748	0.9613	0.9981	0.9925	0.9963	16.9925	11.0453	20.0199

Table 2. Results using the validation dataset. Validation set is composed by 66 brains.

	Validation Phase											
	Dice			Sensitivity			Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Mean	0.5719	0.7992	0.6285	0.6836	0.8714	0.7339	0.9955	0.9863	0.9919	12.9637	12.4414	17.7162
StdDev	0.2749	0.1557	0.2590	0.2782	0.1711	0.3083	0.0038	0.0094	0.0078	15.7325	19.2750	17.5612
Median	0.6910	0.8458	0.7103	0.7776	0.9412	0.8697	0.9960	0.9883	0.9936	6.5556	6.4031	15.106
25q	0.4934	0.7862	0.5365	0.5661	0.8435	0.6204	0.9942	0.9821	0.9903	3.2008	3.6395	9
50q	0.7669	0.8826	0.8196	0.8947	0.9695	0.9615	0.9980	0.9927	0.9963	17.9093	12.4078	20.7175

5 Conclusions

We present an approach for multimodal brain tumour segmentation using an efficient feature encoding technique based on sparse dictionaries and supervoxels. BraTS 2018 database is adopted to evaluate the performance of our method. The method combines the advantages of the supervoxels and sparse coding techniques to generate feature vectors which can be employed to assign each tissue to its respective class using Extremely Randomized Trees and Conditional Random Field algorithms. According to the numerical results the performance of our approach can be compared to the performance of other state-of-the-art algorithms which have been evaluated using the same dataset.

References

1. Soltaninejad, M., Yang, G., Lambrou, T., Allinson, N., Jones, T.L., Barrick, T.R., Howe, F.A., Ye, X.: Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in flair mri. *International Journal of Computer Assisted Radiology and Surgery* **12**(2) (Feb 2017) 183–203
2. Havaei, M., Larochele, H., Poulin, P., Jodoin, P.M.: Within-brain classification for brain tumor segmentation. *International journal of computer assisted radiology and surgery* **11**(5) (2016) 777–788
3. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, ., Durst, C.R., Dojat,

- M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging* **34**(10) 1993–2024 (Oct 2015) DOI: 10.1109/TMI.2014.2377694
4. Zimeras, S., Gortzis, L., Pylarinou, C. In: *Shape Analysis in Radiotherapy and Tumor Surgical Planning Using Segmentation Techniques*. Springer International Publishing, Cham (2014) 159–173
 5. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
 6. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive* (2017) DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
 7. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *The Cancer Imaging Archive* (2017) DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
 8. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: *Advances in neural information processing systems*. (2007) 801–808
 9. Stutz, D., Hermans, A., Leibe, B.: Superpixels: An evaluation of the state-of-the-art. *Computer Vision and Image Understanding* (2017)
 10. Veksler, O., Boykov, Y., Mehrani, P. In: *Superpixels and Supervoxels in an Energy Optimization Framework*. Springer Berlin Heidelberg, Berlin, Heidelberg (2010) 211–224
 11. Van den Bergh, M., Boix, X., Roig, G., Van Gool, L.: Seeds: Superpixels extracted via energy-driven sampling. *International Journal of Computer Vision* **111**(3) (Feb 2015) 298–314
 12. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. *Machine Learning* **63**(1) (Apr 2006) 3–42
 13. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE transactions on pattern analysis and machine intelligence* **26**(9) (2004) 1124–1137

Brain Tumor Segmentation using 2D U-net

Hyung Eun Shin* Moo Sung Park*

* Deepnoid, 1305, E&C Venture Dream Tower 2, 55, Digital-ro 33-gil, Guro-gu, Seoul, South Korea

hshin255@deepnoid.com, mspark@deepnoid.com

Abstract. A brain tumor is one of the most challenged segmentation work in neural network. To segment a brain tumor affectively, 2-D U-net is used for a base architecture. A brain tumor is described in four multimodality MRI scans, i.e. native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), T2 Fluid Attenuated Inversion Recovery (FLAIR). These 3D multimodal scans are converted to 2D images for training. The label consists of background (BG – label 1), GD-enhancing tumor (ET – label 4), the peritumoral edema (ED – label 2) and the necrotic and non-enhancing tumor core (NCR/NET – label 1), and it's converted to following three sub-regions: whole tumor, tumor core, enhancing tumor. A training data is 5-fold cross validated generating 168 HGG training data and 42 HGG validation data for each split. While training, relatively large batch size is applied by virtue of lighter 2D models. Once segmentation is over, survival prediction is performed. A survival dates are classified to following three survivor classes: short-survivors (<10 months), mid-survivors (≥ 10 months & < 15 months), long-survivors (≥ 15 months). Only resection status of GTR is used for prediction, and total of 59 subjects are 5-fold cross validated. Volumetric features and radiomic features are extracted for prediction. Total of 1412 features are screened via recursive feature elimination, variable importance and forward selection. An elastic net is used to predict survivor class using selected variables.

Keywords: 2D U-net, Resnet, Elastic net,

1 Introduction

A brain tumor is a mass of unnecessary cells growing in the brain or central spine canal. Brain tumors are graded into four, and this grade indicates a degree of malignancy. Grade I and II are called as a lower grade glioma (LGG), and Grade III and IV are called as a higher grade glioma (HGG). Well-known glioblastoma (GBM) is the most common example of grade IV tumor, and it is considered as one of the most lethal cancer types in adults. LGG is less malignant and grow slower than HGG. Tumors contains several grades of cells, so the grade of tumors are determined by highest grade of cells regardless of ratios among grades of cells. A glioma has various heterogeneous histological sub-regions such as peritumoral edema, enhancing tumore, necrotic core

and necrotic tumor. They are portrayed across multimodal MRI scans with varying intensity and tumor properties. According to the American Brain Tumor Association (ABTA), nearly 50% of GBM patients die within a year of diagnosis. A median survival is about 14.6 month with a two-year median survival rate of 27%. Furthermore, five-year survival rate is only 10%.

With multimodal MRI scans and labeled sub-regions, segmentation of brain tumors can be performed. Labels in segmentation is transformed to enhancing tumor, tumor core and whole tumor and it will be discussed in detail later. Instead of using 3-D data itself, 2-D axial scans are used to segment tumors in this paper because 2D layers are lighter than 3D enabling large batch size. The model is trained via 2-D U-nets with residual block, dense block, mobile-net, transposed convolution and resize convolution. Once label is created, image features are extracted from sub-regions and multimodal MRI scans. These features are used to predict survival class of patients, i.e. short survivor (<10 months), mid survivor (≥ 10 months & < 15 months) and long survivor (≥ 15 months).

2 Methods

2.1 Data

The data sets used for this work are obtained from BraTS 2018 Challenge. The data consists of ample multi-institutional routine clinically-acquired pre-operative multimodal MRI scans of glioblastoma (HGG) and lower grade glioma (LGG). For each patient, four MRI modality is provided which are native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), T2 Fluid Attenuated Inversion Recovery (FLAIR). Each modality has a volume of (240, 240, 155). In addition to these multimodal scans, ground truth labels which have been manually segmented by experienced neuro-radiologists are provided. Annotations include the GD-enhancing tumor (ET – label 4), the peritumoral edema (ED – label 2) and the necrotic and non-enhancing tumor core (NCR/NET – label 1). Anything else except these labels 1 2 4, is labeled as 0. The training set consists of 210 HGG patients and 75 LGG patients. In this paper, 5-fold cross-validation is used to avoid bias. Each time, 168 HGG patients and 60 LGG patients are trained and left 42 HGG patients and 15 LGG patients are used for validation. Along with MRI scans and segmentations, `survival_data.csv` file is provided which contains information of patient ID, Age, resection status, and survival dates. Here, only subjects with resection status of GTR is used for survival prediction. The validation set provided contains 66 subjects but it doesn't specify grade of glioma. Unlike the training set, the validation set doesn't involve segmentation labels and survival dates in provided csv file.

2.2 Pre-processing

A provided data has a volume of (240,240,150) for each nifty file, but it contains an excessive amount of uninformative parts. These unnecessary parts dramatically increased true negative. This relatively large proportion of background affected not only

background but also brain tumor to be recognized as a background. These distortions generated sensitivity close to 0, specificity close to 1, and accuracy close to 1. To solve this problem, every single volume is consistently cropped to (160,192,150) instead of cropping and resizing it according to the brain size per patient to get location information. The amount of dropped background is decided by checking bounding boxes of brains. Each brain would have average margin of five to fifteen. Every volume is sliced to 150 of 2-D axial images, so 600 scans size of (160, 192) are considered for one patient. A label volume is also converted to (160, 192, 150), and sliced into axial images.

For the task of survival prediction, A numerical variable ‘Survival’ is converted to a categorical variable ‘survival class’. Patients having survival data less than 300 days (< 10 months) are classified to ‘short survivor’, greater or equal to 300 (\geq 10 months) and 450 days (< 15 months) are classified to ‘mid survivor’ and greater or equal to 450 days (\geq 15 months) are classified to ‘long survivor’. This newly created variable is used as a response variable.

2.3 Segmentation

For brain tumor segmentation, 2D U-Net [1] is used as a base architecture. U-net is composed with a contracting path and an expanding path. Feature extraction and down-sampling is performed at the contracting path, and decoding is performed at the expanding path. Other convolutional substitutes plain convolution-batch normalization – ReLu layer to get better performance.

To train 2D U-net, four modalities with shape of (160, 192, 150) for individual modality are transformed into (Batch size, 192, 160, 4) shape, and these are utilized as an input. To fully take an advantage of batch normalization, mini-batch sizes of 30, 50, 75 are applied. Relatively large batch sizes could be applied by virtue of lighter 2D model than 3D model. After some experiments, it turns out that larger mini-batch size brings significantly better result. A shape of a final input data is set as (75, 192, 160, 4). This part will be discussed in detail later.

A residual-block from Resnet [2], dense-block from Dense-Net [3], which are well known for high efficient feature extraction, and depth-wise-separable convolutional layer from Mobile-Net which is lighter than other convolutional layers [4] are considered for convolutional layer at the contracting path. Continuous convolutional layers are applied before down-sampling. The last convolutional layer is followed by down-sampling layer which reduces the size of feature map. Instead of pooling, Nearest-neighbor algorithm is used for down-sampling operation to preserve features effectively.

After four times of down-sampling, four groups of feature maps from a contracting path are concatenated with corresponding groups of feature maps from an expanding path by skip-connections. Each group of feature maps is up-sampled to be concatenated with a group of larger feature maps. Transposed convolutions [5] and resize convolutions [6] are used for up-sampling. Transposed convolutions are optimized for catching a boundary of object, nevertheless it’s less competent at restoring images in details well

resulting in checkerboard pattern. On the other hand, Re-size convolutions restore details more elaborate than transposed convolutions, however it's challenged at catching a boundary. Instead of making decision between two convolutions, both are used to get advantages from two different methods.

As four labels (0, 1, 2, 4) are given, model outputs consist of four classes (Background, NCT/NET, ED, ET). Thus, output data has a same shape with input data as (75, 192, 160, 4). When calculating losses, background loss plummets making it hard to train stably. To train successfully, relatively small weights are applied while pretraining and background loss are ruled out while fine tuning.

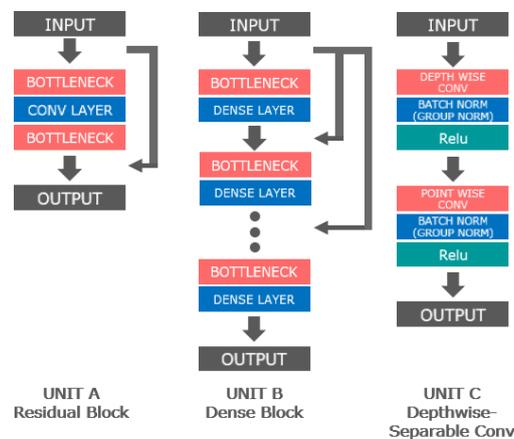


Fig. 1. Three convolutional units used in 2D U-Net contracting path: Unit A is a residual-block which has two bottleneck layer and one convolutional layer with batch normalization, ReLu. Unit B is a dense block consist of dense layer which has two convolutional layers followed by batch normalization, ReLu, dropout. Unit C is a depth-wise separable convolutional layer which has depth-wise and point-wise convolutional layers.

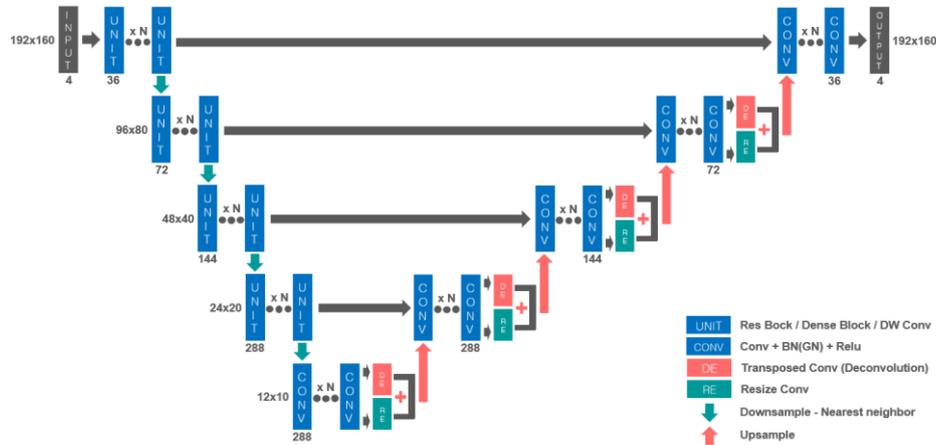


Fig. 2. Modified 2D U-Net used for segmentation:

When training, 5-fold-cross-validation is used to avoid bias. Each time, 168 HGG patients and 60 LGG patients are trained and left 42 HGG patients and 15 LGG patients are used for validation. After extracting label, label 0, i.e. background, are padded back into image to make original size.

2.4 Prediction

For survival prediction, only patients with resection status of GTR (Gross Total Resection) are considered. In the training set, there are total of 59 subjects, and it's 5-fold cross-validated. For validation set, only 28 patients are marked as a GTR. To predict survival classification of patients, 1412 features are considered. These features include volumetric features such as volume, centroid, ratios between sub-regions from four modalities and three sub-regions (ET, TC, WT). In addition to that, radiomic features are also considered, which includes co-occurrence matrix, size zone matrix, run length matrix, neighboring gray tone difference matrix, dependence matrix. To be specific, 23 features are extracted from gray level co-occurrence matrix (GLCM), 16 features are extracted from run length matrix (GLRLM) for each modality and sub-regions.

Among total of 1412 features, 300 features are selected through recursive feature elimination. These 300 selected features are used to predict survivor class using gradient boost. Fifty variables having greatest variable importance are selected, and selected ones are again screened by forward selection with AIC.

To predict survivor class, random forest, elastic net, ridge, lasso and gradient boosting methods are considered. Among these methods, an elastic net works best on predicting survivor class, therefore elastic net is finally selected as a predicting model.

3 Result

3.1 Segmentation

Table1 shows the results of differently modified 2D U-Net. A significant difference is made by mini-batch size. As a batch size grows larger, result gets better. It means that even when training 3D shaped data, 2D segmentation model is significant in terms of large batch size. Moreover, at the point that a specificity which is relative to true negative is high, and a sensitivity which is relative to true positive is low, importance of balancing non-label and label loss could be emphasized while training segmentation with larger non-label area. This points out that elaborate tuning is needed not only for feature extraction but also for loss regardless of which convolution unit is chosen among residual block, dense block, depth-wise separable convolution.

Table 1. Segmentation results on Brats2018 training data set

	DSC			Sensitivity			Specificity			Hausdorff		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Unit C (Batch 75)	61.5	76.8	70.0	62.0	74.1	69.2	99.9	99.8	99.9	15.1	22.0	19.0
Unit B (Batch 50)	59.0	72.5	66.6	65.6	85.2	73.2	99.8	99.3	99.8	17.0	29.1	23.0
Unit C (Batch 50)	45.0	52.1	65.4	47.0	76.6	56.2	99.9	99.4	99.8	13.6	28.7	24.2
Unit B (Batch 50)	52.9	59.6	70.2	60.5	65.2	82.4	99.8	99.4	99.8	16.2	22.4	28.1
Unit A (Batch 30)	51.6	70.6	60.0	56.7	75.5	65.6	99.8	99.5	99.8	15.7	25.1	20.5
Unit C (Batch 15)	26.9	38.9	37.0	39.2	40.0	39.3	99.9	99.8	99.8	4.8	9.3	7.3
Unit A (Batch 50)	55.1	49.0	54.6	58.9	60.1	68.0	99.9	99.1	99.2	8.6	30.4	27.6

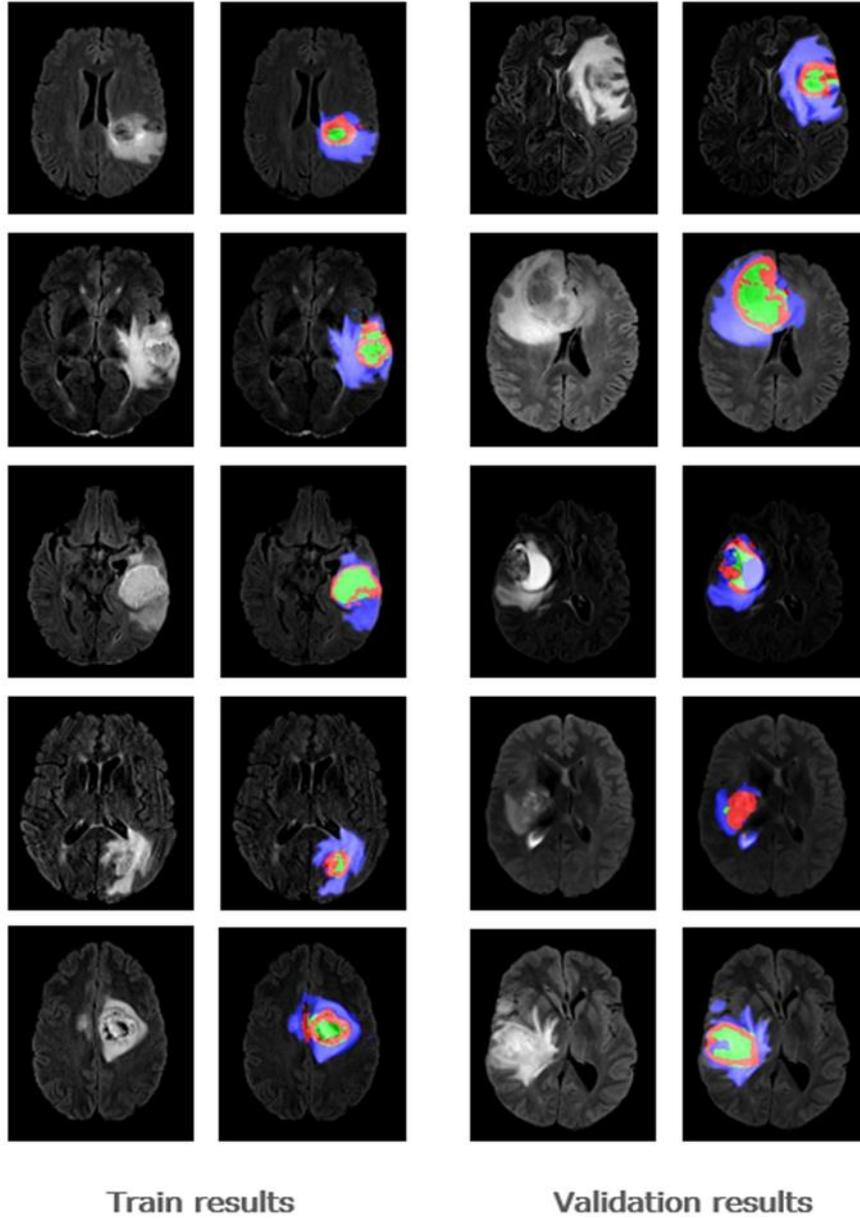


Fig. 3. Original and masked Images of Train and validation results

3.2 Survival prediction result

For survival prediction, an elastic net, random forest and gradient boost etc. are used and an elastic net yields best results. A mean training accuracy was 0.6748 (high as 0.7238 and low as 0.5712) with lambda 0.00367578 and alpha 0.1, however the best tune changes every time if seed are not set so that lambda and alpha is not guaranteed as all-time best tune.

4 Discussion

For segmentation, residual block, dense block, depth-wise separable convolutions are tried to extract more precise feature. And two up-sampling methods, transposed convolutions and resize convolutions are both used and merged for enhanced up-sampling. It turns out that modified U-Net with these functions gives better results. But most of all, mini-batch size contributes to significantly improve results. It means that 2D model is competitive when learning 3D data. However, when TN is much more than TP, sophisticated design of loss functions is very important for successful learning.

For survival prediction, KNN, random forest, naïve bayes, gradient boost, ridge, lasso, and elastic net are tried. As cross-validation is not fixed, the results are changed every time. If variables are added or dropped for classifier, an accuracy is apt to change. Sometimes other methods out-performed elastic net, however elastic net worked better than other methods in overall.

5 Conclusion

In this paper, a brain tumor segmentation and a survival prediction are performed. First for the segmentation, five-fold cross validation is applied generating 168 training data and 42 validation data for each fold. For the base architecture, 2-D U-net is applied instead of 3D models. To run a 2D model, 3D volumes are converted into 2D axial slices with shape of (160,192,150). Instead of using convolutional layers from formal U-net, various types of convolutional layers are used. For instance, residual block, dense block, and depth-wise separable layers are applied to while down sampling. For expanding path, both transposed convolutions and resize convolutions are used. In general, transposed convolutions are efficient at catching a boundary and resize convolutions are competent at restoring details. Applying both convolutions, advantages from two convolutions are maximized. Relatively large batch size could be applied on account of light 2D inputs. Four provided labels were background, enhancing tumor, necrotic/non-enhancing tumor, and edema, however it's transformed into background, enhancing tumor, tumor core, and non-enhancing tumor for survival prediction and evaluation. Results of whole tumor are relatively higher than tumor-core and enhancing tumor as expected. Secondly for survival prediction, survivor dates are classified into three survivor classes, which are short survivors (< 10 months), mid survivors (≥ 10 months & < 15 months), and long survivors (≥ 15 months). Four modality MRI scans

and three labels are used to extract features to predict survivor class. Features consists of volumetric features such as centroid, volume, length, and width etc. and radiomic features such as gray level co-occurrence matrix, gray level size zone matrix and neighboring gray tone difference matrix etc. Total of 1412 features are extracted, and it's screened by recursive feature elimination, variance importance, and forward selection. Mainly, an elastic net is utilized to predict survivor class using selected features.

References

1. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, 2015
2. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
3. Huang G, Liu Z, Weinberger KQ, van der Maaten L. "Densely connected convolutional networks." arXiv preprint arXiv:1608.06993. 2016 Aug 25.
4. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861, 2017.
5. Zeiler, M. D., Taylor, G. W., and Fergus, R. (2011). Adaptive deconvolutional networks for mid and high level feature learning. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2018–2025. IEEE.
6. Odena, Augustus, Vincent Dumoulin, and Chris Olah. "Deconvolution and checkerboard artifacts." *Distill* 1, no. 10 (2016): e3
7. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp I, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
8. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
9. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
10. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of

the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI:
10.7937/K9/TCIA.2017.GJQ7R0EF

Leveraging a DenseNet Encoder Pre-Trained on ImageNet for Brain Tumor Segmentation.

Jean Stawiaski¹

Stryker Corporation, Navigation. Freiburg im Breisgau, Germany.
jean.stawiaski@stryker.com

Abstract. This article presents a convolutional neural network for the automatic segmentation of brain tumors in multi-modality 3D MR images based on a U-net architecture. We evaluate the use of a densely connected convolutional network encoder (DenseNet) which was pre-trained on the ImageNet data set. We detail the network architecture and propose several means to take into account multi modality 3D images as inputs. This approach aims to identify if generic pre-trained networks can be used for very specific medical applications where the targets data differs both in the number of spatial dimensions as well as in the number of inputs channels. We push the limit of transfer learning by only training the decoder part of the U-net architecture. We evaluate the effectiveness of the proposed approach on the BRATS 2018 segmentation challenge [1–4] where we obtained preliminary dice scores of 0.78, 0.88, 0.85 and 95% Hausdorff distance of 2.9mm, 4.3mm, and 6.6mm for enhanced tumor core, whole tumor and tumor core respectively on the validation set.

Keywords: Brain tumor, Convolutional neural network, Densely connected network, Image segmentation..

1 Introduction

Automatic segmentation of brain tumor structures is particularly important in order to quantitatively assess the tumor geometry. It has also a great potential for surgical planning and intraoperative surgical resection guidance. Automatic segmentation still poses many challenges because of the variability of appearances and sizes of the tumors. Moreover the differences in the image acquisition protocols, the inhomogeneity of the magnetic field and partial volume effects have also a great impact on the image quality obtained from routinely acquired 3D MR images. However brain gliomas can be well detected using modern magnetic resonance imaging. The whole tumor is particularly visible in T2-FLAIR, the tumor core is visible in T2 and the enhancing tumor structures as well as the necrotic parts can be visualized using contrast enhanced T1 scans. An example is illustrated in figure 1.

In the recent years, deep neural networks have shown to provide state-of-the-art performance for various challenging image segmentation and classification

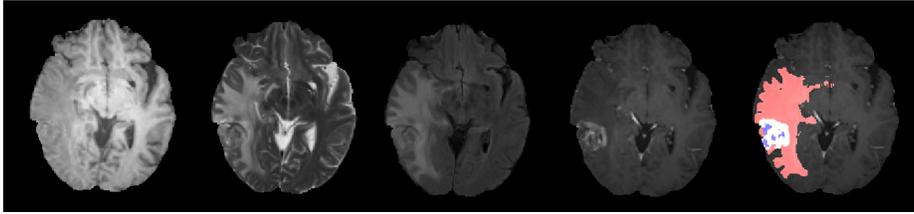


Fig. 1: Example of images from the BRATS 2018 dataset. From left to right: T1 image, T2 image: the whole tumor and its core are visible, T2 FLAIR image: discarding the cerebrospinal fluid signal from the T2 image highlights the tumor region only, T1ce: contrast injection permits to visualize the enhancing part of the tumor as well as the necrotic part. Finally the expected segmentation result is overlaid on the T1ce image. The edema is shown in red, the enhancing part in white and the necrotic part of the tumor is shown in blue.

problems [5, 6, 8, 9, 7]. Medical image segmentation problems have also been successfully tackled by such approaches [10, 11, 13, 14, 18]. Moreover it has been shown that initializing weights of a convolutional network that has been pre-trained on a large data set improves its accuracy on specific tasks where a limited number of training data is available [15]. We evaluate in this work the accuracy of a U-net architecture [10, 11] where the encoder is a densely connected convolutional network [16] which has been pre-trained on the ImageNet data set [17]. We study an extreme case of transfer learning where we fix the weights of the pre-trained DenseNet encoder. Moreover we consider a segmentation problem where the input data dimensionality does not match the native input dimensions of the pre-trained network. More precisely we will make use of a fixed pre-trained convolutional network trained on 2D color images in order to segment 3D multi channel medical images.

2 Methods

This section details the proposed network architecture, the loss function used to train the network as well as the training data preparation.

2.1 Convolutional Neural Network Architectures

The network processes multiple 2D images of size $(224, 224)$ pixels containing three channels. An input image is composed of three equidistant slices of the input volume along one of the three anatomical orientations: either along the coronal, the sagittal or the transverse plane. The considered pre-trained network has been designed to take a single three channels 2D image as input. In order to be able to process multi modal input data, we have tested two distinct encoders:

- the first solution (M1) consists in removing the stem of the original DenseNet and only make use of the following convolutional layers which input is a

tensor of size (64,112,112). This architecture is illustrated in figure 2. The proposed network is thus composed of a so called "precoder" which produces an adequate high dimensional input tensor for the pre-trained network. The precoder architecture is illustrated in figure 3. It processes independently each input images and concatenates the resulting tensors. This approach is very flexible and could take as input an image of any dimensions. For instance it would be possible to increase the number of equidistant slices (i.e. the number of channels) of the input images.

- the second solution (M2) consists in evaluating the input images separately through the original DenseNet encoder. Each input image modality is processed with the same encoder which shares its weights across the different modalities. Outputs at different scales are then concatenated and fed to the decoder. This architecture is illustrated in figure 5 and 6. This architecture does not permit to vary the number of input slices but has the advantage to fully leverage the original DenseNet weights.

For both architectures, the decoder consists in upsampling a low resolution layer, concatenate it with a higher resolution layer before applying a sequence of convolutional layers. The first convolutional layer reduces the number of input channels by applying a (1x1) convolution. Later layers are composed of spatial (3x3) convolutions and residual connections.

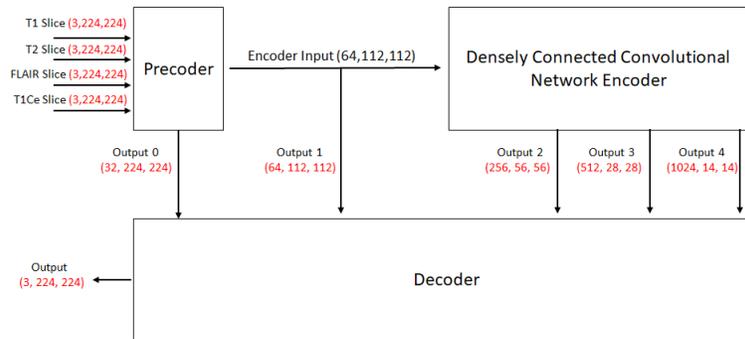


Fig. 2: Network architecture (M1). The network is composed of a "precoder" producing a high order tensor which is fed to a pre-trained densely connected convolutional network. Several intermediate layers are then used so reconstruct a high resolution segmentation map. Images dimensions are highlighted in red.

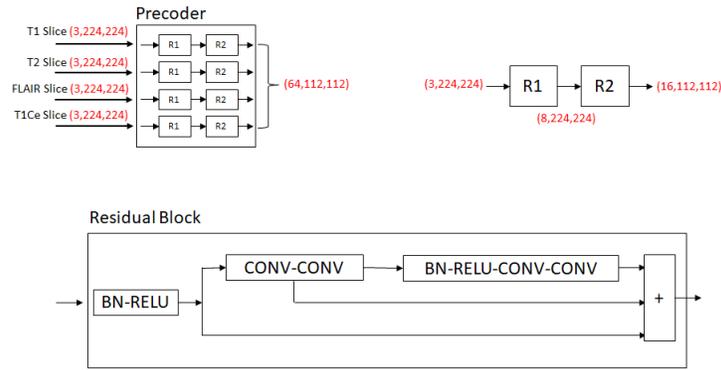


Fig. 3: Precoder architecture (M1). The precoder architecture process independently the input images by a sequence of multiple residual blocks (R1, R2) and concatenate the resulting output tensors. A residual block (R) is also illustrated. All convolution operations are computed with (3x3) kernels. Images dimensions are highlighted in red.

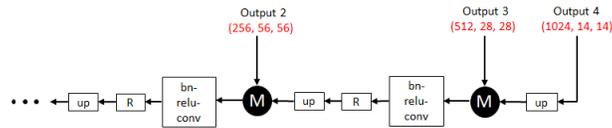


Fig. 4: Decoder architecture (M1). The decoder architecture follows the principle of the U-net architecture. It upsamples low resolution layers, concatenate them with a higher resolution layer before producing new features using a sequence of multiple residual blocks. Images dimensions are highlighted in red.

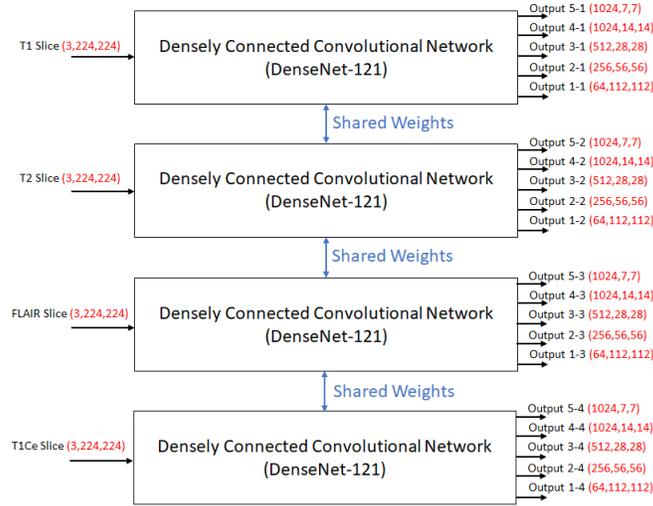


Fig. 5: Encoder architecture (M2). The network processes the different input image modality with the same encoder, a DenseNet composed of 121 layers. Intermediate layers of the encoder are used to feed the decoder network. Images dimensions are highlighted in red.

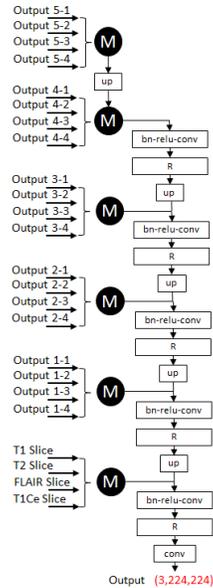


Fig. 6: Decoder architecture (M2). The decoder network processes the different feature maps produced by an the DenseNet encoder. Features are computed independently for all input image modalities. Images dimensions are highlighted in red.

We give here additional details about the network architectures:

- each sample image y is normalised so that voxels values falls in the interval

$$0, 1$$

- batch normalisation is performed after each convolutional layer using a running mean and standard deviation computed on 5000 samples:

$$by = \frac{(y - m_b)}{(\sigma_b + \epsilon)} \times \gamma + c, \quad (1)$$

where m_b and σ_b is the mean and variance of the minibatches and γ and c are learnable parameters,

- each layer is composed of residual connections as illustrated in figure 4,
- the activation function used in the network is a rectified linear unit,
- convolution kernels are (3x3) kernels,
- convolutions are computed using reflective border padding,
- upsampling is performed by nearest neighbor interpolation.

2.2 Training

We used the BRATS 2018 training and validation sets for our experiments [1–4]. The training set contains 285 patients (210 high grade gliomas and 75 low grade gliomas). The BRATS 2018 validation set contains 66 patients with brain tumors of unknown grade with unknown ground truth segmentations. Each patient contains four modalities: T1, T1 with contrast enhancement, T2 and T2 FLAIR. The aim of this experiment is to segment automatically the whole tumor, the tumor core and the tumor enhancing parts. Note that the outputs of our neural network corresponds directly to the probability that a pixel belongs to a tumor, the core of a tumor and the enhancing part of the tumor. The last layer of the proposed architecture is thus composed of three independent (1x1) convolutional layers because we directly model the problem as a multi-label segmentation problem where a pixel can be assigned to multiple classes. Note that only weights of the "precoder" and the decoder are learned. Original weights of the pretrained DenseNet are fixed.

The network produces a segmentation maps by minimizing a loss function defined as the combination of the mean cross entropy (mce) and the mean Dice coefficients (dce) between the ground truth class probabilities and the network estimates:

$$ce = \sum_k \left(\frac{-1}{n} \sum_i y_i^k \log(p_i^k) \right) \quad (2)$$

where y_i^k and p_i^k represent respectively the ground truth probability and the network estimate for the class k at location i .

$$dce = \sum_{k \neq 0} \left(1.0 - \frac{1}{n} \left(\frac{2 \sum_i p_i^k y_i^k}{\sum_i (p_i^k) + \sum_i (y_i^k)} \right) \right). \quad (3)$$

Note that we exclude the background class for the computation of the dice coefficient. The network is implemented using Microsoft CNTK ¹. We use stochastic gradient descent with momentum to train the network. We use a decaying cyclic learning rate where the learning rate varies from 0.0002 to 0.00005. An example of the evolution of the accuracy and the learning rate is illustrated in figure 7. We train the network for 160 epochs. A full epoch consists in analyzing all images of the BRATS training data set and extract 20 2D random samples from the 3D MR volumes.

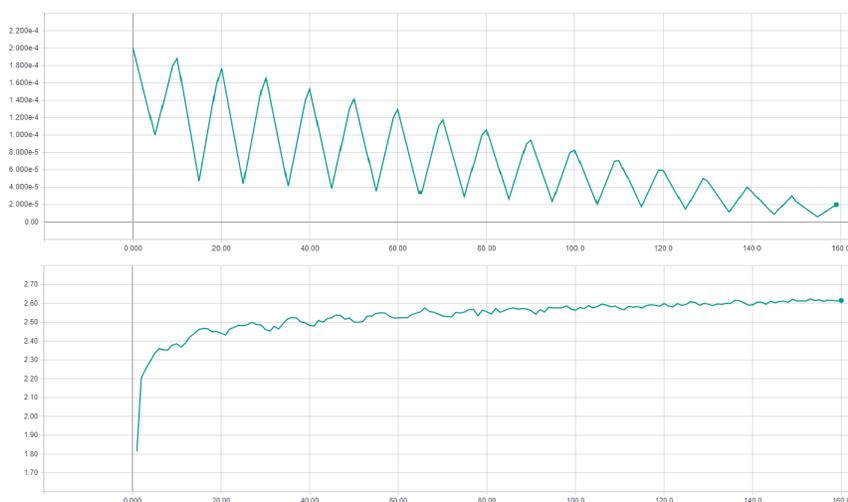


Fig. 7: Network training. Illustration of the cyclic learning rate schedule (top). Evolution of the sum of the dice coefficients of the three classes during training (bottom).

2.3 Testing

Segmentation results are obtained by evaluating the network along slices extracted from the three anatomical orientations and averaging the results. A segmentation map is then obtained by assigning to each voxel the label having the maximum probability among the three classes: tumor, tumor core or enhancing tumor. Finally connected components composed of less than 100 voxels are removed.

¹ <https://www.microsoft.com/en-us/cognitive-toolkit/>

3 Results

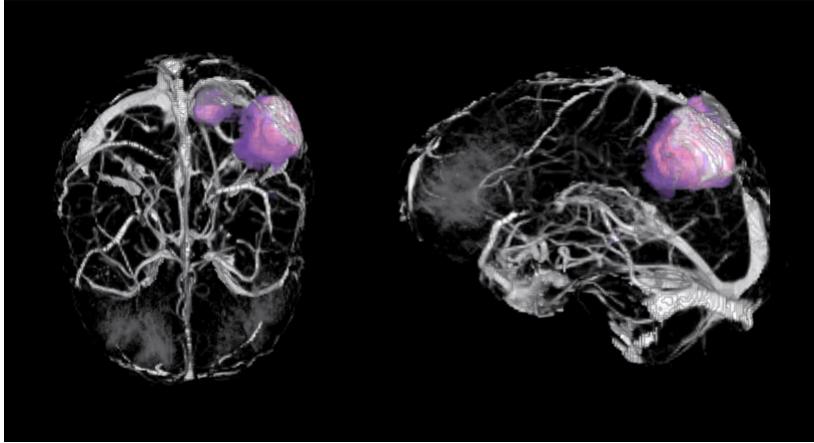


Fig. 8: Segmentation result obtained on an image of the training data.

We uploaded our segmentation results to the BRATS 2018 server ² which evaluates the segmentation and provides quantitative measurements in terms of Dice scores, sensitivity, specificity and Hausdorff distances of enhanced tumor core, whole tumor, and tumor core. Results of the BRATS 2018 validation phase are presented in Table 1.

	Dice ET	Dice WT	Dice TC	Dist. ET	Dist. WT	Dist. TC
Mean M1	0.768	0.892	0.815	3.85	4.85	7.56
Mean M2	0.778	0.881	0.845	2.95	4.37	6.62
StdDev M1	0.241	0.065	0.187	5.43	4.28	12.56
StdDev M2	0.223	0.104	0.130	3.35	3.27	12.21
Median M1	0.849	0.905	0.889	2.23	3.67	3.74
Median M2	0.845	0.909	0.887	2	3.60	3.60
25 quantile M1	0.792	0.881	0.758	1.68	2.23	2
25 quantile M2	0.785	0.877	0.787	1.41	2.23	2
75 quantile M1	0.888	0.933	0.930	3.16	5.65	8.71
75 quantile M2	0.888	0.932	0.930	2.82	5.09	5.83

Table 1: BRATS 2017 Validation scores, dice coefficients and the 95% Hausdorff distances. Our results corresponds to the team name "stryker". (M1) results corresponds to the precoder approach, (M2) corresponds to the direct use of a fixed pretrained DenseNet-121.

² <https://www.cbica.upenn.edu/BraTS18/lboardValidation.html>

4 Discussion

The validation results obtained on the BRATS segmentation challenge show that the proposed approaches are indeed efficient. Despite the fact that the used encoder has been trained on natural color images, it turns out that the learned features can be leveraged for a large class of applications including segmentation of images having different dimensionality and content. The two approaches produce comparable results and have both advantages and drawbacks. The model (M1) is more versatile since it can use any number of input modalities (channels) and any number of spatial dimensions. However current experiments shows that the model (M2), despite its simplicity, produces slightly better results.

5 Conclusion

We have studied an extreme version of transfer learning by using a fixed pre-trained network trained on 2D color images for segmenting 3D multi modal medical images. We have presented two simple approaches for leveraging pre-trained networks in order to perform automatic brain tumor segmentation. We obtained competitive scores on the BRATS 2018 segmentation challenge ³. Future work will concentrate on several possible improvements by additionally fine tuning the pre-trained encoder and making use of extensive image augmentation.

References

1. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp , Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharruddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor T.J, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015)
2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, 2017.
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017.

³ <https://www.cbica.upenn.edu/BraTS17/lboardValidation.html>

4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017.
5. J. Long, E. Shelhamer, and T. Darrell, Fully convolutional networks for semantic segmentation, arXiv:1605.06211, <https://arxiv.org/abs/1605.06211>, 2016.
6. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L.Yuille, Semantic image segmentation with deep convolutional nets and fully connected CRF, arXiv:1412.7062, <https://arxiv.org/abs/1412.7062>, 2014.
7. Hyeonwoo Noh, Seunghoon Hong, Bohyung Han, Learning Deconvolution Network for Semantic Segmentation, arXiv:1505.04366, <https://arxiv.org/abs/1505.04366>, 2015.
8. Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla, SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, arXiv:1511.00561, <https://arxiv.org/abs/1511.00561>, 2015.
9. Fisher Yu, Vladlen Koltun, Multi-Scale Context Aggregation by Dilated Convolutions, arXiv:1511.07122, <https://arxiv.org/abs/1511.07122>, 2016.
10. Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv:1505.04597, <https://arxiv.org/abs/1505.04597>, 2015.
11. Fausto Milletari, Nassir Navab, Seyed-Ahmad Ahmadi, V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, arXiv:1606.04797, <https://arxiv.org/abs/1606.04797>, 2016.
12. Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, Zhuowen Tu, Deeply-Supervised Nets. arXiv:1409.5185, <https://arxiv.org/abs/1409.5185>, 2014.
13. Qikui Zhu, Bo Du, Baris Turkbey, Peter L . Choyke, Pingkun Yan, Deeply-Supervised CNN for Prostate Segmentation. arXiv:1703.07523, <https://arxiv.org/abs/1703.07523>, 2017.
14. Lequan Yu, Xin Yang, Hao Chen, Jing Qin, Pheng-Ann Heng, Volumetric ConvNets with Mixed Residual Connections for Automated Prostate Segmentation from 3D MR Images. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17). 2017.
15. V. Iglovikov, A. Shvets, TerausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation, <https://arxiv.org/abs/1801.05746>, 2018.
16. G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger Densely Connected Convolutional Networks, <https://arxiv.org/abs/1608.06993>, 2017.
17. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg and Li Fei-Fei, ImageNet Large Scale Visual Recognition Challenge, <https://arxiv.org/abs/arXiv:1409.0575>, 2014.
18. Christ P.F., Elshaer M.E.A., Ettliger F., Tatavarty S., Bickel M., Bilic P., Remper M., Armbruster M., Hofmann F., Anastasi M.D., Sommer W.H., Ahmadi S.a., Menze B.H., Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields. arXiv:1610.02177, <https://arxiv.org/abs/1610.02177>, 2016.

Multi-view 3D CNN with Dense CRF for Brain Tumor Segmentation and Survival Prediction

Li Sun and Songtao Zhang

Southern University of Science and Technology, Shenzhen, China, 518055
11510389@mail.sustc.edu.cn

Abstract. Every year, about 238,000 patients are diagnosed with brain tumors in the world. Accurate and robust segmentation of brain tumor can provide quantitative characterization of tumor sub-region, thus it is important for diagnosis, treatment planning and outcome prediction. While manual segmentation is time consuming and error-prone, we present an automatic framework for brain tumor segmentation and survival prediction here. For segmentation, we propose a novel convolutional neural network (CNN) architecture that incorporates multi-view fusion, multi-scale prediction as well as skip connection. In addition, we refine the result using dense Conditional Random Fields as post-processing step. For survival prediction, we extract 3394 radiomic features from segmented brain tumor sub-regions and perform LASSO for feature selection, then we utilize Ridge regression to predict overall survival. Our automatic framework for brain tumor segmentation achieves average Dice scores of 0.75081, 0.86494, 0.71968 for enhancing tumor core, whole tumor and tumor core, respectively on BraTS validation set. It also achieves competitive result for survival prediction.

Keywords: Brain tumor segmentation · Survival prediction · 3D CNN · Deep learning.

1 Introduction

Brain tumor is cancerous or noncancerous mass or growth of abnormal cells in the brain, malignant brain tumor is one of the most aggressive and fatal tumors. Originated in the glial cells, gliomas are the most common brain tumors. [5] Depending on the pathologic evaluation of the tumor, gliomas can be categorized into glioblastoma (GBM/HGG) and lower grade glioma (LGG). Gliomas contain various heterogeneous histological sub-regions, including peritumoral edema, necrotic core, enhancing and non-enhancing tumor core. Magnetic resonance imaging (MRI) is commonly used in radiology to portray the phenotype and intrinsic heterogeneity of gliomas, since multimodal MRI scans, such as T1-weighted, contrast enhanced T1-weighted (T1c), T2-weighted and Fluid Attenuation Inversion Recovery (FLAIR) images, provide complementary profiles for different sub-regions of gliomas. For example, the enhancing tumor sub-region is described by areas that show hyper-intensity in T1Gd scan when compared to T1 scan.

Accurate and robust segmentation of gliomas in pre-operative MRI scans, conventionally done by expert board-certified neuroradiologists, can provide quantitative morphological characterization and measurement of gliomas sub-regions. This quantitative analysis has great potential for diagnosis and research, as it can be used for grade assessment of gliomas and planning of treatment strategies. But this task is challenging due to the high variance in appearance and shape, ambiguous boundaries and imaging artifacts. Until now, automatic segmentation of brain tumors in multimodal MRI scans is still one of the most difficult tasks in medical image analysis.

In recent years, deep convolutional neural networks (CNNs) have achieved great success in the field of computer vision. Inspired by the biological structure of visual cortex, CNNs are artificial neural networks with multiple hidden convolutional layers between the input and output layers. They have non-linear property and are capable of extracting higher level representative features. CNNs have been applied into a wide range of fields and achieved state-of-the-art performance on tasks such as image recognition, instance detection, and semantic segmentation.

In this paper, we present a novel deep-learning based framework to segment brain tumor and its sub-region from MRI scans, as well as to perform survival prediction based on the segmented tumor sub-regions. Our automatic framework for brain tumor segmentation and survival prediction achieves excellent performance on BraTS dataset.

2 Related work

2.1 Early methods for brain tumor segmentation

To date, many automatic methods have been developed for brain tumor segmentation. Most of early reported methods are based on extracting features that represent different tissue or by registration to an anatomical template. Prastawa et al. [16] proposed a method that segment brain tumors simultaneously with the detection of edema. Gooya et al. [6] presented an approach for joint segmentation and deformable registration of brain scans of glioma patients to a normal atlas. Although these traditional segmentation methods achieve acceptable performances, new trend of deep-learning based brain tumor segmentation techniques are also emerging with the state-of-the-art results.

2.2 Deep-learning based methods for brain tumor segmentation

With the advancement of deep learning in recent years, many methods based on convolutional neural network have been developed for brain tumor segmentation. These methods generally adopt an end-to-end structure and perform pixelwise prediction. Kamnitsas et al. [12] presented a method to ensemble different models and architectures for robust performance through combination of predictions from different methods. Wang et al. [21] developed a cascaded approach to decompose the multi-class segmentation problem into a sequence of three binary

segmentation problems according to the hierarchy of glioma sub-region. Isensee et al. [11] developed a model based on U-Net [17] structure for brain tumor segmentation. These methods achieve current state-of-the-art result for brain tumor segmentation.

3 Methodology

3.1 Brain tumor segmentation

Data preprocessing Since the intensity value of MRI is dependent on the imaging protocols and scanners used, we applied intensity normalization to reduce the bias in imaging. More specifically, the intensity value of each MRI is subtracted the mean and divided by the standard deviation of the brain region. In order to reduce overfitting, we applied random flipping and random gaussian noise to augment the training set.

Network architecture In order to perform accurate and robust brain tumor segmentation, we designed the network architecture by incorporating multi-view fusion, multiple-scale prediction as well as skip connection. In addition, the segmentation probability map is post-processed by dense CRF [14]. The overall structure of our network is depicted as follows:

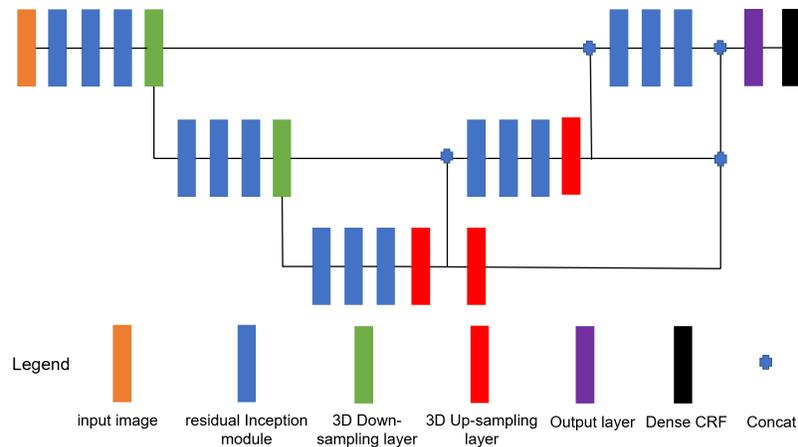


Fig. 1. Architecture overview

The general structure of our proposed network is adapted from the U-Net [17], one of the most commonly used networks in semantic segmentation. Our network is composed of a contracting part and expanding part. The contracting part is used to capture context and encode deep level features, it contains six

residual Inception modules which will be described later, and two downsampling layers, which are used to reduce the number of parameters in network and computational cost. The expanding part is used to restore details and produce the segmentation probability map the same size as the input image, it contains nine residual Inception modules and two upsampling layers. Since deep features in the network encode the abstract overall representations of input, while shallow features have higher resolution and can be used to precisely localize the structures of interest. We also introduce skip connections to integrate both high-level and low-level features. Since features extracted by deep layer encode abstract information about the structure while shallow level features have higher resolution, we adopt a multi-scale approach to combine features from different levels together, as illustrated in the figure above.

The structure of residual Inception module used in our network is illustrated as follows:

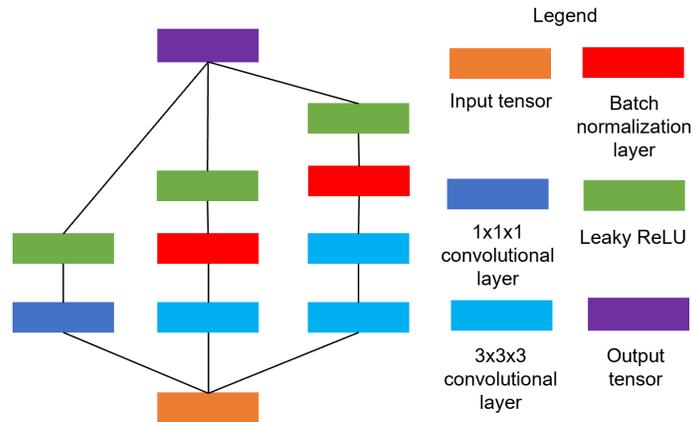


Fig. 2. Structure of residual Inception module

The residual Inception module is the basic building block of our proposed network. Inspired by Inception modules presented by Szegedy et al. [19], our proposed residual Inception module is composed of three 3D convolutional pathways, each with different filter size thus they have different receptive field. After convolution, their results will be concatenated, and a convolutional layer is utilized to integrate the results from three pathways as output. This structure enables the network to automatically select the optimal receptive field. In order to achieve effective training, we also introduce residual connections to each residual Inception module, which directly add the input of residual Inception module to the output, bypassing the parameterized layers in a network. This identity mapping connection can make information propagation smooth and enable the

training of deeper networks. [8] Batch normalization [10] and exponential linear units [4] are also used to speed up learning.

Multi-view fusion and dense CRF refinement 3D contextual information can be used in complementary to reduce false-positive in brain tumor segmentation, we train the network in axial, sagittal and coronal views respectively. Then we fuse the segmentation results from these three views to get the final segmentation during testing time. More specifically, we average the softmax outputs in these three views for fusion. After segmentation, the result is further refined using dense CRF as post-processing step. CNNs tend to predict similar values for spatially near voxels because they share substantial spatial context. In addition, artifact in imaging and noise may cause CNNs to predict false positive isolated tumor regions. [13] Thus, we incorporate dense CRF to improve the performance of segmentation.

For an input image I and the label segmentation z , denote $P(z_i|I)$ as the tumor probability reported by CNN for voxel i , we can have the Gibbs energy in a CRF model:

$$E(z) = \sum_i \psi_u(z_i) + \sum_{ij, i \neq j} \psi_p(z_i, z_j) \quad (1)$$

where the unary potential $\psi_u(z_i) = -\log P(z_i|I)$, while the pairwise potential $\psi_p(z_i, z_j) = \mu(z_i, z_j)k(f_i, f_j)$ for between any pair of voxels, in which f_i and f_j are feature vectors of the pair of voxels.

Loss function We found that the brain tumor segmentation is a highly imbalanced problem. for example, there are 166 times as much background voxels as enhancing tumor voxels in the BraTS 2018 dataset [1–3, 15]. In order to address this class imbalance, we utilized the Generalised Dice overlap proposed by Sudre et al. [18] as the loss function. It is defined as:

$$GDL = 1 - 2 \frac{\sum_{l=1}^2 \omega_l \sum_n r_{ln} p_{ln}}{\sum_{l=1}^2 \omega_l \sum_n r_{ln} + p_{ln}} \quad (2)$$

where r_n represents the voxel values of reference foreground segmentation and p_n represents the predicted probabilistic map for the foreground label.

3.2 Prediction of patient overall survival (OS)

In order to achieve accurate and robust prediction on patient’s overall survival, we extract radiomic features from different tumor’s sub-regions based on segmentation result. Then we use LASSO [20] to select a subset of relevant features and reduce overfitting. Finally, we utilize ridge regression [9] to build prediction model based on selected features. This task is difficult because we only have patient’s imaging data and age without any knowledge about patient’s treatment.

Feature extraction Quantitative phenotypic features from MRI scans can reveal the characteristics of brain tumors. The feature types we extracted include 16 shape descriptors and features extracted from original and derived images (LoG with 5 sigma levels, 1 level of Wavelet decompositions yielding 8 derived images and images derived using Square, Square Root, Logarithm and Exponential filters) using the *PyRadiomics* toolbox. [7] Moreover, we extract features for three kinds of segmented brain tumor sub-region, including non-enhancing tumor core, peritumoral edema and GD-enhancing tumor respectively. Since enhancing tumor sub-region is described by areas that show hyper-intensity in T1Gd scan when compared to corresponding T1 scan, we extract enhancing tumor related features from segmented sub-region of T1Gd scan. Similarly, since the appearance of the necrotic and the non-enhancing tumor core is typically hypo-intense in T1-Gd when compared to T1, we extract necrotic and the non-enhancing tumor related features from segmented sub-region of T1Gd scan. In addition, since peritumoral edema is typically depicted by hyper-intense signal in FLAIR, we extract enhancing tumor related features from segmented sub-region of FLAIR scan. In total, we extracted 3394 radiomic features plus the age of patient. The values of these features were normalized by removing the mean and scaling to unit variance.

Feature selection and survival prediction In order to remove the redundant or irrelevant features we extracted, as well as to reduce overfitting in model, we applied least absolute shrinkage and selection operator (LASSO) [20] to select a subset of features that have the most predictive power. It was put forwarded by Tibshirani et al. to perform parameter estimation and variable selection simultaneously in regression. LASSO minimizes the absolute sum of the coefficients with L1 regularization. The LASSO is defined by

$$\min_{\beta_0, \beta} \left\{ \frac{1}{N} \sum_{i=1}^N (y_i - \beta_0 - x_i^T \beta)^2 \right\} \text{ subject to } \sum_{j=1}^p |\beta_j| \leq t. \quad (3)$$

We applied feature selection by LASSO regression. More specifically, we remove the features whose weights in cost function are less than 0.1. After filtering, we ended up in 157 features. These features were feed into Ridge regression [9] for survival prediction.

4 Experiments

4.1 Dataset

We utilize the BraTS 2018 dataset [1–3, 15] to evaluate the performance of our methods. The training set contains images from 285 patients, including 210 HGG and 75 LGG. The BraTS 2018 validation set contains MRI scans from 66 patients with brain tumors of unknown grade. Each patient was scanned with four

sequences: T1, T1c, T2 and FLAIR. All the images were skull-stripped and re-sampled to an isotropic $1mm^3$ resolution, and the four sequences of the same patient had been co-registered. The ground truth was obtained by manual segmentation results given by experts. Segmentation annotations comprise of the following tumor subtypes: Necrotic/non-enhancing tumor (NCR), peritumoral edema (ED), and Gd-enhancing tumor (ET). Resection status and patient age are also provided. The overall survival (OS) data, defined in days are also included in training set.

4.2 Segmentation result

We train the model using the BraTS 2018 dataset and methods mentioned above. Then we applied the trained model to the BraTS 2018 validation set. The predicted segmentation labels are uploaded to the CBICA's Image Processing Portal (IPP) for evaluation. The evaluation result of segmentation on training set and validation set are listed as follows:

Table 1. Segmentation evaluation result

Phase	Training	Validation
Dice_ET	0.77616	0.75081
Dice_WT	0.85711	0.86494
Dice_TC	0.79446	0.71968
Sensitivity_ET	0.76425	0.78154
Sensitivity_WT	0.88202	0.90872
Sensitivity_TC	0.81931	0.74837
Specificity_ET	0.998	0.998
Specificity_WT	0.991	0.99144
Specificity_TC	0.996	0.99656

4.3 Survival prediction result

Based on the segmentation result of brain tumor sub-regions, we extract features from brain tumor sub-regions segmented from MRI scans and trained the survival prediction model as described above. Then we use the model to predict patient's overall survival on validation set. The evaluation result is listed as follows:

Table 2. Survival prediction evaluation result

Phase	mean squared error	median squared error
Validation	199476.283	107752.542

4.4 Implementation details

To implement the algorithm, we use Tensorflow framework and a NVIDIA M40 GPU accelerator. We use stochastic gradient descent optimizer with momentum 0.9, and set initial learning rate to 0.001. We also set the batch size to 3 for training.

5 Conclusion

In this paper, we present an automatic framework for brain tumor segmentation and survival prediction. Our proposed network architecture for segmentation incorporates multi-view fusion, multi-scale prediction as well as skip connection. In addition, we refine the segmentation result using dense CRF. For survival prediction, we extract the radiomic features from the segmented brain tumor sub-regions. Then we remove irrelevant and redundant features by LASSO, and finally perform survival prediction using Ridge regression. Our automatic framework for brain tumor segmentation and survival prediction achieves excellent performance on BraTS 2018 challenge. In the future, we will explore different network architecture and training strategy to further improve our result. We will also design new features and optimize our feature selection methods for survival prediction.

References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive* **286** (2017)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *The Cancer Imaging Archive* **286** (2017)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data* **4**, 170117 (2017)
4. Clevert, D.A., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289* (2015)
5. Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C., Parkin, D.M.: Estimates of worldwide burden of cancer in 2008: Globocan 2008. *International journal of cancer* **127**(12), 2893–2917 (2010)
6. Gooya, A., Pohl, K.M., Bilello, M., Biros, G., Davatzikos, C.: Joint segmentation and deformable registration of brain scans guided by a tumor growth model. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 532–540. Springer (2011)

7. van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G., Fillion-Robin, J.C., Pieper, S., Aerts, H.J.: Computational radiomics system to decode the radiographic phenotype. *Cancer research* **77**(21), e104–e107 (2017)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
9. Hoerl, A.E., Kennard, R.W.: Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**(1), 55–67 (1970)
10. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015)
11. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In: *International MICCAI Brainlesion Workshop*. pp. 287–297. Springer (2017)
12. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., et al.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: *International MICCAI Brainlesion Workshop*. pp. 450–462. Springer (2017)
13. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis* **36**, 61–78 (2017)
14. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected crfs with gaussian edge potentials. In: *Advances in neural information processing systems*. pp. 109–117 (2011)
15. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging* **34**(10), 1993 (2015)
16. Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. *Medical Image Analysis* **8**(3), 275 – 283 (2004). <https://doi.org/https://doi.org/10.1016/j.media.2004.06.007>, <http://www.sciencedirect.com/science/article/pii/S1361841504000295>, *medical Image Computing and Computer-Assisted Intervention - MICCAI 2003*
17. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 234–241. Springer (2015)
18. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 240–248. Springer (2017)
19. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 2818–2826 (2016)
20. Tibshirani, R.: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 267–288 (1996)
21. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: *International MICCAI Brainlesion Workshop*. pp. 178–190. Springer (2017)

End-to-End Deep Learning versus Classical Regression for Brain Tumor Patient Survival Prediction

Yannick Suter¹, Alain Jungo¹, Michael Rebsamen¹, and Mauricio Reyes¹

Institute for Surgical Technology and Biomechanics, University of Bern, Bern,
Switzerland

yannick.suter@istb.unibe.ch

Abstract. End-to-end deep learning for regression tasks on medical imaging data shows promising results. We evaluate 3D-Convolutional Neural Networks (CNNs) and classical regression methods with hand-crafted features for survival regression of patients with high grade brain tumors. The tested 3D-CNNs for regression showed promising but unstable results, but were outperformed by linear regression using 16 radiomic features. Among the neural network methods, a multi-layer perceptron (MLP) performed best. The accuracy of linear regression on the BraTS validation set was 0.571, while the MLP reached 0.515. If the same CNNs architectures are used for classification instead of regression, we get a comparable accuracy on our validation data split, but again with unstable results between training epochs. A CNN combined with age and resection status information reached an accuracy of 0.515 on our validation split, followed by a k-Nearest-Neighbor regressor with 0.424. More training data is necessary for a stable performance of a CNN model for direct regression from MR images.

Keywords: Brain tumor · Survival prediction · Regression · 3D-CNNs

1 Introduction

High-grade gliomas are the most frequent primary brain tumors in humans. Due to their rapid growth and infiltrative nature, the prognosis for patients with gliomas ranking at grade III or IV on the WHO grading scheme [13] is very poor, with a median survival time of only 14 months. Finding biomarkers based on Magnetic Resonance (MR) imaging data could lead to an improved disease progression monitoring and support clinicians in treatment decision-making.

The progress in the fields of automated tumor segmentation and radiomics have led to many different approaches to predict the survival time of high-grade glioma patients. The introduction of the survival prediction task in the BraTS challenge 2017 [14] makes a direct performance comparison possible.

The current state of the art approaches can roughly be classified into

1. Classical radiomics: Extracting intensity features and/or shape properties from segmentations and use regression techniques such as Random Forest

(RF) regression [5], Logistic regression, or sparsity enforcing methods such as LASSO [17].

2. Deep features: Neural networks are used to extract features, which are subsequently fed into a classical regression method such as Logistic Regression. [6] and Support Vector and Regression (SVR) [10].
3. A combination thereof (e.g., [11]).
4. End-to-end survival regression from MR data using neural networks with or without additional non-imaging input (e.g. [12]).

Predicting the survival from pre-treatment MR data is inherently difficult, due to the high impact of the extent of resection (e.g., [16]) and response of the patient to chemo- and radiation therapy.

To be able to benchmark new approaches against established techniques, we present experiments ranging from simple linear models to end-to-end 3D-Convolutional Neural Networks (CNNs) and combinations of classical radiomics with Deep Learning. We believe that a thorough comparison and discussion will provide a good baseline for future investigations of survival prediction tasks.

We only present experiments for the survival prediction challenge and rely on an established tumor segmentation method (see below).

Experiments with 3D-CNNs for end-to-end regression confirmed observations made by other groups in last year’s competition (e.g. [12]), that these models tend to converge and overfit extremely fast on the training set, but show poor generalization when tested on the hold-out samples. The relatively few samples to learn from may be the reason that the top-ranked methods from last year’s competition used techniques that perform better with sparse training data, such as Random Forest regression, with fewer learnable parameters compared to a 3D-CNN.

2 Methods

2.1 Data

The provided training and validation datasets for the survival prediction task consist of 163 and 53 subjects, respectively. A data point (i.e. subject) contains imaging and clinical data.

The imaging data comprises the four standard magnetic resonance image (MRI) sequences (T1-weighted (T1), T1-weighted post-contrast (T1c), T2-weighted, T2 fluid-attenuated inversion-recovery (FLAIR)). All images in the datasets are resampled to isotropic voxel size ($1 \times 1 \times 1 \text{ mm}^3$), size-adapted to $240 \times 240 \times 155 \text{ mm}^3$, and skull-stripped. Moreover, the sequences are co-registered.

The clinical data comprises a subject’s age and resection status. The three possible resection statuses are: (a) gross total resection (GTR), (b) subtotal resection (STR), and (c) not available (NA).

2.2 Survival Features

Segmentation We rely on brain tumor segmentations for the experiments that use traditional survival features. In the validation and testing dataset, the segmentation is not provided due to overlap with the data of the BraTS18 segmentation task. To obtain the required segmentations, we thus employ the method presented by Wang et al. [18]. Their method is publicly available¹ and contains models pre-trained on the BraTS17 [2–4] training dataset. This dataset is identical to the BraTS18 [2–4] training dataset which enables us to compute the segmentations with the available pre-trained models without retraining a new segmentation network.

Feature Extraction We extract 4465 survival features from the computed segmentation together with the four sequences (i.e., T1, T1c, T2, FLAIR). A majority of the features (4451 out of 4465) are obtained by the open-source *pyradiomics*² python package [8] and include various shape, gray-level, gray-tone, and first-order intensity features for all tumor sub-compartments (i.e., enhancing tumor, edema, necrosis with non-enhancing tumor) and different image types (i.e., original, Laplacian of Gaussian). The remaining features consist of 12 additional enhancing tumor shape features previously used as predictors for survival [9, 15]. These features are the rim width of the enhancing tumor, geometric heterogeneity, and combinations of rim width quartiles.

By adding the two provided clinical features (i.e., resection status, age) to the extracted features, the features count sums to 4465.

Feature Selection Since the number of features is much higher than the available samples, a subset of features is used. Apart from being necessary for many machine learning methods, a reduction of the feature space improves the interpretability of possible markers regarding survival. Based on a step wise forward/backward selection [7] with a linear model, we selected the feature set listed in table 1.

2.3 Analyzed Survival Prediction Methods

To have an ample choice of approaches to benchmark, we investigated both classical radiomic models with hand-crafted features, end-to-end CNNs, and a combination thereof.

3D-CNNs Two different CNNs were built for the survival regression task (see figure 1). CNN1 consists of three convolutional layers with decreasing kernel sizes with intermediary max-pooling, followed by fully-connected layers connected to the single value regression target. For a combination with hand-crafted features

¹ <https://github.com/taigw/brats17>

² <https://github.com/Radiomics/pyradiomics>

Table 1. Reduced Feature Set with Sequence (Seq.) and Tumor Compartments.

Seq.	Feature
<i>Contrast-enhancing tumor</i>	
T1c	log-sigma-3-0-mm-3D_gldm_DependenceNonUniformity
T2	original_gldm_HighGrayLevelEmphasis
T2	log-sigma-1-0-mm-3D_gldm_Correlation
Flair	log-sigma-1-0-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis
T1c	log-sigma-5-0-mm-3D_gldm_ClusterShade
<i>Edema</i>	
T1c	log-sigma-3-0-mm-3D_gldm_DependenceNonUniformity
T2	original_gldm_HighGrayLevelEmphasis
T2	log-sigma-1-0-mm-3D_gldm_Correlation
Flair	log-sigma-1-0-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis
T1c	log-sigma-5-0-mm-3D_gldm_ClusterShade
<i>Necrosis and non-enhancing tumor</i>	
T1c	log-sigma-3-0-mm-3D_gldm_DependenceNonUniformity
T2	original_gldm_HighGrayLevelEmphasis
T2	log-sigma-1-0-mm-3D_gldm_Correlation
Flair	log-sigma-1-0-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis
T1c	log-sigma-5-0-mm-3D_gldm_ClusterShade
<i>Subject Information</i>	
-	Age

or demographic information, features may be appended to the first fully connected layer.

CNN2 consist of five blocks with an increasing number of filters, each block has two convolutional layers and a max pooling operation. These blocks are again connected to two fully connected layers in front of the output.

With the four MR sequences and the segmentation labels, five channels were used at the input of the networks.

FeatNet The third neural network approach is a multi-layer perceptron (MLP). This experiment was motivated by (a) enforcing sparsity regarding features if fed the full hand-crafted feature vector and (b) to test the performance of a model with much fewer parameters compared to CNN1 and CNN2 if a feature subset is used.

Regression Methods With Hand-Crafted Features Since the number of samples is relatively low, we included the following methods

- Linear Regression
- Random Forest Regression
- k-Nearest-Neighbor Regression (kNN) [1]
- Support Vector Regression (SVR)

For a subset of methods, a direct classification alongside the regression was tested. Since the performance on the BraTS validation set can only be assessed through a regression result, these classifiers were only evaluated on our validation split on the training data.

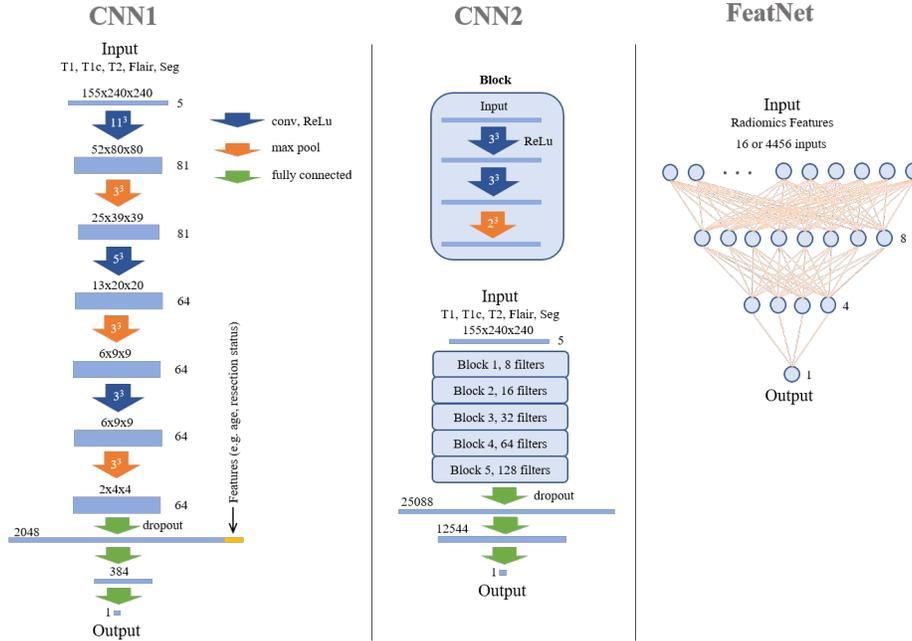


Fig. 1. The CNNs and MLP used for survival regression. In the experiment where CNN1 was used for classification, three outputs with a softmax operation was used instead of the single output value for regression.

2.4 Evaluation

All presented methods were evaluated on a split of the training data of 33 subjects, and on the BraTS validation set. For the CNNs, Spearman’s Rank coefficient was used on our validation set as a model selection criterion during training. The regression output was subsequently thresholded to classify each subject as short, medium or long survivor, with survival times of less than ten, between ten and 15, and longer than 15 months respectively.

The evaluation on the BraTS validation data is performed only on the subset of subjects with resection status GTR (28 out of the 53 validation subjects).

3 Preliminary Results

3.1 Regression

The two convolutional neural networks overfit very fast on the training data, and showed highly variable performance between epochs. Model selection during training was therefore a challenge, since the accuracy and Spearman’s rank coefficient were very unstable as well. The best accuracy of CNN2 on our validation split could not be replicated on the BraTS validation set, where it performed poorly with an accuracy of 0.37. CNN1 showed worse results on our validation split than the deeper CNN2, but performed better on the BraTS validation data. Adding age and resection status to the fully connected layer led to a faster convergence during training, but not to a better final accuracy. Including hand-crafted features showed lower accuracy in preliminary experiments.

The MLP ”FeatNet” with the top 16 radiomic features as input reached the highest accuracy on the BraTS validation set among the neural network models.

Linear regression performed best over all tested methods on the BraTS validation data when the top 16 features were used.

3.2 Direct Classification

Of the tested direct classification methods, CNN1 reached the highest accuracy of 0.515 on our validation split, followed with 0.424 by a kNN Classifier using the top four features.

For this seemingly easier task, no higher or more stable accuracy could be achieved.

4 Conclusions

More data would be needed to fully benefit from direct survival estimation with 3D-CNNs. When inspecting the filters of CNN1 and CNN2, most of the learning seems to take place at the fully connected layers. The fast overfitting of the CNN models led to a poor performance on unseen data.

Using classical regression techniques with hand-crafted features has the advantage of better interpretability. Models with fewer learnable parameters, such as the classical regression methods tested and FeatNet achieve more robust results on unseen data when only few training samples are available.

A robust stopping criterion during training with a small validation set poses a challenge.

More research should go into the combination of deep features and classical hand-crafted features. We expect the performance of survival regression on pre-resection MR data to reach a clinically possible limit, since no treatment information is available.

Method	Validation on 33 subjects split from training data			BraTS Validation Set		
	Acc.	MSE / days ²	r _S	Acc.	MSE / days ²	r _S
Regression						
Lin. Reg. (top 16 features)	0.333	49372	0.3845	0.571	99370	0.251
Random Forest Reg.	0.394	101287	0.199	0.357	152130	0.082
kNN Regressor (top 16 features)	0.303	63542	-0.021	0.5	104923	0.17
SVR (top 4 features)	0.303	63542	-0.021	0.5	104923	0.17
FeatNet (top 16 features)	0.515	29916	0.353	0.536	114850	0.19
FeatNet (all features)	0.364	50050	0.2684	0.429	103878	0.153
CNN1	0.424	56496	0.235	0.394	157617	-0.112
CNN1 with age & resection status	0.394	61798	-0.194	0.444	137912	-0.005
CNN2	0.515	50598	0.298	0.37	172821	0.104
Classification						
Random Forest Cl.	0.364	-	-	-	-	-
kNN Classifier (top 4 features)	0.424	-	-	-	-	-
CNN1 Classifier with age & resection status	0.515	-	-	-	-	-

5 Acknowledgements

We gladly acknowledge the support of the Swiss Cancer League (grant KFS-3979-08-2016) and the Swiss National Science Foundation (grant 169607). We are grateful for the support of NVIDIA corporation for the donation of a Titan Xp GPU. Calculations were partly performed on UBELIX, the HPC cluster at the University of Bern.

References

1. Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* **46**(3), 175–185 (1992)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. *The Cancer Imaging Archive* (2017)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. *The Cancer Imaging Archive* (2017)
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data* **4**, 170117 (9 2017). <https://doi.org/10.1038/sdata.2017.117>

5. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and regression trees (1984)
6. Cox, D.R.: The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 215–242 (1958)
7. Friedman, J., Hastie, T., Tibshirani, R.: The elements of statistical learning, vol. 1. Springer series in statistics New York, NY, USA: (2001)
8. van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G., Fillion-Robin, J.C., Pieper, S., Aerts, H.J.: Computational radiomics system to decode the radiographic phenotype. *Cancer research* **77**(21), e104–e107 (2017)
9. Jungo, A., McKinley, R., Meier, R., Knecht, U., Vera, L., Pérez-Beteta, J., Molina-García, D., Pérez-García, V.M., Wiest, R., Reyes, M.: Towards uncertainty-assisted brain tumor segmentation and survival prediction. In: *International MICCAI Brainlesion Workshop*. pp. 474–485. Springer (2017)
10. Lampert, C.H., et al.: Kernel methods in computer vision. *Foundations and Trends® in Computer Graphics and Vision* **4**(3), 193–285 (2009)
11. Lao, J., Chen, Y., Li, Z.C., Li, Q., Zhang, J., Liu, J., Zhai, G.: A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Scientific reports* **7**(1), 10353 (2017)
12. Li, Y., Shen, L.: Deep learning based multimodal brain tumor diagnosis. In: *International MICCAI Brainlesion Workshop*. pp. 149–158. Springer (2017)
13. Louis, D.N., Perry, A., Reifenberger, G., Von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., Ellison, D.W.: The 2016 world health organization classification of tumors of the central nervous system: a summary. *Acta neuropathologica* **131**(6), 803–820 (2016)
14. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging* **34**(10), 1993–2024 (10 2015). <https://doi.org/10.1109/TMI.2014.2377694>
15. Pérez-Beteta, J., Martínez-González, A., Molina, D., Amo-Salas, M., Luque, B., Arregui, E., Calvo, M., Borrás, J.M., López, C., Claramonte, M., Barcia, J.A., Iglesias, L., Avelillas, J., Albillo, D., Navarro, M., Villanueva, J.M., Paniagua, J.C., Martino, J., Velásquez, C., Asenjo, B., Benavides, M., Herruzo, I., Delgado, M.d.C., del Valle, A., Falkov, A., Schucht, P., Arana, E., Pérez-Romasanta, L., Pérez-García, V.M.: Glioblastoma: does the pre-treatment geometry matter? A postcontrast T1 MRI-based study. *European Radiology* (2017). <https://doi.org/10.1007/s00330-016-4453-9>
16. Sanaï, N., Polley, M.Y., McDermott, M.W., Parsa, A.T., Berger, M.S.: An extent of resection threshold for newly diagnosed glioblastomas. *Journal of neurosurgery* **115**(1), 3–8 (2011)
17. Tibshirani, R.: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 267–288 (1996)

18. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: International MICCAI Brainlesion Workshop. pp. 178–190. Springer (2017)

End-to-End Cascade Network for 3D Brain Tumor Segmentation in MICCAI 2018 BraTS Challenge

Kuan-Lun, Tseng¹, Winston Hsu¹

National Taiwan University

Abstract. In this paper we present our solution for brain tumor segmentation in MICCAI 2018 BraTS data set. We handle the imbalance problem with cascade model architecture and further use focal loss to focus on difficult region. The preliminary results in validation data shows great potential.

Keywords: Brain Tumor Segmentation · Image Segmentation · Convolutional Neural Network

1 Introduction

Brain tumor segmentation is a challenging task in biomedical image research field. Recently, deep learning based on convolutional neural network achieved huge success. Since fully convolutional networks like FCN [11] and UNet [15] was proposed, basically all the top performing methods adopted similar architecture design.

Inspired by [17] and [9], we use 3D convolutional network with a cascade and multi-stage framework to alleviate class imbalance which is the main difficulty in brain tumor segmentation. We used the three stages for segmenting whole, core and enhancing tumor region into one cascade structure. In this way, the later stage network can focus on learning difficult area.

This cascade framework for semantic segmentation was originally proposed by [9] but they use difficulty to separate stage2 while we use the structure of the tumor. 3D convolutional network can aggregate depth information directly but the memory consumption is too high. To directly combine three network and maintain the necessary network capacity can caused significant memory usage. To alleviate this problem, we use anisotropic convolutional network suggested by [17].

With the observation in [9], the difficulty in each voxel is different and learning those difficult voxels is important to improve semantic segmentation result. In brain tumor segmentation, the healthy tissue and background take more than 80% of the volume which can be easily classified while the contour of core and enhancing tumor is much harder. Since we choose to use tumor structure instead of difficulty to separate the cascade layer, we opt to use loss function to address this problem. We use binary focal loss [10] in all stages to down weighted

easy examples. In the following, we describe related research and our method in detail.

2 Related Work

Havaei et al. [4] use 2D convolutional neural network with a multi-scale architecture by combining features from different patch resolution. In this way, both local features as well as more global contextual features can be learned despite the input patch size is small. In addition, they also explore a cascade architecture in which the output of a first network is treated as input for subsequent networks. Pereira et al. [14] use more small kernel size instead of one large kernel to increase effective receptive field which is a common choice in modern CNN architecture. They further boost the performance by using two networks training on high grade glioma (HGG) and low grade glioma (LGG) separately because the two types of glioma are different in nature and the training cases is imbalance. Kamnitsas et al.[6] use 3D CNN and training with 3D patches to achieve high efficiency. Two parallel pathways with small 3D kernel and residual connection that take down-sampled context as input to avoid to computation on large 3D volumes at full resolution. They also developed 3D version of fully connected CRF [8] as post-processing to improve segmentation result. Milletari et al. [13] proposed dice loss to better handle the class imbalance. Dice loss and its variant is commonly used in biomedical segmentation task such as [5] [16]. Triple Cascaded Framework [17] use three networks target for three different regions where whole tumor, core tumor and enhancing tumor. Since tumor core is contained in the whole tumor region and enhancing tumor is contained in core region by definition, we can segment each part in a cascade manner. They also trained three network for different orthogonal view and fuse them together to better leveraging contextual information of 3D volume.

3 Method

3.1 Data

The BraTS challenge [12] [2] [3] [1] training dataset includes 210 different MRI files from high grade glioma (HGG) and 75 MRIs from low grade glioma (LGG) from 19 institutions. The MRI have four modalities for each patient which are T1, T1 contrast, T2 and FLAIR. The annotation was annotated manually by one to four raters and all segmentations were approved by expert raters. The annotations separate the brain tumor into four different labels: Background and healthy tissue, Non-Enhancing tumor, Edema and Enhancing tumor.

3.2 Data Preprocess

We crop the brain region inside the whole 3D volume and randomly crop the tumor region to feed into the network. We normalize each modality of each case individually, with the mean and standard deviation of the brain intensities.

3.3 Network Architecture

End-to-End Cascade Network Inspired By Triple Cascaded Framework [17] and [9], We propose to use cascade framework but combine the three networks to an end-to-end trainable architecture. The design of [17] has great advantages but separate training is time-consuming and end-to-end training allows subsequent networks to provide more information.

The three networks are trained together and produce three predictions and uses three loss function to supervise (see Fig. 1). Each stage is responsible for whole, core and enhancing region respectively.

The input for first stage are the patch cropped to the brain tissue bounding box. The mask for first is also generated from the same bounding box. First stage have two down sampling convolution layers with stride 2. Note that in down-sampling layer we follow the same design choice as [17] which maintain the stride of depth dimension to one. In the following stages, we maintain the feature map resolution to preserve structure informations and used dilated convolution to increase effective receptive field instead as in [18] [17]. The prediction is produced by two consecutive convolution transpose operations to restore the original resolution (the green part in Fig.1).

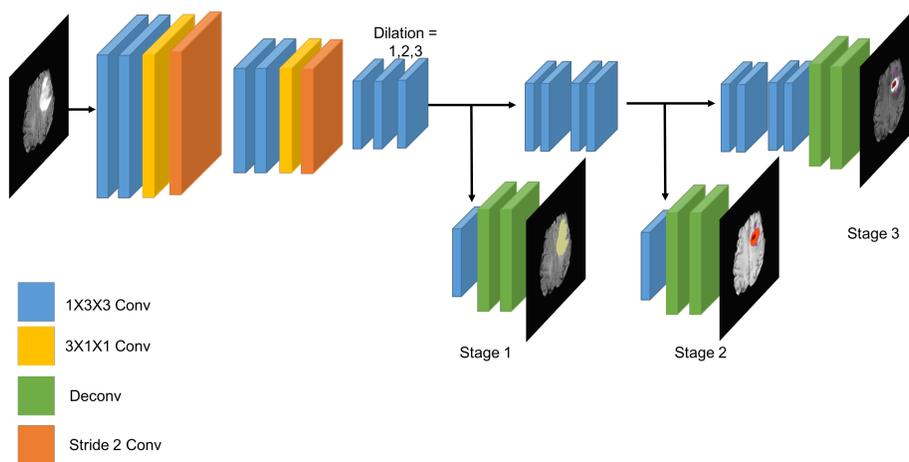


Fig. 1. Our cascade model design. All the filter size is 32 and all convolution is 3D. Each stage is an binary classification problem optimized with focal loss.

3.4 Focal Loss

Focal loss proposed by [10] is used in object detection to alleviate the imbalance of positive and negative samples. The same idea can also apply to semantic segmentation of brain tumor since most of the brain tissues are healthy and

easy to be classified. The easy examples will occupy the total loss causing the difficult examples such as enhancing tumor hard to learn. We penalize the easy example which classified with very high probability as Equation. 1.

$$FocalLoss = \begin{cases} -(1-y)^\gamma \log(y), & \text{if } y = 1 \\ -y^\gamma \log(1-y), & \text{if } y = 0, \end{cases} \quad (1)$$

4 Experiment

4.1 Evaluation

We use the official online evaluation system to test our results. The evaluation system separates the tumor structure into three regions due to practical clinical applications.

- Whole Tumor (WT). It considers all tumor areas and evaluates all labels 1, 2, 4 (1 for non-enhancing core, 2 for edema and 4 for enhancing core).
- Tumor Core (TC). It only takes tumor core region into account and measures the labels 1, 4.
- Enhancing Tumor (ET). It only measures label 4.

4.2 Metrics

Three Evaluation metrics are used for the above mentioned regions:

- Dice Coefficient. The Dice-Coefficient is the main metric and commonly used in biomedical image segmentation. This measure states the similarity between prediction and annotations. It penalize false positive and false negative.

$$Dice = \frac{2TP}{2TP + FN + FP} \quad (2)$$

- Sensitivity and Specificity. Sensitivity indicates the ratio between correct predicted voxels and true positive voxels. Specificity indicates the percentage of true negatives we predicted. These metrics can help us to determine whether our method is over-segmenting or under-segmenting of the tumor regions.

$$Sensitivity = \frac{TP}{TP + FN} \quad \text{and} \quad Specificity = \frac{TN}{TN + FP} \quad (3)$$

- Hausdorff Distance. Hausdorff distance measures how far two subsets (our prediction and ground truth) of a metric space are from each other.

$$d_H(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b) \right\}. \quad (4)$$

4.3 Baseline

We compare our method with 3D-Unet and the official release version of [17]. We build 3D-Unet with 2 down-sampling layer and 2 up-sampling layer which is a popular design choice (see Fig. 2). We trained 3D-Unet with patch size $32 \times 128 \times 128$ and batch size 5. We use Adam optimizer [7] with learning rate 1e-3 for 30k iterations.

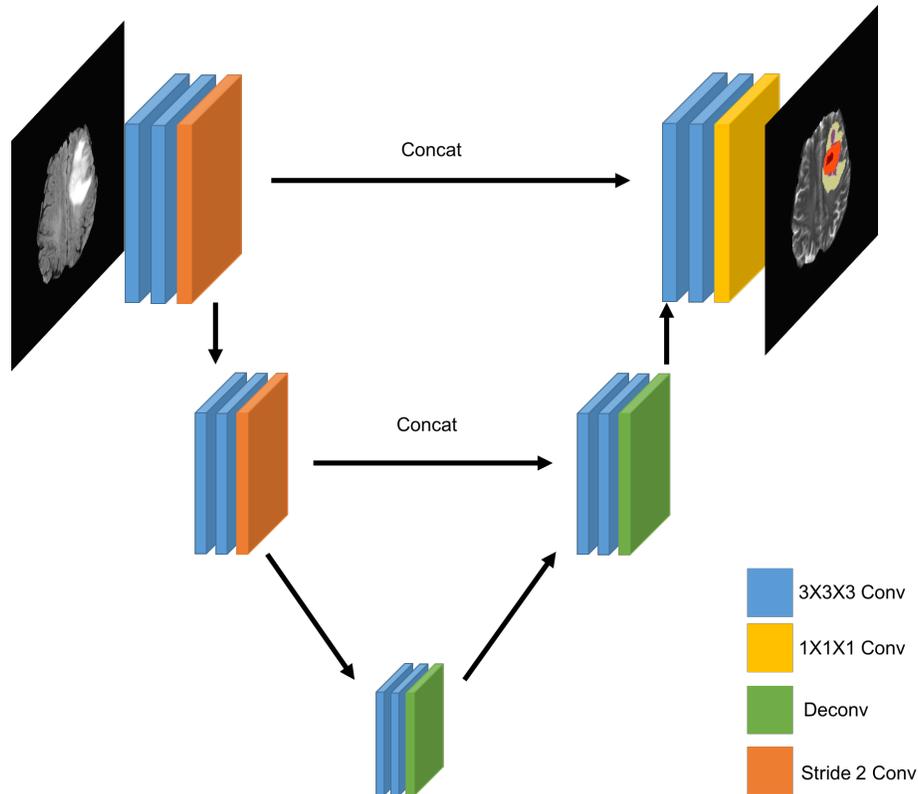


Fig. 2. 3D Unet baseline architecture. All the operations are 3D convolution. The final prediction is obtained by convolving $1 \times 1 \times 1$ kernel. 3D Unet directly output multi-class prediction and optimized with multi-class dice loss [16].

For [17], we preserve the default parameters since the training and testing data is similar.

4.4 Training

We first trained the three networks separately with learning rate 1e-3 for 30k iterations with patch size $19 \times 144 \times 144$. And then fine-tuned the whole network

jointly for another 10k iterations. Noted that We use the focal loss parameter gamma equals to 2 in all experiments (following [10]).

4.5 Results

We report our preliminary results on Brats2018 validation data. All training data were used and without any post-processing (see Table.1). Unet3D is based on our own implementation trained with Dice loss [13] as an baseline. We also used the Cascade [17] with their official released code to train and test since it is the most relevant work to ours. Our approach is fast to train and test although the tumor core performance is worse than [17].

Table 1. Dice score evaluation results on Brats2018 validation data

	Dice_ET	Dice_WT	Dice_TC
Unet3D	0.76397	0.86799	0.78161
Cascade [17]	0.79627	0.90287	0.85425
Ours	0.79855	0.90498	0.80621

Table 2. Sensitivity evaluation results on Brats2018 validation data

	Sensitivity_ET	Sensitivity_WT	Sensitivity_TC
Unet3D	0.77773	0.92016	0.80491
Cascade [17]	0.81572	0.93049	0.85369
Ours	0.84031	0.91737	0.8351

Table 3. Specificity evaluation results on Brats2018 validation data

	Specificity_ET	Specificity_WT	Specificity_TC
Unet3D	0.99796	0.99295	0.99698
Cascade [17]	0.99807	0.99336	0.99785
Ours	0.99773	0.99442	0.99693

5 Conclusions

Following [17], we combined the three networks as one which achieve better training efficiency. We further experimented focal loss function to better handle the imbalance problem. Noted that [17] used multi-view ensemble and we are not. We achieved similar results in whole and enhancing tumor region but not

core region. Meaning that our model cannot separate edema with core tissue well. One possible solution is to increase the receptive field of second stage network since that in [17], the network for core region uses more dilation operations. Another possibility is that our network is over-fitting since the total training iterations are more than [17] (joint training plus separate training). With the preliminary results, we will further investigate the better training strategy and parameter tuning to verify our assumption.

In this paper, we have presented preliminary results in the Brats2018 Validation set which shows great efficiency and achieve competitive results.

References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. The Cancer Imaging Archive (2017). <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. The Cancer Imaging Archive (2017). <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data* **4**, 170117 (2017)
4. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H.: Brain tumor segmentation with deep neural networks. *Medical image analysis* **35**, 18–31 (2017)
5. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In: *International MICCAI Brainlesion Workshop*. pp. 287–297. Springer (2017)
6. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis* **36**, 61–78 (2017)
7. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
8. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected crfs with gaussian edge potentials. In: *Advances in neural information processing systems*. pp. 109–117 (2011)
9. Li, X., Liu, Z., Luo, P., Loy, C.C., Tang, X.: Not all pixels are equal: Difficulty-aware semantic segmentation via deep layer cascade. In: *IEEE Conference on Computer Vision and Pattern Recognition*. vol. 1, p. 2 (2017)
10. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. *arXiv preprint arXiv:1708.02002* (2017)
11. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 3431–3440 (2015)

12. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging* **34**(10), 1993 (2015)
13. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *3D Vision (3DV), 2016 Fourth International Conference on*. pp. 565–571. IEEE (2016)
14. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in mri images. *IEEE transactions on medical imaging* **35**(5), 1240–1251 (2016)
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 234–241. Springer (2015)
16. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 240–248. Springer (2017)
17. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: *International MICCAI Brainlesion Workshop*. pp. 178–190. Springer (2017)
18. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122* (2015)

Brain Tumor Segmentation using Bit-plane and U-NET

Tran Anh Tuan^{1*}, Tran Anh Tuan¹ and Pham The Bao¹

¹ Faculty of Math and Computer Science, University of Science, Vietnam National University

*tatuan@hcmus.edu.vn

tuantran261083@gmail.com

ptbao@hcmus.edu.vn

Abstract. The extraction of brain tumor tissues in 3D Brain Magnetic Resonance Imaging plays an important role in diagnosis gliomas. In this paper, we use clinical data to develop an approach to segment Enhancing Tumor, Tumor Core, and Whole Tumor which are the sub-regions of glioma. Our method starts with Bit-plane to get the most significant and least significant bits which can cluster and generate more images. And then, U-NET, a popular CNN model for object segmentation, is applied to segment all of the glioma regions. In the process, U-NET is implemented by multiple kernels to acquire more accurate results. We evaluated the proposed method with the database BRATS challenge in 2018. Without using additional data, on validation data, our method achieves a performance of 82%, 68% and 70% Dice scores for the whole tumor, enhancing tumor and tumor core, respectively

Keywords: 3D Brain Magnetic Resonance Imaging, Brain Tumor, Bit-plane, 2D U-Net, CNN, BRATS challenge in 2018

1 Introduction

Accurate extraction of brain tumor types plays an important role in diagnosis and treatment planning. Neuro-imaging methods in Magnetic Resonance Imaging (MRI) provide anatomical and pathophysiological information about brain tumors and aid in diagnosis, treatment planning and follow-up of patients. Manual segmentation of brain tumor tissue is a difficult and time-consuming job. Therefore, brain tumor segmentation from 3D Brain MRI automatically will help doctors to overcome these problems. Among many types of brain tumor, Gliomas are the most common primary brain malignancies, with different degrees of aggressiveness, variable prognosis and various heterogeneous histological sub-regions. In this paper, we focus on Enhancing Tumor, Tumor Core, and Whole Tumor segmentation which are the sub-regions of gliomas segmentation.

Segmentation of brain tumors in multimodal MRI scans is one of the most challenging tasks in medical image analysis. Currently, many automatic tumor segmentation methods have been proposed. Many of them use hand-designed features method

and then given to the classifier. However, in recent years, Convolution Neural Networks (CNNs) which have been shown to excel learning a hierarchy task-adapted complex feature are seen prominent success in image classification, object detection and image semantic segmentation [1-3]. Many of the brain tumor segmentation methods based on CNNs or combining CNNs with traditional method are also proposed [4-6].

In this study, we combine the Bit-plane [7] and U-Net [8] method for tumor segmentation. First, we use Bit-plane to divide images into many images by determining significant bits. Second, the images with the significant bits can be used to segment the object boundary. Finally, original images and images with least significant bits can be used to determine tissues inside the boundary. Both stages used the U-NET with multiple kernels to segment the tissues more accurately.

2 Our method

Our method is illustrated in Figure 1. There are three main stages: Preprocess, Object boundary segmentation and Tissues segmentation. As shown in Figure 1, the first U-NET predicts object boundary of the whole tumor and the other U-NET utilizes features to predicts the label of all pixel inside the boundary.

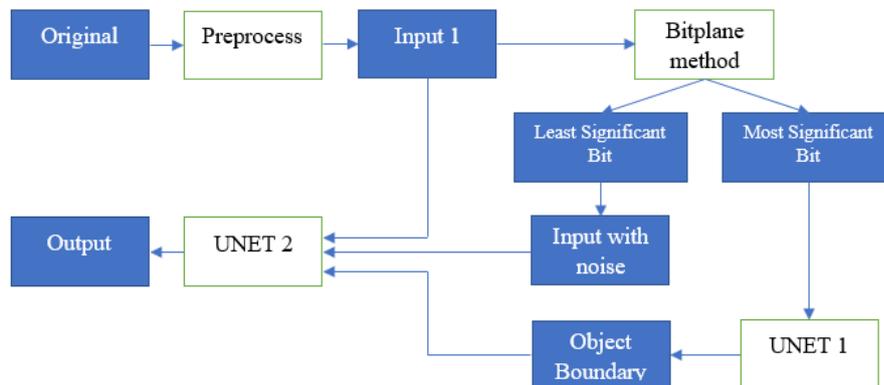


Fig. 1. The overall of our proposed method for brain tumor segmentation

2.1 Preprocessing

The preprocessing is the necessary stage before any tissues segmentation. We implement three main steps

- Scale Range: each individual 3D image is scaled to the range [0-255]
- Category Brain Slices: We group the slices which can contain the tumors together to get the accuracy better. The implementation can be done au-

tomatically by learning feature or set manually by omitting some first and end slices. Here, we detect the tumor from slices 40-140

- **Narrow Object Region:** each 2D image can be cropped to implement deep learning effectively. Here, we cropped the image size from (256,256) to (176,176)

Majority of the volumes in the dataset were acquired along the axial plane and hence had the highest resolution this plane. Therefore, all 3D brain MRI is transformed to 2D brain slices on axial slice extracted from all four sequences. After the preprocessing stage, all the 2D slices is from (155,256,256) to (100,176,176) with value range (0-255)

2.2 Object Boundary Segmentation

The bit plane method [7] is based on decomposing a multilevel image into a series of binary images. The intensities of an image are based on the Eq (1):

$$a_{m-1}2^{m-1} + a_{m-2}2^{m-2} + \dots + a_12^1 + a_02^0 \quad (1)$$

We realize that the final plane contains the most significant bit. We eliminate the last 6 bits to remove the noise and only used the first 2 bits to keep the significant data to generate images for training to detect the object boundary. The U-NET method is proposed to segment the boundary

U-NET is used to segment the background and the whole tumor by combining the 2D slices input and the image which contains 2 significant bits. Fig. 2 shows the results from an example of experiments in the samples of image scans on the real-time data of the BraTS'18. The top and the second row of Fig. 1 are the original images, from the left to the right: (a) FLAIR (b) T1 (c) T1ce and (d) T2). The last row is the result for object boundary segmentation

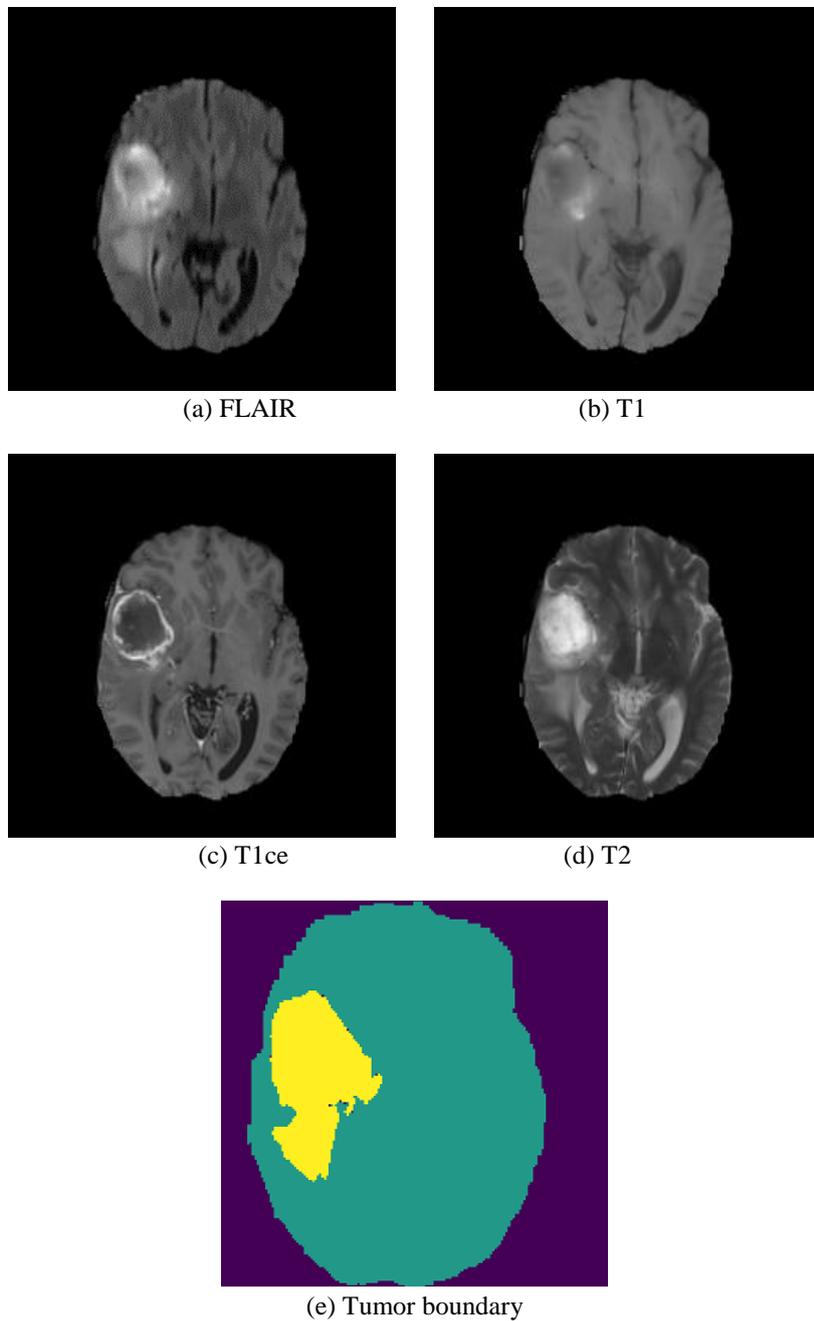


Fig. 2. The implementation results from an example of experiments on the real-time data of the BraTS'18.

2.3 Tissues segmentation

After segment the tumor boundary, we can extract different types of tumors inside the boundary by using U-NET which is shown in Figure 3. The training data is the data which is preprocessed from the first stage.

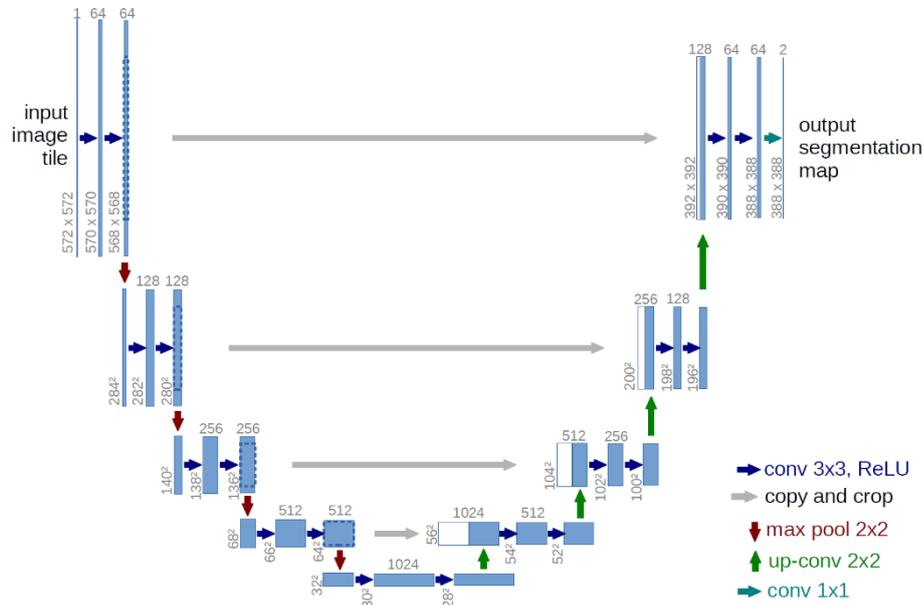


Fig. 3. U-net architecture (example of 32x32 pixels in the lowest resolution) [3]

To get the better results as shown in Figure 4, we have two main contributions to enhance the segmentation:

- Another training data are the images with noises which are generated from the last 3 bits (the least significant bit).
- Implementing U-Net with multiple kernel size to get the better segmentation. Let $K = \{(K_1, (a_1, b_1)), \dots, (K_n, (a_n, b_n))\}$ is the set of n filters K with size (a, b) . The output of layer i is the merge of feature maps that the layer i generate $\cup_{j=1}^n K_j$. Here, in the experiments, the numbers within each Conv block comprises of 2 sets of convolutions by 3x3 kernels and 2 sets of convolutions by 3x5 kernel.

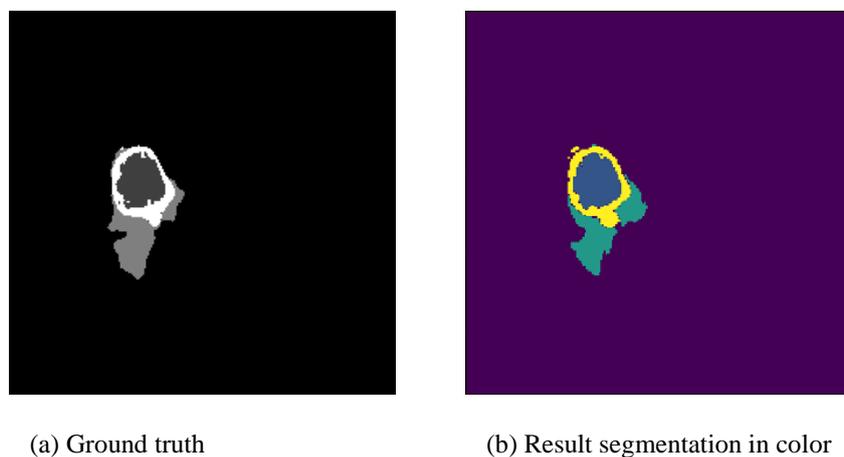


Fig. 4. The results from an example of brain tumor segmentation on the real-time data of the BraTS'18.

3 Results

We use BraTS'2018 training data [4-7], consisting of 210 pre-operative MRI scans of subjects with glioblastoma (HGG) and 75 scans of subjects with lower grade glioma (LGG). These multimodal scans describe a) native (T1) and b) post-contrast T1-weighted (T1Gd), c) T2-weighted (T2), and d) T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes and were acquired with different clinical protocols and various scanners from multiple (n=19) institutions. Ground truth annotations comprise the GD-enhancing tumor (ET — label 4), the peritumoral edema (ED — label 2), and the necrotic and non-enhancing tumor core (NCR/NET — label 1)

Because the testing data has not been available, we only evaluate on the validation data. In Table 1, we show the average performance for each label and score for all the validation patients. Here, we only used the UNET with one kernel (3,3)

Table 1. Dice score, and Hausdorff distance for enhancing tumor (ET), whole tumor (WT) and tumor core (TC).

Label	Dice_ET	Dice_WT	Dice_TC	Hausdorff ET	Hausdorff WT	Hausdorff TC
Mean	0.68252	0.81871	0.69986	7.01652	9.42113	12.46282
StdDev	0.28138	0.16968	0.2913	9.53618	9.74773	14.68491
Median	0.80902	0.88296	0.82567	2.82843	6.04138	6.16441

Expanding upon this evaluation scheme, we will also use the metrics of Sensitivity and Specificity, allowing to determine potential over- or under-segmentations of the tumor sub-regions by participating methods

Table 2. Sensitivity and Specificity for enhancing tumor (ET), whole tumor (WT) and tumor core (TC).

Label	Sen_ET	Sen_WT	Sen_TC	Spec_ET	Speci_WT	Spec_TC
Mean	0.70254	0.77338	0.64729	0.99783	0.99525	0.99862
StdDev	0.25413	0.20257	0.30542	0.00403	0.00589	0.00197
Median	0.7804	0.83364	0.75828	0.9989	0.9967	0.99905

4 Conclusion

We observed that even in the medical image, the data which is generated from the original data is a good approach for segmentation. In this paper, we propose using Bit-plane to generate more image but still remains the significant features. Besides, we also implement the U-NET with multiple kernels to get the better performance. The result is evaluated without additional data and is shown with a promising performance.

Despite getting the good result, we need to process the boundary after the final segmentation as the post-processing. Next, when we implement the data generation, the combination with GAN is essential to get a better result

References

1. Y. Lecun, L. Bottou, Y. Bengio, P. Haffner: Gradient-based learning applied to document recognition. In: Proceedings of the IEEE, pp. 2278–2324, (1998), doi: 10.1109/5.726791
2. Evan Shelhamer, Jonathan Long, Trevor Darrell: Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 39, issue 4, 640-651, (2017), doi: 10.1109/CVPR.2015.7298965
3. Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on pattern analysis and machine intelligence, vol 39, issue 12, 2481-2495, (2017), doi: 10.1109/TPAMI.2016.2644615
4. Lu Guo, Ping Wang, Ranran Sun, Chengwen Yang, Ning Zhang, Yu Guo, Yuanming Feng: A fuzzy feature fusion method for auto-segmentation of gliomas with multi-modality diffusion and perfusion magnetic resonance images in radiotherapy. Scientific Reports volume 8, Article number: 3231, (2018)

5. Shukla G, Alexander GS, Bakas S, Nikam R, Talekar K, Palmer JD, Shi W. Advanced magnetic resonance imaging in glioblastoma: a review. *Chin Clin Oncol*. 2017 Aug;6(4):40. doi: 10.21037/cco.2017.06.28.
6. Bakas S, Zeng K, Sotiras A, Rathore S, Akbari H, Gaonkar B, Rozycki M, Pati S, Davatzikos C. GLISTRboost: Combining Multimodal MRI Segmentation, Registration, and Biophysical Tumor Growth Modeling with Gradient Boosting Machines for Glioma Segmentation. *Brainlesion* (2015). 2016;9556:144-155. doi: 10.1007/978-3-319-30858-6_1
7. R.C. Gonzalez, and R.E. Woods: 'Digital Image Processing'. Prentice Hall, Inc., New Jersey, 2002.
8. Ronneberger, O.; Fischer, P.; Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234-241. Springer, Cham. (2015)
9. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Γ, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
10. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
11. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
12. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF

Brain tumor segmentation with capsule networks versus fully convolutional neural networks

Chiatse J. Wang¹, Yuhsiang M. Tsai¹, Chengen Lee¹, Yuehchou Lee¹, Anthony Costa², Cheyu Hsu¹, Eric Oermann² and Weichung Wang¹

¹ National Taiwan University, Taipei, Taiwan

² Department of Neurosurgery, Mount Sinai Health System, 1468 Madison Avenue, New York, New York 10029, USA

Abstract. Fully convolutional neural network (CNN) based architectures have demonstrated state-of-the-art results in image segmentation across a variety of domains. The generalizability of these models to out of sample data, however, is unknown. In the case of magnetic resonance imaging (MRI) data, where acquisition protocols are heterogenous and results can vary significantly, the ability of models to robustly generalize is critical. Capsule networks might be a solution. In the present work, we compare the performance of Capsule Network based segmentation to a standard fully convolutional neural network (SegNet).

Keywords: Capsule, CNN, SegCaps, SegNet.

1 Introduction

Semantic segmentation is one of the fundamental tasks of computer vision. Since the development of Fully Convolutional Neural Networks (FCNN), FCNN based approaches have become a mainstay of competitive results on numerous segmentation challenges: MSCOCO, PASCAL-VOC, and BrATS. . Recently, Sabour et al. proposed a new type of network architecture, the capsule network, comprised of capsule layers rather than the standard convolution and maxpooling layers found within standard CNN architectures [5]. Capsule layers, due to their ability to store vector representations, in theory can account for translational and rotational variance in features without resorting to maxpooling which necessitates information loss from downsampling. In theory, capsule networks for semantic segmentation may be able to yield superior results due to their preservation of granular information as well as their ability to natively account for spatial variance. In the present paper, we explore and compare the use of capsule networks for performing semantic segmentation on the 2018 BraTS dataset to a standard FCNN implementation (SegNet).

2 Methods

2.1 Data

BraTS-2018 Dataset. The BraTS-2018 dataset is divided into two sets: training and validation. The training dataset includes 210 high grade glioma (HGG) cases and 75 low grade glioma (LGG) cases. All cases in training set include four modalities: T1, T1 contrast-enhanced, T2, FLAIR and segmentation label. In addition, segmentation labeled in four rate 1, 2, 4, 0 are necrotic/non-enhancing (NCR/NET), peritumoral edema (ED), enhancing tumor (ET) and otherwise, respectively. The union of the label 1, 2 and 4 and label 1 and 4 are the whole tumor (WT) and tumor core (ET), respectively. There are 66 cases without distinguish grading in validation dataset. Every cases include the same four modalities but without segmentation label.

Preprocessing. We stack the same slice in four modalities into a tensor and this tensor has four channels in order of T1, T1ce, T2 and FLAIR. The size of each input tensor is $240 \times 240 \times 4$. Before fitting our model, we convert label 4 to 3 then use one-hot encoding on the segmentation label.

2.2 Model Architecture

SegNet. The architecture is first proposed by Vija et al. [7]. SegNet is a deep convolution encoder-decoder neural network. In the encoder part, it's the topologically same as the VGG16 [] but removing the fully connected layer so that makes the SegNet smaller and easier to train. The key point of the Segnet is the decoder, its idea was inspired by an architecture of unsupervised feature learning.

We modified some part of the architecture as shown in Fig. 1. In our network, the encoder part consists 12 convolution layer. After every convolution layer, there are batch normalized and an element-wise rectified-linear non-linearity (ReLU). And do the max-pooling with a 2×2 window and stride 2 every three layers.

The decoder part is to upsample the input feature map by the memorized max-pooling which is corresponding to the encoder part. Thus, it also consists 12 convolution layer. Finally, connects a softmax layer at the end of the SegNet to get the probability of each label.

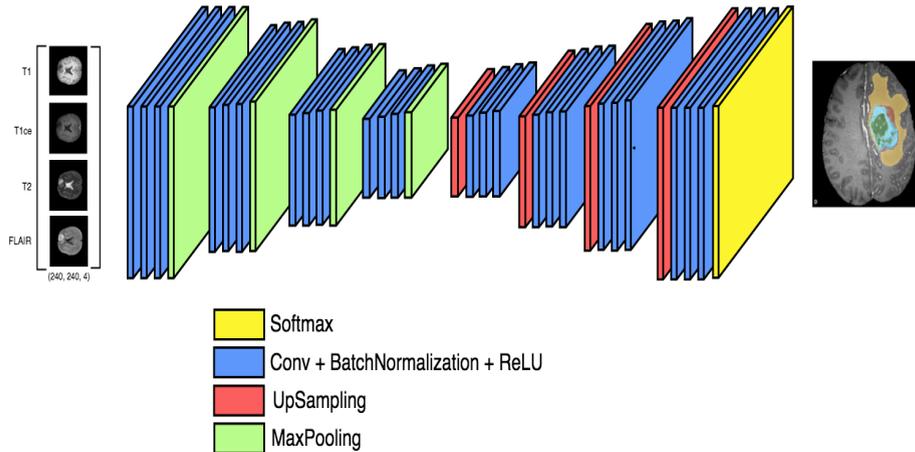


Fig. 1. SegNet architecture.

SegCaps. The architecture is first proposed by Rodney LaLonde and Ulas Bagci [6]. It is based on the U-Net [12] architecture and add the idea of the capsule network. SegCaps used less parameters than U-Net but have slightly improvement for lung segmentation on LUNA16 dataset [8-11]. We extend the idea to make this network capable to multi-label segmentation.

First, in the last two layers, we remove the limitation which only allows each pixel has one capsule. Thus, we increase the capability of SegCaps, to make the last layer to handle 4 capsule neurons and compute the length of each capsule neuron as our segmentation result.

Second, We choose the union of the label 1, 2, and 4 in BraTS 2018, as the reconstruction mask, that is, the net reconstruct the whole tumor as the regularization.

Third, we use the margin loss mentioned in (capsule net). we summarize the loss of each capsule on each pixel and average the loss of each pixel.

(Deconv) Capsule Layer. Because convolution capsule layer has more one dimension than traditional layer. Combine batch_size and capsules as batch_size, make the convolution function also works on this convolution capsule layer. And then apply capsule dynamic routing on this layer.

Loss Function. We use the weighted margin loss for segmentation. We calculate the number of four labels in the training data and use its inverse as our weight on the margin loss.

$$\text{weighted margin loss} = \text{mean}_{\text{pixel}}(\sum_{\text{capsule } i} \text{weight}_i (T_i \max(0, 0.9 - \|v_i\|)^2 + 0.5(1 - T_i) \max(0, \|v_i\| - 0.1)^2))$$

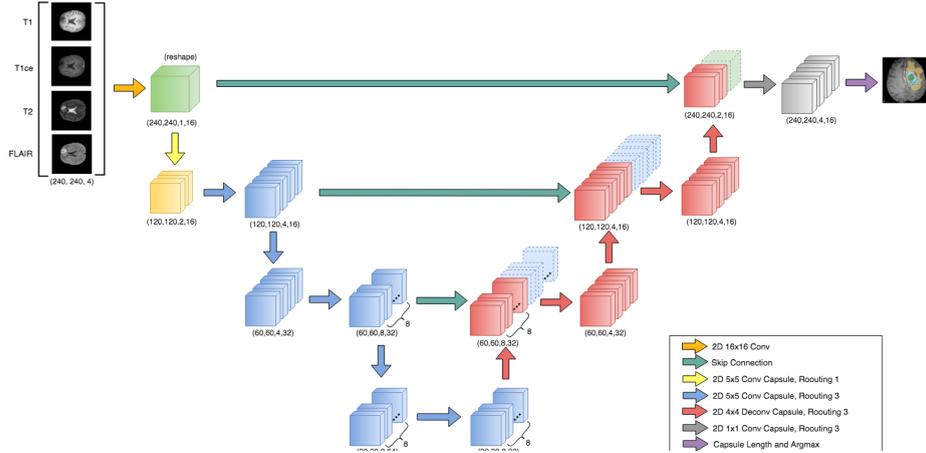


Fig. 2. SegCaps architecture.

2.3 Evaluation Metrics

There are three metrics used to evaluate the results.

Dice Score. The original formula is intended to be applied to binary data. Given two sets, X and Y , it is defined as

$$Dice\ Score = \frac{2|X \cap Y|}{|X| + |Y|}$$

where $|X|$ and $|Y|$ are the cardinalities of the two sets. When this formula is applied to our boolean data, it will be written as

$$Dice\ Score = \frac{2TP}{2TP + FP + FN}$$

where TP, FP and FN are true positive, false positive and false negative respectively.

Hausdorff 95 Distance. To see the distance between each voxel inside the original label image to the closest voxel inside the predictive label image, we use Hausdorff distance and a more robust percentile (95%) rather than the max to avoid issues with noisy segmentations.

Let X and Y be two non-empty subsets of a metric space (M, d) , then the Hausdorff distance $d_H(X, Y)$ is defined as

$$d_H(X, Y) = \max\{\sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y)\}$$

X, Y are the points on the boundary of the region. After computing all Hausdorff distance between true label and predictive label in ET, WT and TC, we take the 95 percentile of the results as Hausdorff 95 result.

Sensitivity and Specificity. Sensitivity is also called true positive rate and defined as

$$Sensitivity = \frac{TP}{TP+FN}$$

measures the proportion of actual positives that are correctly identified. Specificity is also called true negative rate and defined as

$$Specificity = \frac{TN}{FP+TN}$$

measures the proportion of actual negatives that are correctly identified.

3 Experiments and Results

We compared the models train by the HGG, LGG separately and the combination. Then, use it to evaluate the performance on HGG and LGG. Fig.

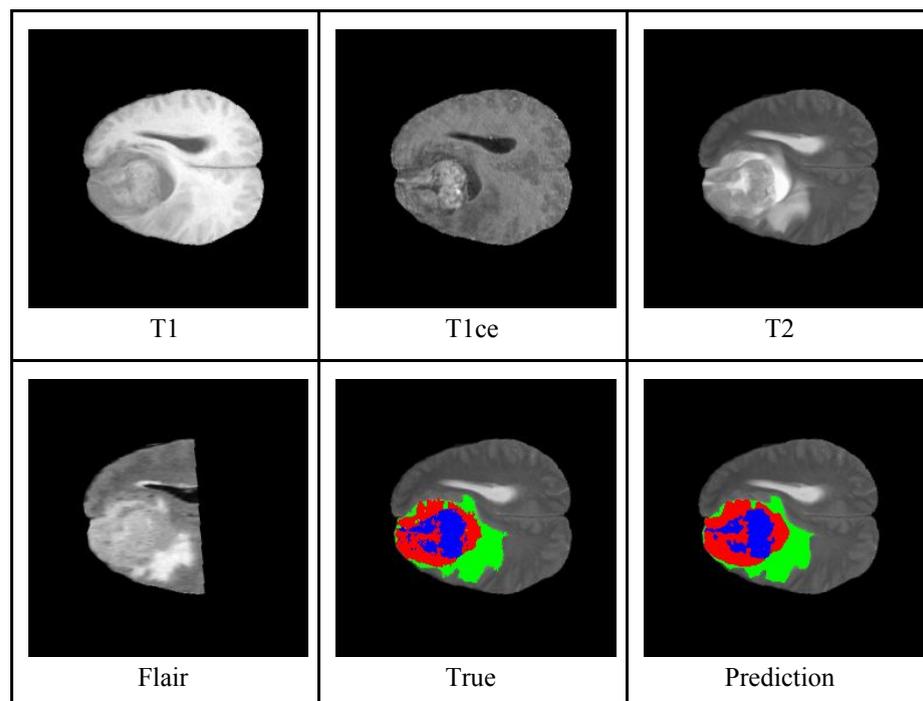


Fig. 3. The results of combination SegNet. (Red: label 1, green: label 2 and blue: label 3)

Table 1. Combination SegNet evaluations.

	Dice			Sensitivity			Specificity			Hausdorff		
	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET
Mean	0.56	0.84	0.74	0.64	0.85	0.75	1	0.99	1	14.8	20.7	20.1
StdDev	0.25	0.11	0.22	0.23	0.15	0.25	0	0.01	0	24.9	27.8	27.3
Median	0.66	0.87	0.83	0.70	0.90	0.86	1	0.99	1	3.74	6.24	7.6
25 quantile	0.48	0.84	0.67	0.6	0.84	0.67	0.99	0.99	0.99	3	3.39	3.81
75 quantile	0.73	0.9	0.89	0.77	0.93	0.92	1	1	1	10.6	21.2	23.9

Table 2. Dice Score Comparison.

Training Inputs	Type	SegNet Dice Score			SegCaps Dice Score		
		WT	TC	ET	WT	TC	ET
HGG	HGG	0.8135	0.7482	0.4949	0	0	0
	LGG	0.6439	0.4770	0.5466	0	0	0.36
LGG	HGG	0.6038	0.3415	0.0170	0.0017	0.0043	0.0051
	LGG	0.7517	0.7128	0.3721	0.0002	0.00002	0.3202
Combine	HGG	0.9660	0.9593	0.9032	6.3e-6	5.13e-6	0.024
	LGG	0.8309	0.6028	0.2755	0.0221	0.0159	0.0074

SegNet achieves excellent scores across the board for segmenting HGG and LGG with mean DICE scores over 0.9 for segmenting HGG (WT, TC, and ET) when trained on a combined dataset.

4 Discussion and Conclusion

In SegNet, the dice score of LGG and ET are always low and the model of LGG will lower the accuracy of the model of HGG. In SegCaps, the result is very terrible, and it need to spend a lot of time to train this model, that is why it's hard to get a descent results. However, it's interesting that the dice score of ET in HGG is higher than others in HGG.

5 References

1. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp

- Γ, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694.
2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117.
 3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q.
 4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF.
 5. Geoffrey E. H, Sara S, Nicholas F. "Dynamic Routing Between Capsules", *NIPS 2017*, arXiv:1710.09829, 2017.
 6. Rodney L, Ulas B, "Capsules for Object Segmentation", *MIDL 2018*.
 7. Vijay B, Alex K, Roberto C. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation", arXiv:1511.00561, 2016.
 8. K. Murphy, B. van Ginneken, A. M. R. Schilham, B. J. de Hoop, H. A. Gietema, and M. Prokop, "A large scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification," *Medical Image Analysis*, vol. 13, pp. 757–770, 2009.
 9. C. Jacobs, E. M. van Rikxoort, T. Twellmann, E. T. Scholten, P. A. de Jong, J. M. Kuhnigk, M. Oudkerk, H. J. de Koning, M. Prokop, C. Schaefer-Prokop, and B. van Ginneken, "Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images," *Medical Image Analysis*, vol. 18, pp. 374–384, 2014
 10. A. A. A. Setio, C. Jacobs, J. Gelderblom, and B. van Ginneken, "Automatic detection of large pulmonary solid nodules in thoracic CT images," *Medical Physics*, vol. 42, no. 10, pp. 5642–5653, 2015.
 11. E. M. van Rikxoort, B. de Hoop, M. A. Viergever, M. Prokop, and B. van Ginneken, "Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection", *Medical Physics*, vol. 4236 no. 10, pp. 2934-2947, 2009.
 12. Olaf Ronneberger, Philipp Fischer, Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv:1505.04597

6 Conclusion

6.1 A Subsection Sample

Please note that the first paragraph of a section or subsection is not indented. The first paragraphs that follows a table, figure, equation etc. does not have an indent, either.

Subsequent paragraphs, however, are indented.

Sample Heading (Third Level). Only two levels of headings should be numbered. Lower level headings remain unnumbered; they are formatted as run-in headings.

Sample Heading (Forth Level). The contribution should contain no more than four levels of headings. The following Table 1 gives a summary of all heading levels.

Table 1. Table captions should be placed above the tables.

Heading level	Example	Font size and style
Title (centered)	Lecture Notes	14 point, bold
1 st -level heading	1 Introduction	12 point, bold
2 nd -level heading	2.1 Printing Area	10 point, bold
3 rd -level heading	Run-in Heading in Bold. Text follows	10 point, bold
4 th -level heading	<i>Lowest Level Heading.</i> Text follows	10 point, italic

Displayed equations are centered and set on a separate line.

$$x + y = z \quad (1)$$

Please try to avoid rasterized images for line-art diagrams and schemas. Whenever possible, use vector graphics instead (see Fig. 1).

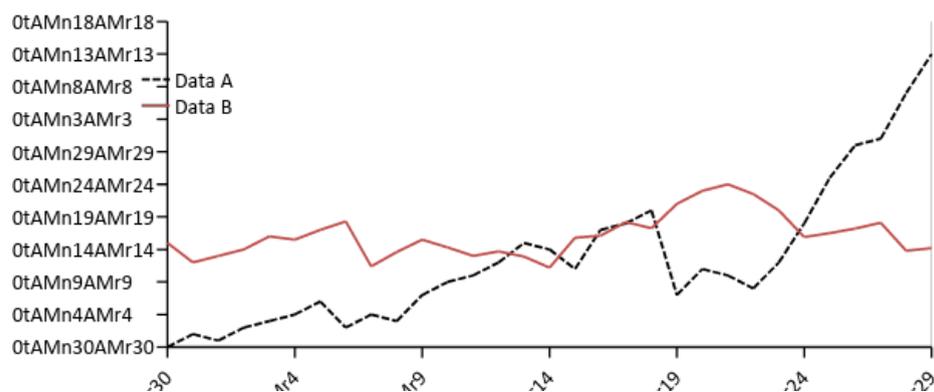


Fig. 1. A figure caption is always placed below the illustration. Short captions are centered, while long ones are justified. The macro button chooses the correct format automatically.

For citations of references, we prefer the use of square brackets and consecutive numbers. Citations using labels or the author/year convention are also acceptable. The following bibliography provides a sample reference list with entries for journal articles [1], an LNCS chapter [2], a book [3], proceedings without editors [4], as well as a URL [5].

Model	Data								
HGG	HGG			LGG			HGG+LGG		
	Dice	Spec	Sens	Dice	Spec	Sens	Dice	Spec	Sence
Complete	0.8135	0.7745	0.8939	0.6439	0.5843	0.8654	0.7689	0.7244	0.8864
Core	0.7482	0.7680	0.7764	0.4770	0.4901	0.6668	0.6768	0.6949	0.7476
Enhance	0.4949	0.5466	0.5064	0.0958	0.2778	0.0923	0.3899	0.4759	0.3974

TABLE 1: Model train with HGG data and predict HGG, LGG, and HGG+LGG data respectively.

Model	Data		
LGG	HGG	LGG	HGG+LGG

	Dice	Spec	Sence	Dice	Spec	Sence	Dice	Spec	Sence
Complete	0.6038	0.7041	0.5712	0.7517	0.9328	0.6754	0.6428	0.7643	0.5986
Core	0.3415	0.4140	0.3739	0.7128	0.8604	0.6712	0.4392	0.5315	0.4522
Enhance	0.0170	0.0090	0.3036	0.3721	0.3453	0.4402	0.1107	0.0979	0.3395

TABLE 2: Model train with LGG data and predict HGG, LGG, and HGG+LGG data respectively.

Model	Data								
HGG+LGG	HGG			LGG			HGG+LGG		
	Dice	Spec	Sence	Dice	Spec	Sence	Dice	Spec	Sence
Complete	0.9660	0.9780	0.9544	0.8309	0.8145	0.8804	0.9305	0.9350	0.9349
Core	0.9593	0.9748	0.9451	0.6028	0.5883	0.7495	0.8655	0.8731	0.8936
Enhance	0.9032	0.9333	0.8769	0.2755	0.3480	0.2719	0.7380	0.7793	0.7177

TABLE 3: Model train with HGG+LGG data and predict HGG, LGG, and HGG+LGG data respectively.

Automatic Brain Tumor Segmentation using Convolutional Neural Networks with Test-Time Augmentation

Guotai Wang^{1,2}, Wenqi Li^{1,2}, Sébastien Ourselin², and Tom Vercauteren^{1,2}

¹Wellcome / EPSRC Centre for Interventional and Surgical Sciences, University College London, London, UK

²School of Biomedical Engineering and Imaging Sciences, Kings College London, London, UK
guotai.wang@kcl.ac.uk

Abstract. Automatic brain tumor segmentation plays an important role for diagnosis, surgical planning and treatment assessment of brain tumors. Deep convolutional neural networks (CNNs) have been widely used for this task. Due to the relatively small data set for training, data augmentation at training time has been commonly used for better performance of CNNs. Recent works also demonstrated the usefulness of using augmentation at test time, in addition to training time, for achieving more robust predictions. We investigate how test-time augmentation can improve CNNs' performance for brain tumor segmentation. We used two underpinning network structures of 3D UNet 2.5D and WNet and augmented the image by 3D rotation, flipping, scaling and adding random noise at both training and test time. Experiments with BraTS 2018 training and validation set show that test-time augmentation helps to improve the brain tumor segmentation accuracy for both network structures.

Keywords: Brain tumor, convolutional neural network, segmentation, data augmentation

1 Introduction

With the development of medical imaging, brain tumors can be imaged by different Magnetic Resonance (MR) sequences, such as T1-weighted (T1w), contrast enhanced T1-weighted (T1wce), T2-weighted (T2w) and Fluid Attenuation Inversion Recovery (FLAIR) images. These sequences provide complementary information for different subregions of brain tumors [14]. Automatic segmentation of brain tumors and substructures from these images has a potential for more efficient and better diagnosis, surgical planning and treatment assessment of brain tumors [14,5].

In recent years, deep convolutional neural networks (CNN) have achieved the state-of-the-art performance for multi-modal brain tumor segmentation. In [8], a

convolutional neural network was proposed to exploit both local and global features for robust brain tumor segmentation. However, their approach works on individual 2D slices without considering 3D contextual information. DeepMedic [11] uses a dual pathway 3D CNN with 11 layers for brain tumor segmentation. It works on local image patches and therefore has a relatively lower inference efficiency. In [17], a 2.5D network was proposed for brain tumor segmentation and demonstrated an advantageous trade-off between receptive field, model complexity and memory consumption. In [10], an ensemble of multiple models and architectures was used for robust brain tumor segmentation.

Training with a large dataset plays an important role for the good performance of deep CNNs. For medical images, collecting a very large training set is usually time-consuming and challenging. Therefore, many works have used data augmentation to partially compensate this problem. Data augmentation applies transformations to the samples in a training set to create new ones, so that a relatively small training set can be enlarged to a larger one. Previous works have used different types of transformations such as flipping, cropping, rotation and scaling training images [2]. In [18], a simple and data-agnostic data augmentation routine termed *mixup* was proposed for training neural networks. Recently, several studies have empirically found that the performance of deep learning-based image recognition methods can be improved by combining predictions of multiple transformed versions of a test image. For example, in [16], a single model was used to predict multiple transformed copies of unlabeled images for data distillation. In [9], augmentation of the samples by rotation and translation was used for pulmonary nodule detection. In [13], test images were geometrically transformed for skin lesion classification. As far as we know, test-time augmentation has not been used for brain tumor segmentation from multi-modal MR images.

In this work, we apply test-time augmentation to automatic brain tumor segmentation. Instead of obtaining a single inference, we augment the input image with different transformation parameters to obtain multiple predictions from the input, with the same network and associated trained weights. The multiple predictions help to obtain more robust inference of a given image. In this work, we use the 3D UNet [2] and WNet [17] as the underpinning network structures. Experiments with BraTS 2018 training and validation set showed that an improvement of segmentation accuracy was achieved by test-time augmentation.

2 Methods

2.1 Network structures

We used two network structures as the underpinning CNN for the brain tumor segmentation task: 3D UNet [2] and 2.5D WNet [17]. The 3D U-Net has a downsampling and an upsampling path each with four resolution steps. In the downsampling path, each layer has two $3 \times 3 \times 3$ convolutions each followed by a rectified linear unit (ReLU) activation function, and then a $2 \times 2 \times 2$ max pooling

layer was used for downsampling. In the upsampling path, each layer uses a deconvolution with kernel size $2 \times 2 \times 2$, followed by two $3 \times 3 \times 3$ convolutions with ReLU. The network has shortcut connections between corresponding layers with the same resolution in the downsampling path and the upsampling path. In the last layer, a $1 \times 1 \times 1$ convolution is used to reduce the number of output channels to the number of segmentation labels, i.e., 4 for the brain tumor segmentation task in the BraTS challenge.

The WNet was proposed in [17]. This is an anisotropic network that considers a trade-off between receptive field, model complexity and memory consumption. It employs dilated convolution, residual connection and multi-scale prediction to improve segmentation performance. The network uses 20 intra-slice convolution layers and four inter-slice convolution layers with two 2D down-sampling layers. Since the anisotropic convolution has a small receptive field in the through-plane direction, multi-view fusion was used to take advantage of the 3D contextual information, where the network was applied in axial, sagittal and coronal views respectively. For the multi-view fusion, the softmax outputs in these three views were averaged. The original implementation of WNet in [17] used three cascaded networks to deal with the whole tumor, tumor core and enhancing tumor core respectively with binary segmentations. Compared with multi-label prediction, it requires longer time for training and testing. To improve the training efficiency, we explore the use of a single WNet for multi-label prediction in each view. Therefore, we change the output channel number of WNet from 2 to 4.

2.2 Augmentation for Training and Testing

From the point view of image acquisition, an observed image is only one of many possible observations of the underlying anatomy that can be observed with different spatial transformations and noise. Direct inference with the observed image may lead to a biased result affected by the specific transformation and noise associated with that image. To obtain a more robust prediction, we consider different transformations and noise during the test time. Let β and e represent the parameters for spatial transformation and intensity noise respectively. We assume that β is a combination of f_l , r and s , where f_l is a random variable for flipping along each 3D axis, r is the rotation angle along each 3D axis, s is a scaling factor. We consider these parameters following some distributions: $f_l \sim \text{Bern}(0.5)$, $r \sim U(0, 2\pi)$, $s \sim U(0.8, 1.2)$. For the intensity noise, we assume $e \sim N(0, 0.05)$ according to the reduced standard deviation of a median-filtered version of a normalized image.

For data augmentation, we randomly sample β and e from the above distributions and use them to transform the image. We use the same distributions of augmentation parameters at both training and test time for the 3D UNet and 2.5D WNet. For test-time augmentation, we obtain N samples from the distributions of β and e , and the resulting transformed version of the input was fed into the network. The N prediction results were combined to obtain the final prediction based on majority voting.

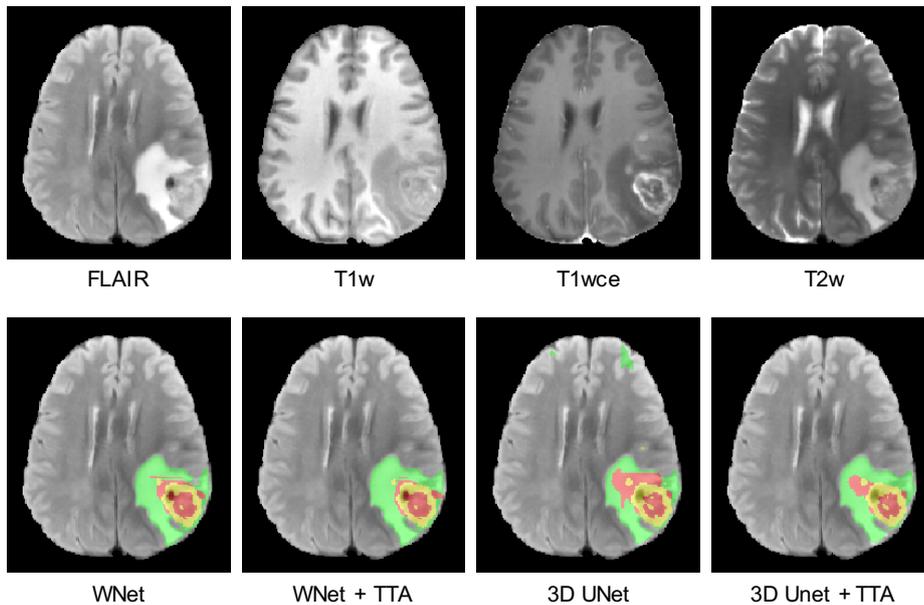


Fig. 1. Visual comparison between WNet, 3D UNet and the use of test-time augmentation (TTA). The first row shows the four modalities of the same patient. The second row shows segmentation results. Green: edema; Red: non-enhancing tumor core; Yellow: enhancing tumor core.

3 Experiments and Results

Data and Implementation Details. We used the BraTS 2018¹ [3,4,5,14] training and validation datasets for experiments. The training set contains images from 285 patients, including 210 cases of high grade glioma (HGG) and 75 cases of low grade glioma (LGG). The BraTS 2018 validation set contains images from 66 patients with brain tumors of unknown grade. Each patient was scanned with four sequences: T1w, T1wce, T2w and FLAIR. As a pre-processing performed by the organizers, all the images were skull-stripped and re-sampled to an isotropic 1mm^3 resolution, and the four modalities of the same patient had been co-registered. The ground truth were provided by the BraTS organizers. We uploaded the segmentation results obtained by our method to the BraTS 2018 server, and the server provided quantitative evaluations including Dice score and Hausdorff distance compared with the ground truth.

We implemented the 3D UNet and 2.5D WNet in Tensorflow² [1] using NiftyNet³ [7]. The Adaptive Moment Estimation (Adam) [12] strategy was used

¹ <http://www.med.upenn.edu/sbia/brats2018.html>

² <https://www.tensorflow.org>

³ <http://niftynet.io>

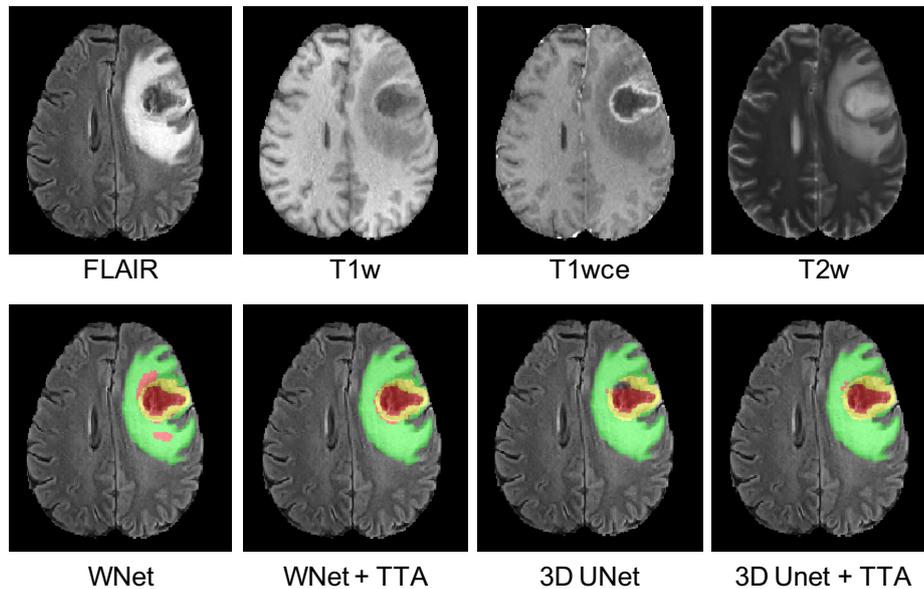


Fig. 2. Visual comparison between WNet, 3D UNet and the use of test-time augmentation (TTA). The first row shows the four modalities of the same patient. The second row shows segmentation results. Green: edema; Red: non-enhancing tumor core; Yellow: enhancing tumor core.

for training, with initial learning rate 10^{-3} , weight decay 10^{-7} , and maximal iteration 20k. The training patch size was $96 \times 96 \times 96$ for 3D UNet and $96 \times 96 \times 19$ for WNet. The batch size was 2 and 4 for these two networks respectively. The training process was implemented on an NVIDIA TITAN X GPU. As a pre-processing, each image was normalized by the mean value and standard deviation. The Dice loss function [15,6] was used for training. At test time, the augmented prediction number was set to $N = 20$ for both network structures. The WNet was trained in axial, sagittal and coronal views respectively, and the predictions in these three views were fused by averaging at test time.

Segmentation Results. Fig. 1 shows an example from the BraTS 2018 validation set. The first row shows the input images of four modalities: FLAIR, T1w, T1wce and T2w. The second row presents the segmentation results of WNet, UNet and their corresponding results with test-time augmentation. It can be observed that the output of WNet seems to be noisy for the non-enhancing tumor core. A smoother segmentation is obtained by WNet with test-time augmentation. The initial output of the 3D UNet also seems to be noisy with some false positives of edema and non-enhancing tumor core. After using test-time augmentation, the result becomes more spatially consistent.

Table 1. Mean values of Dice and Hausdorff measurements of different methods on BraTS 2018 validation set. EN, WT, TC denote enhancing tumor core, whole tumor and tumor core, respectively. TTA: test-time augmentation.

	Dice (%)			Hausdorff (mm)		
	ET	WT	TC	ET	WT	TC
WNet	75.70	88.98	72.53	4.24	4.99	12.13
WNet + TTA	77.07	89.56	73.04	4.44	4.92	11.13
3D UNet	73.44	86.38	76.58	9.37	12.00	10.37
3D UNet + TTA	75.43	87.31	78.32	4.53	5.90	8.03

Fig. 2 shows another example from the BraTS 2018 validation set. It can be observed that the initial prediction by WNet seems to have an over segmentation of the non-enhancing tumor core. After using test-time augmentation, the over-segmented regions become smaller, leading to a better accuracy. The 3D UNet obtains a hole in the tumor core, which seems to be an under-segmentation. The hole is filled after using test-time augmentation and the result looks more consistent with the input images.

A quantitative evaluation of our different methods is shown in Table 1. For the initial output of WNet, the Dice score was 75.70%, 88.98% and 72.53% for enhancing tumor core, whole tumor and tumor core respectively. After using test-time augmentation, an improvement was achieved, and the Dice score was 77.70%, 89.56% and 73.04% for these three targets respectively. The initial output of 3D UNet achieved Dice scores of 73.44%, 86.38% and 76.58% for these structures respectively. 3D UNet with test time augmentation achieved a better performance than the baseline of 3D UNet, leading to Dice scores of 75.43%, 87.31% and 78.32% respectively.

4 Discussion and Conclusion

For test-time augmentation, we only used flipping, rotation and scaling for spatial transformations. It is also possible to employ more complex transformations such as elastic deformations used in [2]. However, such deformations take longer time for testing and have less efficiency. The results show that test-time augmentation leads to an improvement of segmentation accuracy for both 3D UNet and 2.5 WNet, and it can be applied to other CNN models as well.

In conclusion, we explored the effect of test-time augmentation for CNN-based brain tumor segmentation. We used 3D U-Net and 2.5D W-Net as the underpinning network structures. For training and testing, we augmented the image by 3D rotation, flipping, scaling and adding random noise. Experiments with BraTS 2018 training and validation set show that test-time augmentation helps to improve the brain tumor segmentation accuracy for both 3D U-Net and 2.5D W-Net structures.

Acknowledgements. We would like to thank the NiftyNet team. This work was supported through an Innovative Engineering for Health award by the Wellcome Trust [WT101957], Engineering and Physical Sciences Research Council (EPSRC) [NS/A000027/1], the National Institute for Health Research University College London Hospitals Biomedical Research Centre (NIHR BRC UCLH/UCL High Impact Initiative), hardware donated by NVIDIA, and the Health Innovation Challenge Fund [HICF-T4-275, WT 97914], a parallel funding partnership between the Department of Health and Wellcome Trust.

References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X., Brain, G.: TensorFlow: A system for large-scale machine learning. In: USENIX Symposium on Operating Systems Design and Implementation. pp. 265–284 (2016)
2. Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 424–432 (2016)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. The Cancer Imaging Archive (2017), <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017), <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>
5. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Nature Scientific Data* p. 170117 (2017)
6. Fidon, L., Li, W., Garcia-peraza herrera, L.C.: Generalised Wasserstein Dice score for imbalanced multi-class segmentation using holistic convolutional networks. arXiv preprint arXiv:1707.00478 (2017)
7. Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D.I., Wang, G., Eaton-Rosen, Z., Gray, R., Doel, T., Hu, Y., Whyntie, T., Nachev, P., Modat, M., Barratt, D.C., Ourselin, S., Cardoso, M.J., Vercauteren, T.: NiftyNet: A deep-learning platform for medical imaging. *Computer Methods and Programs in Biomedicine* 158, 113–122 (2018)
8. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H.: Brain tumor segmentation with deep neural networks. *Medical Image Analysis* 35, 18–31 (2016)
9. Jin, H., Li, Z., Tong, R., Lin, L.: A deep 3D residual CNN for false positive reduction in pulmonary nodule detection. *Medical Physics* 45(5), 2097–2107 (2018)
10. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., Glocker, B.: Ensembles of multi-

- ple models and architectures for robust brain tumour segmentation. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 450–462. Springer International Publishing (2018)
11. Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis* 36, 61–78 (2017)
 12. Kingma, D.P., Ba, J.L.: Adam: A method for stochastic optimization. In: *ICLR* (2015)
 13. Matsunaga, K., Hamada, A., Minagawa, A., Koga, H.: Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble. *arXiv preprint arXiv:1703.03108* (2017), <http://arxiv.org/abs/1703.03108>
 14. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekeley, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging* 34(10), 1993–2024 (2015)
 15. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: *International Conference on 3D Vision*. pp. 565–571 (2016)
 16. Radosavovic, I., Dollár, P., Girshick, R., Gkioxari, G., He, K.: Data distillation: towards omni-supervised learning. *arXiv preprint arXiv:1712.04440* (2017), <http://arxiv.org/abs/1712.04440>
 17. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 178–190. Springer International Publishing (2018)
 18. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: Mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412* pp. 1–11 (2017), <http://arxiv.org/abs/1710.09412>

Segmentation of Brain Tumors in 3D-MRI Data and Patient Survival Prediction: Methods for the BraTS 2018 Challenge

Leon Weninger, Oliver Rippel, Simon Koppers, and Dorit Merhof

Institute of Imaging & Computer Vision
RWTH Aachen University
Aachen, Germany
{last name}@ifb.rwth-aachen.de

Abstract. Brain tumor localization and segmentation is an important step in the treatment of brain tumor patients. It is the base for later clinical steps, e.g., a possible resection of the tumor. Hence, an automatic segmentation algorithm would be preferable, as it does not suffer from inter-rater variability, and results could be available directly after the brain imaging procedure. Using this automatic tumor segmentation, it could also be possible to predict the survival of patients. The BraTS 2018 challenge consists of these two tasks: tumor segmentation in 3D-MRI images of brain tumor patients and survival prediction based on these images. For the tumor segmentation, we utilize a two-step approach: First, the tumor is located using a 3D U-net. Second, another 3D U-net – more complex, but with a smaller field-of-view – detects subtle differences in the tumor volume, i.e., it segments the located tumor into tumor core, enhanced tumor, and peritumoral edema. The survival prediction of the patients is done with a rather simple, yet accurate algorithm which outperforms other tested approaches.

Keywords: BraTS 2018, Brain Tumor, Automatic Segmentation, Survival Prediction, Deep Learning.

1 Introduction

Brain tumors can appear in different forms, shapes and sizes and can grow to a considerable size until they are discovered. They can be distinguished into glioblastoma (GBM/HGG) and low grade glioma (LGG). A common way of screening for brain tumor is with MRI-scans, where even different brain tumor regions can be determined. In effect, MRI scans of the brain are not only the basis for tumor screening, but are even utilized for pre-operative planning. Thus, an accurate, fast and reproducible segmentation of brain tumors in MRI scans is needed for several clinical applications.

HGG patients have a poor survival prognosis, as metastases often develop even when the initial tumor was completely resected. Whether patient overall

survival can be accurately predicted from pre-operative scans, i.e., knowing factors such as radiomics features, tumor location and tumor shape, remains an open question.

The BraTS challenge [6] addresses these problems, and is one of the biggest and best-known machine learning challenges in the field of medical imaging. Last year around 50 different competitors from around the world took part. The challenge is divided in two parts: First, tumor segmentation based on 3D-MRI images, and second, survival prediction of the brain tumor patients based on only the pre-operative scans and the age of the patients.

Similar to the BraTS 2017 dataset, the BraTS 2018 training dataset consists of images of 285 brain tumor patients from 19 different contributors. The dataset includes T1, T1 post-contrast, T2, and T2 Fluid Attenuated Inversion Recovery (Flair) volumes, as well as hand-annotated expert labels for each patient [3] [2] [1].

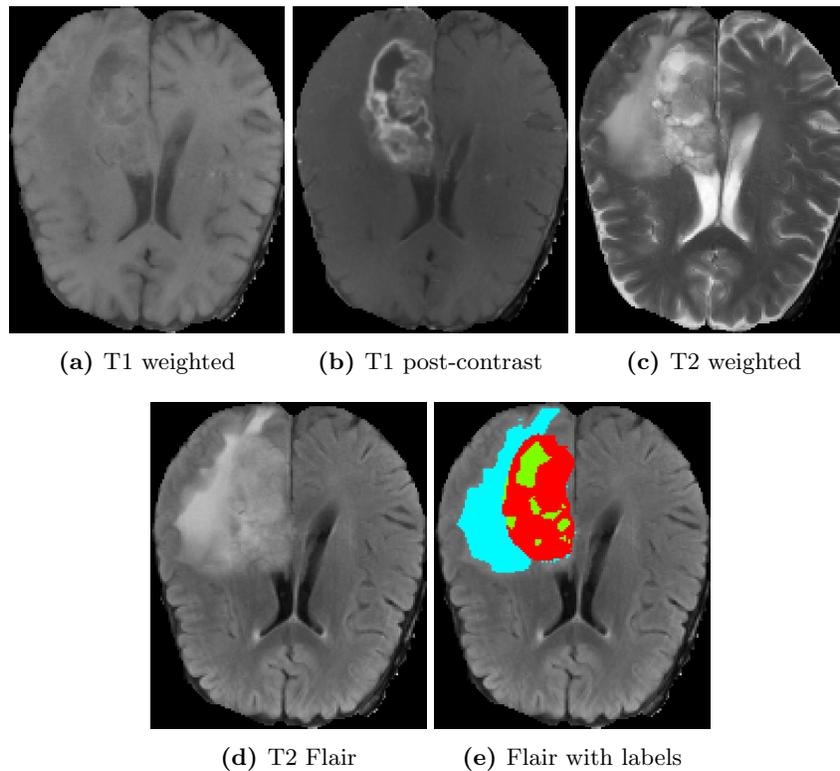


Fig. 1: Different MRI-modalities and groundtruth-labels in the BraTS 2018 dataset. Blue indicates the peritumoral edema, green the necrotic and non-enhancing tumor, and red the GD-enhancing core, as described in the BraTS paper [6].

Motivated by the success of the U-net [9] in biomedical image segmentation, we choose the 3D-adaptation [4] of this network to tackle the segmentation part of the BraTS challenge. Two different versions of this architecture are used, a first one to coarsely locate the tumor, and a second one to accurately segment the located tumor into different areas.

Concerning the survival prediction, we found that complex models using different types of radiomics features such as shape and texture of the tumor and the brain could not outperform a simple linear regressor based on just a few basic features. Using only the patient age and tumor region sizes as features, we achieve competitive results.

The code developed for this challenge will be available online after the final deadline: <https://github.com/weningerleon/BraTS2018>

2 Methods

2.1 Segmentation

We tackle the segmentation task in a two-step approach: First, the location of the brain tumor is determined. Second, this region is segmented into the three different annotations: *peritumoral edema (ed)*, *necrotic tumor (nec)*, and *GD-enhancing core (gde)*.

Preprocessing We first define a brain mask based on all voxels unequal to zero, on which all preprocessing is carried out. Before we calculate the mean and standard deviation of the brain, we clip the values of intensities at 2.5% and 97.5%. Using this preprocessing technique, very high and very low values, often occurring due to imaging artifacts, have less influence on the mean and standard deviation. Since different MRI-scanners and sequences are used, we independently normalize each image and modality based on the obtained values. Non-brain regions remain zero.

The whole tumor is strongly visible in T1, T2 and Flair MRI-images, so we discard all other MRI-modalities for the tumor localization step. We construct a cuboid bounding box around the brain, and crop the non-brain regions to facilitate training. The training target is constructed by merging the three different tumor classes of the groundtruth labels.

For training of the tumor segmentation step, the 3D-images are cropped around a padded tumor area, which is defined as the area of 20 voxels in every direction around the tumor.

Network architectures and employed hardware For both steps, a 3D U-net [4] with a depth of 4 is employed.

The first U-net uses padding in every convolutional layer, such that the input size corresponds exactly to the output size. Every convolutional layer is followed by a ReLU activation function. 16 feature maps are used in the first layer, and the number of feature maps doubles as the depth increases. For normalization

between the different layers, instance-norm layers [10] are used, as they seem to be better suited for normalization in segmentation tasks and for small batch sizes. Testing different training hyperparameters, the Adam optimizer [5] with an initial learning rate of 0.001 together with a binary cross entropy loss was chosen for the tumor localization step. An L2-regularization of $1e-5$ is applied to the weights, and the learning rate was reduced by a factor of 0.015 after every epoch. One epoch denotes a training step over every brain.

The U-net utilized in the second step has a similar architecture as the previous one, but with double as many feature maps per layer. To counteract the increased memory usage, no padding is used, which drastically reduces the size of the output as well as the memory consumption of later feature maps.

Here, we apply a multiclass dice loss to the output of our 3D U-net and the labels for training, as described in [7]. A learning rate of 0.005 was chosen, while weight decay and learning rate reduction remain the same as in step 1.

Our contribution to the BraTS challenge was implemented using pyTorch [8]. Training and prediction is carried out on a Nvidia 1080 Ti GPU with a memory size of 11 Gb.

Training In the first step, we train with complete brain images cropped to the brain mask. The brain mask is determined by all voxels not equal to zero. Using a rather simple U-net, a training pass with a batch-size of one fits on a GPU even for larger brains. Due to the bounding box around the brain, different sizes need to be passed through the network. In practice this is possible using a fully convolutional network architecture and a batch size of one.

For the second step, we choose the input to be fixed to $124 \times 124 \times 124$. Due to the unpadding convolutions, this results in an output shape of $36 \times 36 \times 36$. Hence, the training labels are the $36 \times 36 \times 36$ sized segmented voxels in the middle of the input. Here, a batch-size of two was chosen.

During training, 25 such patches are chosen at random from inside the padded tumor bounding box for each patient. Having 286 training datasets, this gives us 7125 training patches.

Inference Similar to the training procedure, the first step is carried out directly on a complete 3-channel (T1, T2, Flair) 3D image of the brain.

Before the tumor / non-tumor segmentation of this step is used as basis in the second step, only the largest connected area is kept. Based on the assumption that there is only one tumorous area in the brain, we can suppress false positive voxels in the rest of the brain with this method.

We then predict $36 \times 36 \times 36$ sized patches with the trained unpadding U-net. Patches are chosen so that they cover the tumorous area, the distance between two neighboring patches was set to 9 in each direction. Several predictions per voxel result. Accordingly, a majority vote over these predictions gives the final result.

2.2 Survival Prediction

According to the information given by the segmentation labels, we count the number of voxels of the tumor segmentation. This volume information about the necrotic tumor core, the GD-enhancing tumor and peritumoral edema as well as the distance between the centroids of tumor and brain and the age of the patient are used as input for a linear regressor.

3 Results

3.1 Segmentation

For evaluation on the training dataset, we split the training dataset randomly into 245 training images and 40 test images to evaluate our approach with ground truth labels. No external data was used for training or pre-training.

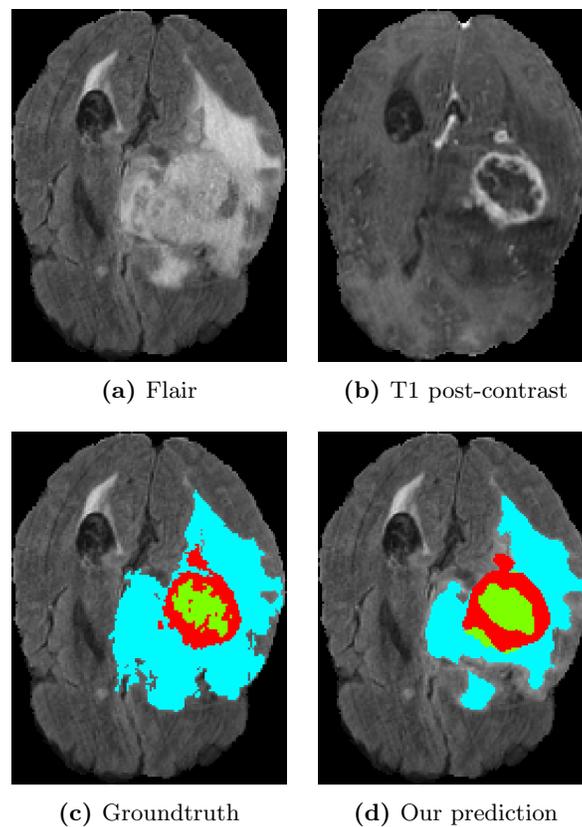


Fig. 2: Comparison of our segmentation result with the groundtruth labels.

Based on our experience with the training dataset, we choose 200 epochs as an appropriate training duration for the first step, and 60 epochs as an appropriate training duration for the second step. We thus train from scratch on all training images for the determined optimal number of epochs, and use the obtained networks for evaluation on the validation set. The results obtained by this method can be seen in Table 1.

Dataset	Dice			Sensitivity			Specificity			Hausdorff Dist.		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Train set	0.763	0.860	0.817	0.747	0.784	0.787	0.998	0.998	0.998	5.63	7.01	7.88
Val set	0.712	0.889	0.758	0.757	0.887	0.735	0.998	0.995	0.998	6.28	6.97	10.91

Table 1: Results for the segmentation challenge. Training and test set errors according to the online submission system.

3.2 Survival Prediction

For evaluating our approach on the training dataset, we fit and evaluate our linear regressor with a 50-fold cross-validation on the training images. We compare the results obtained by solely using the age of the patient versus using the age with a subset of the tumor region sizes as features. On top, we consider the distance between the centroid of the tumor and the centroid of the brain as a feature. Our finding is that all features other than the age of the patient increase the error on left-out images. In Tables 2 and 3, we show the exact results for the different input features.

Features	MSE	Median Err.
Age	87089	206
Age + gde	93599	215
Age + ed	91767	212
Age + nec	92320	207
Age + dist	95070	207
Age+gde+ed+nec	98053	222

Table 2: Training Data: Mean Squared Error and Median Error for 50-fold cross-validation of the linear regressor. The different features considered are the age of the patient, the volume in voxels of the enhancing tumor (*gde*), of the necrotic tumor (*nec*), of the edema (*ed*) as well as the distance between the centroid of the tumor and the centroid of the brain (*dist*).

Features	Accuracy	MSE	Median SE	stdSE	SpearmanR
Age	0.5	97759.5	46120.5	139670.7	0.267
Age+gde+ed+nec	0.536	101012.0	51006.5	140511.5	0.258

Table 3: Validation Data: Accuracy metrics according to the online portal.

4 Discussion & Conclusion

Our contribution submitted to the BraTS challenge 2018 was summarized in this paper. We used a two-step approach for tumor segmentation, which already gives promising results. In the near future we will evaluate a broader variety of different network architectures, and will also include 3D data-augmentation techniques into our framework.

Our algorithm for the survival analysis task is a straight-forward approach. We considered other, more complex approaches, which were however not able to beat this baseline algorithm.

So far, our survival prediction algorithm ranks among the top submissions, e.g., the age-only approach achieves the lowest MSE and second highest accuracy on the validation set. From these observations, it can be concluded that pre-operative scans are not well suited for survival prediction. Here, other datasets could be better suited for survival prediction, e.g., post-operative or follow-up scans of the patient.

References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freyman, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lyg collection. The Cancer Imaging Archive (2017)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. The Cancer Imaging Archive (2017)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific Data* **4**, 170117 EP – (09 2017), <http://dx.doi.org/10.1038/sdata.2017.117>
4. Çiçek, Ö., Abdulkadir, A., Lienkamp, S., Brox, T., Ronneberger, O.: 3d u-net: Learning dense volumetric segmentation from sparse annotation. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. LNCS, vol. 9901, pp. 424–432. Springer (Oct 2016), <http://lmb.informatik.uni-freiburg.de/Publications/2016/CABR16>, (available on arXiv:1606.06650 [cs.CV])
5. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *CoRR* **abs/1412.6980** (2014), <http://arxiv.org/abs/1412.6980>
6. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, ., Durst, C.R., Dojat,

- M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging* **34**(10), 1993–2024 (Oct 2015). <https://doi.org/10.1109/TMI.2014.2377694>
7. Milletari, F., Navab, N., Ahmadi, S.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. *CoRR* **abs/1606.04797** (2016), <http://arxiv.org/abs/1606.04797>
 8. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch. In: *NIPS-W* (2017)
 9. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. pp. 234–241. Springer International Publishing, Cham (2015)
 10. Ulyanov, D., Vedaldi, A., Lempitsky, V.S.: Instance normalization: The missing ingredient for fast stylization. *CoRR* **abs/1607.08022** (2016), <http://arxiv.org/abs/1607.08022>

Multimodal Brain Tumor Segmentation Using U-Net

Shaocheng Wu¹, Hongyang Li¹ and Yuanfang Guan¹

¹ University of Michigan, Ann Arbor, MI, 48109, USA
gyuanfan@umich.com

Abstract. Accurate segmentation of gliomas on routine magnetic resonance image (MRI) scans plays an important role in disease diagnosis, prognosis, and treatment decision. The varying intensity and shape of the pathology of glioma pose great challenges to automated segmentation. Here, we present our methods for brain tumor segmentation in Multimodal Brain Tumor Segmentation Challenge (BraTS) in 2018. A U-Net based segmentation was achieved by training a 10-layer deep convolutional neural network (CNN) on 2-D slices extracted from patient volumes. The network was trained on slices extracted from 210 HGG and 75 LGG patients and validated on 50 patients. On the BraTS 2018 validation set, the proposed algorithm achieved mean dice coefficients of 0.9119, 0.8307, and 0.7976 for whole tumor, tumor core and enhancing tumor respectively. Our proposed method ranked first in segmenting the whole tumor, and second in segmenting the enhancing tumor, as of July 9th.

Keywords: Brain Glioma, MRI, U-Net, Deeping Learning.

1 Introduction

Brain Magnetic Resonance Imaging (MRI) is the most common imaging diagnostic technique in the clinical diagnosis of gliomas [1]. In current clinical practice, visual assessment of brain glioma in MRI scans by radiologists is challenging due to its experience-dependent nature with significant inter-person variation [2] and the varying intensity, shape, and location of brain glioma [3]. Many computational approaches trying to automate this process have been developed but they have not reached the expected performance to assist radiologists' interpretation and identification of subtle changes of brain glioma [4–7]. Recent availability of public datasets and online benchmarks [8] provide opportunities for researchers to focus on optimizing model architectures to achieve high segmentation performance.

Here we present our computational model for the detection of brain glioma in MRI scans. It is based on a 2D end-to-end symmetric U-Net architecture, which outputs and visualizes the segmentation of 3 subregions (enhancing tumor, tumor core, and the whole tumor) of brain glioma simultaneously. The model is trained and validated using

the BraTS 2018 dataset [8-11]. It achieves Mean Dice Coefficients of 0.9123, 0.8274, and 0.7912 for whole tumor, tumor core and enhancing tumor respectively. This result indicates an excellent statistical performance score.

Our U-Net-based model consistently segments subregions of brain glioma with high accuracy and reliability with screening images from a large population, contributing to our top performing algorithm in the BraTS 2018.

2 Methods

2.1 Preprocessing

Because BraTS 2018 training dataset contains only the axial slices with the shape of 240 by 240, we then generated slices from other 2 planes (coronal and sagittal) and they were zero-padded into 240 by 240 as the original axial slices. For each modality (channel) of MRI scans (T2, T2-FLAIR, T1 and T1ce), we calculated the means and standard deviations of pixel array for each patient and normalized it through subtracting the mean and then divided by the standard deviations. All normalized slices from 3 planes were treated as inputs into the model (Fig. 1).

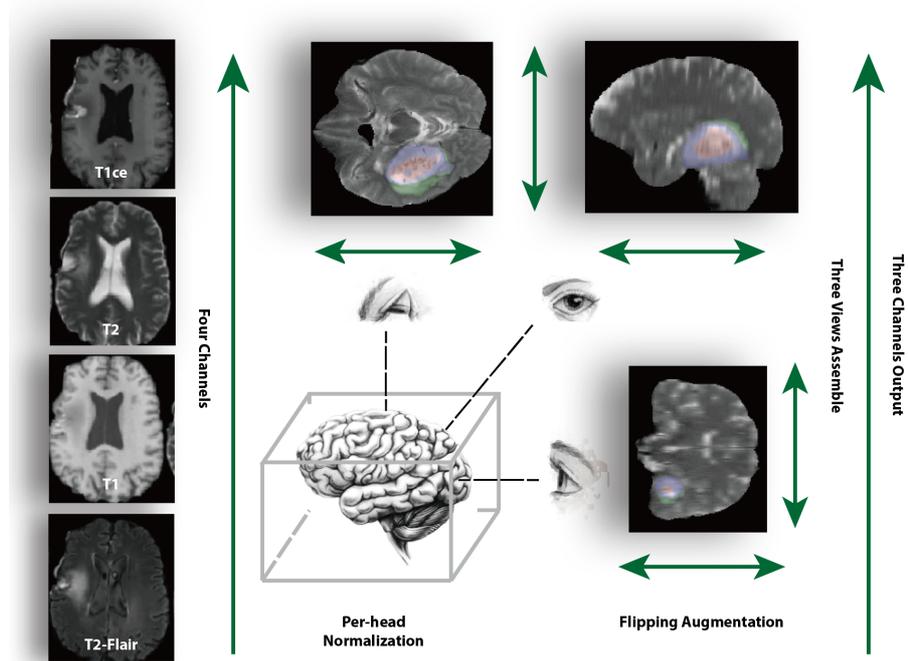


Fig. 1. Slices preprocessing. Slices of four channels (T1ce, T1, T2, and T2-Flair) and from three views (axial, coronal, and sagittal) are input into the model simultaneously.

2.2 Data Augmentation

To avoid overfitting, we deployed two types of data augmentation. First, we will flip a slice in either left/right or up/down direction at a random chance and input it as a new slice. The label for that slice is also processed in the same manner. Second, to address the class imbalance, we oversample slices with positive labels to balance positive/negative ratio.

2.3 Network Architecture

We use typical U-Net architecture [10] (Fig. 2), which takes in 2D image arrays with four channels (T2-FLAIR, T2, T1 and T1ce). Then the inputs pass through a series of convolutional and pooling layers and turn into feature maps with smaller size. The resulting feature maps then pass through a series of up-convolutional and concatenating layers and output a segmentation mask with 4 label channels simultaneously (for 4 sub-regions: 1 for necrotic (NCR) and the non-enhancing (NET) tumor core; 2 for edema; 4 for enhancing tumor; all 1, 2 and 4 for the whole tumor).

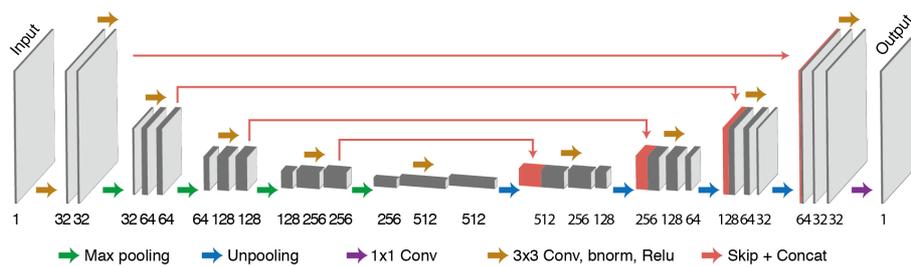


Fig. 2. U-Net Architecture.

2.4 Network Training

We used slices of all 4 channels and 3 planes from both HGG and LGG patient as inputs. Our model is trained with randomly sampled slices of size 240 by 240 with batch size 16 and for a total of 10 epoches. Training is done using the ADAM optimizer with an initial learning rate $l_{r_{init}} = 3e-5$. We conduct 5-fold nested training to choose the best epoch during our training and the loss is assessed by dice coefficient.

2.5 Evaluation

We evaluated our algorithm on the validation dataset in the 2018 BraTS and the performance was assessed by dice score [11].

3 Results

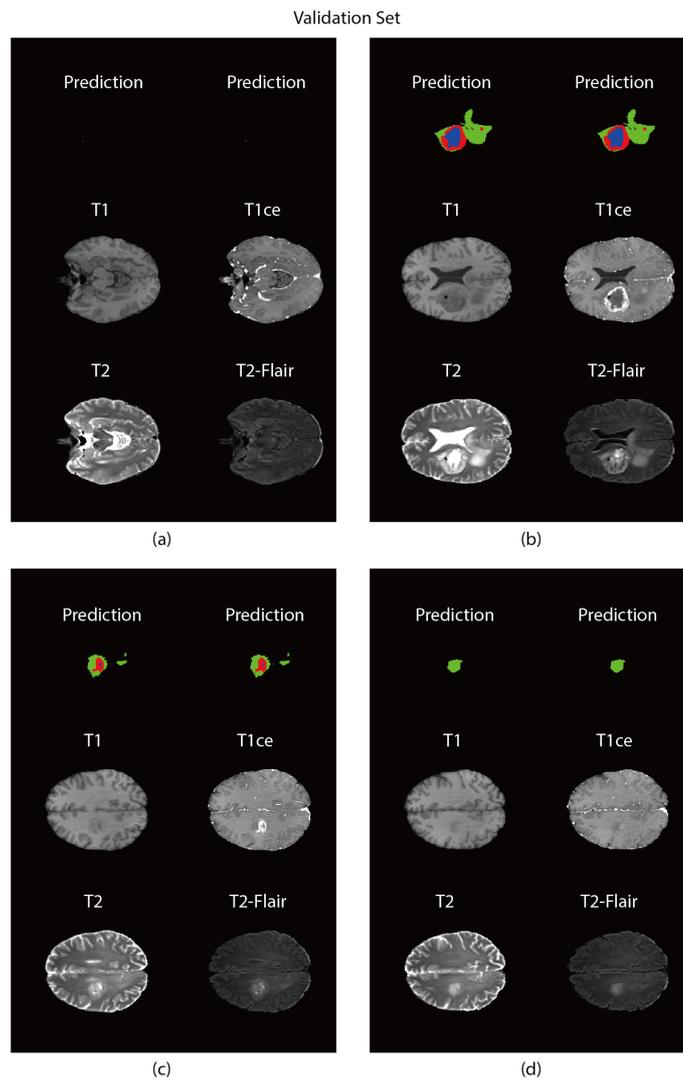


Fig. 3. Prediction examples of validation set.

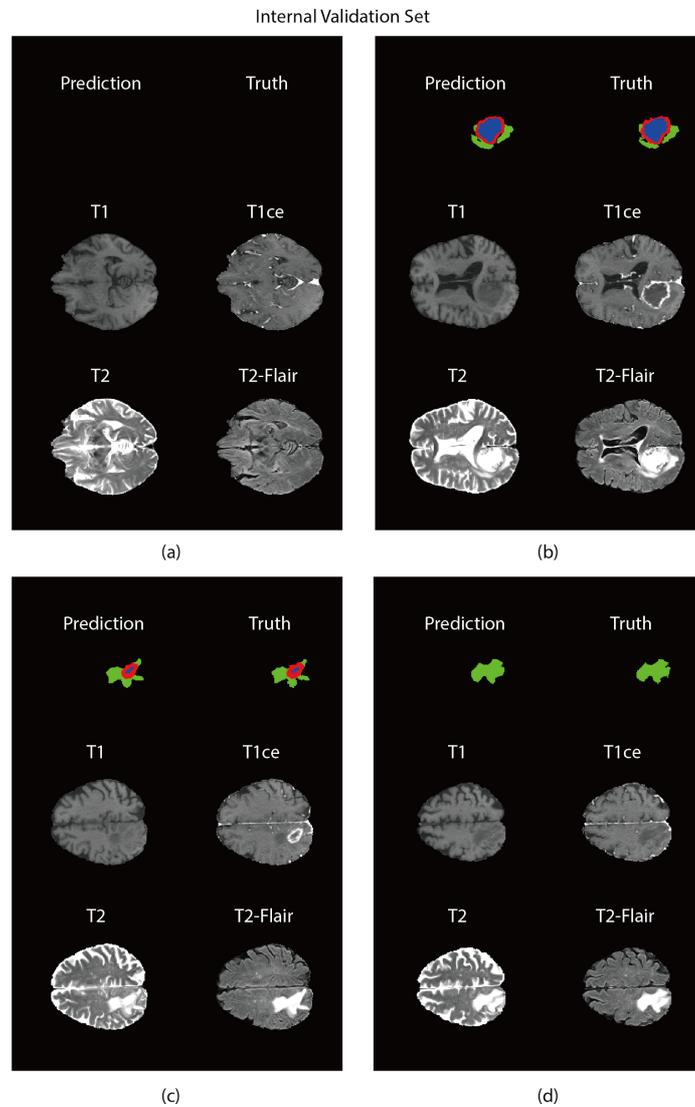


Fig. 4. Prediction examples of internal validation set.

For the training phase in internal training cases, our whole tumor, tumor core, and enhancing tumor segmentation achieved a mean dice coefficients of 0.9058, 0.8111, and 0.7685, respectively (Table 1). We submitted segmentations for all available cases in both training and validation.

Prediction examples of validation set and internal validation set are shown in Fig. 3 and Fig. 4. On the BraTS 2018 validation set, our proposed model achieved a mean

dice coefficients of 0.9119, 0.8307, and 0.7976 for whole tumor, tumor core and enhancing tumor respectively (Table 1).

Table 1. Performance of our model on the BraTS 2018 validation set.

	Whole Tumor	Tumor Core	Enhancing Tumor
Mean Dice Coefficient (Last year best performance)	0.8970	0.8254	0.7640
Mean Dice Coefficient (Our training performance)	0.9058	0.8111	0.7685
Mean Dice Coefficient (Our validation performance)	0.9119	0.8307	0.7976

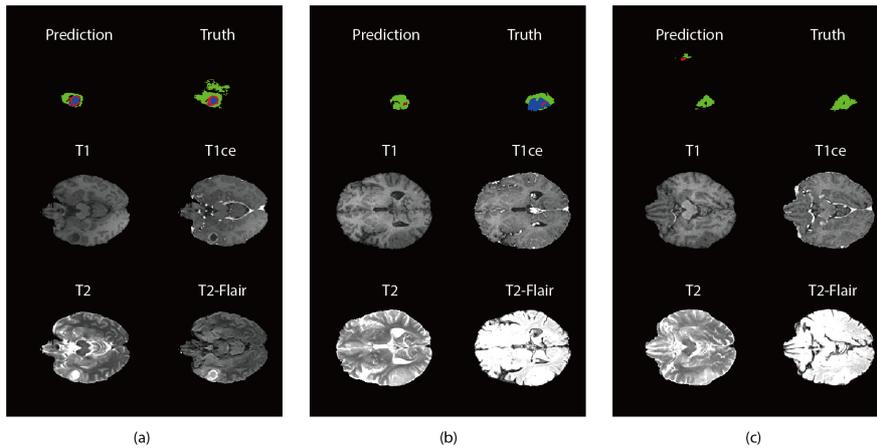


Fig. 5. Prediction examples with low dice coefficients (a, b, c).

Discussion

In the 2018 BraTS Competition, we proposed our brain tumor segmentation model based on 2D U-Net and it achieved first place in segmenting the whole tumor, and second in segmenting the enhancing tumor, as of July 9th.

While implementing a 3D U-Net could potentially improve the segmentation performance [12], the 3D approach faces a major challenge, the limitation of current GPU memory. This restrains the depth of U-Net architecture, the patch size of training samples, and the speed of training process, which may result in loss of information, overfitting, and inefficiency [13]. On the contrary, 2D U-Nets are much efficient to train with many training samples and can achieve the accuracy similar to experienced radiologists on brain tumor MRI scans.

Our model achieved high accuracy on most of training and validation cases but there are still some difficulties (Figure 4). As for Figure 4a, it's possible that the tumor size of this case is so larger than the others that our model failed to predict. For Figure 4b, the enhancing tumor region (labeled blue) is hardly distinguishable in T1ce and therefore, a more precise ground-truth label is needed to achieve high accuracy in this case. For Figure 4c, some parts of predictions are false positive and we plan to utilize connected component as post processing to reduce these.

Conclusions

Our proposed model demonstrated its potential and scope in segmenting multimodalities of brain tumor and it can be of immediate use to neuro-oncologists or radiologists, improving the diagnosis of brain tumor in terms of not only speed but also accuracy.

References

1. Ohgaki, H., Kleihues, P.: Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. *J. Neuropathol. Exp. Neurol.* 64, 479–489 (2005).
2. Huang, M., Yang, W., Wu, Y., Jiang, J., Chen, W., Feng, Q.: Brain tumor segmentation based on local independent projection-based classification. *IEEE Trans. Biomed. Eng.* 61, 2633–2645 (2014).
3. Sachdeva, J., Kumar, V., Gupta, I., Khandelwal, N., Ahuja, C.K.: Segmentation, feature extraction, and multiclass brain tumor classification. *J. Digit. Imaging.* 26, 1141–1150 (2013).
4. Phillips, W.E., 2nd, Velthuizen, R.P., Phuphanich, S., Hall, L.O., Clarke, L.P., Silbiger, M.L.: Application of fuzzy c-means segmentation technique for tissue differentiation in MR images of a hemorrhagic glioblastoma multiforme. *Magn. Reson. Imaging.* 13, 277–290 (1995).
5. Mehnert, A., Jackway, P.: An improved seeded region growing algorithm. *Pattern Recognit. Lett.* 18, 1065–1071 (1997).
6. Beare, R.: Efficient Implementation of The locally Constrained Watershed Transform and Seeded Region Growing. In: *Computational Imaging and Vision.* pp. 217–226.
7. Adams, R., Bischof, L.: Seeded region growing. *IEEE Trans. Pattern Anal. Mach. Intell.* 16, 641–647 (1994).

8. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.-A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.-C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging*. 34, 1993–2024 (2015).
9. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci Data*. 4, 170117 (2017).
10. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017.
11. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", *The Cancer Imaging Archive*, 2017.
12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Lecture Notes in Computer Science*. pp. 234–241 (2015).
13. Davis, J., Goadrich, M.: The relationship between Precision-Recall and ROC curves. In: *Proceedings of the 23rd international conference on Machine learning - ICML '06* (2006).
14. Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* 36, 61–78 (2017).
15. Simi, V.R., Joseph, J.: Segmentation of Glioblastoma Multiforme from MR Images – A comprehensive review. *The Egyptian Journal of Radiology and Nuclear Medicine*. 46, 1105–1110 (2015).

A two-step cascaded strategy for automatic brain tumor segmentation in MICCAI 2018 BraTS Challenge

Peiyuan Xu¹, Yifan Hu¹, Kai Ma¹, Yefeng Zheng¹

¹ Tencent Youtu Lab, Shenzhen 518057, China
ivanyfhu@tencent.com

Abstract. In this paper, we propose a fully automatic multi-tumor segmentation (whole tumor, tumor core and enhanced tumor core) based on multimodal MR images. Two stages are included in the strategy: detection of the approximate tumor position and segmentation via a cascaded and ensemble strategy. In the first step, a detection algorithm developed by [1] is exploited to find the rough position of all the tumors. In the second step, we divide the 4-class segmentation task into 3 binary segmentation tasks due to the different properties of 3 tumor regions. For each task, an ensemble of different 2D, 2.5D and 3D convolutional neural nets with dense block is trained, individually. The results of average Dice scores of 0.903, 0.851, 0.789 for whole tumor (WT), tumor core (TC) and enhanced tumor core (ET) are achieved on BraTS 2018 validation set with this strategy.

Keywords: Brain tumor, segmentation, cascade, ensemble, U-net, dense block.

1 Introduction

Since gliomas are the most common primary brain malignancies [9], segmentation of brain tumors in image scans is one of the most important and challenging tasks in medical image analysis. As the medical imaging developed, a lot of Magnetic Resonance modalities can be exploited to scan brain tumors, such as T1, T1ce, T2 and Fluid Attenuation Inversion Recovery (FLAIR). For different tumors, different modalities may provide complementary information.

A good end-to-end segmentation strategy can give us more accurate, quick and consistent measures of different tumors than human, which can be further utilized on the diagnosis and treatment for the brain tumors [2-3]. In the recent years, semantic segmentation algorithms have developed quickly and achieved a high performance in natural image tasks, as well as some applications in medical image segmentation tasks [4, 10-12]. The key ideas of the innovative literatures include the cascaded strategy for different tumors, maintaining multi-scale features with dense connection, eliminating the training imbalance error by using weighted Dice loss and focal loss and using ensemble strategy to make the model more robust.

Inspired by the ideas above, we propose a two-step cascade CNN strategy for this multimodal MRI brain segmentation task. All the advantages of dense connection, new loss strategy [6-8] and multi-dimensional model ensemble will be included in our model. Moreover, an ensemble model from 3 directions (for 2D and 2.5D CNN models) is also designed in this paper in order to consider all the information from different resolutions corresponding to each direction, respectively.

2 Methods

The framework of the proposed algorithm is shown in Fig 1. There are two stages in total: 1. We use a fast algorithm to detect the position of the tumors; 2. A cascade ensemble method is designed to segment the tumors.



Fig. 1. Two stages of our segmentation method. We firstly detect the region of the whole tumor. The detected region will be served as the input of the segmentation task thereafter.

2.1 Detection of the Tumor and Self-Adaptive Patch Size

Segmenting the whole image directly may not be a good idea because of the following reasons. Firstly, data imbalance issue always plays an important role in machine learning tasks. As a small part inside the brain, tumor segmentation on the whole brain is an extremely imbalanced problem and this factor might deteriorate the final result a lot. Secondly, computational efficiency is another big issue. The full brain image size might be practical for a 2D net but it is unrealistic to input such a large volume into a deep 3D net.

Since the sizes of tumors are limited, it will be beneficial to only segment an area around the tumor. In the training step, an $86 \times 86 \times 86$ area is set to be the “tumor bounding box” and the size of the training data (86×86 for 2D net). The only difficulty is how to define the bounding box of the testing dataset. Fortunately, the proposed network in [1] of the BraTS 2017 Challenge is a proper algorithm for tumor “detection”. The central point of the segmented area of this algorithm will either inside the tumor or near the tumor and the boundary of each dimension will have a 10-pixel extension based on the segmentation output of the algorithm. Finally, we can make a bounding box that approximately contains the whole tumor.

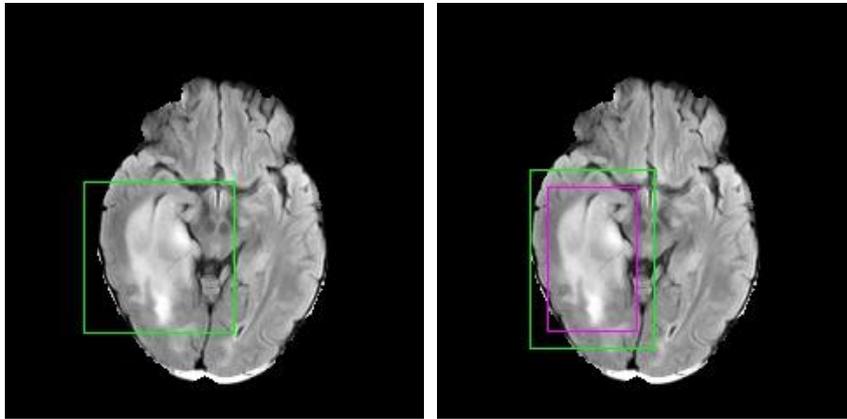


Fig. 2. A figure caption is always placed below the illustration. Short captions are centered, while long ones are justified. The macro button chooses the correct format automatically.

2.2 Multi-dimensional Models Ensemble: Three Individual Convolutional Neural Networks

Recently, a new kind of layer connection called dense connection becomes increasingly popular in CNN based detection and segmentation. Some previous work in natural image processing showed that DenseNet [14] can be seen as an extension of ResNet [13]. Thus, we decide to use this kind of architecture in our method.

One widely used network architecture with dense block is Tiramisu Neural Network [5]. It composes the advantage of dense block and U-net, which is proved to be a effective architecture in medical image processing. Therefore, we design three different neural nets based on the “dense block and U-net” architecture in three different dimensions: 2D, 2.5D and 3D.

2D Net. Our 2D net is similar to Tiramisu-103, but not that deep because the input size of our images are smaller than the natural images used in [5]. The advantage of this 2D net is that more detailed boundary information can be grasped by this network, while the drawback will be too much false positives in the meantime. The total number of dense block is 11 and the layers inside each block (5, 5, 6, 7, 7) are shallower than Tiramisu-103.

3D Net. Our 3D net is an extension based on the 2D net. The main difference between them is that all the input images, output images and kernels are 3D volumes instead of 2D images. Moreover, all the depth and weights of the 3D network are different from the 2D network. The advantage of the 3D net is that it considers the tumor as an integral whole. The consistency between consecutive slices of 3D model will be much better than the 2D model and the segmented boundary becomes smoother. However, it also

sacrifices some details in each slice. The total number of dense block of our network is 7.

2.5D Net. The framework of this network is shown in Fig. 3. The highlight of this network is that the $3 \times 3 \times 3$ kernel is decomposed into $3 \times 3 \times 1$ kernel (inter-slice) and $1 \times 1 \times 3$ kernel (intra-slice). This network can balance the weights of inter-slice connections and intra-slice connections by using different numbers of kernels, respectively. As we know, the resolutions of x-y-space, x-z-space and y-z-space are not the same. Thus, it is reasonable and necessary to have different weights on inter-slice pixels and intra-slice pixels.

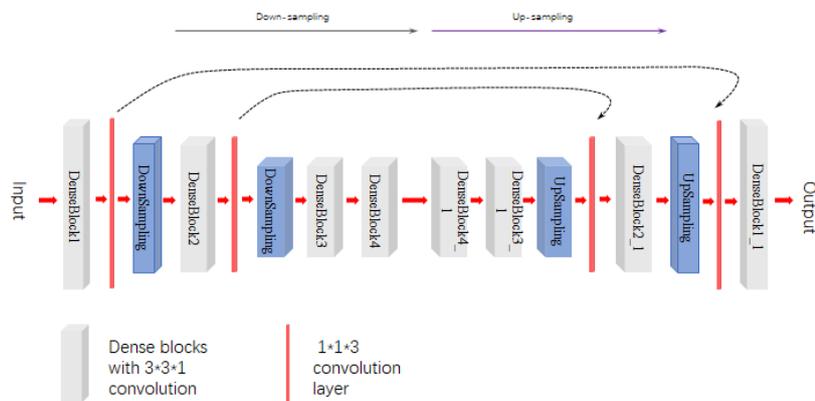


Fig. 3. Architecture of our 2.5D network.

Finally, we ensemble the outputs of 3 networks via a majority voting. This is a good trade-off strategy by involving all the 2D, 3D and resolution information as shown in Fig. 4.

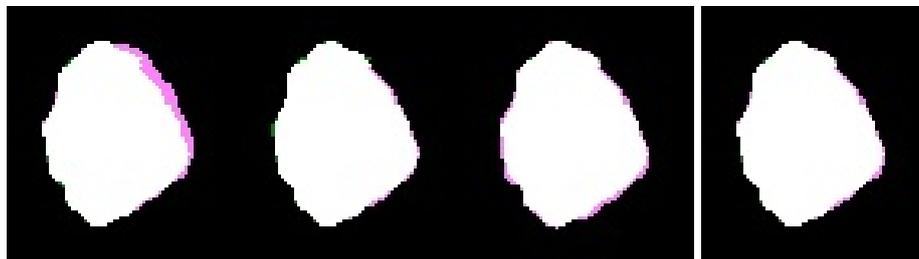


Fig. 3. From left to right: 2D result, 2.5D result, 3D result and ensemble result from one scanning direction. Green pixels are false negatives and pink pixels are false positives. It shows that the ensemble result achieves fewer false negatives and false positives.

2.3 Multi-direction Ensemble

As we mentioned above, the resolution of different scanning directions varies a lot and the information contained in three directions are also different. Thus, it is necessary to train 3 models based on 3 directions. A final ensemble of those outputs can be considered as the connection and balance between different information from 3 directions.

2.4 Cascaded Network

In this task, the enclosing relationship of 3 tumor regions is taken into consideration for segmentation. Since the whole tumor include all the tumor core area and the tumor core contains all the enhanced tumor core area, we can perform 3 binary segmentations step by step: 1. background vs. whole tumor; 2. background vs. tumor core; 3. background vs. enhanced tumor core. The background of each segmentation step will become smaller and smaller, which eliminates part of the background noises. Meanwhile, the depths and weights will be changed by a new trained model in adapt to the sizes and properties of different tumors.

3 Methods

3.1 Data Preprocessing and Implementation Details

We use the BraTS 2018 [15-18] dataset for experiments. 285 image volumes from patients, which include 210 HGG and 75 LGG, are served as the training dataset. 66 patients' volumes are contained in the validation set. Each patient has been scanned by 4 registered MRI modality sequences: T1, T1ce, T2 and FLAIR. All the images are skull-stripped and interpolated into an isotropic resolution. The ground truth for the segmentation in the training set is drawn manually by the doctors and no segmentation is given for the validation set. In this paper, all the results, including Dice, Sensitivity, Specificity and Hausdorff distance, are calculated by the server of BraTS 2018 after uploading the segmentation results of our algorithm.

Our networks are mainly implemented by revising the Tiramisu net in PyTorch. Training is implemented on an NVIDIA Tesla P40. For preprocessing of the images, all are normalized by the mean and standard variation of the training images.

3.2 Preliminary Segmentation Results

Many kinds of losses are explored to achieve the best segmentation result. Finally, we apply the Dice loss and hard-example-mining as the loss function when training the final model.

The training patch size is 86×86 for the 2D net and $86 \times 86 \times 86$ for the 2.5D and 3D nets. The testing patch size is self-adaptive according to the output of rough detection step of our pipeline.

We use 241 patients (210 HGG and 31 LGG) to train and evaluate our models on the rest 44 patients of LGG. The results show that the 3D network outperforms others with the Dice of 0.91 on whole tumor. The Dice scores achieved by the 2D and 2.5D nets are 0.85 and 0.89, respectively. Fig.4 shows an example of outputs of these networks. As can be seen, the 3D network achieves a better performance with a smaller false positive rate. The ensemble of these results could get a little better result.

Table 1. shows the performance of our best model on the validation set. Our model achieves average Dice scores of 0.788, 0.903, 0.851 for enhancing tumor core, whole tumor and tumor core, respectively.

Table 1. The result of our 3D models evaluated on the validation set.

	Dice			Hausdorff		
	ET	WT	TC	ET	WT	TC
Mean	0.788	0.9036	0.851	3.39	5.28	6.76
StdDev	0.23	0.057	0.142	4.34	8.26	9.23
Median	0.86	0.916	0.897	1.866	3	3.35

4 Conclusions and Discussion

Considering the model complexity, data imbalance issues and memory consumption, we propose a two-step cascaded CNN strategy which can improve the accuracy and reduce the computation cost. To reduce the difficulty of this task, we train different networks on different dimensions to segment the whole tumor, tumor core and enhancing tumor core, respectively.

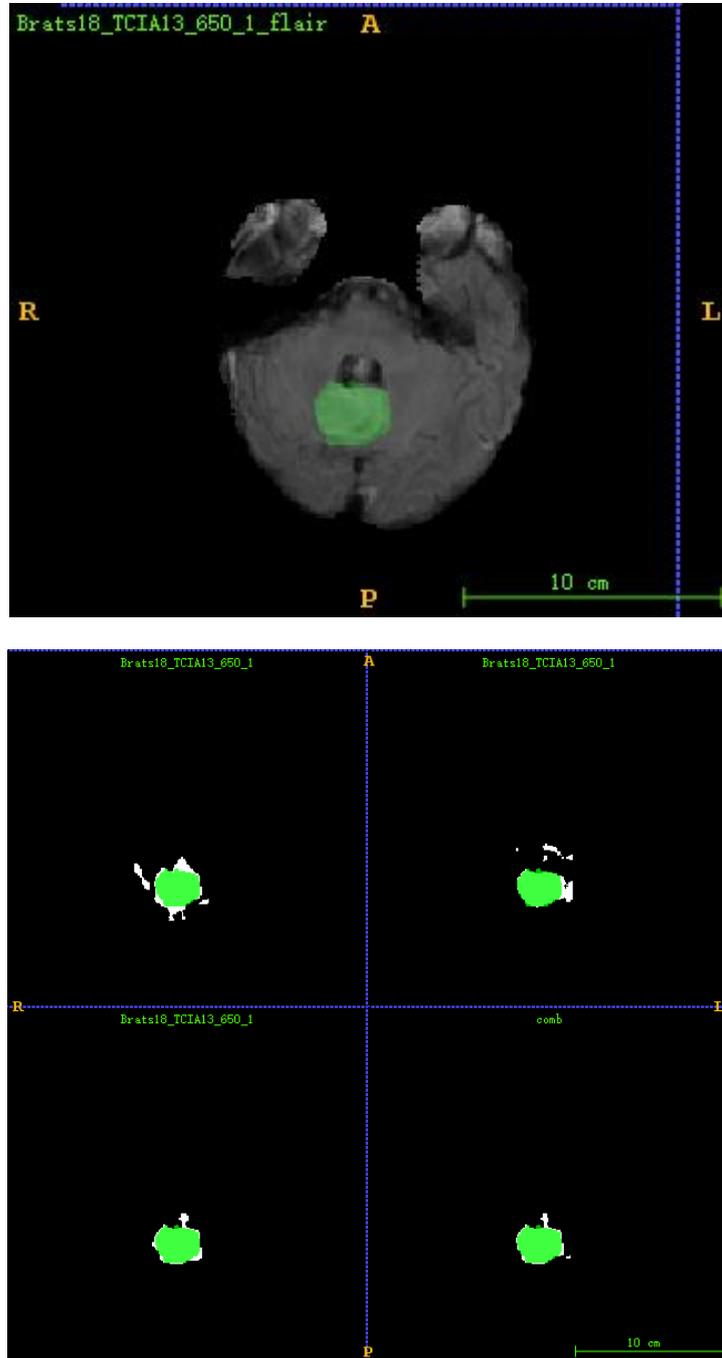


Fig. 5. An example from training set. The first one is a FLAIR image with ground truth and the rest are outputs of 2D, 2.5D, 3D and ensemble network.

In the future study, we will continue our experiments on fine-tuning all the networks, including seeking for the optimal weights and optimal blocks inside the network and a proper loss function.

References

1. Wang G., Li W., Ourselin S., Vercauteren T. Automatic Brain Tumor Segmentation Using Cascaded Anisotropic Convolutional Neural Networks. *Lecture Notes in Computer Science*, vol 10670. Springer, Cham. (2018)
2. Menze, B.H.. et al.: The multimodal brain tumor image segmentation benchmark (BRATS). *TMI* 34(10), 1993-2024 (2015)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., & Kirby, J. S., et al.. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific Data*, 4(170117), 170117. (2017)
4. Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: *MICCAI*. pp. 424-432 (2016)
5. Jegou, S., Drozdal, M., Vazquez, D., Romero, A., & Bengio, Y.. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. *computer vision and pattern recognition*, 1175-1183. (2017)
6. Fidon, L., Li, W., Garcia-peraza herrera, L.C.: Generalised Wasserstein Dice score for imbalanced multi-class segmentation using holistic convolutional networks. *arXiv preprint arXiv:1707.00478* (2017)
7. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. pp. 240-248 (2017)
8. Salehi, S. S., Erdogmus, D., & Gholipour, A.. Tversky loss function for image segmentation using 3D fully convolutional deep networks. *arXiv: Computer Vision and Pattern Recognition*, 379-387. (2017)
9. DeAngelis, L.M.: Brain tumors. *New Engl. Journ. of Med.* 344(2), 114-123 (2001).
10. Urban, G., Bendszus, M., Hamprecht, F., Kleesiek, J.: Multi-modal brain tumor segmentation using deep convolutional neural networks. *BRATS-MICCAI* (2014).
11. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in mri images. *IEEE TMI* 35(5), 1240-1251 (2016).
12. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* 36, 61-78 (2017).
13. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, (2015).
14. G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten. Densely connected convolutional networks. *CoRR*, abs/1608.06993 (2016).
15. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE TMI* 34(10), 1993-2024 (2015).
16. Akbari, H., Bakas, S., Pisapia, J. M., Nasrallah, M. P., Rozycki, M., & Martinez-Lage, M., et al.. In vivo evaluation of egfrviii mutation in primary glioblastoma patients via complex multiparametric mri signature. *Neuro-oncology* (3) (2018).

17. Bakas, Spyridon & Akbari, Hamed & Sotiras, Aristeidis & Bilello, Michel & Rozycki, Martin & Kirby, Justin & Freymann, John & Farahani, Keyvan & Davatzikos, Christos.. Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. 10.7937/K9/TCIA.2017.KLXWJJ1Q (2017).
18. Bakas, Spyridon & Akbari, Hamed & Sotiras, Aristeidis & Bilello, Michel & Rozycki, Martin & Kirby, Justin & Freymann, John & Farahani, Keyvan & Davatzikos, Christos.. Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. 10.7937/K9/TCIA.2017.GJQ7R0EF (2017).

Brain Tumor Segmentation and Survival Prediction Based On Extended U-Net Model and XGBoost

Xiaowen Xu ^{*1}, Xiangmao Kong ^{*1}, Guoxia Sun¹, Fengming Lin¹, Xiaomeng Cui¹, Shuang Sun¹, and Qiang Wu^{1,2} ^{**} and Ju Liu^{1,2}

School of Information Science and Engineering, Shandong University, Jinan, China, 250100

Institute of Brain and Brain-Inspired Science, Shandong University, Jinan, China, 250100

Abstract. In this paper, we propose our method for gliomas segmentation and overall survival prediction in MICCAI 2018 BraTS datasets. U-net and feature pyramid network are mainly used in the segmentation network. The dice-loss value of the segmentation test set reached 0.894 on the whole, 0.743 on the enhancing and 0.786 on the core. The method of multilevel feature fusion is much better than the traditional U-net method. In the survival prediction tasks, we propose a overall framework based on *XGBoost*. In this framework, we extract five kinds raw features from four modality MRI images and segmentation groundtruth. After harvesting those significant features using the *Kaplan – Merier* algorithm, we employ the *XGBoost* algorithm to complete the regression task and perform the survival prediction on the BraTS dataset. Finally, we obtain the accuracy of 65% with ten cross-validation on the training dataset.

Keywords: U-net · Feature Pyramid · Segmentation · Survival Predict · XGBoost.

1 Introduction

Gliomas is a type of brain tumor that grows from glial cells. According to different aggressiveness, brain tumors are broadly divided into high-grade and low-grade[1]. Low-grade gliomas is well-differentiated and exhibits benign tendencies, which portend a longer survival time for the patient. High-grade gliomas are undifferentiated and invasive, which carry a worse prognosis, leading the patient to death. For both groups, intensive methods are used to evaluate the survival time of the patients.

* Xiaowen Xu and Xiangmao Kong are co-first authors and contribute equally to this study.

** Corresponding author: wuqiang@sdu.edu.cn

Recently in the medical field, machine learning (ML) and deep learning (DL) techniques have been used widely in automatic detection, classification, and regression [5]. Different methods have been proposed for predicting the overall survival of patients with brain tumors by several studies [7], [1], [11]. In [11], the author predicts the survival by training an ensemble of a random forest regressor and a multi-layer perceptron on shape features describing the tumor subregions. As stated in [11], an overall survival prediction method was proposed by Random forest regression model based on different structural multi-resolution texture features, volumetric, and histogram features. In this paper, we employ the *XGBoost* model to achieve the survival prediction task based on different texture, histogram and volumetric features.

2 Segmentation Model

To achieve the segmentation task, we proposed a model based on U-net [10] and feature pyramid network [8]. All deep learning networks are accomplished with Keras library in Python.

Our model includes two parts, the first part is U-net structure and the second part is feature pyramid network. The traditional U-net was used to segment medical image, which was composed of encoder part and decoder part. The architecture of our proposed model is shown in Figure 1. It consists of 3 modules, a downsampling path with convolution and max pooling, an upsampling path with convolution and up-sample, and a feature pyramid path. The downsampling path is mainly to enlarge the receptive field to extract high-level and global contextual features of the tumor. And the upsampling path is used to reconstruct the object details. As we know, the high-level path has much semantic information and low-level path has much location information, the auxiliary path is used to extract multi-scale information and make full use of multiple levels of information and combine semantic and location information in the upsampling process to help the model complete segmentation for objects of different scales.

Additionally, in the second part we have used feature pyramid network to integrate the low-level and high-level features. Traditional structure of segmentation networks neglect the effect of low-level feature maps. In the feature pyramid network part, we select different scale feature maps from different layers. First we sample up them to the same scale. Then we add up the pixels in the same position from different images. Therefore we get a feature map combined with high-level and low-level features.

3 Survival Prediction Method

3.1 Feature Extraction

Texture Feature We extract 208 features from multimodality MRI data (T1, T2, T1ce, Flair), including Fris-t-order histogram, Gray-Level Co-occurrence Matrix

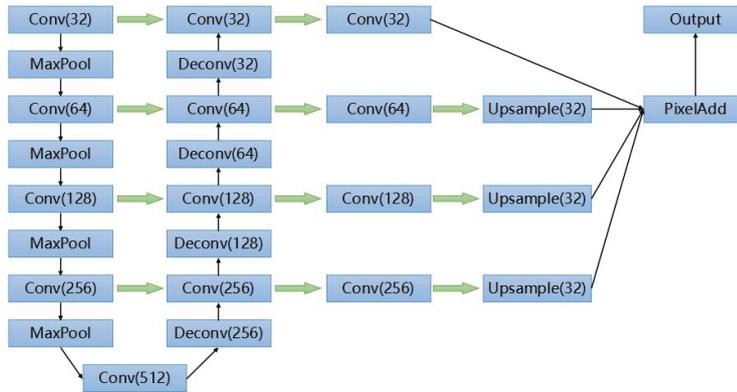


Fig. 1. Architecture of the segmentation network.

(GLCM), Gray-Level Run-Length Matrix (GLRLM), Gray-Level Size Zone Matrix (GLSZM), Neighborhood Gray-Tone Difference Matrix (NGTDM) [12] and their own attributions.

Volumetric feature In addition to the above features, 7 volumetric features are extracted in our experiment. The volumetric features we extracted in this paper including the volume of whole tumor (WT), the volume of the whole tumor with respect to the brain, the volume of sub-regions (ED, ET, NCR/NET) with respect to whole tumor, and the volumes of ET and NCR/NET with respect to ED, the volumes of ED and NCR/NET with respect to ET, and the volumes of ET and ED with respect to NCR/NET [11].

Tumor Profile Area Feature We also extract some tumor profile area features as following [11], the area centroid, perimeter, major axis length, min axis length, eccentricity, orientation from x, y, z axis. We use those variable of fitted ellipse to instead of this features. A total of 84 features characterize the profile area feature in this part.

First-order Statistics Feature We also extract 42 first-order statistic features in this section. There are mean, variance, skewness, kurtosis, energy and entropy of tumor and different tumor sub-regions.

Location Feature The tumor locations and the spread of the tumor in the brain are also considered. Features extracted from the histogram of the four modalities of the whole tumor and its subregions are also considered. These features represent the frequency at different intensity bins (number of bins = 11, 23), and the bins of the max frequency.

3.2 Feature Selection

Since we collect too many features in the Feature Extraction, we need to select effective features. We use the *Kaplan – Merier* algorithm to evaluate the performance of each features. Our goal is to find the significant features that can separate the samples into short vs long survival as follows [11]. In this section, we only take one round to select significant feature. Fig.2 shows the Kaplan-Merier of five significant features.

3.3 Our Regression Framework

In this paper, we use the XGboost [6] to complete the regression task. The overall Regression framework is shown in Fig. 3. We extracted five kinds features from four different modality MRI images and respective segmentation groundtruth. After harvesting those protogenetic features, we employ the *Kaplan – Merier* algorithm to select the significant features to achieve the better regression performance. Finally we use the *XGBoost* algorithm to complete the regression task and obtain the survival prediction on every sample.

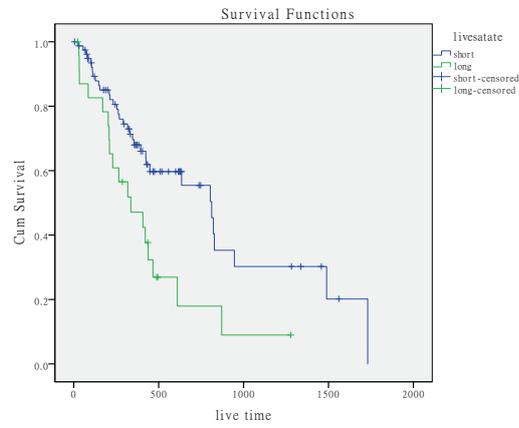
4 Experiments

4.1 Segmentation

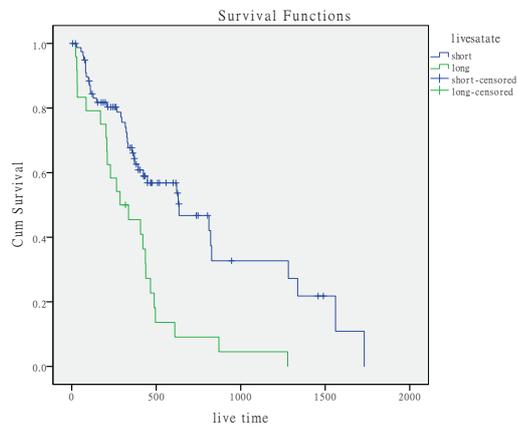
We normalized each subject data with zero mean and unit standard deviation. Then the slices that do not contain tumor information were removed from training data. We augmented the MRI data by left rotating the first half and right rotating the other half to construct a new dataset that is twice times larger than the original dataset. We used the keras library with tensorflow as the backend. The model was trained with standard back-propagation using Adam as an optimizer, and all parameters are initialized using he normal. The training time on the augmented data is about ten hours with 70 rounds run using a workstation computer with a NVIDIA Titan X GPU. The dice changes with the epoch are shown in Fig. 4. And Table 1 shows the result of PPV, Sensitivity and dice results. We can see from Fig. 5 that the segmentation result is very close to the groundtruth.

Table 1. The result of segmentation

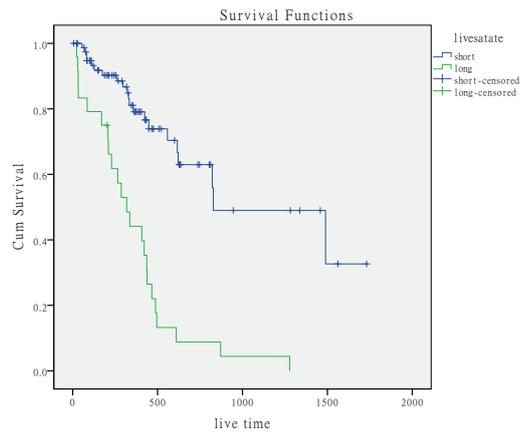
	Enh.	Whole	Core
PPV	0.92899	0.93487	0.93125
Sensitivity	0.66739	0.96641	0.85381
Dice	0.74309	0.89409	0.78629



(a)



(b)



(c)

Fig. 2. The visual result of three features in our Feature Selection section. (a) The Skewness of Flair global ($\text{sig}=0.007$). (b) The Hrhge of Flair GLRLM ($\text{sig}=0.0003$). (c) The Lzhge of Flair GLSZM ($\text{sig}=0.0003$).

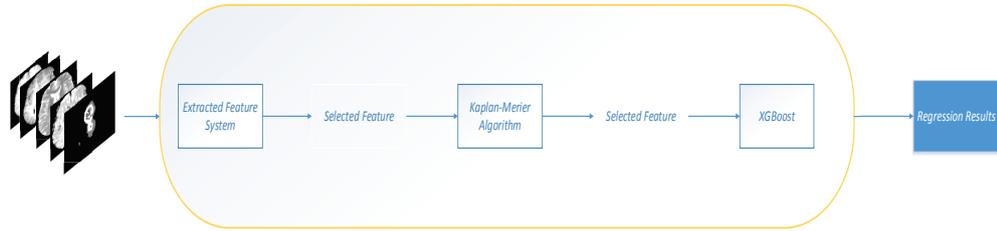


Fig. 3. Our Framework of Survival Prediction Based on XGBoost

4.2 Survival Prediction

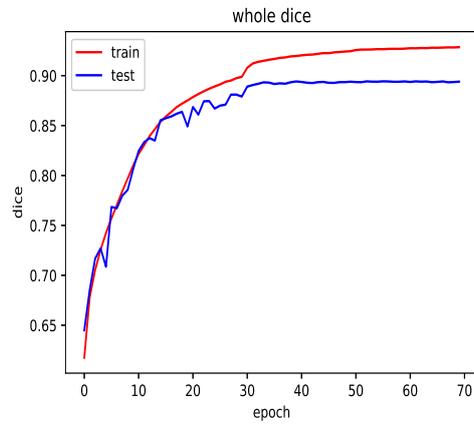
In this section, we use 160 samples in Brats2018 [9],[4],[2][3] training data to train our framework for the survival prediction task with the 10-fold cross-validation. We use the mean square error (MSE) of the criterion to train the *XGBoost* for regression tasks. In order to intuitively evaluate the performance of our framework in survival prediction task, we take the overall survival data into three classes (long, medium and short) survival corresponds to (≥ 15 months, ≥ 10 months and ≤ 15 months, ≤ 10 months), respectively. The classification results in our framework is in Table.2

Table 2. The result of survival prediction result with ten-fold cross-validation

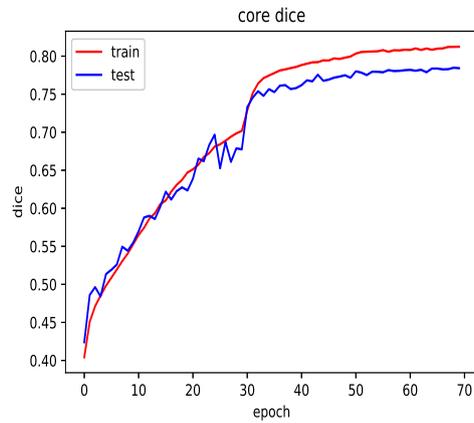
Data	Result
Brats2018	0.65

5 Conclusion

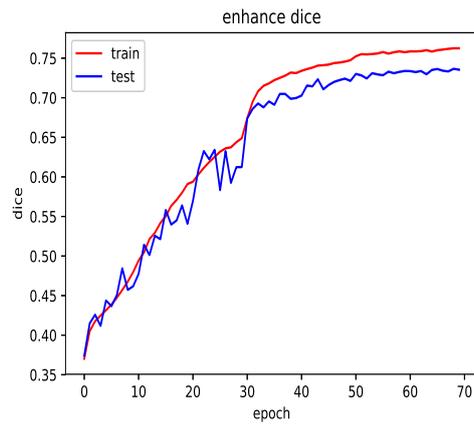
In this paper, we propose an end-to-end brain tumor segmentation model based on extended UNet. Our model includes a downsampling path and an upsampling path and a pyramidal hybrid path to extract multiscale information. Going deeper made the dice improved, and adding the image pyramid model also improved the segmentation result. Our model achieved the effect of SOA, and later we will consider evaluating on larger dataset. In the survival prediction task, we proposed an overall framework based on *XGBoost*. In this framework, we extracted five kinds of features from four different modality MRI images and segmentation groundtruth. After harvesting those protogenetic features, we use the *Kaplan – Merier* algorithm to select the significant feature to have the better regression performance. Finally we use the *XGBoost* to complete the regression task and get the survival prediction on every sample. As a result, we get the accuracy of 65% with ten cross-validation on the training data.



(a)



(b)



(c)

Fig. 4. The performance curves of dice. From left to right: complete, core and enhancing. The vertical axis is Dice while horizontal axis is the number of epochs.

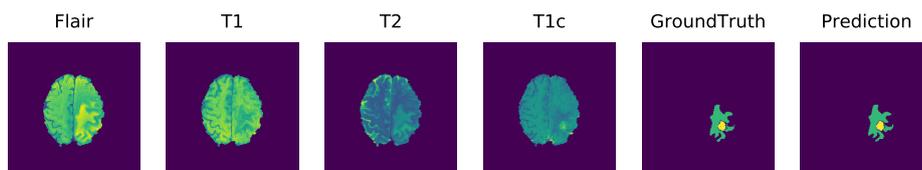


Fig. 5. Segmentation results.

References

1. https://www.cbica.upenn.edu/sbia/Spyridon.Bakas/MICCAI_BraTS/MICCAI_BraTS_2017_proceedings_shortPapers.pdf
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive* **286** (2017)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *The Cancer Imaging Archive* (2017)
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data* **4**, 170117 (2017)
5. Chato, L., Latifi, S.: Machine learning and deep learning techniques to predict overall survival of brain tumor patients using mri images. In: *Bioinformatics and Bioengineering (BIBE), 2017 IEEE 17th International Conference on*. pp. 9–14. IEEE (2017)
6. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. pp. 785–794. ACM (2016)
7. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In: *International MICCAI Brainlesion Workshop*. pp. 287–297. Springer (2017)
8. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: *CVPR*. vol. 1, p. 4 (2017)

9. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging* **34**(10), 1993 (2015)
10. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 234–241. Springer (2015)
11. Shboul, Z.A., Vidyaratne, L., Alam, M., Iftekharuddin, K.M.: Glioblastoma and survival prediction. In: *International MICCAI Brainlesion Workshop*. pp. 358–368. Springer (2017)
12. Vallières, M., Freeman, C.R., Skamene, S.R., El Naqa, I.: A radiomics model from joint fdg-pet and mri texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Physics in Medicine & Biology* **60**(14), 5471 (2015)

Cascaded Mask V-Net

Yanwu Xu¹, Mingming Gong^{1,3}, Huan Fu², Dacheng Tao², Kun Zhang³, and Kayhan Batmanghelich¹

¹ University of Pittsburgh, Department of Biomedical Informatics yanwuxu@pitt.edu

² UBTECH Sydney AI Centre, SIT, FEIT, The University of Sydney

³ Carnegie Mellon University, Philosophy Department

Abstract. As for an accurate segmentation task for tumor detection of multi-class sub-regions which are composed of peritumoral edema, necrotic core, enhancing and non-enhancing tumor core using multimodal MRI scans, this task is very challenging due to its intrinsic highly heterogeneity of appearance and shape as well as a 3-D unnatural image. However, recently, with the development of deep model and calculating resources, deep neural network has shown its effectiveness on brain tumor segmentation and get the top performance in MICCAI BraTS challenge [1–4]. In this paper we propose a patches training method which extracted 3-D patches from original scanning images, and our networks is based on V-Net [5]. The basic idea is that we classify each voxels into four categories which are Whole Tumor(WT), Tumor Core(TC), Enhancing Tumor(ET) and Background(BC). And we report our preliminary result of the online validation system which shows the efficiency of our V-Net based method.

Keywords: Brain Tumor Segmentation · Patches Training · V-Net

1 Introduction

Gliomas are the most common primary brain malignancies, with different degrees of aggressiveness, variable prognosis and various heterogeneous histological sub-regions, i.e. peritumoral edema, necrotic core, enhancing and non-enhancing tumor core. Accurate segmentation and measurement of the different tumor sub-regions is critical for monitoring progression, surgery or radiotherapy planning and follow-up studies. However, the distinction between tumor and normal tissue is difficult as tumor borders are often fuzzy and there is a high variability in shape, location and extent across patients. All the imaging datasets have been segmented manually and offered by BraTS organizer, by one to four raters, following the same annotation protocol, and their annotations were approved by experienced neuro-radiologists. Annotations comprise the GD-enhancing tumor (ET label 4), the peritumoral edema (ED label 2), and the necrotic and non-enhancing tumor core (NCR/NET label 1), as described in the BraTS reference paper [2], so Whole Tumor(WT) is including label 1,2 and 4, Tumor Core (TC) is including label 1 and 4 and then Enhancing Tumor (ET) is including label 1.

There are many different algorithms being proposed to solve this problem such as U-Net [6] and V-Net [5]. As we observed that these two structures are similar and U-Net is used for brain tumor segmentation and V-Net is the upgraded version for detecting lung tissue lesion. Here we apply 3-D version V-Net as our basic model. And we extract 3-D $64 \times 64 \times 64$ patches around the tumor area. Furthermore, we can observe the relationship between each tumor that WT contains TC and TC contains ET. Thus, we design our network for discriminating these three types of area and we apply mask strategy among these tumor which we detect TC only in the area of WT and ET only in the area of TC.

As for survival task, we classify each patient to three survival length, which are short survivors ($p < 3months$), mid survivors ($3months < p < 6months$) and long survivors ($p > 6months$). We extract features from the segmented label rather than original MRI images and then we concatenate extracted features with normalized ages. At the end, we feed final feature vector to a simple fully-connected classified neural network.

2 Proposed Method

The proposed method is as follows:

1. Pre-processing of data
2. Training stage
3. Post-processing
4. Feature extraction for survival net training and survival rate prediction.

2.1 Data Preprocess and Patches extraction

The proposed V-Net like neural network was trained on BraTS2018 training data and validated on BraTS2018 validation data. The training data comprises 210 HGG volumes and 75 LGG volumes collected from multiple centers. Each patient comprises of FLAIR, T2, T1, T1 post contrast and the associated ground truth labeled by experts. Each sequence was skull stripped and was re-sampled to $1mm \times 1mm \times 1mm$ (isotropic resolution).

To compensate for the MR inhomogeneity, we apply the bias correction algorithm based on N4ITK library to the T1 and T1ce images. To reduce the effect of the absolute pixel intensities to the model, an intensity normalization step is applied to each volume of all subjects by subtracting the mean and dividing by the standard deviation so that each MR volume will have a zero mean and unit variance. In practice, as the original uncropped volume is used in which the brain only takes the central region, the mean and standard deviation are estimated using the central region (0.25 - 0.75 on each dimension) of the volume. And we extract the patches around the Tumor area, when the tumor ratio in the $64 \times 64 \times 64$ volume is $r = 0.75$ the choosing chance will be the maximum. And we extract 400 patches from each patient in total.

2.2 Model Structure

As shown in Fig. 1, we apply this typical network structure as our segmentation generation model, as for tumor segmentation we firstly train WT ,TC and ET simultaneously to get the best WT tumor mask for TC, secondly we train TC and ET together under the mask of WT and thirdly we train ET alone under the mask of TC. This training strategy is namely cascaded V-Net. The idea is inspired by remove redundant background information for each tumor training focusing more on vital area.

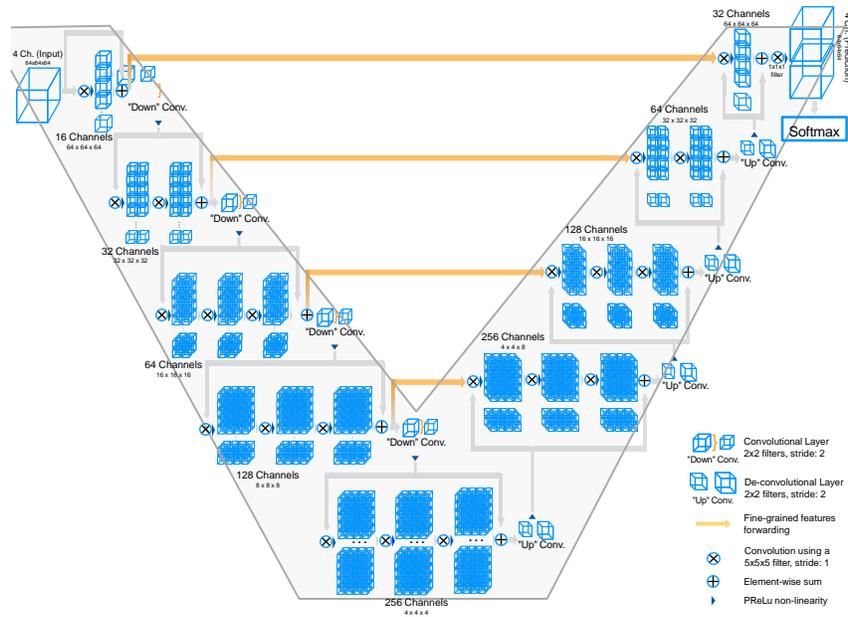


Fig. 1. The typical V-Net structure.

The details of our mask strategy is shown in Fig. 2. We apply tumor flair t1 t1ce and t2 as our input concatenating in channel dimension, so our input is 4-D patches, cause we consider that multimodal can help train each model.

2.3 Model Training

We train our model with Adam optimizer and apply dice loss and Cross Entropy loss in the same time for optimize our parameters of model. The dice loss can be written as follows:

$$D = \frac{\sum_{i=1}^N p_i \cdot l_i}{\sum_{i=1}^N p_i + \sum_{i=1}^N l_i} \tag{1}$$

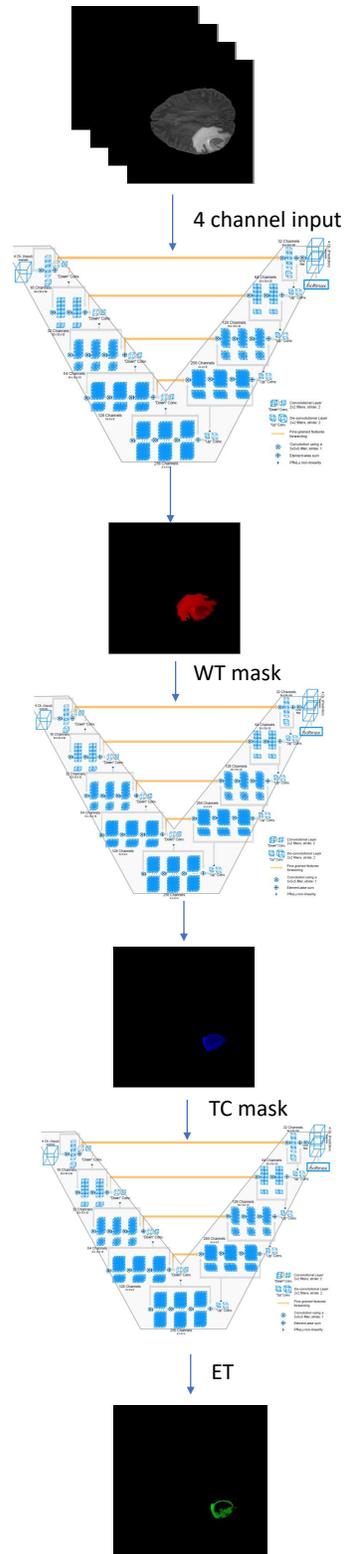


Fig. 2. Mask structure

N represents the number of voxels, p is the output probability of each voxel and l denotes the ground truth label. The idea is different from typical classification loss cross entropy. It focus more on the overlap between output and label rather than background.

3 Results and discussion

We introduce the evaluation metric and validation result of segmentation below:

3.1 Evaluation Metrics

Dice Coecient The Dice-Coecient (Equation 2) is calculated as performance metric. This measure states the similarity between clinical Ground Truth annotations and the output segmentation of the model. Afterwards, we calculate the average of those results to obtain the overall dice coecient of the models.

$$D = \frac{2|A \cap B|}{|A| + |B|} \quad (2)$$

Hausdor Distance The Hausdor Distance (Equation 2) is mathematically dened as the maximum distance of a set to the nearest point in the other set [15], in other words how close are the segmentation and the expected output.

$$H(A, B) = \max(\min(d(A, B))) \quad (3)$$

Table 1. BraTS18 tumor validation segmentation results.

Tumor type	Dice Score	Hausdor95	Sensitivity	Specificity
WT	0.9020	6.03694	0.91691	0.9941
TC	0.8200	6.78376	0.81829	0.99783
ET	0.7861	4.06025	0.84103	0.99745

we also visualize several patients' segmentation result from validation data which are shown as Fig. 3

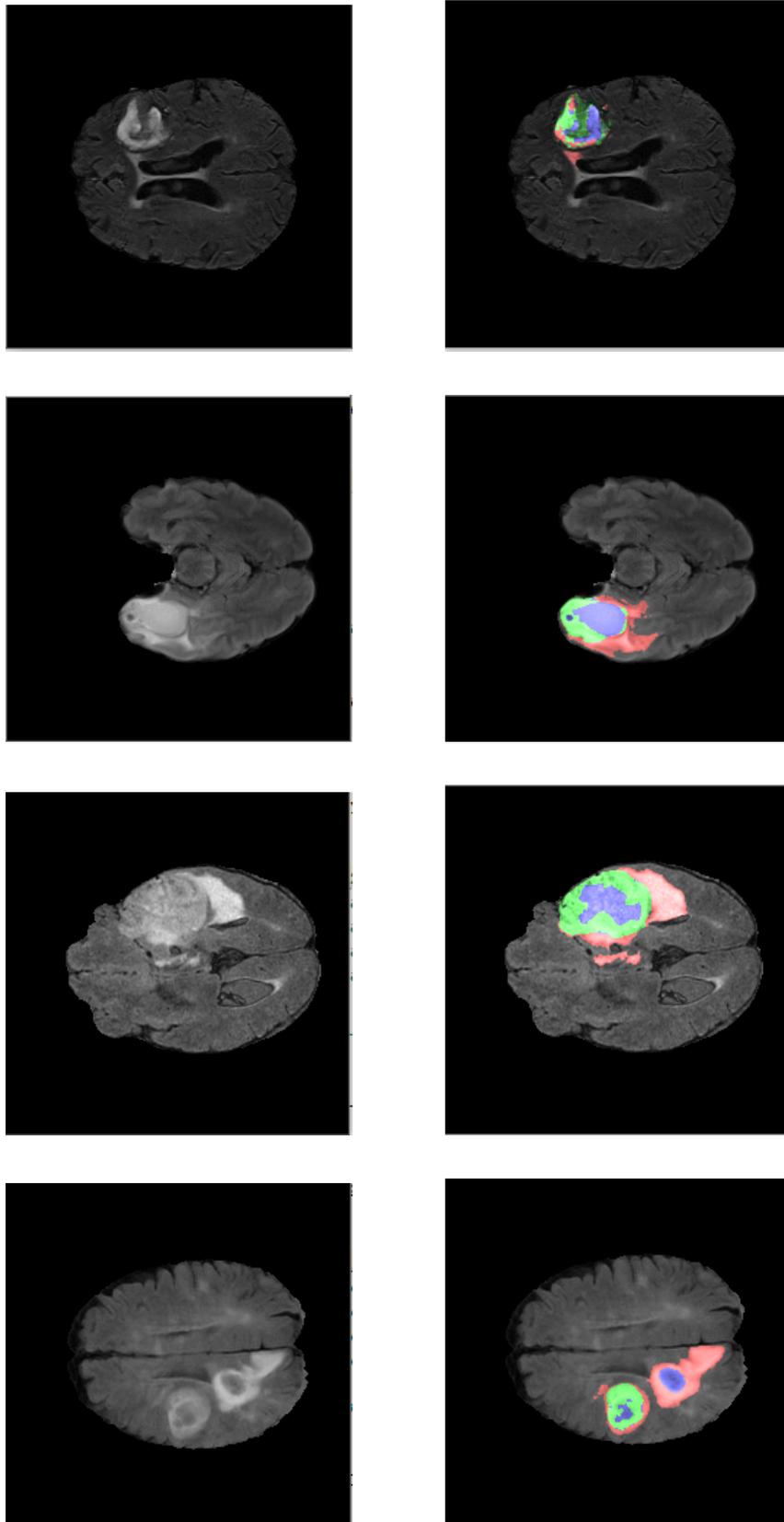
4 Conclusions

Based on V-Net, we proposed a cascaded mask neural network training structure. Our method receives as input 3D patches extracted from the dataset volumes and we train three networks seperately based on our cascaded mask strategy, nally, we also extract patches from the validation data and get the segmentation for each patch and we combine this segmentation patches back to normal size with a stride. In this paper, we have presented preliminary results in the 2018 BraTS

Challenge dataset (Validation) reaching an average dice coefficient of 0.902 on WT without data augmentation. We think the result will be better if implementing augmentation and network with modification.

References

1. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp , Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015)
2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
5. Milletari, Fausto, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. *3D Vision (3DV)*, 2016 Fourth International Conference on. IEEE, 2016.
6. Ronneberger, O., Fischer, P., et al.: U-net: Convolutional networks for biomedical image segmentation. In: *MICCAI*. pp. 234241. Springer (2015)



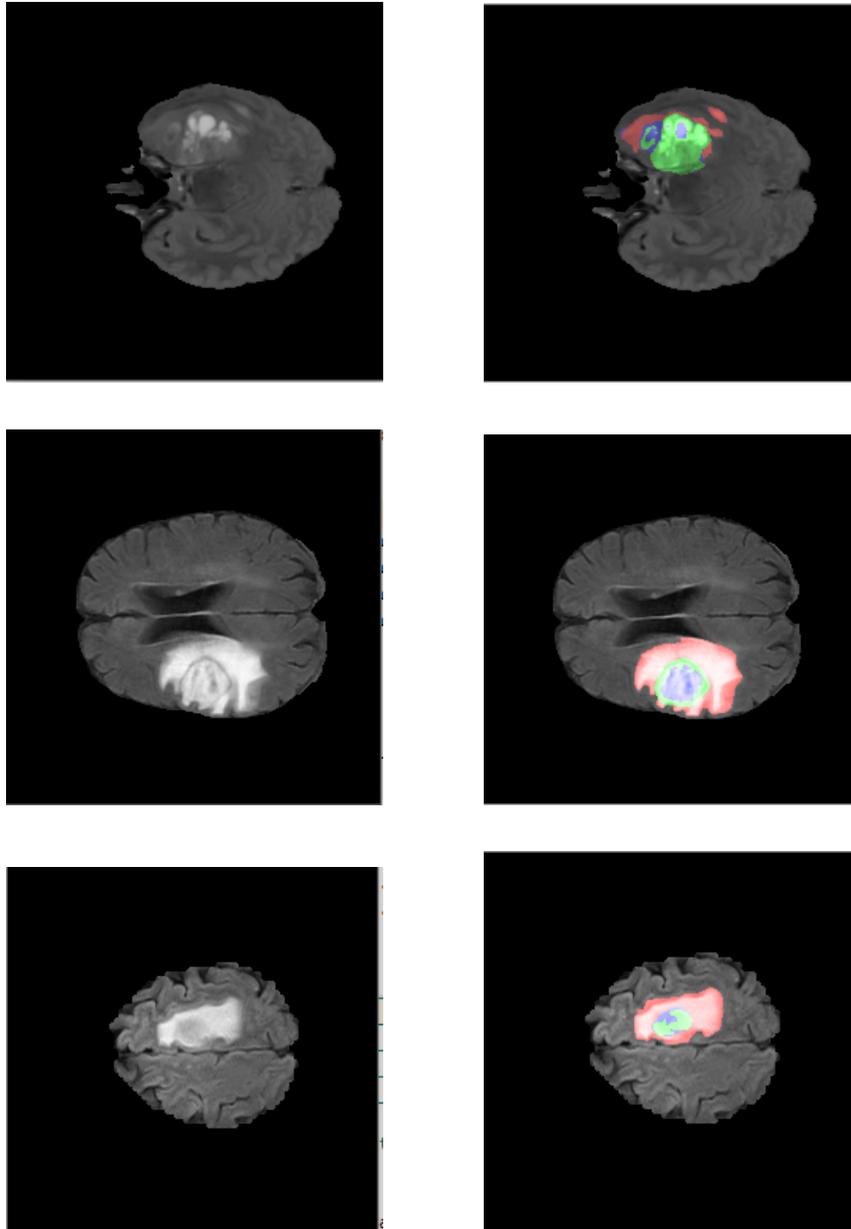


Fig. 3. Pink represent whole tumor ,green represents tumor core and purple represents enhancing tumor.

Automatic Brain Tumor Segmentation with Contour Aware Residual Network and Adversarial Training

Hao-Yu Yang^{1,2} and Junlin Yang¹

¹ Yale University, New Haven, CT 06511, USA

² Cura Cloud Cooperation, Seattle, WA, 98104, USA

Abstract. In this paper, we present a contour-aware 3D convolution neural network (CNN) with adversarial training to improve segmentation results. Image derived contours are added as an auxiliary loss in addition to the pixel-wise classification loss to ensure the segmentation results mimics the contours of the ground truth annotation. For adversarial training, a discriminator network is trained alongside the segmentation network. The classification loss from the discrimination network provide guidance for the generating realistic segmentation results. The proposed network is evaluated on the BraTS 2018 training and validation dataset. We demonstrate that by incorporating extra constraints the raw segmentation results can be improved.

Keywords: Neural Networks · Adversarial Training · Contour Aware

1 Introduction

Glial cells-arose brain tumors are one of the most type of common brain tumors. The segmentation of gliomas and intra-tumor substructures are essential to progression monitor and treatment assessment of the disease. [5]

There are several difficulties that lay in these tasks. First, the gliomas varies significantly in shape, size and location. Second, the normal brain anatomy structure varies from patients to patients. Besides, the proportion of brain tumors to normal brain tissue is quite low, resulting in extreme class imbalance for tumor segmentation.

Previous studies have found successes on automated brain tumor segmentation. Among existing methods, Convolution Neural Networks (CNN) based methods have achieved state-of-the-art performance, especially U-Net type architecture [13]. The rise of residual CNNs grants the possibility of building a deeper network with higher segmentation accuracy, which gives networks more representational power to capture the variations in complex data. [9]. Additionally, adversarial training has been recently applied successfully in image segmentation, using a discriminator network trained to distinguish annotations from predicted segmentations [11]. It enforces CNNs to learn long range spatial label contiguity.

In this work, we tackled two tasks in BraTS 2018. For tumor segmentation task, we employ adversarial training to train a 3D Residual U-Net based architecture. Our segmentation model is designed to be contour aware in a multi-task framework. For survival prediction task, we proposed an ensemble of models including convolutional neural network and random forest to predict the overall survival time. Our models are evaluated on the validation set from BRATS 2018 and demonstrate outstanding performances.

2 Method

2.1 Data

The proposed model is trained and evaluated on the The Multimodal Brain Tumor Segmentation Challenge (BraTS) [1], [2] [3] [4] dataset. The training set contained a total of 285 cases. Each case contained 4 MRI modalities: T1-weighted (T1), T1-weighted with gadolinium enhancing contrast (T1ce), T2-weighted(T2) and Fluid Attenuation Inversion Recovery(FLAIR). Of the 285 cases, 210 cases are High Grade Gliomas(HGG) and 75 cases are Low Grade Gliomas (LGG).

2.2 Preprocess

Appropriate image processing is essential for the following computing. The official dataset provided by the organizers have been skull-stripped and co-registered. Bias correction in [8] is performed on raw data before segmentation. Intensity normalization (scaling and z-score) is also applied for better performance [14].

2.3 Gliomas Segmentation

The Gliomas segmentation task Pixel-wise annotation labels that included GD-enhancing tumor (ET), peritumoral edema(ED) and necrotic /non-enhancing tumor(NCR/NET) were provided alongside the image data. The segmentation results are evaluated by several matrices including Dice score, Hausdorff distance, sensitivity and specificity of the respective tumor tissue.

3D Residual Unet Due to the 3D nature of the MRI and the memory limitation of the current GPU, it is infeasible to feed the proposed network multiple modalities with entire volume. During training phase, a single volume is broken down to smaller patches. The stride size for the patch division is 24 and each the dimension of each patch is 96 by 96 by 96 voxels. Therefore, a single modality image with dimension of 155 by 240 by 240 will generate a total of 72 patches.

The backbone of the segmentation network is a 3D U-net. The U-net consists of both down sampling and up sampling pathways. The down sampling pathway is made up of multiple residual blocks. Each residual block contains two convolution layer, batch normalization, activation and residual connection.

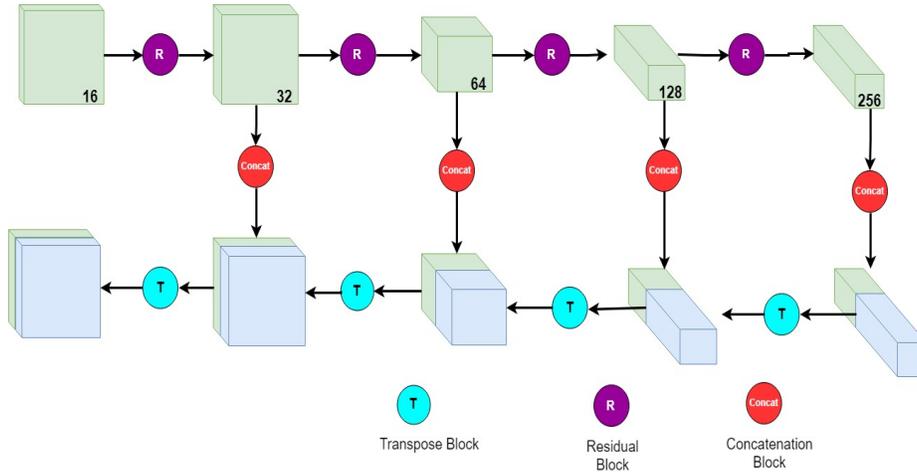


Fig. 1. Schematic of the Segmentation Network. The each circle represents a different set of operations.

Contour Aware In order to obtain a robust segmentation that captures the highly varying multi-class tumor contours, we added a auxiliary constraining factor derived from the ground truth segmentation. For each ground truth mask, an elastic snake (also known as active contour) [10]. The total energy of the deformable spine can be written as:

$$E_{snake} = \int_0^1 E_{snake}(v(s))ds = \int_0^1 (E_{internal}(v(s)) + E_{image}(v(s)) + E_{con}(v(s)))ds \tag{1}$$

Where the set of n points v_i denotes the snake itself. The total energy of the snake is a combination of internal elastic energy and the inherit energy of the image. In traditional computer vision framework, boundary finding tasks involving active contours usually have objectives that minimizes the total energy of the snake. However, these methods need higher-level supervision for the actual contours. Ground truth and neural-network generated segmentation results can serve as essential information for the active contour method. In our approach, the energy derived from the mask is a weighted loss term that is added to the total segmentation loss.

Adversarial Training In each iteration, the segmentation network will generate a proposed segmentation. In order to capture the true distribution of the ground truth segmentation, we added a adversarial training in every iteration. This is implemented by a discriminator network, a shallow 7-layer network containing 3 3D convolution layer, each followed by a maxpooling layer. The objective of the discriminator network is to distinguish between generated masks and ground truth masks.

The original Generative Adversarial Network [12] solution can be written as:

$$\min_G \max_D E_{x \sim p_{data}} [\log D(x)] + E_{z \sim p(z)} [\log(1 - D(G(z)))] \quad (2)$$

In our case, the generator is actually the segmentation network that produces synthetic samples. At each iteration, the segmentation network is back-propagated twice: once from the segmentation loss from that includes the contour constraint and the second time for the discriminator loss.

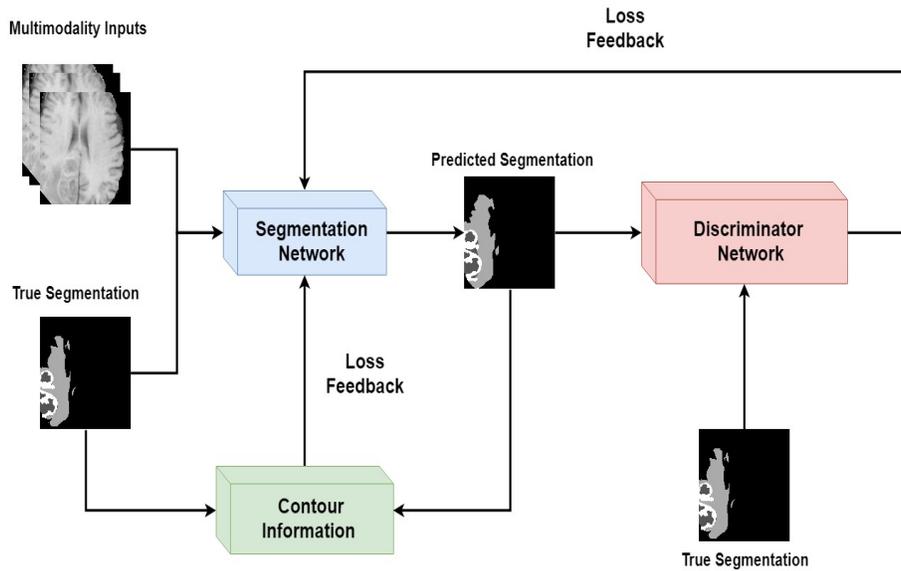


Fig. 2. Overview of the network pipeline

2.4 Overall Survival Prediction

An ensemble of models including Convolution Network-based regression network and feature-based random forest are proposed to predict the overall survival (OS) of High Grade Gliomas (HGG) patients. Second part of the BraTS challenge involves the prediction of overall survival of patients. We have developed an CNN based regression network and a feature-based random forest model. During validation and test phase, we assign different weight to combine the above models according to training loss to get the final prediction.

Regression Network The CNN takes the four modalities MRI volumes and the tumor segmentation mask from the first task as input. The last layer of the regression contains a single node with no activation function. This allows

the network to directly outputs the survival days. The segmentation results are essential to predicting survival time as it contains extracted information about the tumor tissue.

Random Forest Random Forest is a powerful regression model. Here, we utilize imaging features extracted from the tumor segmentation mask, including the relative size and number of different subtype tumors. The age of the patients is also incorporated as a non-imaging feature in the RF model.

Ensemble The variability of a single model can be quite high. In order to reduce the prediction variance, we ensemble results from regression network and random forests at inference phase. Each model is assigned a weight according to training data cross validation results.

2.5 Implementation Details

The proposed model was constructed in pytorch and training is done on NVIDIA Tesla V100.

3 Preliminary Results

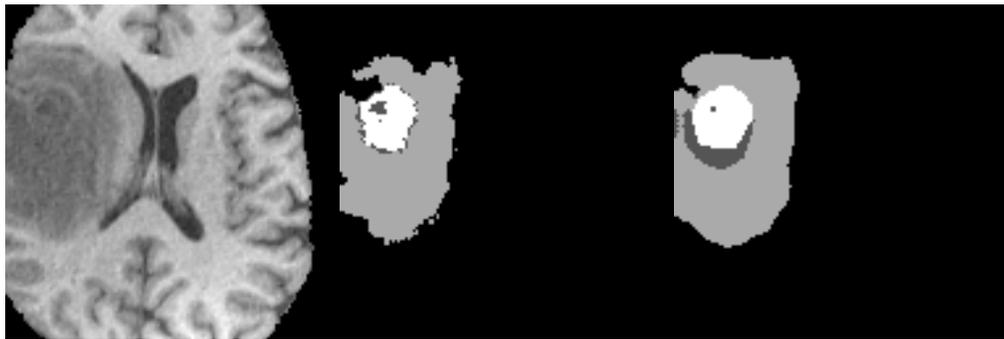
We have tested our segmentation and survival prediction results on the BraTS validation dataset. The following table shows the validation results according to the BraTS on-line evaluation system.

	ET Dice	WT Dice	TC Dice	Hausdorff 95 ET	Hausdorff 95 WT	Hausdorff 95 TC
Training	0.7814	0.8375	0.7861	4.743	5.9031	7.082
Validation	0.7149	0.8011	0.667	5.136	6.231	7.871

Table 1. Preliminary Segmentation Results on the BraTS training and validation set

	Accuracy	MSE	medianSE	stdSE	SpearmanR
Training	0.475	44724.102	12996	79142.403	0.666
Validation	0.321				0.051

Table 2. Preliminary Survival Prediction Results on the BraTS training and validation set



(a) fig 1.1 Input Image (b) fig 1.2 Ground Truth Mask (c) fig 1.3 Model Prediction



(d) fig 2.1 Model Prediction (e) fig 2.2 Model Prediction (f) fig 2.3 Model Prediction

Fig. 3. Visualization of Patch-Based Segmentation results

4 Conclusion

In this paper, We've proposed a novel method for accurate pixel-wise classification and survival prediction. For segmentation task, we utilize adversarial training to train a contour aware 3D Residual U-Net. By introducing extra constraints to the neural network, we are able to generate segmentation results that approximate the ground truth well. For Survival prediction, we proposed a ensemble of models including neural network and random forest to predict survival days.

References

1. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp , Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
5. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Nature Scientific Data*
6. Olaf Ronneberger, Philipp Fischer, Thomas Brox U-Net: Convolutional Networks for Biomedical Image Segmentation
7. Author, F.: Article title. *Journal* **2**(5), 99–110 (2016)
8. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* 29, 13101320 (2010)
9. Tran Minh Quan, David G. C. Hildebrand, Won-Ki Jeong FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics

10. Kass, M.; Witkin, A.; Terzopoulos, D. (1988). "Snakes: Active contour models". *International Journal of Computer Vision*. 1 (4): 321. doi:10.1007/BF00133570.
11. Pauline Luc, Camille Couprie, Soumith Chintala, Jakob Verbeek Semantic Segmentation using Adversarial Networks
12. Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative Adversarial Networks
13. Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, Pheng Ann Heng H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes
14. Nyul, L.G., Udupa, J.K., Zhang, X.: *IEEE Trans. Med. Imaging* 19, 143150 (2000)
15. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun Deep Residual Learning for Image Recognition

Automatic Segmentation of Brain Tumor using 3D SE-Inception Networks with Residual Connections

Hongdou Yao, Xiaobing Zhou^(✉) and Xuejie Zhang

School of Information Science and Engineering,
Yunnan University, Kunming 650091, P.R.China
zhouxb@ynu.edu.cn

Abstract. Nowadays, there are various kinds of methods in medical image segmentation tasks, in which Cascaded FCN is an effective one. The idea of this method is to convert multiple classification tasks into a sequence of two categorization tasks, according to a series of sub-hierarchy regions of multi-modal Magnetic Resonance images. We propose a model based on this idea, by combining the mainstream deep learning models for two dimensional images and modifying the 2D model to adapt to 3D Medical image data set. Our model use the Inception model, 3D Squeeze and Excitation structures, and dilated convolution filters, which are well known in 2D image segmentation tasks. Residual connections can ensure the information effectively transmitted and prevent over fitting. When segmenting the whole tumor, we set the bounding box of the result, which is used to segment tumor core, and the bounding box of tumor core segmentation result will be used to segment enhancing tumor. We not only use final output of the model, but also combine the results of intermediate output. In MICCAI BraTs 2018 gliomas segmentation task, we achieve a competitive performance without data augmentation.

Keywords: 3D-SE-Inception-ResNet · Cascaded FCN · Anisotropic · Medical Image Segmentation.

1 Introduction

Image segmentation has always been a challenging task in the field of computer vision. Especially in medical image field, multi-modal Magnetic Resonance images are used to segment human body pathological tissue. Many medical committees such as MICCAI, have always been focusing on the evaluation of state-of-the-art methods for the segmentation of brain tumors in multi-modal magnetic resonance imaging (MRI) scans. Automatic segmentation of brain tumor has great values both in theory and practice, and has great potential to evaluate brain tumors with better diagnosis, surgical planning and treatment[1]. In 2D image processing fields, many effective models were proposed. AlexNet, presented by Krizhevsky et al[2], won the image classification task of ImageNet 2012.

The method of deep learning aroused researchers' attention. Later, Deep learning models have been kept explosive growth. VGGNet[3], used a series of small convolution filters to substitute for large convolution filters. GoogleNet [4] proposed a multi scale concept, by using different size filters to extract information, and its improved version Inception[5], creatively used 1*1 convolution filters to reduce the number of model parameters, while ensuring the model depth without increasing the parameters of the model. Squeeze and Excitation Networks[6], a kind of attention mechanism, introduced the Attention mechanism into the spatial dimension, further improving the performance of the model. However, using multi-modal magnetic resonance images to segment human tissue has been very challenging. Because medical image data is more complex than ordinary image data, both plane information and spatial information should be considered. So some researchers try to solve the problem of medical image segmentation by using deep learning method. In the first attempt, the modified variants of 2D CNN was adopted, by using aggregated adjacent slices[7] or orthogonal planes[8, 9], but this method did not take into account space information, it couldn't segment object accurately. Recently, a variety of 3D models had been developed to segment object from volumetric data and gained competitive performance. For examples, 3D U-Net [10] allows end-to-end training and testing for volumetric image segmentation. VoxResNet [11], a deep voxelwise residual network, improves the volumetric segmentation performance by seamlessly integrating the low-level image appearance features, implicit shape information and high-level context together.

The contribution of this paper are four-fold. First, we combine the mainstream segmentation models of 2D CNNs[12] and modified Inception structure to deal with 3D images. In the process of designing the model, we also consider the computation performance, and design two kinds of Inception layer, which are named as Lower Inception and Higher Inception. Second, we apply the 3D Squeeze and Excitation structure to our model. Third, we use multi-scale filters to downsample the 3D feature maps, the loss of valid information can be better reduced when resizing the 3D feature maps. Fourth, our model uses the residual connection to make sure the information can transfer better and the training process of the model can be accelerated.

2 Methods

2.1 Cascaded Framework

The cascaded framework of the model is designed to simplify segmentation problems[13, 14]. We use triple cascaded networks to segment substructures of brain tumor, each network can be seen as a binary segmentation network. While the first network segments the whole tumor task according to the MRI, a bounding box of the whole tumor is obtained. The region of the input images is cropped based on the bounding box, and the cropped result is used as the input of the second network to segment tumor core. After segmenting tumor core, another smaller bounding box is obtained. The image region is resized according to the

smaller bounding box of the tumor core. Then the resized image region is used as the input of the third network to segment the enhancing tumor core. During the training phase, the bounding boxes are decided by the ground truth. In the testing stage, the bounding boxes are generated based on the segmentation results.

2.2 Neural Networks Architecture

The overall architecture of the model we proposed is shown in Fig.1. It contains a great deal of 2D image mainstream model structures. Considering the huge advantages of their own structure in 2D images, modifying them to adapt the 3D medical image data can have better effects.

Low-Level Inception The Low-level Inception structure is shown in Fig.2. Why do we design model like this? There are several model general design principles[12]. The first principle is to avoid representational bottlenecks, especially early in the network. Any feed-forward networks can be seen an acyclic graph from input to output. Once the model is defined, the flow direction of information will be decided. When the information passes the model, information is fading. To avoid bottlenecks with extreme compression, we use the multi large receptive fields early in the network. In AlexNet[2], Krizhevsky et al use the 11*11 receptive fields. However large convolution filters have a serious shortcoming, i.e., large convolution filters have a huge number of training parameters. The parameters of 7*7 receptive fields are 5 times as much as the parameters of 3*3 receptive fields. But large receptive fields can better capture the space information. We should consider the trade-off between computation performance and model complexity, so we apply the large convolution filters only to the first four layers. We use different multi-scale size filters to better capture the space information, while avoiding representation bottleneck.

3D Squeeze and Excitation Structure Squeeze and Excitation structure was proposed by Jie Hu et al[6] in 2017, they used the SENet to get a top performance in the ImageNet 2017. Although there is not much innovation in the structure of SENet, two concepts of squeeze and excitation were put forward. In recent years, convolution neural networks have made great breakthroughs in many fields. The convolution kernel, as the core of the convolution neural network, is often seen as an aggregations of information aggregated on the spatial information and the feature dimension (channel-wise) on the local receptive field. However, it is quite difficult to learn a very strong network, and its difficulties come from many aspects. Then it is natural to think that this network can be modified to improve performance, such as considering the relationship between feature channels. Based on this point, Squeeze-and-Excitation Networks (SENet) is proposed. The motivation is to explicitly model the interdependence between feature channels. Specifically, it is important to acquire each characteristic channel automatically through the way of learning, improve the useful features and

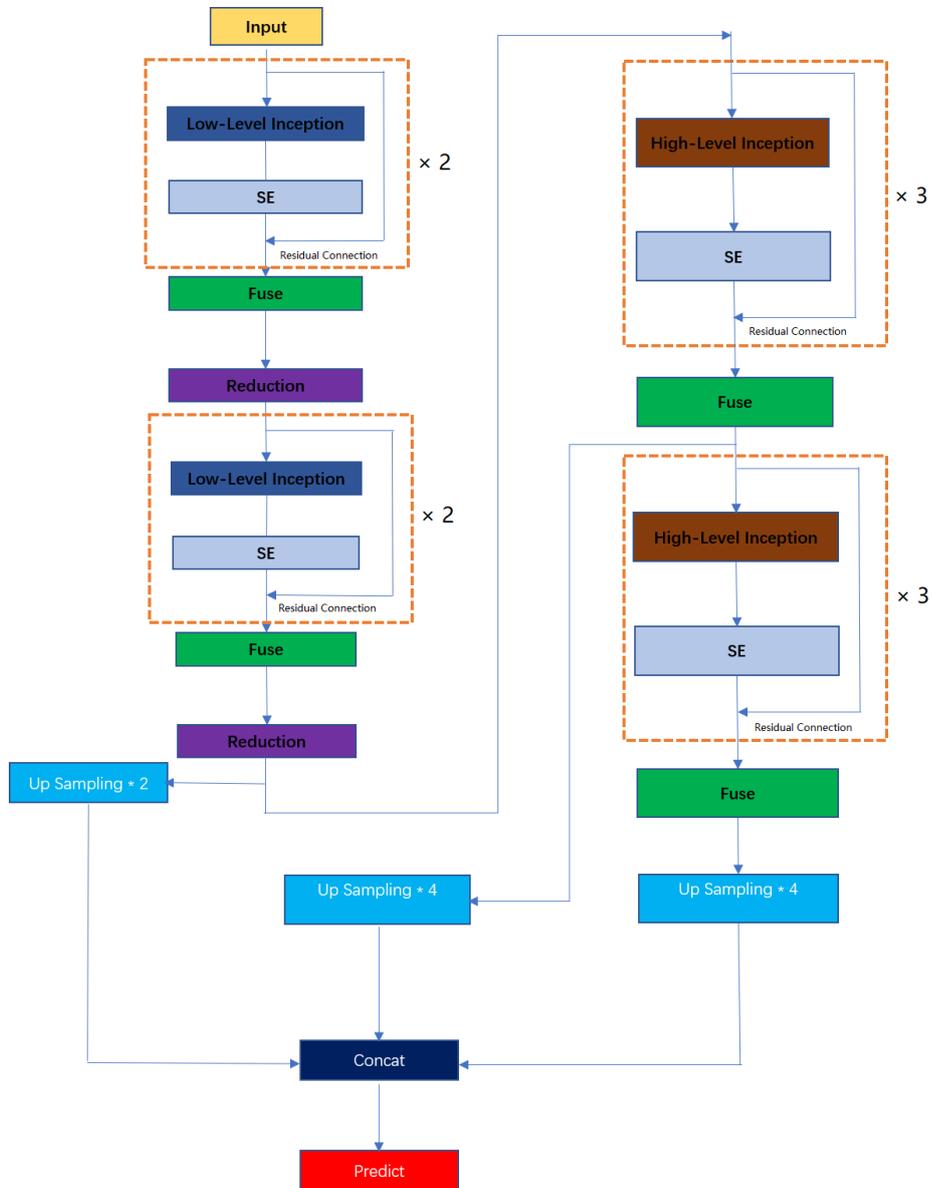


Fig. 1. This is the architecture of the model we proposed. It includes inception layers, SE structures, reduction layers and residual connection. High-Level Inception uses dilated convolution.

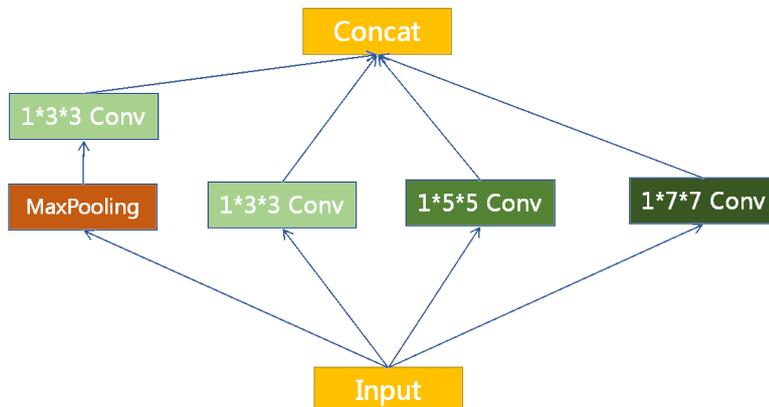


Fig. 2. We use 1*3*3, 1*5*5 and 1*7*7 convolution filters to better capture the information of feature maps early in the networks.

restrain the small features of the current task in accordance with its importance. Based on this idea, we redefine the squeeze and excitation operation in our model. For any given transformation $F_{tr} : X \rightarrow U$, $X \in \mathbb{R}^{D' \times W' \times H'}$, $U \in \mathbb{R}^{D' \times W' \times H'}$. We take F_{tr} as a standard 3D convolution operator. $V = [V_1, V_2, \dots, V_C]$ denotes the learned set of filter kernels, where V_C refers to the parameters of the c -th filter. We denote $U = [u_1, u_2, \dots, u_c]$ as the output of F_{tr} , where

$$u_c = v_c * X = \sum_{s=1}^{C'} v_c^s * x^s \quad (1)$$

Here $*$ denotes convolution operation, $v_c = [v_c^1, v_c^2, \dots, v_c^{C'}]$ and $X = [x^1, x^2, \dots, x^{C'}]$, v_c^s is a 3D spatial kernel, and therefore represents a single channel of v_c , which acts on the corresponding channel of X .

3D Squeeze: We perform feature compression along the space dimension, turning each of the three dimensional characteristic channels into a real number. This real number has a global receptive field to some extent, and the output dimension matches the number of input characteristic channels. It represents the global distribution of responses on characteristic channels. Moreover, the whole receptive field can be obtained near the input layer.

$$z_c = F_{sq}(u_c) = \frac{1}{D \times W \times H} \sum_{i=1}^D \sum_{j=1}^W \sum_{k=1}^H u_c(i, j, k) \quad (2)$$

Here, a statistic $z \in \mathbb{R}^c$ is generated by shrinking U through spatial dimensions $D \times W \times H$, z_c denotes the c -th element of z .

3D Excitation: Excitation operation is a mechanism similar to recurrent neural network's middle gate. Parameters are used to generate weights for each characteristic channel, the parameters are learned to explicitly model the correlation between feature channels. To meet these criteria, we use the sigmoid activation as a simple gating mechanism:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (3)$$

where δ refers to the *ReLU* function[13], $W_1 \in \mathbb{R}^{\frac{c}{r} \times C}$ and $W_2 \in \mathbb{R}^{C \times \frac{c}{r}}$. After *ReLU* function, we add two fully-connected layers to limit model complexity and aid generalisation. r denotes the reduction ration.

Output: The final output is a reweight operation. It's obtained by rescaling the transformation output U with the activations:

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \quad (4)$$

where $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_c]$, and $F_{scale}(u_c, s_c)$ refers to channel-wise multiplication between the feature map $u_c \in \mathbb{R}^{D \times W \times H}$ and the scalar s_c . The SE structure is shown in Fig.3.

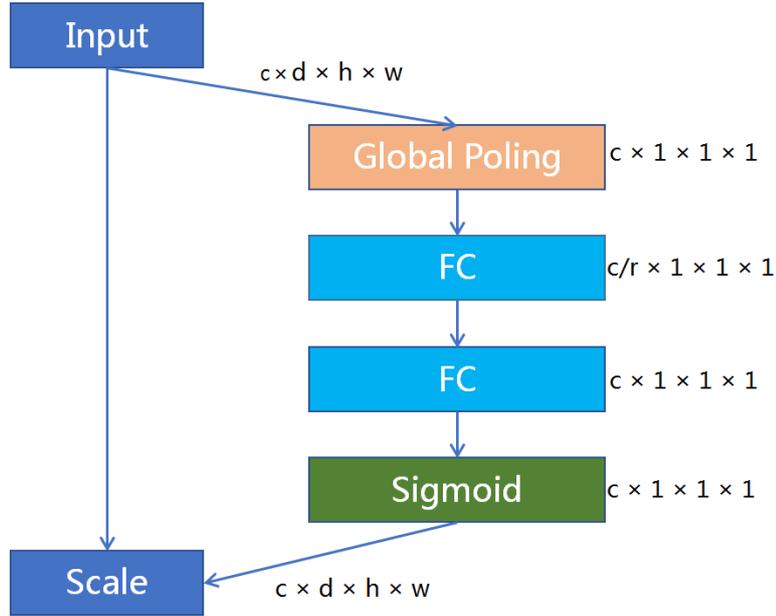


Fig. 3. 3D SE structure is a kind of attention mechanism that can pay attention to 3D channels relationship. c denotes the channels, r denotes ration (In our model, the ration $r=4$).

Reduction Structure Reduction structure is used for reduction feature maps. As mentioned before, using multi-scale can capture more spatial information. Different variants of this blocks (with various number of filters) can set by users, here we set the number of m , n , o , k and I as 8. As shown in Fig.4, we use $1*1*1$ convolution. In the design principles we mentioned earlier[15], the second principles is intent to let the spatial aggregation be done over lower dimensional embeddings without affecting representational power. Considering that these signals are easy to be compressed, dimensionality reduction will speed up the learning process. We redesign the reduction structure according to this idea.

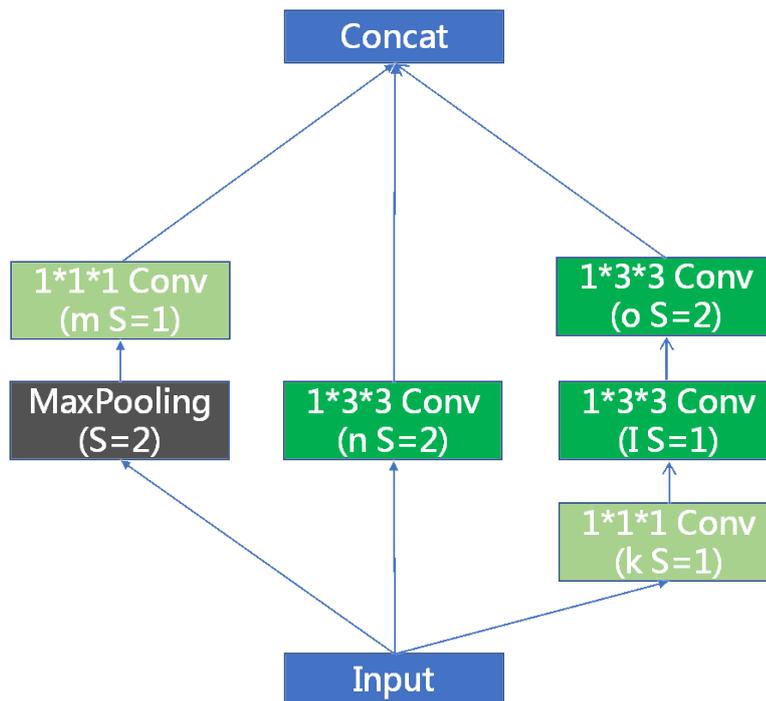


Fig. 4. Reduction Structure can use multi-scale convolution to capture the information from input feature maps. m , n , o , k and i can be set arbitrarily. We consider the simplified model, so set all the variables to the same number 8.

High-Level Inception The High-level Inception structure is shown in Fig.5. The third principles is to factorize into smaller convolutions. Convolutions with large filters have a huge computation complexity. For example, in the case of the same number of convolution kernel, $1*5*5$ convolution is $25/9 = 2.78$ times more computationally complex than that of $1*3*3$. But simply reducing the size of the convolution core will cause information loss. However $1*5*5$ convolution

can be replaced by multi-layer small convolution networks. Look at the $1*5*5$ network as full convolution, each output is a convolution kernel slipping on the input, it can be replaced by a two $1*3*3$ convolutional layer. The High-Level Inception architecture we designed can be seen in Fig.5.

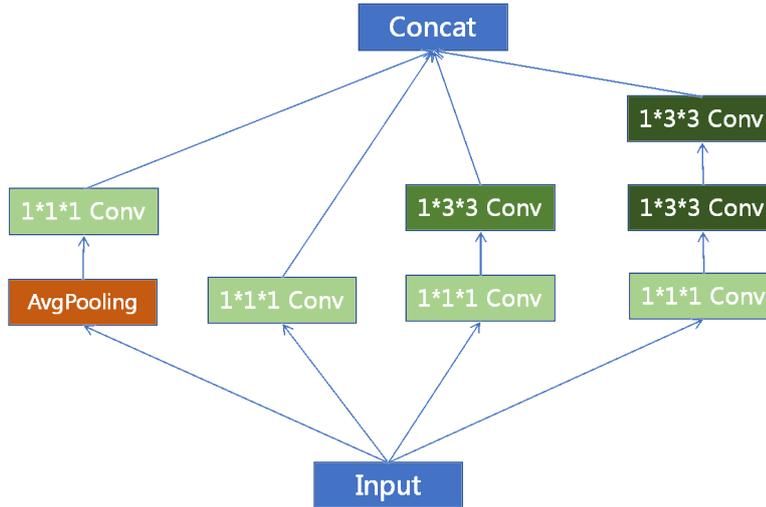


Fig. 5. The convolution of High-Level Inception uses the dilated convolution kernels. The dilated convolution use small filters but has a larger receptive fields, without increasing the parameters. We set the dilation rate 1,2,3 and 3,2,1 corresponding to each High-Level Inception layers in order.

Residual Connection ResNet was put forward in 2015 by Kaiming He et.al[16], it won the first place in the classification competition of ImageNet. With the increasing of network depth, the problem of the disappearance of the gradient is becoming more and more obvious. The training of the network has become quite difficult. The basic idea of ResNet is to introduce "shortcut connection" that can skip one or more layers. ResBlock can be defined as:

$$y = F(x, w_i) + x \quad (5)$$

Here x and y are the input and output vectors of the layers considered. The function $F(x, w_i)$ represents the residual mapping to be learned. If the dimensions of x and F don't equal, we can perform a linear projection W_s by the shortcut connections to match the dimensions:

$$y = F(x, w_i) + W_s x \quad (6)$$

W_s is used only when matching dimensions.

Prediction and Fusion In the prediction phase, we not only use the final result but also use the intermediate output results, and concatenating them as the final prediction result. In the training phase, each neural network is trained in axial, sagittal and coronal views. During the test phase, predictions are fused to get the final segmentation. We average the softmax outputs in these cascade networks. Fusion structure is a simple $3*1*1$ convolution. The model decomposes $3*3*3$ convolution kernels to $1*3*3$ convolution and $3*1*1$ convolution.

3 Experiments and Results

Brain tumor segmentation is a challenging task, which has attracted a lot of attentions in the past few years. We use the BRATS 2018 dataset [17, 18], which is composed of multiple segmentation subproblems. The whole tumor region is identified in a set of multi-modal images, tumor core areas and active tumor regions, referred as "whole tumor", "tumor core" and "enhancing tumor core" as per[19].

Medical Image Data Brats 2018 dataset contains real volumes of 210 high-grade and 75 low-grade glioma subjects. For each patient, T1Gd, T1, T2, FLAIR and Ground Truth MR volumes are available. These 285 subjects are used in training set, and there are 66 other subjects as the validation dataset. Considering the unbalance distribution of the training data, we expand the LGG dataset 3 times based on the original one. When training the network, we randomly choose 5 subjects as the input. All of these volume average size is $155 * 240 * 240$, we resize the volume and extract the voxel of specified shape in the middle volume as the final training input.

Training Details Our network is implemented in Tensorflow and NiftyNet. We use Adam optimizer to train, and use PReLU[20] as the activation, set the batch-size = 5, weight decay = 10^{-7} , learning rate = 10^{-3} , max-iteration = $20k$. We train on the GTX 1080Ti GPU. For the data pre-processing, the images are normalized by the mean and standard deviation. And we use the Dice coefficient as the model loss function.

Segmentaion Results The detail results of our model predict are shown as follows.

Table 1. Dice coefficient results of our model's prediction.

Data Set	Dice-ET	Dice-WT	Dice-TC
Training	0.72867	0.88453	0.83376
Validation	0.7838	0.87815	0.80731

Table 2. Sensitivity results of our model's prediction.

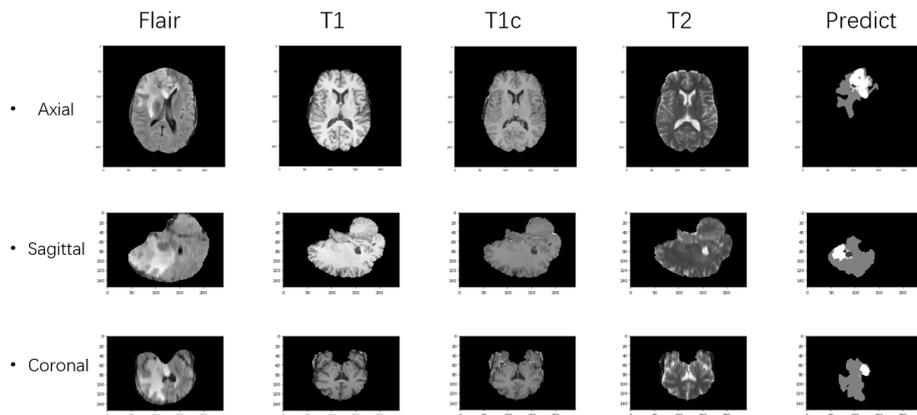
Data Set	Sensitivity-ET	Sensitivity-WT	Sensitivity-TC
Training	0.80547	0.92371	0.8778
Validation	0.8138	0.93498	0.84425

Table 3. Specificity result of our model's prediction.

Data Set	Specificity-ET	Specificity-WT	Specificity-TC
Training	0.99814	0.99163	0.99582
Validation	0.99817	0.99096	0.99665

Table 4. Hausdorff95 results of our model's prediction.

Data Set	Hausdorff95-ET	Hausdorff95-WT	Hausdorff95-TC
Training	5.65727	16.99984	7.08216
Validation	4.37991	19.03361	9.40808

**Fig. 6.** Example of one patient's segmentation result of our model.

4 Conclusion

Due to the limitation of time, we don't perform enough parameter adjustment experiments and don't use other optimization algorithms. In addition, in the stage of the data processing, we don't test other volume size of the input data, only use single volume size. We believe that the results can be better if we perform enough experiments. In the future, we plan to integrate convolution CRFs[21] and self-attention of 2D image segmentation into our model.

Acknowledgements We would like to thank the team of Guotai Wang et.al[14] for their source code, with which we can focus more on the innovation of the model and do our work more easily. And we'd also thank the NiftyNet team, they developed the deep learning tools for processing medical image data, which made us more efficiently to construct our model. This work was supported by the Natural Science Foundations of China under Grants No.61463050, No.61702443, No.61762091, the NSF of Yunnan Province under Grant No. 2015FB113, the Project of Innovative Research Team of Yunnan Province.

References

1. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, and et al., "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," *Scientific Data*, vol. 4, no. 170117, p. 170117, 2017.
2. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
3. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: <http://arxiv.org/abs/1409.1556>
4. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, and et al., "Going deeper with convolutions," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
5. C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, inception-resnet and the impact of residual connections on learning," *CoRR*, vol. abs/1602.07261, 2016. [Online]. Available: <http://arxiv.org/abs/1602.07261>
6. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," *CoRR*, vol. abs/1709.01507, 2017. [Online]. Available: <http://arxiv.org/abs/1709.01507>
7. H. Chen, L. Yu, Q. Dou, L. Shi, V. C. T. Mok, and P. A. Heng, "Automatic detection of cerebral microbleeds via deep learning based 3d feature representation," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, April 2015, pp. 764–767.
8. A. Prason, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 246–253.

9. H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers, "A new 2.5d representation for lymph node detection using random sets of deep convolutional neural network observations," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, Eds. Cham: Springer International Publishing, 2014, pp. 520–527.
10. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation from sparse annotation," *CoRR*, vol. abs/1606.06650, 2016. [Online]. Available: <http://arxiv.org/abs/1606.06650>
11. H. Chen, Q. Dou, L. Yu, and P. Heng, "Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation," *CoRR*, vol. abs/1608.05895, 2016. [Online]. Available: <http://arxiv.org/abs/1608.05895>
12. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
13. P. F. Christ, M. E. A. Elshaer, F. Ettliger, S. Tatavarty, and et al., "Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3d conditional random fields," *CoRR*, vol. abs/1610.02177, 2016. [Online]. Available: <http://arxiv.org/abs/1610.02177>
14. G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi, S. Bakas, H. Kuijf, B. Menze, and M. Reyes, Eds. Cham: Springer International Publishing, 2018, pp. 178–190.
15. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
16. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
17. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, "Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection," *The Cancer Imaging Archive*, 2017, DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q.
18. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, "Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection," *The Cancer Imaging Archive*, 2017, DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF.
19. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, and R. Wiest, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, p. 1993, 2015.
20. K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.
21. M. T. T. Teichmann and R. Cipolla, "Convolutional crfs for semantic segmentation," *CoRR*, vol. abs/1805.04777, 2018. [Online]. Available: <http://arxiv.org/abs/1805.04777>

3D Dense U-Nets For Brain Tumor Segmentation

Xiaoyue Zhang¹, Weijian Jian² and Kun Cheng³

¹ QED Technologies. Co., Ltd, Beijing, China

² Beihang University, Beijing, China

³ Beijing University of Posts and Telecommunications University, Beijing, China
laurelcheung@126.com

Abstract. Segmentation of brain glioma is a challenging task in field of medical image processing due to its diversity of intensity and complex shapes. This paper presents a method which combines U-net and DenseNet to efficiently segment the brain gliomas. This Dense U-net use skip connections densely which increases the number of convolutional layers to improve the performance and avoid overfitting. We use a two-step strategy: firstly segment whole the tumor from a low resolution volume and then feed with tumor patch to second step which refine segmentation. Preliminary results from the current version on validation data had mean dice coefficients of 0.75, 0.89, and 0.74 for enhancing tumor, whole tumor and tumor core respectively.

Keywords: BraTS, brain tumor segmentation, U-Net, DenseNet.

1 Introduction

Gliomas originating from glial cells are one type of the most common tumors in human brain. Accurate segmentation and measurement is needed since early diagnosis of gliomas is helpful prolonging the survival time of patients. However, the Magnetic Resonance Images of gliomas has plenty problems such as the unclear tumor boundaries, irregular shapes and image discontinuities [1-4].

With the improvements of computer hardware and network structure, deep learning algorithms, especially convolutional networks, have quickly become the preferred method for images processing. They are practical tools helping us to solve medical image analysis problems, especially for accurately detection, segmentation and lesion classification. In segmentation task, Jonathan Long [5] proposed that by rewriting the fully connected layers as convolutions, the FCN can produce a likelihood map, rather than an output for a single pixel. Ronneberger et al. (2015) [6] further developed the idea of FCN architecture and proposed the U-net. They use up-sampling to increase the image size, and add skip connections from the encoder features to the corresponding decoder activations. This model consists of two parts, the one is an encoder part where the spatial dimension of feature maps is gradually reduced and thus longer range information is more easily captured in the deeper encoder output, and the other is a decoder

part where object details and spatial dimension are gradually recovered. Cicek et al. (2016) [7] took a further step by modifying U-net to fit for 3D data.

Deeper model can always have better performance. However, the optimization problem of gradients vanishing prevent the model going deeper. ResNet [10] is proposed to solve this problem and makes it possible to train up to hundreds or even thousands of layers and still achieves compelling performance. Similar with ResNet structure, DenseNet [11] takes advantage of an observation that densely connected layers may leads a deeply, accurately and effectively performance by using a different connectivity pattern.

In this paper, we propose a network which combines the advantage of DenseNet and U-net to participate in 2018 BraTS Challenge. Every participant is required to complete a task to challenge multi-modality tumor segmentation for "enhancing tumor"(ET), the "tumor core"(TC), and the "whole tumor"(WT) of glioblastoma (GBM/HGG) and lower grade glioma (LGG) from pre-operative brain scans of patients. The experiment results show that Dense U-net can effectively cope with the optimization problem of gradients vanishing when training a 3D deep model, accelerate the convergence speed and simultaneously improve the precision of segmentation.

2 Methods

2.1 Data Preprocessing

In the applications of medical image processing, data augmentation is very important to avoid overfitting of the model due to the limited size of medical image dataset. The main augmentation method we use is elastic deformation. The elastic deformation was performed by defining a random smooth displacement field $u(x, y, z)$ [12]. Suppose R_o describes the location (x, y, z) of the pixels in the original volume, and R_w is the location (x', y', z') of the corresponding pixels in the warped volume. The relationship between R_o and R_w is defined as

$$R_w = R_o + \partial u \quad (1)$$

where ∂ is the strength of the displacement. To make the warped volume more similar to the real MR data, the displacement field u should be smoothed by Gaussian filter. The smoothness of the displacement field is controlled by the standard deviation σ of the Gaussian. Fig.1 shows the results of data augmentation by elastic deformation. Beside elastic deformation, we also use random crop, flipping along the left-right axis in the axial plane, zoom and etc. for data augmentation.

In order to reduce the effect of the absolute pixel intensities of the model input, intensity normalization step is applied to each volume by subtracting the mean and dividing by the standard deviation. So each input volume will have a zero mean and unit variance. Besides, zero-padding in MR data should be removed for the limit of GPU memory.

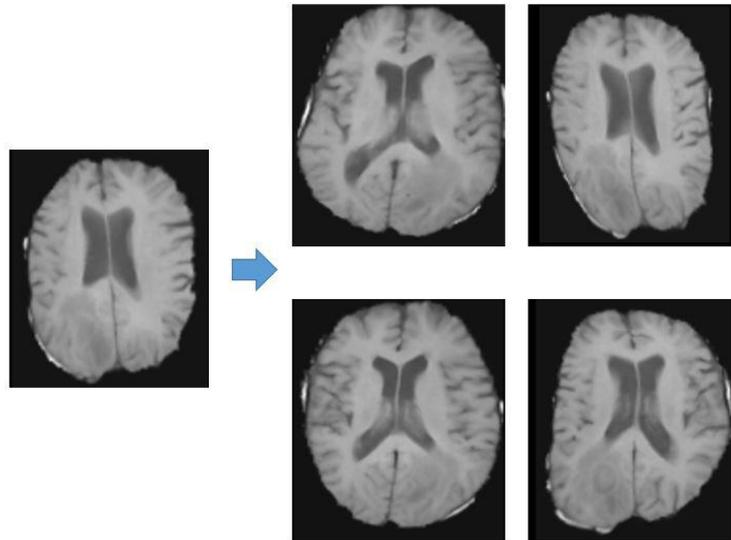


Fig. 1. The results of data augmentation by elastic deformation

2.2 Dense U-net

Current research has shown that if a convolution network contains a shorter connection between a near-input layer and a near-output layer, the convolution network can be trained more deeply, accurately and effectively. DenseNet takes advantage of this observation by using a different connectivity pattern. DenseNets have several compelling advantages: they alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters.

In the task of image segmentation, DenseNet based Fully Convolutional Networks (FCN) has been proved to be effective in natural image [13]. However, the traditional FCN only replace the fully connected layers with convolution operation and it's hard to achieve a satisfactory result on the details of segmentation. U-net, which use encoder to gradually enlarge the field of view and decoder recovers the object details, has been widely used in medical image segmentation. Here we insert dense blocks to U-net structures so that the network can be much deeper and more effective. Here is the details of Dense U-net:

Similar with traditional U-net, Dense U-net has two paths: a downward path (left) and an upward path (right). The networks have 4 levels encoder in the downward path, 4 levels decoder in the upward path and a base level. In the encoder path, each encoder level has a dense blocks (DB). Each layer in the dense block can use the feature-maps of all preceding layers as inputs, and use its own feature-maps as inputs into all subsequent layers. The dense blocks are made up of repeated Batch Normalization layers, ReLU, $3 \times 3 \times 3$ convolutions and concatenate operations (as shown in Fig.2). Zero padding is used at every layer to equalize input and output dimensions. Dropout at a rate of 50% was used after the fourth and fifth dense block. The output feature map has

$1 \times 1 \times 1$ convolution and pooling operation, then was passed to the next dense block. In the decoder path, each compact layer is consisted of two convolutional layers following by an upsampling layer. The upsampling layer uses simple resizing by trilinear interpolation. Like the common U-net, a feature map from the last layer of each pooling step is concatenated depth-wise with the upsampled layer.

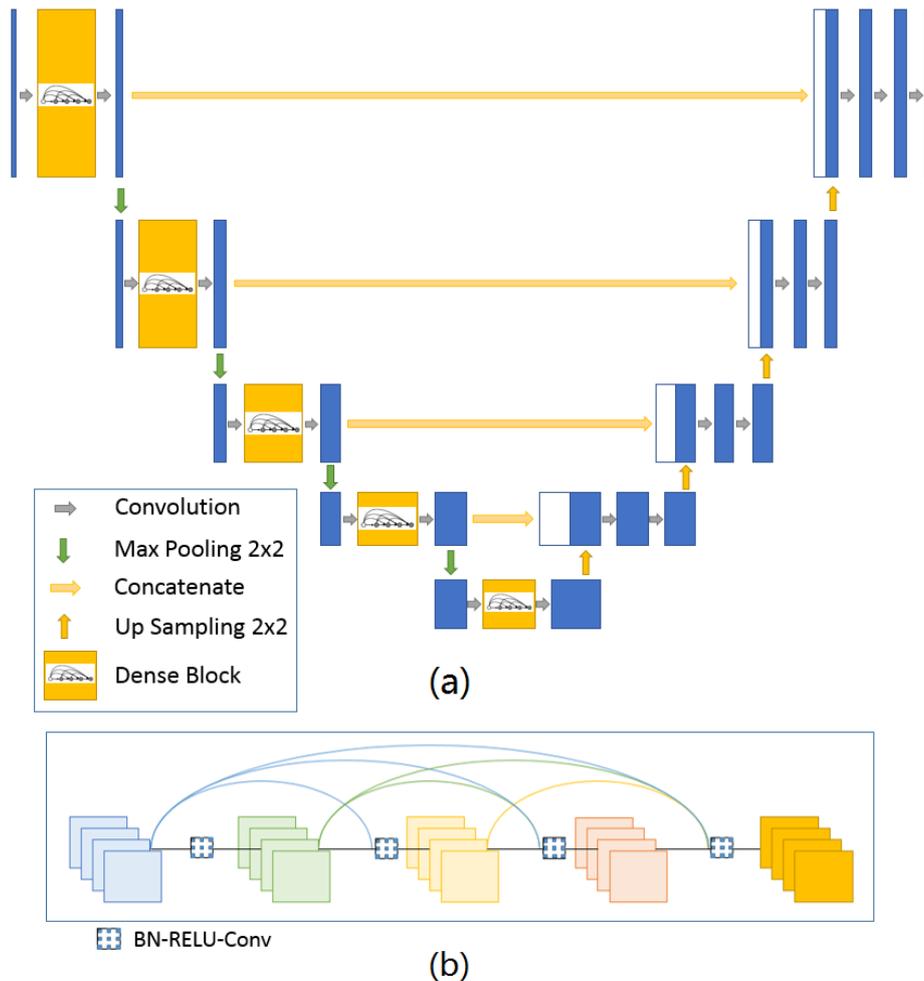


Fig. 2. (a) Structure of Dense U-net, dense blocks are used in each level of encoder part. (b) Structure of dense block

2.3 Two-Step Segmentation

The segmentation was divided into 2 steps: coarse segmentation and fine segmentation. Both steps are based on a 3-D Dense U-net structure. The whole procession is described in Fig.3. First of all, we prepare the input volumes containing four modalities: native

(T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR). For the limit of GPU memory, we crop and downsize the volume. In the first step, the output of coarse segmentation network is the mask of lesion area.

From the first step, we get the position of gliomas and then exact the slices which possibly containing lesion area. As input of step 2, the lesion slices is feed to the Fine Segmentation Network which generates the masks of enhancing tumor (ET), tumor core (TC) and whole tumor (WT). Some particularities of each network are explained in the following sub-sections.

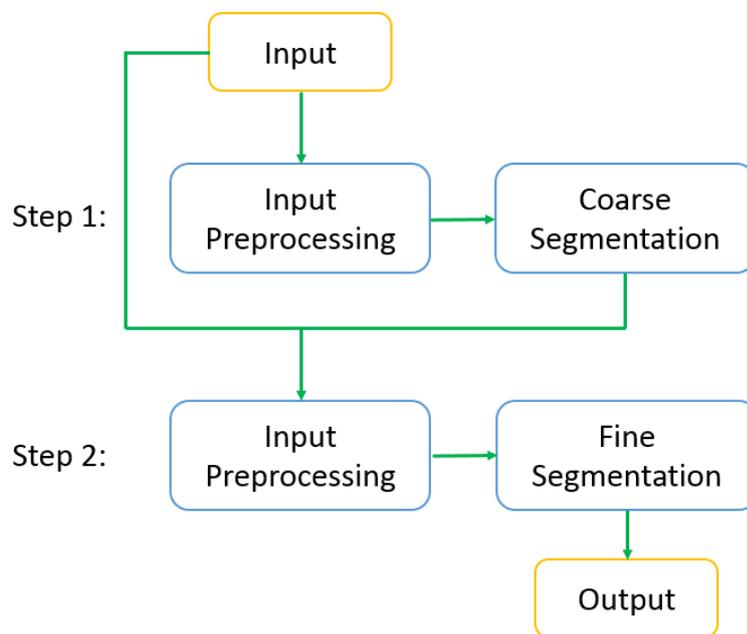


Fig. 3. Block diagram of two-step segmentation.

Coarse Segmentation Network This network input $112 \times 128 \times 64$ volume with 4 channels. This volume is generated by down sampling the original MR data. Due to the limitation of GPU memory (one GPU has memory of 12 GB), this down sampling is very necessary. The model is finished by using sigmoid function to distinguish background and lesion area. The output of the network is a one-channel $112 \times 128 \times 64$ volume.

Fine Segmentation Network This network have an input of 4-channel $144 \times 144 \times 48$ volume. It was noted that symmetry in axial view is an important cue for brain tumor segmentation as tumors usually break symmetric appearance in a healthy brain [15]. So we only remove the slices without tumor regions along z-axis. The final segmentation is operated by a convolutional layer followed by a softmax activation among the 4 classes.

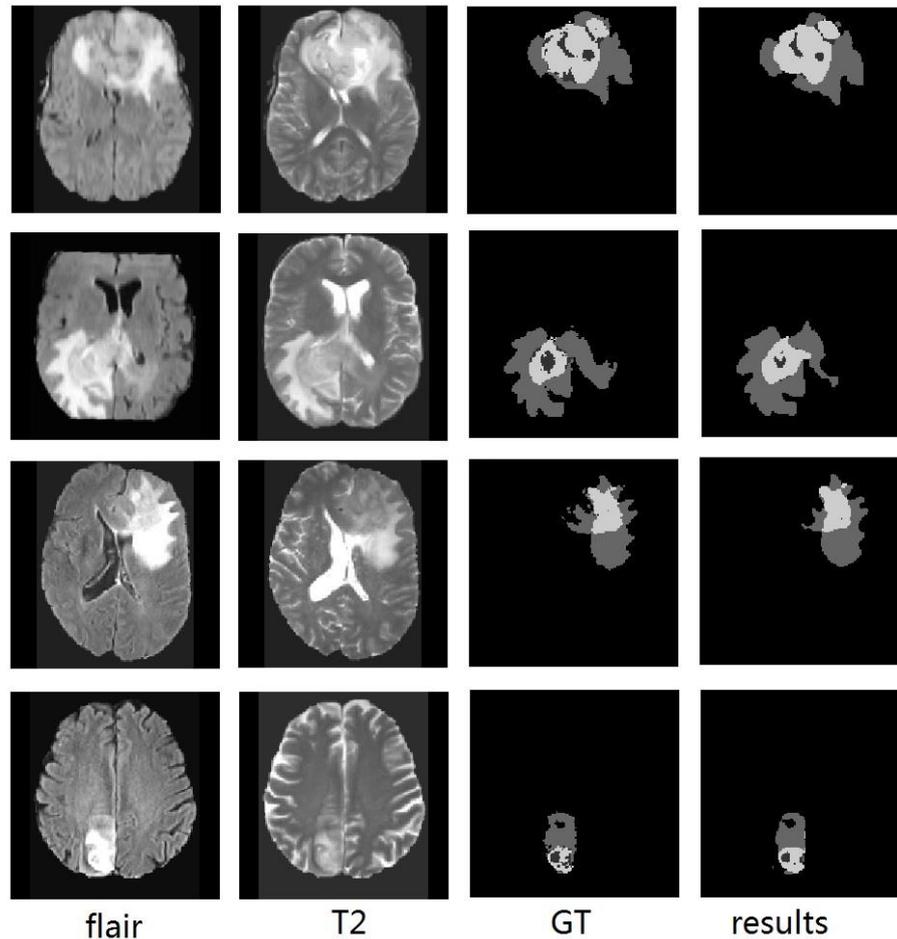


Fig. 4. A figure caption is always placed below the illustration. Short captions are centered, while long ones are justified. The macro button chooses the correct format automatically.

3 Experiments and Results

Training Details: The models were trained using Keras framework with TensorFlow backend and four NVIDIA TitanX GPUs. The models in Coarse Segmentation Network and Fine Segmentation Network have the same structure and the training takes 4 days on 4000 3D volumes with 4 channels. All models used a NADAM optimizer with a learning rate between 0.002 and 0.00002. For we have 4 GPUs, the batch size was set as 4. We used categorical cross-entropy as the loss function and weight decay (L2 weighting factor = 0.00001) for regularization. The network was trained for 30 epochs. 265 of the training datasets are used for model training and the rest 20 datasets are used for validation. With multiple submissions to the CBICA's Image Processing Portal, we

are able to compare the performances with other teams and improve our models. The results are tested on all 66 validation dataset. The evaluation metrics includes Dice Coefficient, Hausdorff Distance and Sensitivity and Specificity. The results are shown in Table 1 and Table 2. Our method performs well among the participations and there's a lot of indications showing that some improvements and further research is needed before it fulfills its potential.

Table 1. Dice and Sensitivity for BraTS 2018 validation dataset.

	Dice			Sensitivity		
	ET	WT	TC	ET	WT	TC
Mean	0.7486	0.8895	0.7395	0.7325	0.8803	0.7209
StdDev	0.2649	0.1212	0.2920	0.2727	0.1358	0.3023
Median	0.8517	0.9100	0.8671	0.8113	0.9181	0.8585
25quantile	0.7822	0.8849	0.6778	0.7102	0.8586	0.6206
75quantile	0.8742	0.9380	0.9212	0.8921	0.9473	0.9242

Table 2. Specificity and Hausdorff95 for BraTS 2018 validation dataset.

	Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC
Mean	0.9832	0.9806	0.9833	21.9397	7.2743	27.3519
StdDev	0.1229	0.1226	0.1229	78.6565	17.0189	78.3757
Median	0.9991	0.9967	0.9990	2.0000	3.2395	4.9268
25quantile	0.9981	0.9942	0.9973	1.4142	2.2361	2.2894
75quantile	0.9996	0.9981	0.9996	3.6723	5.0000	11.1298

4 Conclusion

Dense U-nets allow layers accessing to feature-maps from all of its preceding layers. This indicates that features extracted by shallow layers can be reused by every deeper layer. What's more, thanks to the U-net structure, the feature maps extracted by every dense block can also reused in the deeper layers. The efficiency of feature reuse is even higher and the number of feature maps can be even smaller. All in all, Dense U-nets can effectively cope with the optimization problems of gradients vanishing and improves the precision of segmentation. The experiments are done on BraTS Challenging 2018 validation data. Currently the result is far from perfect, but this networks have

huge potential. We believe that by parameters tuning the model can have a better performance and achieve an outstanding ranking score.

References

1. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Γ , Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftikharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
5. Long J, Shelhamer E, Darrell T: Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431-3440 (2015).
6. Ronneberger O, Fischer P, Brox T.: U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer, Cham, 234-241 (2015).
7. Çiçek Ö, Abdulkadir A, Lienkamp S S, et al.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 424-432 (2016).
8. Pereira, S., Pinto, A., Alves, V., et al.: Brain Tumor Segmentation Using Convolutional Neural Networks in MRI images. *Medical Imaging*, 35(5), 1240-1251 (2016)
9. Havaei, M., Davy, A., Warde-Farley, D., et al.: Brain Tumor Segmentation with Deep Neural Networks. *Medical Image Analysis*, 35, 18-31 (2017)
10. He K, Zhang X, Ren S, et al.: Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778 (2016).
11. Huang G, Liu Z, Weinberger K Q, et al.: Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1(2): 3 (2017).
12. Wong S C, Gatt A, Stamatescu V, et al.: Understanding data augmentation for classification: when to warp?. *arXiv preprint arXiv:1609.08764* (2016).

13. Jégou S, Drozdal M, Vazquez D, et al.: The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on. IEEE*, 1175-1183(2017).
14. Badrinarayanan, V., Kendall, A. and Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for scene segmentation. *IEEE transactions on pattern analysis and machine intelligence* (2017).
15. Shen H, Zhang J, Zheng W.: Efficient symmetry-driven fully convolutional network for multimodal brain tumor segmentation. *Image Processing (ICIP), 2017 IEEE International Conference on. IEEE*, 3864-3868(2017).

Learning Contextual and Attentive Information for Brain Tumor Segmentation

Chenhong Zhou^{1*}, Shengcong Chen^{1*}, Changxing Ding^{1**}, and Dacheng Tao²

¹ School of Electronic and Information Engineering,
South China University of Technology, Guangzhou, China
chxding@scut.edu.cn

² UBTECH Sydney AI Centre, SIT, FEIT, University of Sydney, Australia

Abstract. Thanks to the powerful representation learning ability, convolutional neural network has been an effective tool for the brain tumor segmentation task. In this work, we design multiple deep architectures of varied structures to learning contextual and attentive information, then ensemble the predictions of these models to obtain more robust segmentation results. In this way, the risk of overfitting in segmentation is reduced. Preliminary results on validation dataset of BRATS 2018 challenge demonstrate that the proposed method can achieve good performance with average DICE scores of 0.90, 0.83 and 0.79 for complete tumor, core tumor and enhancing tumor, respectively.

1 Introduction

Brain tumor is one of the most fatal cancers, which consists of uncontrolled, unnatural growth and division of the cells in the brain tissue [1]. The most frequent types of brain tumors in adults are gliomas that arise from glial cells and infiltrating the surrounding tissues [2]. According to the malignant degree of gliomas and their origin, these neoplasms can be categorized into Low Grade Gliomas (LGG) and High Grade Gliomas (HGG) [2, 3]. The former is slower-growing and comes with a life expectancy of several years, while the latter is more aggressive and infiltrative, having a shorter survival period and requiring immediate treatment [2]. Therefore, timely diagnosis of brain tumor is of critical importance for assisting the doctors to perform surgery and make treatment planning.

In recent years, convolutional neural networks (CNNs) have been widely applied to automatic brain tumor segmentation tasks. Pereira et al. [14] and Havaei et al. [12] respectively trained a CNN to predict the label of the central voxel only within a patch, which causes that they suffer from high computational cost and time consumption during inference. To reduce the computational burden, Kamnitsas et al. [5] propose an efficient model named DeepMedic that can predict the labels of voxels within a patch simultaneously, in order to achieve

* stands for equal contribution.

** corresponding author.

dense predictions. Recently, fully convolutional networks (FCNs) have achieved promising results. Shen et al. [6] and Zhao et al. [10] allow end-to-end dense training and testing for brain tumor segmentation at the *slice* level to improve computational efficiency. With a large variety of CNN architectures proposed, the performance of automatic brain tumor segmentation from Magnetic Resonance imaging (MRI) images has been improved greatly.

In this work, we construct multiple different CNN architectures and approaches to ensemble their prediction results, in order to create stable and robust segmentation performance. In this preliminary work, we evaluate our approaches on the validation set of 2018 Brain Tumor Segmentation (BRATS) challenge, where we obtain the good performance with average DICE scores of 0.90, 0.83 and 0.79 for complete tumor, core tumor and enhancing tumor, respectively.

2 Data

We use the training set of 2018 Brain Tumor Segmentation challenge [2, 4, 7, 8], which contains 210 high-grade glioma (HGG) and 75 low-grade glioma (LGG) cases, whose corresponding manual segmentations are provided. As shown in Fig. 1, the provided manual segmentations include four labels: 1 for necrotic (NCR) and the non-enhancing (NET) tumor, 2 for edema (ED), 4 for enhancing tumor (ET), and 0 for everything else, i.e. normal tissue and background (black padding). The validation set consists of 66 cases with unknown grade and hidden segmentations. The valuation for validation set is conducted via an online system. The official evaluation is calculated by merging the predicted labels into three regions: whole tumor (1,2,4), tumor core (1,4) and enhancing tumor (4). Besides, each patient has four MRI sequences that are named T1, T1 contrast enhanced (T1ce), T2 and FLAIR. These datasets are provided after their pre-processing, i.e. co-registered to the same anatomical template, interpolated to the same resolution (1 mm^3) and skull-stripped. Dimensions of each MRI sequence are $240 \times 240 \times 155$.

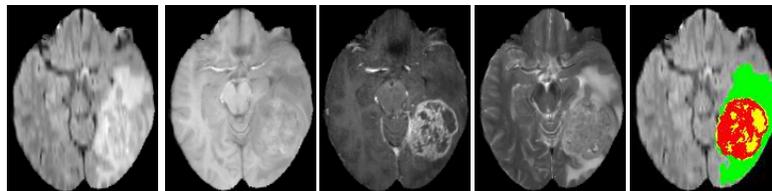


Fig. 1. Example of images from the BRATS 2018 dataset. From left to right: Flair, T1, T1ce, T2 and manual annotation overlaid on the Flair image: edema (green), necrosis and non-enhancing (yellow), and enhancing (red).

3 Methods

As is well known, brain tumor segmentation from MRI images is a very tough and challenging task due to the severe class imbalance problem. Following [13], we decompose the multi-class brain tumor segmentation into three different but related sub-tasks to handle the class imbalance problem. 1) *Coarse segmentation to detect complete tumor*. In this sub-task, the region of complete tumor is located. To reduce overfitting, we define the first task being the five-class segmentation problem. 2) *Refined segmentation for complete tumor and its intra-tumoral classes*. The above coarse tumor mask is dilated by 5 voxels as the ROI for the second task. In this sub-task, the intra-tumoral classes of the dilated region are predicted. 3) *Precise segmentation for enhancing tumor*. We specially design the third sub-task to segment the enhancing tumor, due to its difficulty of segmentation.

3.1 Model Cascade

In view of the above three sub-tasks, it is probably easy to train a CNN individually for each sub-task, which is the currently popular Model Cascade (MC) strategy. We use a 3D variant of the FusionNet [9], as illustrated in Fig. 2. The network architecture consists of an encoding path (upper half of the network) to extract complex semantic features and a symmetric decoding path (lower half of the network) to produce the same resolution as the input for voxel-to-voxel predictions. The network is constructed by four types of basic building blocks, as shown in Fig. 2. In addition, the network has not only the short shortcuts in residual blocks, but also three long skip connections to merge the feature maps from the same level in the encoding path during decoding by using a voxel-wise addition. We employ the identical network architecture for each sub-task, except for the final convolutional classification layer. The number of channels of last classification layer is equal to 5, 5 and 2 for the first, second and third sub-tasks, respectively. Size of input patches for the network is $32 \times 32 \times 16 \times 4$, where the number 4 indicates the four MRI modalities.

3.2 Multi-scale Contextual Information

To deal with the 3D medical scans, we employ the above 3D CNNs that process small 3D patches. However, small patches cause the network to lean the limited local information. It seems necessary to introduce large contextual information, in order to provide larger receptive fields and more contextual information to the network. Therefore, based on the above 3D CNNs, we design a two parallel pathway architecture that processes the input images at multiple scales simultaneously. As shown in Fig. 3, we incorporate both local and larger contextual information to the above 3D CNNs, which not only extracts semantic features at a higher resolution, but also considers larger contextual information from lower resolution levels. It can provide rich information to discriminate voxels that appear very similar when considering only local appearance, avoiding making wrong predictions.

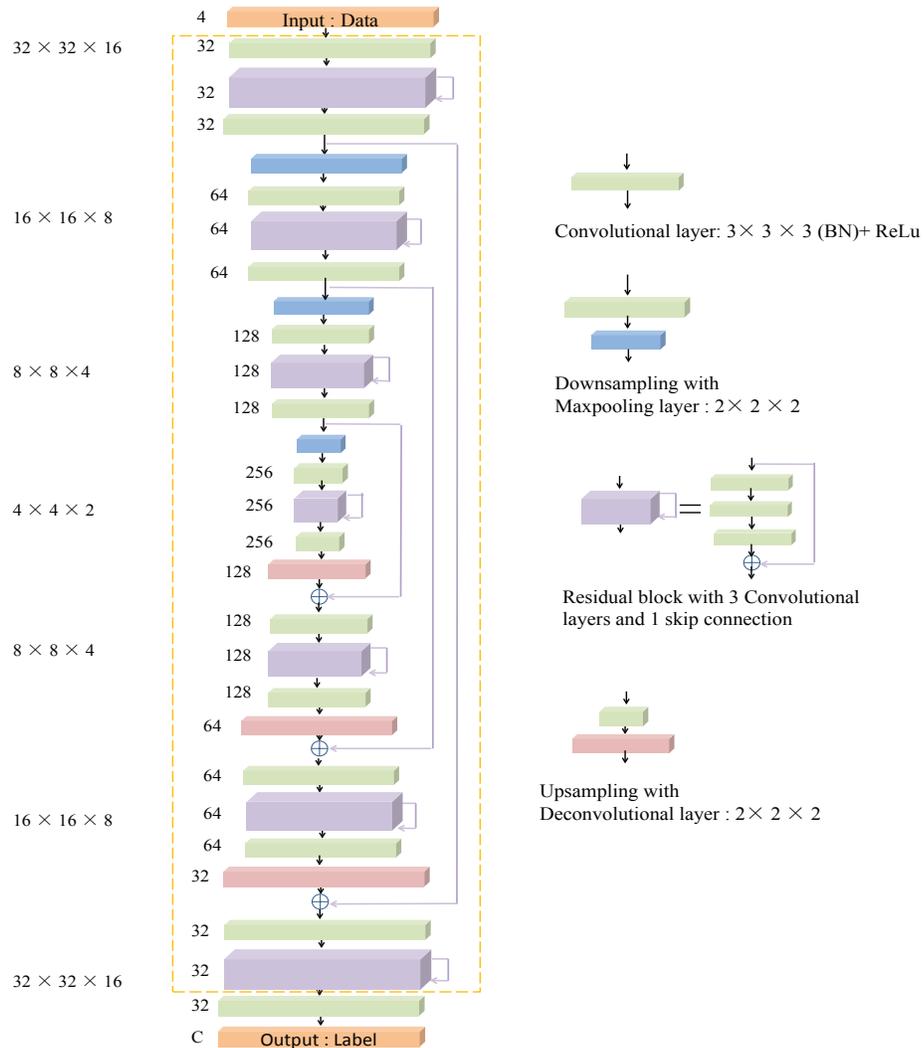


Fig. 2. Network architecture used in each sub-task. The building blocks are represented by colored cubes with numbers nearby being the number of feature maps. C equals to 5, 5, and 2 for the first, second, and third task, respectively. (Best viewed in color) This figure is reproduced from [13].

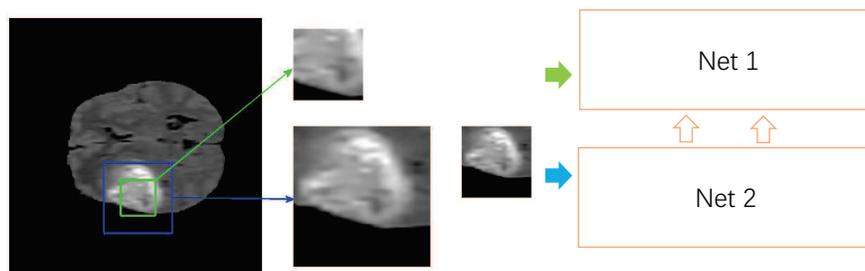


Fig. 3. The proposed network architecture to introduce multi-scale contextual information.

3.3 One-pass Multi-task Network

The above proposed model cascade approaches have obtained promising segmentation performance, which is attributed to relieve the class imbalance problem. Evidently, model cascade approaches need to train a series of deep model individually for the three different sub-tasks. However, they would lead to large memory cost and system complexity in practice during training and testing. In addition, we observe that the networks used for three tasks are almost the same except for the training data. It is obvious that the three sub-tasks are relative to each other.

Therefore, we employ the one-pass multi-task network (OM-Net) proposed in [13], which is a multi-task learning framework that incorporates the three sub-tasks into an end-to-end holistic network, to save a lot of parameters and exploit the underlying relevance among the three sub-tasks. The OM-Net proposed in [13] is described in Fig. 4, which is composed of the sharable parameters and task-specific parameters. Specially, the shared backbone model refers to the network layers outlined by the yellow dashed line in Fig. 2, while three respective branches for different sub-tasks are designed after the shared parts.

In addition, inspired by the curriculum learning theory proposed by Bengio et al. [11] that humans can learn a set of concepts much better when the concepts to be learned are presented by gradually increasing the difficulty level, we adopt the curriculum learning-based training strategy in [13] to train OM-Net more effectively, so that data sharing and parameters sharing are achieved in a single deep model. The training strategy of our framework is to start training the network on the first easiest sub-task, then gradually add the more difficult sub-tasks and their corresponding training data to the model. This is a process from easy to difficult, highly consistent with the thought of manual segmentation of the tumor. Besides, the training data conforming to the sampling strategy of the other sub-tasks can be transferred to achieve data sharing. Eventually, the OM-Net is a single deep model to solve three sub-tasks simultaneously in one-pass.

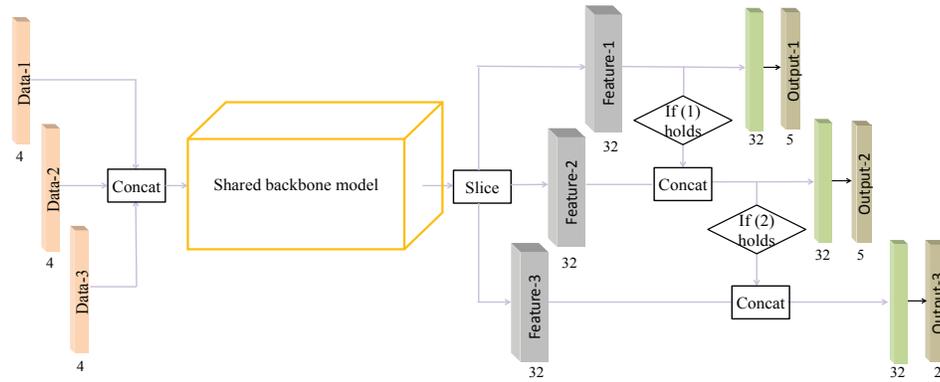


Fig. 4. Architecture of OM-Net. Data- i , Feature- i , and Output- i denote training data, feature, and classification layer for the i -th task, respectively. The shared backbone model refers to the network layers outlined by the yellow dashed line in Fig. 2. This figure is reproduced from [13].

3.4 One-pass Multi-task Network with Attention Mechanisms

Convolutional neural networks have powerful ability to extract informative features by fusing rich information together within local receptive fields through the convolution operation. To boost the representational power of a CNN, Squeeze-and-Excitation (SE) block is proposed to adaptively perform channel-wise feature recalibration by explicitly modelling interdependencies between channels in [15]. Inspired by it, we employ the SE blocks to the OM-Net, in order to re-calibrate the feature maps and further improve the learning and representational properties of OM-Net. The SE block is described in Fig. 5. Similar to [15], the SE block focuses on channels to adaptively recalibrates channel-wise feature responses in two steps, squeeze and excitation. It helps the network to increase the sensitivity to informative features and suppress less useful ones.

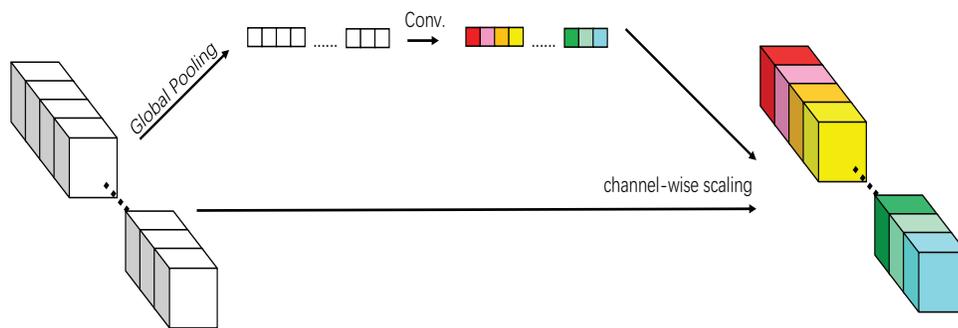


Fig. 5. The adopted Squeeze-and-Excitation (SE) block.

3.5 Ensembles of the Above Multiple Models

By bagging together all the above models, we obtain more robust performance to fuse the segmentation results from the above multiple CNNs.

4 Experiments

4.1 Pre-processing

We adopt the minimal pre-processing operation to the BRATS 2018 data. That is, each sequence is individually normalized by subtracting its mean value and dividing by its standard deviation of the intensities within the brain area in that sequence.

4.2 Results

Preliminary results on the validation set of 2018 BRATS challenge were presented in Table 1.

Table 1. A summary of the individual net results and the results of ensembles (%)

Method	DSC			Sensitivity			Specificity			Hausdorff95		
	Enh.	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core
MC-Net (multi-scale)	77.65	90.74	82.26	83.19	90.80	82.08	99.75	99.52	99.76	3.72	4.48	6.37
OM-Net	78.82	90.34	82.73	79.53	90.84	78.83	99.84	99.48	99.85	3.10	6.52	7.20
OM-Net with attention	79.10	90.12	83.34	80.87	90.62	80.60	99.82	99.48	99.82	3.86	15.85	17.78
Ensembles	79.22	90.74	83.58	81.35	90.80	81.01	99.82	99.52	99.82	2.80	4.48	7.07

5 Conclusion

In this work, we propose to ensemble multiple CNNs to produce robust performance for brain tumor segmentation. In the future, we would investigate the more powerful network architectures and post-processing to further promote the segmentation performance.

References

1. A. Işın, C. Direkoğlu, and M. Şah, “Review of MRI-based brain tumor image segmentation using deep learning methods,” *Procedia Comput. Sci.*, vol. 102, pp. 317–324, 2016.
2. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, et al. ”The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)”, *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694

3. S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies", *Phys. Med. Biol.*, vol. 58, no. 13, pp. R97–R129, 2013.
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
5. K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Imag. Anal.*, vol. 36, pp. 61–78, 2017.
6. Shen, H., Wang, R., Zhang, J., McKenna, S.J.: Boundary-aware fully convolutional network for brain tumor segmentation. In: MICCAI, Part II. LNCS, vol. 10434, pp.433–441 (2017)
7. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
8. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
9. T. M. Quan, D. G. Hilderbrand, and W.-K. Jeong, "Fusionnet: A deep fully residual convolutional neural network for image segmentation in connectomics." [Online]. Available: <https://arxiv.org/abs/1612.05360>
10. X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Med. Imag. Anal.*, vol. 43, pp. 98–111, 2018.
11. Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proc. 26th Int. Conf. Mach. Learn. (ICML)*, 2009, pp. 41–48.
12. M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Med. Imag. Anal.*, vol. 35, pp. 18–31, 2017.
13. C. Zhou, C. Ding, et al. "One-pass Multi-task Convolutional Neural Networks for Efficient Brain Tumor Segmentation," In: MICCAI 2018 [In Press].
14. S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1240–1251, 2016.
15. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," *arXiv preprint arXiv:1709.01507*, vol. 7, 2017.