MULTIMODAL BRAIN TUMOR SEGMENTATION USING VARIOUS 2-D U-NET ARCHITECTURES

Ahmed Gouda, Tewodros W. Arega, Zohaib Salahuddin

Università degli Studi di Cassino e del Lazio Meridionale

ABSTRACT

Brain Tumor Segmentation is a challenging task because of tumor variability and heterogeneity. Multi-modal Magnetic Resonance Images (MRIs) are used for the diagnosis and segmentation of gliomas. Manual Brain Tumor Segmentation not only requires extensive knowledge of the anatomy but is also susceptible to human error. Hence, automatic brain tumor segmentation technique will help counter these difficulties in diagnosis and treatment. In this paper, we implemented 2D U-Net Architectures to segment the tumor into three subregions. The results of the various U-Net architectures for segmentation of brain tumors into whole tumor, tumor core and enhancing tumor are very encouraging. Our 2-D U-Net Architecture with a traditional encoder provides very good dice cores of 0.86, 0.72 and 0.81 with augmentation while U-Net Architecture with VGG-16 encoder provides scores of 0.86, 0.73 and 0.67 for whole tumor, tumor core and enhancing tumor respectively.

***Index Terms*—** BraTS 2018, High Grade Glioblastomas, Low Grade Gliomas, VGG-16, 2D U-NET

1. INTRODUCTION

An estimated 86,970 new cases of primary malignant and non-malignant brain and other tumors are expected to be diagnosed in the United States in 2019. [1] Early detection is critical in the case of High Grade Glioblastoma (HGG) tumors as it has an average survival rate of less than 2 years while Low Grade Gliomas having survival rate of several years. Tumor detection and characterization is not only important during diagnosis but also during treatment. Manual delineation is a time intensive task prone to human error. Hence, Automatic Brain tumor Segmentation is of great significance. Deep Learning Techniques which have been disrupting the field of Computer Vision can be utilized for this problem.

In this paper, we investigate various 2-D U-Net Architectures for Brain Tumor Segmentation. The choice of 2D U-Net architecture was made due its superiority in semantic segmentation as well as limited processing power at hand. The BraTS 2018 data set consisting of 285 patients was utilized. It includes 210 patients having high grade glioblastomas and 75 patients having low grade gliomas. The data for each patient contains 155 axial slices and four modalites namely: FLAIR, T1, T2 and T1ce. [3] Labels 1,2 and 4 correspond to the necrotic and non-enhancing tumor core, the peritumoral edema and enhancing tumor respectively.Labels 1,2,4 combined give the whole tumor, labels 1 and 2 give the tumor core while label 4 gives the enhancing tumor.

The framework that we have used to train our model is **Keras** on top of **tensorflow**. The computing resources that we have used are **Kaggle**(1xTesla K80 GPU, 13GB GDDR5) and **Google Colab** (1xTesla K80, 12GB GDDR5) and one session runs 9 hours and 12 hours respectively.

2. METHODOLOGY

2.1. Data Pre-Processing

All the four modalities had to be normalized before training because of the difference in their intensity distribution. The pre-processing step consisted of applying *mean and standard deviation* normalization. The normalization was applied on each slice of the image separately. Mean and Standard deviation Normalization was only restricted to the brain region of each image. The black background of the image was not changed in any way. The slices that were blank and did not contain any information of the brain region were eliminated. In case for binary segmentation, the labels were converted into whole tumor, tumor core and enhancing tumor. However, in case of the multi-class segmentation, these labels were used as such.

2.2. Data Augmentation

The data from 285 patients already consisted of 176700 images. Due to the limited GPU power available at disposal, we had limited capacity to try limited augmentation. Various Augmentation techniques like rotation, flipping and scaling were experimented with. In our experiments, the data was augmented by constructing coronal and saggital slices from the already available axial data. We compared our results before and after data augmentation. Our final training set consisted of $176700 * 3$ images.

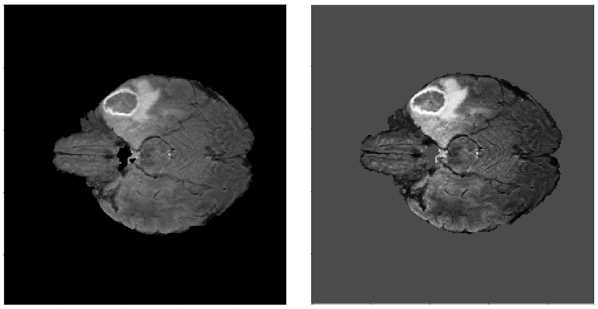

Fig. 1
The figure on the left shows original image of one of the four modalities and the figure on the right shows the normalized image.

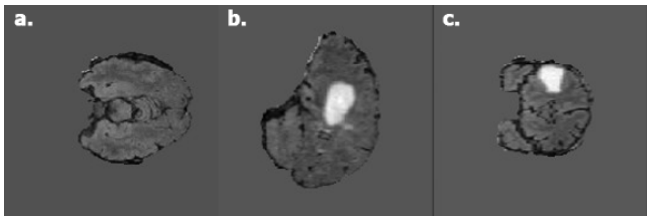

Fig. 2
From Axial Slice (a), saggital (b) and Coronal (c) were constructed for Data Augmentation.

2.3. Training

The dataset provided by BraTS includes 285 cases(210 HGG and 75 LGG), each with four 3D MRI modalities (T1, T1c, T2 and FLAIR). Annotations include 3 tumor subregions: the enhancing tumor, the peritumoral edema, and the necrotic and non-enhancing tumor core. The annotations were combined into 3 nested sub-regions: whole tumor (WT), tumor core (TC) and enhancing tumor (ET).[5, 6]

We divided our dataset into training, validation and testing. From the 285 patients,we chose 10 patients for testing (7 from HGG and 3 from LGG). The remaining 275 patients were split in the training validation ratio of $0.8 : 0.2$. *5-fold* cross-validation was applied to train the 2-D network without augmentation due to time constraints. The Batch size was constrained by the memory of the GPU with the upper limit being 48. After hyper-parameter tuning, the batch size was fixed to be 32. Experiments were made using dice/multi-dice [7] loss as well as weighed binary/categorical cross entropy. Dice loss was fixed for all the architectures because of its superior performance. Each Epoch took 30 mins to train without augmentation and 80 mins to train with augmentation. All the models were evaluated after 30 epochs. For Validation and Testing sets, the dice scores were evaluated separately for all the slices with the brain region in it. All the four modalities were stacked together to make the input image of dimensions $240 * 240 * 4$. Drop-out Regularization technique was employed in order to counter the over-fitting problem.

When training all models, we used a learning rate of *1e-4*. We chose this learning rate after we tried learning rates less than and greater than the above value. When the learning rate is very small, the model learns slowly and gets stuck on a sub-optimal value. If the learning rate is larger it learns fast but sometimes it oscillates over training epochs.Therefore, after many trails, we discovered that *1e-4* learning is the best one for our case.

3. MODEL ARCHITECTURES

3.1. Binary Segmentation - Combination of three 2D U-Net with VGG-16 Encoder models

VGG-16 architecture with its pre-trained weights from training on ImageNet dataset is used for image classification for RGB images. These pre-trained weights were used to initialize the VGG-16 encoder weights. VGG-16 is mainly used for RGB images but we modified the input to use all the four modalities for training at once(four channel image). All three classes are trained separately namely whole tumor, tumor core and enhancing tumor. The outputs of all the three models after training is combined to give the final result of the segmentation of three classes.

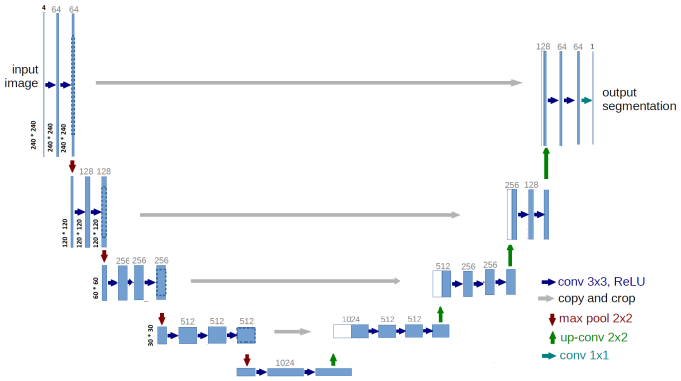

Fig. 3
UNET with VGG encoder

3.2. Multiclass Segmentation

For multiclass segmentation, multi-dice loss was used. The input was the same as used for the binary segmentation with the dimensions $240 * 240 * 4$. With three classes to segments: Whole Tumor, Tumor Core and Enhancing Tumor, we also set background as another class. Hence there were four classes in total. The inclusion of background as an extra class gives more weights to the tumor region segmentation. This helps in catering the class imbalance between the background class and the tumor regions.

3.2.1. *Traditional 2D U-Net*

The traditional 2D U-Net was utilized for multiclass segmentation. The choice of batch size as well as the encoder and decoder depth is limited by the amount of GPU memory available. Batch Normalization was added after each layer in the Encoder section of the U-Net. The Batch Normalization helped us in countering spurious changes in the weights which otherwise resulted in the creation of NAN in the layers. The results of this 2D U-Net were investigated before and after augmentation.

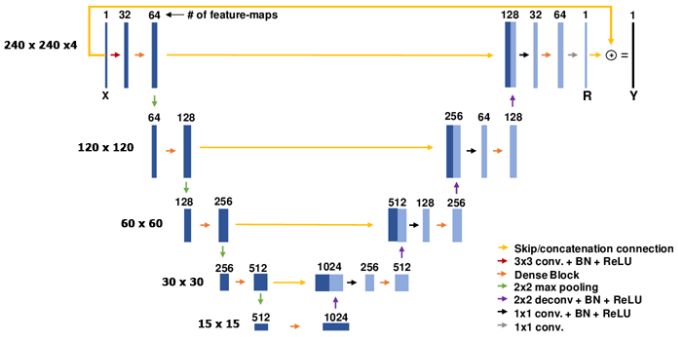

Fig. 4
This figure shows the traditional U-Net Architecture used for multiclass segmentation.[2]

4. RESULTS AND DISCUSSION

4.1. Results

The results shown by both models are encouraging with the multi-class UNET model with augmentation having the superior performance. However, this performance can be further improved by adding more kinds of augmentation and using other encoders like INCEPTIONV3, DENSENET encoders but we were limited by the computational power available at our end.It can be clearly seen that Multi-class 2D U-Net model has superior performance after careful augmentation. The combination of binary segmentation with VGG-16 2D U-Net performed better than Multi-class 2D U-Net without segmentation. Results can be improved by adding more augmentation and using more deep networks.

4.2. Problems Faced

4.2.1. *Large Dataset*

The large dataset of 285 patients consisting of 176700 images required a lot of computational power to train. All the four modalities had to be used at once because each modality contributed some specific information. All the images could not be loaded into the RAM at once hence a generator had to be utilized. Google Cloud Plateform and Kaggle were utilized for train. Since normalization had to be applied to the images, this large dataset had to be saved after applying normalization. Google Cloud Plateform becomes unresponsive after a day due to the presence of large amount of images. On the other hand, kaggle didn't allow enough storage. This limitation added 2-3 hours of data generation in each training session.

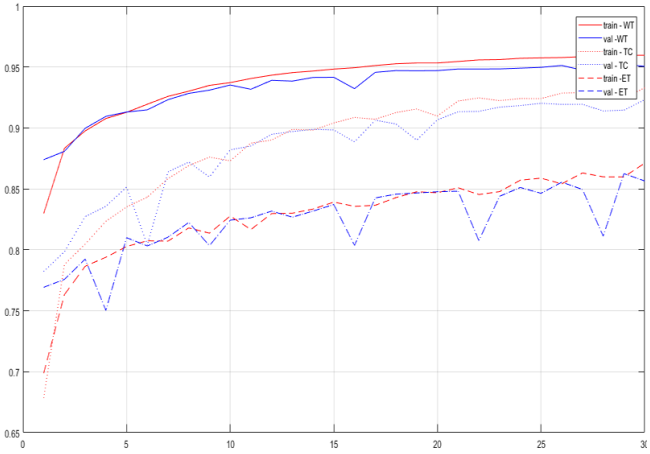

Fig. 5
This figure shows the training and validation dice coefficient vs Number of Epochs for VGG-16 Binary Segmentation. WT, TC and ET correspond to whole tumor, tumor core and enhancing tumor respectively

Table 1
Results of Various Models on the Testing Data

Model	**WT**	**TC**	**ET**
Binary: **VGG-16**	0.86	0.73	0.67
Multiclass: **U-Net**	0.83	0.74	0.69
Multiclass: **U-Net with Augmentation**	0.86	0.72	0.81
State of the Art	0.91	0.86	0.82

4.2.2. *Creation of NaNs in model weights - Bug in Keras*

During training, there was a sudden drop of validation loss and training loss to maximum because of the creation of NANs in the model weights. This issue was investigated to be an internal bug of keras. This sudden drop occurred varied with the learning rate. After extensive trials, this error was fixed by adding Batch Normalization in the encoder layers of the network.

4.2.3. *Limited in Computing Power*

After Augmentation, the total number of training images increases to $530,100$. The training time for each epoch using

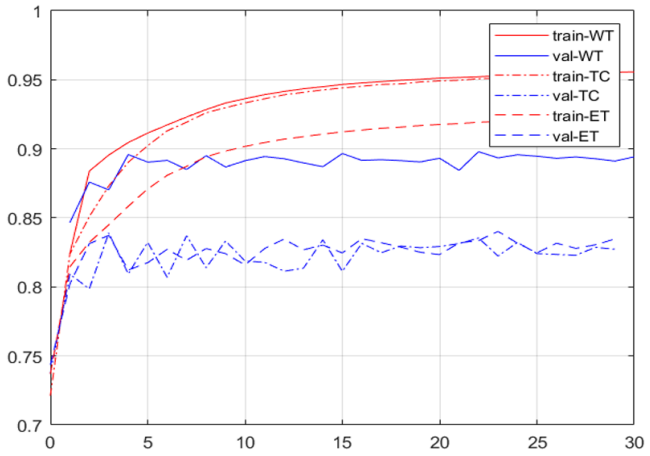

Fig. 6
This figure shows the training and validation dice coefficient vs Number of Epochs for Multi-class Unet Segmentation with Augmentation. WT, TC and ET correspond to whole tumor, tumor core and enhancing tumor respectively

NVIDIA Tesla P100 GPU provided by kaggle is 70 mins. In order to train the system for 30 epoch, it required almost 30 hours. The session allowed by kaggle is for 9 hours. To perform five fold cross-validation, five different models had to be trained separately.

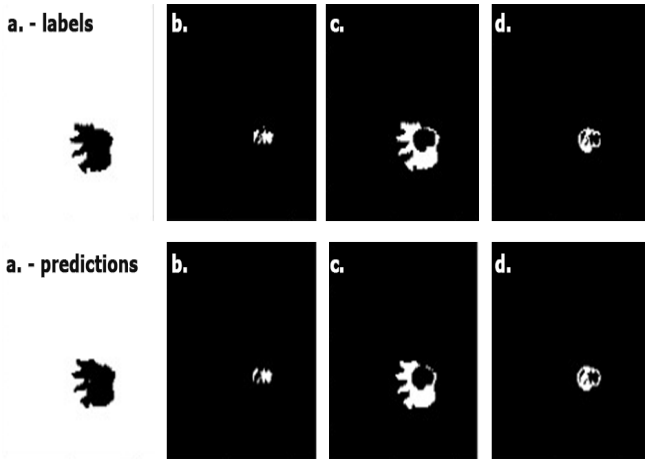

Fig. 7
The first row of images correspond to the original image labels and the second row are the predictions.a is the background mask where as b, c and d correspond to the necrotic and non-enhancing tumor core, the peritumoral edema and enhancing tumor respectively.

5. CONCLUSION

Tumor Segmentation is a 3D Segmentation problem. The choice of 2D models at our ends was dictated by limited computations resources available. 3D U-Nets have an extra dimension in the filters which exponentially increases number of parameters. Hence, significantly more memory is applied. The next step of this project would be to apply these concepts for 3 D segmentations and compare our results with the state of the art models.

6. REFERENCES

[1] Ostrom QT, Gittleman H, Truitt G, Boscia A, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: *Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011-2015.* Neuro Oncol. 2018.

[2] Olaf Ronneberger, Philipp Fischer, Thomas Brox, (MICCAI), Springer, LNCS, Vol.9351: 234–241, 2015 https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/

[3] Multimodal Brain Tumor Segmentation Challenge 2018 https://www.med.upenn.edu/sbia/brats2018/data.html

[4] Center for Biomedical Image Computing and Analytics University of Pennsylvania, Philadelphia PA, USA, https://www.cbica.upenn.edu/sbia/Spyridon.Bakas/MICCAI_BraTS/MICCAI_BraT_2018_proceedings_shortPapers.pdf

[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", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117

[6] 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, 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

[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.