Brain Tumor Segmentation using 3D U-Net

Sachin Salim Shrikant Arvavasu Nowrin Mohamed

University of Michigan, Department of EECS

Abstract

Among brain tumors, gliomas are the most common and aggressive, having extreme variations in shape, size and appearance. Automatic and reliable segmentation methods are important because the large amount of data produced by MRI prevents manual segmentation in a reasonable time. In this paper we aim to develop a deep learning model using a 3D U-Net with adaptations in the preprocessing, training and testing strategies. In addition to this, we created an ensemble of multiple models trained with different hyper-parameters that are used to reduce random errors from each model and yield improved performance. Given the limited computational power, three different 3D U-Net architectures are implemented where each model performs better than the other in its own aspects. Furthermore, the ensemble provides better results.

1. Introduction

Gliomas have 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 radiographic phenotypes, as their sub-regions are depicted by different intensity profiles reflecting differences in tumor biology[7, 15]. Because of the heterogenity, segmentation is challenging[1] and thus, developing a reliable machine learning model that can accurately predict the genetics of this tumor could significantly speed up the diagnosis process and avoid the requirement of multiple invasive surgeries and therapies.

U-Net is a widely used network structure that consists of a contracting and a symmetric expanding path that enables segmentation for the entire input image[12]. In practice, it is very challenging to achieve a single "optimized" model and thus, an ensemble of multiple models can generally improve the segmentation accuracy[14]. In this paper, we propose three different 3D U-Nets with different hyper-parameters and also an ensemble to 3D U-Nets. For each 3D U-Net, the smaller 3D patches will be extracted to minimize memory overhead and also a data reshaping is done where all four modalities of the MRI volume are concatenated together. Furthermore, during testing, a sliding window approach is used to predict class labels with overlap between patches as a testing augmentation method to improve accuracy. Even though many new methods show superior performance, a recent paper claimed that vanilla U-Net can yield robust and superior performance[8].

Our Contributions. Feng et al.[7] explored the idea of ensemble learning by incorporating predictions from multiple models. Though we followed most of their work in this project, there were certain areas where we improvised based on our available resources, time and knowledge. Firstly, the patch extraction method used [7] using a probabilistic approach with heuristic weights while extracting random patches from the scan, whereas we first cropped the scans around the neighborhood of the tumor and then extracted uniformly random patches of a fixed size. Secondly, the model was trained on the "dice-loss" averaged over the classes with a 1:3 ratio of weights for the background and foreground classes, which penalizes the model from overly generalizing the segmentation in case of a class imbalance. We decided the weights empirically based on class frequency and relevance. Moreover, our model achieved a good dice score in 30 epochs of training when we incorporated the scheduling of learning rate for the model.

We also increased the number of dropout layers, but used a lower dropout ratio to keep the overfitting in check. We got rid of the parametric ReLU and used ordinary ReLU instead. We used a batch size of 5 in contrast to the original paper where they'd used 1 during training.

2. Related Work

2.1. CNN and 3D U-Net

Brain tumor segmentation methods include generative and discriminative approaches.The biggest breakthrough in this area was introduced by DeepMedic[2] a 3D CNN that exploits multi-scale features using parallel pathways and incorporates a fully connected conditional random field

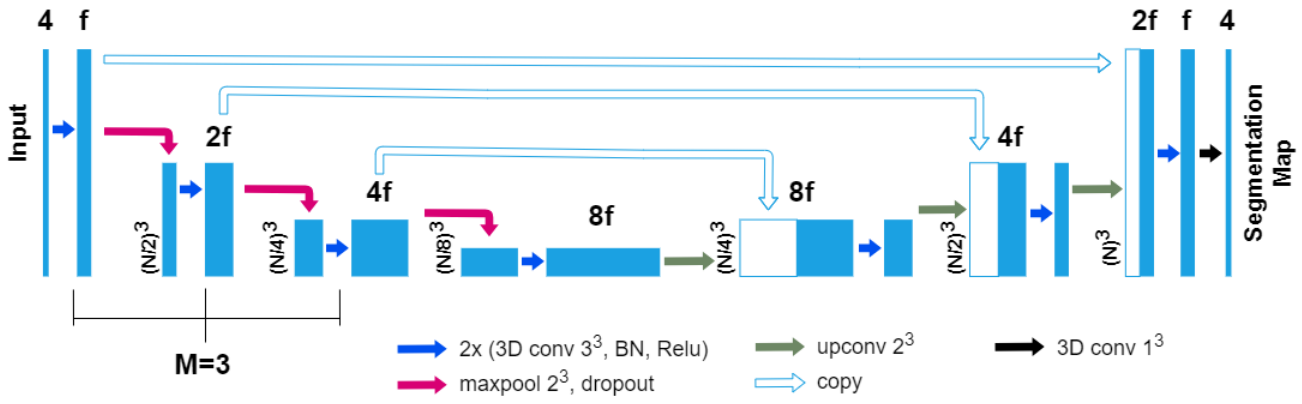

Figure 1. 3D U-Net Architecture

(CRF) to remove false positives[10]. FCN is the main architecture for many semantic segmentation tasks. Furthermore, residual connection is used in the UNet, which is called Res-U-Net[4]. Chen et al.[3] suggested an approach based on atrous convolutional and pooling layers for segmentation. This is called DeepLab. DeepLabV3 is another form of DeepLab. They used few convolutional layers in the decoder path. A FCNN was proposed in [11] for segmenting isointense phase brain MR images. Instead of simply stacking the three modalities at the network input, the network in [11] processes each modality within an independent path. The final segmentation is obtained by fusing the ensuing paths. These approaches have some important drawbacks. These networks use 2D patches as input. In paper [2, 6] authors consider 3D data as input rather than slice-by-slice to improve the segmentation accuracy. 3D-UNet was introduced by Cicek [5], in which contracting skip layers and learned up-sampling parts to get the full resolution of segmentation were proposed.

2.2. Ensemble models

Kao et al. [9] proposed a three step process for survival prediction of patients. In the first step, an ensemble of lesion occurrence probabilities in structural regions with MR images and a patch-based neural network were proposed for the brain tumor segmentation. Ali et al. [1] proposed an ensemble model of 2D and 3D Convolutional Neural Network (CNN) for brain tumor segmentation. Then multiple radiomic and image-based features were extracted from MRI images and segmented regions. Finally, a classification algorithm was applied to predict the overall survival of the patient.

3. Methodology

The steps in our proposed method include data loading, pre-processing of the images, training multiple models using a generic 3D U-Net structure with different hyper-parameters, deployment of each model and the final ensemble step. The description of the methodological details are in the following sections.

3.1. Preprocessing

Patch Extraction. There are several challenges in directly using the whole image as the input to a 3D UNet: (1) the memory of a moderate GPU is often 12Gb so fitting the model into the GPU might affect performance; (2) the training time will be greatly prolonged; (3) as the background voxels dominate the whole image, there will be class imbalance. Therefore, to more effectively utilize the training data, smaller patches were extracted from each subject[7].

For data preparation, all four modalities of the MRI volume are concatenated together along the dimensions(B, H, W, D, Channels). The segmentation data is encoded to one-hot tensors. During each epoch of the training process, a random patch was extracted from each subject using the foreground tumor mask with uniform probabilities. This patch varies every time because the tumor might be in different forms, regions, size and shape. The resultant patch is now resized to 64x64x64 by either cropping or padding. This padding is done dependent on the position of the cropped image. If it's towards the left, adding padding to the left adds zero data which doesn't add any value to the model; rather we will add padding to the right so the model can learn the healthy background voxels and vice versa. The patch size was decided after considering average tumor size, and the computation power available to us.

3.2. Model architecture

A 3D U-Net based network was used as the base structure, as shown in Fig1 For each encoding block, a VGG-like network[13] with two consecutive 3D convolutional layers of kernel size 3 followed by the activation function(ReLU)

Model	N	M	f
Model1	3	64	64
Model2	3	96	48
Model3	4	96	24

Table 1. Table of Hyperparamaters

and batch normalization layers were used. The ReLU function is defined as,

$$f(x) = max(0, x) \tag{1}$$

Similar to the conventional UNet structure, the number of features were doubled while the spatial dimension was halved with every encoding block. A dropout layer with ratio 0.2 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. Weighted dice loss was used as the loss function.

Wu et al.[16]suggested that a wider network with large number of features and a deeper network can increase the expressiveness and thus performance of the network ; furthermore, the larger the patch size, the more spatial information to be used in one patch; however, as mentioned before, the memory of the GPU is often a limiting factor with 3D inputs and we have limited GPU powers. In our study, we balanced the three parameters to make sure that the GPU memory is sufficient while favoring one in one model. The exact choice of these parameters was made empirically. Given the limited time for training and testing, a total of three models was selected, with detailed parameters shown in Table 1. N denotes the input patch size, M denotes the number of encoding/decoding blocks and f denotes the input features at the first layer. Dice loss is separately computed for each class and combined using weights which is the inverse of the class size.Dice loss is calculated as follows,

$$DL = 1 - DSC \tag{2}$$

As already mentioned, even though an entire image can be given as input to the model, there are computational limitations and the input cannot fit into the memory during deployment. Thus a sliding window approach needs to be used to get the output for each subject. A stride size at a fraction of the window size was used and the output probability was averaged. In implementation, the deployment window size was chosen to be the same as the training window size, and the stride was chosen as ½ of the window size. Although smaller stride sizes can be used to further improve the accuracy with more averages, the deployment time will be increased 8 times for every ½ reduction of the window size and thus quickly becomes unmanageable. Using the parameters as mentioned on the same GPU, it took about 1 min to generate the output for the entire volume per subject. Instead of performing a thresholding on the probability output to get the final labels, the direct probability output after the last convolutional layer was saved for each model as a measure of "confidence" for each model. The ensemble modeling process was rather straightforward. The probability output of all classes from each model was averaged to get the final probability output. The class with the highest probability was selected as the final segmentation label for each voxel.

3.3. Post Processing.

Our initial attempt to down-scale the input image to NxNxN and up-scaling the segmented results didn't provide desired results. Hence, we used an approach where we extract the patches in a sliding window and feed to the model, and finally the results are stitched together. Binary closing is done on the result to fill the holes. Segmentation classes are combined to obtain "Enhancing tumor" (ET), the "Tumor core" (TC=ET+NCR), and the "Whole tumor" (WT=TC+ED).

Model #	**Dice Score**			**Hausdroff Distance**		
	WT	**TC**	**ET**	**WT**	**TC**	**ET**
1	0.756	0.761	0.711	4.56	6.98	5.03
2	**0.812**	0.756	0.702	4.02	8.22	4.78
3	0.804	0.732	0.724	4.33	7.64	4.56
Ensemble	0.805	**0.769**	**0.735**	**3.84**	**6.72**	**4.23**

Table 2. Performance of the models

4. Experiments and Results

4.1. DataSet

The datasets used in this study are collected by signing up in synapse.org. The data set was made available for research by the BraTS challenge organizers and contains clinically-acquired preoperative 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 from multiple institutions. They were acquired following different clinical protocols and from various scanners. All the datasets were 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). The latest challenge provides 1251 cases, but from previous works,

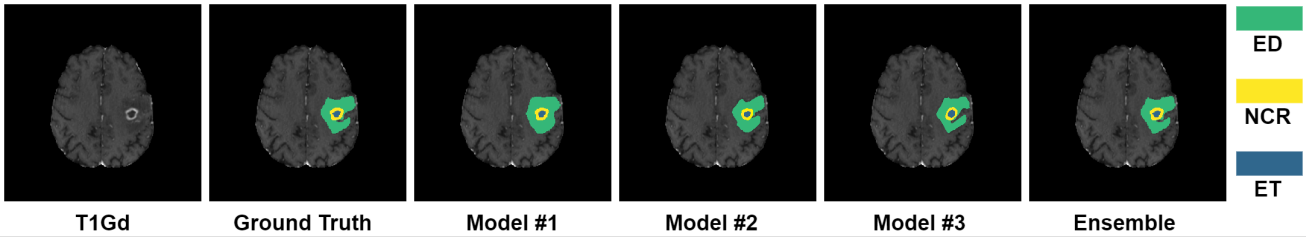

Figure 2. Segmented sub-regions from models 1-3 and the ensemble model is compared with the ground truth for case BraTS2021-00346.

decent results were obtained from lesser data. By keeping our computational power in mind, we have used 300 cases for training, 70 for validating the model performance at each iteration, and 200 for testing the model performance. The batch size chosen is 5 for the data sequence generator. For data preparation, all four modalities of the MRI volume of a particular case were concatenated together along the dimensions(B, H, W, D, Channels). The segmentation data is encoded to one-hot tensors of shape (B, H, W, D,Classes).

4.2. Implementation details

As mentioned above, the UNET architecture was configured with three different settings of hyperparameters as tabulated in Table1 Training was performed on a Google Colab Pro with 80 Gb of GPU-RAM. For each hyperparameter setting, the model was trained for 30 epochs.. Subject orders were randomly permuted every epoch. The model initialization and training were done on Python 3.8 + Tensorflow 2.x framework. Batch size was set to 5 during training. The Adam optimizer was used with an initial learning rate of 10^{-4} along with a reduction of learning rate when the loss stops decreasing. The best model was saved at each step of improvement. The total training time was about 10 hrs.

All 300 training subjects were used in the training process. 70 subjects were provided as validation. The dice indexes are calculated as follows,

$$DSC = \frac{2TP + \epsilon}{FP + 2TP + FN + \epsilon} \tag{3}$$

and 95% Hausdorff distances of ET, WT and TC are calculated as follows,

$$h(A, B) = \max_{a \in A} \{ \min_{b \in B} \{ d(a, b) \} \} \tag{4}$$

$$H(A, B) = \min\{ h(a, b) , h(b, a)\} \tag{5}$$

ET corresponds to label 4 in the direct output label maps; WT is the union of all non-background label maps including label 1, 2, and 4; TC is the union of ET and NCR/NET, or label 1 and 4.

4.3. Results

Table2, shows the mean Dice scores (Dice) and 95% Hausdorff distances of ET, WT and TC in mm for the 3 individual models and the ensemble of them. The model with the best performance of each metric is highlighted. For WT, all 3D U-Net models perform similarly. However, model 3 has the highest Dice for ET. The rankings based on Dice scores are also not consistent with the rankings based on the distance measures. This shows that no single parameter set has a clear advantage over others. However, the ensemble of them has the best overall Dice scores as compared with each individual model. The distance metrics show a wider range and the ensemble does not achieve the smallest values. However, as the Hausdorff distance is largely determined by the "worst" pixels, it may be less reliable in obtaining an overall performance evaluation as compared with Dice scores. Despite this, the metrics in the ensemble method for all three sub-regions are all on the lower end, showing increased robustness.

Fig.2 shows the result of the automatically segmented brain tumors from all the 3 models and its ensemble. A single model may suffer from under- or over-segmentation while the average of multiple models achieves a more stable performance, which is also closer to the ground-truth, as shown with the improved Dice scores. Furthermore, the ensemble of all 3 models yields a much smoother boundary for different sub-regions and eliminates a few isolated regions, which are likely false positives.

5. Discussions and Conclusion

From the experiments in training and ensembling we can conclude that, the hyper-parameters (N, M and F) can significantly affect the segmentation performance of the for different classes differently. Thus by ensembling the results of these three models, we strike a balance between errors in segmented labels of different models. Due to time and computational constraints, we did ensembling of only three models whereas including a grid search might provide better hyper-parameters which is on ongoing research however, one possible concern is that this may lead to overfitting as the validation set is much smaller (70 cases) compared with the training and testing dataset.

References

[1] Muhammad Junaid Ali, Muhammad Tahir Akram, Hira Saleem, Basit Raza, and Ahmad Raza Shahid. Glioma segmentation using ensemble of 2d/3d u-nets and survival prediction using multiple features fusion. In Alessandro Crimi and Spyridon Bakas, editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 189–199, Cham, 2021. Springer International Publishing.

[2] Adrià Casamitjana, Santi Puch, Asier Aduriz, and Verónica Vilaplana. 3d convolutional neural networks for brain tumor segmentation: A comparison of multi-resolution architectures. In Alessandro Crimi, Bjoern Menze, Oskar Maier, Mauricio Reyes, Stefan Winzeck, and Heinz Handels, editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 150–161, Cham, 2016. Springer International Publishing.

[3] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587, 2017.

[4] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2018.

[5] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation. In Sebastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, and William Wells, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 424–432, Cham, 2016. Springer International Publishing.

[6] Jose Dolz, Christian Desrosiers, and Ismail Ben Ayed. 3d fully convolutional networks for subcortical segmentation in mri: A large-scale study. *NeuroImage*, 170:456–470, 2018. Segmenting the Brain.

[7] Xue Feng, Nicholas J. Tustison, and Craig H. Meyer. Brain tumor segmentation using an ensemble of 3d u-nets and overall survival prediction using radiomic features. *CoRR*, abs/1812.01049, 2018.

[8] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F. Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, and Klaus H. Maier-Hein. nnu-net: Self-adapting framework for u-net-based medical image segmentation, 2018.

[9] Po-Yu Kao, Thuyen Ngo, Angela Zhang, Jefferson W. Chen, and B. S. Manjunath. Brain tumor segmentation and tractographic feature extraction from structural MR images for overall survival prediction. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 128–141. Springer International Publishing, 2019.

[10] Andriy Myronenko. 3d MRI brain tumor segmentation using autoencoder regularization. *CoRR*, abs/1810.11654, 2018.

[11] Dong Nie, Li Wang, Yaozong Gao, and Dinggang Shen. Fully convolutional networks for multi-modality isointense infant brain image segmentation. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 1342–1345, 2016.

[12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[13] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2014.

[14] Aik Choon Tan and David Gilbert. Ensemble machine learning on gene expression data for cancer classification. *Appl Bioinformatics*, 2(3 Suppl):S75–83, 2003.

[15] Margaret Wrensch, Yuriko Minn, Terri Chew, Melissa Bondy, and Mitchel S. Berger. Epidemiology of primary brain tumors: Current concepts and review of the literature. *Neuro-Oncology*, 4(4):278–299, 10 2002.

[16] Zifeng Wu, Chunhua Shen, and Anton van den Hengel. Wider or deeper: Revisiting the resnet model for visual recognition. *CoRR*, abs/1611.10080, 2016.