Multimodal Brain Tumor Segmentation and Survival Prediction

EN 520.623 Project 2 Final Report

Chenyu Gao (cgao17@jhu.edu)

Abstract Quantitative analysis of brain tumors is critical for clinical decision making. While manual segmentation is tedious, time consuming and subjective, this task is at the same time very challenging to solve for automatic segmentation methods. In this report we present our implementations of U-Net and Cascaded Anisotropic Convolutional Neural Networks (CACNN) for brain tumor segmentation. On the validation sets, the CACNN gets higher dice coefficients than the U-Net in all subregions of tumor. Based on the segmentation result, we furthermore extract radiomic features to train a regression tree ensemble to predict on the survival time of the patient. We observe a high error on the testing set and will discuss possible causes in this report.

Keywords: *Brain Tumor Segmentation, U-Net, Regression Tree Ensemble, Deep Learning*

1 Introduction & Background

Brain tumor segmentation is crucial for monitoring tumor growth or shrinkage in patients during therapy, and it plays an important role in surgical planning or radiotherapy planning[1]. In current clinical practice, the segmentation is usually still done manually, which is time consuming and tedious for the radiologists and is also of limited use for an objective quantitative analysis. 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 each subregions[4].

However, this task is challenging first because the size, shape, and localization of brain tumors have considerable variations among patients. This limits the usability and usefulness of prior information about shape and location that are widely used for robust segmentation of many other anatomical structures[3]. Second, the boundaries between adjacent structures are often ambiguous due to the smooth intensity gradients, partial volume effects and bias field artifacts[8].

In the past ten years, deep convolutional network models have emerged as promising solution to the segmentation task and outperformed traditional methods in many challenges. This is because when large training set is available, deep convolutional network can encode the spatial relationship and generate feature maps in different scale for a better classification[4].

U-Net[5] is one of the most widely used convolutional networks in medical image analysis. It was designed for semantic segmentation given training set with relatively small sizes, which is often the case in biomedical application. Inspired by the success of U-net, an increasing number of U-net related automatic segmentation algorithms have been proposed. Özgün ÇİÇEK et al.[2] generalized the application of U-net from 2D to 3D, by replacing the 2D kernels with the 3D ones. This allows end-to-end training and testing for volumetric image segmentation.

In addition to U-net, convolutional neural networks with different architectures have also shown their performances in challenges. One good example is the cascaded anisotropic convolutional neural networks (CACNN) proposed by Wang et al.[8]. It separates the complex problem of multiple class segmentation into three simpler binary segmentation problems, and take advantage of the hierarchical structure of tumor subregions to reduce false positives. As the result, their method is one of the leading methods on the BRATS 2017 validation set.

Here we implement both the 3D U-net and the CACNN for brain tumor segmentation and compare their performances. Based on our segmentation results, we build regression tree ensemble for survival prediction. Our approach achieves 0.17, 0.129, 0.81 Dice scores on the testing set for the necrotic and non-enhancing tumor core, the peritumoral edema, and the enhancing tumor respectively. The error of survival prediction is the highest among all teams and we will analyze the potential causes in the following sections.

2 Methods

Dataset and Preprocessing We used a dataset that contains multimodal magnetic resonance imaging (MRI) scans from 100 patients for experiments. Each patient was scanned with four sequences: T1-weighted, T2-weighted, Post-contrast T1-weighted, and T2 Fluid Attenuated Inversion Recovery (FLAIR). All the images were skull-stripped and re-sampled to an isotropic 1 mm^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, comprising 4 labels: the enhancing tumor (ET), the peritumoral edema (ED), the necrotic and non-enhancing tumor core (NCR/NET), and the background. For 41 patients of the 100, we had access to their age and survival time, which were used in combination with the segmentation results to build a regression model for survival prediction.

Tumor Segmentation We implemented two convolutional neural networks respectively to segment the brain tumor. One is a 3D multimodal U-net inspired by Özgün ÇİÇEK et al.[2], and the other is a cascaded anisotropic convolutional neural networks (CACNN) proposed by Wang et al.[8] during the BRATS 2017 Challenge.

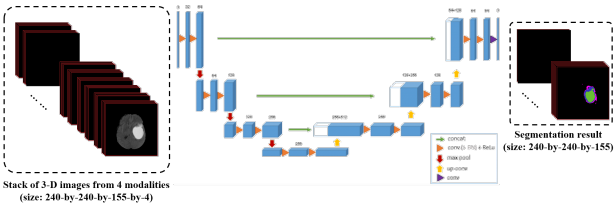

Fig. 1: Schematic of the 3D multimodal U-net.

The 3D U-net[2] is based on the original U-net architecture[5], which consists of a contracting encoder part to analyze the whole image and a successive expanding decoder part to produce a full-resolution segmentation. The author replaced the corresponding kernels and operations from 2D to 3D, thus generalizing its application to volumetric data.

To accomplish multi-modal fusion, i.e., to take the advantages of each modality, we modified the input layer of the network by stacking the 3D MR images of each patient together to a 4D array with a size of $240 \times 240 \times 155 \times 4$, as shown in Fig.1.

One challenge of medical image segmentation is the amount of memory needed to store and process 3-D volumes. Training a network on the full input volume is impractical due to GPU resource constraints. Here we solve the problem by training the network on image patches with a size of $132 \times 132 \times 132$. And to increase the training set, we performed data augmentation (same for the CACNN) by rotation, shifting, flipping, and so on.

The cascaded anisotropic convolutional neural networks (CACNN)[8] segment brain tumor subregions sequentially, as shown in Fig.2. The highlight of this model is the attempt to separate the complex problem of multiple class segmentation into three simpler binary segmentation problems, and to take advantage of the hierarchical structure of tumor subregions to reduce false positives. Also, the author proposes to fuse the output of CNNs in three orthogonal views for more robust segmentation of brain tumor. We replicated this model and finished the training process in Python using PyTorch.

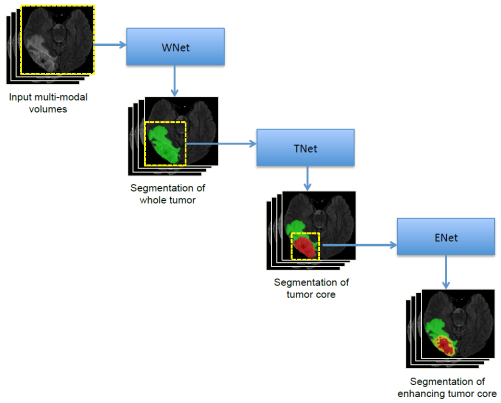

Fig. 2: The triple cascaded framework for brain tumor segmentation.

Survival Prediction The survival prediction is very challenging due to the absence of treatment information and the small size of the available dataset. For only 41 patients, the age and the survival information are provided in addition to their MR image data.

Our approach to survival prediction is based on Radiomics[7], a package providing MATLAB programming tools for radiomics analysis, see https://github.com/mvallieres/radiomics for more details. We computed the texture features and non-texture features for each subregion of tumors and for each volume from different modalities. After append-

ing the age information and removing repeated non-texture features, we extracted a total of 701 features.

These features were then used for training a regression tree ensemble for survival prediction. We performed Principal Component Analysis (PCA) and kept enough components to explain 95% of the variance. We used bagging as the ensemble method and searched for the optimized hyperparameters that minimize the mean squared error (MSE), using 5-fold cross-validation.

3 Results

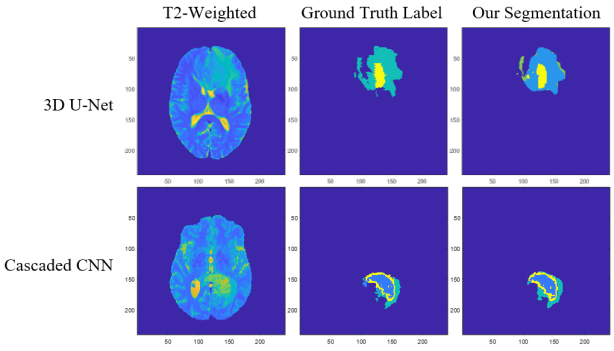

Fig. 3: Segmentation result by the 3D U-Net and the Cascaded Anisotropic CNN.

	Dice Coefficient		
	NCR/NET	ED	ET
U-Net	0.1287	0.2178	0.1943
Cascaded CNN	0.6577	0.8432	0.8102

Table 1: Results of the U-Net and the Cascaded Anisotropic CNN on the validation set.

Segmentation Fig. 3 shows an example of the segmentation results by the 3D U-net and the CACNN. The CACNN captures the fine structure of the subregions of the tumor, and has a significantly better performance than the U-net. Table 1 shows the quantitative segmentation performances of the two models. Compared with the 3D U-net, the CACNN gets higher dice coefficients for all tumor subregions on the validation set.

	Dice Coefficient		
	NCR/NET	ED	ET
U-Net	0.170	0.129	0.810

Table 2: Results of the U-Net on the testing set.

However, we failed to export the CACNN model from Python to MATLAB and it is impractical to write MATLAB code from scratch and finish the training in the given time. Therefore, we used the 3D U-net model for the final evaluation, and Table 2 shows the performance of the 3D U-net on the testing set. Our model ranks at 7^{th} for the NCR/NET, 9^{th} for the ED, and 2^{nd} for the ET.

Survival Prediction We compared our regression tree ensemble with other regression models such as linear regression, support vector machine, single regression tree and so on. We optimized the parameters of these models all by running 5-fold cross-validation on the 41 provided training cases. The optimized single regression tree has the lowest root mean square error (RMSE), which is 358.98. And the optimized regression tree ensemble has the second-lowest RMSE, which is 380.79. Nevertheless, we still submitted the regression tree ensemble for the final evaluation, because it is more robust to overfitting.

However, the survival error of our model is still the highest among all teams. And there are many potential causes to this.

4 Discussion

On the testing set, our approach for brain tumor segmentation achieves a high Dice coefficient on the ET, but shows weak performance on the NCR/NET and the ED. And we get an abnormally low score in the survival prediction task. The potential causes for these problems are the imbalanced label distribution, overfitting, and insufficient feature selection.

Imbalanced Label Distribution The distribution of labels that corresponds to different subregions of tumors is imbalanced, i.e., the number of voxels that are labeled as ET is much larger than NCR/NET

and ED. This can be the main reason why our network performs well on the ET but weakly on the NCR/NET as well as the ED. Carole H. Sudre et al.[6] proposed to use the class re-balancing properties of the Generalized Dice overlap as loss function for unbalanced data. We should explore this method in addition to the weighted multi-class Dice loss function we have implement.

Insufficient Feature Selection In the survival prediction task, we extracted 701 features. Although we performed PCA to reduce dimensionality, the number of remaining features is still large enough to cause overfitting. Moreover, the regression tree ensemble was trained based upon the segmentation results from the CACNN but tested based upon the segmentation results from the 3D U-net, which is less robust. Because we included many texture features, which is highly dependent on the segmentation of the region of interest (ROI), the survival prediction is expected to be affected by the variation of the segmentation result. Isensee et al.[4] provide a good example of feature selection in their survival prediction section. To make the model more robust, we need to consider different combination of features as well.

Overfitting Overfitting exists in both the 3D U-net and the regression tree ensemble, and there are many strategies we can use to overcome the problem. For example, we did data augmentation in the segmentation task, and use bagging to ensemble regression trees in the survival prediction task. There are still many methods we haven't implement, one of which is the optimization of hyperparameters.

With more available time and computation power, we can explore deeper and come up with a more robust approach for the tasks.

5 Contribution

Angelina Zhu and I mainly focused on this project. After discussion and choosing the architecture of the neural networks, we split the coding work into segmentation part and survival prediction part. I was fully in charge of the survival prediction part and Angelina was fully in charge of the segmentation part. We contributed equally to this project.

References

[1] Stefan Bauer, Roland Wiest, Lutz-P Nolte, and Mauricio Reyes. A survey of mri-based medical image analysis for brain tumor studies. *Physics in Medicine & Biology*, 58(13):R97, 2013.

[2] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.

[3] Damien Grosgeorge, Caroline Petitjean, J-N Dacher, and Su Ruan. Graph cut segmentation with a statistical shape model in cardiac mri. *Computer Vision and Image Understanding*, 117(9):1027–1035, 2013.

[4] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H Maier-Hein. Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In *International MICCAI Brainlesion Workshop*, pages 287–297. Springer, 2017.

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[6] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep learning in medical image analysis and multi-modal learning for clinical decision support*, pages 240–248. Springer, 2017.

[7] Martin Vallieres, Emily Kay-Rivest, Léo Jean Perrin, Xavier Liem, Christophe Furstoss, Hugo JWL Aerts, Nader Khaouam, Phuc Felix Nguyen-Tan, Chang-Shu Wang, Khalil Sultanem, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Scientific reports*, 7(1):1–14, 2017.

[8] Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In *International MICCAI brainlesion workshop*, pages 178–190. Springer, 2017.