3D Semantic Segmentation of Brain Tumor for Overall Survival Prediction

Rupal R. Agravat and Mehul S. Raval
Ahmedabad University, Ahmedabad, Gujarat, India
rupal.agravat@iet.ahduni.edu.in
mehul.raval@ahduni.edu.in

Abstract. Glioma, the malignant brain tumor, requires immediate treatment to improve the survival of patients. Glioma’s heterogeneous nature makes the segmentation difficult, especially for sub-regions like necrosis, enhancing tumor, non-enhancing tumor, and Edema. Deep neural networks like full convolution neural networks and ensemble of fully convolution neural networks are successful for Glioma segmentation. The paper demonstrates the use of a 3D fully convolution neural network with a three-layer encoder-decoder approach for layer arrangement. The encoder blocks include the dense modules, and decoder blocks include convolution modules. The input to the network is 3D patches. The loss function combines dice loss and focal loss functions. The validation set dice score of the network is 0.74, 0.88, and 0.73 for enhancing tumor, whole tumor, and tumor core, respectively. The Random Forest Regressor uses shape, volumetric, and age features extracted from ground truth for overall survival prediction. The regressor achieves an accuracy of 44.8% on the validation set.

Keywords: Brain Tumor Segmentation · Deep Learning · Dense Network · Overall Survival · Radiomics Features · Random Forest Regressor · U-net.

1 Introduction

Early-stage brain tumor diagnosis can lead to proper treatment planning, which improves patient survival chances. Out of all types of brain tumors, Glioma is one of the most life-threatening brain tumors. It occurs in the glial cells of the brain. Depending on its severity and aggressiveness, Glioma has grades ranging from grade I to grade IV. Grade I, II are Low-Grade Glioma(LGG), and grade III and IV are High-Grade Glioma(HGG). It can further be divided into constituent components like - Necrosis, Enhancing tumor, Non-enhancing tumor, and Edema. The core consists of necrosis, enhancing tumor, non-enhancing tumor. In most cases, LGG does not contain enhancing tumor, whereas HGG contains necrosis, enhancing, and non-enhancing sub-components. Edema occurs from infiltrating tumor cells and biological response to the angiogenic and vascular permeability.
factors released by the spatially adjacent tumor cells\cite{4}. Non-invasive Medical Resonance Imaging (MRI) is the most advisable imaging technique as it captures the functioning of soft tissue adequately compared to other imaging techniques. MR images are prone to inhomogeneity introduced by the surrounding magnetic field, which introduces the artifacts in the captured image. Besides, the appearance of various brain tissues is different in various modalities. Such issues increase the time in the study of the image. The treatment planning is highly dependent on the accurate sub-component segmentation, but due to the heterogeneous nature of Glioma, the segmentation task becomes difficult. Furthermore, the human interpretation of the image is non-reproducible as well as dependent on the expertise. It requires computer-aided MR image interpretation to locate the tumor. Also, even the initially detected tumor is completely resected, such patients have poor survival prognosis, as metastases may still redevelop, which leads to an open question to the accurate overall survival prediction.

Authors in \cite{1} discussed the basic, generative, and discriminative techniques for brain tumor segmentation. Nowadays, Deep Neural Network (DNN) has gained more attention for the segmentation of biological images. In which, Convolution Neural Networks (CNN), like DeepMedic \cite{15}, U-net \cite{20}, V-Net \cite{19}, SegNet \cite{5}, ResNet \cite{12}, DenseNet \cite{13} give state-of-the-art results for semantic segmentation. Out of all these methods, U-net is a widely accepted end-to-end segmentation architecture for brain tumors. U-net is an encoder-decoder architecture, which reduces the size of feature maps to half and doubles the number of feature maps at every encoder layer. The process is reversed at every decoder layer. The skip connections between the peer layers of U-net help in proper reconstruction of the features.

1.1 Literature Review: BraTS 2019

**Segmentation** Authors in \cite{14} used the ensemble of twelve encoder-decoder models, where each model is made up of cascaded network. The first network in a model finds the coarse segmentation, which was given as input in the second network and the input images to predict all the labels. The network losses combine at different stages for better network parameter tuning. The Dice Similarity Coefficient (DSC) for the validation set is 0.91, 0.87, 0.80 for Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET), respectively.

Authors in \cite{25} applied various data processing methods, network design methods, and optimization methods to better learn the segmentation labels at every iteration by the student models combined at the teacher-level model with successive output merging. The loss function is the combination of dice loss and cross-entropy loss for the networks trained on various input patch sizes. They achieved the DSC of 0.91, 0.84, and 0.75 for WT, TC, and ET.

The approach demonstrated in \cite{17} used thirty Heteroscedastic classification models to find the variance of all the models for the ensemble. The focal loss forms the loss function. Various post-processing techniques were applied to fine-tune the network segmentation. The DSC achieved for the approach was 0.91, 0.83, and 0.77 for WT, TC, and ET.
Survival Prediction Authors in [2] implemented 2D U-net with dense modules at encoder part and convolution modules at the decoder part along with focal loss function at training time. The segmentation results fed into Random Forest Regressor (RFR) to predict the overall survival of the patients. The RFR trains on the age, shape, and volumetric features extracted from the ground truth provided with the training dataset. They achieved 58.6% accuracy on the validation set.

Authors in [24] used vanilla U-net and U-net with attention blocks to make the ensemble of six models based on various input patches and the presence/absence of attention blocks. The linear regressor trains selected radiomics features along with the relative invasiveness coefficient. The DSC achieved on the validation set was 0.90, 0.83, and 0.79 for WT, TC, and ET, respectively, and the OS accuracy was 59%.

Authors in [11] implemented the ensemble of six models, which are the variation of U-net with different patch sizes, feature maps with several layers in the encoder-decoder architecture. For OS prediction, six features were extracted from the segmentation results to train the linear regression. The DSC achieved on the validation set was 0.91, 0.80, and 0.74 for WT, TC, and ET, and the OS accuracy was 31%.

Authors in [23] used the U-net variation, where the additional branch of prediction uses Variational Encoder. The OS prediction used the volumetric and age features to train ANN with two layers, each with 64 neurons.

Except the approach demonstrated in [2,23], all the other approaches use an ensemble of the segmentation prediction networks. There are certain disadvantages of ensemble approaches: (1) ensemble methods are usually computationally expensive. Therefore, they add learning time, and memory constraints to the problem, (2) using ensemble methods reduces the model interpret-ability due to increased complexity and makes it very difficult to understand.

Moreover, according to [3], inductive transfer learning improves network performance. It implements the U-net of [10,20] with reduced network depth. Reduction in network depth has reduced the number of network parameters. In addition to the depth reduction, the dense module at the encoder replaces the convolution module. The network training uses a combination of dice loss and focal loss functions.

The remaining paper is as follows: section [2] of the paper focuses on the BraTS 2020 dataset, section [3] demonstrates the proposed method, section [4] provides implementation details, and section [5] shows the results followed by the conclusion and future work.

2 Dataset

The dataset [8,9,18] contains 293 HGG and 76 LGG pre-operative scans. All the images have been segmented manually, by one to four raters, following the same annotation protocol to generate the ground truths. The annotations were
approved by experienced neuro-radiologists. Annotations have the enhancing tumor (ET label 4), the peritumoral edema (ED label 2), and the necrotic and non-enhancing tumor core (NCR/NET label 1). Images are co-registered to the same anatomical template, interpolated to the same resolution (1mm x 1mm x 1mm), and skull-stripped. Features like age, survival days, and resection status for 237 HGG scans are provided separately for Overall Survival (OS). The validation dataset consists of 125 scans, with the same preprocessing as well as additional features, as mentioned for OS.

3 Proposed Method

3.1 Task 1: Tumor Segmentation

A Fully Convolution Neural Network (FCNN) provides end-to-end semantic segmentation for the input of the arbitrary size and learns global information related to it. Moreover, the 3D FCNN gathers a new spatial relationship between the voxels. Our network has a basis from the network proposed by with 3D variation. The network uses three-layer encoder-decoder architecture with the dense connections between the successive convolution layers and skip-connections across peer layers at the encoder side, as shown in Fig. 1. The network contains three dense modules and two convolution modules. Each convolution layer in the dense module uses the ReLU activation function. Dense connections between the layers in the dense module allows to obtain additional inputs (collective knowledge) from all earlier layers and passes on its feature-maps to all subsequent layers. Dense connections allow the gradient to flow to the earlier layers directly, which provides in-depth supervision on preceding layers by the classification layer. Also, dense connections provide diversified features to the layers, which leads to having detailed patterns identification capabilities. Each dense module generates 64, 128, and 256 feature maps, respectively. Each convolution module generates 128 and 64 feature maps applying 1x1x1 convolution at the end to generate a single probability map for multi-class classification for the labels.

Brain tumor segmentation task deals with a highly imbalanced dataset where tumorous slices are less than non-tumorous slices; such an imbalance dataset reduces network accuracy. The patch-based input to the network guarantees that the network does not overlearn the background voxels. The network trains with the combination of following loss functions.

- Soft Dice Loss: is a measure to find overlap between two regions.

\[
\text{SoftDiceLoss} = 1 - \frac{2 \sum_{\text{voxels}} y_{true} y_{pred}}{\sum_{\text{voxels}} y_{true}^2 + \sum_{\text{voxels}} y_{pred}^2}
\]  

(1)

\(y_{true}\) represents ground truth and \(y_{pred}\) represents network output probability. The dice loss function directly considers the predicted probabilities without converting into binary output. The numerator provides standard
correct predictions between input and target, whereas the denominator provides individual separate correct predictions. This ratio normalizes the loss according to the target mask and allows learning even from the minimal spatial representation of the target mask.

- **Focal Loss** [16]: It is dependent on the network probability $p_t$. It balances negative and positive samples by tuning $\alpha$. It also deals with easy and hard examples by focusing on parameter $\gamma$.

$$FL(p_t) = -\alpha t (1 - p_t)^\gamma \log(p_t)$$  \hspace{1cm} (2)

The modulating factor $(1 - p_t)^\gamma$ adjusts the rate at which easy examples are down-weighted.

### 3.2 Task 2: Overall Survival prediction

OS prediction deals with predicting the number of days for which patients survive after providing appropriate treatment. We have used the following features to train RFR:

- **Statistical Features**: the amount of edema, amount of necrosis, amount of enhancing tumor, the extent of tumor and proportion of tumor
- **Radiomic Features** [22] **for necrosis**: Elongation, flatness, minor axis length, primary axis length, 2D diameter row, 2D diameter column, sphericity, surface area, 2D diameter slice, 3D diameter
- **Age** (available with BraTS dataset)
Necrosis plays a significant role in the treatment of tumors. Gross Total Resection (GTR) of necrosis is comparatively easy vis a vis enhancing tumor. Considering this, shape features of necrosis are extracted using a radiomics package [22]. In addition to those features, whole tumor statistical features from the segmentation results and age are considered to train RFR.

4 Implementation Details

4.1 Pre-processing

Pre-processing boosts network training and improves performance. All four modality images are biased field corrected followed by denoising, and Z-score normalization on individual MR sequence is applied where each sequence was subtracted by its mean from the data and divided by its standard deviation. Data augmentation happens by flipping the patches around the vertical axis.

4.2 Training

Input to the network is 3D patches from four modalities (T1, T2, T1c, FLAIR). The network trains on the entire training image dataset. The network uses two different loss functions: 1) dice loss function and 2) focal loss function with $\alpha = 1$ and $\gamma = 2$. The network trains for 610 epochs with batch size 1.

The sliding window approach provides the output for each subject. The stride size is reduced to half of the training window size to overcome the boundary voxels’ unstable prediction issue. The output of the original patch and flipped patch is predicted and averaged to generate the final output. The prediction of a single image takes around one minute.

4.3 Post-processing

Enhancing tumor is formed in surrounding of the necrosis tumor sub-component. Empirically it’s size cannot be very small. In post-processing such small size enhancing tumor is converted to necrosis sub-components. The threshold for the conversion is set to three hundred.

5 Results

Segmentation: Various evaluation metrics like DSC, Hausdorff95, Sensitivity and Specificity for training set are in Table 1 and Table 2 and for validation set are in Table 3 and Table 4.

Fig. 2 shows the successful segmentation of the tumor and Fig. 3 shows the segmentation failure of the network. The network fails to segment the necrosis and enhancing tumor from the image.

RFR trains on features extracted from the 213 ground truth images. In the trained RFR, features of network segmented images predict OS days. If the
Table 1: DSC and Hausdorff95 for BraTS 2020 training dataset.

|          | DSC |       |       |       |       |       |
|----------|-----|-------|-------|-------|-------|-------|
|          | ET  | WT    | TC    | ET    | WT    | TC    |
| Mean     | 0.763 | 0.890 | 0.820 | 29.559 | 7.070 | 7.179 |
| StdDev   | 0.261 | 0.103 | 0.213 | 93.329 | 11.151 | 21.226 |
| Median   | 0.863 | 0.925 | 0.909 | 1.732  | 3.162 | 3.464 |
| 25quantile | 0.758 | 0.876 | 0.790 | 1.414  | 2.000 | 1.732 |
| 75quantile | 0.906 | 0.947 | 0.947 | 4.000  | 7.348 | 7.162 |

Table 2: Sensitivity and Specificity for BraTS 2020 training dataset.

|          | Sensitivity |       |       |       |       |       |
|----------|-------------|-------|-------|-------|-------|-------|
|          | ET          | WT    | TC    | ET    | WT    | TC    |
| Mean     | 0.787       | 0.864 | 0.776 | 0.999 | 0.999 | 0.999 |
| StdDev   | 0.264       | 0.143 | 0.244 | 0.000 | 0.000 | 0.000 |
| Median   | 0.881       | 0.915 | 0.878 | 0.999 | 0.999 | 0.999 |
| 25quantile | 0.780 | 0.825 | 0.731 | 0.999 | 0.999 | 0.999 |
| 75quantile | 0.940 | 0.955 | 0.940 | 0.999 | 0.999 | 0.999 |

Table 3: DSC and Hausdorff95 for BraTS 2020 validation dataset.

|          | DSC |       |       |       |       |       |
|----------|-----|-------|-------|-------|-------|-------|
|          | ET  | WT    | TC    | ET    | WT    | TC    |
| Mean     | 0.738 | 0.876 | 0.725 | 34.191 | 9.475 | 14.538 |
| StdDev   | 0.307 | 0.093 | 0.284 | 109.143 | 15.215 | 38.067 |
| Median   | 0.835 | 0.914 | 0.866 | 2.828  | 4.000 | 5.099 |
| 25quantile | 0.614 | 0.863 | 0.595 | 1.414  | 2.236 | 2.236 |
| 75quantile | 0.889 | 0.932 | 0.924 | 10.770 | 7.550 | 10.724 |

Table 4: Sensitivity and Specificity for BraTS 2020 validation dataset.

|          | Sensitivity |       |       |       |       |       |
|----------|-------------|-------|-------|-------|-------|-------|
|          | ET          | WT    | TC    | ET    | WT    | TC    |
| Mean     | 0.756       | 0.858 | 0.675 | 0.999 | 0.999 | 0.999 |
| StdDev   | 0.321       | 0.138 | 0.312 | 0.000 | 0.001 | 0.000 |
| Median   | 0.852       | 0.900 | 0.819 | 0.999 | 0.999 | 0.999 |
| 25quantile | 0.594 | 0.825 | 0.432 | 0.999 | 0.998 | 0.999 |
| 75quantile | 0.926 | 0.952 | 0.915 | 0.999 | 0.999 | 0.999 |
network fails to identify / segment necrosis from the image, then the feature extractor considers the absence of the necrosis and marks all the features except age as zero. OS accuracy for training and validation datasets of the images whose resection status is GTR is in Table 5.

| Dataset   | Accuracy | MSE       | MedianSE | StdSE       | SpearmanR |
|-----------|----------|-----------|----------|-------------|-----------|
| Training  | 0.508    | 94338.575 | 28561.640 | 198488.993 | 0.525     |
| Validation| 0.448    | 122377.477| 60123.040 | 201199.356 | 0.226     |

According to the study [21], gender plays a vital role in response to tumor treatment. The females respond to the post-operative treatment better compared to males, which improve their life expectancy. The inclusion of the gender feature into the existing feature list can significantly improve OS accuracy.

6 Conclusion

The proposal uses three-layer deep 3D U-net based encoder-decoder architecture for semantic segmentation. Each encoding and decoding layer modules incorpo-
rates dense modules. The network achieves comparable DSC for training datasets with other leader board methods but generates slightly poor results for the validation dataset. In the future, better preprocessing techniques, augmentation, better design of the layer modules, and further post-processing will improve the results. The improved network output will be compared with state-of-the-art and the results will be discussed in the updated version of the paper. Age, statistical, and necrosis shape features of the ground truth, train RFR with five-fold cross-validation for OS prediction. Later, network segmentation for cases with GTR tests RFR for OS prediction.

Acknowledgement

The authors would like to thank NVIDIA Corporation for donating the Quadro K5200 and Quadro P5000 GPU used for this research, Dr. Krutarth Agrawat (Medical Officer, Essar Ltd) for clearing our doubts related to medical concepts, Po-yu Kao, Ph.D. Candidate, Vision Research Lab, University of California, Santa Barbara and Ujjawal Baid for their continuous guidance during implementation difficulties. The authors acknowledge continuous support from Professor Sanjay Chaudhary, Professor N. Padmanabhan, and Professor Manjunath Joshi for this work.

References

1. Agrawat, R.R., Raval, M.S.: Deep learning for automated brain tumor segmentation in mri images. In: Soft Computing Based Medical Image Analysis, pp. 183–201. Elsevier (2018)
2. Agrawat, R.R., Raval, M.S.: Brain tumor segmentation and survival prediction. In: International MICCAI Brainlesion Workshop. pp. 338–348. Springer (2019)
3. Agrawat, R.R., Raval, M.S.: Prediction of overall survival of brain tumor patients. In: TENCON 2019-2019 IEEE Region 10 Conference (TENCON). pp. 31–35. IEEE (2019)
4. Akbari, H., Macyszyn, L., Da, X., Wolf, R.L., Bilello, M., Verma, R., ORourke, D.M., Davatzikos, C.: Pattern analysis of dynamic susceptibility contrast-enhanced mr imaging demonstrates peritumoral tissue heterogeneity. Radiology 273(2), 502–510 (2014)
5. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence 39(12), 2481–2495 (2017)
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)
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 286 (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., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M., Rozycki, M., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. arXiv preprint arXiv:1811.02629 (2018)

10. Dong, H., Yang, G., Liu, F., Mo, Y., Guo, Y.: Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In: annual conference on medical image understanding and analysis. pp. 506–517. Springer (2017)

11. Feng, X., Dou, Q., Tustison, N., Meyer, C.: Brain tumor segmentation with uncertainty estimation and overall survival prediction. In: International MICCAI Brain lesion Workshop. pp. 304–314. Springer (2019)

12. 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)

13. Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., Keutzer, K.: Densenet: Implementing efficient convnet descriptor pyramids. arXiv preprint arXiv:1404.1869 (2014)

14. Jiang, Z., Ding, C., Liu, M., Tao, D.: Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part I 5. pp. 231–241. Springer International Publishing (2020)

15. 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)

16. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)

17. McKinley, R., Rebsamen, M., Meier, R., Wiest, R.: Triplanar ensemble of 3d-to-2d cnns with label-uncertainty for brain tumor segmentation. In: International MICCAI Brainlesion Workshop. pp. 379–387. Springer (2019)

18. 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 (2014)

19. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571. IEEE (2016)

20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedial image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)

21. Sun, T., Plutynski, A., Ward, S., Rubin, J.B.: An integrative view on sex differences in brain tumors. Cellular and molecular life sciences 72(17), 3323–3342 (2015)

22. 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)
23. Wang, F., Jiang, R., Zheng, L., Meng, C., Biswal, B.: 3d u-net based brain tumor segmentation and survival days prediction. In: International MICCAI Brainlesion Workshop. pp. 131–141. Springer (2019)
24. Wang, S., Dai, C., Mo, Y., Angelini, E., Guo, Y., Bai, W.: Automatic brain tumour segmentation and biophysics-guided survival prediction. arXiv preprint arXiv:1911.08483 (2019)
25. Zhao, Y.X., Zhang, Y.M., Liu, C.L.: Bag of tricks for 3d mri brain tumor segmentation. In: International MICCAI Brainlesion Workshop. pp. 210–220. Springer (2019)