# IMPROVING 3D U-NET FOR BRAIN TUMOR SEGMENTATION BY UTILIZING LESION PRIOR

*Po-Yu Kao*[⋆]      *Jeffereson W. Chen*[†]      *B.S. Manjunath*[⋆]

[⋆] Vision Research Lab, University of California, Santa Barbara, California, United States
[†] UCI Health, University of California, Irvine, California, United States

## ABSTRACT

We propose a novel, simple and effective method to integrate lesion prior and a 3D U-Net for improving brain tumor segmentation. First, we utilize the ground-truth brain tumor lesions from a group of patients to generate the heatmaps of different types of lesions. These heatmaps are used to create the volume-of-interest (VOI) map which contains prior information about brain tumor lesions. The VOI map is then integrated with the multimodal MR images and input to a 3D U-Net for segmentation. The proposed method is evaluated on a public benchmark dataset, and the experimental results show that the proposed feature fusion method achieves an improvement over the baseline methods. Besides, our proposed method also achieves competitive performance compared to state-of-the-art methods.

***Index Terms***— Brain tumor segmentation, feature fusion, volume-of-interest, 3D U-Net, lesion prior

## 1. INTRODUCTION

Primary central nervous system (CNS) tumors refer to a heterogeneous group of tumors arising from cells within the CNS and can be benign or malignant. Malignant primary brain tumors remain among the most difficult cancers to treat, with a 5-year overall survival rate no greater than 35%. The most common malignant primary brain tumors in adults are gliomas. In a patient with a suspected brain tumor, magnetic resonance imaging (MRI) with gadolinium is the investigation tool of choice [1]. Manual segmentation of brain tumors on MR images is a challenging and time-consuming task. Therefore, an automatic and accurate brain tumor segmentation tool benefits radiologists and physician on both diagnosis and treatment planning.

Convolutional neural networks have achieved state-of-the-art performance in the recent Multimodal Brain Tumor Image Segmentation Benchmarks (BraTS) [2, 3, 4, 5]. These works focus on designing a new network architecture, loss function, data augmentation, and training and testing procedure in order to improve the performance of brain tumor

segmentation. Another method proposed by Kao et al. [6] utilizes an existing brain parcellation to bring location information of brain into patch-based neural networks that improves the brain tumor segmentation performance of networks. Inspired by their work, we directly integrate lesion prior with multimodal MR images and input the fused information to a 3D U-Net. The proposed lesion prior fusion method includes two steps: (i) we first create a volume-of-interest (VOI) map from the ground-truth brain tumor lesions, and (ii) this VOI map is then integrated with the multimodal MR images and input to a 3D U-Net for the brain tumor segmentation. The main contribution of this paper is the integration of lesion prior to a 3D U-Net architecture that improves the brain tumor segmentation performance of the 3D U-Net.

## 2. MATERIALS AND METHODS

### 2.1. Dataset

Multimodal Brain Tumor Image Segmentation Benchmark (BraTS) 2017 [7, 8, 9, 10] provides 285 subjects in the training set and 46 subjects in the validation set. Multimodal MR images are provided for each subject, but ground-truth lesion mask is only available for the training subject. These MR images include T1-weighted, contrast-enhanced T1-weighted, T2-weighted, and fluid-attenuated inversion recovery scans, and the ground-truth lesion mask comprises the enhancing tumor (ET), edema (ED), and necrotic & non-enhancing tumor (NCR/NET). The dimension of each image is $240 \times 240 \times 155$ in the $x, y$ and $z$ direction, and the voxel resolution is $1mm^3$. The provided data are intra-subject registered, interpolated to the same resolution and skull-stripped.

### 2.2. Volume-of-interest Map

The volume-of-interest (VOI) map is built in the Montreal Neurological Institute (MNI) 1mm space [11], and each voxel of the VOI map has a label ranging from 0 to 9, which represents different probabilities of observing the brain tumor lesions. First, we build the heatmaps of different types of brain tumor lesions in the MNI space. We apply inter-subject registration which registers the ground-truth lesions of each BraTS

2017 training subject from the subject space to the MNI space using FLIRT [12] from FSL [13]. We then split the brain lesions of each subject into three binary masks, and each binary mask only contains information of one type of lesion. For each type of lesion, we apply element-wise summation to the binary masks of all 285 training subjects and create the heatmap of this type of lesion. Fig. 1 shows the heatmaps of different brain tumor lesions from BraTS 2017 training subjects in the MNI space.
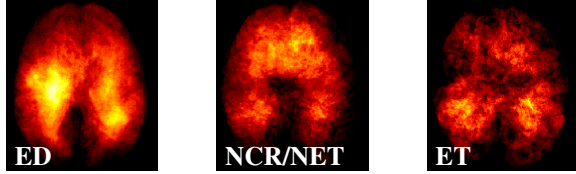


**Fig. 1**. The heatmaps of different brain tumor lesions. The brighter voxels (yellow) represent higher intensity values. **Best viewed in color.**

The heatmaps of different brain lesions are then used to create the VOI map. The VOI map construction accounts for the fact that the whole tumor is a superset of ET, NCR/NET and ED, and the tumor core includes ET and NCR/NET. In addition, ETs are usually observed in patients with high-grade gliomas whose survival rate is considerably lower than patients with low-grade gliomas. Based on these observations, we create Algorithm 1 to generate the VOI map and prioritize the order of the VOI labels.

Note that the VOI labels are based on the thresholds which are chosen from the percentiles of non-zero voxels of heatmaps. For each lesion type, we sort the frequency counts of the non-zero voxels, and the heatmaps are used to generate these frequency counts. The percentile thresholds $(h_{ed}, h_{ncr}, h_{et})$ are selected from these sorted frequency counts. We then use these percentile thresholds to create the VOI label mapping. Any given voxel location in the VOI map has probabilities of being different types of lesion. We examined different thresholds, and $(\alpha, \beta, \gamma) = (50, 65, 80)$ percentiles yield the best overall segmentation performance. Fig. 2 shows the VOI map and the distribution of brain tumor lesions occurring in the different labels of VOI map. This distribution is computed by dividing the total voxel value of lesions in the heatmaps by the total volume of the corresponding VOI label. This distribution shows that (i) the prior probabilities of different lesions depend on their corresponding labels in the VOI label map, and (ii) lesions have higher probabilities to happen in the larger VOI labels.

## 2.3. 3D U-Net

### 2.3.1. Data pre-processing

Intensity normalization is the procedure of mapping intensities of different MR images into a standard scale, and it is an

---

**Algorithm 1:** Build the VOI map from the heatmaps of lesions.

**input** : A heatmap $H_{ed}$ of ED of size $w \times l \times d$
A heatmap $H_{ncr}$ of NCR/NET of size $w \times l \times d$
A heatmap $H_{et}$ of ET of size $w \times l \times d$
**output:** The VOI map $V$ of size $w \times l \times d$

$h_{ed,1}, h_{ed,2}, h_{ed,3} \leftarrow \alpha, \beta, \gamma$ percentile of non-zero voxels of $H_{ed}$;
$h_{ncr,1}, h_{ncr,2}, h_{ncr,3} \leftarrow \alpha, \beta, \gamma$ percentile of non-zero voxels of $H_{ncr}$;
$h_{et,1}, h_{et,2}, h_{et,3} \leftarrow \alpha, \beta, \gamma$ percentile of non-zero voxels of $H_{et}$;

**for** $i \leftarrow 1$ **to** $w$ **do**
  **for** $j \leftarrow 1$ **to** $l$ **do**
    **for** $k \leftarrow 1$ **to** $d$ **do**
      **if** $H_{et}[i,j,k] \geq h_{et,3}$ **then**
        $V[i,j,k] \leftarrow 9$;
      **else if** $H_{ncr}[i,j,k] \geq h_{ncr,3}$ **then**
        $V[i,j,k] \leftarrow 8$;
      **else if** $H_{ed}[i,j,k] \geq h_{ed,3}$ **then**
        $V[i,j,k] \leftarrow 7$;
      **else if** $H_{et}[i,j,k] \geq h_{et,2}$ **then**
        $V[i,j,k] \leftarrow 6$;
      **else if** $H_{ncr}[i,j,k] \geq h_{ncr,2}$ **then**
        $V[i,j,k] \leftarrow 5$;
      **else if** $H_{ed}[i,j,k] \geq h_{ed,2}$ **then**
        $V[i,j,k] \leftarrow 4$;
      **else if** $H_{et}[i,j,k] \geq h_{et,1}$ **then**
        $V[i,j,k] \leftarrow 3$;
      **else if** $H_{ncr}[i,j,k] \geq h_{ncr,1}$ **then**
        $V[i,j,k] \leftarrow 2$;
      **else if** $H_{ed}[i,j,k] \geq h_{ed,1}$ **then**
        $V[i,j,k] \leftarrow 1$;
      **else**
        $V[i,j,k] \leftarrow 0$;
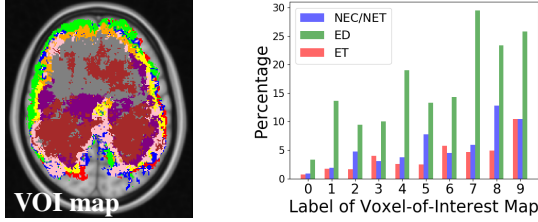      **end**
    **end**
  **end**
**end**

**Fig. 2**. The VOI map (background-0, red-1, green-2, blue-3, yellow-4, orange-5, pink-6, purple-7, grey-8, and brown-9) and the distribution of brain tumor lesions (green-ED, blue-NEC/NET, and red-ET) observed in the different labels of VOI map from BraTS 2017 training subjects. **Best viewed in color.**

essential step to avoid initial biases and improve the performance of the network. For each MR image, we first clip it at [0.2 percentile, 99.8 percentile] of non-zero voxels to remove the outliers and subsequently normalize every voxel within the brain with respect to their mean and standard deviation. That is, $\bar{x}_i = (x_i - \mu)/\sigma$ where $i$ is the index of voxel inside the brain, $\bar{x}_i$ is the normalized voxel, $x_i$ is the corresponding raw voxel, and $\mu$ and $\sigma$ are the mean and standard deviation of the raw voxels inside the brain, respectively.

### 2.3.2. Network architecture

The proposed network architecture shown in Fig. 3 is based on 3D U-Nets [14, 15]. Different colors of blocks represent different types of layers. The number of convolutional kernels is indicated within the white box. Group normalization [16] is used, and the number of groups is set to 4. Trilinear interpolation is used in the upsampling layer.

### 2.3.3. Training and testing procedure

The proposed network is trained with randomly cropped patches of size $128 \times 128 \times 128$ voxels and batch size 2. A larger input patch capture more contextual information of the brain. In every epoch, a cropped patch is randomly extracted from each subject. The network is trained for a total of 300 epochs. The weights of network are updated by Adam algorithm [17] with an initial learning rate $l_0 = 10^{-3}$ following the schedule of $l_0 \times 0.1^{\text{epoch}}$, L2 penalty weight decay of $10^{-4}$, and AMSGrad [18]. For the loss function, the standard multiclass cross-entropy loss with the hard negative mining is used to solve the class imbalance problem of the dataset. We only back-propagate the negative (background) voxels with the largest losses (hard negative) and the positive (lesions) voxels to the gradients. In our implementation, the number of selected negative voxels is at most three times more than the number of positive voxels. In addition, data augmentation is not used for both training and testing. At the testing time, we input the entire image of size $240 \times 240 \times 155$ voxels into the

trained 3D U-Net for each patient to get the predicted lesion mask. Training takes approximate 12.5 hours, and testing takes approximate 1.5 seconds per subject on an Nvidia 1080 Ti GPU.

### 2.4. Integrate the VOI Map and a 3D U-Net

Fig. 4 shows the pipeline of integrating the VOI map and a 3D U-Net for brain tumor segmentation. First, we register the VOI map from the MNI 1mm space to the subject space using FLIRT [12] from FSL [13], and this registered VOI map is then split into 9 binary masks. Each binary mask only contains information of one VOI label. Afterward, these binary masks are concatenated with the multimodal MR images. In the end, we input this 13-channel (4 image channels + 9 VOI channels) image to a 3D U-Net for both training and testing.

### 2.5. Evaluation Metrics

The employed evaluation metrics are the (i) Dice similarity coefficient (DSC) and the (ii) 95 percentile of the Hausdorff distance (H95). DSC is the quotient of similarity and ranges between 0 and 1 which is defined as

$$\text{DSC} = \frac{2|G \cap P|}{|G| + |P|}$$

where $|G|$ and $|P|$ are the number of voxels in the ground-truth label and predict label, respectively. Hausdorff distance $d_H(X, Y)$ measures how far two subsets $\{X, Y\}$ of a metric space are from each other which is defined as

$$d_H(X, Y) = \max\{\sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y)\}$$

where $d$ is the Euclidean distance, $\sup$ is the supremum, and $\inf$ is the infimum.

## 3. EXPERIMENTAL RESULTS AND DISCUSSION

First, we examine if the proposed lesion prior fusion method improves the brain tumor segmentation performance of the proposed 3D U-Net. Therefore, we train two identical 3D U-Nets with and without additional VOI map using 285 subjects of BraTS 2017 training set. BraTS 2017 validation set is used to evaluate the performance of these networks. The quantitative results are shown in Table 1. From the first two rows of Table 1, our proposed lesion prior fusion method improves the performance of 3D U-Net, particularly for the DSC of ET (3.5%), and H95 of ET (2.56) and whole tumor (2.39).

Second, we examine if the proposed lesion prior fusion method improves the performance of the ensemble of 3D U-Nets. Thus, we train two identical ensembles with and without additional VOI map using 285 subjects of BraTS 2017 training set. Each ensemble has five identical networks with different seed initializations, and the output of ensemble is
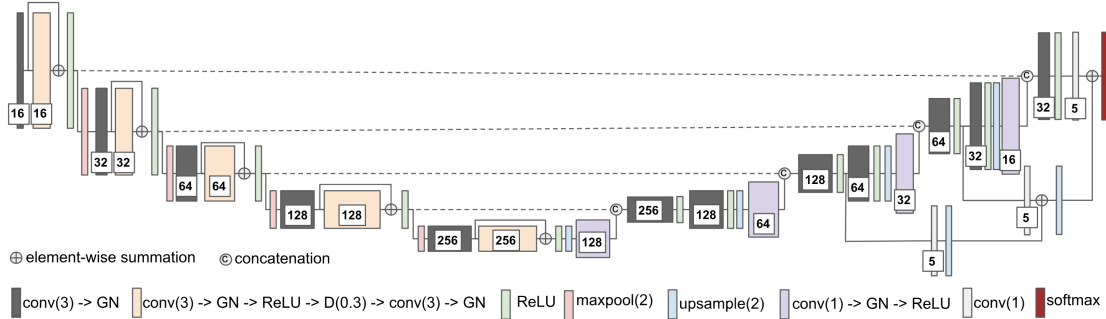
**Fig. 3**. The proposed network architecture. conv(3): $3 \times 3 \times 3$ convolutional layer, GN: group normalization, D(0.3): dropout layer with 0.3 dropout rate, maxpool(2): $2 \times 2 \times 2$ max pooling layer, and conv(1): $1 \times 1 \times 1$ convolutional layer. **Best viewed in color.**

**Table 1**. Quantitative results of the different models on BraTS 2017 validation set. Higher DSC and lower H95 indicate better segmentation performance. These results are given by the official online evaluation website. Results are reported as mean. Tumor core (TC) is the union of necrosis & non-enhancing tumor and enhancing tumor (ET). Whole tumor (WT) is the union of edema, necrosis & non-enhancing tumor and enhancing tumor. The underlined numbers highlight the improvement of VOI map, and the bold numbers highlight the best performance.

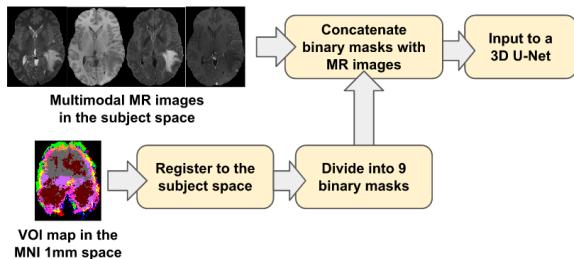| Model Descriptions | DSC | | | H95 | | |
|---|---|---|---|---|---|---|
| | ET | WT | TC | ET | WT | TC |
| Single 3D U-Net (baseline) | 0.695 | 0.896 | 0.762 | 6.79 | 6.92 | 11.38 |
| Single 3D U-Net + VOI (proposed) | <u>0.730</u> | <u>0.899</u> | <u>0.764</u> | <u>4.23</u> | <u>4.53</u> | <u>10.93</u> |
| Ensemble of five 3D U-Nets (baseline) | 0.723 | 0.902 | 0.763 | 5.99 | 4.75 | 10.58 |
| Ensemble of five 3D U-Nets + VOI (proposed) | **<u>0.744</u>** | **<u>0.903</u>** | <u>0.780</u> | <u>5.01</u> | **<u>3.86</u>** | <u>9.71</u> |
| Isensee et al. [2] | 0.732 | 0.896 | **0.797** | 4.55 | 6.97 | 9.48 |
| Kamnitsas et al. [4] | 0.738 | 0.901 | **0.797** | **4.50** | 4.23 | **6.56** |



**Fig. 4**. The pipeline of integrating the VOI map and a 3D U-Net.

averaged from five networks. BraTS 2017 validation set is used to evaluate the performance of ensembles, and the quantitative results are shown in Table 1. From the middle two rows of Table 1, our proposed lesion prior fusion method also improves the tumor segmentation performance of the ensemble of five 3D U-Nets, particularly for the DSC of ET (2.1%) and tumor core (1.7%). The reason why the VOI map has the greatest improvement on the ET is that the percentiles of ET heatmap have the highest priorities while we create the VOI map. In addition, the proposed VOI map, directly built

from the heatmaps of brain lesions, has inhomogeneous labels within neighboring voxels that carry more precise information of brain tumor lesions to the 3D U-Net.

In the end, we compare the performance of our proposed method with the state-of-the-art methods [2, 4]. From Table 1, the baseline model has worse performance than the state-of-the-art methods but it achieves a competitive performance by integrating the proposed VOI map. It is noted that the ensemble of Kamnitsas et al. [4] contains 7 different types of models but our proposed ensemble only consists of five 3D U-Net.

## 4. CONCLUSION

We have proposed a novel method to integrate prior information about the lesion probabilities into a 3D U-Net for improving brain tumor segmentation. Our experimental results demonstrate that the proposed lesion prior fusion approach improves the segmentation performance of the baseline model. Moreover, the proposed lesion prior fusion method can be easily integrated with other network architectures to further potentially enhance their segmentation performance.

# 5. REFERENCES

[1] Sarah Lapointe et al., "Primary brain tumours in adults," *The Lancet*, 2018.

[2] Fabian Isensee et al., "Brain tumor segmentation and radiomics survival prediction: contribution to the brats 2017 challenge," in *International MICCAI Brainlesion Workshop*. Springer, 2017, pp. 287–297.

[3] Fabian Isensee et al., "No new-net," in *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 234–244.

[4] Konstantinos Kamnitsas et al., "Ensembles of multiple models and architectures for robust brain tumour segmentation," in *International MICCAI Brainlesion Workshop*. Springer, 2017, pp. 450–462.

[5] Andriy Myronenko, "3d mri brain tumor segmentation using autoencoder regularization," in *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 311–320.

[6] Po-Yu Kao et al., "Brain tumor segmentation and tractographic feature extraction from structural mr images for overall survival prediction," in *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 128–141.

[7] Spyridon Bakas et al., "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," *Scientific data*, vol. 4, pp. 170117, 2017.

[8] Spyridon Bakas et al., "Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection," The Cancer Imaging Archive., 2017.

[9] Spyridon Bakas et al., "Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection," The Cancer Imaging Archive., 2017.

[10] Bjoern 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.

[11] GŘnther Grabner et al., "Symmetric atlasing and model based segmentation: an application to the hippocampus in older adults," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2006, pp. 58–66.

[12] Mark Jenkinson and Stephen Smith, "A global optimisation method for robust affine registration of brain images," *Medical image analysis*, vol. 5, no. 2, pp. 143–156, 2001.

[13] Mark Jenkinson et al., "Fsl," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.

[14] Özgün Çiçek et al., "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.

[15] Jiaxiang Jiang et al., "Accurate 3d cell segmentation using deep feature and crf refinement," *arXiv preprint arXiv:1902.04729*, 2019.

[16] Yuxin Wu and Kaiming He, "Group normalization," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.

[17] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2015.

[18] Sashank J. Reddi et al., "On the convergence of adam and beyond," in *International Conference on Learning Representations*, 2018.