Multimodal Brain Tumor Segmentation Using Cascaded V-Nets

Rui Hua[1,2], Quan Huo[2], Yaozong Gao[2], Yu Sun[1], and Feng Shi[2(✉)]

[1] School of Biomedical Engineering, Southeast University, Nanjing, China
[2] United Imaging Intelligence, Shanghai, China
feng.shi@united-imaging.com

Abstract. In this work, we propose a novel cascaded V-Nets method to segment brain tumor substructures in multimodal brain magnetic resonance imaging (MRI). Although V-Net has been successfully used in many segmentation tasks, we demonstrate that its performance could be further enhanced by using a cascaded structure and ensemble strategy. Briefly, our baseline V-Net consists of four levels with encoding and decoding paths and intra- and inter-path skip connections. Focal loss is chosen to improve performance on hard samples as well as balance the positive and negative samples. We further propose three preprocessing pipelines for multimodal MRI images to train different models. By ensembling the segmentation probability maps obtained from these models, segmentation result is further improved. In other hand, we propose to segment the whole tumor first, and then divide it into tumor necrosis, edema, and enhancing tumor. Experimental results on BraTS 2018 online validation set achieve average Dice scores of 0.9048, 0.8364 and 0.7748 for whole tumor, tumor core and enhancing tumor, respectively. The corresponding values for BraTS 2018 online testing set are 0.8761, 0.7953 and 0.7364, respectively. We further make a prediction of patient overall survival by ensembling multiple classifiers for long, mid and short groups, and achieve accuracy of 0.519, mean square error of 367239 and Spearman correlation coefficient of 0.168.

Keywords: Deep learning · Brain tumor · Segmentation · V-Net

1 Introduction

Gliomas are the most common brain tumors and comprise about 30% of all brain tumors. Gliomas occur in the glial cells of the brain or the spine [1]. They can be further categorized into low-grade gliomas (LGG) and high-grade gliomas (HGG) according to their pathologic evaluation. LGG are well-differentiated and tend to exhibit benign tendencies and portend a better prognosis for the patients. HGG are undifferentiated and tend to exhibit malignant and usually lead to a worse prognosis. With the development of the Magnetic Resonance Imaging (MRI), multimodal MRI plays an important role in disease diagnosis. Different MRI modalities are developed sensitive to different tissues. For example, T2-weighted (T2) and T2 Fluid Attenuation Inversion Recovery (FLAIR) are sensitive to peritumoral edema, and post-contrast T1-weighted (T1Gd) is sensitive to necrotic core and enhancing tumor core. Thus, they can provide complementary information about gliomas.

Segmentation of brain tumor is a prerequisite while essential task in disease diagnosis, surgical planning and prognosis [2]. Automatic segmentation provides quantitative information that is more accurate and has better reproducibility than conventional qualitative image review. Moreover, the following task of brain tumor classification heavily relies on the results of brain tumor segmentation. Automatic segmentation is considered as a powered engine and empower other intelligent medical application. However, the segmentation of brain tumor in multimodal MRI scans is one of the most challenging tasks in medical imaging analysis due to their highly heterogeneous appearance, and variable localization, shape and size.

As the rapid development of deep leaning techniques, state-of-the-art performance on brain tumor segmentation have been achieved. For example, in [3], an end-to-end training using fully convolutional network (FCN) showed a satisfactory performance in the localization of the tumor, and patch-wise convolutional neural network (CNN) was used to segment the intra-tumor structure. In [4], a cascaded anisotropic CNN was designed to segment three sub-regions with three Nets, and the segmentation result from previous net was used as receptive field in the next net.

Inspired by the good performance of V-Net in segmentation tasks and the cascaded strategy, we propose a cascaded V-Nets method to segment brain tumor into three substructures and background. In particular, the cascaded V-Nets not only take advantage of residual connection but also use the extra coarse localization and ensemble of multiple models to boost the performance.

2 Method

2.1 Dataset and Preprocessing

The data used in experiments come from BraTS 2018 training set and validation set [5–8]. The training set includes totally 210 HGG patients and 75 LGG patients. The validation set includes 66 patients. Each patient has five MRI modalities including T1-weighted (T1), T2, T1Gd, FLAIR, and a ground truth label of tumor substructures. We use 80% of the training data as our training set, other 20% of the training data as our local testing set. All data used in the experiments are preprocessed with special designed procedures. A flow chart of the proposed preprocessing procedures is shown in Fig. 1, as follows:

(1) Apply bias field correction N4 [9] to T1 and T1Gd images, normalize each modality using histogram matching with respect to a MNI template image, and rescale the images intensity value into range of −1 to 1.
(2) Apply bias field correction N4 to all modalities, compute the standardized z-scores for each image and rescale 0–99.9 percentile intensity values into range of −1 to 1.
(3) Follow the first method, and further apply affine alignment to co-register each image to the MNI template image.
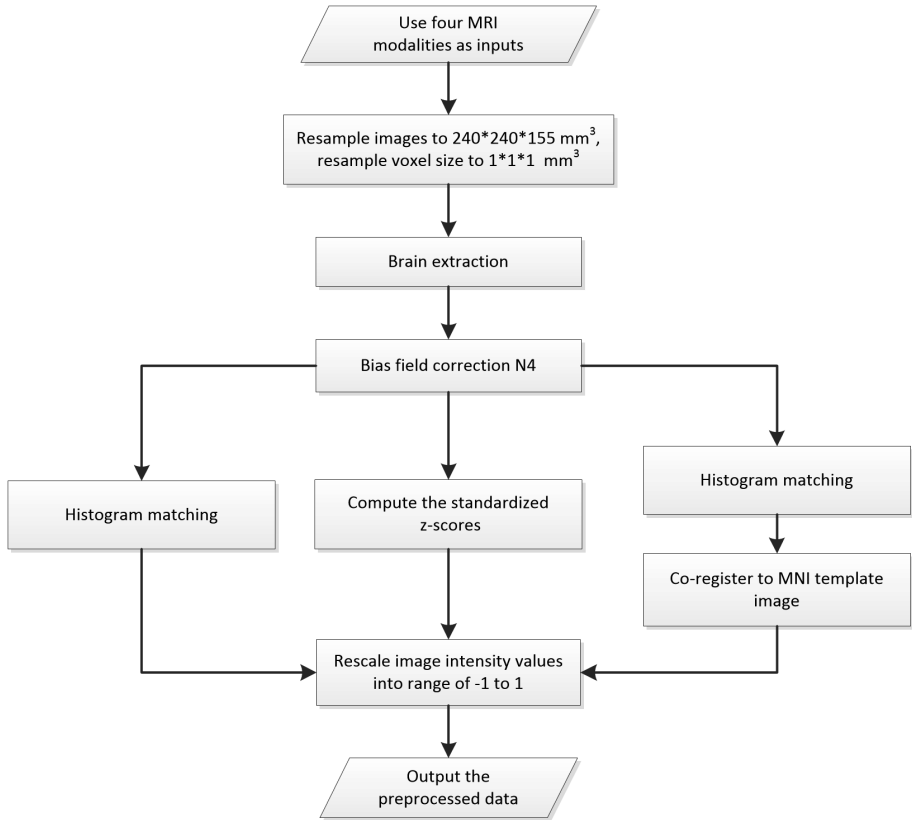
Fig. 1. The flow chart of the preprocessing procedures.

2.2 V-Net Architecture

V-Net was initially proposed to segment prostate by training an end-to-end CNN on MRI [10]. The architecture of our V-Net is shown in Fig. 2. The left side of V-Net reduces the size of the input by down-sampling, and the right side of V-Net recovers the semantic segmentation image that has the same size with input images by applying de-convolutions. The detailed parameters about V-Net is shown in Table 1. By means of introducing residual function and skip connection, V-Net has better segmentation performance compared with classical CNN. By means of introducing the 3D kernel with a size of 1 * 1 * 1, the numbers of parameters in V-Net is decreased and the memory consumption is greatly reduced.
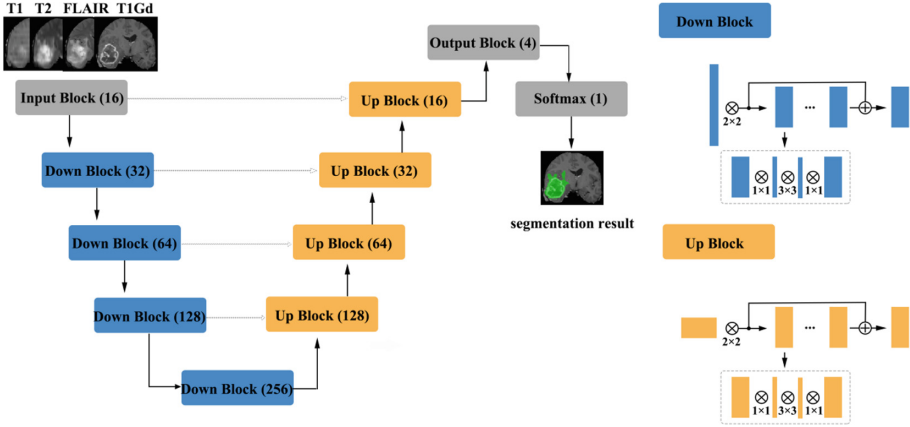
Fig. 2. The architecture of the used V-Net.

2.3   Proposed Cascaded V-Nets Framework

Although V-Net has demonstrated promising performances in segmentation tasks, it could be further improved if incorporated with extra information, such as coarse localization. Therefore, we propose a cascaded V-Nets method for tumor segmentation. Briefly, we (1) use one V-Net for the brain whole tumor segmentation; (2) use a second V-Net to further divide the tumor region into three substructures, e.g., tumor necrosis, edema, and enhancing tumor. Note that the coarse segmentation of whole tumor in the first V-Net is also used as receptive field to boost the performance. Detailed steps are as follows.

The proposed framework is shown in Fig. 3. There are two networks to segment substructures of brain tumors sequentially. The first network (V-Net 1) includes models 1–3, designed to segment the whole tumor. These models are trained by three kinds of preprocessed data mentioned in part of 2.1, respectively. V-Net 1 uses four modalities MR images as inputs, and outputs the mask of whole tumor (WT). The second network (V-Net 2) includes models 4-5, designed to segment the brain tumor into three sub-structures: tumor necrosis, edema, and enhancing tumor. These models are trained by the first two kinds of preprocessed data mentioned in part of 2.1, respectively. V-Net 2 also uses four modalities MR images as inputs, and outputs the segmented mask with three labels. Note that the inputs of V-Net 2 have been processed by using the mask of WT as region of interest (ROI). In other words, the areas out of the ROI are set as background. Finally, we combine the segmentation results of whole tumor obtained by V-Net 1 and the segmentation results of tumor core (TC, includes tumor necrosis and enhancing tumor) obtained by V-Net 2 to achieve more accurate results about the three

substructures of brain tumor. In short, the cascaded V-Nets take advantage of segmenting the brain tumor and three substructures sequentially, and ensemble of multiple models to boost the performance and achieve more accurate segmentation results.

Table 1. The detailed parameters of the used V-Net, as shown in Fig. 2. The symbol '-' means the output dimensions are the same with input dimensions.

Blocks	Sub-blocks or layers		Input dimensions	Output dimensions
Input Block	Conv(k = 3, p = 1, s = 1) + BN + ReLU		96 * 96 * 96 * 4	96 * 96 * 96 * 16
Down Block 1	Conv(k = 2, p = 0, s = 2) + BN + ReLU		96 * 96 * 96 * 16	48 * 48 * 48 * 32
	Residual Block	Conv(k = 3, p = 1, s = 1) + BN	48 * 48 * 48 * 32	-
		(input + output) + ReLU	48 * 48 * 48 * 32	-
Down Block 2	Conv(k = 2, p = 0, s = 2) + BN + ReLU		48 * 48 * 48 * 32	24 * 24 * 24 * 64
	Residual Block	Conv Block * 2	24 * 24 * 24 * 64	-
		(input + output) + ReLU	24 * 24 * 24 * 64	-
Down Block 3	Conv(k = 2, p = 0, s = 2) + BN + ReLU		24 * 24 * 24 * 64	12 * 12 * 12 * 128
	Residual Block	Conv Block * 3	12 * 12 * 12 * 128	-
		(input + output) + ReLU	12 * 12 * 12 * 128	-
Down Block 4	Conv(k = 2, p = 0, s = 2) + BN + ReLU		12 * 12 * 12 * 128	6 * 6 * 6 * 256
	Residual Block	Conv Block * 3	6 * 6 * 6 * 256	-
		(input + output) + ReLU	6 * 6 * 6 * 256	-
Up Block 1	Conv(k = 2, p = 0, s = 2) + BN + ReLU		6 * 6 * 6 * 256	12 * 12 * 12 * 128
	Residual Block	Cat(output, skip)	12 * 12 * 12 * 128	12 * 12 * 12 * 256
		Conv Block * 3	12 * 12 * 12 * 256	-
		(input + output) + ReLU	12 * 12 * 12 * 256	-
Up Block 2	Conv(k = 2, p = 0, s = 2) + BN + ReLU		12 * 12 * 12 * 256	24 * 24 * 24 * 64
	Residual Block	Cat(output + skip)	24 * 24 * 24 * 64	24 * 24 * 24 * 128
		Conv Block * 3	24 * 24 * 24 * 128	-
		(input + output) + ReLU	24 * 24 * 24 * 128	-
Up Block 3	Conv(k = 2, p = 0, s = 2) + BN + ReLU		24 * 24 * 24 * 128	48 * 48 * 48 * 32
	Residual Block	Cat(output + skip)	48 * 48 * 48 * 32	48 * 48 * 48 * 64
		Conv(k = 3, p = 1, s = 1) + BN + ReLU	48 * 48 * 48 * 64	-
		Conv(k = 3, p = 1, s = 1) + BN	48 * 48 * 48 * 64	-
		(input + output) + ReLU	48 * 48 * 48 * 64	-
Up Block 4	Conv(k = 2, p = 0, s = 2) + BN + ReLU		48 * 48 * 48 * 64	96 * 96 * 96 * 16
	Residual Block	Cat(output + skip)	96 * 96 * 96 * 16	96 * 96 * 96 * 32
		Conv(k = 3, p = 1, s = 1) + BN	96 * 96 * 96 * 32	-
		(input + output) + ReLU	96 * 96 * 96 * 32	-
Out Block	Conv(k = 1, p = 0, s = 1) + BN + ReLU		96 * 96 * 96 * 32	96 * 96 * 96 * 4
	Softmax		96 * 96 * 96 * 4	96 * 96 * 96 * 1

Note: Each Conv sub-block contains three convolution layers: Conv1(k = 1, p = 0, s = 1), Conv2(k = 3, p = 1, s = 1), and Conv3(k = 1, p = 0, s = 1). k, kernel size; p, padding; s, stride.
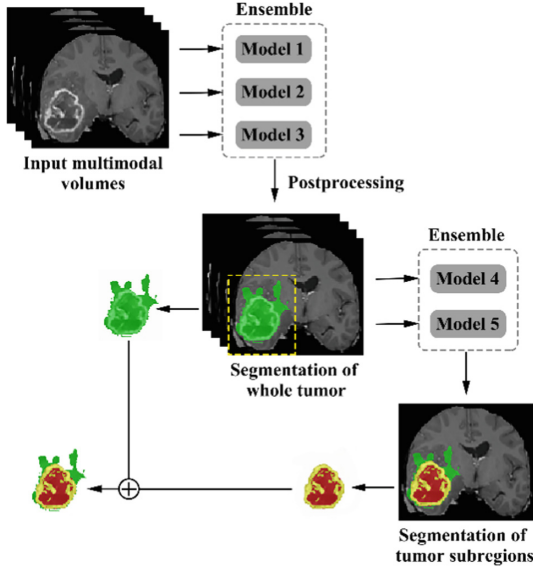
Fig. 3. The proposed framework of cascaded V-Nets for brain tumor segmentation.

2.4   Ensemble Strategy

Our ensemble strategy is simple but efficient. It works by averaging the probability maps obtained from different models. We use ensemble strategy twice in the two-step segmentation of the brain tumor substructures. For example, in V-Net 1, the probability maps of WT obtained from Model 1, Model 2, and Model 3 are averaged to get the final probability map of WT. In V-Net 2, the probability maps of tumor necrosis, edema, and enhancing tumor obtained from Model 4 and Model 5 are averaged to get final probability maps of brain tumor substructures, respectively.

2.5   Network Implementation

Our cascaded V-Nets are implemented in the deep learning framework PyTorch. In our network, we initialize weights with kaiming initialization [11], and use focal loss [12] illustrated in formula (1) as loss function. Adaptive Moment Estimation (Adam) [13] is used as optimizer with learning rate of 0.001, and batch size of 8. Experiments are performed with a NVIDIA Titan Xp 12 GB GPU.

$$\text{Focal Loss } (p_t) = -\alpha(1 - p_t)^r \log(p_t) \tag{1}$$

where, α denotes the weight to balance the importance of positive/negative samples, and r denotes the factor to increase the importance of correcting misclassified samples. p_t is the probability of the ground truth.

In order to reduce the memory consumption in the training process, 3D patches with a size of 96 * 96 * 96 are used. And the center of the patch is confined to the bounding box of the brain tumor. Therefore, every patch used in training process contains both tumor and background. The training efficiency of the network has been greatly improved.

2.6 Post-processing

The predicted segmentation results are post-processed using connected component analysis. We consider that the isolated segmentation labels with small size are prone to artifacts and thus remove them. After the V-Net 1, the components with total voxel number below a threshold (T = 1000) are discarded and these over a threshold (T = 15000) are retained in the binary whole tumor map. For others, their average segmentation probabilities are calculated, and will be retained if over 0.85. After the V-Net 2, masks of different labels are used in the connected component analysis. Moreover, if all the connected components are less than 1000 voxels, we will retain the largest connected component.

2.7 Prediction of Patient Overall Survival

Overall survival (OS) is a direct measure of clinical benefit to a patient. Generally, brain tumor patients could be classified into long-survivors (e.g., >15 months), mid-survivors (e.g., between 10 and 15 months), and short-survivors (e.g., <10 months). From the multimodal MRI data, we propose to use our tumor segmentations and generate imaging markers through Radiomics method to predict the patient OS groups.

From the training data, we extract 40 hand-crafted features and 945 radiomics features in total. The detailed extracted features are shown in Table 2. All features are normalized into range of 0 to 1. Pearson correlation coefficient is used for feature selection. We use support vector machine (SVM), multilayer perceptrons (MLP), XGBoost, decision tree classifier, linear discriminant analysis (LDA) and random forest (RF) as our classifiers in an ensemble strategy. F1-score is used as the evaluation standard. The final result is determined by the vote on all classification results. In order to reduce the bias, a ten-fold cross-validation is used. For the validation and testing data, these selected features are extracted and prediction is made using the above model.

Table 2. Selected features in the training data for the prediction of patient overall survival.

| Features	Number of features
Age	1
Volume of whole brain	1
Volume of whole tumor	1
Volumes of three tumor substructures	3
Ratio of the whole tumor in whole brain	1
Ratios of three tumor substructures in whole tumor	3
Extent of lesion in x, y, z directions	3
Center coordinates of the whole tumor	3
Means and variances of three tumor substructures in four MR modalities	24
First order statistics features of three tumor substructures	411
Shape-based features of three tumor substructures	78
Gray level cooccurence matrix features of three tumor substructures	180
Gray level run length matrix features of three tumor substructures	96
Neigbouring gray tone difference matrix features of three tumor substructures	96
Gray level dependence matrix features of three tumor substructures	84

3   Experimental Results

3.1   Segmentation Results on Local Testing Set

We use 20% of all data as our local testing set, which includes 42 HGG patients and 15 LGG patients. Representative segmentation results are shown in Fig. 4. The green shows the edema, the red shows the tumor necrosis, and the yellow shows the enhancing tumor. In order to evaluate the preliminary experimental results, we calculate the average Dice scores, sensitivity and specificity for whole tumor, tumor core and enhancing tumor, respectively. The results are shown in Table 3. The segmentation of whole tumor achieves best results with average Dice score of 0.8505.

3.2   Segmentation Results on MICCAI BraTS 2018 Validation Set of 66 Subjects

The segmentation results on BraTS 2018 online validation set achieve average Dice scores of 0.9048, 0.8364, 0.7768 for whole tumor, tumor core and enhancing tumor, respectively. That performance is slightly better than that in local testing set, while the whole tumor still has best results and enhancing tumor is the most challenging one. The details are shown in Table 4.
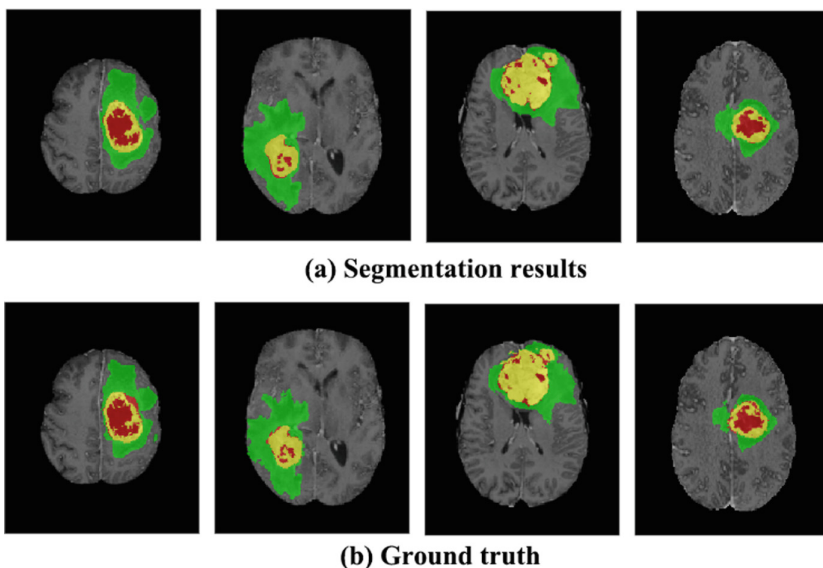
Fig. 4. The comparison of segmentation results and ground truth on four representative cases from local testing set. (a) The segmentation results of brain tumor. (b) The ground truth of the brain tumor. (Color figure online)

Table 3. Dice, Sensitivity and Specificity measurements of the proposed method on local testing set.

	Whole tumor	Tumor core	Enhancing tumor
Dice mean $\pm$ SD	0.8505 ± 0.0972	0.7842 ± 0.1919	0.7426 ± 0.2080
Sensitivity mean $\pm$ SD	0.9180 ± 0.1091	0.7596 ± 0.2199	0.7174 ± 0.2337
Specificity mean $\pm$ SD	0.9981 ± 0.0012	0.9996 ± 0.0008	0.9997 ± 0.0003

3.3 Segmentation and Prediction Results on MICCAI BraTS 2018 Testing Set of 191 Subjects

The segmentation results on BraTS 2018 online testing set achieve average Dice scores of 0.8761, 0.7953, 0.7364 for whole tumor, tumor core and enhancing tumor, respectively. Compared with the Dice scores on MICCAI BraTS 2018 validation set, the numbers are slightly dropped. The details are shown in Table 5. The prediction of patient OS on BraTS 2018 testing set achieve accuracy of 0.519 and mean square error (MSE) of 367239. The details are shown in Table 6. The BraTS 2018 ranking of all participating teams in the testing data for both tasks has been summarized in [14], where our team listed as "LADYHR" and ranked 18 out of 61 in the segmentation task and 7 out of 26 in the prediction task.

Table 4. Dice, Sensitivity, Specificity and Hausdorff95 measurements of the proposed method on BraTS 2018 validation set.

	Whole tumor	Tumor core	Enhancing tumor
Dice mean ± SD	0.9048 ± 0.0648	0.8364 ± 0.1609	0.7768 ± 0.2355
Sensitivity mean ± SD	0.9146 ± 0.0949	0.8453 ± 0.1781	0.8166 ± 0.2382
Specificity mean ± SD	0.9945 ± 0.0041	0.9971 ± 0.0041	0.9977 ± 0.0032
Hausdorff95 mean ± SD (mm)	5.1759 ± 7.3622	6.2780 ± 7.7681	3.5123 ± 4.5407

Table 5. Dice and Hausdorff95 measurements of the proposed method on BraTS 2018 testing set.

	Whole tumor	Tumor core	Enhancing tumor
Dice mean ± SD	0.8761 ± 0.1247	0.7953 ± 0.2543	0.7364 ± 0.2592
Hausdorff95 mean ± SD (mm)	7.0514 ± 11.5935	6.7262 ± 11.8852	3.9217 ± 6.1934

Table 6. The prediction of patient OS on BraTS 2018 testing set.

	Scores
Accuracy	0.519
Mean squared error (MSE)	367239.974
Median square error (Median SE)	38416
Standard deviation square error	945593.877
Spearman R	0.168

4   Discussion

In this paper, we propose a cascaded V-Nets framework to segment brain tumor. The V-Nets are trained only using provided data, data augmentation and a focal loss formulation. We achieve state-of-the-art results on BraTS 2018 validation set. The experimental results on BraTS 2018 online validation set achieve average Dice scores of 0.9048, 0.8364, 0.7768 for whole tumor, tumor core and enhancing tumor respectively. The corresponding values for BraTS 2018 online testing set are 0.8761, 0.7953 and 0.7364, respectively. Generally, all the three average Dice scores degenerate in testing set compared with validation set. Three are two possible reasons: (1) the testing set includes more cases than validation set, and (2) the thresholds in post-processing maybe more suitable for validation set. Therefore, our future work is to make the models to be more robust.

There are several benefits of using a cascaded framework. First, the cascaded framework breaks down a difficult segmentation task into two easier subtasks. Therefore, a simple network V-Net can have excellent performance. In fact, in our experiment, V-Net does have better performance when segment the tumor substructures step by step than segment background and all the three tumor substructures

together. Second, the segmentation results of V-Net 1 helps to reduce the receptive field from whole brain to only whole tumor. Thus, some false positive results can be avoid.

In addition to cascaded framework, ensemble strategy contributes to the segmentation performance. In our cascaded framework, V-Net 1 includes models 1–3 and V-Net 2 includes models 4–5. Every model uses the same network structure V-Net. However, the training data is preprocessed with different pipelines mentioned in part of 2.1. According to our experimental experience, the Dice scores will greatly decrease due to the false positive results. While we did try several ways to change the preprocessing procedures for the training data, or change the model used in the segmentation task, the false positive results always appear. Interestingly, the false positive results appear in different areas in terms of different models. Therefore, ensemble strategy works by averaging probability maps obtained from different models.

Moreover, we find three interesting points in the experiment. Firstly, for multimodal MR images, the combination of data preprocessing procedures is important. In other words, different MRI modalities should be preprocessed independently. For example, in our first preprocessing pipeline, bias field correction only applied to T1 and T1Gd images. The reason is that the histogram matching approach may remove the high intensity information of tumor structure that has negative impact to the segmentation task. Secondly, we use three kinds of preprocessing methods to process the training and validation data, and compared their segmentation results. As a result, there is almost no difference between preprocessing methods in the three average Dice scores for whole tumor, tumor core and enhancing tumor, respectively. However, after the ensemble of the multiple models, the three average Dice scores all rose at least 2%. This suggests that data preprocessing methods is not the most important factor for the segmentation performance, while different data preprocessing methods are complementary and their combination can boost segmentation performance. Thirdly, the postprocessing method is also important that it could affect the average Dices scores largely. If the threshold is too big, some of small clusters will be discarded improperly. If the threshold is too small, some false positive results will be retained. In order to have a better performance, we test a range of thresholds and choose the most suitable two thresholds as the upper and the lower bounds. For the components between upper and lower bounds, their average segmentation probabilities are calculated as a second criterion. Of course, these thresholds may not be suitable for all cases.

5   Conclusions

In conclusion, we propose a cascaded V-Nets framework to segment brain tumor into three substructures of brain tumor and background. The experimental results on BraTS 2018 online validation set achieve average Dice scores of 0.9048, 0.8364, 0.7768 for whole tumor, tumor core and enhancing tumor, respectively. The corresponding values for BraTS 2018 online testing set are 0.8761, 0.7953 and 0.7364, respectively. The state-of-the-art results demonstrate that V-Net is a promising network for 3D medical imaging segmentation tasks, and the cascaded framework and ensemble strategy are efficient for boosting the segmentation performance.

References

1. Mamelak, A.N., Jacoby, D.B.: Targeted delivery of antitumoral therapy to glioma and other malignancies with synthetic chlorotoxin (TM-601). Expert Opin. Drug Deliv. **4**, 175–186 (2007)
2. Bakas, S., et al.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci. Data **4**, 170117 (2017)
3. Cui, S., Mao, L., Jiang, J., Liu, C., Xiong, S.: Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded neural network. J. Healthc. Eng. **2018**, 4940593 (2018)
4. Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.): BrainLes 2017. LNCS, vol. 10670. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9
5. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**, 1993–2024 (2015)
6. Bakas, S., et al.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Nat. Sci. Data **4**, 170117 (2017)
7. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017)
8. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. The Cancer Imaging Archive (2017)
9. Tustison, N.J., et al.: N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging **29**, 1310–1320 (2010)
10. Milletari, F., Navab, N., Ahmadi, S.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, pp. 565–571 (2016)
11. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, pp, 1026–1034 (2015)
12. Lin, T., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. In: 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp, 2999–3007 (2018)
13. Kingma, D., Ba, J.: Adam, a method for stochastic optimization. In: International Conference on Learning Representations (ICLR), vol. 5 (2014)
14. Bakas, S., 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)