Multi-scale Masked 3-D U-Net for Brain Tumor Segmentation

Yanwu Xu[1], Mingming Gong[1,3], Huan Fu[2], Dacheng Tao[2], Kun Zhang[3], and Kayhan Batmanghelich[1(✉)]

[1] Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, USA
kayhan@pitt.edu
[2] UBTECH Sydney AI Centre, SIT, FEIT, The University of Sydney, Sydney, Australia
[3] Philosophy Department, Carnegie Mellon University, Pittsburgh, USA

Abstract. The brain tumor segmentation task aims to classify sub-regions into peritumoral edema, necrotic core, enhancing and non-enhancing tumor core using multimodal MRI scans. This task is very challenging due to its intrinsic high heterogeneity of appearance and shape. Recently, with the development of deep models and computing resources, deep convolutional neural networks have shown their effectiveness on brain tumor segmentation from 3D MRI cans, obtaining the top performance in the MICCAI BraTS challenge 2017. In this paper we further boost the performance of brain tumor segmentation by proposing a multi-scale masked 3D U-Net which captures multi-scale information by stacking multi-scale images as inputs and incorporating a 3-D Atrous Spatial Pyramid Pooling (ASPP) layer. To filter noisy results for tumor core (TC) and enhancing tumor (ET), we train the TC and ET segmentation networks from the bounding box for whole tumor (WT) and TC, respectively. On the BraTS 2018 validation set, our method achieved average Dice scores of 0.8094, 0.9034, 0.8319 for ET, WT and TC, respectively. On the BraTS 2018 test set, our method achieved 0.7690, 0.8711, and 0.7792 dice scores for ET, WT and TC, respectively. Especially, our multi-scale masked 3D network achieved very promising results enhancing tumor (ET), which is hardest to segment due to small scales and irregular shapes.

Keywords: Brain tumor segmentation · Multi-scale · ASPP · U-Net

1 Introduction

Multimodal Brain Tumor Segmentation Challenge (BraTS) [16] provides an excellent platform to boom the development of methods for segmenting tumor regions from 3D MRI scans as well as data [3,4]. As explained in [16], gliomas are the most common primary brain malignancies, with different degrees of aggressiveness, variable prognosis and various heterogeneous histological sub-regions,

A. Crimi et al. (Eds.): BrainLes 2018, LNCS 11384, pp. 222–233, 2019.
https://doi.org/10.1007/978-3-030-11726-9_20

i.e. peritumoral edema, necrotic core, enhancing and non-enhancing tumor core. In this challenge, our goal is to segment whole tumor (WT), tumor core (TC) and enhancing tumor (ET) from the other patterns. The dataset provided by [16] is composed of annotated low grade gliomas (LGG) and high grade glioblastomas (HGG), where LGG tends to be benign tendencies, while HGG denote the tumors which can grow rapidly and spread fast. For both LGG and HGG, four modals of scanning images are given, including Fluid Attenuation Inversion Recovery (FLAIR), T1-weighted (T1), contrast enhanced T1-weighted (T1ce) and T2-weighted (T2) images. Each modality supplies complementary information and they together provide more complete description of the tumor patterns. For example, the contours of whole tumor detected in FLAIR and T2 are more distinctive from the background than those in T1 and T1ce. Similarly, TC and ET can be easily distinguished from background in T1 and T1ce, the bounding information of which can be restricted by Flair and T2 by segmenting WT first. For instance, as mentioned in [16,21], T2 and FLAIR highlight the tumor with peritumoral edema, designated "whole tumor" as per [16]. T1 and T1ce highlight the tumor without peritumoral edema, designated "tumor core" as per [16]. An enhancing region of the tumor core with hyper-intensity can also be observed in T1ce, designated as "enhancing tumor core" [16].

Nowadays, we have a considerable quantity of diagnostic cases using Magnetic Resonance (MR) images. Moreover, we have the capacity to train very deep neural networks with the development of computing resources from these MR images, which makes automated disease diagnosis possible [2,16]. Automatic brain tumor segmentation can be much faster than manual segmentation; however, due to the irregular characteristics of brain tumor, the possibly subtle distinction between tumor and normal tissue, as well as a high variability in shape, location, and extent across patients, the accuracy of the current brain tumour segmentation algorithms needs further improvement so that they can be deployed in real systems.

There have been many methods for segmenting brain tumor which is detailed in [15]. Recently, the deep convolutional neural networks (CNNs) have shown promising performance in medical image segmentation and other related tasks [9–12,14,21]. DeepMedic [12] is one of the deep model-based method which combines patches with multiple resolutions as inputs to capture fine details and global information. They further introduced an enhancing structure which adds residual connection from previous feature layers. The 3-D U-Net [18] uses a compact encoder-decoder structure, which utilizes the features from several encoder layers twice by concatenating them with the decoder layers. Isensee et al. apply a U-Net based network to capture large scale information by large input patch size [10]. Additional works focus on the modification on the choice of convolutional kernal and loss function, such as the mixture of convolutional kernel and downsampling strategy [8,12]. Coping with unbalanced data, specific loss function [6,19] and sampling strategy [6] are introduced to train networks.

In this work, we focus on extracting multi-scale information from a single patch input instead of using multi-resolution inputs. Our contributions are

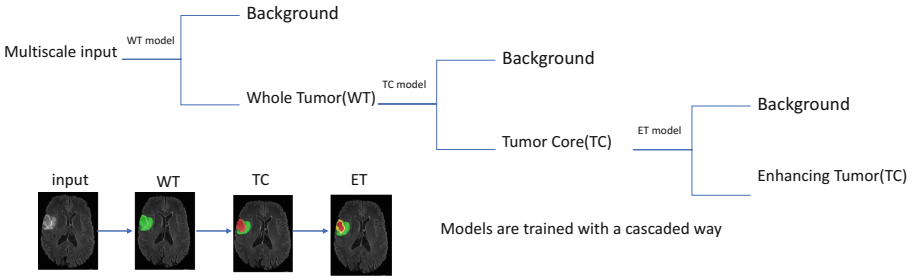


Fig. 1. The diagram depicts the training strategy for three different tumor with a cascaded masked way.

three fold. First, we extend 2D Atrous Spatial Pyramid Pooling (ASPP) [5] to 3D ASPP for extracting multi-scale information from feature maps of the neural network. By making use of the ASPP layer, we are able to enlarge the receptive field and thus capture larger scale information without introducing extra parameters. Second, we adjust basic structure of U-Net for small tumor segmentation by removing subsampling layers in specific layers of U-Net. This could help detect small tumors which are usually ignored in the original U-Net due to too many subsampling layers. Finally, we apply the cascaded masked strategy [21] for tumor segmentation training. Specifically, we segment WT, TC, and ET sequentially and use the bounding box from the former ones to restrict the search space for the following ones. This strategy could help remove false positive detections from the background regions. Our paper is collected in [1].

2 Methods

In this section, we will introduce the details of our method. First, we will describe the data preprocessing and patch extraction methods. Second, we will present the details of our network structure and training strategies.

2.1 Data Preprocess and Patch Extraction

We follow the standard procedure to preprocess the input images. To compensate for the MR inhomogeneity, we apply the bias correction algorithm based on N4ITK library [20] to the T1 and T1ce images. To reduce the effect of the absolute pixel intensities to the model, an intensity normalization step is applied to each volume of all subjects by subtracting the mean and dividing them by the standard deviation so that each MR volume will have a zero mean and unit variance. In practice, as the original uncropped volume is used but the brain only takes the central region, the mean and standard deviation are estimated from the brain area. Because of the GPU memory limitation and insufficient training data, we extract 400 patches per patient with patch size $64 \times 64 \times 64$ and take these patches as network inputs.

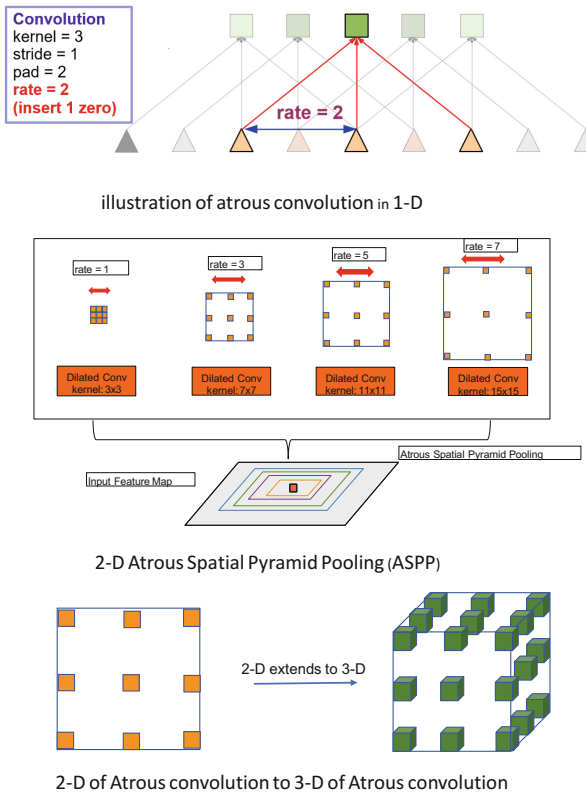


Fig. 2. The proposed extending ASPP layer. By the order of top to bottom, the dilated convolution, 2-D ASPP layer and the extending of 3-D ASPP layer from 2-D ASPP layer are well depicted.

2.2 Cascaded Masked Strategy

To remove false positive detections from background, we apply the cascaded masked strategy as [21]. The training strategy is shown in Fig. 1. By doing so, we can reduce the multiclass segmentation problem as a binary segmentation problem. Specifically, we train the WT network only with WT labeled data. Then we keep the segmented WT tumor as a mask for TC training. Similarly, we set segmented TC as the mask for ET training. Note that we use the groundtruth masks in the training phase, but the predicted masks in the test phase.

2.3 Extended 3-D ASPP Layer

Atrous Spatial Pyramid Pooling (ASPP) is first introduced in [5] for semantic segmentation in 2D natural scene images. ASPP layer consists of multiple scales of Atrous layers, also called dilated convolution layers. Figure 2 shows a one-dimensional Atrous convolutional operation. With the annotation for the output

$y[i]$ with respect to the 1-D input signal $x[i]$ and convolutional kernel $w[k]$, the formula is formed as follows:

$$y[i] = \sum_{k=1}^{K} x[i + r \cdot k]w[k] \tag{1}$$

Rate r denotes the dilated rate and dilated rate $r = 1$ is the normal convolution. Then the 2-D ASPP is displayed in Fig. 2, we feed feature maps into several Atrous layer with different rates and then combine these feature maps in channel dimension. Finally, we obey the same strategy and extend the 2-D ASPP layer to 3-D ASPP layer which is then applied on U-Net. We propose to apply ASPP layer here for capturing multi-scale objects and context employing multiple 3-D atrous convolutional layers with different sampling rates, and this is implemented with a parallel way. The advantage is that we can capture multi-scale information without introducing additional parameters and thus avoid overfitting.

2.4 Multi-scale Input

Additional, we try to include as much information as possible in the input. We rescale the images by multiple scales and then feed multi-resolution patches as input. Thus we apply a multi-scale input patches rather than only the patches of original size. In this work, the scales chosen are $\times 0.5$, $\times 1$ (original size) and $\times 2$, and these patches are concatenated in channel dimension. By doing so, we can extract global and local information even when the patch size is small.

2.5 Network Structure

Now we can stack the building blocks together to form the final network structure. Our network is based on U-net with 3-D ASPP layer for trade-off between scale information and receptive field as well as memory usage. The designation of network structure is shown in Fig. 3. Compared to the original U-Net, we remove the third downsampling and upsampling layers for WT network and TC network, furthermore, we add the ASPP layer between the encoder and decoder in U-Net. We observe that ET is small with respect to WT and TC and maybe evanescent after downsampling of encoder, which is not able to be recovered by upsampling of decoder. Thus, we only keep the second downsampling and upsampling layers for the ET network.

As for training network, we apply ADAM optimizer [13] and set the parameters of ADAM as $lr = 0.0002$, $\beta_1 = 0.5$ and $\beta_2 = 0.9999$, which is the unified setting. We apply Xavier initialization [7] to initialize the network parameters.

2.6 Loss Function

We adopt the cross entropy loss to train our networks. We classify each voxels to a binary label (1, 0: 1 means tumor and 0 means background), when

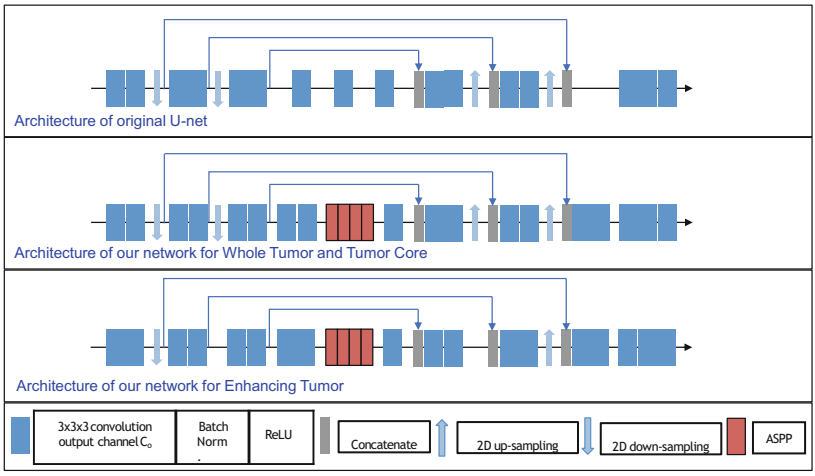


Fig. 3. Our modified U-Net based neural networks. We show the original U-Net on the top and the networks for our contribution is below. The networks trained by whole tumor and tumor core share the same network structure but do not share parameters. As for network designed for enhancing tumor training, we only keep one downsampling layer.

training network for WT, TC and ET separately. The cross entropy loss can be written as

$$loss = \sum (y' \log(y) + (1 - y') \log(1 - y)), \tag{2}$$

where y' represents ground truth label and y represents predicted label.

3 Experiments

Data. We got all our training data from BraTS web[1] to evaluate our method. The training data consist of 285 patients including segmented masks annotated by human experts. These training data are separated into two categories including HGG and LGG, containing 210 HGG and 75 LGG. There is an unbalance between HGG and LGG, and the data distributions of HGG and LGG are also different, especially for TC and ET. Each patient has four sequences, which are FLAIR, T2, T1, and T1ce. We feed all of the sequences into our network by combining them in channel dimension. Thus, our input data are 5-D, the dimension of which are batch, sequences, width, length, and depth. Regarding the validation data and testing data, they are the same as given training data, however the segmentation labels are not released. We finally receive validation data and testing data which are composed of 66 patients and 191 patients, respectively.

We train our whole network using Pytorch [17], which is a new hybrid front-end seamlessly transitions between eager mode and graph mode to provide both

[1] https://www.med.upenn.edu/sbia/brats2018/data.html.

flexibility and speed. We set our training batch size as 24 and training image size as $64 \times 64 \times 64$. We extract 400 patches for each patient, each patch consists of all of the FLAIR, T2, T1, and T1ce sequences as well as multiscale stacked patches. We choose NVIDIA TITAN XP GPU for training our network and it costs about 11 gigabytes GPU RAM. The whole training process is finished with 2 days with 10 epochs, and each epoch will traverse the whole training dataset.

3.1 Evaluation Metrics

Dice Coefficient. The Dice-Coefficient (Eq. 3) is calculated as performance metric. This measure states the similarity between clinical Ground Truth annotations and the output segmentation of the model. Afterwards, we calculate the average of those results to obtain the overall dice coefficient of the models.

$$D = \frac{2|A \bigcap B|}{|A| + |B|} \tag{3}$$

Hausdorff Distance. The Hausdorff Distance (Eq. 4) is mathematically defined as the maximum distance of a set to the nearest point in the other set [15], in other words how close are the segmentation and the expected output.

$$H(A, B) = max(min(d(A, B))) \tag{4}$$

Table 1. Mean values of Dice and Hausdorff measurements of the proposed method on BraTS 2018 validation set. ET, WT, TC denote enhancing tumor core, whole tumor and tumor core, respectively.

	Dice			Hausdorff (mm)		
	ET	WT	TC	ET	WT	TC
Original U-Net	0.739	0.882	0.788	5.329	7.356	10.243
Our network without ASSP layer	0.773	0.899	0.820	4.259	6.374	6.404
Our network with ASSP layer	0.809	0.903	0.832	3.780	6.022	7.091

Segmentation Results. To provide qualitative results of our method, we random choose two segmented images from validation data which are shown in Figs. 4 and 5 as well as in Appendix I. Figure 4 is suspected as one of the HGG data and Fig. 5 is suspected as one of the LGG data. Because the border for Fig. 4 is clear and the red non-enhancing tumor is inside the yellow enhancing tumor core. Furthermore, in Fig. 5, the border for tumor is quite blurred and there is almost no yellow enhancing tumor and it is in accordance with the feature of LGG data from training dataset. As we can observe from Fig. 4, the network with ASPP layer performs better than network without ASPP layer in that network with ASPP layer segments more local information that corresponds to the details shown in original sequences. As shown in Fig. 5, there

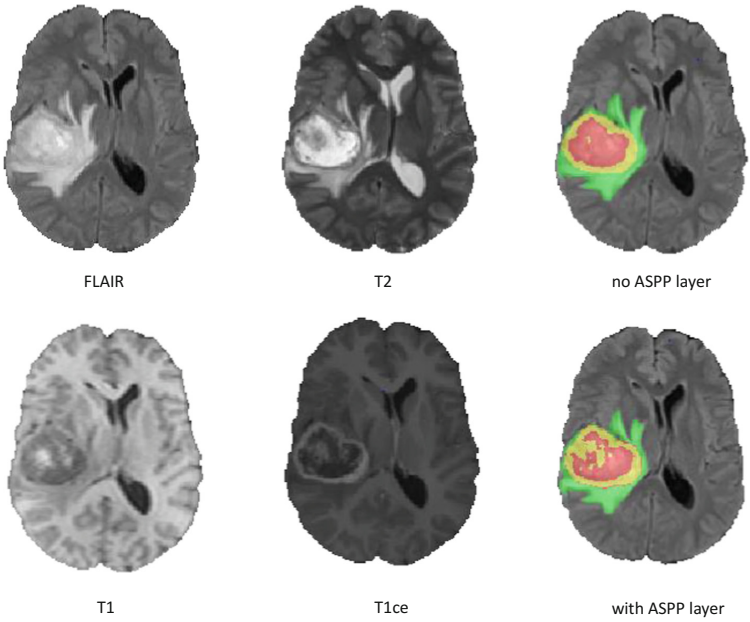


Fig. 4. Segmentation result of the brain tumor (suspected HGG) from a validation image. Green: edema; Red: non-enhancing tumor core; Yellow: enhancing tumor core. On the left, the original images are shown, and on the right, we show the segmented result of network without ASPP layer and network with ASPP layer. (Color figure online)

is a suspected wrong segmented area for red non-enhancing tumor, but the original sequences are blurred as well. In comparison, our network with ASPP layer can perform better on more local details and decrease the wrong classification for each voxel.

We show our quantitative results in Table 1. For comparison with existing methods, we list the result of the original U-Net, our modified U-Net without ASPP layer and our modified U-Net with ASPP layer. As can be seen from Table 1, our baseline of modified U-Net perform much better than the original U-Net in terms of all of the evaluation metrics. If just comparing the effect of ASPP layer, we find that assembling with ASPP layer can help improve TC and ET, especially improving ET by a large margin. However, they almost achieve the same performance on WT. We can also find that our method concentrate on detecting with multi-scale information that can help improve the ability for detecting small tumor area such as ET and TC. In this way, we achieve dice score above 0.8 for ET on testing data (Table 2).

As for testing data, we list the details of the result of mean value, standard deviation, median, 25% ranking and 75% ranking of Dice score and Hausdorff distance. Due to possible overfit on the validation data, we achieve a relatively lower performance on testing data; however our method still obtains rank 9^{th} out of all the submitted methods on testing data.

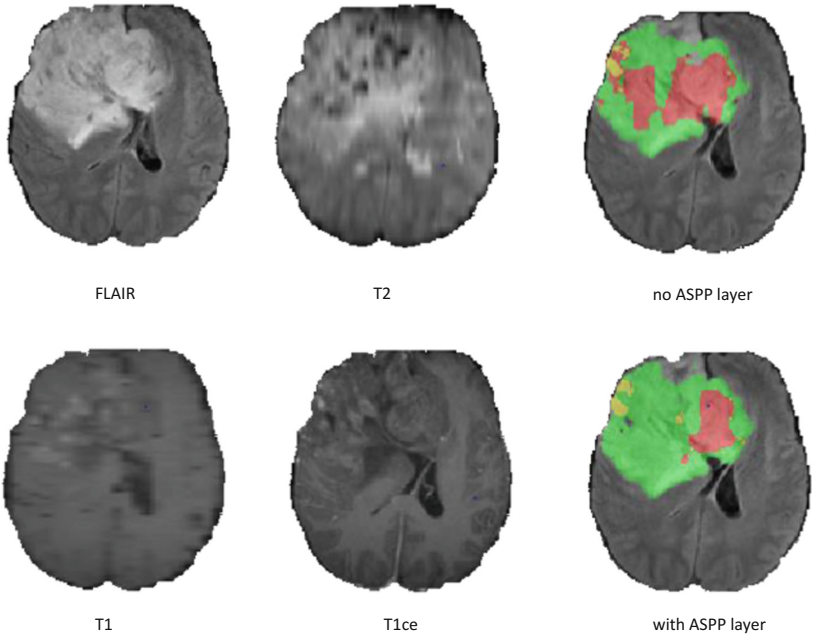


Fig. 5. Segmentation result of the brain tumor (suspected LGG) from a validation image. Green: edema; Red: non-enhancing tumor core; Yellow: enhancing tumor core. On the left, the original images are shown, and on the right, we show the segmented result of network without ASPP layer and network with ASPP layer. (Color figure online)

Table 2. Dice and Hausdorff measurements of the proposed method on BraTS 2017 testing set. EN, WT, TC denote enhancing tumor core, whole tumor and tumor core, respectively.

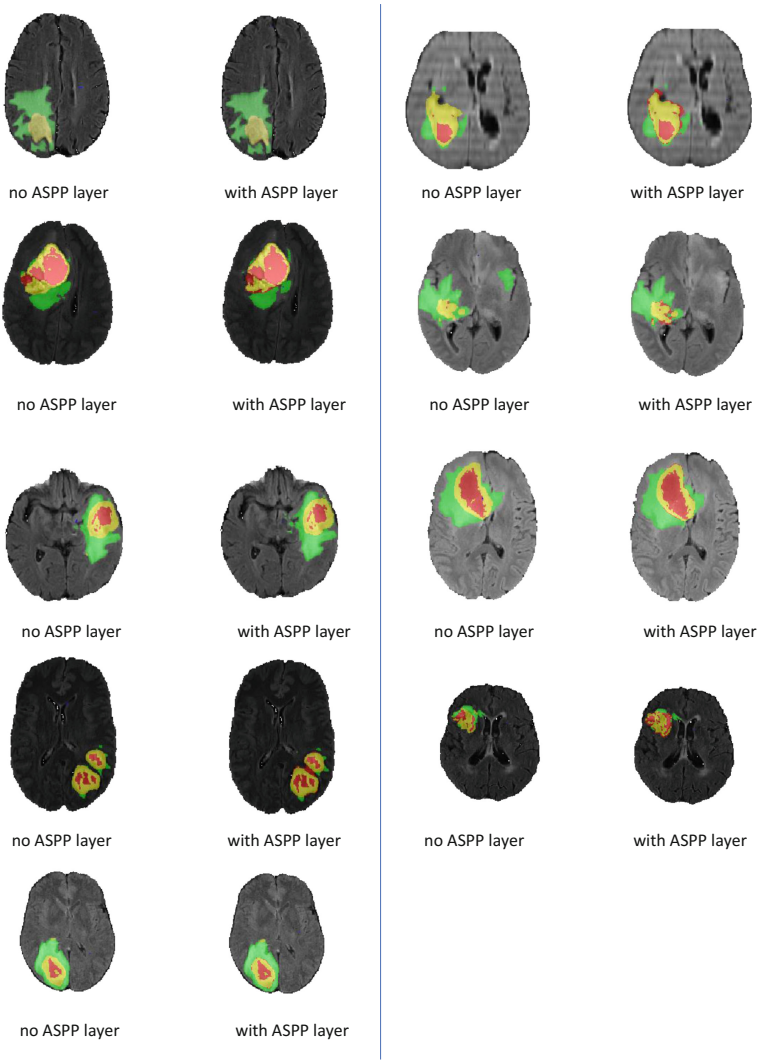
	Dice			Hausdorff (mm)		
	ET	WT	TC	ET	WT	TC
Mean	0.769	0.871	0.779	4.799	9.523	7.186
Standard deviation	0.240	0.129	0.274	9.293	16.822	10.900
Median	0.842	0.915	0.900	2.000	3.464	3.162
25 quantile	0.747	0.860	0.758	1.414	2.236	2.0000
75 quantile	0.892	0.939	0.936	3.000	6.364	7.280

4 Conclusions

We proposed a multi-scale neural network with a cascaded masked training structure for segmenting glioma subregions from multi-modal brain MR images. Our method receives as input multi-scale 3D patches extracted from the dataset volumes and we train three networks separately based on our cascaded

mask strategy. To further incorporate multi-scale information, we also incorporate the 3D ASPP layer which contains filter with various receptive field size without introducing many additional parameters. Our method achieves good results in the BraTS challenge. Future work would be incorporating attention in the network to aggregate multi-scale information.

A More Example



References

1. 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 e-prints, November 2018
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. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. The Cancer Imaging Archive 2017. https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q
4. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. The Cancer Imaging Archive 2017. https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF
5. Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. CoRR abs/1606.00915 (2016). http://arxiv.org/abs/1606.00915
6. Fidon, L., et al.: Generalised Wasserstein dice score for imbalanced multi-class segmentation using holistic convolutional networks. CoRR abs/1707.00478 (2017)
7. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS 2010). Society for Artificial Intelligence and Statistics (2010)
8. Havaei, M., et al.: Brain tumor segmentation with deep neural networks. CoRR abs/1505.03540 (2015)
9. Isensee, F., Jaeger, P., Full, P.M., Wolf, I., Engelhardt, S., Maier-Hein, K.H.: Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. CoRR abs/1707.00587 (2017)
10. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: contribution to the brats 2017 challenge. CoRR abs/1802.10508 (2018)
11. Kamnitsas, K., et al.: Ensembles of multiple models and architectures for robust brain tumour segmentation. CoRR abs/1711.01468 (2017)
12. Kamnitsas, K., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. **36**, 61–78 (2017). https://doi.org/10.1016/j.media.2016.10.004
13. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. CoRR abs/1412.6980 (2014)
14. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.W., Heng, P.A.: H-denseunet: hybrid densely connected UNet for liver and tumor segmentation from ct volumes. IEEE Trans. Med. Imaging (2018)
15. Liu, J., Li, M., Wang, J., Wu, F., Liu, T., Pan, Y.: A survey of MRI-based brain tumor segmentation methods. Tsinghua Sci. Technol. **19**(6), 578–595 (2014)
16. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2015)
17. Paszke, A., et al.: Automatic differentiation in pytorch. In: NIPS-W (2017)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. CoRR abs/1505.04597 (2015). http://arxiv.org/abs/1505.04597

19. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS -2017. LNCS, vol. 10553, pp. 240–248. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_28
20. Tustison, N.J., et al.: N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging **29**(6), 1310–1320 (2010). http://dblp.uni-trier.de/db/journals/tmi/tmi29.html#TustisonACZEYG10
21. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. CoRR abs/1709.00382 (2017). http://arxiv.org/abs/1709.00382