Multimedia Tools and Applications (2021) 80:22951–22964
https://doi.org/10.1007/s11042-020-08795-9

Path aggregation U-Net model for brain tumor segmentation

Fengming Lin[1,2] · Qiang Wu[1,2] · Ju Liu[1,2] · Dawei Wang[3] · Xiangmao Kong[1,2]

Received: 29 July 2019 / Revised: 29 January 2020 / Accepted: 24 February 2020 /
Published online: 19 March 2020

Abstract
The deep neural network has been widely used in semantic segmentation, especially in tumor image segmentation. The segmentation performance of traditional methods cannot meet the high standard of clinical application. In this paper, we propose a new neural network model called path aggregation U-Net (PAU-Net) model for brain tumor segmentation with multi-modality magnetic resonance imaging (MRI). Specifically, we shorten the distance between output layers and deep features by bottom-up path aggregation encoder (PA), reducing the introduction of noises. We present the enhanced decoder (ED) to reserve more intact information. The efficient feature pyramid (EFP) is used to improve mask prediction further, using fewer resources to complete the feature pyramid effect. Finally, experiments in BraTS2017 and BraTS2018 datasets are performed. The results show that the proposed method outperforms state-of-the-art methods.

Keywords Deep neural network · Brain tumor · Segmentation · Path aggregation U-Net · Multimodal MRI · Deep learning

1 Introduction

Harmful to human life and health, brain tumor [4] needs to be diagnosed by more accurate and efficient methods. The semantic segmentation algorithm based on deep learning has gradually become the mainstream method to tackle this disease. Multi-modality magnetic resonance imaging (MRI) images are applied to segment specific lesion areas of the brain tumor. The segmentation task of brain tumor needs to include multiple different lesion areas and require multi-modality data for voxel classification. Single-modality data do not take full advantage of modality correlations Bjoern H. Menze et al. [19] point out that different modalities specialize in different segmentation tasks.

✉ Qiang Wu
wuqiang@sdu.edu.cn

1 School of Information Science and Engineering, Shandong University, Qingdao, China

2 Institute of Brain and Brain-Inspired Science, Shandong University, Qingdao, China

3 Department of Radiology, Qilu Hospital, Shandong University, Jinan, China

Compared with natural images, medical images have fewer data samples and higher complexity [26], i.e., large gray value range and complex boundary. Based on deep learning, brain image registration [5], multimodal data analysis [29, 30] and brain function network construction [27] have all improved. Among the numerous segmentation methods based on convolutional neural networks, fully convolutional networks (FCN) [18] and U-Net [22] stand out, which are very suitable for medical image segmentation.

The deconvolution layer was first used in FCN to implement segmentation. FCN can take arbitrary scale images as input to solve the problem of repeated calculation of adjacent pixels. Besides, FCN only uses one upsampling layer, so it does not introduce many noises. However, the capacity of the decoder with only one upsampling layer is insufficient. It can not fuse a large amount of useful information.

Therefore, U-Net was proposed to fuse more information to improve segmentation accuracy. This structure increases the spatial consistency by combining features of the same scale between upsampling blocks and downsampling blocks. There are many variants of U-Net, such as combinations with modules like ResNet [8, 20], DenseNet [9, 11, 23]. However, excessive noises are introduced in multiple upsampling processes with deep structure. Also, the decoder in U-Net, which is symmetric with the encoder, is insufficient to accommodate features from the encoder and the features from the previous block of decoder. In addition, U-Net cannot achieve the fusion of detailed information and spatial information.

Feature pyramid networks (FPN) [16], RefineNet [15] and Laplacian pyramid [6] use feature pyramid structure [1] for object detection. HPU-Net [13] and FMNet [14] apply the idea of multi-layer feature fusion to brain tumor segmentation, achieving outstanding performance. However, the upsampling process needs too much graphic memory resources, which is not conducive to segmentation when computing resources are limited.

Furthermore, the path aggregation network [17] is improved based on the FPN and the structure of Mask R-CNN [7]. Firstly, the new bottom-up branches are added in the structure of the feature pyramid to reduce the loss of deep information. Also, adaptive feature pooling is used to achieve feature fusion by operating the maximum pixel value of features in each layer. Then, it combines the FCN with the fully connected layer. The idea of path aggregation has important guiding significance for brain tumor segmentation. Nevertheless, this approach still cannot tackle the problem of consuming too many computing resources.

In general, the image segmentation algorithms based on deep learning have been improved in the field of medical imaging, while there are still three main problems. (1) Excessive noises are introduced in multiple upsampling processes with deep structure. (2) The capacity of the decoder is insufficient. (3) The feature pyramid upsampling process needs too much graphic memory resources.

In this paper, we propose a path aggregation U-Net (PAU-Net) model to improve the brain tumor segmentation performance. These three problems above have been solved by using the following three structures. (1) The path aggregation encoder is employed to facilitate the dissemination of deep features and shorten the distance between deep features and the output layer in the model. (2) The enhanced decoder is used to improve the performance of the decoder compared with the traditional structure symmetric with encoder. (3) The efficient feature pyramid is employed to connect multi-level features with less memory resources and output the segmentation results.

The rest of this paper is organized as follows. Section 2 presents the proposed method, including path aggregation encoder, enhanced decoder, and efficient feature pyramid. Section 3 describes the datasets we use and the setup of the experiments. The experimental results are also provided in this section. Section 4 comprises the discussion. Finally, Section 5 provides the conclusion of this paper.

2 Materials and methods

The network structure of the PAU-Net model is shown in Fig. 1. The model is composed of four parts, namely encoder, enhanced decoder, path aggregation encoder, and efficient feature pyramid. As shown in Fig. 1a, five downsampling blocks are set up in the encoder to downsample the original image through the convolutional layer and pooling layer. In Fig. 1b, the enhanced decoder with four upsampling blocks improves the traditional encoder and increases the number of channels for the decoder. In a conventional U-Net, the number of channels for the decoder is equal to the number of channels in the encoder. The channel number same with encoder cannot accommodate features from encoder and features from the previous block of decoder. Therefore, we use an enhanced decoder to accommodate much more information. In Fig. 1c, path aggregation encoder contains three downsampling blocks. In the feature extraction of the above two structures, the features in each layer extracted by the encoder can only reach the final output layer after going through multiple convolutional layers. Therefore, we propose a path aggregation encoder, which shortens the distance of feature propagation and reflects the function of features in each layer more directly. As shown in Fig. 1d, an efficient feature pyramid is employed to reduce the usage of memory.

2.1 Encoder

In the encoder part, the training data are downsampled, and the context information is extracted. The encoder consists of five downsampling blocks, the first four of which are downsampling blocks, and the last one is a non-downsampling block. The downsampling region consists of two convolutional layers and one downsampling layer. The kernel size of the convolutional layer is 3, and the step size is 1. The last non-downsampling block is the one that removes the pooling layer. Its convolution kernel number is 512. A batch normalization [10] layer is added after each convolutional layer to remedy the vanishing gradient problem. The features after batch normalization are activated by Rectified Linear Unit (ReLU). When the data pass through the downsampling block, both the width and height become 1/2 of the original scale. The output channel number of the encoder is 512.

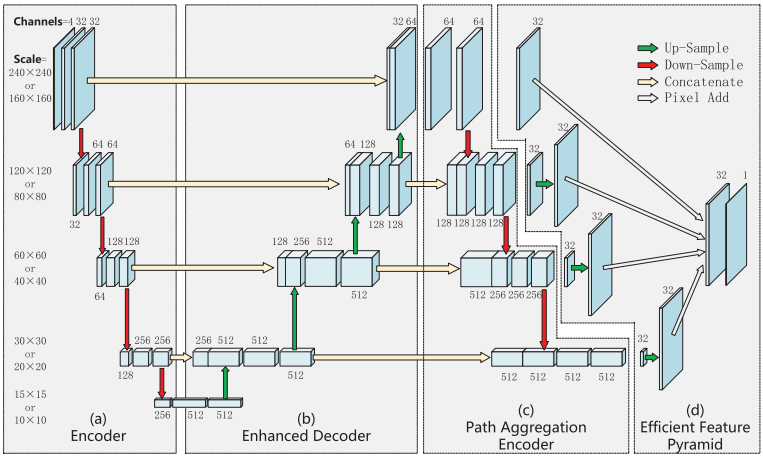

Fig. 1 The structure of PAU-Net

The width and height of output are 1/16 of the initial features. The implementation is shown in Fig. 2.

2.2 Enhanced decoder (ED)

The enhanced decoder consists of four upsampling blocks; each block is composed of an upsampling layer, a convolutional layer, a concatenate layer and two convolutional layers. The upsampling layer adopts the interpolation algorithm. The concatenate layer connects features of the same scale in the encoder part with features after upsampling and convolution in the enhanced decoder. The size of the convolution kernels of the three convolutional layers is 3, and the number of convolution kernels from the first to the fourth upsampling block is 512, 512, 128 and 64, respectively. The asymmetric number of convolution kernels is adopted to enhance the feature resolution of the decoder. After each convolutional layer, a batch normalization layer and a ReLU activation layer are added. When the feature passes through the upsampling area, both the width and height are doubled, and the output channel

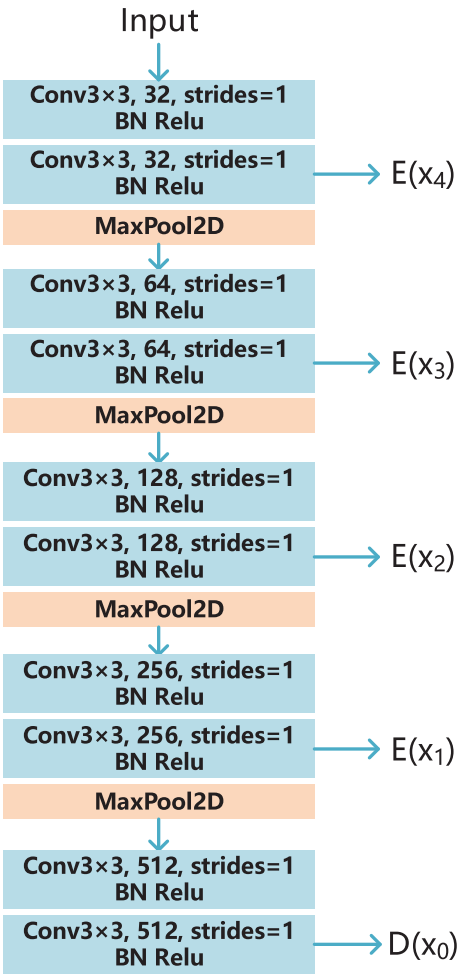

Fig. 2 The implementation of encoder in PAU-Net

number of the final upsampling block is 64. The width and height of the output are the same as the original scale. The implementation is shown in Fig. 3, and the calculation process is as follows:

$$D(x_i) = D(x_{i-1}) \oplus E(x_i). \tag{1}$$

where $D(x_i)$ is output of the ith upsampling block in enhanced decoder. $D(x_{i-1})$ is output of the $(i-1)$th upsampling block in enhanced decoder. $E(x_i)$ is output of the ith downsampling block in encoder. $\oplus$ means concatenate operation. We regard $D(x_4)$ as $A(x_4)$.

2.3 Path aggregation encoder (PA)

The path aggregation encoder contains three downsampling blocks. Each block consists of a downsampling layer, a concatenate layer, and two convolution layers. In this case, the convolutional layer with a step size of two is adopted for the downsampling process. The concatenate layer connects the features of the same scale in the enhanced decoder with the features in the path aggregation encoder. The size of the convolution kernel is 2 and 3. The number of convolution kernels from the first to the third downsampling block is 128, 256, and 512, respectively. A batch normalization layer follows each convolutional layer, and the ReLU activation layer is used after the batch normalization layer. When the features pass through the downsampling block, the width and height become 1/2 of the original scale. The

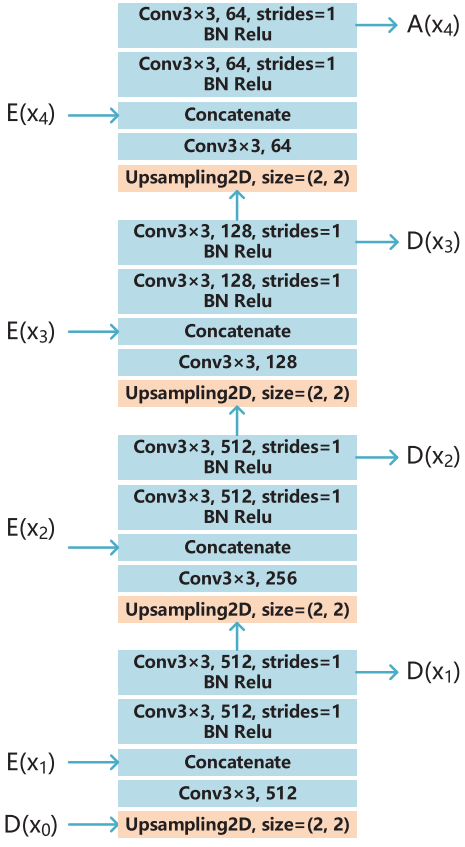

Fig. 3 The implementation of enhanced decoder in PAU-Net

training data pass through three path aggregation blocks in total. We also regard the input $A\,(x_4)$ of the path aggregation encoder as one of the outputs of it. Therefore, we get four features with the length and width of 1/1, 1/2, 1/4, and 1/8 of the original scale. The number of channels is 64, 128, 256, and 512, respectively. These features are the output of path aggregation encoder. The implementation is shown in Fig. 4, and the calculation process is as follows:

$$A\,(x_i) = A\,(x_{i+1}) \oplus D\,(x_i)\,. \tag{2}$$

where $A\,(x_i)$ is output of the ith downsampling block in path aggregation encoder. $A\,(x_{i+1})$ is output of the $(i+1)$th downsampling block in path aggregation encoder.

2.4 Efficient feature pyramid (EFP)

In the efficient feature pyramid, we fuse the output features of different levels from the path aggregation encoder. We firstly compress the channel number of the features to the minimum value, which is 32. Then we upsample them to the original scale. After that, we add them up in pixels and take them through the convolutional layer and the softmax layer. If the features are directly up-sampled without channel compression, we call this a feature pyramid (FP). Features with more channels in the FP will take up more storage space after upsampling. The upsampling method is an interpolation algorithm. The convolutional layer in the upsampling block follows the batch normalization layer and the ReLU layer. After adding these pixel values, the output feature is activated by using the softmax function to get the final result of segmentation. The implementation is shown in Fig. 5, and the calculation process is as follows:

$$P(x) = \sum_{i=1}^{n} A\,(x_i)\,. \tag{3}$$

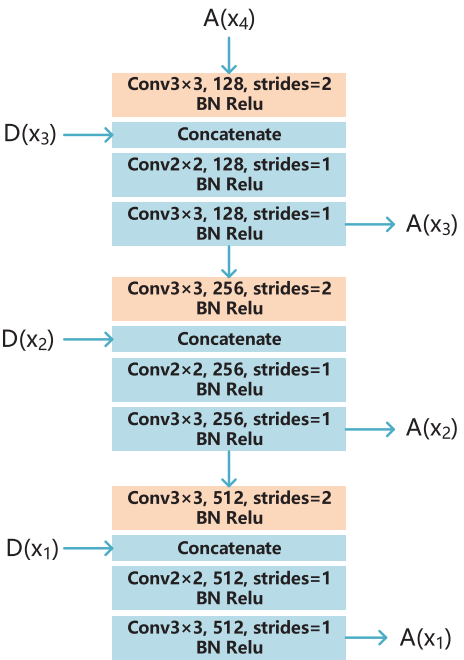

Fig. 4 The implementation of path aggregation encoder in PAU-Net

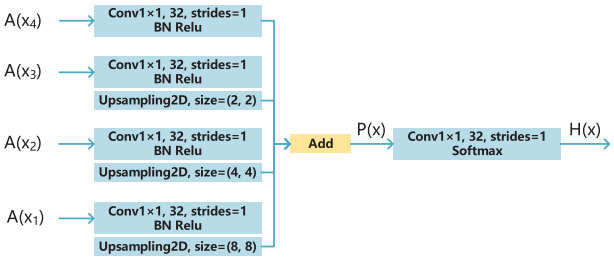

Fig. 5 The implementation of efficient feature pyramid in PAU-Net

where $P(x)$ is the output of an efficient feature pyramid.

The final output of the PAU-Net model is shown as follows:

$$\begin{aligned} H(x) &= Softmax\left[P(x)\right] \\ &= Softmax\left[\sum_{i=1}^{n} A\left(x_i\right)\right] \\ &= Softmax\left[\sum_{i=1}^{n}\left[A\left(x_{i+1}\right) \oplus D\left(x_i\right)\right]\right] \\ &= Softmax\left[\sum_{i=1}^{n}\left[A\left(x_{i+1}\right) \oplus D\left(x_{i-1}\right) \oplus E\left(x_i\right)\right]\right]. \end{aligned} \tag{4}$$

where $H(x)$ is the output of the PAU-Net model.

3 Experiments

In this section, the proposed path aggregation U-Net model is evaluated on the brain tumor segmentation 2017 dataset (BraTS2017) and brain tumor segmentation 2018 dataset (BraTS2018) [19] [3] compared with FCNN [2] , VGG [24], DUNet [12].

3.1 Dataset and preprocessing

We use BraTS2018 and BraTS2017 datasets to evaluate the performance of the baseline methods and our proposed method. Annotations comprise the GD-enhancing tumor (ET-label 4), the peritumoral edema (ED-label 2), and the necrotic and non-enhancing tumor core (NCR/NET-label 1). ET corresponds to Enhancing Tumor. ED corresponds to Whole Tumor. NCR/NET corresponds to Tumor Core. The brain MRI data contain four modalities: T1, T1c, T2, and Flair.

The boundary of the medical image is complex and the distribution of gray value is unbalanced. Therefore, we need to leverage multi-modality features in MRI and preprocess these images [29]. At first, the images of four modalities are normalized to adjust the gray value distribution of MRI. Then the images of the four modalities are fused to form a four-channel array as the final multi-modality array. Finally, the dataset is shuffled.

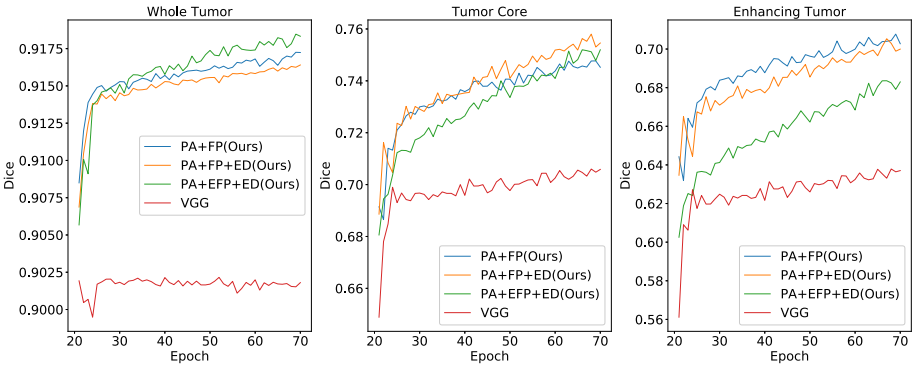

Fig. 6 Dice on BraTS2017 dataset. VGG had the best performance in the baseline comparison experiments in this paper, so VGG was used to compare with the three proposed methods. The three methods are better than VGG in different tasks

3.2 Baseline system

In the comparison experiments, VGG applies VGGNet [25] structure on brain tumor segmentation. DUNet is a method to deepen the model of U-Net, which is more effective than the traditional U-Net segmentation method. DUNet may not converge in cross-validation, therefore we use the convergence results to calculate the mean value in the experiments of DUNet. FCNN uses the traditional fully convolutional neural network structure on brain tumor segmentation.

The experiments are performed by a GTX-1080ti graphics card with 11G memory. Since in the BraTS2017 dataset, the scale of brain image does not exceed 160×160. We resample it from 240×240 to 160×160. Therefore, in the experiments with BraTS2017 dataset, the batch size of all experiments is 16. However, in the experiment with the BraTS2018 dataset, the data size is 240×240. Therefore, in FCNN, VGG, DUNet, PA+FP, and PA+FP+ED, the batch is 8. The efficient feature pyramid is used to reduce resource usage in PA+EFP+ED. Therefore, we can set the batch as twice of the comparison experiments as 16. We use Adam

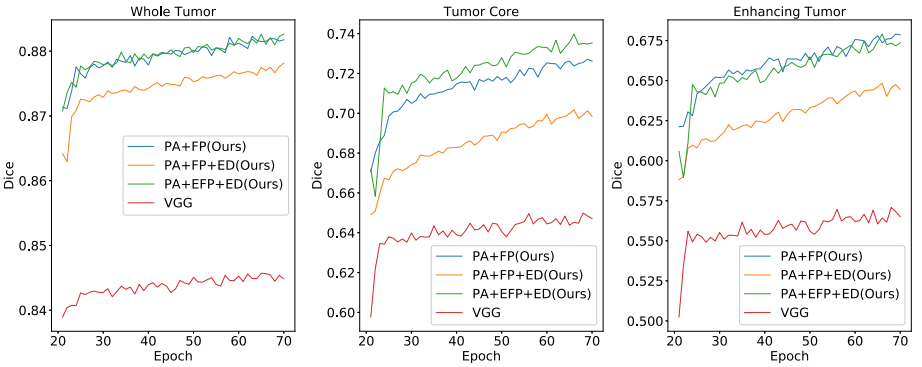

Fig. 7 Dice on BraTS2018 dataset. After preprocessing, the size of each slice in BraTS2018 is more than twice that of BraTS2017. The advantage of efficient feature pyramids in reducing features is evident when the data size is large

as an optimizer, and use categorical cross-entropy as the loss function. The epoch of each comparison experiment is 70. The learning rate is defined as *LR*. Instead of setting the *LR* as a fixed value, we adopt a piecewise function as follows:

$$LR = \begin{cases} 0.003 - \frac{(0.003-0.00003)*Epochs}{24} & 0 \leqslant Epochs \leqslant 24, \\ 0.00003 & 25 \leqslant Epochs \leqslant 70. \end{cases} \tag{5}$$

3.3 Results

We employ cross-validation to evaluate the performance of all methods. All the datasets are divided into five parts, and four parts are selected as the training set, and the remained part is the testing set.

Figures 6 and 7 show the dice results on BraTS2017 and BraTS2018 datasets. The influence of path aggregation encoder, enhanced decoder, and efficient feature pyramid on dice is demonstrated respectively. It can be seen from Tables 1 and 2 that in the baseline experiment of VGG, DUNet and FCNN, VGG has the best segmentation performance with DICE, PPV and Sensitivity in WT, TC and ET. Therefore, in Figs. 6 and 7, we took the result of VGG as the representative of the baseline experiment and compared it with PA+FP, PA+FP+ED, PA+EFP+ED. The result of VGG is shown in the red curve. The results of PA+FP, PA+FP+ED and PA+EFP+ED are represented by blue, yellow and green curves respectively. In the curve chart, the dice value of our method is higher than VGG, which proves that our method is more effective than state-of-the-art methods.

Table 1 shows the comparison experiment on the BraTS2018 dataset. Table 2 shows the comparison experiment on the BraTS2017 dataset. In the above two datasets, our method is better than state-of-the-art methods.

Both Tables 1 and 2 show that path aggregation encoder and enhanced decoder significantly improve segmentation performance. The enhanced decoder is valid for the segmentation of TC and ET. In Table 1, for the PA+FP and PA+FP+ED experiments, the batch size is 8. This reaches the upper limit of a single GTX-1080Ti. Because the efficient feature pyramid reduces the occupation of computing resources, we can expand the batch to 16 for the PA+EFP+ED experiment. However, in Table 2, we adopt different data preprocessing methods. The image is cropped to 160 × 160. This approach reduces the image

Table 1 Comparison with state-of-the-art methods on the testing set of BRATS2018 dataset

BraTS2018		VGG	DUNet	FCNN	PA+FP	PA+FP+ED	PA+EFP+ED
Dice	WT	0.8725	0.8563	0.8743	0.8796	0.8797	**0.8827**
	TC	0.6876	0.6751	0.6674	0.7060	0.7044	**0.7257**
	ET	0.6551	0.6002	0.6040	0.6518	0.6507	**0.6738**
PPV	WT	0.9237	0.9214	0.9244	0.9258	0.9266	**0.9320**
	TC	0.9214	0.9190	0.9210	0.9224	0.9238	**0.9287**
	ET	0.9172	0.9139	0.9137	0.9189	**0.9213**	0.9205
Sen	WT	0.9596	0.9540	0.9585	0.9604	0.9640	**0.9654**
	TC	0.8286	0.8155	0.8285	0.8330	0.8394	**0.8487**
	ET	0.6277	0.6102	0.6295	0.6333	0.6423	**0.6628**

PA+FP, PA+FP+ED, and PA+EFP+ED are our methods. The bold entries are the maximum value of the line. The batch of PA+FP and PA+FP+ED both reach the upper limit of 8, while the batch of PA+EFP+ED can reach the upper limit of 16. Under the same graphic memory condition, PA+EFP+ED performs better

Table 2 Comparison with state-of-the-art methods on the testing set of BRATS2017 dataset

BraTS2017		VGG	DUNet	FCNN	PA+FP	PA+FP+ED	PA+EFP+ED
Dice	WT	0.9047	0.9000	0.9065	0.9157	**0.9166**	0.9154
	TC	0.7014	0.7095	0.6619	0.7308	**0.7473**	0.7386
	ET	0.6569	0.6357	0.5810	0.6751	**0.6892**	0.6742
PPV	WT	0.9250	0.9241	0.9237	0.9282	**0.9291**	0.9284
	TC	0.9223	0.9206	0.9204	0.9250	**0.9256**	0.9248
	ET	0.9182	0.9147	0.9149	**0.9220**	0.9208	0.9206
Sen	WT	0.9608	0.9583	0.9581	0.9634	**0.9637**	0.9632
	TC	0.8322	0.8281	0.8271	0.8408	**0.8424**	0.8408
	ET	0.6326	0.6318	0.6254	0.6447	**0.6487**	0.6455

PA+FP, PA+FP+ED, and PA+EFP+ED are our methods. The bold entries are the maximum value of the line. When the batch is large enough, the efficient feature pyramid is not needed to increase the batch. At this point, the batch size for all experiments is 16. Therefore, the performance of the feature pyramid is better than that of the efficient feature pyramid. Under the same batch size condition, PA+FP+ED performs better

size by almost half compared to BraTS2018 dataset. Therefore, the batch sizes of PA+FP, PA+FP+ED and PA+EFP+ED were all raised to 16. This indicates that when the batch is large enough, the efficient feature pyramid is not needed to increase the batch, and the performance of the feature pyramid is better than that of the efficient feature pyramid.

4 Discussion

In MRI images, the region size of normal tissue is much larger than that of tumor lesions. Therefore, this classification is an unbalanced problem. The deep neural networks are used to mine deeper and more representative features, and the path aggregation encoder is used

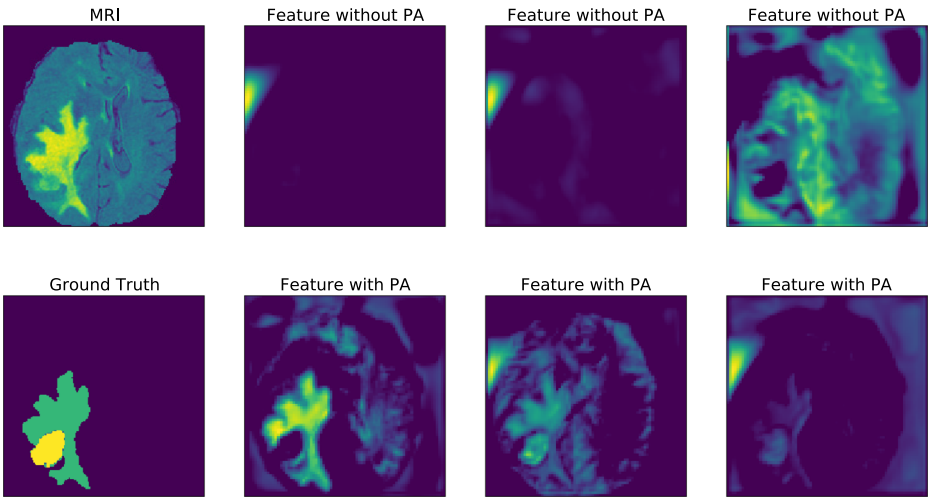

Fig. 8 The comparison of the intermediate features of whether there is a path aggregation encoder or not

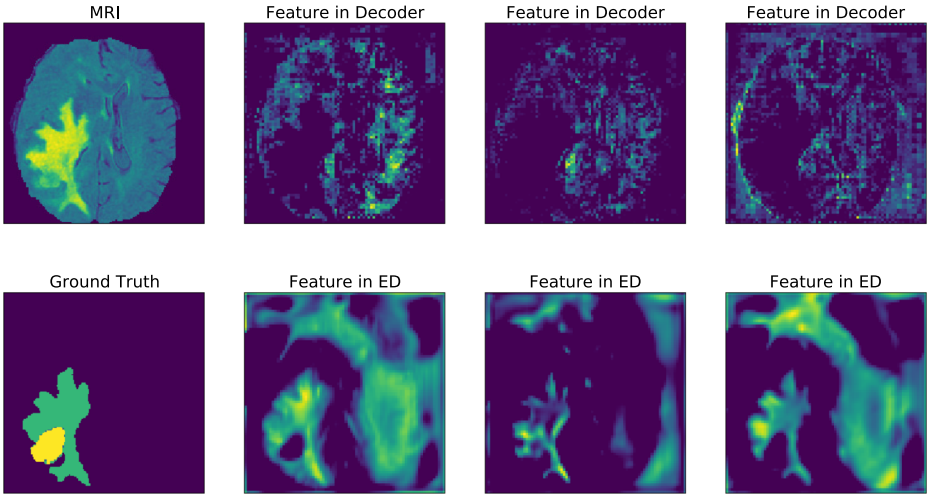

Fig. 9 The comparison of the intermediate features of whether we use enhancing decoder or use common decoder

to promote the propagation of these deep features. We solve the problem of unbalanced classification by increasing the importance of deep features.

It can be seen that in Fig. 8, features without PA were randomly extracted from $D(x_3)$ which is illustrated in Fig. 3. Features with PA were randomly extracted from $A(x_3)$ which is illustrated in Fig. 4. The boundary contour of feature maps without path aggregation encoder is fuzzy. After adding the path aggregation encoder, the extracted features are closer to the ground truth. The concatenate layers in the structure without path aggregation encoder can connect these features in the middle of the network. However, these features cannot be directly connected with the final layer. Path aggregation encoder can facilitate the dissemination of these features and shorten their propagation distance.

In the traditional U-Net, the encoder and decoder have the same number of channels. Common decoders cannot accommodate features from the same layer of the encoder and features from the previous layer of the decoder. In Fig. 9, features in ED were randomly extracted from $D(x_3)$ which is illustrated in Fig. 3. Features in Decoder were randomly extracted from the location of $D(x_3)$ in the common decoder. The location of the features taken from the common decoder and the enhanced decoder are the same. The features in

Table 3 he size of BraTS2018 data is large after preprocessing, so an efficient feature pyramid structure is needed to reduce the size of features

BraTS2018 NVIDIA1018ti	PA+FP	PA+FP+ED	PA+EFP+ED
Max batch size	8	8	16
Total params	9,593,124	24,640,004	24,637,956
Trainable params	9,584,292	24,625,540	24,624,516
Non-trainable params	8,832	14,464	13,440

This table illustrates the differences of parameters and maximum batch size between the efficient feature pyramid and the feature pyramid

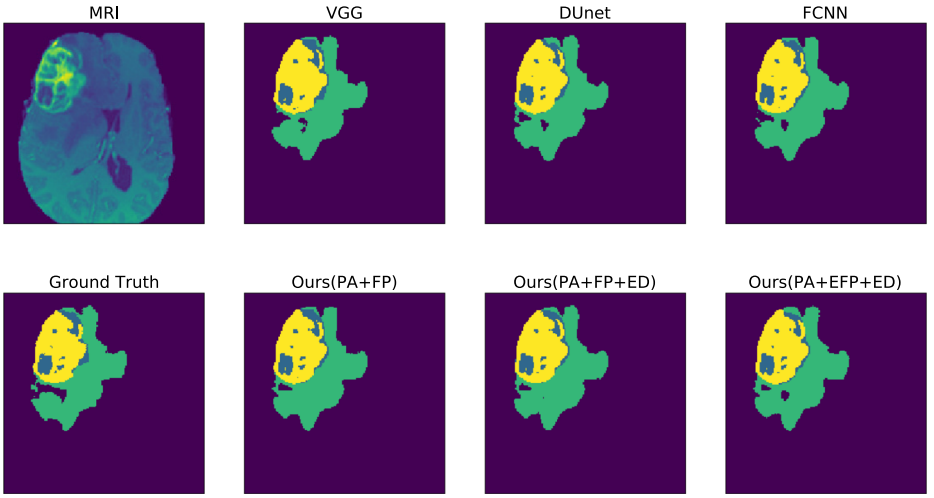

Fig. 10 Segmentation results on BraTS2017 dataset. PA+FP, PA+FP+ED, and PA+EFP+ED are our methods

(common) decoder are ambiguous and not smooth. The features in the enhanced decoder are closer to the ground truth and the information is richer.

E-Net [21] and ICNet [28] point out that efficiency can be improved in three ways: reducing the input resolution, reducing the downsampling, and compressing the network model. The size of BraTS2018 data is large after preprocessing, so an efficient feature pyramid structure is needed to reduce the size of features. Table 3 illustrates the differences of parameters and maximum batch size between the efficient feature pyramid and the feature pyramid. It can be seen that the batch size can be enlarged after using the efficient feature pyramid.

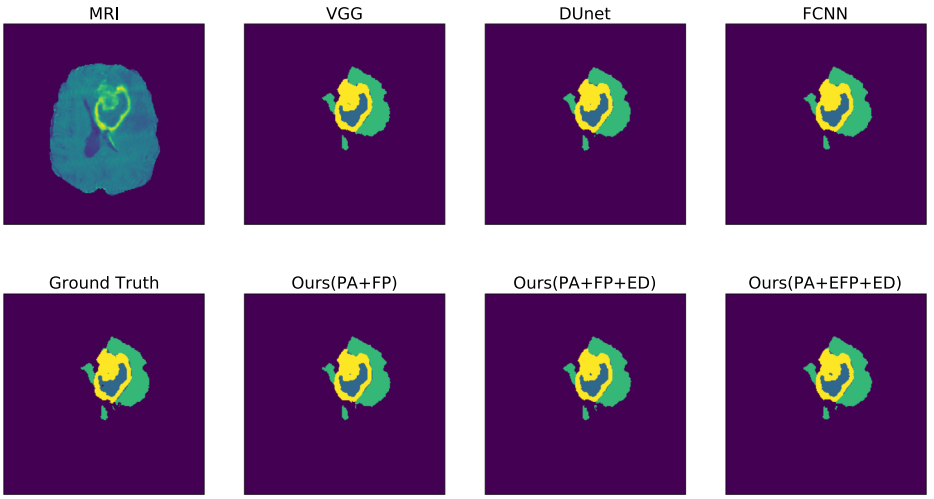

Fig. 11 Segmentation results on BraTS2018 dataset. PA+FP, PA+FP+ED, and PA+EFP+ED are our methods

In the segmentation result, the green area is whole tumor, the yellow area is tumor core, the blue area is enhancing tumor and the purple area is background. In Fig. 10, the enhancing tumor areas are marked in blue, which are unevenly distributed, and their areas are small. It can be seen that for the segmentation of the blue area, the segmentation result of our method is closer to the ground truth than that of the state-of-the-art methods. In Fig. 11, the green mask is the edema area. At the bottom of the entire image of the tumor, there is a tiny area of edema, like a scatter point. It can be seen that the state-of-the-art methods can hardly divide these scattered points. Our proposed method can segment these scattered points. The segmentation results are very close to the ground truth.

5 Conclusion

In this paper, we propose a path aggregation U-Net for brain tumor segmentation. The innovation lies in the following three aspects. Above all, path aggregation encoder facilitates the dissemination of deep information, shortening the distance between deep layers and the output layer in the network. Furthermore, an enhanced decoder is proposed to employ more channels corresponding to the accommodation requirements. Then, an efficient feature pyramid is proposed to use a small number of memory resources to connect multi-level features and output the segmentation results. In the future, we will enhance our framework for merging multiple relevant components and try to investigate the problem of weak supervision.

Acknowledgments The work is supported by the Fundamental Research Funds of Shandong University (Grant No. 2017JC013), the Shandong Province Key Innovation Project (Grant No. 2017CXGC1504, 2017CXGC1502) and the Natural Science Fundation of Shandong Province (Grant No. ZR2019MH049).

References

1. Adelson EH, Anderson CH, Bergen JR, Burt PJ, Ogden JM (1984) Pyramid methods in image processing. RCA engineer 29(6):33–41
2. Alex V, Safwan M, Krishnamurthi G (2017) Brain tumor segmentation from multi modal mr images using fully convolutional neural network. In: International MICCAI BraTS challenge, pp 1–8
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C (2017) Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Sci Data 4:170117
4. Bauer S, Wiest R, Nolte L-P, Reyes M (2013) A survey of mri-based medical image analysis for brain tumor studies. Phys Med Bio 58(13):R97
5. Fan J, Cao X, Yap PT, Birnet DS (2019) Brain image registration using dual-supervised fully convolutional networks. Med Image Anal 54:193–206
6. Ghiasi G, Fowlkes CC (2016) Laplacian pyramid reconstruction and refinement for semantic segmentation. In: European conference on computer vision, pp 519–534. Springer
7. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: Computer Vision (ICCV), 2017 IEEE international conference on computer vision, pp 2980–2988 IEEE
8. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
9. Huang G, Liu Z, Maaten LVD, Weinberger KQ (2017) Densely connected convolutional networks. In CVPR 1:3
10. Ioffe S, Szegedy C (2015) Batch normalization Accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167

11. Jégou S., Drozdzal M, Vazquez D, Romero A, Bengio Y (2017) The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: Computer vision and pattern recognition workshops (CVPRW), 2017 IEEE conference on, pp 1175–1183. IEEE
12. Kim G (2017) Brain tumor segmentation using deep u-net. In: International MICCAI BraTS challenge, pp 154–160
13. Kong X, Sun G, Wu Q, Liu J, Lin F. (2018) Hybrid pyramid u-net model for brain tumor segmentation. In: International conference on intelligent information processing, pp 346–355. Springer
14. Lin F, et al. (2019) FMNet: Feature Mining Networks for Brain Tumor Segmentation. In: 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), pp. 555–560. Portland, OR, USA,
15. Lin G, Milan A, Shen C, Reid ID (2017) Refinenet multi-path refinement networks for high-resolution semantic segmentation. In Cvpr 1:5
16. Lin Tsung-Yi, Dollár P., Girshick R, He K, Hariharan B, Belongie SJ (2017) Feature pyramid networks for object detection. In CVPR 1:4
17. Liu S, Qi LU, Qin H, Shi J, Jia J (2018) Path aggregation network for instance segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8759–8768
18. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3431–3440
19. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, et al. (2014) The multimodal brain tumor image segmentation benchmark brats. IEEE trans Med Imaging 34(10):1993–2024
20. Milletari F, Navab N, Ahmadi S (2016) V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3D Vision (3DV), 2016 fourth international conference on, pp 565–571. IEEE
21. Paszke A, Chaurasia A, Kim S, Culurciello E (2016) Enet: A deep neural network architecture for real-time semantic segmentation. arXiv:1606.02147
22. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional Networks for Biomedical Image Segmentation. In: International conference on medical image computing and computer-assisted intervention, pp 234–241. Springer
23. Shaikh M, Anand G, Acharya G, Amrutkar A, Alex V, Krishnamurthi G (2017) Brain tumor segmentation using dense fully convolutional neural network. In: International MICCAI brainlesion workshop, pp 309–319. Springer
24. Shen H, Zhang J, Zheng W (2017) Efficient symmetry-driven fully convolutional network for multimodal brain tumor segmentation. In: Image processing (ICIP), 2017 IEEE international conference on, pp 3864–3868. IEEE
25. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
26. Zhang X, Yao L, Wang X, Monaghan J, Mcalpine D (2019) A survey on deep learning based brain computer interface: Recent advances and new frontiers. arXiv:1905.04149
27. Zhang YU, Zhang H, Chen X, Liu M, Zhu X, Lee SW, Shen D (2019) Strength and similarity guided group-level brain functional network construction for mci diagnosis. Pattern Recogn 88:421–430
28. Zhao H, Qi X, Shen X, shi J, jia j (2017) Icnet for real-time semantic segmentation on high-resolution images. arXiv:1704.08545
29. Zhou T, Thung KH, Zhu X, Shen D (2019) Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. Hum Brain Mapp 40(3):1001–1016
30. Zhou T, Fu H, Shen J, Chen G, Shao L Hi-Net: Hybrid-fusion Network for Multi-modal MR Image Synthesis, IEEE transactions on medical imaging, 2020.