

Improved MRI Brain Tumor Segmentation Method Based on UNet++

Ke Tong^{1,2}, Jurong Ding^{1,2,*}, Xia Li^{1,2}

¹ School of Automation and Information Engineering, Sichuan University of Science and Engineering, Zigong, China

² Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Zigong, China

* Corresponding author: Ju-Rong Ding (Email: jurongding@gmail.com)

Abstract: Brain tumors are one of the most common tumors with high mortality. Accurate detection and quantitative evaluation of brain tumors are necessary for patient treatment. A reliable and effective method for automatic segmentation of brain tumors by magnetic resonance imaging (MRI) plays a crucial role in patient treatment. In recent years, the UNet++ network structure has shown better performance in image segmentation. However, it is deficient in small-scale feature extraction and segmentation accuracy. In this study, the AC-ResUNet++ network was proposed to address the problems of UNet++ network mentioned above, which combined residual module, convolutional block attention module (CBAM) and Atrous spatial pyramidal pooling (ASPP). The residual module can solve the problem of gradient explosion in the deep network, CBAM can focus on local feature information and enhance the feature learning ability of the network for the target range, and ASPP module can generate different receptive fields and obtain multi-scale contextual feature information. The BraTS 2018 and BraTS 2019 datasets were used to test the proposed network. The Dice similarity coefficient of AC-ResUNet++ in whole tumor, tumor core and enhanced tumor are 0.8521, 0.8696 and 0.7868, respectively. Compared with UNet++ network, the Dice similarity coefficient increased by 1.22%, 5.08% and 2.03% on the baselines. The comprehensive results demonstrate that AC-ResUNet++ performs more effectively in improving the segmentation accuracy of brain tumor.

Keywords: Brain Tumor Segmentation, UNet++, Residual Module, CBAM, ASPP.

1. Introduction

Brain tumors are one of the most dangerous diseases endangering human life and health [1]. Brain tumors are caused by the abnormal proliferation of cells in the brain, which seriously damage the nervous system and nearby healthy brain tissues. Glioma is the most common brain tumor [2]. According to the severity of tumor cells, brain tumors are divided into malignant and non-malignant. The World Health Organization (WHO) classifies glioma into four grades according to the behavior of tumor cells. Grades 1 and 2 are low-grade gliomas (LGG), and grades 3 and 4 are high-grade gliomas (HGG)[3]. Therefore, accurate clinical detection and quantitative evaluation of brain tumors are necessary for the treatment of patient. However, accurate detection of tumors is a challenging task because they vary in size and location.

Medical imaging technology plays a crucial role in the diagnosis and evaluation of the treatment effect of brain tumors. Magnetic resonance imaging (MRI) is the most widely used mode, which has the characteristics of non-injury and high resolution, and can obtain different anatomical information by setting different parameters [4]. At present, the segmentation of brain tumors is mainly performed manually based on MRI. Manual segmentation is time-consuming and has a strong subjectivity, so that the reproducibility is poor [5]. Therefore, it is necessary to study the method that can be used to segment brain tumors accurately and automatically. The MICCAI has launched a Multi-modal Brain Tumor Segmentation Challenges (BraTS) since 2012 to promote study in this area. The challenges indicated that compared with the traditional segmentation methods, the segmentation method based on deep learning showed superiority [6].

2D UNet++ network [7] has good segmentation performance for lesion areas, which indirectly fuses multiple

different levels of features. As one of the important improved networks of UNet, the UNet++ network introduces a built-in UNet collection with variable depth in the network. This kind of collection allows the decoder to perceive targets with different sizes under different receptive fields. The skip connection is redesigned for more flexible feature fusion. However, too many jump connections and nesting would lead to the loss of features and prevent the network from getting enough multi-scale information. In addition, as the network gets deeper, there is a problem of gradient disappearance in the process of down-sampling, which will lead to a reduction in the network's ability to extract brain tumor features. During the process of up-sampling, the extracted feature information is incomplete, and the edge information and location information are lost, which will affect the accuracy of the network segmentation on small-scale brain tumors.

In this study, the AC-ResUNet++ network for medical image segmentation is proposed to solve the problem of the UNet++ network. A residual block is added to each down-sampling of the network to solve the problem of network gradient disappearance. Convolutional block attention module (CBAM) is fused in each up-sampling to enhance the network extraction ability of local features and suppress redundant information. The inclusion of the Atrous spatial pyramidal pooling (ASPP) module between the encoder and decoder provides a different perceptual field for the feature map, preventing the loss of detailed information and promoting the network to capture more useful multi-scale information. Finally, the performance of AC-ResUNet++ model in this study is evaluated on the BraTS 2018 and BraTS 2019 public dataset.

2. Method

2.1. AC-ResUnet++

Unet++ is an improved network based on Unet, which is structured as a deeply supervised encoder-decoder network essentially [8]. The encoder and decoder subnetworks are connected through a series of nested dense jumps [9] to capture more feature layers. The Unet++ will continuously deepen the depth of the network during the down-sampling process, resulting in the phenomenon of gradient degradation. Therefore, a residual block is added to each down-sampling to address the encoder degradation problem and improve the network's capacity for deeper feature extraction and representation. Brain tumors are complex in shape and vary in size. During the up-sampling process, the spatial and location details of the feature map are easily lost, which makes network difficult to segment the location and size of small-scale brain tumors. This is an important factor that restricts the improvement of segmentation accuracy. Thus, CBAM is incorporated in each up-sampling to enhance the network's ability to extract and represent local features. The spatial and channel attention can make the network more focused on useful feature mapping, suppress redundant information, and solve the problem of small-scale tumors being ignored [10]. Brain tumors differ in location, size and shape, and are often surrounded by edema, which leads to insufficient information extraction during segmentation. Hence, ASPP module is added between the network encoder and decoder to generate different receptive fields, which makes the network obtain image multi-scale context feature information and solve the problem of image boundary detail information loss. The AC-ResUnet++ network architecture is shown in Fig. 1.

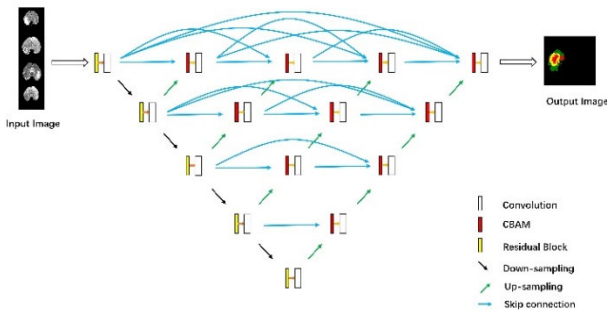


Figure 1. AC-ResUnet++ Network Architecture

2.2. Residual Module

The deeper the level of a convolutional network, the better the ability of the network to extract features from images. However, with the deepening of the network, the gradient will disappear when the accuracy reaches saturation. He et al. [11] proposed a residual network consisting of multiple residual blocks to solve this problem. Fig. 2. shows the schematic diagram of residual block.

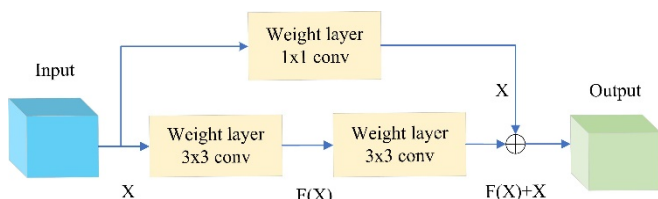


Figure 2. Residual Block

A neural network supposes X as an input and the desired underlying mapping is represented as $H(X)$. The formula is as follows:

$$F(X) = H(X) - X \quad (1)$$

where, $F(X)$ is a residual unit. Due to the skip connection, X skips the intermediate layers and is transmitted to the deeper network, then the original function becomes $F(X)+X$. The input and output are superimposed, which propagates larger gradients to the initial layers and allows these layers to learn as quickly as the final layers, and then allows the training of deeper networks [12]. The network uses ReLu as the activation function to improve the linear fitting of the network.

The residual module enables the deep network to transmit information better and solves the problem of information loss caused by the disappearance of gradients in the network training process. This module is added to the up-sampling of the network to improve feature extraction and expression capabilities at deeper levels of the network.

2.3. Convolutional Block Attention Module

CBAM is a lightweight attention module that can be integrated into any convolutional neural network framework [13]. CBAM emphasizes the meaning features in both channel and space dimensions. Each convolutional block of a deep neural network can be adaptively mapped with intermediate features by the CBAM module. There are two attention modules in CBAM, named channel attention and spatial attention. Feature images are first input to passed through the channel attention module for Max and Avg pooling, the results of pooling are summed by the mlp layer respectively. Then the channel feature maps are generated by the sigmoid activation function, and then input into the spatial attention module. The above operations are repeated to generate the final feature images. After this process, the image loss caused by network jumps can be effectively reduced and the accuracy of feature maps can be improved [14].

Assuming an intermediate feature $F \in R^{C \times H \times W}$ as an input feature, then a channel attention $M_C \in R^{C \times 1 \times 1}$ and a spatial attention $M_S \in R^{1 \times H \times W}$ can be inferred sequentially. Where denotes element-based multiplication. The feature map after channel attention is given by the equation $F' = M_C(F) \otimes F$ and the refined feature is obtained by the equation $F'' = M_S(F') \otimes F'$. The CBAM is shown in Fig. 3.

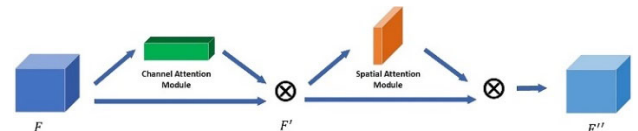


Figure 3. Convolutional Block Attention Module

2.4. Atrous Spatial Pyramid Pooling

ASPP is composed of multiple parallel atrous convolutions with different sampling rates and proportions [15]. It consists of convolution kernels with different receptive fields, captures contextual information, and accurately obtains multi-scale information. The ASPP module can solve the problem of the loss of image boundary details caused by the decrease of feature map resolution. It can also provide different receptive fields for image feature extraction to prevent the loss of too much information.

In this proposed network, the ASPP module acts as a bridge between encoder and decoder to capture multi-scale useful information in semantic segmentation tasks [16], which can effectively improve the accuracy of segmentation.

2.5. Combined Loss Function

The phenomenon of class imbalance is challenging for medical image segmentation [17]. Although the cross-entropy loss function is popular in conventional classification, it cannot obtain the same desirable dice score performance as the dice loss function. This study proposed a mixed loss function by combining Dice loss and Cross-entropy loss, and the mixed loss function can consider the weight information of the categories of brain tumors, and fully suppress the effect of category imbalance on brain tumor segmentation. The definition formula of the mixed loss function is as follows:

$$\varepsilon(Y, P) = -\frac{1}{N} \sum_{c=1}^C \sum_{n=1}^N (y_{n,c} \log p_{n,c} + \frac{2y_{n,c}P_{n,c}}{y_{n,c}^2 + P_{n,c}^2}) \quad (2)$$

Where $y_{n,c} \in Y$, Y and P are the retinal image truth value and prediction result, respectively, C and N represent the number of classes and pixels of the dataset in batch processing, respectively.

3. Experiments And Analysis

3.1. Datasets

The experimental data sets are the BraTS2018 and BraTS2019 public datasets provided by BraTS [18, 19]. The BraTS2018 dataset contains a training dataset of 210 HGG cases and 75 LGG cases. The BraTS2019 dataset adds 49 HGG cases and 1 LGG case to the BraTS2018 dataset. Each patient has four MRI modalities (Flair, T1, T1ce, and T2), and the size of each MRI image in the dataset is $240 \times 240 \times 155$. Images with different modes show different parts of the tumor. Flair image clearly shows the whole tumor, but cannot distinguish the necrotic components. T1 image shows the basic brain tumor imaging. T1ce image can distinguish necrotic components, but cannot identify edema around tumors from normal tissue. T2 image are sensitive to edema around the tumor.

Brain tumor labels are divided into three classes: necrotic and non-enhancing tumor core (NET, label1), peritumoral edema (ED, label2), and enhanced tumor (ET, label4). Ground truth (GT) is manual segmentation of brain tumors by experienced experts. To better evaluate the segmentation effect, it is necessary to segment whole tumor (WT, NET + ED + ET), tumor core (TC, NET + ET), and enhanced tumor (ET) (see Table 1). Fig. 4 illuminates a typical case of MRI brain image and GT.

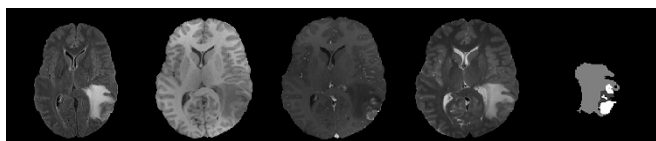


Figure 4. Example of the brain MRI data from a patient in the BraTS2018 dataset. From left to right: Flair modality, T1 modality, T1ce modality, T2 modality and Ground Truth.

3.2. Datasets

First, the contrast of MRI images of different modalities is different, and it is easy to produce overfitting in the training process. In this study, the Z-score normalization method is

used to normalize the mean value of the four modalities of patient data. Secondly, the size of brain tumor images in the BraTS dataset is $155 \times 240 \times 240$, there are a large number of background regions that do not contain brain tumor information, and they account for a large proportion in the overall image, which will affect the extraction of brain tumor feature information. Therefore, the image slices are cropped to filter useless information, and finally 155 image slices with a size of 160×160 are obtained. Finally, data enhancement was performed on the image slices, and the slices that did not contain brain tumors were removed to solve the problem of class imbalance. After preprocessing, the four modal outputs are merged and saved as npy format.

We processed the BraTS2018 dataset (210 HGG cases and 75 LGG cases) to generate a training set, and the newly added BraTS2019 dataset (49 HGG cases and 1 LGG case) to generate a test set.

3.3. Experimental Settings

The PyTorch deep learning framework was employed to conduct the experiments. Software environment: Win10 and Python 3.6 and CUDA10.0, CUDNN7.6.5. Hardware environment: CPU I5-7 generation (4 cores), 16G memory, GPU 12GB NVIDIA GTX1080Ti. The training set consisting of the BraTS2018 dataset was randomly divided, with 75% as the training set and 25% as the verification set. Epoch was 10000, weight_decay was 0.0001, early stopping was 10, and training was performed using the Adam optimizer [20].

All of the models were trained by using the same datasets, the same hardware environment, and their optimal parameter settings.

3.4. Evaluation Metric

Dice similarity coefficient (DSC) and Hausdorff distance (HD) coefficient are used to assess the proposed approach, which are the two most commonly used evaluation metrics for brain tumor segmentation [21]. DSC is used to evaluate the similarity between segmentation prediction and ground truth. The calculation formulas of the evaluation metrics are as follows:

$$DSC = \frac{2TP}{FN + FP + 2TP} \quad (3)$$

where TP represents the area correctly detected as a positive sample, FP represents the area incorrectly detected as a positive sample, and FN represents the area incorrectly detected as a negative sample.

Besides, HD represents the maximum mismatch between the tumor segmentation results and the labels, and the smaller the HD, the higher the accuracy. The calculation formulas of the evaluation metrics are as follows:

$$HD = \max \left\{ \max_{x \in X} \min_{y \in Y} \|x - y\|, \max_{y \in Y} \min_{x \in X} \|y - x\| \right\} \quad (4)$$

Where x represents the true result image and y represents the predicted result image.

3.5. Experimental Results

In order to evaluate the effectiveness of the proposed methods objectively, four network models, named Unet++, ResUnet++, CBAM-Unet++ and AC-ResUnet++, are tested in the present study. The newly added data set (49 cases of high-grade gliomas and 1 case of low-grade gliomas) of BraTS 2019 is used as the test set. The DSC scores of WT, TC

and ET regions of brain tumors were 0.8418, 0.8276, 0.7712 for Unet++, 0.8385, 0.8390, 0.7712 for ResUnet++, 0.8508, 0.8153, 0.7822 for CBAM-Unet++, and 0.8521, 0.8696, 0.7868 for AC-ResUnet++, respectively. The experimental results are shown in Table 1 and Fig. 5. Compared with the Unet++ network, the DSC of the AC-ResUnet++ network proposed in this paper in the three different subregions of WT, TC, and ET increased by 1.22%, 5.08%, and 2.03%, respectively. Our results show that the model has a good segmenting performance for all regions, especially for the tumor core area, which gets the highest score. In addition to

the DSC, this study also took Hausdorff coefficient to evaluate the segmentation. Compared with the Unet++ network, the Hausdorff coefficients of the AC-ResUnet++ network in the WT, TC and ET increased by 1.41%, 9.77%, and 1.77%, respectively.

In order to intuitively view the segmentation results and observe the spatial location of the brain tumor, the segmentation results from one patient are visualized (Fig. 6). The loss rates of the training and test sets of the network are plotted in Fig. 7.

Table 1. Evaluation results of four segmented models on the BraTS 2019 dataset.

Methods	DSC			HD		
	WT	TC	ET	WT	TC	ET
Unet++	0.8418	0.8276	0.7712	2.6222	1.7309	2.8039
ResUnet++	0.8404	0.8414	0.7693	2.6216	1.7330	2.8288
CBAM-Unet++	0.8508	0.8153	0.7822	2.6114	1.7345	2.7755
Our	0.8521	0.8696	0.7868	2.5854	1.5618	2.7544

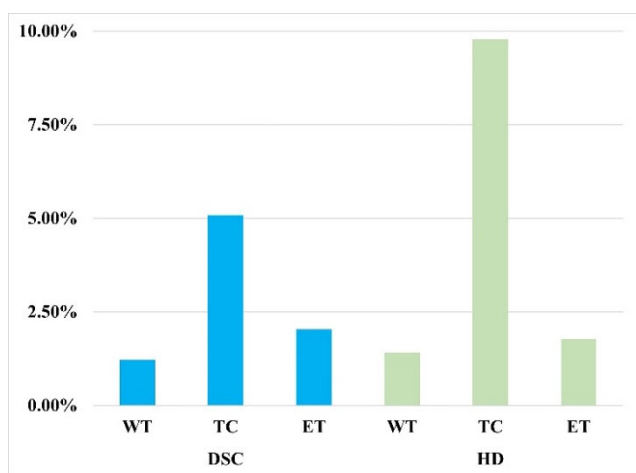


Figure 5. The percentages of performance improvement of AC-ResUnet++ in DSC and HD evaluation index compared with Unet++

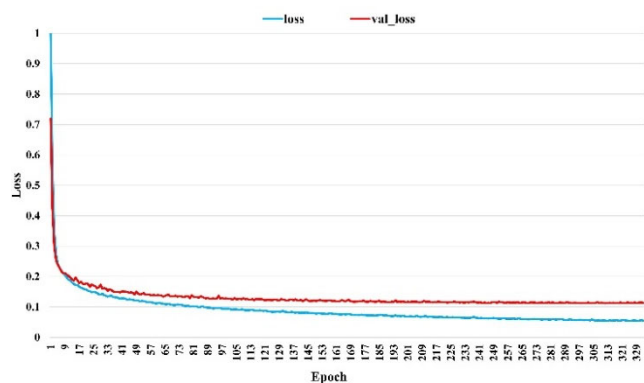


Figure 7. Loss graph of the training and validation sets of AC-ResUnet++. Val_loss is represented by a red curve that represents the loss change of the validation set during model training, and loss is represented by a blue curve that represents the loss change of the training set during model training.

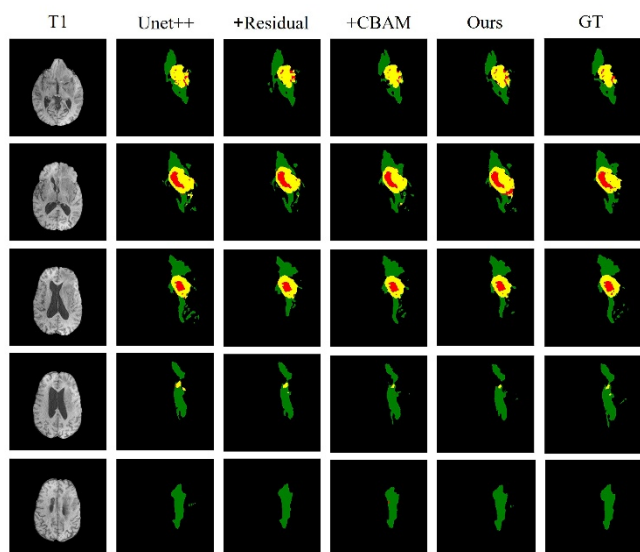


Figure 6. The brain tumor segmentation results in BraTS2019. Red, non-enhanced and necrotic tumor areas. Yellow, enhanced tumor areas. Green, edema areas

4. Conclusion

In this study, we proposed an improved segmentation model based on Unet++, AC-ResUnet++ network, which combines residual module, CBAM and ASPP module. In the proposed network, CBAM and residual module were added to the network skip connections, enhancing the ability of local feature expression and small feature extraction. ASPP module was added between encoder and decoder to extract effective multi-scale information. The BraTS 2018 and BraTS 2019 datasets were used to evaluate the performance of the proposed network. Compared with Unet++, AC-ResUnet++ improves the DSC score of WT, TC and ET by 1.22%, 5.08% and 2.03%, respectively. The results confirm the effectiveness of our proposed segmentation model.

References

- [1] MAGADZA T, VIRIRI S. Deep Learning for Brain Tumor Segmentation: A Survey of State-of-the-Art [J]. Journal of imaging, 2021, 7(2).W.-K. Chen, Linear Networks and Systems (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
- [2] MOHAN G, SUBASHINI M M. MRI based medical image analysis: Survey on brain tumor grade classification [J]. Biomedical Signal Processing and Control, 2018, 39: 139-61.

- [3] TANG J, LI T, SHU H, et al. Variational-Autoencoder Regularized 3D MultiResUNet for the BraTS 2020 Brain Tumor Segmentation; proceedings of the Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Cham, F 2021//, 2021 [C]. Springer International Publishing.
- [4] COLMAN J, ZHANG L, DUAN W, et al. DR-Unet104 for Multimodal MRI Brain Tumor Segmentation; proceedings of the Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Cham, F 2021//, 2021 [C]. Springer International Publishing.
- [5] RIBALTA LORENZO P, NALEPA J, BOBEK-BILLEWICZ B, et al. Segmenting brain tumors from FLAIR MRI using fully convolutional neural networks [J]. *Computer methods and programs in biomedicine*, 2019, 176: 135-48.
- [6] GAN X, WANG L, CHEN Q, et al. GAU-Net: U-Net Based on Global Attention Mechanism for brain tumor segmentation [J]. *Journal of Physics: Conference Series*, 2021, 1861(1): 012041.
- [7] ZHOU Z, RAHMAN SIDDIQUEE M M, TAJBAKHS N, et al. UNet++: A Nested U-Net Architecture for Medical Image Segmentation; proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Cham, F 2018//, 2018 [C]. Springer International Publishing.
- [8] CHEN L-C, ZHU Y, PAPANDREOU G, et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation; proceedings of the Computer Vision – ECCV 2018, Cham, F 2018//, 2018 [C]. Springer International Publishing.
- [9] HOORALI F, KHOSRAVI H, MORADI B. Automatic microscopic diagnosis of diseases using an improved UNet++ architecture [J]. *Tissue & cell*, 2022, 76: 101816.
- [10] ZHAO Z, CHEN K, YAMANE S. CBAM-Unet++:easier to find the target with the attention module "CBAM"; proceedings of the 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE), F 12-15 Oct. 2021, 2021 [C].
- [11] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition [J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-8.
- [12] WANG J, GAO J, REN J, et al. DFP-ResUNet:Convolutional Neural Network with a Dilated Convolutional Feature Pyramid for Multimodal Brain Tumor Segmentation [J]. *Computer methods and programs in biomedicine*, 2021, 208: 106208.
- [13] WOO S, PARK J, LEE J-Y, et al. CBAM: Convolutional Block Attention Module; proceedings of the 2018 European Conference on Computer Vision (ECCV) Cham, F 2018//, 2018 [C]. Springer International Publishing.
- [14] GUAN X, YANG G, YE J, et al. 3D AGSE-VNet: an automatic brain tumor MRI data segmentation framework [J]. *BMC medical imaging*, 2022, 22(1): 6.
- [15] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834-48.
- [16] JHA D, SMEDSRUD P H, RIEGLER M A, et al. ResUNet++: An Advanced Architecture for Medical Image Segmentation [J]. *IEEE International Symposium on Multimedia (ISM)*, 2019: 225-2255.
- [17] ZHAO Y, LIU Y, HUANG Z, et al. GCAUNet: A group cross-channel attention residual UNet for slice based brain tumor segmentation [J]. *Biomedical Signal Processing and Control*, 2021, 70: 102958.
- [18] MENZE B H, JAKAB A, BAUER S, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) [J]. *IEEE transactions on medical imaging*, 2015, 34(10): 1993-2024.
- [19] BAKAS S, AKBARI H, SOTIRAS A, et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features [J]. *Scientific Data*, 2017, 4(1): 170117.
- [20] KINGMA D, BA J J C S. Adam: A Method for Stochastic Optimization [J]. *International Conference on Learning Representations*, 2014.
- [21] SUDRE C H, LI W, VERCAUTEREN T, et al. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations; proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Cham, F 2017//, 2017 [C]. Springer International Publishing.