

Brain Tumor Image Segmentation Method Based on Multi-scale and Attention

Bowen Wang*

Department of Computer Science and Engineering, Northwest Normal University, Lanzhou, Gansu, China

Abstract. Brain tumor, as a high-risk disease of the brain, has been a threat to human life and health. In order to help doctors diagnose some parts of brain tumor accurately in hospitals, multi-scale fusion brain tumor image segmentation network has shown strong feature extraction ability and image segmentation accuracy improvement. In the original Unet network, only the feature information of the current layer is used in the jump connection layer, and the relevant feature information of the shallow network is ignored, so the segmentation accuracy will be affected accordingly. We use an improved segmentation network to solve this problem. Firstly, the multi-scale feature fusion module MFF is added to the encoder to fuse the features of different scales to improve the segmentation ability of the network. Secondly, the attention module ResCBAM is added to the jump connection layer of the encoder and decoder to guide the encoder to adaptively learn the important feature information in the jump connection. The BraTS2020 dataset in MICCAI competition was used for ablation experiments and contrast experiments, and Dice coefficient and HD95 were used as evaluation indicators. Through the experimental results, it can be seen that the improved network can extract more features in the whole tumor, tumor core and enhanced tumor region, and the segmentation effect of brain tumors is good. At the same time, the model parameters and the number of iterations are reduced.

1. Introduction

Brain tumors, also known as intracranial tumors, refer to abnormally developed cell populations that grow in the inner regions of the brain¹. If malignant tumors are not found and treated in time, they are likely to lead to human death². Since the brain is not open, with the increasing size of the tumor, it will continuously compress other brain tissues in the brain, causing a gradual increase in blood pressure in the brain, which will have a certain adverse effect on the central nervous system and cause irreversible damage to the patient's health³. An article published by the journal CA in 2021 showed that the incidence of malignant tumors in people under the age of 18 is increasing at a rate of 0.5% to 0.7% per year⁴, and an article in Neurology also showed that globally, the incidence and mortality of brain tumors are the highest. Glioma is composed of four regions, and different regions will have different effects on patients⁵. Because the structure of the brain is complex, and the size, shape and location of brain tumors are highly complex and prone to errors, so the diagnosis and identification in clinical practice are still very difficult. Therefore, imaging can provide doctors with different information about different parts of brain tumors and determine the types and malignant degrees of diseases as early as possible, which is of great significance for the treatment and recovery of patients⁶.

In medical imaging, Magnetic resonance imaging (MRI) is more prominent in imaging effect than computed Tomography (CT) and X-ray, which can provide good contrast. At the same time, it does not cause a large burden on patients, so it is a commonly used detection method for brain tumors. At the same time, MRI can observe coronal, sagittal, transverse and other directions, and will not produce artifacts that may occur in CT scanning⁷. In the process of MRI imaging, multiple different sequences can be obtained according to the influence of different images. The information contained in the four sequences is different. FLAIR sequence can better observe the boundary of the tumor, T1 sequence can observe the anatomical map of the fault, and T1c sequence can further observe the internal situation of the tumor. The lesion tissue could be observed by T2 sequence. Although different sequences of MRI images can contain different information about the tumor, which provides a great help to the accuracy of doctors' judgment, missed diagnosis and misdiagnosis often occur due to the variability of brain images and the influence of doctors' personal subjective factors. In order to make full use of medical image information, deep learning network models can be used to segment the key regions in brain tumor images, which is of great significance not only for the orientation analysis of brain tumors, but also for the evaluation of the feasibility of surgery and the prediction of the life span of patients.

* Corresponding author: 13149618035@163.com

Brain tumor segmentation is a branch of medical segmentation. Due to the multi-modal characteristics of brain tumor MRI images, it is generally necessary to comprehensively process four MRI sequences to obtain better segmentation results. Therefore, improving the accuracy of brain tumor image segmentation model in medical image segmentation is one of the important research directions of medical image segmentation.

The main contributions of this paper are the following three parts: 1) By introducing the MFF module, the feature information fusion of different scales in the network is improved, the segmentation ability is improved, and the Dice index is improved. 2) The residual attention module ResCBAM was added to make the network pay more attention to the detail features in image segmentation. The experimental results of BraTS2020 dataset verify that the network model proposed in this paper achieves good segmentation effect in the comparison of different models of the same image.

2. Related work

2.1. Brain tumor image segmentation task

The BraTS2020 dataset was used for the experiments in this paper, which has the following types: Edema (ED), Necrotic (NCR), Non-enhancing Tumor (NET) and enhancing Tumor (ET). Among them, NCR and NET are fused to the same label and are no longer distinguished. In addition, according to the competition criteria, segmentation performance was evaluated according to three officially defined tumor subregions, namely: Whole Tumor/Complete Tumor (WT/CT), the region containing all tumors, Tumor Coire (TC), the region containing all tumors but without edema, and enhanced tumor (ET), the region containing the category of enhanced tumors.

2.2. Unet and improvement

In 2015, Ronneberger proposed Unet network on the basis of FCN. Unet network is a model of coding-decoding structure, which has good information global ability and can realize relatively fine segmentation. A schematic representation of the structure of the Unet network is shown in Figure 1.

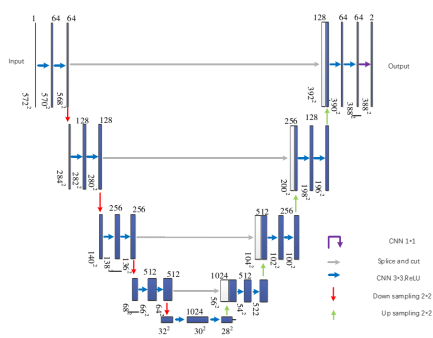


Fig. 1. Schematic diagram of the Unet network structure

As can be seen from Figure 1, the Unet network has five layers, including convolutional layer, pooling layer, deconvolution layer and ReLU activation function. The left part of the Unet network in the figure is the coding stage. Four groups of convolution operations are the core of the coding operation, and each group contains two convolution layers with the same number and size of convolution kernels. The pooling layer is used to reduce the dimension of all feature maps and reduce the loss of computer resources. The right part of the Unet network was the decoding stage. While feature extraction was carried out by four groups of convolutions, the confounding effect caused by upsampling was eliminated, and the spatial dimension was raised between adjacent decoding layers by deconvolution. In the middle jump connection part, because the convolution operation will change the size of the feature map, so that the input size of the decoding part of the same order is different from the output of the coding part, so the feature map of the coding part needs to be trimmed, and then the feature map of the same layer is combined in a cascade form to obtain the fused feature, which is used as the input of the next convolution layer. The above operations can make the features better fused and improve the segmentation ability of the network.

2.3. Multi-scale feature segmentation network

The multi-scale feature segmentation network enhances the perception ability of the model to various scale structures in medical images by fusing the feature information of different scales.

Such networks are typically designed with multiple parallel paths or branches, each responsible for capturing features at a particular scale. Then, these features are fused in some way (such as concatenation, weighted sum, etc.) to generate more comprehensive and accurate segmentation results.

In recent years, researchers have proposed a variety of multi-scale feature segmentation network architectures, such as U-Net++ and Attention U-Net. These networks perform well in several medical image segmentation tasks, such as lesion region detection, organ segmentation, and so on. They not only improve the segmentation accuracy, but also enhance the robustness of the model to noise and artifacts.

2.4. Multi-scale feature segmentation network

Attention mechanism is gradually becoming a key technology in image segmentation processing. Traditional medical image segmentation methods often rely on complex preprocessing and post-processing steps, as well as extensive manual feature engineering. With the continuous development of deep learning, especially the introduction of attention mechanism, the performance and accuracy of medical image segmentation have been significantly improved.

The application of attention mechanism in medical image segmentation is mainly reflected in two aspects: feature attention and spatial attention. Feature attention

allows the model to automatically select the most important features in the task, and spatial subject is for the key regions in the image.

Combining these two types of attention in the model enables accurate identification of diseased areas or structures in complex medical images.

3 Brain tumor image segmentation network based on multi-scale fusion

The main task is to add two modules to the basic Unet network. The first part is to add a multi-scale feature fusion module to the encoder to fuse the features of each layer to obtain more detailed features. In the second part, an improved residual convolution attention module ResCBAM is added to improve the conventional

attention module while reducing the number of parameters and improving the segmentation accuracy. The specific structure is shown in Figure 2.

3.1. Encoder module

Firstly, the multimodal brain tumor MRI images were input, and the features were encoded using convolution and 4 down-sampling operations in the encoder. Features at each level are used with two batch normalization, mish activation functions, and convolution operations to obtain feature information in different dimensions. Each downsampling uses a maximum pooling layer to reduce the size of the feature map, so that the network pays more attention to the feature information of the brain tumor image.

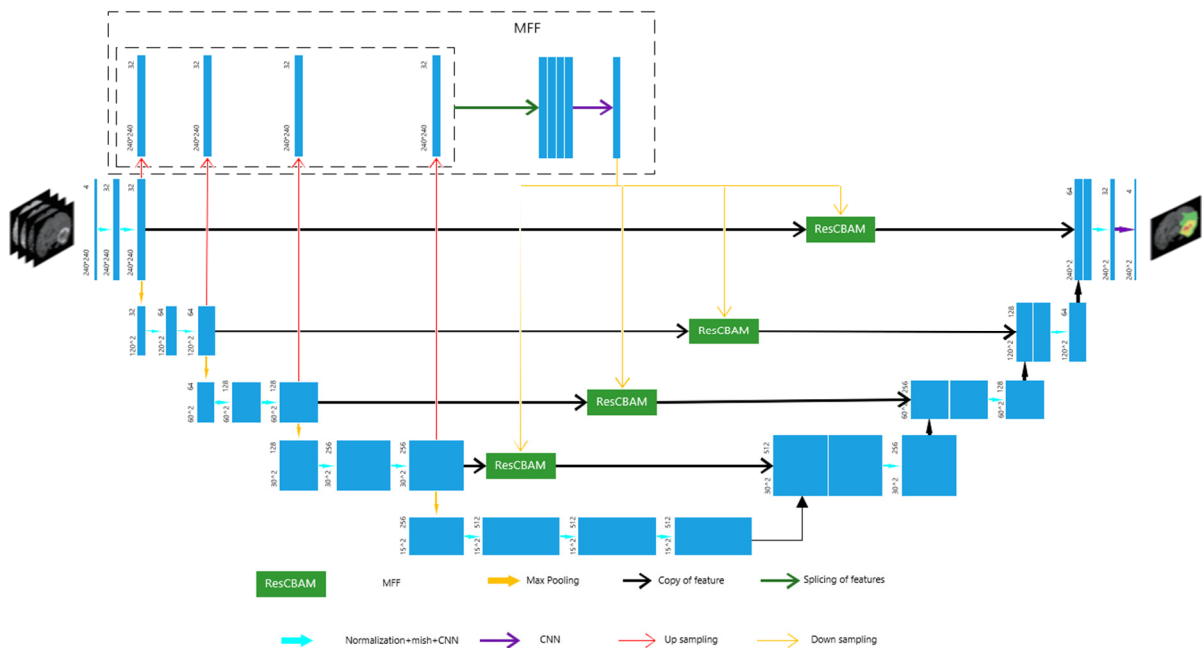


Fig. 2. Multi-scale feature fusion network

3.2. Multi-scale feature fusion module MFF

Due to a series of convolutional and downsampling operations in the encoder, the receptive field of the feature map gradually becomes larger, and the shallow network mainly focuses on the local features and texture information of the image, which helps to identify the edges, shapes, and small differences from other tissues of the tumor. Deep networks pay more attention to the global features and semantic information of images. In brain tumor segmentation, deep networks can learn the overall structure, location and relationship with surrounding tissues of tumors. Through multi-level convolutional and pooling operations, deep networks are able to progressively extract higher-level and more abstract feature representations, which are essential for accurate tumor segmentation. Therefore, in Unet, the features at different levels need to be upsampled into the encoder, the maximum resolution of the features is used

for feature splicing and fusion, and then downsampled into the attention module of each level. The Multi-scale Feature Fusion (MFF) module contains the feature information of shallow and deep networks.

3.3. Fusion features plus attention module ResCBAM

Unet uses jump connection to different levels of features. Although this practice can obtain more relevant information, low-level features have certain interference to the final segmentation. Therefore, inspired by the attention module, it is understood that the structure of AAM⁸ can capture global context information in low-order features while retaining high-order semantic information. However, CBAM⁹ focuses on features from two directions of channel and space, which can improve the feature expression ability of the network. As a lightweight module, it will not add additional burden to computer resources. Therefore, in order to better extract

features, on the basis of AAM and CBAM, this paper proposes an improved ResCBAM whose structure is

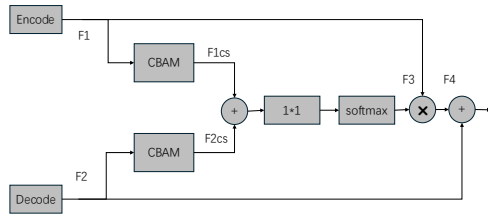


Fig. 3. Structure diagram of the improved ResCBAM

As can be seen from Figure 3, ResCBAM has two parts, encoding and decoding. The encoding part outputs the global context information, while the decoding stage outputs the semantic information, and each part is processed through the CBAM module. The attention module CBAM in the figure is shown in Figure 4.

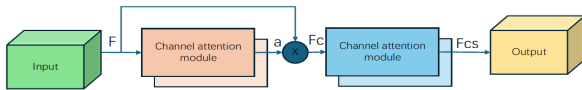


Fig. 4. Schematic representation of the CBAM structure

ResCBAM improves the feature extraction ability of the network by using the residual module to focus on spatial and channel attention.

4 Experiments and results analysis

4.1. Experimental data set

The BraTS2020 dataset is used for the experiments in this paper.

The data in the dataset are all MRI images of multimodal brain tumors. For MRI images, there will be different sequence slices containing FLAIR, T1 sequence, T1c sequence, and T2 sequence 3D images. The size of the image is 240*240*155. Before the experiment, 300 cases were used as the training data set, 100 cases as the validation data set, and 50 cases as the test data set. There are 232500 images in the training set, 77500 images in the validation set, and 38750 images in the test set. The image enhancement method used in the model training process was random inversion of images in the left and right and up and down directions, with a probability of 50%.

The images of four MRI sequences are shown in FIG. 5, a) FLAIR sequence, b) T1 sequence, c) T1c sequence, and d) T2 sequence.

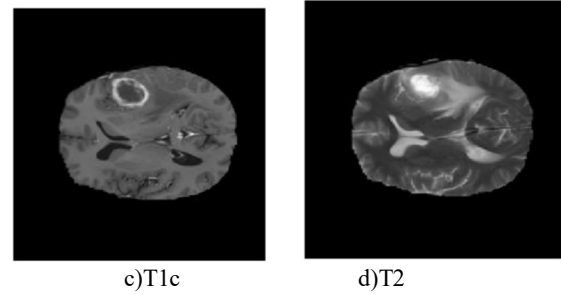
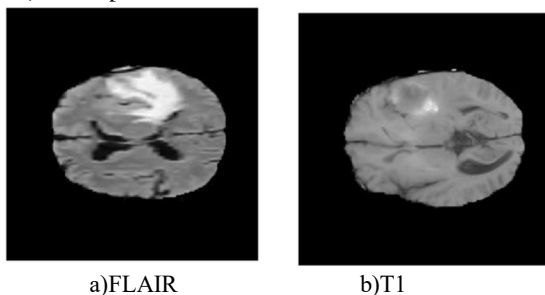


Fig. 5. Four image sequences of MRI

It can be seen from Figure 4-1 that different sequences of images present different states, so different sequences can be processed to obtain different information.

4.2. Indicators of evaluation

The results of model segmentation are compared with the real results, which are divided into four categories, as shown in Figure 6 below.

		Results of prediction	
		Obj	Back
GT	Obj	TP	FN
	Back	FP	TN

Fig. 6. Plot of the prediction versus the gold standard

It can be seen from Figure 6 that if the segmentation target is accurately segmented, it is called true positive; if the background is accurately segmented, it is called true negative; if the target is wrongly segmented, it is called false negative;

For the judgment of the segmentation effect of the model, Dice coefficient and HD95 (95% Hausdorff) were used as the evaluation indexes, which will be introduced in the following.

Dice coefficient is mainly used to evaluate the degree of coincidence between the real results and the segmentation results of the model. The closer the Dice coefficient is to 1, the better the segmentation accuracy is.

$$\text{Dice} = \frac{2TP}{FP + 2TP + FN} \quad (1)$$

Where: TP represents that the pixel is true positive and TN represents true negative. Similarly, FP represents pixels that are false positive and FN represents false negative.

Hausdorff distance (HD) is used to represent the maximum mismatch between two sets of samples. The closer the distance is, the smaller the HD value is, the better the segmentation effect is. The purpose of Hausdorff95 is to eliminate the influence of outliers on the calculation results, which is calculated as follows.

$$HD(P, T) = \max_{p \in P} (\min_{t \in T} \| p - t \|, \max_{t \in T} (\min_{p \in P} \| t - p \|)) \quad (2)$$

Where: P represents the set of labels of actual segmentation prediction results, and T represents the set of labels of real segmentation results.

Dice coefficient and HD95 were used as evaluation indexes in the brain tumor image segmentation model, and the performance of the model was evaluated from the overall similarity and boundary accuracy, respectively. The combined use of these two metrics can more comprehensively evaluate the performance of the model in the brain tumor image segmentation task, and provide valuable guidance for the optimization and improvement of the model.

4.3. Implementation details

The network model in this paper is tested on Ubuntu20.04 operating system using Python language and based on pytorch framework. The basic learning rate is 0.003, the batch number is set to 16, the iteration period is set to 800, and the early stopping method is used to prevent overfitting in the training process. The network performance was evaluated using Dice and HD95 in the evaluation metrics.

4.4. Function of loss

Binary Cross-Entropy (BCE) loss function can ensure the correct segmentation of background, and the BCE loss function is as follows.

$$loss_{BCE} = - \sum_{i=1}^n \sum_{j=1}^L g_{ij} \log p_{ij} + (1 - g_{ij}) \log(1 - p_{ij}) \quad (3)$$

Where: n is the set of pixels in the segmented image, L is the set of gold standard label pixels

4.5. Experimental results and analysis

In this section, the BraTS2020 dataset is used to verify the network segmentation effect in this paper. Subsequently, the effectiveness of the added modules was verified by ablation experiments. Finally, the superiority of the proposed network is verified by comparative experiments with other models.

4.5.1 Results of the experiment

The segmentation results in this paper are shown in Table 1, where the average Dice of ET is 0.7378, the average Dice of WT is 0.8852, and the average Dice of TC is 0.7378. It can be seen that the proposed network model has a good segmentation effect.

Table 1. Segmentation results of the proposed network

statistic	Dice			HD95		
	ET	WT	TC	ET	WT	TC
Mean	0.7235	0.8852	0.7378	40.91	15.3	16.37
StdDev	0.3137	0.0811	0.3189	86.24	22.13	46.33
Median	0.8447	0.9223	0.8881	2.43	4.22	4.23
25quantile	0.722	0.8736	0.5966	1.4	2.61	2.6
75quantile	0.899	0.9476	0.933	8.77	10.56	14.1

4.5.2 Ablation experiments

Ablation experiments were performed on the BraTS2020 validation dataset. Group A experiments were the basal Unet network. Group B experiments were used to verify the segmentation effect after adding the MFF module to the base network. Group C experiment was used to verify the segmentation effect after adding ResCBAM module to the base network. Group D experiment was used to verify the segmentation effect after adding MFF and ResCBAM modules simultaneously.

The results of ablation experiments are shown in Table 2 by adding modules to the proposed network module to the base Unet. Comparing the segmentation results of group A and group B, it was found that the segmentation accuracy of adding MFF module was improved. Among them, the average Dice of ET, WT and TC were increased by 2.35%, 2.08% and 0.99%, respectively. Group C represents the addition of ResCBAM on the basis of group B experiments. Comparing the segmentation results of group B and group C, it was found that the Dice coefficient of the model segmentation was improved after adding the attention module. Comparing the segmentation results of group A and group D, it was found that the average Dice of ET, WT and TC in group D were increased by 3.12%, 2.31% and 2.63%, respectively. Compared with the basic network, the average Dice of ET, WT and TC of the proposed network model was increased by 6.88%, 3.72% and 5.97%, respectively. These results demonstrate that the MFF and ResCBAM modules can improve the segmentation accuracy of brain tumor MRI images.

Table 2. Changes in model structure and comparison of evaluation indicators

Indicators Model	A	B	C	D
MFF		√		√
ResCBAM			√	√
WT Dice	0.824	0.832	0.831	0.841
TC Dice	0.783	0.820	0.816	0.843
ET Dice	0.765	0.758	0.760	0.817
WT HD95	2.773	2.713	2.496	2.556
TC HD95	1.672	1.770	1.657	1.523
ET HD95	2.898	2.886	2.887	2.890

4.5.3 Comparative experiment

The segmentation performance of the network proposed in this paper can be reflected by comparing different network models. The setting of the contrast experiment is to segment the same picture, and the segmentation performance of the model is judged from two aspects: qualitative analysis of human eye observation and quantitative analysis of evaluation indicators. TrUE-Net network proposed by Vaanathi et al.¹⁰, improved Transformer model TransBTS network and OM-Net network are used in the comparison model. These models are all good segmentation models for brain tumor segmentation at present, so the comparison experiment is more rigorous and accurate. The parameter Settings of the model used are the same as those in the literature, and the results are shown in Figure 7.

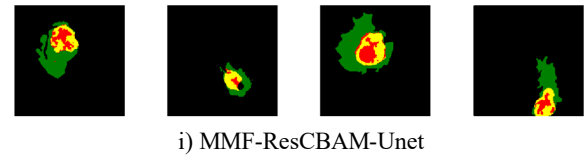
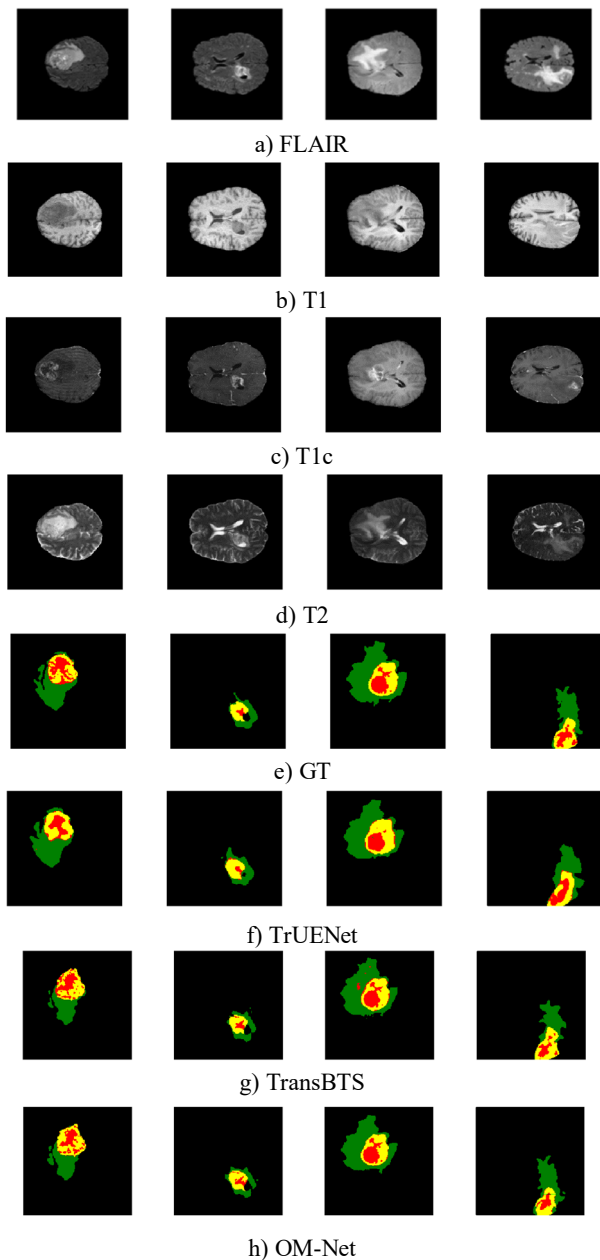


Fig. 7. Comparison of segmentation results of different models

Figure 7. The first four columns are the images of different modalities in the BraTS2020 dataset as well as the gold standard image, the output of which is a four-channel output, so the whole area of the tumor includes different small regions of the tumor in green, yellow, and red. Where f) is the segmentation result of TrUE-Net, g) is the segmentation result of TransBTS, h) is the segmentation result of OM-Net, and i) is the segmentation result of the proposed model MMF-ResCBAM-Unet.

By comparing the segmentation results of different models, the segmentation effect maps of different models are different. By comparing the segmentation results of OM-Net and the segmentation results of this paper, the segmentation effect and detail features are relatively similar, and the segmentation results can basically correspond to the gold standard, so the segmentation effect is relatively good. For TrUE-Net, it can be seen that the segmentation effect is poor, the edge information of the segmented image is poor, and the relevant feature information of the shallow network may be ignored. For TransBTS, some detailed features in brain tumors are not segmented more finely. Therefore, according to the qualitative judgment of human eyes, the network proposed in this paper has good segmentation ability and segmentation accuracy.

Since human eye observation is only a qualitative judgment, the model is quantitatively evaluated by Dice coefficient and HD95, and uploaded to the evaluation platform of BraTS2020 to obtain consistent evaluation indicators, which ensures the fairness and rigor of the experiment. The comparative experimental results of different model evaluation indexes are shown in Table 3.

Table 3. Comparison of evaluation indexes of different models

Indicators Modal	TrUE-Net	TransBTS	OM-Net	MMF-ResCBAM-Unet
WT Dice	0.86	0.86	0.90	0.90
TC Dice	0.84	0.85	0.87	0.89
ET Dice	0.80	0.81	0.81	0.82
WT HD	5.21	6.19	4.72	2.23
TC HD	6.33	13.18	7.16	2.31
ET HD	4.15	14.76	2.98	2.72

Table 3 shows that in the segmentation results of MMF-ResCBAM-Unet network, the Dice scores of WT, TC and ET were 0.90, 0.89 and 0.82, respectively. Compared with other methods, the average Dice index was increased by 0.021-0.030. The HD scores of WT, TC and ET were 2.54, 1.61 and 2.62, respectively. Compared with other methods, HD was reduced by 2.77-11.66. In conclusion, compared with other methods, the

proposed MFF-ResCBAM-Unet network has the best performance and the best segmentation effect.

5. Conclusion

For the lack of multi-scale feature information in the Unet network, this paper proposes a multi-scale feature fusion module MFF with complementary information. Then, in the jump connection layer, the key details of each scale feature are judged by the residual attention module ResCBAM, which makes the model pay more attention to the segmentation region. By adding these two modules, the performance of the network in the details of image segmentation is improved. At the same time, the introduction of the attention module reduces the number of network parameters and makes the training speed faster. Through the training, testing and verification on BraTS2020 dataset, as well as the results of ablation experiments and comparative experiments, it can be seen that the segmentation accuracy of the proposed network module has been significantly improved, and its performance is better than that of the current well-performing segmentation networks such as OM-Net, which can more accurately segment the relevant regions of brain tumors in MRI images. The more accurate the segmentation of brain tumor images, the more accurate it is to help doctors find the accurate location of the tumor, which will be of great help for later treatment and surgery. Therefore, the performance of network segmentation helps patients and doctors to solve the difficulty of early judgment and treatment to a certain extent.

Although the network model in this paper has achieved a significant segmentation effect, in the daily use of brain tumor models, a faster and less resource-consuming network model is needed, the speed of training is improved, and the parameters of the model are reduced to reduce computer resources. The encoder and decoder in the network structure in this paper still have relatively complex convolution modules. The size and speed of the model will also be affected accordingly; therefore, the convolution module will be improved in the future, and modules such as inverted residuals will be used to lightweight the model and speed up the model. The model can be trained and segmented in daily use. The lightweight model is more suitable for daily use.

References

1. Deangelis L M. Brain Tumors[J]. New England Journal of Medicine, 2001, **344**(02): 114-123.
2. Tan A C, Ashley D M, Giselle Y López, et al. Management of Glioblastoma: State of The Art and Future Directions[J]. CA A Cancer Journal for Clinicians, 2020, **70**(04): 299-312.
3. Zarnie L. From Survivorship to End-of-Life Discussions for Brain Tumor Patients [J]. Neuro-Oncology Practice, 2021, **8**(03): 231-232.
4. Gorp Marloes, Erp Loes M E, Maas Anne, et al. Increased Health-related Quality of Life Impairments of Male and Female Survivors of Childhood Cancer: Dccss Later 2 Psycho-oncology Study[J]. Cancer, 2021, **128**(05): 1074-1084.
5. Corso J J, Sharon E, Dube S, et al. Efficient Multilevel Brain Tumor Segmentation with Integrated Bayesian Model Classification[J]. IEEE Trans Med Imaging, 2008, **27**(05): 629-640.
6. Narayanan V. High Grade Glioma: Pathogenesis, Management and Prognosis[J]. Advances in Clinical Neuroscience and Rehabilitation, 2012, **12**(04): 24-29.
7. Wahab R A, H Albasha, Martin J, et al. Characterization of Common Breast Mri Abnormalities: Comparison Between Abbreviated and Full Mri Protocols[J]. Clinical Imaging, 2021, **79**(15): 125-132.
8. Ni ZL, Bian GB, Zhou XH, et al. RAUNet: Residual Attention U-Net for Semantic Segmentation of Cataract Surgical Instruments[J]. Lecture Notes in Computer Science, 2019, **11954**(04): 139-149
9. Trebing K, T Stańczyk, Mehrkanon S. SmaAt-UNet: Precipitation Nowcasting Using A Small Attention-Unet Architecture[J]. Pattern Recognition Letters, 2021, **145**(06): 178-186.
10. Sundaresan V, Griffanti L, Jenkinson M. Brain Tumour Segmentation Using A Triplanar Ensemble of U-Nets on Mr Images[J]. Lecture Notes in Computer Science. Springer, Cham, 2021, **12658** (02): 340-353.