EFU Net: Edge Information Fused 3D Unet for Brain Tumor Segmentation

Yu WANG, Hengyi TIAN, Minhua LIU**

School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China

wangyu@btbu.edu.cn, thy8562@sina.com, liuminhua@btbu.edu.cn

Submitted October 23, 2023 / Accepted May 28, 2024 / Online first June 7, 2024

Abstract. *Brain tumors refer to abnormal cell proliferation formed in brain tissue, which can cause neurological dysfunction and cognitive impairment, posing a serious threat to human health. Therefore, it becomes a very challenging work to full-automaticly segment brain tumors using computers because of the mutual infiltration and fuzzy boundary between the focus areas and the normal brain tissue. To address the above issues, a segmentation method which integrates edge features is proposed in this paper. The overall segmentation architecture follows the encoder decoder structure, extracting rich features from the encoder. The first two layers of features are input to the edge attention module, and to extract tumor edge features which are fully fused with the features of the decoder segment. At the same time, an adaptive weighted mixed loss function is introduced to train the network by adaptively adjusting the weights of different loss parts in the training process. Relevant experiments were carried out using the public brain tumor data set. The Dice mean values of the proposed segmentation model in the whole tumor area (WT), the core tumor area (TC), and the enhancing tumor area (ET) reach 91.10%, 87.16%, and 88.86%, respectively, and the mean values of Hausdorff distance are 3.92, 5.12, and 1.92 mm, respectively. The experimental results showed that the proposed method can significantly improve segmentation accuracy, especially the segmentation effect of the edge part.*

Keywords

Deep learning, brain tumor segmentation, encoder decoder structure, edge attention mechanism, hybrid loss function

1. Introduction

As a common brain disease, glioma has a high incidence, accounting for about 40% of all brain tumors and 78% of all malignant brain tumors, which seriously endangers human health [1]. Brain magnetic resonance imaging (MRI) is widely used in clinical practice because of its advantages such as good soft tissue contrast, multi-parameter imaging in arbitrary direction, non-injury, and non-bone artifacts. In order to better understand the patient's condition and develop a reasonable treatment plan, doctors need to accurately classify the size, shape and location of the tumor. However, manually segmentation on the location of brain tumors and different focal areas relying only on doctors is not only time-consuming and laborious, but also requires high diagnostic experience, and is in possession of great subjectivity. Automatic segmentation of brain tumors by computer technology can significantly improve the working efficiency of doctors, with advantages such as high efficiency and strong objectivity [2].

With the development of deep learning technologies, convolutional neural networks (CNNs) have gained good image segmentation performance and achieved good results in the field of medical image segmentation. Xu et al. [3] used transfer learning to pre-train network weights on ImageNet, then to fine-tune network weights with brain tumor images. Since the UNet network was proposed, Dong et al. [4] applied the UNet to the segmentation task of brain tumors. However, for three-dimensional MRI images of brain tumors, because only each slice is taken as input by the two-dimensional network, information between slices could not be extracted. To solve this problem, Beers et al. [5] proposed a three dimensional (3D) UNet network architecture, in which complete 3D MRI images could be directly inputted into the network. Based on the extraction of slice information in the image, the 3D convolution kernel is used. Interlayer features can also be extracted at a deeper level. Due to its good segmentation performance, 3D UNet has become the most widely used basic model in the field of brain tumor segmentation, and some advanced deep learning networks are improved based on this network [6–9].

In the brain of a patient, there is mutual infiltration and blurred boundary between the normal soft tissue and the lesion area [10]. Therefore, in the brain MRI, the difference between foreground and background pixels is small, and the edge is fuzzy [11]. However, the edge of a lesion is an important feature in medical images, and human doctors pay much attention to the edge information when reading MRI. Correctly segmented tumor edges can reflect the contours and morphological characteristics of

DOI: 10.13164/re.2024.0387

tumors, which is crucial for the subsequent formulation of treatment plans [12]. Therefore, accurate prediction of lesion edge by CNN network has always been the focus and difficulty of this research field. At the same time, brain tumors are usually small in size, approximately accounting for only 1.54% of the whole brain image, and the volume of different sub-regions within the tumor is even smaller, which makes it difficult for the model to learn enough features from fewer positive samples in the training stage, and reduces the optimization quality of the model.

To solve the above problems, researchers have made a lot of improvements based on 3D UNet. Zhang et al. [13] proposed the attention gate residual U-Net model (AGResU-Net), in which the attention gating mechanism is used to eliminate irrelevant and noisy features, and further to enhance the characteristic expression of tumor region. Aboelenein et al. [14] proposed a hybrid two track UNet (HTTU-Net), in which the first track focuses on the shape and size of tumors, while the second track captures contextual information. The segmentation accuracy of tumor margin was improved effectively. Zhang et al. [15] proposed the edge attention Network (ET-NET), which is used for two-dimensional (2D) image segmentation of retinal blood vessels. Shallow features are inputted into the edge supervision module, then the edge prediction results are transmitted to the weighted aggregation module, so that the edge information can be used more effectively. Lee et al. [16] proposed a boundary key point selection method to fit the target region with embedded expert knowledge. In this method, the key point graph generator is used to effectively constrain the selected key points on the structure boundary to obtain more accurate segmentation results. Zhou et al. [17] proposed a fusion loss function for the brain tumor segmentation task, which was composed of cross entropy loss and Dice loss. Cross-entropy losses are used to maintain stability during training, while Dice losses are used to reduce the effects of data imbalances. However, the current method has the following problems. First, the superficial features of the network are not all related to the edge, and the irrelevant features will reduce the ability of the network to recognize the tumor edge. Secondly, when fusing edge information with semantic segmentation backbone, in the most of the above methods fusion strategy is used only in full-resolution feature maps, so that edge information is not fully utilized in the tumor localization process. At the same time, in the current methods loss weighting sum is applied when learning multiple tasks, but the weight setting is manually set according to experience, and it is not effective enough to balance multiple losses.

Therefore, in this paper a new segmentation framework for brain tumors segmentation of 3D MRI images is proposed. Specifically, the ideas behind the development model are as follows. First, the shallow features extracted from the first two stages of 3D UNet network encoder are inputted into the designed edge attention module. The input feature map adjusts channel weights internally within the edge attention module, and features relevant to edges are filtered to generate more accurate edge prediction maps. Subsequently, the edge prediction map is fused with decoded segments at multiple levels, which, compared to the conventional fusion method typically performed only at the final level, enables a more effective integration of edge information. Finally, a fusion loss function is proposed, which is composed of cross entropy loss, Dice loss and edge loss. Meanwhile an adaptive weight determination method is proposed, and the weights of the three parts are adjusted adaptively with the training process using trainable parameters.

The rest of this paper is organized as follows. In Sec. 2, the proposed method is depicted in detail. Experimental implementation and results and are given in Sec. 3. Finally, Section 4 concludes the paper.

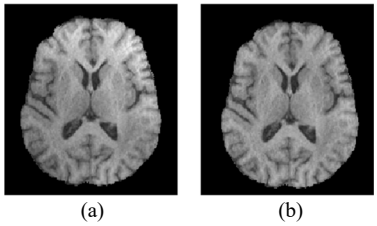
2. Method

2.1 Data Preprocessing

During the collection of medical images, due to the influence of imaging equipment, imaging principle and individual differences, the original images generally contain degradation phenomena such as uneven brightness and noise [18]. Therefore, before segmentation, the data provided by brain tumor segmentation (BraTS) need to be preprocessed, such as bias field correction and standardization, to reduce misdiagnosis and to improve diagnostic accuracy. In this paper, N4ITK bias field correction method [19] was used to remove image inhomogeneity. The examples of brain tumor images before and after processing are shown in Fig. 1. Since the original data came from different institutions and were obtained by different scanning instruments, the value range of each group of data would also be inconsistent. Z-Score standardization method should be used to unify the values of all data into a small range and make the values of images present a normal distribution, so as to facilitate the numerical calculation of the model in the training process. In this study, the four modal images of each patient were standardized separately, and the Z-Score standardization was shown by

$$\mathbf{X} = \frac{\mathbf{X} - \bar{\mathbf{X}}}{\mathbf{X}_{\text{std}}} \tag{1}$$

where $\mathbf{X}$ denotes a modal image of the patient, $\bar{\mathbf{X}}$ is the mean of all voxels of $\mathbf{X}$, and $\mathbf{X}_{\text{std}}$ is the standard deviation value of all voxels of $\mathbf{X}$.



(a) (b)

Fig. 1. Bias field correction results: (a) Original image, (b) corrected result.

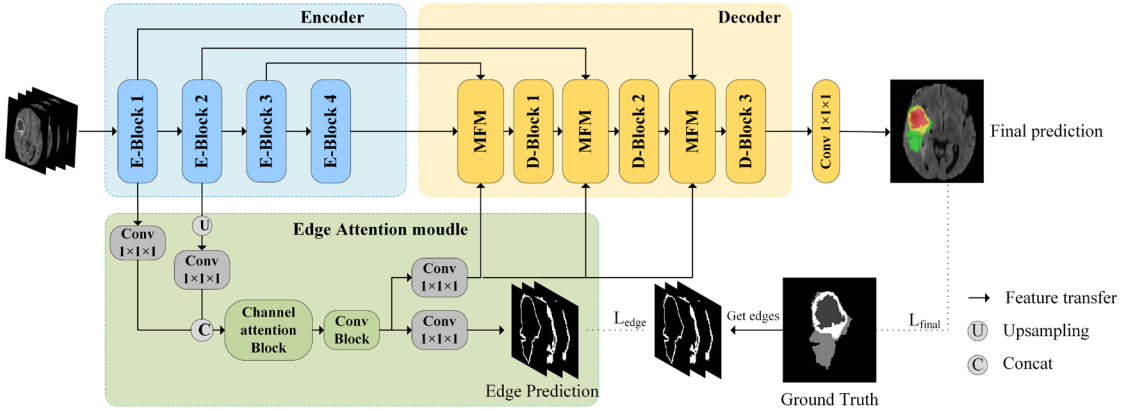


Fig. 2. Network architecture.

2.2 Network Structure Overview

Figure 2 shows the overview of the proposed brain tumor segmentation framework. In the encoder, a convolutional module with a residual structure [20] is used for feature extraction. The network consists of four coding blocks, each of which corresponds to a feature map of different resolution. The choice of a four-stage encoder structure is grounded in preliminary analysis and investigation of experimental results. By systematically experimental comparisons, we have found that the four-stage encoder structure yields optimal performance for the given task. Fewer stages may result in insufficient model capacity to capture complex features and patterns, while more stages could escalate computational costs, rendering the model cumbersome and impractical.

The shallow features extracted from the first two coding blocks are inputted into the edge attention module. The edge map of the brain tumor is predicted by the edge attention module, the shape prior of the segmentation target for the segmentation network is provided, and the poor performance of the network in boundary segmentation is improved. In the decoder, three decoding blocks are used to locate the tumor according to the extracted features. Before each decoding block, a multi-feature fusion module (MFM) is used to fuse the feature maps from the encoder, front layer and edge attention modules at the feature level. The final convolution layer unifies feature graphs with three convolution kernel size of 1, corresponding to the three segmented sub-regions.

2.3 Encoder Network

Using the encoder the low-level and high-level semantic features of the inputted image are extracted successively. The first three coding blocks are composed of residual convolution module and downsampling, and the fourth coding block has only one residual convolution module. The Swin Transformer consists of four stages, as shown in Fig. 2. The residual convolution module is shown in Fig. 3(b), which consists of two $3 \times 3 \times 3$ convolution layers with stride 1 and residual connection. The number of

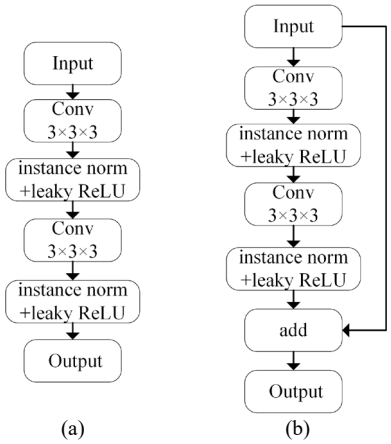


Fig. 3. The structure schematic of convolution module: (a) Convolution module, and (b) residual convolution module.

convolutional kernels in the first layer is 32. After every downsampling, the number of convolution kernels doubles. After convolution, instance normalization (IN) and leaky Relu activation function are adopted. The subsampling module is a $3 \times 3 \times 3$ convolution with stride 2. After 4 encoding blocks, the generated feature map size is 1/8 of the inputted image.

2.4 Decoder Network

The decoder consists of multi-feature fusion module and decoding block. The multi-feature fusion module is used to upsample the features from the previous layer, then to fuse the features from the coding block and edge attention module. Upsampling is a $2 \times 2 \times 2$ deconvolutional layer with stride 2. According to the extracted features, the decoding block relocates the positions of different semantic regions. The decoding block is composed of a convolution module, as shown in Fig. 3(a), which contains two $3 \times 3 \times 3$ convolution layers with step size 1. After convolution, case standardization and leaky Relu activation function are also adopted. The decoding block uses four upsampling operations to convert the low-resolution feature map extracted from the coding segment into a high-resolution feature map. A $1 \times 1 \times 1$ convolutional layer with stride 1

is added to the network as the segmentation layer, and the results are sent to the sigmoid layer to obtain the output probability graph of the model, and to achieve end-to-end segmentation.

2.5 Edge Attention Module

In the process of manually segmenting brain tumors and their sub-regions, tumor margins provide experts with extremely important information. In order to improve the deep learning network's attention to edge information, and further to improve the segmentation ability of tumor regions, in this paper the ways to supplement more brain tumor edge information to the network are studied in depth. In order to select the features with edge information, the outputted feature maps are visualized by the four encoding blocks, as shown in Fig. 4.

As can be seen from Fig. 4, edge information is considered to be a low-level shallow feature which is only captured at the shallow layer of the network. Therefore, the feature maps of the first two levels of the code segment are inputted into the edge attention module in order to better capture and to utilize these key features. The specific structure of the module is shown in Fig. 2. In this module, the inputted feature graph is first up-sampled to the same size, then the number of feature channels is adjusted after the convolution of two convolution cores with the size of 1, and the channel dimension is finally superimposed.

However, not all shallow feature maps are associated with edge information. Some irrelevant features may affect the edge attention module's prediction of the edge graph. Therefore, it is necessary to input the superimposed feature map into the channel attention module and to process it. In the channel attention module a structure similar to the squeeze and excitation (SE) attention module is used [21], which can judge the importance of each channel so as to obtain the attention weights of each channel in the feature map, enabling the network to ignore irrelevant features when

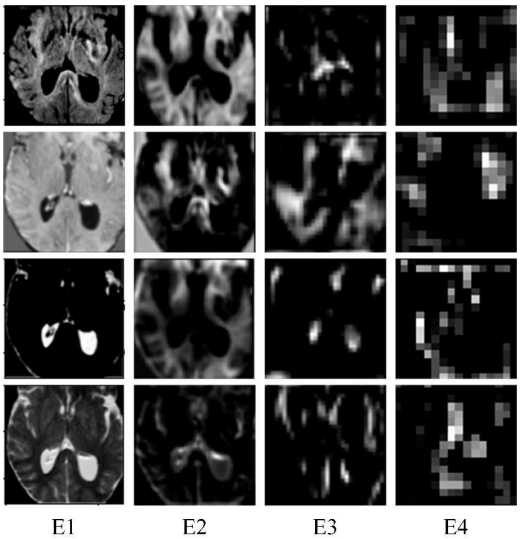


Fig. 4. The outputted feature maps of four encoders.

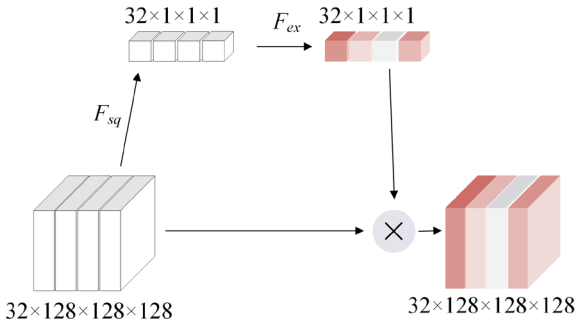


Fig. 5. The framework of channel attention module.

predicting the edge map. The specific module structure is shown in Fig. 5.

First, a global average pooling layer is used to compress the feature map, and its calculation process is shown by

$$F_{\mathrm{sq}} = \frac{1}{H \times W \times D} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{k=1}^{D} x_C(i,j,k) \tag{2}$$

where C, H, W and D represent the number of channels, height, width and depth of the feature map, respectively.

Then feature excitation is used to adaptively recalibrate the channel relationship, and to generate the channel dimension attention weight matrix. The weight calculation process of the C channel is described by

$$F_{\mathrm{ex}} = \sigma(W_1(\delta(W_0(input)))) \tag{3}$$

where σ and δ respectively represent sigmoid and ReLU activation functions. $W_0 \in \mathrm{R}^{C \times \frac{C}{r}}$ and $W_1 \in \mathrm{R}^{\frac{C}{r} \times C}$ respectively represent two fully connected layers, where r is the hyperparameter used for dimensionality reduction, which is set to 8 in this experiment.

Finally, the output weight matrix is multiplied by the channel dimension of the original feature graph, and the feature graph after the channel weight correction is obtained.

Subsequently, the processed feature maps are inputted into the convolution module. The feature map is transferred into two branches, the first of which takes it as the supervised prediction image and optimizes the network by calculating the loss value against the standard edge image. When getting the standard edge image, Sobel operator is used to extract the edge. The second branch fuses it with the feature map in the decoder to provide more edge information for the split network.

2.6 Multi-feature Fusion Module

In order to merge edge features with semantic features extracted from UNet network more effectively, a multi-feature fusion module is proposed in this paper. At each resolution level of the decoding path, multi-feature fusion module is used to fuse features from the front layer, jump

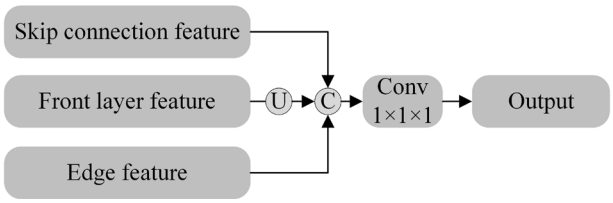


Fig. 6. The framework of multi-feature fusion module.

connections from the encoder, and edge attention module to improve the representation ability of the model. The specific structure is shown in Fig. 6. First, the features of the front layer are up-sampled so that they have the same scale as the other two features. Then the channel dimension is spliced. Next, in order to better fuse these features and to reduce the number of parameters, 3D convolution with convolution kernel size of 1 is used for feature fusion. Finally, the fused feature map is used as the output.

2.7 Adaptive Weighted Hybrid Loss Function

Binary cross entropy (BCE) loss is widely used in medical image segmentation tasks to measure the difference between the two distributions by calculating the cross entropy between the model prediction and the ground truth (GT) label for each pixel in the image. However, since the proportion of foreground region in medical images is much smaller than that of background region, a serious data imbalance phenomenon exist. This loss function is calculated based on cross entropy, and it will pay more attention to categories with a large number of samples, resulting in poor prediction effect for categories with a small number of samples. Dice loss measures the prediction effect of the model by calculating the overlap between the prediction results and the real labels, and encourages the model to segment a few categories more accurately. As a result, Dice loss has a certain relieving effect on the imbalance of medical images in categories. The cross entropy loss function and Dice loss function are shown in (4) and (5) respectively,

$$L_{\mathrm{BCE}}=-\frac{1}{N}\sum_{i=1}^{N} y_i \log(p_{i,c})+(1-y_i)\cdot\log(1-p_{i,c}), \quad (4)$$

$$L_{\mathrm{Dice}}=1-\left(2\times\sum_{i=1}^{N}\frac{y_i\times p_{i,c}}{y_i+p_{i,c}}\right) \quad (5)$$

where N represents the number of voxels of MRI data input to the model, y_i represents the true category of pixel i, and $p_{i,c}$ represents the probability that the model predicts the i-th voxel as the true category c of that voxel.

In the segmentation network of fused edge attention proposed in this paper, the weighted sum of three loss functions is used, and the total loss function is shown in (6). The three loss functions are respectively the cross entropy loss function of the segmentation region, the Dice loss function of the segmentation region, and the cross entropy loss function of the edge. The cross entropy loss of the segmentation region (L_{BCE}) measures the classification accuracy of the prediction. The Dice loss of the segmentation region (L_{Dice}) is a measure of how similar the predicted split region is to the real label. The cross entropy loss of the edge (L_{Edge}) is used to measure the difference between the edge prediction graph and the real edge label:

$$L_{\mathrm{total}}=\alpha_1 L_{\mathrm{BCE}}+\alpha_2 L_{\mathrm{Dice}}+\alpha_3 L_{\mathrm{Edge}}. \quad (6)$$

By weighting the three losses, the importance of the three loss functions can be balanced according to the task requirements and data characteristics. At the same time, the weight of the loss affects the training effectiveness of the model. However, in previous studies, the weight loss was simply set manually by experience or selected by simple comparative experiments, which would consume a lot of computing resources, and make it difficult to determine the best weight combination. To solve the above problems, in this paper an adaptive weighting method is proposed, which introduces trainable weight parameters into the loss function, sets α_1, α_2 and α_3 as trainable parameters, updates the weights by backpropagation and optimization algorithm, and adaptively learns the optimal value of the weights according to the characteristics of the training data.

However, in the process of optimization of network parameters in backpropagation algorithm, simply using linear sum of losses to learn the weight will cause the weight α to rapidly converge to zero. Therefore, by experimental verification, in this paper a penalty factor α is added to the loss function, which punishes the optimization direction making α decaying rapidly. Finally, the adaptive weighted hybrid loss function proposed in this paper is shown in (7), which is smooth and differentiable, and the task weight will not converge to zero:

$$L_{\mathrm{final}}=\alpha_1 L_{\mathrm{BCE}}+\alpha_2 L_{\mathrm{Dice}}+\alpha_3 L_{\mathrm{Edge}}+\sum_{j}^{3}\frac{1}{\alpha_j}. \quad (7)$$

3. Experimental Results and Analysis

In order to verify the effectiveness of the proposed algorithm in brain tumor segmentation task, a series of comparative experiments were designed to observe the influence of different algorithms on the experimental results, and to analyze the reasons. The experimental environment of this study was installed on a Baode server with 12 GB GPU, 1 TB hard disk capacity, 32 GB RAM and Windows 10 operating system. The software platform used was PyTorch1.9, Python3.7 and related Python function libraries.

3.1 Data Set and Experimental Settings

In this paper, the BraTS2019 and BraTS2020 public training dataset [22], [23] was adopted to train and test the model. The BraTS2019 included 335 cases of brain gliomas, including 76 cases of low-grade gliomas and 259 cases of high-grade gliomas. The BraTS2020 included 369 cases of brain gliomas, including 76 cases of low-grade gliomas and 293 cases of high-grade gliomas. Each case

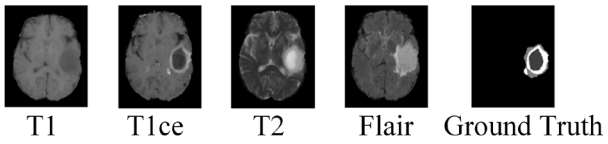


Fig. 7. Four modal MRI images and physician-labeled brain tumor image of a patient.

contains T1-weighted MRI (T1), T1-weighted MRI with contrast enhancement (T1ce), T2-weighted MRI (T2), fluid-attenuated inversion recovery (Flair) and a ground truth label manually segmented by experts, which is shown in Fig. 7.

The experiment uses 5-fold-cross-validation to divide the BraTS2019 and BraTS2020 data set according to 4 : 1. The size of the input image block is 64 × 128 × 128 voxels. The network is optimized by Adam optimizer. The initial learning rate of the weight of the network model is 3e–04, the initial learning rate of the weight of the loss function is 1e–06, the weight decay rate is 1e–05, the dropout rate is 0.5, and the batch size is 2. The maximum number of epochs is 400, and the early stop method is used to end the training process.

The basic process of the experiment was as follows. Step 1 is that bias field correction and standardized preprocessing were performed on MRI images in BraTs2019 dataset. The second step is to perform basic data enhancement on the pre-processed MRI images, such as translation, flipping, rotation, mirroring, etc. The default usage probability of each operation is 0.3. In the third step, Sobel operator is used to get the edge supervised label image. The fourth step is to use the processed data to train the proposed fusion edge attention mechanism model. The fifth step is to test the performance of the model using test data and evaluate the segmentation results.

3.2 Evaluation Metrics

Using the segmentation algorithm the WT, TC and ET regions from the patient's multimodal MRI image will be segmented. WT contains TC, and TC contains ET, and the three regions present a hierarchical nested topology. In this paper, Dice coefficient, sensitivity, specificity and 95% Hausdorff distance are used to evaluate the segmentation performance of the model for the three regions, in which Dice coefficient is the main evaluation index, and Hausdorff distance is mainly used to evaluate the segmentation accuracy of the model for tumor boundaries. Sensitivity, also known as recall rate, can show the degree of influence of loss function on segmentation results. The expressions of the three indicators are respectively shown in (8)–(11):

$$Dice(\mathbf{P},\mathbf{T}) = \frac{|\mathbf{P} \wedge \mathbf{T}|}{(|\mathbf{P}| + |\mathbf{T}|)/2}, \tag{8}$$

$$Sensitivity(\mathbf{P},\mathbf{T}) = \frac{|\mathbf{P} \wedge \mathbf{T}|}{|\mathbf{T}|}, \tag{9}$$

$$Specificity(\mathbf{P}_0,\mathbf{T}_0) = \frac{|\mathbf{P}_0 \wedge \mathbf{T}_0|}{|\mathbf{T}_0|}, \tag{10}$$

$$Hausdorff(A,B) = \max\left\{\max_{a\in A}\min_{b\in B} d(a,b), \max_{b\in B}\min_{a\in A} d(a,b)\right\} \tag{11}$$

where **P** denotes the tumor area predicted by the model. **T** correspondingly is the real tumor area. $\mathbf{P}_0$ represents the background area predicted by the model. $\mathbf{T}_0$ represents the real background area. $\wedge$ represents a logical 'and' operation, and $|\cdot|$ represents the size of sets. The a is the point on the surface A of the region **T**, and b is the point on the surface B of the region **P**. The function $d(\cdot)$ is used to calculate the distance between the points a and b.

3.3 Ablation Experiments

3.3.1 Ablation of Different Modules

To demonstrate the effectiveness of adding modules to the proposed method, ablation experiments were performed on the baseline model 3DUNet on the BraTs2019 dataset. And the addition of edge attention (EA) mechanism and multi-feature fusion module (MFM) respectively were tested. The effectiveness of EA, MFM and the adaptive weighted hybrid Loss (AWL) function are shown in Tab. 1.

As can be seen from Tab. 1, after the edge attention module is added to the 3DUNet model, all its indicators are improved, especially the Dice score of the ET region is increased by 3.93%, while the Mean Hausdorff of the three subregions is significantly reduced, which indicates that after adding prior knowledge of edge importance to the model. The model learns more features to distinguish the

Models	Mean Dice (%)			Mean Sensitivity (%)			Mean Hausdorff (mm)		
	WT	TC	ET	WT	TC	ET	WT	TC	ET
UNet(base)	89.79	85.13	80.90	94.08	85.46	80.77	13.61	7.47	5.45
UNet+EA	90.54	86.97	84.83	92.43	89.45	86.28	6.15	5.64	3.79
UNet+EA+MFM	90.07	86.64	86.79	92.78	88.12	87.42	4.74	5.86	3.65
UNet+EA+AWL	90.43	87.47	85.67	91.38	86.43	84.27	6.37	5.07	3.44
Proposed	**91.10**	**87.16**	**88.86**	**94.50**	**93.17**	**88.86**	**3.92**	**5.12**	**1.92**

Tab. 1. Performance comparison of different improvements.

region near the edge, which improves the segmentation accuracy. After the multi-feature fusion module is used, the edge information is fused with the reference semantic segmentation network at the feature level, which improves the application effect of the features. The training process of the model is optimized by using the proposed adaptive weighted mixed loss idea in the loss function. When the three improvements are used at the same time, the model can effectively combine the advantages of the three improvements, and can achieve excellent results on all indicators, which shows that the improvements of edge attention module, multi-feature fusion module and adaptive weighted hybrid loss function are feasible and very effective.

3.3.2 Ablation of Edge Attention Module

In order to demonstrate the role of edge attention module and channel attention during the model prediction process, a series of comparative experiments on the BraTs2019 dataset were conducted, comparing the baseline model 3DUNet, the channel-enhanced edge attention module (CEEA), and the non-channel-enhanced edge attention module (NCEA). The experimental results are shown in Fig. 8.

As shown in Fig. 8(a) and 8(d), the 3DUnet network was accurate in predicting the gross tumor region without marginal supervision, but lacked more powerful recognition ability in the marginal region, resulting in large false positive areas. As shown in Fig. 8(b) and 8(e), after adding the idea of edge supervision, the model learned more edge-related features and guided the segmentation network, thus limiting the segmented tumor region and significantly reducing the false-positive region. As shown in Fig. 8(c) and 8(f), after adding channel weights in the edge attention module and redistributing ideas, the model's prediction of tumor edge is significantly more accurate, and the segmentation of tumors by the guiding model is also the closest to the results labeled by experts.

3.3.3 Ablation of Loss Function

In order to prove the validity of the loss function proposed in this paper, ablation experiments were conducted on the BraTs2019 data set for each part of the loss function, and the experimental results are shown in Tab. 2.

In Tab. 2, the training effects of the model are compared using the single loss function, the average weighted sum, and the proposed adaptive weighted mixed loss. These automatic weights can converge to similar optimal values during training iterations. Throughout the entire training process, the final task weights may vary. For our final model, at the end of training, the three parts of the loss were weighted in a ratio of 2.6 : 6.3 : 1.06, respectively. As can be seen from Tab. 3, the accuracy of training model using mixed loss function is significantly higher than that of single loss. For example, the average segmentation accuracy can be improved by 5.31% only using simple linear average. Especially for small targets in TC

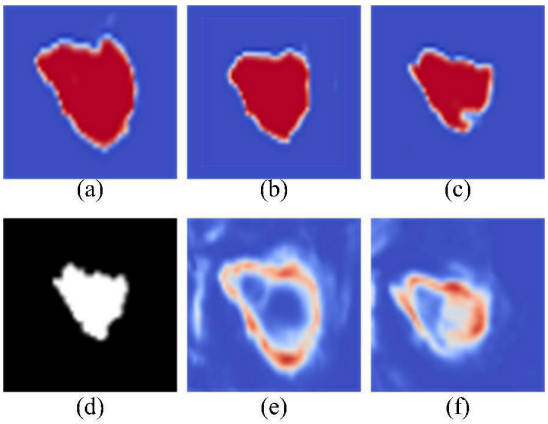


Fig. 8. Heat map comparison of experimental results on prediction of (a) baseline model, (b) CEEA, (c) NCEA, (d) ground truth, (e) edge with NCEA, and (f) edge with CEEA.

region and ET region, the improvement amplitude is larger, which indicates that the fusion of cross entropy loss and dice loss effectively improves the segmentation effect of the model on small volume targets. After adding edge loss, the model improves the importance of edge features in the learning process under supervised conditions, and the segmentation effect of edge parts is significantly improved. When using the determined weights, the model cannot adjust the weights of each part in real time according to the training effect, so the performance is poor. When the trainable parameter is used to determine the weight, the model will reduce the weight of part with high loss, so as to avoid the parameter adjustment with too large step in the wrong direction. When the three-part adaptive weighted loss function is used, the model achieves the best segmentation effect.

3.4 Results Comparison of the Advanced Models

To demonstrate the effectiveness of the proposed algorithm, we first compared it with 7 state-of-the-art CNN methods on the BraTs2019 dataset, including 2DUNet [10], 3DUNet, Attention Unet [24], UNet++ [7], ET-Net [15], Point-unet [25], and TransBTS [26]. For the fairness of comparison, all the results of different models are the best ones obtained by optimization. And the same data preprocessing and training steps are used. The experimental results are shown in Tab. 3 and Tab. 4. Best results are highlighted in bold.

As can be seen from the experimental results in Tab. 3, on the BraTs2019 dataset, the average Dice scores of the model proposed in this paper are as high as 91.10%, 87.16% and 88.86%, respectively, which are 1.31%, 2.03% and 7.96% higher than that of the baseline model 3DUNet. This shows that the improved network architecture can effectively improve the ability of the model to divide the TC region and the ET region, especially for the small ET region. It is worth noting that due to the addition of the boundary attention module, the model pays more attention

to boundary-related features, which greatly improves the boundary perception measure Hausdorff, which is significantly better than other methods. As can also be seen from Tab. 4, the proposed model also demonstrates excellent performance, particularly on the Hausdorff distance, which indicates that the proposed method is highly effective in edge constraints.

Figures 9 and 10 show the segmentation results using models such as 3DUNet compared to labeled images. It can be found from Fig. 9 and Fig. 10 that the fuzzy boundary between the lesion area and the healthy voxels often leads to errors in the model's classification of voxels near the boundary. Using the guidance of boundary attention, the model pays more attention to distinguishing the features of the boundary, which is more in line with the habit of manual annotation by experts, and also significantly improves the segmentation accuracy, making the segmentation result closer to the real label.

4. Summary

Precise segmentation of brain MRI images is of great significance to guide clinical diagnosis and surgical planning, and automatic segmentation technology can greatly alleviate the adverse effects of manual segmentation, which is time-consuming and labor-intensive, and depends on the subjective cognition of doctors. In this paper, an effective

Loss	Weights distribution			Mean Dice (%)			Mean Hausdorff (mm)		
	L_{BCE}	L_{Dice}	L_{Edge}	WT	TC	ET	WT	TC	ET
Only L_{BCE}	1	0	0	86.36	81.26	79.64	9.46	7.46	7.41
Only L_{Dice}	0	1	0	85.41	82.89	81.46	12.46	14.56	11.58
L_{BCE} & L_{Dice}	1	1	0	88.11	85.69	86.79	8.16	7.99	6.87
Unweighted sum	1	1	1	89.16	86.48	87.54	5.15	5.93	3.36
Experience tuning	2	3	2	90.37	86.64	86.79	4.74	5.86	3.65
Two loss automatic weighting	√	√	-	90.87	**87.24**	86.34	7.86	9.02	4.46
	√	-	√	89.12	85.56	86.41	5.46	7.56	2.47
	-	√	√	87.64	85.48	87.57	7.96	6.48	7.46
Proposed	√	√	√	**91.10**	87.16	**88.86**	**3.92**	**5.12**	**1.92**

Tab. 2. Experimental results using different loss functions.

model	Mean Dice (%)			Mean Sensitivity (%)			Mean Specificity (%)			Mean Hausdorff (mm)		
	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET
2DUNet	88.16	82.85	79.94	92.36	86.24	80.56	99.84	99.93	99.94	16.35	11.27	9.16
3DUNet	89.79	85.13	80.90	94.08	87.46	80.77	99.84	99.94	99.95	13.61	7.47	5.45
Attention Unet	88.98	82.36	81.23	91.45	86.45	76.57	99.85	99.94	99.95	14.12	10.17	8.53
UNet++	90.27	84.06	79.55	94.23	85.77	80.41	99.88	99.95	99.95	10.21	9.72	6.78
ET-Net	90.78	86.39	84.73	92.54	89.46	86.15	99.86	99.95	99.96	6.57	5.49	3.19
Point-unet	**91.61**	87.07	86.42	91.61	87.94	87.18	**99.89**	**99.96**	99.96	9.16	7.13	8.75
TransBTS	91.15	86.43	86.47	92.41	**94.92**	86.39	99.88	99.95	99.96	6.57	9.16	6.65
Proposed	91.10	**87.16**	**88.86**	**94.50**	93.17	**88.86**	99.88	99.96	**99.96**	**3.92**	**5.12**	**1.92**

Tab. 3. Comparison of segmentation results of different model on BraTs2019.

model	Mean Dice (%)			Mean Sensitivity (%)			Mean Specificity (%)			Mean Hausdorff (mm)		
	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET
3DUNet	85.45	82.87	78.86	91.15	85.72	79.86	99.82	99.93	99.93	18.34	10.53	10.04
Point-unet	89.34	**85.34**	83.37	89.46	85.61	85.37	**99.86**	**99.95**	99.94	13.25	9.23	9.52
TransBTS	89.04	84.16	83.94	88.31	89.32	**85.42**	99.85	99.95	99.95	7.34	9.14	9.37
Proposed	**90.23**	84.04	**85.67**	**92.50**	**90.43**	85.11	99.85	99.94	**99.96**	**5.25**	**8.94**	**6.16**

Tab. 4. Comparison of segmentation results of different model on BraTs2020.

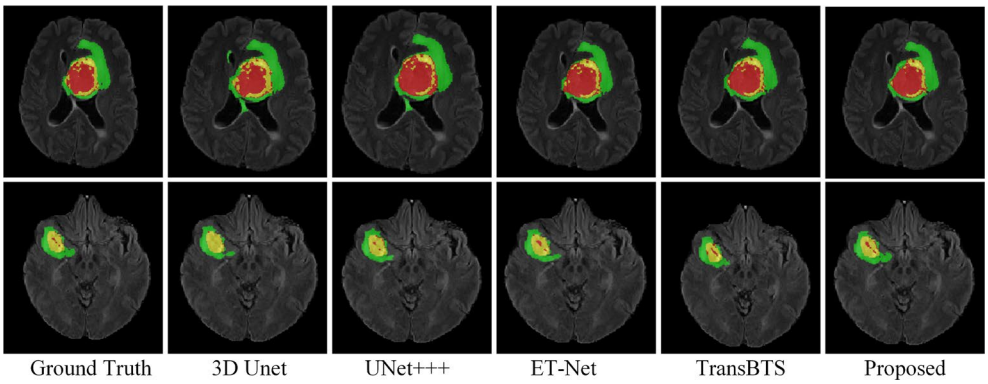


Fig. 9. 2D visualization of segmentation results of different algorithms.

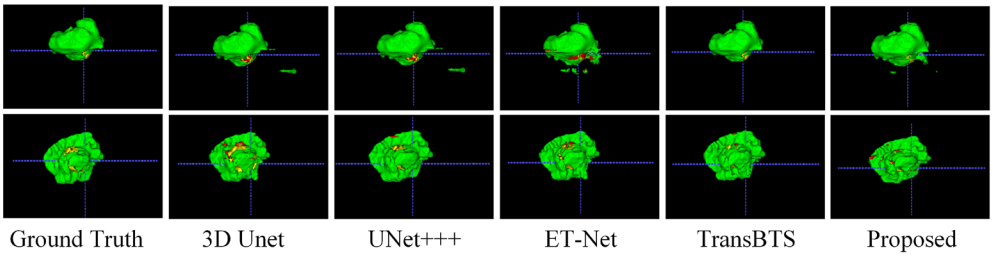


Fig. 10. 3D visualization of segmentation results of different algorithm.

segmentation algorithm of 3D brain tumor based on fusion of edge attention mechanism is proposed. The network generally follows the encoder-decoder structure, in which the features are extracted by the first two encoding block, and are inputted into the edge attention module to predict the tumor edge. Then the above prediction from encoder is connected the prediction from the decoder for feature fusion. Finally, the proposed adaptive weighted hybrid loss function is used to adjust the weight of the three-part loss adaptively and to train the network, which can avoid too much adjustment in the wrong direction during the training process. The test results of BraTs2019 data set show that the average Dice value of the algorithm in WT, TC and ET regions can reach 91.10%, 87.16% and 88.86%, respectively, which proves the feasibility and effectiveness of the method. Our code is available at https://github.com/tianhengyi/EFUNET.

Acknowledgments

This study is supported by the Joint Project of Beijing Natural Science Foundation and Beijing Municipal Education Commission (No. KZ202110011015).

References

[1] KLEIHUES, P., BURGER, P. C., SCHEITHAUER, B. W. The new WHO classification of brain tumors. *Brain Pathology*, 1993, vol. 3, no. 3, p. 255–268. DOI: 10.1111/j.1750-3639.1993.tb00752.x

[2] SUN, H., YANG, S., CHEN, L., et al. Brain tumor image segmentation based on improved FPN. *BMC Medical Imaging*, 2023, vol. 23, no. 1, p. 1–10. DOI: 10.1186/s12880-023-01131-1

[3] XU, Y., JIA, Z. P., AI, Y. G., et al. Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brisbane (Australia), 2015, p. 947–951. DOI: 10.1109/ICASSP.2015.7178109

[4] DONG, H., YANG, G., LIU, F., et al. Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. In *Proceedings of Annual Conference on Medical Image Understanding and Analysis*. Edinburgh (Scotland), 2017, p. 506 to 517. DOI: 10.1007/978-3-319-60964-5_44

[5] BEERS, A., CHANG, K., BROWN, J., et al. Sequential 3D U-nets for biologically-informed brain tumor segmentation. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Hawaii (USA), 2017, p. 235–242. DOI: 10.48550/arXiv.1709.02967

[6] DIAKOGIANNIS, F. I., WALDNER, F., CACCETTA, P., et al. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, vol. 162, no. 1, p. 94–114. DOI: 10.1016/j.isprsjprs.2020.01.013

[7] ZHOU, Z., SIDDIQUEE, M. M. R., TAJBAKHSH, N., et al. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 2019, vol. 39, no. 6, p. 1856–1867. DOI: 10.1109/TMI.2019.2959609

[8] ISENSEE, F., JÄGER, P. F., FULL, P. M., et al. The nnU-net for brain tumor segmentation. In *Proceedings of the International Conference on Brainlesion-Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries.* Lima (Peru), 2020, p. 118–132. DOI: 10.48550/arXiv.2011.00848

[9] WADHWA, A., BHARDWAJ, A., VERMA, V. S. A review on brain tumor segmentation of MRI images. *Magnetic Resonance Imaging*, 2019, vol. 1, no. 61, p. 247–259. DOI: 10.1016/j.mri.2019.05.043

[10] AL NASIM, M. A., AL MUNEM, A., ISLAM, M., et al. Brain tumor segmentation using enhanced u-net model with empirical

analysis. In *Proceedings of the 25th International Conference on Computer and Information Technology (ICCIT)*. Cox's Bazar (Bangladesh), 2022, p. 1027–1032. DOI: 10.1109/ICCIT57492.2022.10054934

[11] VADHNANI, S., SINGH, N. Brain tumor segmentation and classification in MRI using SVM and its variants: A survey. *Multimedia Tools and Applications*, 2022, vol. 81, no. 22, p. 31631–31656. DOI: 10.1007/s11042-022-12240-4

[12] SHEN, H., ZHANG, J., ZHENG, W. Efficient symmetry-driven fully convolutional network for multimodal brain tumor segmentation. In *Proceedings of IEEE International Conference on Image Processing*. Beijing (China), 2017, p. 3864–3868. DOI: 10.1109/ICIP.2017.8297006

[13] ZHANG, J., JIANG, Z., DONG, J., et al. Attention gate resU-Net for automatic MRI brain tumor segmentation. *IEEE Access*, 2020, vol. 8, p. 58533–58545. DOI: 10.1109/access.2020.2983075

[14] ABOELENEIN, N. M., SONGHAO, P., KOUBAA, A., et al. HTTU-Net: Hybrid two track U-Net for automatic brain tumor segmentation. *IEEE Access*, 2020, vol. 8, p. 101406–101415. DOI: 10.1109/ACCESS.2020.2998601

[15] ZHANG, Z., FU, H., DAI, H., et al. ET-Net: A generic edge-attention guidance network for medical image segmentation. In *Proceedings of the 22nd International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Shenzhen (China), 2019, p. 442–450. DOI: 10.1007/978-3-030-32239-7_49

[16] LEE, H. J., KIM, J. U., LEE, S., et al. Structure boundary preserving segmentation for medical image with ambiguous boundary. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle (USA), 2020, p. 4817–4825. DOI: 10.1109/CVPR42600.2020.00487

[17] ZHOU, X., LI, X., HU, K., et al. ERV-Net: An efficient 3D residual neural network for brain tumor segmentation. *Expert Systems with Applications*, 2021, vol. 170, p. 1–13. DOI: 10.1016/j.eswa.2021.114566

[18] LIN, T. Y., GOYAL, P., GIRSHICK, R., et al. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, vol. 42, no. 2, p. 318–327. DOI: 10.1109/TPAMI.2018.2858826

[19] LU, J. L., WANG, Z. Y., BIER, E., et al. Bias field correction in hyperpolarized Xe-129 ventilation MRI using templates derived by RF-depolarization mapping. *Magnetic Resonance in Medicine*, 2022, vol. 88, no. 2, p. 802–816. DOI: 10.1002/mrm.29254

[20] HE, K., ZHANG, X., REN, S., et al. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas (USA), 2016, p. 770–778. DOI: 10.1109/CVPR.2016.90

[21] HU, J., SHEN, L., SUN, G. Squeeze-and-excitation networks. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City (USA), 2018, p. 7132–7141. DOI: 10.1109/CVPR.2018.00745

[22] MENZE, B. H., JAKAB, A., BAUER, S., et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 2014, vol. 34, no. 10, p. 1993 to 2024. DOI: 10.1109/TMI.2014.2377694

[23] BAKAS, S., AKBARI, H., SOTIRAS, A., et al. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data*, 2017, vol. 4, no. 1, p. 1–13. DOI: 10.1038/sdata.2017.117

[24] OKTAY, O., SCHLEMPER, J., LE FOLGOC, L., et al. Attention u-net: Learning where to look for the pancreas. In *Proceedings of International Conference on Medical Imaging with Deep Learning (MIDL)*. Amsterdam (Netherlands), 2018, p. 1–10. DOI: 10.48550/arXiv.1804.03999

[25] HO, N. V., NGUYEN, T., DIEP, G. H., et al. Point-unet: A context-aware point-based neural network for volumetric segmentation. In *Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Strasbourg (France), 2021, p. 644–655. DOI: 10.1007/978-3-030-87193-2_61

[26] WANG, W., CHEN, C., DING, M., et al. TransBTS: Multimodal brain tumor segmentation using transformer. In *Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Strasbourg (France), 2021, p. 109–119. DOI: 10.1007/978-3-030-87193-2_11

About the Authors ...

Yu WANG (corresponding author) was born in 1977. She received her Ph.D. degree from the University of Science and Technology Beijing in 2009. She was engaged in scientific research as a post-doctoral in the Beijing Key Laboratory of Multidimensional and Multiscale Computing Photography, Tsinghua University from 2009 to 2011. She is now a Professor and doctoral supervisor of the Beijing Technology and Business University. Her research interests include pattern recognition, medical image processing and computer vision.

Hengyi TIAN was born in 1999. He is now a candidate of master degree in the School of Artificial Intelligence, Beijing Technology and Business University, China. His research interests include pattern recognition, 3D multimodal medical image segmentation and computer vision.

Minhua LIU (corresponding author) was born in 1976. He received his Ph.D. degree from Tsinghua University. Now he is the vice president of the Beijing Technology and Business University. His research interests include image processing and the modeling and analysis of complex systems.