Full-automatic Segmentation Algorithm of Brain Tumor Based on RFE-UNet and Hybrid Focal Loss Function

Yu WANG, Hengyi TIAN, Yarong JI, Minhua LIU[*]

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

wangyu@btbu.edu.cn, thy8562@sina.com, 571529288@qq.com, liuminhua@btbu.edu.cn*

Submitted October 22, 2023 / Accepted March 20, 2024 / Online first April 24, 2024

Abstract. *Semantic segmentation of glioma and its subregions plays a critical role in the entirely clinical workflow of brain cancer diagnosis, monitoring, and treatment planning. Recently, automatic tumor segmentation has attracted a lot of attention, especially supervised learning methods based on neural networks, and the popular "U-shaped" network architecture has achieved state-of-the-art performance in many fields of medical image segmentation. Despite the success of these models, the commonly used small convolution kernel can only extract local features, and more global contextual features cannot be learned, resulting in the disappointing performance of modeling long-range information. At the same time, due to the difficulty of obtaining medical image data, and the imbalance of tumor data in which tumor usually occupies a relatively small volume compared with the background, the adverse influence on the training of the model occurs. In this paper, a novel segmentation framework including TensorMixup data augmentation, improved Receptive Field Expansion UNet (RFE-UNet) and hybrid loss function is designed. Specifically, the TensorMixup algorithm in the data preprocessing phase is used to provide more high-quality training data. In the training phase, both a RFE-UNet network and a hybrid loss function are proposed respectively. RFE-UNet network adds Receptive field expansion module based on Dilated convolution in the first three stages of skip connection, which is used to learn more local and global features. In addition, hybrid loss function is mainly composed of focal loss and focal Tversky loss, focal loss increasing the weight of fewer samples and focal Tversky loss focusing on learning the characteristics of samples with incorrect predictions, which is adopted to alleviate data imbalance. The experimental results on the BraTs2019 dataset show that the average Dice value of the proposed algorithm in the intact tumor, tumor core, and enhanced tumor region can reach 91.55%, 89.23%, and 84.16% respectively, which proves the feasibility and effectiveness of using the proposed architecture.*

Keywords

Segmentation, brain tumor, magnetic resonance imaging, dilated convolutions, three-dimensional CNN

1. Introduction

As a common brain disease, brain tumor has a serious threat to human health because of high fatality rate. Brain tumors are uncontrollably abnormal cell growth in the brain. They grow rapidly, are closely connected with the surrounding normal tissues, and infiltrate each other, resulting in blurred boundaries and unlimited proliferation. Malignant tumors are highly invasive and can quickly destroy the normal function of the human body [1]. Currently, magnetic resonance imaging (MRI) is commonly used in the medical diagnosis and labeling of brain tumors. Segmentation of tumor locations from MRI is a key step in the diagnosis and treatment plan of brain tumors. However, relying only on manual segmentation of the location of brain tumors and different lesion areas by doctors is not only time-consuming and laborious, but also requires higher diagnostic experience, and has great subjectivity. With the development of computer technology, the use of computer and MRI-related to assist the diagnosis of brain tumors has many advantages such as high efficiency and strong objectivity [2].

In recent years, automatic segmentation is a research hotspot in the field of medical image analysis. Machine learning techniques, especially deep learning methods, can automatically extract feature representations to achieve accurate and stable segmentation performance [3]. A deep learning model called UNet has been proposed, and its performance has been significantly improved. Dong [4] used the two-dimensional (2D) UNet network for brain tumor segmentation earlier. But the 2D UNet model could not effectively establish the connection between the slices of three-dimensional MRI image data, and each slice could only be processed separately, which would affect the segmentation accuracy of the model to a certain extent. Myronenko [5] proposed a three-dimensional (3D) work architecture, which achieved high segmentation accuracy in BraTS2017 datasets. Although these methods have yielded good results, many problems exist.

When 3D UNet is used for brain tumor segmentation, because its input data are 3D MRI images of four modalities, a large number of parameters are used to calculate for the training of model. Thus, the amount of computation is

DOI: 10.13164/re.2024.0253

increased, and more memory and computing resources are consumed [6]. For avoiding this problem, convolution kernels with small size is generally used in 3D UNet [7]. However, a small-size convolution kernel will narrow the perception range of MRI image information, resulting in a smaller range of the model's field of perception of the image, so that the model is more inclined to extract the local features of the MRI image, and to ignore the global features in the image, affecting the segmentation accuracy of the model [8]. Inspired by the development of dilated convolution, various methods of expanding the receptive field have been proposed. Lopez [9] firstly proposed the dilated residual network for brain tumor segmentation, and proved the effectiveness of expansive convolution by patch extraction training and testing. Ding [10] proposed a residual network combining convolution and dilated convolution, which improved the feature extraction capability of the network. Bala [11] proposed an initial module of multiscale expansion based on expansive convolution to make the model wider, and to solve the problem of vanishing gradient.

During the training phase, because of the problem of class imbalance in brain tumor MRI, difficulties in model optimization can be met. Since the volume of the tumor region is much smaller than the one of the normal tissue in the patient's MRI, typically only 1.54% of the whole brain image, and the volume of different sub-regions such as edema, necrosis, enhancing and non-enhancing part within the tumor is even smaller [12]. These phenomena make the negative samples including normal tissue and background dominate in the model during the training phase, and the model has difficulty in extracting enough effective features from the smaller number of positive samples to shift the optimization route, further reducing the optimization quality of the model. In addition, the model is prone to false positive and false negative predictions when classifying the input samples, known as the output imbalance problem, which in turn leads to incorrect segmentation of the input images, and reduces the segmentation accuracy. In view of the above data imbalance problem, researchers usually select the appropriate loss function for the model. The model minimizes the loss function value using backpropagation, which can make its segmentation performance tend to be optimal. In the field of image segmentation, the focus loss function [13] is often used to alleviate class imbalance. Abraham [14] proposed a focused Tversky loss function based on Tversky index to solve the output imbalance problem. In recent years, the mixed loss function has been widely used. It combines various independent loss functions, makes full use of the advantages of each loss function, and has a good development potential. Taghanaki [15] proposed a Combo loss function combining cross entropy and Dice loss, and Wang [16] proposed a DiceFocal loss function combining Dice loss and cross entropy loss. The problem of output imbalance can be effectively solved by all these methods. However, the loss function used in most studies can only solve the problem of single data imbalance. At present, Yeung [17], [18] proposed a new mixed focus loss function based on focus loss and focus Tversky loss, which was used in the dichotic medical image segmentation task. His research proved that the function could effectively alleviate the problems of class imbalance and output imbalance.

Significantly, when the brain tumor segmentation model is constructed by deep learning method, a large number of annotated data is required to train the network. The acquisition of clinical MRI data annotated by doctors is very expensive and time-consuming. However, for data-driven methods, insufficient annotated data will severely limit the performance of deep learning network. To solve this problem, basic data augmentation techniques such as flipping, rotation, mirroring, and scaling are generally used to transform the original data to expand the data set [19], [20], but studies have shown that these methods can only bring a slight improvement to the model performance [21], [22]. Recently, Mixup, a data augmentation technology based on interpolation, was proposed [23]. This method combines two randomly selected images and their unique thermal coding labels in a convex combination to synthesize new images and their label sequences. Good data augmentation results can be obtained by training the model with the synthesized new data. Wang [24] proposed TensorMixup in which firstly image blocks from MRI sequences are mixed. Then, a tensor with all elements independently sampled in beta distribution is used to mix the information of image blocks. Finally, the above tensor is mapped into a matrix for mixing the unique thermal coding label sequence of image blocks to form new images and their annotation data.

Inspired by the above ideas, in this paper a new automatic segmentation architecture is designed for brain tumors. Specifically, we developed the model in the following way. Firstly, the TensorMixup data augmentation algorithm is used to expand the data set, which effectively solved the overfitting problem caused by insufficient annotated data. Secondly, the proposed UNet with feature extraction paths for receptive field expansion, called RFE-UNet network is used to effectively improve the inadequately and globally contextual feature extraction. Thirdly, the proposed hybrid focus loss function including focal loss and focal Tversky loss can effectively alleviate the input-output imbalance problem.

2. Method

2.1 Data Preprocessing

When medical images are acquired, the original images usually contain brightness unevenness and noise due to the influence of imaging equipment, imaging principles and individual's own differences [25]. Therefore, before segmentation, the data provided by brain tumor image segmentation (BraTS) requires pre-processing such as bias field correction and normalization to reduce misdiagnosis

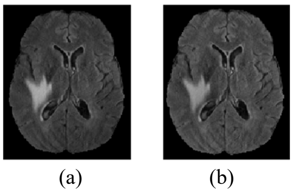

Fig. 1. Bias field correction results: (a) Original image, (b) corrected result.

and to improve diagnostic accuracy. In this paper, we use the N4ITK bias field correction method to remove the inhomogeneity of the images, and an example of the brain tumor images before and after processing is shown in Fig. 1. Since the raw data are from different institutions, and are obtained by different scanning instruments, the range of values of each group of data can be inconsistent, which requires the use of the Z-Score normalization method to unify the values of all data into a smaller range while making the values of the images show a normal distribution to facilitate the numerical calculation of the model during the training process. In this study, the four modal images of each patient were normalized individually, and the Z-Score normalization formula is shown in (1):

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

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 $\mathbf{X}$.

2.2 TensorMixup

As far as we know, a network with more parameters requires more training data to solve the overfitting problem. Since the images in the database are constant, the amount of data in the training set is increased by a data augmentation algorithm.

The TensorMixup algorithm can generate high-quality brain tumor images from the original dataset. Firstly image patches of the tumor region were obtained from the MRI brain images of the same modality of the two patients, respectively. The execution of this process mainly depends on the boundary box information of tumor regions which could be acquired from the ground truth labels among original images. And the acquired image patches need to be resized to 128 × 128 × 128 voxels, denoted by $\mathbf{X}_1$ and $\mathbf{X}_2$. Next, the information of the image blocks is mixed using a tensor $\mathbf{\Lambda}$ with all elements independently sampled in a beta distribution, which in turn synthesizes a new image sequence. The image mixing process is shown in (2):

$$\mathbf{X} = \mathbf{\Lambda} \odot \mathbf{X}_1 + (1 - \mathbf{\Lambda}) \odot \mathbf{X}_2 \tag{2}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ denote the two image blocks to be mixed, the synthetic image block is denoted by $\mathbf{X}$, $\mathbf{\Lambda}$ is a tensor in which all elements are sampled in the Beta(α, α) distribution, and whose dimensions are identical to those of the image block, and $\odot$ denotes the Hadamard product operation.

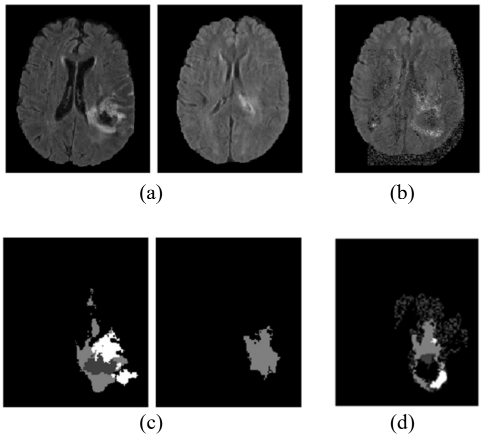

Fig. 2. Fusion results of TensorMixup: (a) Two original MRI images, (b) mixed image, (c) two original label images, (d) mixed label.

Then the label images of $\mathbf{X}_1$ and $\mathbf{X}_2$ are transformed into the onehot encoding sequences $\mathbf{Y}_1$ and $\mathbf{Y}_2$, which are both of size $128^3 \times 4$. The number of matrix rows is the total number of voxels in the label images of $\mathbf{X}_1$ and $\mathbf{X}_2$, and the four elements of each row of the matrix indicate the true probability that the voxels corresponding to that row in $\mathbf{X}_1$ and $\mathbf{X}_2$ belong to each of the four categories. Since $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are two-dimensional matrices, and tensor $\mathbf{\Lambda}$, $1 - \mathbf{\Lambda}$ are three-dimensional tensors, $\mathbf{\Lambda}^*$ is used to mix $\mathbf{Y}_1$ and $\mathbf{Y}_2$ when mixing the unique thermal encoding labels. $\mathbf{\Lambda}^*$ is related to $\mathbf{\Lambda}_\text{v}$ as in (3), and $\mathbf{\Lambda}_\text{v}$ is related to $\mathbf{\Lambda}$ as shown in (4). vec($\mathbf{\Lambda}$) denotes the vectorization operation on tensor $\mathbf{\Lambda}$.

$$\mathbf{\Lambda}^* = f(\mathbf{\Lambda}_\text{v}) = [\mathbf{\Lambda}_\text{v}, \mathbf{\Lambda}_\text{v}, \mathbf{\Lambda}_\text{v}, \mathbf{\Lambda}_\text{v}], \tag{3}$$

$$\mathbf{\Lambda}_\text{v} = \text{vec}(\mathbf{\Lambda}). \tag{4}$$

The label data can be mixed by transforming the original label with the mixing tensor accordingly, and the mixing process is described by (5). After the new data ($\mathbf{X}$, $\mathbf{Y}$) are synthesized, they can be directly fed into the neural network to obtain the loss function values, and to optimize the network parameters. The images before and after mixing are shown in Fig. 2 in which Figure 2(a) shows two Flair modal images to be mixed, and Figure 2(b) shows the mixed image, Figure 2(c) shows two original label images to be mixed, and Figure 2(d) shows the mixed label.

$$\mathbf{Y} = \mathbf{\Lambda}^* \odot \mathbf{Y}_1 + (1 - \mathbf{\Lambda}^*) \odot \mathbf{Y}_2 \tag{5}$$

2.3 Improved 3D RFE-UNet

2.3.1 3D UNet

According to the characteristics of brain tumor MRI images, for ensuring the segmentation effect, in this paper 3D UNet is introduced as the base network, the network architecture is shown in Fig. 3, which consists of the encoder and the decoder. The encoder is used to extract the high level semantic features of the input image, and the decoder is used to relocate the location of different semantic regions based on the extracted features. The encoder mainly consists of a downsampling module and a residual convolution module. The downsampling module is a 3 × 3 × 3 convolution with a step size of 2. The residual convolution module contains two 3 × 3 × 3 convolution layers with step size 1 and residual connections as shown in Fig. 4. The number of convolution kernels in the first layer is 16, and the number of convolution kernels is doubled after each downsampling. The convolution is followed by instance normalization (IN) and Leaky Relu activation function. The decoder mainly consists of a convolution module and an upsampling module. The convolution module contains two 3 × 3 × 3 convolution layers with step size of 1 as shown in Fig. 4. The same instance normalization and leaky Relu activation function are used after convolution. The upsampling module contains a 3 × 3 × 3 deconvolution layer with step size of 1. The decoder uses four upsampling operations to convert the low-resolution feature map extracted from the encoding segment into a high-resolution feature map, and a 1 × 1 × 1 convolutional layer with a step size of 1 is added to the network as a segmentation layer. Finally, the results are fed to the sigmoid layer to obtain the output probability map of the model [26] to achieve end-to-end segmentation.

In addition, one of the reasons for the great results obtained is the skip connection structure of the 3D UNet, which splices the output of the same stage encoder feature map with the upsampled feature map of the decoder in the channel dimension. In this way, shallow, low-level, local features in the encoder feature map are combined with deep, high-level, global features in the decoder feature map. Thus, the gradual loss of shallow information in the feature map is compensated after flowing to the decoder. However, the skip connection structure in the 3D UNet simply concatenates the high and low level feature maps, and the concatenated features contain only single-scale features from the encoder with small perceptual fields. So the encoder features which contain different scales cannot be fully captured by the original skip-connected structure.

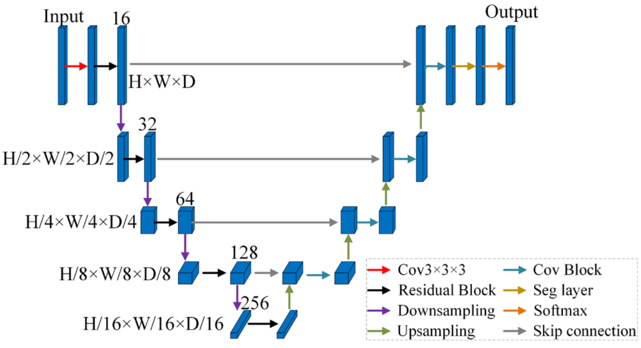

Fig. 3. The framework of 3D UNet.

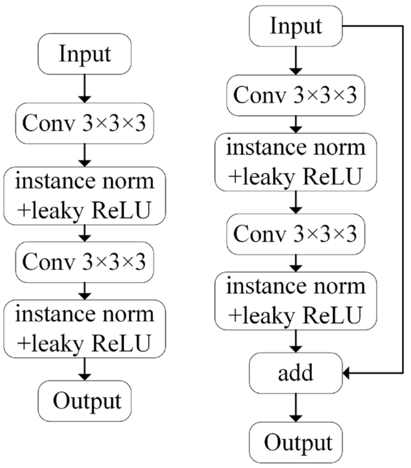
Input
Conv 3×3×3
instance norm +leaky ReLU
Conv 3×3×3
instance norm +leaky ReLU
Output
Input
Conv 3×3×3
instance norm +leaky ReLU
Conv 3×3×3
instance norm +leaky ReLU
add
Output

Fig.4. The structure schematic of convolution module.

2.3.2 The Designed RFE Module

The main role of the receptive field expansion (RFE) module is to capture more global contextual features with a large perceptual field and different scales. The module mainly contains three parallel dilated convolution pathways which use dilated convolution layers with a convolution kernel size of 3 × 3 × 3 and step size of 1. In addition, the dilation ratios are 1, 2, and 3 in order, and each dilated convolution layer is connected with an IN layer and a leaky ReLU layer in turn. The dilated convolution is used to capture the contextual information of the image at a larger ratio without adding any parameters, and to increase the perceptual field of the model. The ratios of the three path-way dilated convolution are set differently in order to resample the input feature map several times using different sampling rates. Thus, features are extracted at different levels of multiple scales. The perceptual field size of the expanded convolution is calculated as shown in (6):

$$r = (n-1)\cdot(ksize+1)+ksize \tag{6}$$

where n denotes the expansion ratio, $ksize$ expresses the convolution kernel size, and r is the size of the receptive field. By equation (2), it is known that the receptive field range of the RFE module in the three pathways in this paper is 3 × 3 × 3, 7 × 7 × 7, and 11 × 11 × 11 respectively.

Since the sampling rates of the input feature maps are different for the dilated convolution in the three paths, the output feature maps of the three paths are not of the same size. For obtaining the output feature maps with the same size to further fuse features and to prevent the loss of image

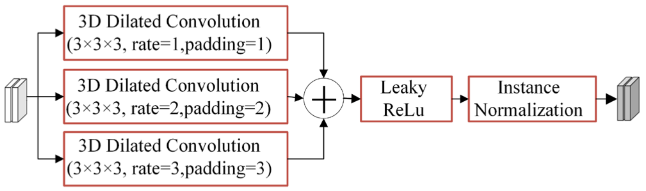
3D Dilated Convolution (3×3×3, rate=1,padding=1)
3D Dilated Convolution (3×3×3, rate=2,padding=2)
3D Dilated Convolution (3×3×3, rate=3,padding=3)
Leaky ReLu
Instance Normalization

Fig. 5. The structure schematic of the RFE module.

edge information, in this study the boundaries of the feature maps are padded with 0 before the convolution operation, and the padding parameters of the convolution layers in the three paths are specifically set to 1, 2, and 3 in turn. The structure of the RFE module is shown in Fig. 5.

The output of RFE module is related to the input x as shown in (7).

$$output=[F_{11}(x),F_{22}(x),F_{33}(x)] \tag{7}$$

where $F_{ln}(x)$ denotes the feature map generated by the lth (l = 1,2,3) pathway, and $F_{ln}(\cdot)$ denotes the composite function which consists of three consecutive operations orderly including IN, leaky ReLU function activation, and 3D dilated convolution calculation with convolution kernel size of 3 and dilation ratio of n. $[F_{11}(x), F_{22}(x), F_{33}(x)]$ denotes that the feature maps generated by the three pathways are concatenated by channel dimension, and the feature maps generated by the RFE module will be concatenate with the feature maps of the decoded part to form fused features containing high and low levels.

2.3.3 3D RFE-UNet

For addressing the problem that the 3D UNet cannot fully extract global features, and considering that the skip connection structure of 3D UNet has certain limitations, in this paper, a receptive field expansion module is designed, referred to as the RFE module whose output and input feature maps have the same dimensions in the wide and high dimensions, so that the module can be added to any of the jump connections in the original network, and also at the input side of the network. The input image can be fused with the output feature map of the decoding part after passing through this module at the network input. By setting the RFE module at different positions of the original network, different brain tumor segmentation effects are produced. Comparison experiments show that the RFE module can be finally introduced into the first, second and third jump connections of the 3D UNet at the same time to form the 3D RFE-UNet, thus achieving an effective combination of high and low level semantic features. The network structure of the 3D RFE-UNet is shown in Fig. 6.

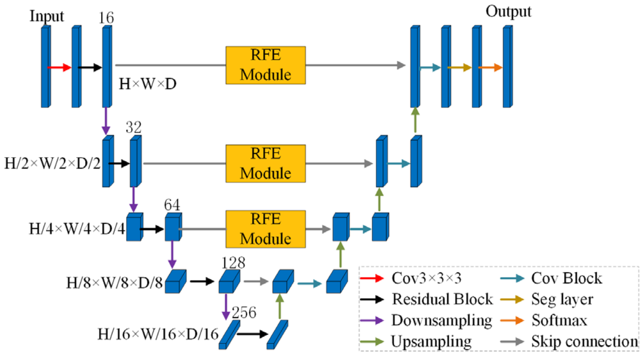

Fig. 6. The framework of 3DRFE-UNet.

2.4 Hybrid Focal Loss Function

In this paper, a hybrid focal loss function is proposed for a multimodal multi-classification brain tumor segmentation task. This hybrid focal loss function is mainly composed of focal loss and focal Tversky loss which will be described in detail below.

2.4.1 Focal Loss

In the deep learning network framework, the traditional focal loss function was proposed by Lin [27] based on the improvement of the cross-entropy loss function, which aims to solve the class imbalance problem of dense target detection task. This kind of focal loss function increases the weights of hard-to-classify samples and few samples, and decreases the weights of easy-to-classify samples and many samples, so that the model can focus more on the learning of tumor lesion features, and also can improve the classification accuracy of the model for hard-to-classify samples. For the brain tumor multi-classification segmentation task studied in this paper, the expression of the focal loss function used by the model is shown in (8):

$$L_{\mathrm{F}}=-\frac{1}{N}\mu\sum_{\mathrm{i}=1}^{N}\left(1-p_{i,c}\right)^{\tau}\cdot\log\left(p_{i,c}\right) \tag{8}$$

where μ denotes the weight of positive samples which can be used to increase the weight of positive samples when the number of positive samples is less than negative samples. Thus the class imbalance problem can be alleviated. τ is called the focusing parameter, and $\tau \geq 0$. $(1-p_{i,c})^{\tau}$ is the modulation coefficient which is mainly used to reduce the weight of easy-to-classify samples, so that the model is more focused on hard-to-classify samples. In addition, $p_{i,c}$ indicates the matrix of predicted values for each class in which indices c and i iterate over all classes and pixels, respectively. N denotes the number of voxels of MRI data input to the model, and in this study $N = 128^3$ and c = 0, 1, 2, or 3 which denote background (BG, 0), edema (ED, 1), necrosis with non-enhanced tumor (NCR/NET, 2) and enhanced tumor (ET, 3), respectively.

2.4.2 Focal Tversky Loss

The focal Tversky loss function originates from the Dice loss function, and is mainly used to reduce the false positive and false negative predictions of the model. By setting the values of weights α and β in this loss function, the weights of false positive and false negative samples in the loss value can be increased, thus enabling the model to focus on learning the features of these incorrectly predicted samples and improving the segmentation accuracy of the model [15]. The expressions of the focal Tversky loss function are shown in (9) and (10):

$$L_{\mathrm{FT}}=\sum_{c=0}^{C}(1-TI_{c})^{\frac{1}{\gamma}}\ , \tag{9}$$

$$TI_c = \frac{\sum_{i=1}^{N} p_{ic} g_{ic}}{\sum_{i=1}^{N} p_{ic} g_{ic} + \alpha \sum_{i=1}^{N} p_{ic} g_{i\bar{c}} + \beta \sum_{i=1}^{N} p_{i\bar{c}} g_{ic}} \qquad (10)$$

where γ is a hyperparameter which allows the model to increase the degree of attention to more difficult regions when $\gamma < 1$. TI_c denotes the Tversky index of class c. p_{ic} is the model's predicted probability of the each voxel belonging to class c, and $p_{i\bar{c}}$ is the model's predicted probability of the voxel which doesn't belong to class c. g_{ic} denotes the true probability of each voxel belonging to class c, and $g_{i\bar{c}}$ denotes the true probability of the voxel which doesn't belong to non-category c. α and β, as a set of hyperparameters, control the weights of false-negative and false-positive samples in the loss values, respectively. By adjusting α and β, the balance between false positives and false negatives can be controlled to alleviate the output imbalance problem.

2.4.3 Improved Hybrid Focal Loss Function

The focal loss function can alleviate the problem of input data class imbalance, while the focal Tversky loss function can alleviate the problem of output data class imbalance. Both of them can make the model focus on the learning of hard-to-classify samples. In order to make the model avoid the problem of input and output data imbalance as much as possible during the construction process, this paper proposes the focal loss and the focal Tversky loss are combined to form a hybrid focal loss function, so that it has the advantages of both loss functions. The expression of the hybrid loss function is shown in (11):

$$L=\lambda L_{\mathrm{F}}+(1-\lambda)L_{\mathrm{FT}} \qquad (11)$$

where λ is a weighting factor between [0,1] for regulating the weight of the focal loss value and the improved focal Tversky loss value during each loss function calculation.

3. Results and Analysis

3.1 Dataset

In this paper, we adopt BraTs2019 provided by the multi-modal brain tumor image segmentation challenge [28], [29]. The BraTs2019 dataset has 335 cases. Each case 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.

All data has been preprocessed including skull-stripping, image registration and spatial normalization by the challenge organizers, and image size is 240 × 240 × 155 voxels. For experiments, we divided the datasets into training and testing sets according to 4 : 1 using the five-fold cross-validation method.

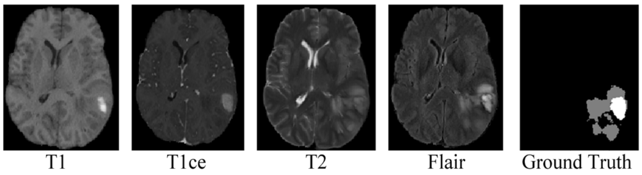

Fig. 7. Four modal MRI images and physician-labeled brain tumor image of a patient.

3.2 Evaluation Metrics

MRI images of four modalities were input as four channels in parallel, and the segmentation algorithm will segment ED, NC&NETC, ET and BG. According to clinical requirements, the performance of the algorithm is evaluated in terms of the accuracy of whole tumor (*WT* = *NCR*/*NET* + *ED* + *ET*), tumor core (*TC* = *NCR*/*NET* + *ET*) and enhancing tumor (*ET*). We employ three widely-used evaluation metrics including Dice, Jaccard, sensitivity, 95% Hausdorff distance to evaluate the model performance. Among these metrics, Dice and Jaccard are the most comprehensive evaluation metric. 95% Hausdorff distance metric measures the segmentation accuracy of the model on the tumor boundary. Sensitivity, also known as recall, shows the degree of influence of the loss function on the segmentation results. The expressions of the four metrics are shown in (12)–(15), respectively:

$$Dice(\mathbf{P},\mathbf{T}) = \frac{|\mathbf{P} \wedge \mathbf{T}|}{(|\mathbf{P}|+|\mathbf{T}|)/2}, \qquad (12)$$

$$Jaccard\left(\mathbf{P},\mathbf{T}\right) = \frac{\left|\mathbf{P} \wedge \mathbf{T}\right|}{\left|\mathbf{P} \vee \mathbf{T}\right|}, \qquad (13)$$

$$Sensitivity(\mathbf{P},\mathbf{T}) = \frac{|\mathbf{P} \wedge \mathbf{T}|}{|\mathbf{T}|}, \qquad (14)$$

$$\begin{aligned} &Hausdorff\left(\mathbf{P},\mathbf{T}\right) = \\ &\max\left\{\max_{a\in\mathbf{P}} \min_{b\in\mathbf{T}} d\left(a,b\right), \max_{b\in\mathbf{T}} \min_{a\in\mathbf{P}} d\left(a,b\right)\right\} \end{aligned} \qquad (15)$$

where **P** denotes the tumor and background area predicted by the model. **T** correspondingly is the real tumor and background area. $\wedge$ represents a logical 'and' operation, and |.| represents the size of sets. 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 Implementation Details

In this paper, the model is based on Pytorch and Python 3.7, and an Nvidia RTX 2080 Ti GPU is used to implement the algorithm for computation. The experiments are validated using the five-fold crossover method. The input image block size is 128 × 128 × 128 voxels, the network is optimized using the Adam optimizer, and the loss value is minimized using the hybrid focus loss function. In addition, the initial learning rate is 1e–04, the momentum is 0.9, the decay rate is 1e–05, the dropout is 0.5, the batch size is 2, and the number of epochs is 190. The detailed process of this algorithm is shown as follows. Firstly, the MRI images in the BraTs2019 are preprocessed with bias field correction and normalization. Secondly, TensorMixup algorithm is used for data augmentation. Then, the image blocks are taken for the training set. Next, the data are used to train and optimize RFE-UNet model. Finally, the segmentation results are evaluated, and the performance of the model using test data is given.

3.4 Parameter Analysis

The setting of λ in the proposed hybrid focal loss function is crucial, and various experimental attempts were made to determine the optimal λ value, and the specific results are shown in Tab. 1.

From the experimental results in Tab. 1, it can be seen that the two parts of the loss function have different values of λ for model optimization, and different values of λ parameters will determine the direction of model optimization when inputting clinical data for training the model. When λ is 0.5, the best experimental results are obtained for regions like TC and ET which are difficult to classify due to few samples, while the results of WT are only slightly decreased. Overall, the best results are obtained. So in the subsequent experiments, λ is set 0.5.

3.5 Comparison with the State-of-the-Arts

To demonstrate the effectiveness of the proposed algorithm, we first compared it with four state-of-the-art CNN methods on the BraTs2019 dataset including 3DUNet, ResUNet [30], UNet++ [31], nnUNet [32], [33], TransUNet [34]. All of these methods do the same data preprocessing and underlying data enhancement including translation, rotation, flipping, and mirroring as the proposed model when experimented. The experimental results are shown in Tab. 2.

As can be seen from Tab. 2, on the BraTs2019 dataset, the average Dice score of the proposed model in this paper are as high as 91.55%, 89.23%, and 84.16%, and the baseline model 3DUNet is improved by 1.76%, 4.1%, and 3.26%, which indicates that the improved network architecture can better improve the segmentation ability of the model for TC regions, and all of them are higher than the currently effective ResUNet, UNet++, nnUNet and TransUNet. In particular, a huge improvement is also achieved in the boundary-aware measure Hausdorff due to the action of RFE Module with expanded perceptual field and hybrid focal loss function, which verifies that the proposed method significantly outperforms other methods.

Figure 9 and Figure 10 show the segmentation results using 3DUNet and other models compared with the labeled images, and it can be found that since small convolution kernel cannot extract global features and has a small receptive field, missegmentation at details and over-segmentation at the edges occur, which turns normal brain areas into edematous regions. Although TransUNet sets the transformer module to expand the receptive field in the deep layer of the network, the detailed information of the deep feature map is lost in the down-sampling process. This leads to poor performance in detail reconstruction. However, 3DRFE-UNet has obtained the information of large receptive field by combining local and global features, and using the shallower feature map with rich details. The

Parameter	Mean Dice (%)			Mean Jaccard (%)			Mean Sensitivity (%)			Mean Hausdorff (mm)		
	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET
$\lambda = 0.1$	87.36	87.58	81.56	86.47	87.14	79.76	91.58	85.41	83.49	13.46	7.89	5.53
$\lambda = 0.3$	88.46	88.44	83.12	86.35	87.35	82.56	94.82	87.15	82.16	14.48	6.16	6.58
$\lambda = 0.5$	89.01	88.67	83.74	87.34	87.98	83.36	90.36	86.43	84.20	14.29	5.01	3.84
$\lambda = 0.7$	88.12	85.64	81.79	86.35	83.64	80.74	89.19	85.17	81.29	15.32	8.32	4.18
$\lambda = 0.9$	89.42	86.48	82.47	87.49	85.23	81.25	90.65	85.42	80.89	13.67	7.46	5.72

Tab. 1. Experimental results of different λ.

Models	Mean Dice (%)			Mean Jaccard (%)			Mean Sensitivity (%)			Mean Hausdorff (mm)			Params
	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	
3DUNet	89.79	85.13	80.90	87.42	84.39	77.07	94.08	87.46	80.77	13.61	7.47	5.45	16322979
ResUNet	89.93	83.55	78.40	87.37	81.19	76.48	93.12	85.25	79.35	12.35	11.20	7.53	17211624
UNet++	90.27	84.16	79.55	88.93	83.05	76.41	94.23	85.77	80.41	10.21	9.72	6.78	26989637
nnUNet	90.22	85.02	81.46	88.21	83.16	82.95	93.79	86.05	83.57	10.03	8.82	5.37	20716134
TransUNet	91.05	87.53	82.65	90.59	86.29	82.16	93.59	86.94	84.47	8.58	10.54	9.51	23418346
Proposed	91.55	89.23	84.16	90.64	87.42	82.03	92.04	86.89	86.91	10.11	4.57	3.21	19435421

Tab. 2. Comparison of segmentation results of different model on BraTs2019.

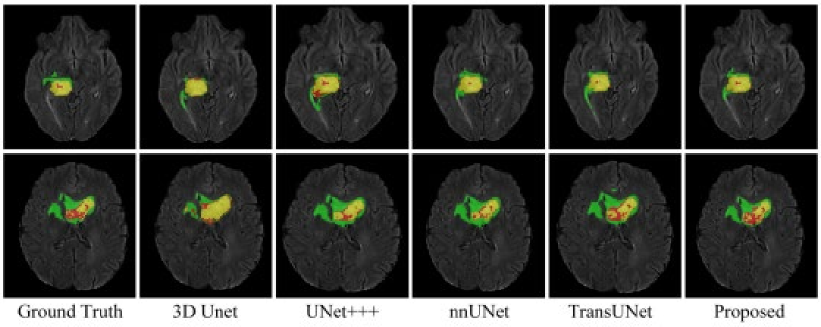

Fig. 9. 2D slices of segmentation results of different algorithms.

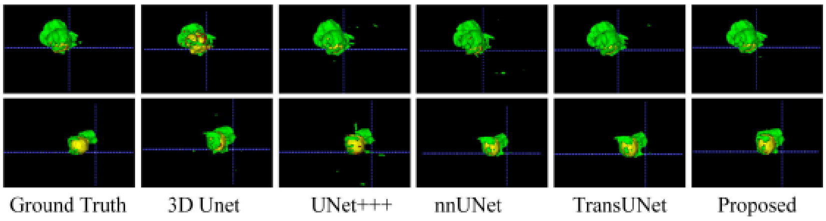

Fig. 10. 3D segmentation results of different algorithms.

Models	Mean Dice (%)			Mean Jaccard (%)			Mean Sensitivity (%)			Mean Hausdorff (mm)		
	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET
UNet (base)	89.79	85.13	80.90	87.15	83.84	81.84	94.08	85.46	80.77	13.61	7.47	5.45
UNet+TM	91.32	85.67	82.20	90.44	83.52	81.67	90.23	85.88	87.47	10.88	8.71	4.73
UNet+RFE	91.47	87.14	83.35	90.73	85.83	81.92	94.50	86.26	86.46	8.17	8.73	6.21
UNet+HFL	89.01	88.67	83.74	87.38	86.82	82.47	90.36	86.43	84.20	14.29	5.01	3.84
Proposed	91.55	89.23	84.16	89.34	88.38	83.76	92.04	86.89	86.91	10.11	4.57	3.21

Tab. 3. Performance comparison of different improvements.

segmentation result is more close to the ground truth, and the false positive area is very small. So the proposed method is more suitable for clinical use.

3.6 Ablation Study

3.6.1 Ablation of Different Modules

In order to demonstrate the effectiveness of adding modules in the proposed method, ablation experiments were conducted on the baseline model 3DUNet on the BraTs2019 dataset, and the effects of adding TensorMixup (TM), RFE Module (RFE), and hybrid focal loss function (HFL) respectively were tested, as shown in Tab. 3.

As can be seen from Tab. 3, after using the TensorMixup algorithm in the data enhancement part of the 3DUNet model, all of criteria were improved, especially the sensitivity in the ET region was improved by 6.7%, which shows that after intercepting and mixing the tumor images, the model can learn more tumor morphological features, and can improve the sensitivity and generalization of the model. After using RFE Module, the accuracy of model segmentation was significantly improved with the help of large receptive field feature maps due to the enlarged receptive field, and the Dice scores in the three regions were increased by 1.68%, 2.01%, and 2.45%, respectively. Hybrid focal loss function enabled the model to focus more on the indistinguishable TC and ET regions during training, which made mean Hausdorff was significantly reduced. And the model predicted the tumor margins more accurately. When the three modules are used simultaneously, the model can effectively combine the advantages of each module, and can achieve excellent results in most metrics, which indicates that the improvements of TensorMixup, RFE Module and hybrid focal loss function are very effective.

3.6.2 Ablation of RFE-Modules on Different Stages

The number of channels may change by the feature map of the RFE module, but it is the same as the input feature map in width and height dimensions. Therefore, various ways to set up the RFE module are used in the original network, such as placing the RFE module at any of the jump connections of the original network, or using the RFE module to connect the input and output of the network, or introducing the RFE module to multiple locations of the network. At the same time, the RFE module can be placed at any hop connection of the original network, or the RFE module can be used to connect the input and output of the network, or the RFE module can be introduced to multiple locations of the network simultaneously.

Considering that different setting methods will bring different brain tumor segmentation effects, four sets of RFE module methods are designed in this paper in 3D UNet. The networks corresponding to each of the four methods are evaluated by training and using BraTs2015 dataset, and the optimal method of RFE module is determined by comparing and analyzing the experimental results so that it can bring out the maximum effect. In this process, the hop connection in the first layer of the 3D UNet is noted as SK1 in this study in the order from top to bottom, and similarly the hop connections in the second, third and fourth layer are noted as SK2, SK3, SK4, etc. In turn, and the connection from the input flowing to the last concat operation in the decoding segment is noted as SK0. The models with the above four methods and the basic 3D UNet model were tested on the BraTs2019. The segmentation results on the test set are shown in Fig. 11, where SK12 represents the joint use of SK1 and SK2. And the other label is the similar meaning.

From the experimental results in Fig. 11, it can be seen that the network adding the RFE module at SK1-SK3 can fuse more comprehensive information of the large sensory field, and can enhance the model feature extraction ability. The 3D RFE-UNet model proposed by this paper achieves the optimal segmentation accuracy in the three regions because the RFE module is added at the appropriate position, and the average Dice coefficient value of the three regions reaches 87.9%, which exceeds the traditional 3D UNet model by 1.34%. These results indicate that the improved network architecture can better improve the segmentation ability of the model. In addition, Figure 11 also shows that the average Hausdorff value of the 3D RFE-UNet model is 8.71 mm in each of the three regions, which indicates that the segmentation accuracy of the 3D RFE-UNet model is also more accurate for the boundaries of the three regions.

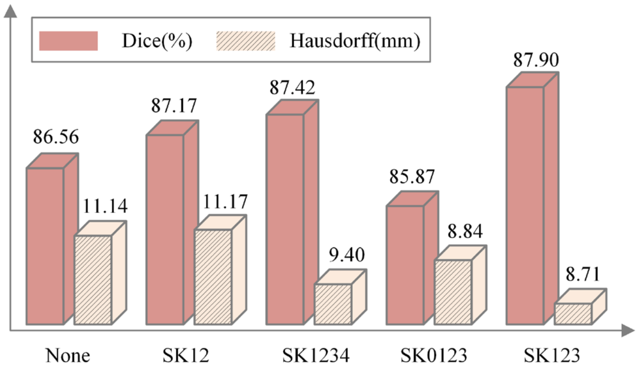

Fig. 11. Comparison of segmentation results of different model on BraTs2019.

3.6.3 Ablation of Different Loss Functions

To demonstrate the effectiveness of the loss function proposed in this paper, three comparative losses are used in this study for verifying the performance of the proposed algorithm, including cross-entropy loss, Dice loss, and focal loss with focal Tversky loss, and the objective evaluation results are shown in Tab. 4.

The results in Tab. 4 show that the use of different loss functions in the brain tumor segmentation model can produce different segmentation results. Among them, the improved focal Tversky loss outperforms the focal Tversky loss in all metrics. The model using the hybrid focal loss function has the best segmentation effect in TC and ET. Compared with the focal loss function alone, the hybrid focal loss function improves the mean values of Dice in TC and ET regions by 3.54% and 2.84%, respectively, and compared with the focal Tversky loss function, the hybrid focal loss function improves by 1.63% and 1.18%, respectively. In addition, the Hausdorff distance values of TC and ET segmentation by the model using the hybrid focus loss function were both the smallest, which shows that using the hybrid focus loss function for brain tumor segmentation is beneficial to improve the segmentation accuracy of the model for TC and ET.

However, compared with the focal loss and focal Tversky loss function, the mean Dice value of the model using the hybrid focal loss function in WT decreased by 0.78% and 0.34%, respectively, which indicates that the hybrid focal loss function is not advantageous for the segmentation of WT. The reason may be that the hybrid focus loss function makes the model focus on learning a smaller number of hard-to-classify samples, such as those of TC and ET, while the samples of WT are easier to identify compared to TC and ET. So the weight in the hybrid focus

Loss functions	Mean Dice (%)			Mean Jaccard (%)			Mean Sensitivity (%)			Mean Hausdorff (mm)		
	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET
Cross-entropy loss	83.24	76.03	72.45	82.41	75.67	71.15	83.67	82.37	76.29	25.87	30.61	21.67
Dice loss	84.57	78.34	73.52	82.19	76.38	72.39	82.54	80.41	75.28	24.39	28.64	20.17
Focal loss	89.79	85.13	80.90	87.73	84.26	78.14	94.08	87.46	80.77	13.61	7.47	5.45
Focal Tversky loss	88.37	86.71	82.37	87.47	85.14	81.49	91.49	85.45	81.16	15.72	7.29	6.78
Proposed	89.01	88.67	83.74	87.29	87.24	82.36	90.36	87.62	84.20	14.29	5.01	3.84

Tab. 4. Experimental results of different loss functions.

loss value will be reduced, making the model's segmentation ability for WT not significantly improved. However, overall, the model using the hybrid focus loss function obtains excellent results in the brain tumor segmentation task, especially in the hard-to-classify core tumor region and enhanced tumor region.

4. Discussion

In the field of medical image segmentation, 3D UNet is a commonly used baseline network that can directly process three-dimensional MRI images. However, this method faces problems of insufficient data volume and limited receptive field. In this paper, a new method which extends the dataset using the TensorMixup algorithm is proposed, and the RFE module is designed to expand the receptive field. Experimental results show that this method improves the mean accuracy of scores, reduces variance in all three lesion areas, and makes segmentation results more stable.

Although some achievements have been obtained, this study still has some limitations. Due to the wide range of medical image sources, different medical institutions and imaging equipment may result in differences on image quality and data distribution. These differences may lead to the model learning biased features, further affecting the accuracy of segmentation. Therefore, when applying the new method to clinical practice, optimization of training data should be conducted to ensure that clinical practice data has the same distribution.

In the future, we hope to collaborate with doctors to further optimize the dataset and post processing stage. We believe that through in-depth exploration and cooperation, our new method can play an effective role in clinical practice, and contribute to the development of medical image analysis.

5. Conclusion

In this paper, a new MRI image segmentation framework for brain tumors is proposed which uses the TensorMixup data enhancement algorithm to provide more high-quality training data, a RFE-UNet network to efficiently fuse local and global features, and a fusion loss function to alleviate the data imbalance phenomenon. The experimental results show that both the proposed RFE-UNet and the hybrid focal loss function have a great effect on the segmentation results in different aspects, and the whole architecture can effectively solve the problems of low data volume, restricted perceptual field, and data imbalance in the field of brain tumor segmentation. The test results in BraTs2019 dataset show that the average Dice values of the proposed algorithm can reach 91.55%, 89.23% and 84.16% in the WT, TC and ET regions, respectively, which proves the feasibility and effectiveness of using the proposed framework.

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] LUNDERVOLD, A. S., LUNDERVOLD, A. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift Für Medizinische Physik*, 2019, vol. 29, no. 2, p. 102–127. DOI: 10.1016/j.zemedi.2018.11.002

[3] PICCIALLI, F., DI SOMMA, V., GIAMPAOLO, F., et al. A survey on deep learning in medicine: Why, how and when? *Information Fusion*, 2021, vol. 66, no. 1, p. 111–137. DOI: 10.1016/j.inffus.2020.09.006

[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 the International Conference on Medical Image Understanding and Analysis*. Cham (Switzerland), 2017, p. 506–517. DOI: 10.1007/978-3-319-60964-5_44

[5] MYRONENKO, A. 3D MRI brain tumor segmentation using autoencoder regularization. In *Proceedings of the International Conference on Brainlesion-Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Berlin (German), 2018, p. 311–320. DOI: 10.1007/978-3-030-11726-9_28

[6] DOSOVITSKIY, A., BEYER, L., KOLESNIKOV, A., et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. In *Proceedings of the International Conference on Learning Representations (ICLR).* Vienna (Austria), 2021, p. 1–22. DOI: 10.48550/arXiv.2010.11929

[7] MAGADZA, T., VIRIRI, S. Deep learning for brain tumor segmentation: A survey of state-of-the-art. *Journal of Imaging*, 2021, vol. 7, no. 2, p. 1–22. DOI: 10.3390/jimaging7020019

[8] LIU, Z., LIN, Y., CAO, Y., et al. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the International Conference on Computer Vision (ICCV).* Montreal (Canada), 2021, p. 10012–10022. DOI: 10.1109/ICCV48922.2021.00986

[9] MORENO LOPEZ, M., VENTURA, J. Dilated convolutions for brain tumor segmentation in MRI scans. In *Proceedings of the Conference on Medical Image Computing for Computer Assisted Intervention (MICCAI).* Granada (Spain), 2018, p. 253–262. DOI: 10.1007/978-3-319-75238-9_22

[10] DING, Y., LI, C., YANG, Q. Q., et al. How to improve the deep residual network to segment multi-modal brain tumor images. *IEEE Access*, 2019, vol. 7, no. 1, p. 152821–152831. DOI: 10.1109/ACCESS.2019.2948120

[11] BALA, S. A., KANT, S. Dense dilated inception network for medical image segmentation. *International Journal of Advanced Computer Science and Applications*, 2020, vol. 11, no. 11, p. 785–793. DOI: 10.14569/IJACSA.2020.0111195

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

[13] MCKINLEY, R., MEIER, R., WIEST, R. Ensembles of densely-connected CNNs with label-uncertainty for brain tumor segmentation. In *Proceedings of the Conference on Medical Image Computing for Computer Assisted Intervention (MICCAI)*. Granada (Spain), 2018, p. 456–465. DOI: 10.1007/978-3-030-11726-9_40

[14] ABRAHAM, N., KHAN, N. M. A novel focal Tversky loss function with improved attention U-Net for lesion segmentation. In *Proceedings of the 16th International Symposium on Biomedical Imaging (ISBI)*. Venice (Italy), 2019, p. 683–687. DOI: 10.48550/arXiv.1810.07842

[15] TAGHANAKI, S. A., ZHENG, Y., ZHOU, S. K., et al. Combo loss: Handling input and output imbalance in multiorgan segmentation. *Computerized Medical Imaging and Graphics*, 2019, vol. 75, no. 1, p. 24–33. DOI: 10.1016/j.compmedimag.2019.04.005

[16] WANG, G. T., LI, W., OURSELIN, S., et al. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In *Proceedings of the Conference on Medical Image Computing for Computer Assisted Intervention (MICCAI)*. Quebec (Canada), 2017, p. 178–190. DOI: 10.1007/978-3-319-75238-9_16

[17] YEUNG, M., SALA, E., SCHÖNLIEB, C. B., et al. Focus U-Net: A novel dual attention-gated CNN for polyp segmentation during colonoscopy. *Computers in Biology and Medicine*, 2021, vol. 137, no. 1, p. 1–11. DOI: 10.1016/j.compbiomed.2021.104815

[18] YEUNG, M., SALA, E., SCHÖNLIEB, C. B., et al. Unified focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics*, 2022, vol. 95, no. 1, p. 1–13. DOI: 10.1016/j.compmedimag.2021.102026

[19] GARCEA, F., SERRA, A., LAMBERTI, F., et al. Data augmentation for medical imaging: A systematic literature review. *Computers in Biology and Medicine*, 2023, vol. 152, no. 1, p. 1–20. DOI: 10.1016/j.compbiomed.2022.106391

[20] MOK, T. C., CHUNG, A. C. Learning data augmentation for brain tumor segmentation with coarse-to-fine generative adversarial networks. In *Proceedings of the Conference on Medical Image Computing for Computer Assisted Intervention (MICCAI)*. Granada (Spain), 2018, p. 70–80. DOI: 10.1007/978-3-030-11723-8_7

[21] NALEPA, J., MARCINKIEWICZ, M., KAWULOK, M. Data augmentation for brain-tumor segmentation: A review. *Frontiers in Computational Neuroscience*, 2019, vol. 13, no. 1, p. 1–18. DOI: 10.3389/fncom.2019.00083

[22] DVORNIK, N., MAIRAL, J., SCHMID, C. On the importance of visual context for data augmentation in scene understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, vol. 43, no. 6, p. 2014–2028. DOI: 10.1109/TPAMI.2019.2961896

[23] ZHANG, H. Y., CISSE, M., DAUPHIN, Y. N., et al. Mixup: Beyond empirical risk minimization. In *Proceedings of the International Conference on Learning Representations (ICLR)*. Vancouver (Canada), 2018, p. 1–13. DOI: 10.48550/arXiv.1710.09412

[24] WANG, Y., JI, Y. R., XIAO, H. B. A data augmentation method for fully automatic brain tumor segmentation. *Computers in Biology and Medicine*, 2022, vol. 149, p. 1–10. DOI: 10.1016/j.compbiomed.2022.106039

[25] RICKMANN, A. M., ROY, A. G., SARASUA, I., et al. Project & excite' modules for segmentation of volumetric medical scans. In *Proceedings of the Conference on Medical Image Computing for Computer Assisted Intervention (MICCAI)*. Shenzhen (China), 2019, p. 39–47. DOI: 10.1007/978-3-030-32245-8_5

[26] ALJABRI, M., ALGHAMDI, M. A review on the use of deep learning for medical images segmentation. *Neurocomputing*, 2022, vol. 506, p. 311–335. DOI: 10.1016/j.neucom.2022.07.070

[27] LIN, T. Y., GOYAL, P., GIRSHICK, R., et al. Focal loss for dense object detection. In *Proceedings of the International Conference on Computer Vision (ICCV)*. Venice (Italy), 2017, p. 2980–2988. DOI: 10.48550/arXiv.1708.02002

[28] 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

[29] 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

[30] 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, p. 94–114. DOI: 10.1016/j.isprsjprs.2020.01.013

[31] 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*, 2020, vol. 9, no. 6, p. 1856–1867. DOI: 10.1109/TMI.2019.2959609

[32] ISENSEE, F., JÄGER, P. F., KOHL, S. A. A., et al. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 2021, vol. 18, no. 2, p. 203 to 211. DOI: 10.1038/s41592-020-01008-z

[33] ISENSEE, F., JÄGER, P. F., FULL, P. M., et al. 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: 10.1007/978-3-030-72087-2_11

[34] CHEN, J., LU, Y., YU, Q., et al. TransUnet: Transformers make strong encoders for medical image segmentation. In *Proceedings of the International Conference on Machine Learning (ICML)*. Vienna (Austria), 2021, p. 1–13. DOI: 10.48550/arXiv.2102.04306

About the Authors ...

Yu WANG 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.

Yarong JI was born in 1997. She is now a candidate of master degree in the School of Artificial Intelligence, Beijing

Technology and Business University, China. Her research interests include pattern recognition, image processing 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 Beijing Technology and Business University. His research interests include image processing and the modeling and analysis of complex system.