

## APPLIED RESEARCH

# SEResU-Net for Multimodal Brain Tumor Segmentation

CHENGDONG YAN<sup>1,2</sup>, JURONG DING<sup>1,2</sup>, HUI ZHANG<sup>1,2</sup>,  
KE TONG<sup>1,2</sup>, BO HUA<sup>1,2</sup>, AND SHAOLONG SHI<sup>3,4,5</sup>, (Member, IEEE)

<sup>1</sup>School of Automation and Information Engineering, Sichuan University of Science and Engineering, Zigong 643099, China

<sup>2</sup>Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Zigong 643009, China

<sup>3</sup>Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou 324000, China

<sup>4</sup>School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610056, China

<sup>5</sup>Department of Computer Science, City University of Hong Kong, Hong Kong

Corresponding authors: Jurong Ding (jurongding@gmail.com) and Shaolong Shi (shi\_shaolong@163.com)


This work was supported in part by the National Natural Science Foundation of China under Grant 81401482 and Grant 62102071, in part by the Scientific Research Foundation of the Sichuan University of Science and Engineering under Grant 2020RC31, in part the Research Foundation of Artificial Intelligence Key Laboratory of Sichuan Province under Grant 2021RYJ05, and in part by the Hong Kong Scholars Program under Grant XJ2021007.

**ABSTRACT** Glioma is the most common type of brain tumor, and it has a high mortality rate. Accurate tumor segmentation based on magnetic resonance imaging (MRI) is of great significance for the diagnosis and treatment of brain tumors. Recently, the automatic segmentation of brain tumors based on U-Net has gained considerable attention. However, brain tumor segmentation is a challenging task due to the structural variations and inhomogeneous intensity of tumors. Existing brain tumor segmentation studies have shown that the problems of insufficient down-sampling feature extraction and loss of up-sampling information arise when using U-Net to segment brain tumors. In this study, we proposed an improved U-Net model, SEResU-Net, which combines the deep residual network and the Squeeze-and-Excitation Network. The deep residual network solves the problem of network degradation so that SEResU-Net can extract more feature information. The Squeeze-and-Excitation Network avoids information loss and enables the network to focus on the useful feature map, which solves the problem of insufficient segmentation accuracy of small-scale brain tumors. Furthermore, a fusion loss function combining Dice loss and cross-entropy loss was proposed to solve the problems of network convergence and data imbalance. The performance of SEResU-Net was evaluated on the dataset of BraTS2018 and BraTS2019. Experimental results revealed that the mean Dice similarity coefficients of SEResU-Net were 0.9373, 0.9108, and 0.8758 for the whole tumor, the tumor core, and the enhanced tumor, which were 7.10%, 11.88%, and 15.33% greater than those of the U-Net benchmark network, respectively. Our findings demonstrate that the proposed SEResU-Net has a competitive effect in segmenting multimodal brain tumors.

**INDEX TERMS** MRI, brain tumor segmentation, deep learning, U-Net, residual network, squeeze-and-excitation network.

## I. INTRODUCTION

Brain tumors, caused by the abnormal growth of brain tissue cells, seriously affect human health [1], [2]. Gliomas are the most common brain tumors and can be classified as low-grade gliomas (LGG) or high-grade gliomas (HGG). Mag-

The associate editor coordinating the review of this manuscript and approving it for publication was Marco Giannelli .

netic resonance imaging (MRI), a typical non-invasive imaging technology, offers high resolution, does not generate skull artifacts, and provides valuable information about anatomical structure. It has become the primary screening method for the diagnosis of brain tumors [3]. Accurately identifying and segmenting brain tumors using multimodal MRI can provide quantitative information, such as tumor volume and maximum diameter, which helps surgeons establish optimal

treatment plans for individual patients. Therefore, accurate segmentation of brain tumors is a key step in the diagnosis and treatment of brain tumors [4].

Traditionally, the brain tumor contour is manually delineated by an experienced radiologist, but this process is time-consuming, involves strong subjectivity, and is likely to result in segmentation errors [5]. Therefore, research on semiautomatic or automatic brain tumor segmentation is valuable for the quantitative analysis and evaluation of brain tumors and normal tissues [6].

With the rapid development of deep learning, from CNN [7] to FCN [8], and then to U-Net [9], ResNet [10], SegNet [11], and DenseNet [12], convolutional neural networks (CNNs) have been successfully utilized in a variety of computer vision tasks and have received extensive attention from academia and industry. Due to the strong feature extraction ability of deep neural networks, they have been rapidly applied to the field of complex medical image processing and analysis [13], [14]. Brain tumor segmentation by multi-modality MRI based on deep learning has also attracted extensive attention. Among them, U-Net, which has a symmetric structure and integrates high-level feature information and low-level feature information through a jump connection, offers outstanding segmentation performance and has quickly become a commonly used benchmark in medical image segmentation [15], [16]. However, during the process of down-sampling, U-Net constantly reduces the dimension of the image, which will result in inadequate segmentation accuracy for small-scale tumors. Meanwhile, when up-sampling and a simple jump connection are used for feature fusion, spatial information and location details of the high-level output maps tend to get lost in the procedures of continuously cascaded convolutions and non-linearities transformations, which seriously reduces the resolution of the feature map. The residual network can prevent network degradation, effectively increase the depth of the network and improve the ability to extract fine feature. Attention mechanism can highlight useful feature information, enhance local feature expression, and suppress redundant information, which can solve the problem of insufficient segmentation accuracy of small-scale brain tumors. Adding these two modules on U-Net may solve the problems during the down-sampling and up-sampling processes.

Therefore, in this study we proposed the SEResU-Net model, which combines ResNet50 in the down-sampling process and the attention module of the Squeeze-and-Excitation Network (SENet) [17] in the up-sampling process of U-Net. Theoretically, SEResU-Net should be superior to U-Net in segmenting different brain tumor subregions (i.e., edema, necrosis, and enhancing and non-enhancing tumor cores) in multimodal MRI. We evaluated the effectiveness of the model on the BraTS2019 dataset. This study's contribution is threefold: (1) an end-to-end SEResU-Net model for multimodal MRI brain tumor segmentation is proposed, which can not only extract more abundant semantic information but also pay attention to small-scale brain tumor information

and improve the segmentation accuracy; (2) down-sampling combined with a ResNet50 depth residual network is used to extract deeper fine information and eliminate the problem of vanishing and exploding gradient; up-sampling combined with SENet channel attention module is used to enhance local feature information expression, strengthen the segmentation of fine information, and improve the accuracy; and (3) the mixed loss function of Dice loss and cross-entropy loss is used to fully suppress the impact of class imbalance on brain tumor segmentation.

The rest of this paper is organized as follows: Section II briefly introduces the related work of brain tumor segmentation. In section III, the principle of the method proposed in this paper is introduced in detail. In section IV, the specific experiments and analyze the results are reported. Finally, the conclusions are presented in section V.

## II. RELATED WORK

In previous studies, to future improve the performance of U-Net, researchers have conducted the following research on U-Net, such as by adding dense connection [18], residual network [19], generative adversarial network (GAN) [20], variational automatic encoder (VAE) [21], or attention mechanism [22], [23], [24], to solve the problems of insufficient feature extraction in the down-sampling process or neglect of small-scale tumors and information loss in the up-sampling process. Increasing the number of network layers is a popular method for enhancing the performance of the network, but the deeper the network layer, the more serious the disappearance gradient. To solve the problem of vanishing gradient, the residual model was proposed [10]. Yang et al. [19] combined U-Net and residual network and proposed Deeper ResU-Net to enhance the feature extraction ability. Furthermore, many attempts have been made to embed the attention module into the deep neural network architecture to enhance the local response and improve the effectiveness of feature extraction and restoration. Zhang et al. proposed AResUNet [25] by adding a series of attention units among corresponding down-sampling and up-sampling processes. Their model adaptively rescales features to effectively enhance local responses of down-sampling residual features, which are utilized for the feature recovery of the following up-sampling process. However, the problems with the down-sampling and up-sampling processes cannot be well solved simultaneously in these models.

Among the state-of-the-art methods, researchers have proposed different methods to effectively improve the segmentation performance of brain tumors. Since U-Net cannot capture long-distance dependence, Gan et al. [26] proposed a global attention mechanism to capture long-distance dependencies and solve the problem that convolution operations can only extract local information. Aboelenein et al [27] proposed a hybrid two-track U-Net (HTTU-Net), in which the first track focuses on the shape and size of the tumor, and the second track captures contextual information. HTTU-Net can extract more semantic information and consider more

information of small-scale brain tumors. Zhou et al. [28] proposed an efficient encoder-decoder architecture for brain tumor segmentation, which uses a lightweight neural network ShuffleNetV2 as the encoder to reduce the number of parameters and obtain a large receptive field. The decoder introduces residual blocks to avoid degradation problems. To improve the ability of the neural network to extract and utilize multiscale image features, Wang et al. [29] proposed a spatial dilated feature pyramid (DFP) module. Most models cannot make full use of the global context information, so Chen et al. [30] presented a two-stage automated brain lesion segmentation framework by integrating cascaded RF and dense CRF, which can effectively integrate the local appearance and global contextual information of multimodal MRI and iteratively improve the segmentation results. To further recover the details of brain tumors and improve the brain tumor segmentation performance, Huang et al. [31] proposed a group cross-channel attention residual U-Net that can make full use of the low-level fine details of tumor regions.

### III. METHOD PRINCIPLE

#### A. RESIDUAL NETWORK

Through entries including AlexNet’s 8-layer neural network [32], VGG’s 19-layer neural network [33], and GoogLeNet’s 22-layer neural network [34], the ImageNet LSVR competition showed that the greater the network depth, the stronger the ability to extract image features. However, if the network depth is continuously increased, once the network depth reaches a certain degree, the network will exhibit gradient dispersion and other problems, and the continuous increase will lead to network degradation and decrease the accuracy of the model. To solve the problem of network degradation and speed up training, He et al. proposed the residual network. When the network has reached the saturated accuracy, an identity shortcut connection, i.e.,  $y = x$ , is added behind it so that the error will not increase even if the depth of the network is increased. As shown in Fig. 1(a), assuming that the input is  $x$  and the expected output is  $H(x)$ , then an identity mapping is added through the jump connection to directly transfer  $x$  to the output as the initial result  $H(x) = F(x) + x$ . Now, what needs to be learned is no longer  $H(x)$  but the difference between  $H(x)$  and  $x$ , namely, the residual:  $F(x) = H(x) - x$ . Therefore, the subsequent training is focused on getting the residual result to approach zero and deepening the network without reducing the accuracy.

To build a deeper network structure, He et al. [10] also proposed a “bottleneck” structure. They reported that the “bottleneck” building block of ResNet50/101/152 reduced the amount of calculation by 16.94 times compared with the building block for ResNet34 (on  $56 \times 56$  feature maps). The bottleneck structure is shown in Fig. 1(b). When input channels = 256, the first conv  $1 \times 1$  can reduce the input dimension to 64, conv  $3 \times 3$  maintains the current channels, and finally, conv  $1 \times 1$  reverts to the original channels.

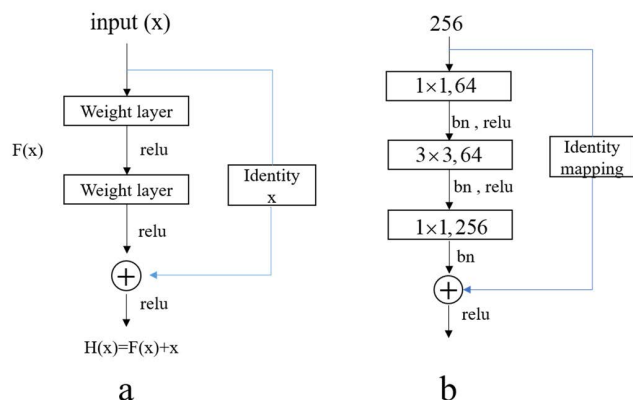


FIGURE 1. Diagrams of the residual network (a), and the bottleneck structure (b).

Thus, using the “bottleneck” structure can greatly reduce the amount of parameter calculation and raise the training speed.

#### B. SQUEEZE-AND-EXCITATION NETWORK (SENet)

Learning from the selective cognitive mechanism of human beings, attention mechanisms can effectively identify and highlight useful information and suppress redundant information. In recent years, they have been extensively applied in the fields of image classification [35], image detection and recognition [36], and image segmentation [37]. To enhance the feature extraction and expression of the network, our experimental method combines the SENet proposed by Hu et al. [17]. Unlike the spatial attention mechanism, to improve spatial coding ability, SENet focuses on exploring the attention on the channel. SENet can learn the feature weight through the network loss and obtain the important degree of each feature map. According to the importance degree, a weight value is assigned to each feature channel so that the neural network can focus on the useful feature map. SENet effectively deals with the information loss caused by the different importance of different channels of feature maps in the process of convolution and pooling. Fig. 2 shows the principal structure of SENet: input  $X$  to obtain a feature map  $U$  after convolution, and squeeze  $U$ . The formula is as follows:

$$Z_c = F_{sq}(U_c) = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H U_c(i, j). \quad (1)$$

Next, compress the spatial information of each channel to a single value of  $1 \times 1 \times C_2$  and perform exception to obtain a channel attention vector  $S$  of  $1 \times 1 \times C_2$  after weight adjustment. The formula is as follows:

$$S = F_{ex}(Z, W) = \sigma(W_2 \delta(W_1 Z)), \quad (2)$$

where  $\delta$  represents the ReLU activation function,  $\sigma$  represents a sigmoid function, and  $W_1$  and  $W_2$  represent two fully connected layers. Finally,  $S$  is used to recalibrate  $U$ , and feature

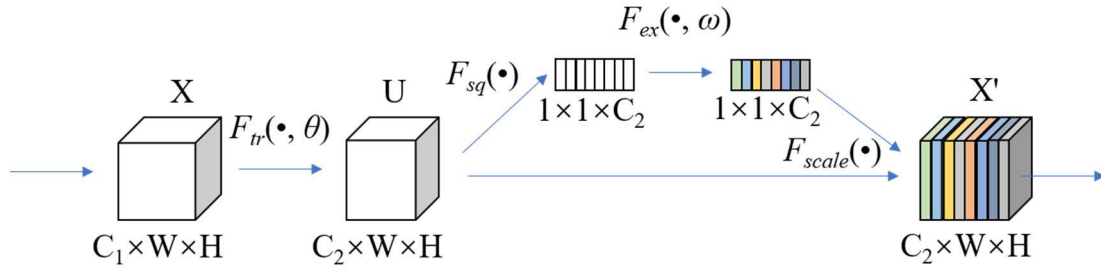


FIGURE 2. Schematic diagram of Squeeze-and-Exception network.

maps with different channel importance are obtained. The formula is as follows:

$$X' = F_{\text{scale}}(U_c, S_c) = S_c \cdot U_c. \quad (3)$$

### C. BATCH NORMALIZATION

Internal covariate shift often appears during network training. When the parameters in the underlying network have a slight change, due to the linear transformation and nonlinear activation mapping of each layer, these small changes will be amplified with the increase of the number of layers in the network. The changed parameters will change the distribution of input data, and further lead to the need for the upper network to constantly adapt to these changes, which will increase the difficulty of model training. Therefore, batch normalization is added after the convolution layer, which makes the input data distribution relatively stable and accelerates the model training speed. The sensitivity of the model to network parameters is simplified and the network learning is more stable. Batch normalization has a regularization effect and relieves the problem of gradient disappearance.

### D. PIXELSHUFFLE

PixelShuffle [38] is an up-sampling method proposed to solve the super-resolution of the image, which fills pixels with the information of channel dimension and can effectively enlarge the reduced feature image. PixelShuffle can solve the problem of checkerboard artifacts and better segment the boundary of brain tumors. As shown in Fig. 3, the main function is to change a low-resolution feature map of  $H \times W$  into a high-resolution feature map of  $rH \times rW$  through the sub-pixel convolution operation. The principle is as follows:

$$N \times (C \times r \times r) \times W \times H \rightarrow N \times C \times (r \times H) \times (r \times W).$$

### E. SEResU-Net

Considering the advantages of deep residual network and SENet channel attention module, as well as the complexity of multimodal MRI brain tumor images, the network structure of our proposed method is shown in Fig. 4. First, the coding part performs down-sampling combined with the ResNet50 model. Testing reveals a three-layer encoder has a better effect than a four-layer encoder. Therefore, both up-sampling

and down-sampling are applied four times. Encoder 1 contains 3 bottlenecks, encoder 2 contains 4 bottlenecks, and encoder 3 contains 6 bottlenecks. SEResU-Net further deepens the network and adds more jump connections than the original U-Net. Therefore, SEResU-Net has better initial feature extraction and expression ability, and it can also better combine the background semantic information of the image for multi-scale segmentation, which will effectively reduce the segmentation error during down-sampling of small-scale brain tumors.

The decoder part adds the attention mechanism of the SENet channel. Each decoder module consists of a conv  $3 \times 3$  with strides of 1 and an up-sampling layer with PixelShuffle. The feature map after skip connection feature fusion is input into the SENet module through the ReLU activation function to obtain the weighted feature map and finally achieve the purpose of improving the segmentation accuracy. Compared with the U-Net decoding part, the addition of SENet can enhance the meaningful feature information, suppress the feature response of irrelevant regions, and reduce the number of redundant features.

### F. COMBINED LOSS FUNCTION

The MRI brain tumor segmentation task exhibits severe class imbalance. To provide better supervision for model training, we utilized the mixed loss function combining Dice loss [39] and cross-entropy loss. The Dice coefficient is a set similarity measurement function, which is typically used to calculate the similarity between two samples, and the value range is [0, 1]. The expression is as follows:

$$S = \frac{2|X \cap Y|}{|X| + |Y|}. \quad (4)$$

The Dice loss expression is as follows:

$$d = 1 - \frac{2|X \cap Y|}{|X| + |Y|}. \quad (5)$$

Cross-entropy loss [40] is mainly used to determine how close the actual output is to the expected output, and the difference is used to update network parameters through reverse



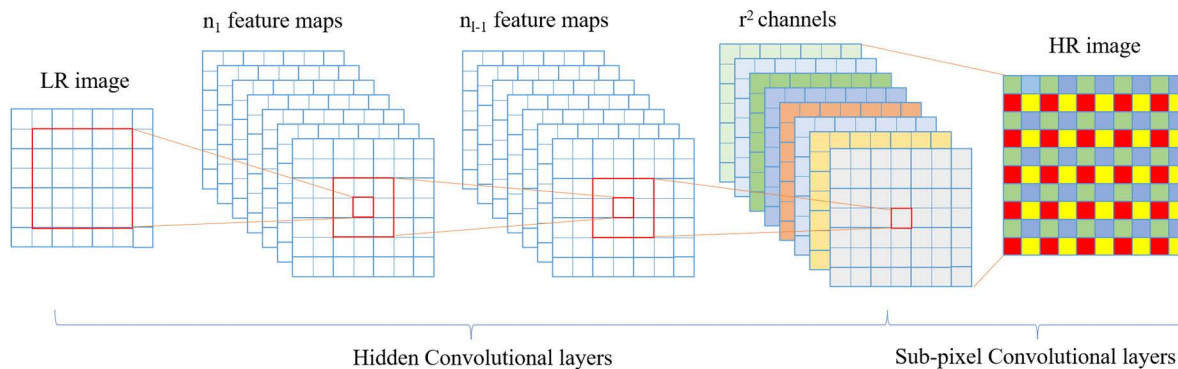


FIGURE 3. Implementation flow of PixelShuffle algorithm.

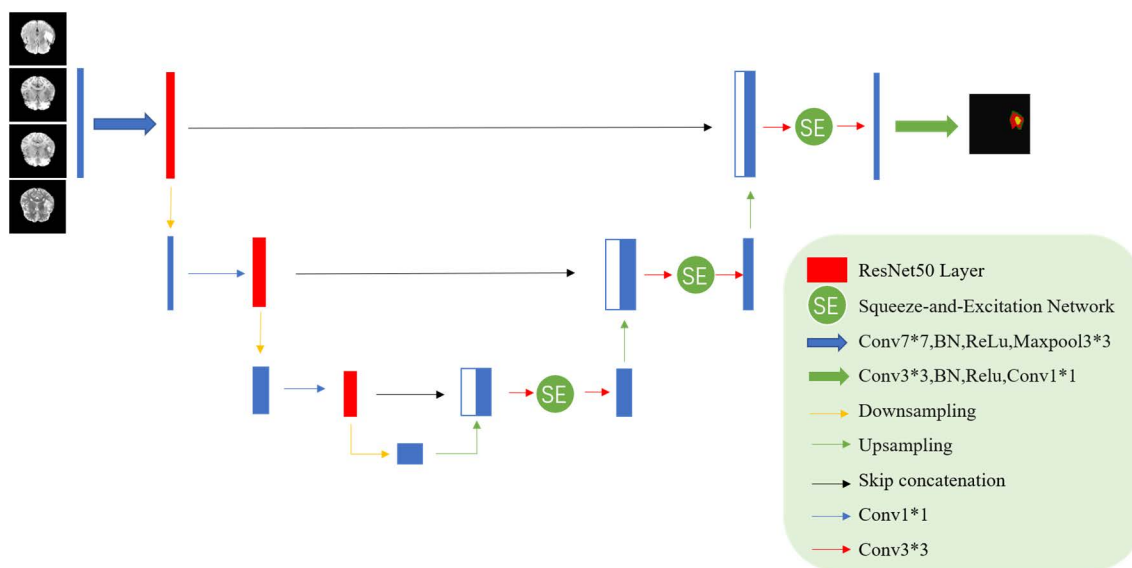


FIGURE 4. Structure diagram of SEResU-Net. SEResU-Net integrates residual modules and Squeeze-and-Excitation Network (SENet) with a primeval and single U-Net architecture, in which ResNet50 is combined with the down-sampling process to focus on small-scale tumors and strengthen the segmentation of small information. SENet is added into the up-sampling process to highlight salient feature information and disambiguate irrelevant and noisy feature responses.

propagation. The expression is as follows:

$$H(p, q) = - \sum_{i=1}^n p(x_i) \log(q(x_i)), \quad (6)$$

where  $p$  presents the expected value, and  $q$  presents the predicted value.

The combined loss function takes into account the advantages of both Dice loss and cross-entropy loss, and it can make a consideration globally and microscopically. When the foreground and background are unbalanced and the segmentation content is unbalanced, the network can still learn well.

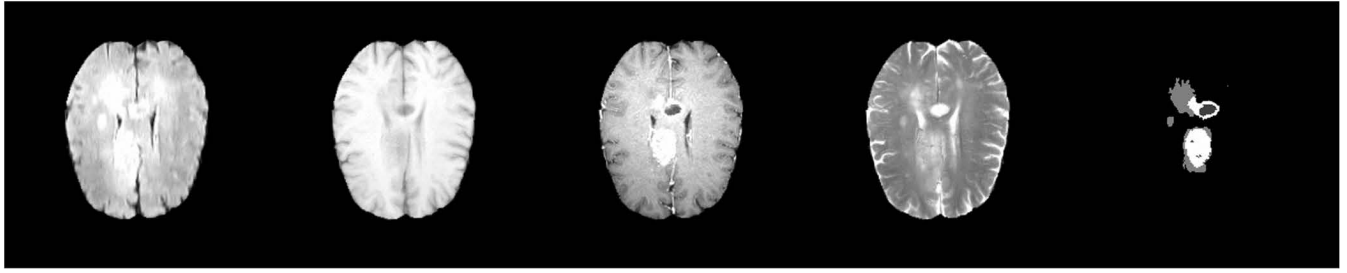
#### IV. EXPERIMENT AND RESULT ANALYSIS

##### A. DATASET AND DATA PREPROCESSING

###### 1) DATASET

The International Association for Medical Image Computing and Computer Assisted Intervention (MICCAI) has held the

Multimodal Brain Tumor Segmentation Challenge (BraTS) since 2012, which has greatly promoted the development of brain tumor segmentation methods based on deep learning. Since then, the BraTS dataset has become an authoritative dataset for evaluating MRI brain tumor segmentation methods. We obtained the data for this study from the BraTS2018 and BraTS2019 public datasets provided by BraTS [4], [41], [42]. The BraTS2018 dataset contains a training dataset of 210 HGG cases and 75 LGG cases. The BraTS2019 database adds 49 HGG cases and 1 LGG case to BraTS2018. Each patient has four MRI modalities (FLAIR, T1, T1ce, and T2), and the size of each MRI image in the dataset is  $240 \times 240 \times 155$ . Brain tumor labels are divided into three classes: necrotic and non-enhancing tumor core (NET, label1), peritumoral edema (ED, label2), and enhanced tumor (ET, label4). Ground truth (GT) is manual segmentation of brain tumors by experienced experts. To better evaluate the segmentation effect, it is necessary to segment whole tumor (WT, NET +



**FIGURE 5.** Example of the brain MRI data from a patient in the BraTS2019 dataset. From left to right: FLAIR modality, T1 modality, T1ce modality, T2 modality and the ground truth.

**TABLE 1.** Tumor segmentation sub-region and label description.

Tumor sub-region	Description	Label
Whole tumor	NET, ED, ET	1, 2, 4
Core tumor	NET, ET	1, 4
Enhancing tumor	ET	4

ED + ET), tumor core (TC, NET + ET), and enhanced tumor (ET) (see Table 1). Fig. 5 illuminates a typical case of MRI brain image and GT.

## 2) DATA PREPROCESSING

Due to contrast variations, uneven intensity, and noise effects, MRI brain tumor segmentation is a challenging task. Although deep learning-based methods are robust to noise, data processing is still a critical and essential step. First, data standardization was performed, that is, the image was normalized by using the average and standard values of the non-zero region. Then the image size was cropped to  $160 \times 160$  and enhanced the image contrast and brightness. Finally, the sections without lesions were removed. We processed the BraTS2018 dataset (210 HGG cases and 75 LGG cases) to generate a training set, and we processed the newly added BraTS2019 dataset (49 HGG cases and 1 LGG case) to generate a test set.

## B. EVALUATION METRICS

To evaluate the experimental results, this study used the same evaluation metrics as the official BraTS website: Dice similarity coefficient (DSC), sensitivity, specificity, and Hausdorff distance (HD), which are the four most authoritative evaluation metrics for brain tumor segmentation. The calculation formulas of the evaluation metrics are as follows (note: TP represents the area correctly detected as a positive sample, FP represents the area incorrectly detected as a positive sample, and FN represents the area incorrectly detected as a negative sample):

$$DSC = \frac{2TP}{FP + 2TP + FN}. \quad (7)$$

DSC is used to evaluate the similarity between segmentation prediction and ground truth.

$$Sensitivity = \frac{TP}{TP + FN}. \quad (8)$$

Sensitivity is also called the true positive rate and the recall rate. It measures the positive voxel part in the real background, that is, it measures the ability to segment the region of interest in the segmentation experiment.

$$Specificity = \frac{TP}{TP + FP}. \quad (9)$$

Specificity is also called the true negative rate, and it measures the negative voxel part in the ground truth segmentation, that is, it measures the ability to correctly judge whether the pixels in the region of interest are not pixels in the segmentation experiment.

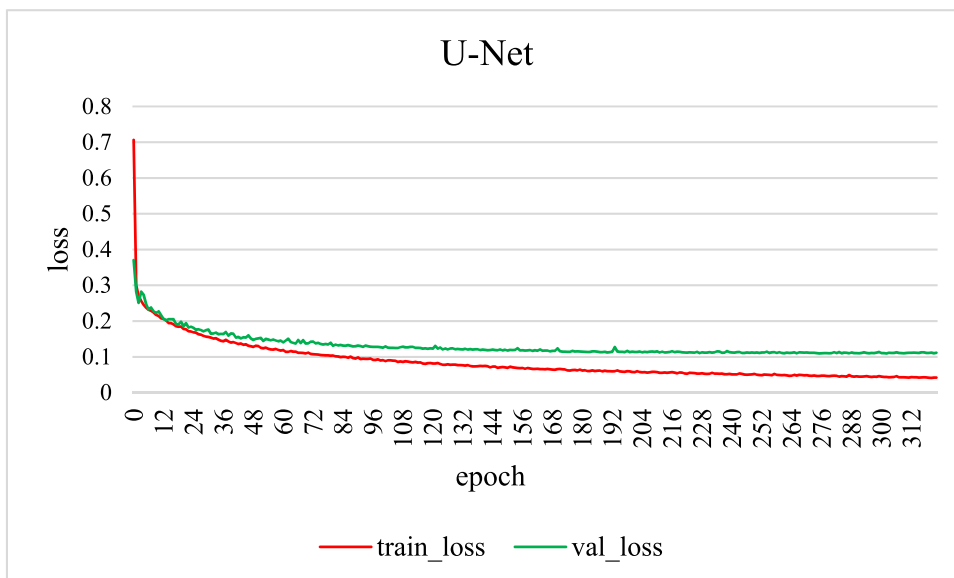
$$HD = \max[d_{XY}, d_{YX}] = \max\{\max_{x \in X} \min_{y \in Y} d[x, y], \max_{y \in Y} \min_{x \in X} d[x, y]\}. \quad (10)$$

HD is a measure to describe the degree of similarity between two sets of points, and it represents the maximum mismatch degree between the predicted result and the real result. The smaller the HD, the higher the segmentation accuracy.

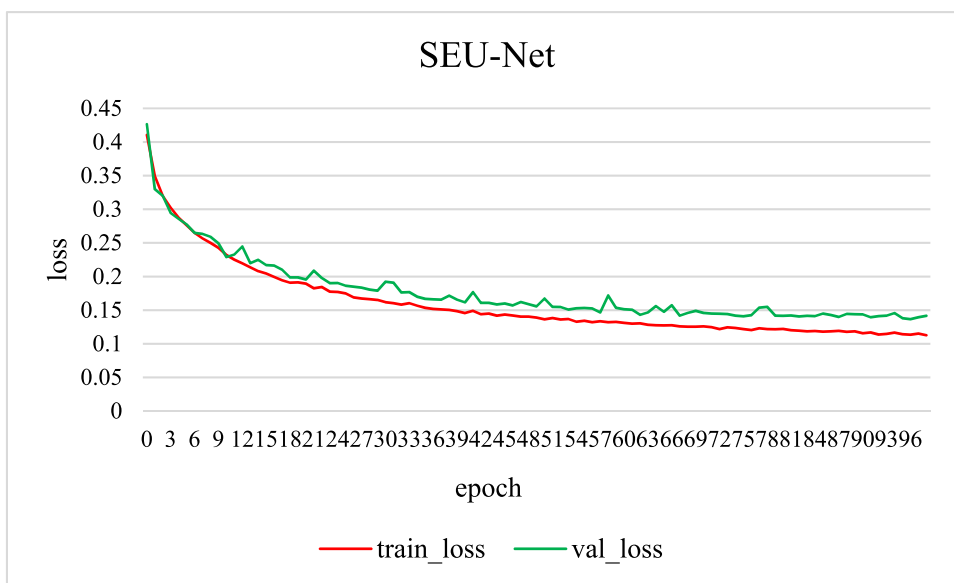
## C. IMPLEMENTATION DETAILS

Experiments were conducted on the PyTorch deep learning framework. Software environment: Win10 and Python 3.6 and CUDA10.0, CUDNN7.6.5. Hardware environment: CPU I5-7 generation (4 cores), 16G memory, GPU 12GB NVIDIA GTX1080Ti. In the training stage, the training set was randomly divided, with 75% as the training set and 25% as the verification set. Learn rate was 0.03, momentum was 0.9, batch\_size was 13, epoch was 10000, weight\_decay was 0.0001, early stopping was 10, and training was performed using the Adam optimizer [43].

To verify the effectiveness of the proposed model, we trained the original U-Net model, SEU-Net model (SE module added only in the up-sampling part of U-Net), ResNet50U-Net model (ResNet50 added only in the down-sampling part of U-Net), and SEResU-Net model. All of the models were trained by using the same datasets, the



**FIGURE 6.** Loss curves of training set and validation set for U-Net. The red curve represents the loss change of the training set during model training, and the green curve represents the loss change of the validation set during the training.



**FIGURE 7.** Loss curves of training set and validation set for SEU-Net. The red curve represents the loss change of the training set during model training, and the green curve represents the loss change of the validation set during the training.

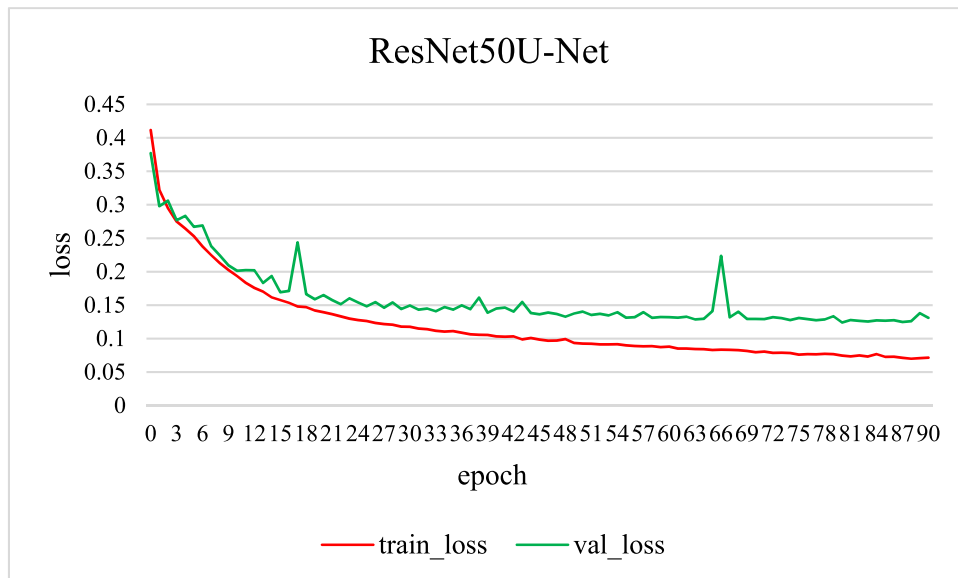
**TABLE 2.** Evaluation results of four models on the BraTS2019 dataset.

Methods	DSC			Sensitivity			Specificity			HD		
	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET
U-Net	0.8752	0.8205	0.7594	0.8908	0.8414	0.8040	0.8848	0.9024	0.7639	2.6326	1.7387	2.9757
SEU-Net	0.9012	0.8780	0.8172	0.9056	0.8896	0.8302	0.9157	0.9357	0.8355	2.4288	1.5502	2.6488
ResNet50U-Net	0.9249	0.9122	0.8594	0.9253	0.9236	0.8653	0.9392	<b>0.9460</b>	0.8777	2.2428	1.3988	2.4043
<b>SEResU-Net</b>	<b>0.9373</b>	<b>0.9180</b>	<b>0.8758</b>	<b>0.9372</b>	<b>0.9418</b>	<b>0.8738</b>	<b>0.9479</b>	0.9351	<b>0.8963</b>	<b>2.1517</b>	<b>1.3743</b>	<b>2.3162</b>

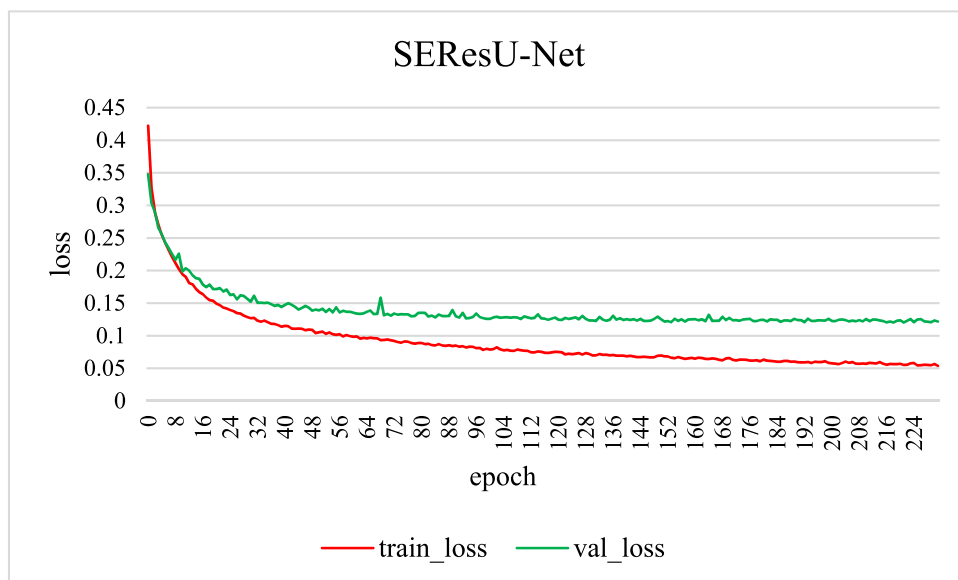
same hardware environment, and their optimal parameter settings.

The change curves of the loss functions during the model training process for the training set and the verification set versus the number of iterations are shown

in Fig. 6 through Fig. 9. As the number of iterations increases, the loss value of the network gradually decreases. When the number of training iterations of SEResU-Net reaches about 220, the loss starts to converge and stabilize.



**FIGURE 8.** Loss curves of training set and validation set for ResNet50U-Net. The red curve represents the loss change of the training set during model training, and the green curve represents the loss change of the validation set during the training.



**FIGURE 9.** Loss curves of training set and validation set for SEResU-Net. The red curve represents the loss change of the training set during model training, and the green curve represents the loss change of the validation set during training.

#### D. EXPERIMENT RESULTS

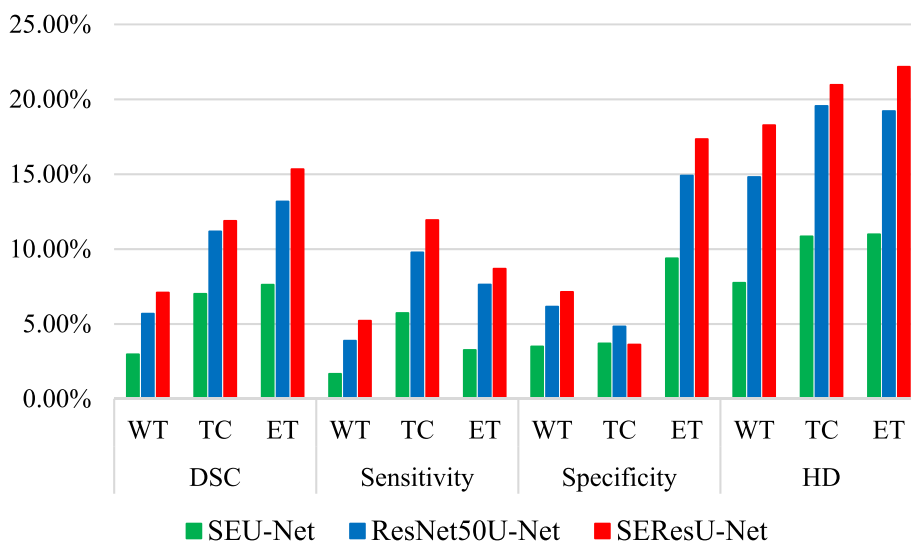
To systematically and scientifically verify the effectiveness of the proposed method, the four network models were tested in this study, including U-Net, SEU-Net, ResNet50U-Net, and SEResU-Net.

The test set were composed of the newly added 49 HGG cases and 1 LGG case in BraTS2019. Table 2 displays the segmentation results of the four models. The mean DSC scores in the WT, TC, and ET subregions of brain tumors were 0.8752, 0.8205, and 0.7594 for U-Net, 0.9012, 0.8780, and 0.8172 for

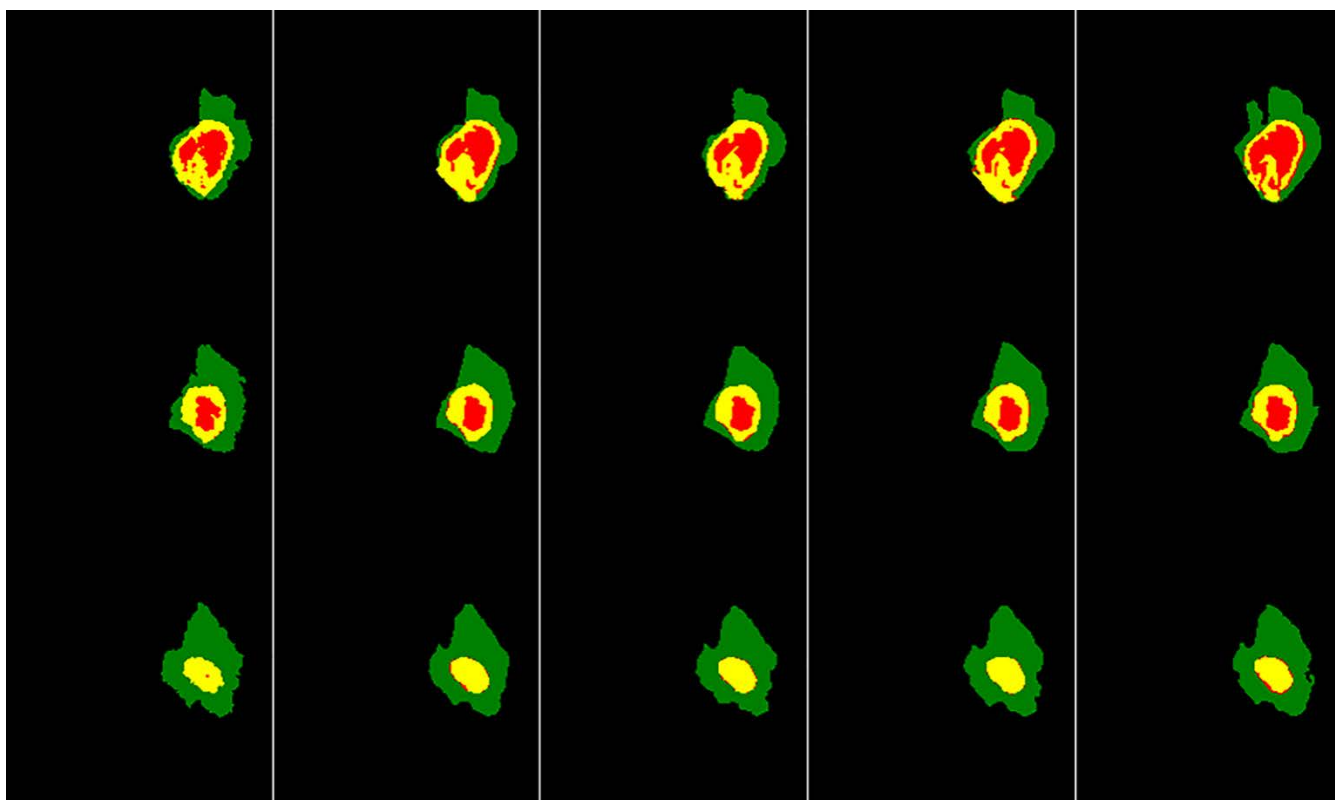
SEU-Net, 0.9249, 0.9122, and 0.8594 for ResNet50U-Net, and 0.9373, 0.9180, and 0.8758 for SEResU-Net. Compared with U-Net, the mean DSC scores of the proposed SEResU-Net increased by 7.10%, 11.88%, and 15.33% for the WT, TC, and ET segmentation (Fig. 10). Furthermore, the proposed model also exhibited higher sensitivity and specificity, and lower HD.

Next, we visualized the brain tumor segmentation results of the four models in a typical case (Fig. 11). To better observe the spatial location of brain tumors, the segmentation





**FIGURE 10.** The percentages of performance improvement of SEResU-Net, ResNet50U-Net, and SEU-Net in DSC, sensitivity, Specificity, and HD evaluation criteria compared to the U-Net baseline network.

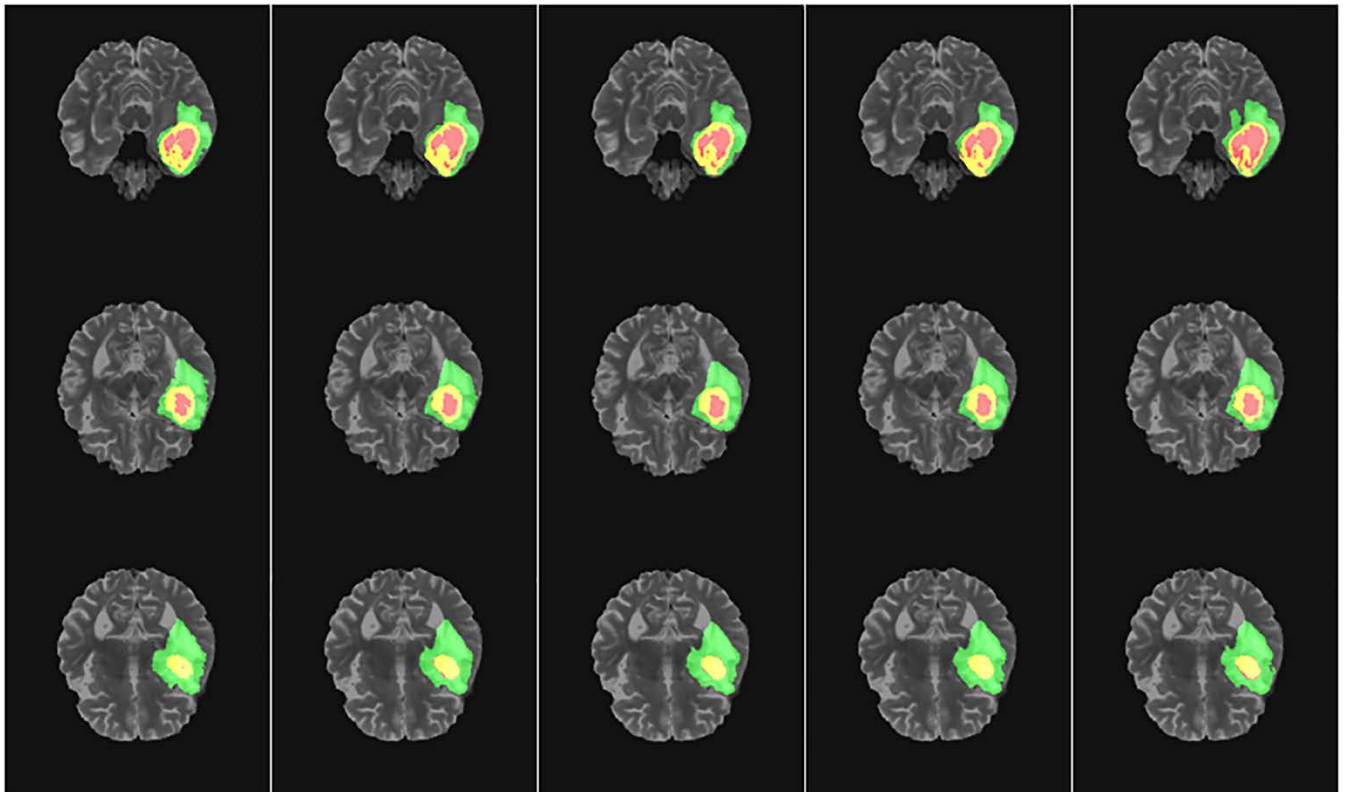


**FIGURE 11.** Example of segmentation results of four models in the BraTS2019 dataset. From left to right: ground truth, SEResU-Net, ResU-Net, SEU-Net, and U-Net segmentation results. Each row represents a different MRI slice. Each color represents a tumor class: red—necrosis and non-enhancing, green—edema, and yellow—enhancing tumor.

results were overlaid on a T2 image of this case (Fig. 12). As seen in Fig. 11 and Fig. 12, our proposed model achieved better segmentation performance than the other three models.

**E. COMPARISON WITH OTHER STATE-OF-THE-ART METHODS**

To better validate the effectiveness and robustness of the proposed method, we compared the segmentation results



**FIGURE 12.** Example of segmentation results of four models overlaid on T2 image in the BraTS2019 dataset. From left to right: ground truth, SEResU-Net, ResU-Net, SEU-Net, and U-Net segmentation results. Each row represents a different MRI slice. Each color represents a tumor class: red—necrosis and non-enhancing, green—edema, and yellow—enhancing tumor.

**TABLE 3.** Comparison of our proposed model with the state-of-the-art model on the BraTS2018 dataset.

Methods	DSC			Sensitivity			Specificity			HD		
	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET
Gan et al. <sup>[26]</sup>	0.8200	0.8500	0.7600	0.8400	0.9000	0.7900	0.8400	0.9100	0.8800	—	—	—
Aboelenen et al. <sup>[27]</sup>	0.8520	0.8120	0.7410	0.8850	0.8200	0.7680	<b>0.9980</b>	0.9970	<b>0.9990</b>	8.2500	3.3010	3.3040
Zhou et al. <sup>[28]</sup>	0.9121	0.8662	0.8181	0.9139	0.8558	0.8520	—	—	—	3.8800	6.8100	2.700
Wang et al. <sup>[29]</sup>	0.8970	0.9060	0.8430	0.9150	0.9010	0.8760	0.9910	<b>0.9980</b>	0.9970	5.2300	6.3650	2.1990
Chen et al. <sup>[30]</sup>	0.8600	0.7900	0.7500	0.9100	0.8600	0.8300	0.8400	0.7800	0.7400	—	—	—
Huang et al. <sup>[31]</sup>	0.9090	0.8450	0.8130	—	—	—	—	—	—	5.0700	6.2000	2.0900
<b>SEResU-Net</b>	<b>0.9413</b>	<b>0.9193</b>	<b>0.8736</b>	<b>0.9384</b>	<b>0.9378</b>	<b>0.8935</b>	0.9579	0.9371	0.8975	<b>2.1317</b>	<b>1.2843</b>	<b>2.1012</b>

achieved by the proposed model with the state-of-the-art methods as mentioned in the related work section on the benchmark BraTS2018 dataset (Table 3). As shown in Table 3, the proposed method offers better performance (higher DSC scores and sensitivity, and lower HD) than other methods and a relatively high specificity.

## V. CONCLUSION AND FUTURE WORK

This study proposed a new neural network model, SEResU-Net, which integrates residual modules and SENet with U-Net architecture to realize the automatic segmentation of ET, WT, and TC subregions of brain tumors. The deep residual network is beneficial for extracting deep-seated information in the down-sampling process. The addition of SENet channel attention mechanism highlights significant feature information and eliminates the ambiguity of irrelevant and noisy feature responses. The proposed model was extensively evaluated using the benchmark BraTS2018 and

BraTS2019 dataset. The experimental results demonstrated that SEResU-Net outperformed the state-of-the-art models, and can improve segmentation accuracy.

In the present study, SEResU-Net is a 2D network. When processing 3D information from MRI data, slicing processing will inevitably lose some context information and local details. Therefore, in the future, a 3D network architecture for SEResU-Net can be considered to better utilize 3D information of MRI data to improve the segmentation accuracy.

## ACKNOWLEDGMENT

The authors thank LetPub ([www.letpub.com](http://www.letpub.com)) for its linguistic assistance during the preparation of this manuscript.

## REFERENCES

- [1] A. Işın, C. Direkoğlu, and M. Şah, "Review of MRI-based brain tumor image segmentation using deep learning methods," *Proc. Comput. Sci.*, vol. 102, pp. 317–324, Jan. 2016.
- [2] W. Anjali, B. Anuj, and V. S. Verma, "A review on brain tumor segmentation of MRI images," *Magn. Reson. Imag.*, vol. 61, pp. 247–259, Sep. 2019.

- [3] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," *Phys. Med. Biol.*, vol. 58, no. 13, pp. R97–R129, Jul. 2013.
- [4] B. H. Menze et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.
- [5] P. R. Lorenzo, J. Nalepa, B. Bobek-Billewicz, P. Wawrzyniak, G. Mrukwa, M. Kawulok, P. Ulrych, and M. P. Hayball, "Segmenting brain tumors from FLAIR MRI using fully convolutional neural networks," *Comput. Methods Programs Biomed.*, vol. 176, pp. 135–148, Jul. 2019.
- [6] M. Soltaninejad, G. Yang, T. Lambrou, N. Allinson, T. L. Jones, T. R. Barrick, F. A. Howe, and X. Ye, "Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 2, pp. 183–203, 2017.
- [7] Y. Lecun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.
- [8] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- [9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [11] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [12] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [13] M. Z. Khan, M. K. Gajendran, Y. Lee, and M. A. Khan, "Deep neural architectures for medical image semantic segmentation: Review," *IEEE Access*, vol. 9, pp. 83002–83024, 2021.
- [14] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Correction: Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 546, no. 7660, p. 686, 2017.
- [15] G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, "Medical image segmentation based on U-Net: A review," *J. Imag. Sci. Technol.*, vol. 64, no. 2, pp. 020508-1–020508-12, 2020.
- [16] D. Hao, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks," in *Proc. Annu. Conf. Med. Image Understand. Anal.* Cham, Switzerland: Springer, 2017, pp. 506–517.
- [17] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.
- [18] M. Shaikh, G. Anand, G. Acharya, A. Amrutkar, V. Alex, and G. Krishnamurthi, "Brain tumor segmentation using dense fully convolutional neural network," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2017, pp. 309–319.
- [19] T. Yang, J. Song, L. Li, and Q. Tang, "Improving brain tumor segmentation on MRI based on the deep U-Net and residual units," *J. X-Ray Sci. Technol.*, vol. 28, no. 1, pp. 95–110, Feb. 2020.
- [20] Q. Li, Z. Yu, Y. Wang, and H. Zheng, "TumorGAN: A multi-modal data augmentation framework for brain tumor segmentation," *Sensors*, vol. 20, no. 15, p. 4203, Jul. 2020.
- [21] K. Li, L. Kong, and Y. Zhang, "3D U-Net brain tumor segmentation using VAE skip connection," in *Proc. IEEE 5th Int. Conf. Image, Vis. Comput. (ICIVC)*, Jul. 2020, pp. 97–101.
- [22] J. Zhang, Z. Jiang, J. Dong, Y. Hou, and B. Liu, "Attention gate ResU-Net for automatic MRI brain tumor segmentation," *IEEE Access*, vol. 8, pp. 58533–58545, 2020.
- [23] T. Zhou, S. Canu, and S. Ruan, "Fusion based on attention mechanism and context constraint for multi-modal brain tumor segmentation," *Comput. Med. Imag. Graph.*, vol. 86, Dec. 2020, Art. no. 101811.
- [24] N. Li and K. Ren, "Double attention U-Net for brain tumor MR image segmentation," *Int. J. Intell. Comput. Cybern.*, vol. 14, no. 3, pp. 467–479, Jul. 2021.
- [25] J. Zhang, X. Lv, H. Zhang, and B. Liu, "AResU-Net: Attention residual U-Net for brain tumor segmentation," *Symmetry*, vol. 12, no. 5, p. 721, May 2020.
- [26] X. Gan, L. Wang, Q. Chen, Y. Ge, and S. Duan, "GAU-Net: U-Net based on global attention mechanism for brain tumor segmentation," *J. Phys., Conf. Ser.*, vol. 1861, no. 1, Mar. 2021, Art. no. 012041.
- [27] N. M. Aboelenen, P. Songhao, A. Koubaa, A. Noor, and A. Afifi, "HTTU-Net: Hybrid two track U-Net for automatic brain tumor segmentation," *IEEE Access*, vol. 8, pp. 101406–101415, 2020.
- [28] X. Zhou, X. Li, K. Hu, Y. Zhang, Z. Chen, and X. Gao, "ERV-Net: An efficient 3D residual neural network for brain tumor segmentation," *Expert Syst. Appl.*, vol. 170, May 2021, Art. no. 114566.
- [29] J. Wang, J. Gao, J. Ren, Z. Luan, Z. Yu, Y. Zhao, and Y. Zhao, "DFP-ResUNet: Convolutional neural network with a dilated convolutional feature pyramid for multimodal brain tumor segmentation," *Comput. Methods Programs Biomed.*, vol. 208, Sep. 2021, Art. no. 106208.
- [30] G. Chen, Q. Li, F. Shi, I. Rekiq, and Z. Pan, "RFDCR: Automated brain lesion segmentation using cascaded random forests with dense conditional random fields," *NeuroImage*, vol. 211, May 2020, Art. no. 116620.
- [31] Z. Huang, Y. Zhao, Y. Liu, and G. Song, "GCAUNet: A group cross-channel attention residual UNet for slice based brain tumor segmentation," *Biomed. Signal Process. Control*, vol. 70, Sep. 2021, Art. no. 102958.
- [32] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 1–14.
- [34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [35] P. Ghosal, L. Nandanwar, S. Kanchan, and A. Bhadra, "Brain tumor classification using ResNet-101 based squeeze and excitation deep neural network," in *Proc. 2nd Int. Conf. Adv. Comput. Commun. Paradigms (ICACCP)*, 2019, pp. 1–6.
- [36] J. D. Bodapati, S. N. Shareef, V. Naralasetti, and N. B. Mundukur, "MSENet: Multi-modal squeeze-and-excitation network for brain tumor severity prediction," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 35, no. 7, pp. 1–22, 2021.
- [37] X. Zhong, O. Gong, W. Huang, L. Li, and H. Xia, "Squeeze-and-excitation wide residual networks in image classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 395–399.
- [38] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.
- [39] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.
- [40] Z. I. Botev, D. P. Kroese, R. Y. Rubinstein, and P. L'ecuyer, "The cross-entropy method for optimization," in *Handbook of Statistics*, vol. 31. Elsevier, 2013, pp. 35–59.
- [41] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Nature Sci. Data*, vol. 4, Sep. 2017, Art. no. 170117.
- [42] S. Bakas et al., "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," 2018, *arXiv:1811.02629*.
- [43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.



**CHENG DONG YAN** is currently pursuing the master's degree with the Artificial Intelligence Key Laboratory of Sichuan Province, School of Automation and Information Engineering, Sichuan University of Science and Engineering.

His research interests include medical image processing, computer vision, and deep learning.



**JURONG DING** received the B.S. degree in biomedical engineering from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2008, and the Ph.D. degree in biomedical engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2012.

From 2013 to 2018, she was an Assistant Professor with the School of Automation and Electronic Information, Sichuan University of Science and Engineering. Since 2018, she has been an Associate Professor with the School of Automation and Information Engineering, Sichuan University of Science and Engineering. She is the author of more than 30 articles. She holds two patents. Her research interests include multimodal neuroimaging analyses and applications, machine learning and deep learning with applications in image processing, and pattern recognition, and feature detection.

Dr. Ding is a Council Member of the Sichuan Youth Association for Science and Technology. She is a Guest Editor of the journal *Frontiers in Human Neuroscience*.



**HUI ZHANG** is currently pursuing the master's degree with the Artificial Intelligence Key Laboratory of Sichuan Province, School of Automation and Information Engineering, Sichuan University of Science and Engineering.

His research interests include medical image processing, computer vision, and deep learning.



**KE TONG** is currently pursuing the master's degree with the Artificial Intelligence Key Laboratory of Sichuan Province, School of Automation and Information Engineering, Sichuan University of Science and Engineering.

His research interests include medical image processing and deep learning.



**BO HUA** received the B.S. degree in communication engineering from the Chengdu University of Information Technology, Chengdu, China, in 2010, and the M.S. degree in pattern recognition and intelligent system from the Sichuan University of Science and Engineering, Zigong, China, in 2020.

Since 2020, he has been a Research Assistant with the School of Automation and Information Engineering, Sichuan University of Science and Engineering. He is the author of seven articles. His research interests include multimodal neuroimaging analyses and applications, image processing, and pattern recognition.



**SHAOLONG SHI** (Member, IEEE) received the B.S. and M.S. degrees in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2014 and 2016, respectively, and the Ph.D. degree in computer science and technology from the Harbin Institute of Technology, and the Southern University of Science and Technology, Shenzhen, China, in 2020.

He is currently a Postdoctoral Fellow with the Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou, China, and the City University of Hong Kong, Hong Kong, supported by the Hong Kong Scholar Fellowship. His research work had been published in prominent journals, including the *IEEE TRANSACTIONS ON CYBERNETICS*, *IEEE TRANSACTIONS ON NANOBIOSCIENCE*, and *IEEE TRANSACTIONS ON NANOTECHNOLOGY*, and flagship conferences, including *IEEE-NANO*, *IEEE ICC*, and *IEEE CEC*. His current research interest includes computational nanobiosensing with applications in cancer detection.

...