ELU-Net: An Efficient and Lightweight U-Net for Medical Image Segmentation

YUNJIAO DENG[1], YULEI HOU[1], JIANGTAO YAN[2], AND DAXING ZENG[3]
[1]School of Mechanical Engineering, Yanshan University, Qinhuangdao 066004, China
[2]Department of Orthopaedics, First Hospital of Qinhuangdao, Qinhuangdao 066004, China
[3]School of Mechanical Engineering, Dongguan University of Technology, Dongguan 523015, China

Corresponding author: Jiangtao Yan (470863606@qq.com)

ABSTRACT Recent years have witnessed a growing interest in the use of U-Net and its improvement. It is one of the classic semantic segmentation networks with an encoder-decoder architecture and is widely used in medical image segmentation. In the series versions of U-Net, U-Net++ has been developed as an improved U-Net by designing an architecture with nested and dense skip connections, and U-Net 3+ has been developed as an improved U-Net++ by taking advantage of full-scale skip connections and deep supervision on full-scale aggregated feature maps. Each network architecture has its own advantages in the use of the encoder and decoder. In this paper, we propose an efficient and lightweight U-Net (ELU-Net) with deep skip connections. The deep skip connections include same- and large-scale skip connections from the encoder to fully extract the features of the encoder. In addition, the proposed ELU-Net with different loss functions is discussed to improve the effect of brain tumor learning including WT (whole tumor), TC (tumor core) and ET (enhance tumor) and a new loss function DFK is designed. The effectiveness of the proposed method is demonstrated for a brain tumor dataset used in the BraTS 2018 Challenge and liver dataset used in the ISBI LiTS 2017 Challenge.

INDEX TERMS Semantic segmentation, medical imaging, deep skip connection, lightweight architecture.

I. INTRODUCTION

The rapid development of deep learning has been widely used in the medical, industrial, agriculture, and transportation fields. It is gradually playing a huge role in people's productivity and life due to its efficient performance, and has received increasing attention and applications from scientific researchers. Its application in medical imaging has also become a current research hotspot.

The most widely used medical imaging technique is image semantic segmentation, which is used in automatic segmentation and recognition of organs and lesions. Typical image semantic segmentation algorithms include FCN [1], SegNet [2], U-Net [3], PSPNet [4], series versions of Deeplab [5]–[7], DANet [8], etc. Among them, U-Net is more suitable for

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

medical segmentation tasks due to its unique architecture. A large number of researchers have made many improvements and attempts on this basis, and have achieved a series of gratifying achievements. For example, U-Net and its improved versions [9] are used to separate out bladder cancer cells [10], predict skin lesions [11], and segment gallstones [12], liver [13], liver tumors [14], and brain tumor [15], [16], etc.

For the series versions of U-Net, U-Net uses skip connections to combine the high-level semantic feature maps from the decoder and corresponding low-level detailed feature maps from the encoder. U-Net++ introduces the nested and dense skip connections from DenseNet [17] to further strengthen the plain skip connections for reducing the semantic gap between the encoder and decoder [18]. In U-Net 3+, each decoder layer incorporates both smaller- and same-scale feature maps from the encoder and larger-scale feature
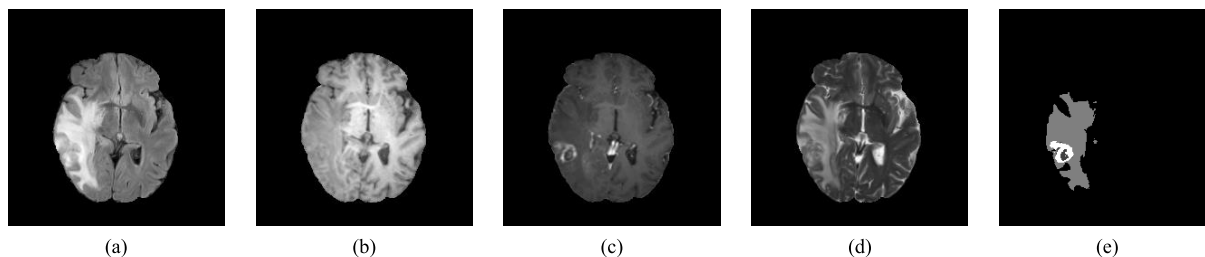
FIGURE 1. Different modalities of MRI in the BraTS 2018 challenge. (a) Flair, (b) T1, (c) contrast enhanced T1, (d) T2, and (e) ground truth of brain tumor and its sub-regions.

maps from the decoder to capture fine-grained details and coarse-grained semantics in full scales [19]. Despite achieving better performance, the series versions of U-Net are still incapable of exploring sufficient information from full scales of the encoder to give full play to the real ability of the U-shaped network in semantic segmentation. In addition, they have a large number of parameters. We focus on the real ability of a U-shaped network with fewer parameters.

Two datasets were selected to verify the effectiveness of our proposed method, one of which is the brain tumor dataset from the Brain Tumor Segmentation Benchmark (BraTS) 2018 Challenge with multi-classification tasks (https://aistudio.baidu.com/aistudio/datasetdetail/64660), and the other is the liver dataset from the Liver Tumor Segmentation Benchmark Challenge organized by the 2017 IEEE International Symposium on Biomedical Imaging (ISBI LiTS 2017 Challenge) with two-classification tasks (https://aistudio.baidu.com/aistudio/datasetdetail/79729). Glioma, a type of brain tumor including LGG (low-grade glioma) and HGG (high-grade glioma), has received general concern from researchers [20] due to its difficulty in recognition and the public challenges. In the BraTS 2018 challenge, there are four modalities of brain MRI, Flair, T1, contrast enhanced T1, and T2, available for prediction, and the segmentation task consists of three nested brain tumor sub-regions: WT (whole tumor), TC (tumor core) and ET (enhance tumor), as shown in Figure 1. In the ISBI LiTS 2017 Challenge, the segmentation task was the liver region.

II. RELATED WORK

It is worth noting that the U-Net improved strategy has received extensive attention. On the basis of U-Net, a residual module was introduced to reduce the complexity of the network architecture [21]; a U-Net network with a residual module was connected in series to avoid or minimize the natural information loss that occurs following image shrinkage [22]; Two-Pathway-Residual (TPR) blocks were designed to replace linear blocks in the U-Net network to solve the gradient degradation problem [23]; non-linear multi-level residual blocks were incorporated into skip connections to reduce the semantic gap [24]; a attention gate was introduced to focus onto the target [25]; a weighted attention mechanism was introduced into the U-Net network containing residual modules [26], and on this basis, a tightly connected network was

introduced to improve the utilization of model feature information and reduce the complexity of network learning parameters [27]; a small change was designed so that each layer in the encoder as connected with the same-size layer in scaled original image pyramids to capture the large-scale detailed information and small-scale contour information [28]; Based on the U-Net with a ResNet50 convolution block, the feature maps of different scales obtained by the decoder can be used to obtain the segmentation output using the feature pyramid network [29]; a dual-channel encoder was designed to obtain a larger receptive field and retain spatial information, including the context channel by multi-scale convolution and the spatial channel by using a large convolution kernel [30]; a dual encoder was designed by simultaneously extracting both the zero filled k-space data and undersampled image for reconstruction, which provides better representation at the bottleneck region and supplements the decoder with skip connections [31].

Furthermore, Abd-Ellah et al. [32] designed a two-parallel U-Net with asymmetric residual blocks to extract local and global features in parallel paths. Wang et al. [33] proposed a wide residual network and pyramid pool network (WRN-PPNet), in which the wide residual network (WRN) was used to extract features of multimodal brain tumor slices and the pyramid pool network (PPNet) was used to obtain the global prior representation with a different level. Tan et al. [34] replaced the ordinary architecture in U-Net with deep separable convolutional layers to distinguish the spatial correlation and appearance correlation of the mapped convolutional channel, and then introduced a residual skip connection to heighten the propagation capacity of features and increase the convergence speed. Myronenko [35] added Resnet-based skip connections and designed a VAE (Variational AutoEncoder) architecture based on the U-Net, which won 1st place on the BraTS 2018 challenge validation dataset with a Dice score value of 0.91000, 0.86680 and 0.82330 for WT, TC and ET, respectively. Ahmad [36] proposed a multi-scale hierarchical-based U-Net, which introduced a hierarchical block for merging features to extract multi-scale information.

In this paper, we proposed an efficient and lightweight U-Net (ELU-Net) with deep skip connections. Our main contributions are three-fold: (i) devising a novel ELU-Net to make full use of the full-scale features from the encoder by

introducing deep skip connections, which incorporate same- and large-scale feature maps of the encoder; (ii) discussing different loss functions and their combination to the effect of feature learning, and designing a new loss function to maximize the performance of the proposed network; (iii) conducting extensive experiments on the brain tumor dataset from the BraTS 2018 Challenge and liver dataset from the ISBI LiTS 2017 Challenge, where ELU-Net with the fewest parameters is not inferior or even better in many typical algorithms.

The remainder of this paper is organized as follows: Section 3 describes the ELU-Net network architecture and the calculated parameters for ELU-Net with Vgg16 and ResNet34. Section 4 discusses the ELU-Net with different loss functions for the brain tumor dataset and the design of a novel loss function by combination. Section 5 conducts extensive experiments on the liver dataset to verify the effectiveness of the proposed network and compares it with other representative state-of-the-art methods. In the final section, some concluding comments are made.

III. METHODS

Figure 2 gives simplified overviews of U-Net, U-Net++, U-Net 3+ and the ELU-Net. The U-Net, including encoder and decoder, is the most popular convolutional network architecture for biomedical image segmentation to predict the segmentation mask at the pixel-level rather than image-level classification. First, the image is taken into the encoder to extract the higher-level features by down-sampling the output of the previous encoder layer. Second, the output from each encoder layer is taken into the corresponding decoder layer to classify the pixels by concatenating the feature maps from the output of the last encoder layer or previous decoder layer by up-sampling to keep the scale consistent. Finally, the output of the last decoder layer is activated by the softmax to output the segmentation result.

The superiority and effectiveness of U-Net are well known. Based on U-Net, U-Net++ replaces the plain skip connection with the nested and dense skip connection, and U-Net 3+ replaces the plain skip connection with the full-scale skip connection. Furthermore, the proposed ELU-Net
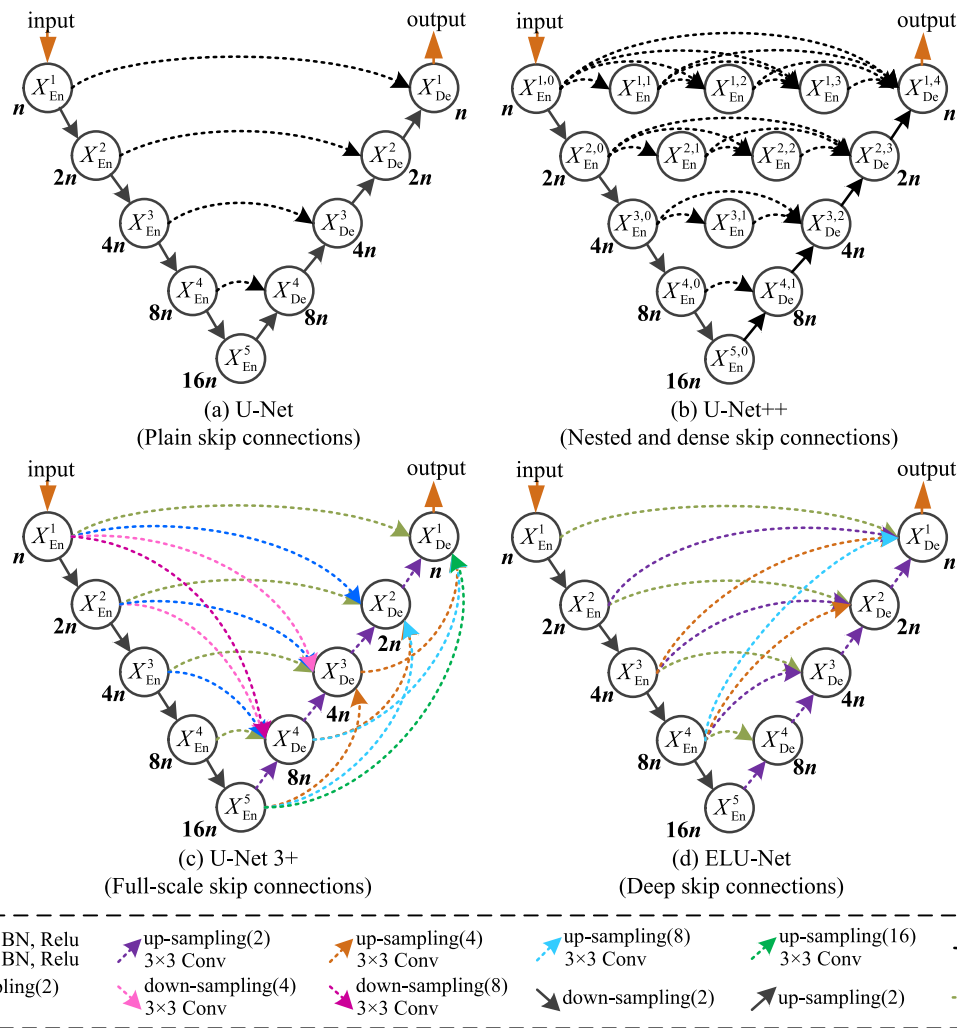

FIGURE 2. Comparison of U-Net (a), U-Net++ (b), U-Net 3+ (c) and the proposed ELU-Net (d). The number of the output channel of each node is presented below the circle.

replaces the initial plain skip connection with a deep skip connection.

By analyzing the series versions of the U-Net architecture, it is not difficult to find that the features of the encoder are the key to image segmentation. To fully extract the features of the encoder, the deep skip connection includes a plain skip connection from the corresponding encoder layer and skip connections from all deeper encoder layers except the last encoder layer. Use of the deep skip connections can enable one to fully capture fine-grained details and coarse-grained semantics in the encoder.

Formally, the parameters of the ELU-Net architecture are formulated as follows: let N refer to the total number of the encoder layer, and i index the down-sampling layer along the encoder. Each encoder layer X_{En}^i extracts the smaller-scale feature maps from the encoder to obtain the higher semantic feature. Each decoder layer X_{De}^i incorporates all the larger- and same-scale feature maps from the encoder except for the last encoder layer, and larger-scale feature maps from the decoder or the last encoder layer to obtain the segmentation result with pixel classification.

Here, each decoder layer X_{De}^i can be computed as follows:

$$
X_{De}^i = \begin{cases} \Theta\left(\begin{bmatrix} \Gamma\left(X_{En}^i\right), \Gamma\left(U\left(X_{En}^k\right)\right)_{i+1}^{N-1}, \\ \Gamma\left(U\left(X_{De}^{i+1}\right)\right) \end{bmatrix}\right), \\ \qquad i = 1, \cdots, N-2 \\ \Theta\left(\left[\Gamma\left(X_{En}^i\right), \Gamma\left(U\left(X_{En}^{i+1}\right)\right)\right]\right), \\ \qquad i = N-1 \end{cases} \tag{1}
$$

where function $\Theta\left(\cdot\right)$ realizes the feature aggregation mechanism with two convolution operations followed by a batch normalization and a ReLU activation function, $\Gamma\left(\cdot\right)$ represents a convolution operation, $U\left(\cdot\right)$ represents an up-sampling operation by utilizing a bilinear interpolation operation, and $[\cdot]$ denotes a concatenation operation.
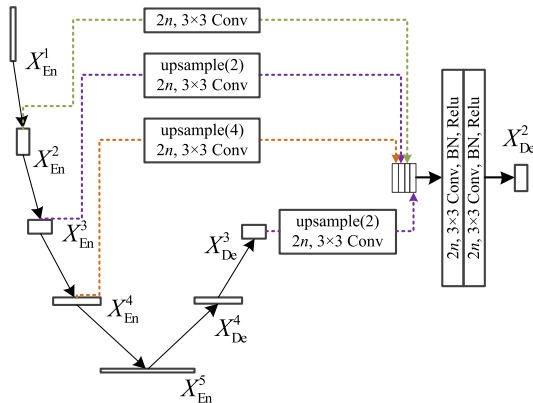

FIGURE 3. Illustration of how to construct the aggregated feature map of the second decoder layer $X_{De}^{2,3}$, where the four dotted lines represent the source.

Preferably, the total number of encoder layers N takes a value of 5. Take X_{De}^2 as an example, as shown in Figure 3.

The feature map of X_{De}^2 is constructed by the same- and larger-scale encoder layer X_{En}^2, X_{En}^3 and X_{En}^4, and the larger-scale decoder layer X_{De}^3. This incorporates the same four resolution feature maps from the encoder and decoder to seamlessly merge the shallow exquisite information with deep semantic information.

Similarly in U-Net, the number of channels of the decoder feature map in ELU-Net are symmetric to that of the encoder, and, thus, X_{En}^i and X_{De}^i both have $2^{i-1}n$ channels.

The number of parameters in the ith decoder layer of ELU-Net, P^i, can be computed as follows:

$$
\begin{aligned}
P^i &= k_e^2 \left[d\left(\sum_{k=i}^{N-1} X_{En}^k + X_{De}^{i+1}\right) d\left(X_{De}^i\right) \right. \\
&\left. + d\left((N-i+1)X_{De}^i\right) d\left(X_{De}^i\right) + d\left(X_{De}^i\right)^2 \right] \\
&= \left(\sum_{k=i}^{N-1} 2^{k-i} + N - i + 4\right) 2^{2i-2} n^2 k_e^2
\end{aligned} \tag{2}
$$

where k_e represents the convolution kernel size, and $d\left(\cdot\right)$ represents the number of channels of the notes.

Vgg16 and ResNet34 are chosen for the backbone network architecture of ELU-Net. The number of parameters for ELU-Net with Vgg16, P_v, can be computed as follows:

$$
\begin{aligned}
P_v &= \left[n_0 d\left(X_{En}^1\right) + d\left(X_{En}^1\right)^2\right] k_e^2 \\
&+ \sum_{i=2}^{N} \left[d\left(X_{En}^{i-1}\right) d\left(X_{En}^i\right) + d\left(X_{En}^i\right)^2\right] k_e^2 \\
&+ \sum_{i=1}^{4} P^i + n_c d\left(X_{De}^1\right) \\
&= 1118 n^2 k_e^2 + n_0 n k_e^2 + n_c n
\end{aligned} \tag{3}
$$

where n_0 represents the number of channels of the input, n_c represents the number of output categories of the segmentation result.

The number of parameters for ELU-Net with ResNet34, P_r, can be computed as follows:

$$
\begin{aligned}
P_r &= n_0 d\left(X_{En}^1\right) k_e^2 + \sum_{i=2}^{N} \left[d\left(X_{En}^{i-1}\right) d\left(X_{En}^i\right)\right] \\
&+ \sum_{i=2}^{N} \left[d\left(X_{En}^{i-1}\right) d\left(X_{En}^i\right) + (2w_i - 1) d\left(X_{En}^i\right)^2\right] k_e^2 \\
&+ \sum_{i=1}^{4} P^i + n_c d\left(X_{De}^1\right) \\
&= 2893 n^2 k_e^2 + n_0 n k_e^2 + 170 n^2 + n_c n
\end{aligned} \tag{4}
$$

where w_i represents the number of Basic blocks for the ith layer in ResNet34 and $w = [w_2 \; w_3 \; w_4 \; w_5] = [3 \; 4 \; 6 \; 3]$.

In order to make the network parameters lightweight and facilitate training and transplantation, the number of channels of the first encoder layer in ELU-Net, n, takes a value of 8, which is much less than the value of 64 or 32 used in the

other series versions of U-Net. The convolution kernel size, k_e, takes a value of 3. P_v and P_r can be computed as follows:

$$\begin{cases} P_v = 643968 + 72n_0 + 8n_c \\ P_r = 1677248 + 72n_0 + 8n_c \end{cases} \quad (5)$$

It is worth mentioning that the fifth encoder layer is not necessarily the deepest level. The value of N can be 5, 6, or any number. Of course, it is not clear if a bigger value of N will lead to a better segmentation result. This depends on the type of datasets and computing power. In this paper, 5 layers are competent for this segmentation task, which is relatively friendly considering the requirements of computing power.

IV. DISCUSSION OF THE LOSS FUNCTION
A. LOSS FUNCTION
In order to give full play to the effect of the proposed ELU-Net as much as possible, it is necessary to choose a suitable loss function [37], which is used to evaluate the matching degree between the predicted label of the segmentation result and the true label.

Cross entropy loss (CE), as a widely used loss function, can be expressed as follows:

$$L_{CE} = -\frac{1}{N_b} \sum_{i=1}^{N_b} \sum_{c=1}^{C} y_{true} \log\left(y_{pred}\right) \quad (6)$$

where y_{true} represents the true label for the segmentation task, y_{pred} represents the predictive label of the segmentation result, N_b represents the number of the batch size used for training, and C represents the number of segmentation classes.

Then, 8 other loss functions including L2 loss (F2), smoothF1 loss (SF), bce loss (BCE), bcewithlogits loss (BCL), dice loss (DL), Tversky loss (TL), focal loss (FL) and KL divergence loss (KL), were selected, which can be expressed as

$$L_{F2} = -\frac{1}{N_b} \sum_{i=1}^{N_b} \sum_{c=1}^{C} \left(y_{true} - y_{pred}\right)^2 \quad (7)$$

$$L_{SF} = -\frac{1}{N_b} \sum_{i=1}^{N_b} \sum_{c=1}^{C} \\ \times \begin{cases} 0.5\left(y_{true} - y_{pred}\right)^2, & \left|y_{true} - y_{pred}\right| < 1 \\ \left|y_{true} - y_{pred}\right| - 0.5, & \text{otherwise} \end{cases} \quad (8)$$

$$L_{BCE} = -\frac{1}{N_b} \sum_{i=1}^{N_b} \sum_{c=1}^{C} \left[\begin{array}{l} y_{true} \log\left(y_{pred}\right) \\ + (1 - y_{true}) \log\left(1 - y_{pred}\right) \end{array} \right] \quad (9)$$

$$L_{BCL} = -\frac{1}{N_b} \sum_{i=1}^{N_b} \sum_{c=1}^{C} \left[\begin{array}{l} y_{true} \log \sigma\left(y_{pred}\right) \\ + (1 - y_{true}) \log\left(1 - \sigma\left(y_{pred}\right)\right) \end{array} \right] \quad (10)$$

$$L_{DL} = \frac{1}{C} \sum_{c=1}^{C} \frac{1}{N_b} \left(1 - \frac{2\sum_{i=1}^{N_b} y_{true} y_{pred}}{\sum_{i=1}^{N_b} y_{true}^2 + \sum_{i=1}^{N_b} y_{pred}^2} \right) \quad (11)$$

TABLE 1. Confusion matrix.

Confusion matrix		True value	
		Positive	Negative
Predictive value	Positive	TP	FP
	Negative	FN	TN

TABLE 2. The Dice coefficients for the segmentation result obtained for the ELU-Net with Vgg16 and ResNet34.

Method	Mean	WT	TC	ET
ELU-Net (Vgg16)	0.83096	0.91397	0.82751	0.75141
ELU-Net (ResNet34)	0.86817	0.93619	0.85526	0.81307

$$\begin{cases} L_{TL} = 1 - \frac{1}{N_b}\left(\frac{t_p}{t_p + \alpha f_p + \beta f_n}\right) \\ t_p = \sum_{i=1}^{N_b} \sum_{c=1}^{C} y_{true} y_{pred} \\ f_p = \sum_{i=1}^{N_b} \sum_{c=1}^{C} (1 - y_{true}) y_{pred} \\ f_n = \sum_{i=1}^{N_b} \sum_{c=1}^{C} y_{true} \left(1 - y_{pred}\right) \end{cases} \quad (12)$$

$$L_{FL} = -\frac{1}{N_b} \sum_{i=1}^{N_b} \sum_{c=1}^{C} \lambda \left(1 - y_{pred}\right)^{\gamma} y_{true} \log\left(y_{pred}\right) \quad (13)$$

$$L_{KL} = \frac{1}{N_b} \sum_{i=1}^{N_b} \sum_{c=1}^{C} y_{true} \left(\log y_{true} - \log\left(\sigma y_{pred}\right)\right) \quad (14)$$

where σ represents the sigmoid function, α and β represent the hyperparameter of the Tversky loss, λ represents the balance factor of the focal loss, and γ represents the exponential factor of the focal loss.

B. DATASET AND EVALUATION METRIC
The brain tumor dataset using in the BraTS 2018 challenge with multi-classification tasks was selected to discuss the effect of the proposed ELU-Net with the 9 loss functions and their combinations. It contains 210 cases for HGG and 75 cases for LGG. According to the size of the brain MRI, all slices were extracted for each case except the slices in whose corresponding segmentation result the number of pixels classified as ET were less than 20, of which 90% slices and 10% slices were used for training and testing for the ELU-Net, respectively.

We utilized the Adam algorithm for optimization. Its learning rate was set to 1e-4, and was reduced to one tenth of the previous rate whenever the evaluation metric of verification dataset had been not updated for 20 consecutive generations. Its weight decay was set to 0.002.

The input of ELU-Net cascades 4 slices corresponding to the four modalities of brain MRI including Flair, T1, contrast enhanced T1, and T2, and each slice has 3 channels, and thus a n_0 value of 12. The output of ELU-Net is 4 channels ($n_c = 4$) corresponding to the background and

TABLE 3. The Dice coefficients for the segmentation result obtained for the ELU-Net with ResNet34 by different loss functions. The best result is highlighted in bold.

				Loss function						Dice		
CE	F2	SL	BCE	BCL	DL	TL	FL	KL	Mean	WT	TC	ET
√									0.86817	**0.93619**	0.85526	0.81307
	√								0.82087	0.91486	0.79757	0.75018
		√							0.78806	0.89835	0.77835	0.68748
			√						0.86560	0.93535	0.85441	0.80705
				√					0.86573	0.93534	0.85395	0.80791
					√				0.86863	0.93537	0.85554	0.81499
						√			0.85842	0.92106	0.85185	0.80235
							√		0.86937	0.93369	0.85793	0.81649
								√	0.86488	0.93114	0.85560	0.80792
√			√						0.86738	0.93583	0.85537	0.81093
√					√				0.86764	0.93599	0.85594	0.81099
√						√			0.86829	0.93553	0.85498	0.81436
√							√		0.87012	0.93563	**0.86115**	0.81358
√								√	0.87009	0.93452	0.85965	0.81608
					√		√		**0.87095**	0.93395	0.85986	**0.81850**
							√	√	0.87006	0.93466	0.85919	0.81633
					√			√	0.86987	0.93514	0.85985	0.81461

three categories including enhancing tumor (ET), peritumoral edema (ED), and non-enhancing tumor (NET). They constitute three nested brain tumor sub-regions, WT, TC and ET, as the image segmentation result:

$$\begin{cases} \text{WT} = \text{ET} + \text{ED} + \text{NET} \\ \text{TC} = \text{ET} + \text{NET} \\ \text{ET} = \text{ET} \end{cases} \quad (15)$$

A confusion matrix was introduced to evaluate the results and the basic parameters are listed in Table 1.

In Table 1, TP refers to the number for which the true value is positive and the predictive value is positive. FN refers to the number for which the true value is positive and the predictive value is negative. FP refers to the number for which the true value is negative and the predictive value is positive. TN refers to the number for which the true value is negative and the predictive value is negative.

The Dice coefficient was used as the evaluation metric for each segmentation result, which is expressed as follows:

$$Dice = \frac{2TP}{FP + 2TP + FN} \quad (16)$$

C. DISCUSSION

First, under the cross entropy loss (CE), we compared the segmentation result obtained for the proposed ELU-Net with Vgg16 and ResNet34. The Dice coefficients of the three nested brain tumor sub-regions on the BraTS 2018 validation dataset are listed in Table 2.

In Table 2, the ELU-Net with ResNet34 achieves a better performance, which exceeds 0.03721 than Vgg16 on average.

In order to give further play to the advantage of the ELU-Net with ResNet34, another 8 loss functions and their combinations were compared and their corresponding Dice

coefficients on the BraTS 2018 validation dataset are listed in Table 3.

As given in Table 3, the combination with dice loss and focal loss achieves a best performance on average, 0.87095, and the focal loss achieves a best performance on average among the alone loss function, 0.86937, followed by the dice loss, 0.86863. As for the sole sub-region, the cross entropy loss, focal loss and the combination with dice loss and focal loss shows the best performance for WT, TC and ET of 0.93619, 0.86115, 0.81850, respectively.

Based on a combination of dice loss and focal loss, a novel loss function with dice loss, focal loss and KL divergence loss, DFK, was designed as follows:

$$L_{DFK} = 3 \log DL^2 + \log FL^2 + 0.1 \log KL^2 \quad (17)$$

The Dice coefficients for the three nested brain tumor sub-regions on the BraTS 2018 validation dataset were obtained through training the proposed ELU-Net with Vgg16 and ResNet34 by the loss function with dice loss, focal loss and KL divergence loss (DFK), which are compared with other state-of-the-art methods in Table 4. Some segmentation results obtained using the ELU-Net with ResNet34 and DFK on the BraTS 2018 validation dataset are shown in Figure 4, which show good performance.

As given in Table 4, with the help of DFK, the ELU-Net with ResNet34 achieves on average the best performance compared to other state-of-the-art methods. For the sole sub-region, the ELU-Net with ResNet34 shows an obvious advantage on WT, 0.93498, and considerable performance on TC and ET, 0.86023 and 0.81779, which are just reduced by 0.00657 and 0.00551, respectively, with fewer parameters compared with the best results obtained by the other state-of-the-art methods. According to Eq. (5), the numbers of
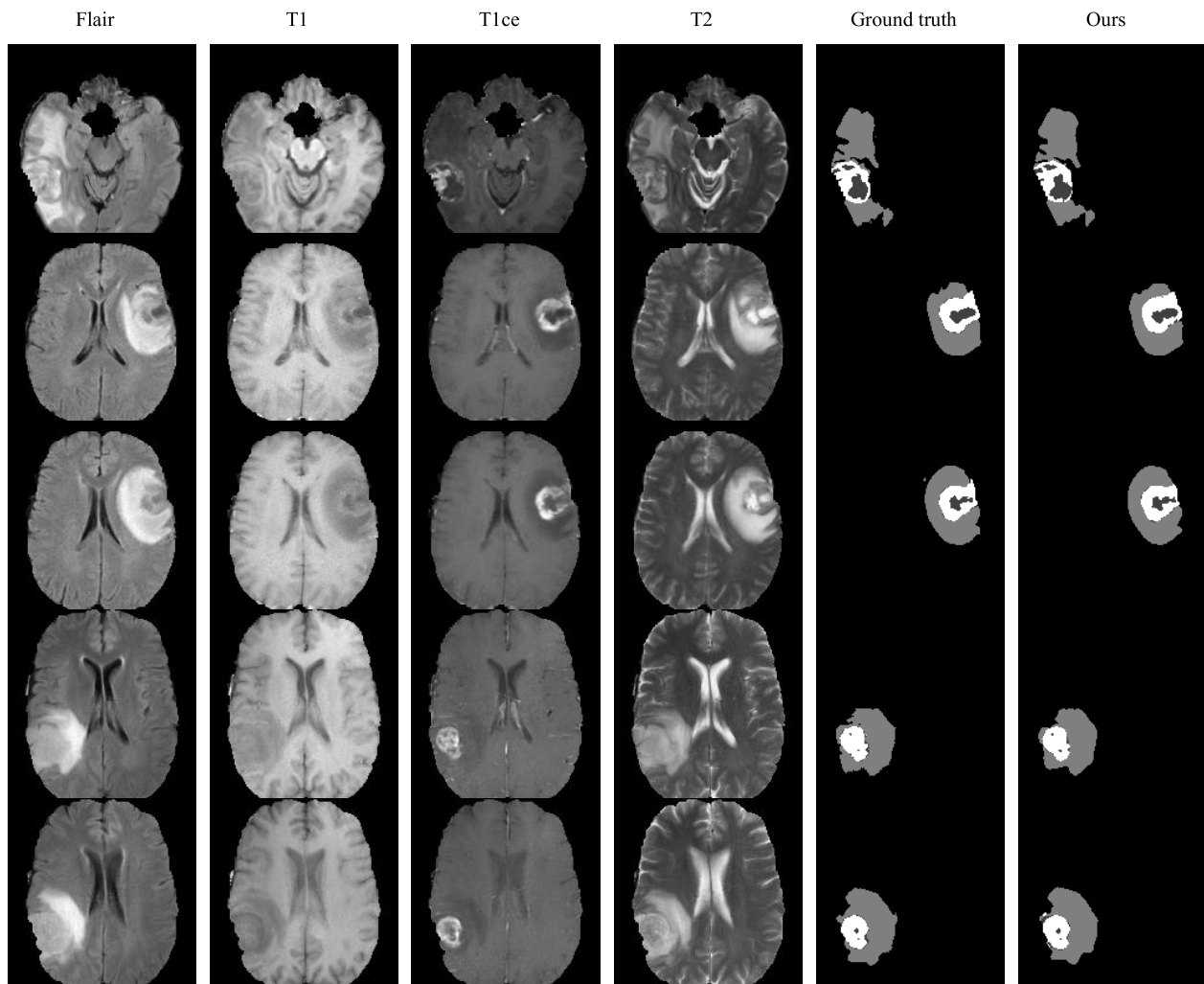
Flair	T1	T1ce	T2	Ground truth	Ours

FIGURE 4. The segmentation result obtained for the BraTS 2018 validation dataset by the ELU-Net with ResNet34 and DFK.

TABLE 4. Comparison of our methods and other state-of-the-art methods on the BraTS 2018 validation dataset. The best result is highlighted in bold.

Method	Mean	WT	TC	ET
McKinley et al. [38]	0.84867	0.90300	0.84700	0.79600
Zhou et al. [39]	0.85880	0.90780	0.85750	0.81111
Kermi et al. [40]	0.81867	0.86800	0.80500	0.78300
Albiol et al. [41] (VGG)	0.79433	0.87200	0.76000	0.75100
Albiol et al. [41] (Ensemble)	0.81033	0.88100	0.77700	0.77300
Feng et al. [42]	0.84577	0.90940	0.83620	0.79170
Ahmad et al. [36]	0.85401	0.90715	0.84115	0.81374
Myronenko. [35] (Single)	0.85943	0.90420	0.85960	0.81450
Myronenko. [35] (Ensemble)	0.86670	0.91000	**0.86680**	**0.82330**
No New-Net [20]	0.86157	0.91260	0.86340	0.80870
ELU-Net (Vgg16+DFK)	0.83391	0.91165	0.82860	0.76147
ELU-Net (ResNet34+DFK)	**0.87100**	**0.93498**	0.86023	0.81779

parameters for the ELU-Net with Vgg16 and ResNet34 are only 644, 864 and 1, 678, 144, respectively.

In addition, the ELU-Net with ResNet34 achieves a better performance than with Vgg16, exceeded by 0.03709 on average and 0.02333, 0.03163 and 0.05632 on WT, TC and ET, respectively, which proves the effectiveness and advantage of the residual network. Based on the ResNet34 backbone network, compared with the result of the combination with dice loss and focal loss, the DFK loss performs better on WT and TC, exceeded by 0.00103 and 0.00037, and a little poor on ET, reduced by 0.00071. In summary, the ELU-Net with ResNet34 and DFK loss shows a better overall performance.

Moreover, with the help of the trained weights for the ELU-Net with ResNet34 and DFK, the Dice coefficients of the three nested brain tumor sub-regions on the HGG and LGG validation dataset were obtained by training, respectively, and are listed in Table 5.

As given in Table 5, the ELU-Net with ResNet34 and DFK shows a better performance on the HGG than that on the LGG, which exceeds that on the whole dataset by 0.00812 and 0.02434 on TC and ET, respectively, and is reduced by 0.00456 on WT. However, for the LGG, a better

TABLE 5. The Dice coefficients for the segmentation result obtained for the ELU-Net with ResNet34 and DFK on HGG and LGG.

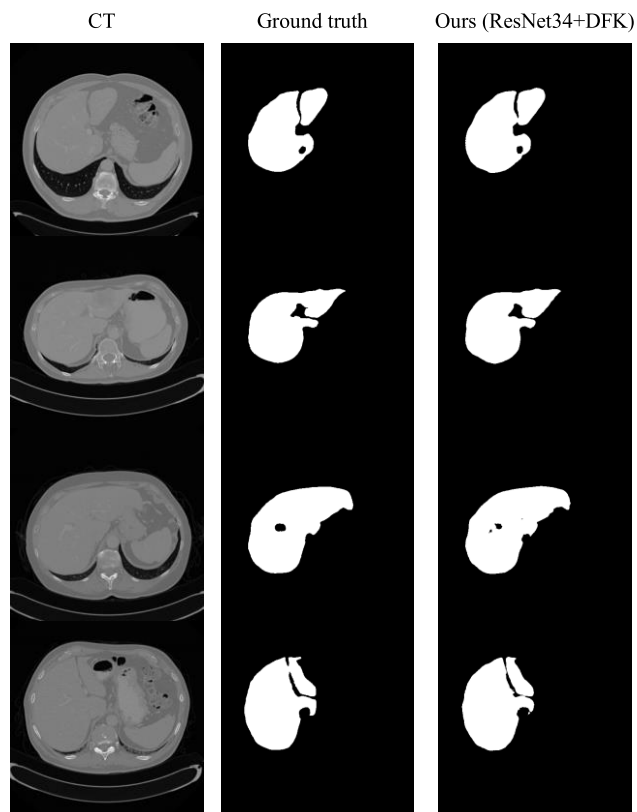|	Mean	WT	TC	ET
HGG	0.88030	0.93042	0.86835	0.84213
LGG	0.78694	0.95611	0.71682	0.68789

FIGURE 5. The segmentation result obtained for the ISBI LITS 2017 validation dataset by the ELU-Net with ResNet34 and DFK.

performance on WT is obtained, which exceeds that on the whole dataset by 0.02113.

V. COMPARISON WITH A LIVER DATASET

The liver dataset was selected to further validate the effectiveness of the proposed ELU-Net, which is obtained from the ISBI LiTS 2017 Challenge. It contains 131 contrast-enhanced 3D abdominal CT scans, of which 120 and 11 volumes were used for training and testing, respectively. The 3 most obvious slices were extracted for each volume according to the size of the liver. The other hyperparameters were consistent with the aforementioned method.

The input of ELU-Net is 1 slice with 3 channels ($n_0 = 3$), and the output is 2 channels ($n_c = 2$) including the background and liver region.

The proposed ELU-Net with Vgg16 and ResNet34 was quantitatively compared with the other representative state-of-the-art methods based on the Dice coefficients.

Moreover, it is worth mentioning that all results were directly obtained from a single-model test without relying on any post-processing tools and each network was optimized by

TABLE 6. Comparison of our methods and other state-of-the-art methods applied to the ISBI LiTS 2017 validation dataset. The best result obtained is highlighted in bold.

Method	*Dice*
PSPNet [4]	0.92420
DeepLabV2 [5]	0.90210
DeepLabV3 [6]	0.92170
DeepLabV3+ [7]	0.91860
U-Net (Vgg16) [3]	0.92060
U-Net (ResNet101) [3]	0.93870
Attention U-Net [25]	0.93410
U-Net++ (Vgg16) [18]	0.92780
U-Net++ (ResNet101) [18]	0.94750
U-Net 3+ (Vgg16) [19]	0.95500
U-Net 3+ (ResNet101) [19]	0.96010
U-Net 3+ (ResNet101+Hybrid loss) [19]	0.96430
U-Net 3+ (ResNet101+Hybrid loss+CGM) [19]	0.96750
Ours (Vgg16+CE)	0.96000
Ours (Vgg16+DFK)	0.96059
Ours (ResNet34+CE)	0.96829
Ours (ResNet34+DFK)	**0.97365**

* CGM is the classification-guided module [19].

the loss function proposed in its own article. The input of our segmentation network is directly obtained from the extracted slices without relying on any filtering and enhancement processing. The comparison result is shown in Table 6.

As given in Table 6, the ELU-Net with ResNet34 and DFK greatly improves the performance of the U-shaped network. It has the fewest parameters (1.68M), but it has an absolute advantage, even for the other state-of-the-art network architectures with ResNet101 and shows some particular improvements. Its Dice value exceeds that of the U-Net 3+ with ResNet101 and Hybrid loss and CGM by 0.615% and exceeds that of the U-Net 3+ with Vgg16 by 1.865%, but the number of its parameters is only 3.86% of that of the U-Net 3+ with ResNet101 (43.55M) and 6.23% of that of the U-Net 3+ with Vgg16 (26.97M), which proves the lightweight and effectiveness of the ELU-Net. Some segmentation results obtained for the ISBI LiTS 2017 validation dataset are shown in Figure 5.

In addition, the ELU-Net with Vgg16 and DFK (644, 200 parameters) also shows a considerable performance, which proves the effectiveness and advantage of our network architecture.

VI. CONCLUSION

In this study, we proposed a novel ELU-Net with deep skip connections to make full use of the features from the encoder for realizing an efficient and lightweight segmentation network architecture. The Vgg16 and ResNet34 backbone network with many loss functions and their combinations is found to give full play to the effect of our methods, and a new loss function with dice loss, focal loss and KL divergence loss, DFK, was designed based on the exponential and logarithmic advantages. The experimental results obtained for

brain tumor and liver datasets demonstrate the effectiveness and outstanding performance of the proposed ELU-Net architecture with fewer parameters. The ELU-Net with ResNet34 and DFK show Dice coefficients of 93.498%, 86.023% and 81.779% for WT, TC and ET with an average value of 87.100% for the BraTS 2018 validation dataset, and a value of 97.365% for the ISBI LiTS 2017 validation dataset.

ACKNOWLEDGMENT

REFERENCES

[1] 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.

[2] 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.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[4] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2881–2890.

[5] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[6] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[7] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, 2018, pp. 801–818.

[8] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 3141–3149.

[9] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021.

[10] H. Hu, Y. Zheng, Q. Zhou, J. Xiao, S. Chen, and Q. Guan, "MC-UNet: Multi-scale convolution UNet for bladder cancer cell segmentation in phase-contrast microscopy images," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, San Diego, CA, USA, Nov. 2019, pp. 1197–1199.

[11] J. Wu, E. Z. Chen, R. Rong, X. Li, D. Xu, and H. Jiang, "Skin lesion segmentation with C-UNet," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Berlin, Germany, Jul. 2019, pp. 2785–2788.

[12] T. Song, F. Meng, A. Rodríguez-Patón, P. Li, P. Zheng, and X. Wang, "U-Next: A novel convolution neural network with an aggregation U-Net architecture for gallstone segmentation in CT images," *IEEE Access*, vol. 7, pp. 166823–166832, 2019.

[13] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[14] W. Zhao, K. Li, D. Zhao, Y. Jiang, J. Wu, J. Chen, X. Quan, X. Li, and F. Xue, "Liver segemtation in CT image with no-edge-cuting UNet," in *Proc. IEEE 9th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC)*, Chongqing, China, Dec. 2020, pp. 2315–2318.

[15] D. Maji, P. Sigedar, and M. Singh, "Attention res-UNet with guided decoder for semantic segmentation of brain tumors," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103077.

[16] Y. Liu, J. Du, C.-M. Vong, G. Yue, J. Yu, Y. Wang, B. Lei, and T. Wang, "Scale-adaptive super-feature based MetricUNet for brain tumor segmentation," *Biomed. Signal Process. Control*, vol. 73, Mar. 2022, Art. no. 103442.

[17] 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)*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269.

[18] Z. Zhou, S. M. Rahman, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA ML-CDS)*. Granada, Spain, 2018, pp. 3–11.

[19] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet 3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 1055–1059.

[20] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "No new-net," in *Proc. Int. MICCAI Brainlesion Workshop*, 2018, pp. 234–244.

[21] P. Xiuqin, Q. Zhang, H. Zhang, and S. Li, "A fundus retinal vessels segmentation scheme based on the improved deep learning U-Net model," *IEEE Access*, vol. 7, pp. 122634–122643, 2019.

[22] X. Xiao, S. Lian, Z. Luo, and S. Li, "Weighted res-UNet for high-quality retina vessel segmentation," in *Proc. 9th Int. Conf. Inf. Technol. Med. Educ. (ITME)*, Hangzhou, China, Oct. 2018, pp. 327–331.

[23] M. Aghalari, A. Aghagolzadeh, and M. Ezoji, "Brain tumor image segmentation via asymmetric/symmetric UNet based on two-pathway-residual blocks," *Biomed. Signal Process. Control*, vol. 69, Aug. 2021, Art. no. 102841.

[24] N. R. Gudhe, H. Behravan, M. Sudah, H. Okuma, R. Vanninen, V.-M. Kosma, and A. Mannermaa, "Multi-level dilated residual network for biomedical image segmentation," *Sci. Rep.*, vol. 11, no. 1, p. 14105, Jul. 2021.

[25] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, Apr. 2018, pp. 1-10.

[26] Z. Luo, Y. Zhang, L. Zhou, B. Zhang, J. Luo, and H. Wu, "Micro-vessel image segmentation based on the AD-UNet model," *IEEE Access*, vol. 7, pp. 143402–143411, 2019.

[27] G. Alfonso Francia, C. Pedraza, M. Aceves, and S. Tovar-Arriaga, "Chaining a U-Net with a residual U-Net for retinal blood vessels segmentation," *IEEE Access*, vol. 8, pp. 38493–38500, 2020.

[28] Y. Tian, Y. Hu, Y. Ma, H. Hao, L. Mou, J. Yang, Y. Zhao, and J. Liu, "Multiscale U-Net with edge guidance for multimodal retinal image deformable registration," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Montreal, QC, Canada, Jul. 2020, pp. 1360–1363.

[29] X. Liu, W. Deng, and Y. Liu, "Application of hybrid network of UNet and feature pyramid network in spine segmentation," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Lausanne, Switzerland, Jun. 2021, pp. 1–6.

[30] B. Wang, S. Wang, S. Qiu, W. Wei, H. Wang, and H. He, "CSU-Net: A context spatial U-Net for accurate blood vessel segmentation in fundus images," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 1128–1138, Apr. 2021.

[31] A. K. Jethi, B. Murugesan, K. Ram, and M. Sivaprakasam, "Dual-encoder-Unet for fast mri reconstruction," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. Workshops (ISBI Workshops)*, Iowa City, IA, USA, Apr. 2020, pp. 1–4.

[32] M. K. Abd-Ellah, A. A. M. Khalaf, A. I. Awad, and H. F. A. Hamed, "TPUAR-Net: Two parallel U-Net with asymmetric residual-based deep convolutional neural network for brain tumor segmentation," in *Proc. Int. Conf. Image Anal. Recognit.* Cham, Switzerland: Springer, 2019, p. 11663.

[33] Y. Wang, C. Li, T. Zhu, and J. Zhang, "Multimodal brain tumor image segmentation using WRN-PPNet," *Comput. Med. Imag. Graph.*, vol. 75, pp. 56–65, Jul. 2019.

[34] L. Tan, W. Ma, J. Xia, and S. Sarker, "Multimodal magnetic resonance image brain tumor segmentation based on ACU-net network," *IEEE Access*, vol. 9, pp. 14608–14618, 2021.

[35] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in *Proc. Int. MICCAI Brainlesion Workshop*, 2018, pp. 311–320.

[36] P. Ahmad, H. Jin, R. Alroobaea, S. Qamar, R. Zheng, F. Alnajjar, and F. Aboudi, "MH UNet: A multi-scale hierarchical based architecture for medical image segmentation," *IEEE Access*, vol. 9, pp. 148384–148408, 2021.

[37] R. El Jurdi, C. Petitjean, P. Honeine, V. Cheplygina, and F. Abdallah, "High-level prior-based loss functions for medical image segmentation: A survey," *Comput. Vis. Image Understand.*, vol. 210, Sep. 2021, Art. no. 103248.

[38] R. McKinley, R. Meier, and R. Wiest, "Ensembles of densely-connected CNNs with label-uncertainty for brain tumor segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*, 2018, pp. 456–465.

[39] C. Zhou, C. Ding, X. Wang, Z. Lu, and D. Tao, "One-pass multi-task networks with cross-task guided attention for brain tumor segmentation," *IEEE Trans. Image Process.*, vol. 29, pp. 4516–4529, 2020.

[40] A. Kermi, I. Mahmoudi, and M. T. Khadir, "Deep convolutional neural networks using U-Net for automatic brain tumor segmentation in multi-modal MRI volumes," in *Proc. Int. MICCAI Brainlesion Workshop*, 2018, pp. 37–48.

[41] A. Albiol, A. Albiol, and F. Albiol, "Extending 2D deep learning architectures to 3D image segmentation problems," in *Proc. Int. MICCAI Brainlesion Workshop*, 2018, pp. 73–82.

[42] X. Feng, N. J. Tustison, S. H. Patel, and C. H. Meyer, "Brain tumor segmentation using an ensemble of 3D U-Nets and overall survival prediction using radiomic features," *Frontiers Comput. Neurosci.*, vol. 14, pp. 279–288, Apr. 2020.

YULEI HOU received the B.S. and Ph.D. degrees in mechatronic engineering from Yanshan University, Qinhuangdao, China, in 2002 and 2007, respectively. He is currently a Professor with the School of Mechanical Engineering, Yanshan University. His research interests include computer vision, intelligent robots, complex electromechanical automation systems, artificial intelligence-assisted diagnosis and treatment technology, and medical robots.
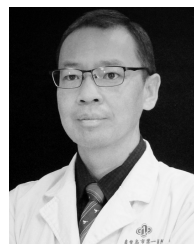
JIANGTAO YAN received the B.S. degree in clinical medicine and the M.S. degree in orthopedics from Hebei Medical University, Shijiazhuang, China, in 2004 and 2007, respectively. He is currently a Chief Physician with the Department of Orthopaedics, First Hospital of Qinhuangdao, China. His research interests include chiropractic and rehabilitation medicine.

YUNJIAO DENG received the B.S. degree in mechanical design, manufacturing, and automation from the North China Institute of Aerospace Engineering, Langfang, China, in 2016. He is currently pursuing the Ph.D. degree in mechatronic engineering with Yanshan University, Qinhuangdao, China. His research interests include deep learning, medical image processing, and computer vision.

DAXING ZENG received the B.S., M.S., and Ph.D. degrees in mechatronic engineering from Yanshan University, Qinhuangdao, China, in 2002, 2004, and 2008, respectively. He is currently a Professor with the School of Mechanical Engineering, Dongguan University of Technology, China. His research interests include theory of mechanism and machine, parallel mechanism, type synthesis, and image processing.

• • •