An Enhanced Residual U-Net for Microaneurysms and Exudates Segmentation in Fundus Images

CAIXIA KOU [ID][1], **WEI LI**[1], **ZEKUAN YU** [ID][2], **AND LUZHAN YUAN**[1]
[1]School of Sciences, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Academy for Engineering and Technology, Fudan University, Shanghai 200433, China

Corresponding author: Zekuan Yu (yzk@fudan.edu.cn)

ABSTRACT Diabetic retinopathy (DR) is a leading cause of visual blindness. However if DR can be diagnosed and treated early, 90% of DR causing blindness can be prevented significantly. Microaneurysms (MAs) and exudates (EXs), as signs of DR, can be used for early DR diagnosis. However, MAs and EXs segmentation is a challenging task due to the low contrast of the lesions, the interference of noises, and the imbalance between the lesion areas and the background. In this paper, an enhanced residual U-Net (ERU-Net) for MAs and EXs segmentation is proposed. ERU-Net obtains three U-paths, which are composed by three upsampling paths together with one downsampling path. With such three U-paths structure, ERU-Net can enhance the corresponding features fusion and capture more details of fundus images. Also, a residual block is constructed in ERU-Net to extract more representative features. In the experiments, we evaluate the performance of ERU-Net for MAs and EXs segmentation on three public datasets, E-Ophtha, IDRiD, and DDR. The ERU-Net obtains the AUC values of 0.9956, 0.9962, 0.9801, 0.9866, 0.9679, 0.9609 for MAs and EXs segmentation on these three datasets, respectively, which are greater than that of the original U-Net. Compared with some traditional methods, convolutional neural networks and other recent U-Nets, ERU-Net also performs competitively. Besides, we have applied ERU-Net to segment optic disc (OD) on the DRISHTI-GS1 dataset, achieving the highest Jaccard index of 0.994 compared with the existing methods. The numerical results indicate that ERU-Net is a promising network for medical image segmentation.

INDEX TERMS U-Net, microaneurysms, exudates, medical image segmentation.
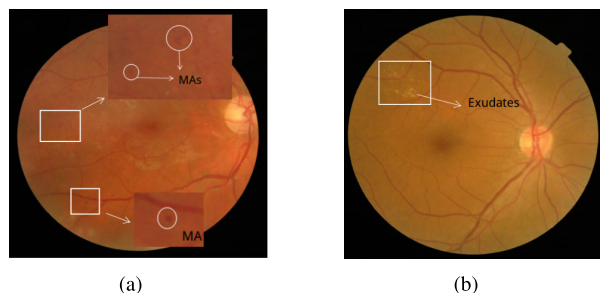
I. INTRODUCTION

Diabetes is a main chronic disease and one of the most serious diseases facing the world's population today. The number of people with diabetes is increasing and is estimated to become 640 million in 2040 [1]. Diabetes can damage blood vessels, nerve cells, and further damage the brain, the heart, and the eyes. In the eye damage caused by diabetes, diabetic retinopathy (DR) is a significant cause of blindness, especially in the working age people [2]. The patients may not have any symptoms of vision problems when in the early stage of DR. If DR can be diagnosed and treated early, 90% of DR causing blindness can be prevented [3]. Thus the early detection of DR is of great significance to human beings.

Fundus images are often used in DR detection to inspect the early signs of DR: microaneurysms (MAs) and exudates (EXs). MAs are the earliest clinical signs of DR caused by

The associate editor coordinating the review of this manuscript and approving it for publication was Chunbo Xiu [ID].

increased blood glucose concentration in diabetic patients. They are small swellings in the retina's tiny blood vessels, which look like small, round red spots in the retinal image (Figure 1(a)). MAs can cause blood vessels to leak in the retina. With the disease developing, lipids and fluids leak from the blood vessels and become EXs. EXs are yellow-white dots in the retinal image and often appear in a ring around the damaged blood vessel with different sizes and sharp edges (Figure 1(b)).

To segment MAs and EXs automatically in fundus images is of great significance for ophthalmologists in early detection of DR. In fact, a large number of people are requiring fundus scans, whereas the number of ophthalmologists is extremely limited. Furthermore, the fundus diagnosis results are much more subjective according to the ophthalmologists' experiences. Thus, an automated MAs and EXs segmentation method not only alleviates the workload of ophthalmologists, but also improves the efficiency and accuracy of diagnosis.

FIGURE 1. (a): Unhealthy retinal fundus images with MAs. (b): Unhealthy retinal fundus images with exudates (EXs for short).

Recently, the deep learning methods have achieved good performance in MA and EX segemention. But there are still some challenges such as: the accurate fine-grained segmentation, the interference of noises, and the poor generalization of the model. To overcome these challenges, in this paper, we develop an enhanced residual U-Net (ERU-Net) to segment MAs and EXs in retinal images with limited ground truth samples. Deep convolution layer can extract more semantic features and global features, but this does not mean that the last layer is the best feature expression. In particular, for MAs and EXs segmentation, the fine-grained segmentation tasks, more local features are needed. The features fusion can enhance the feature representation. Inspired by this, we designed the structure of ERU-Net as follows: Three upsampling paths combined with one down-sampling path compose three U-paths in the network. At each downsampling step, ERU-Net combines the high-resolution features into the corresponding upsampling features in the first U-path. Then the combined features are combined into the corresponding upsampling features in the second U-path, and so on in the third U-path. The fusion operation between the U-paths is connection. With its unique structure, ERU-Net can enrich the corresponding fusion feature maps in the upsampling path. The experiment results show that the proposed ERU-Net can capture more details and segment MAs and EXs more accurately.

The remainder of this paper is organized as follows. In Section 2, a brief review of different MAs and EXs segmentation methods are presented. In Section 3, the proposed ERU-Net is introduced in detail. The experiment results and discussion are presented in Section 4 and 5, respectively. We end the paper in Section 6 with a short conclusion.

II. RELATED WORK

Automatic MAs and EXs segmentation has been widely studied. These segmentation methods mainly include two categories: traditional image processing methods and recent deep learning methods. The traditional image processing methods mostly use morphological methods, threshold methods and classifier based methods. In the morphological methods [4]–[8], in order to reduce the interference with EXs and MAs, one often need to first identify and remove the dominant structures of images, such as blood vessels, optic

dice from the images. The threshold methods are based on global or local image gray-level and clustering according to the image threshold. Most classifiers based methods apply different classifiers to the candidate of the lesions, such as [9]–[13]. For different segmentation tasks, the traditional image processing methods are different.

For MAs segmentation, traditional image processing methods mainly include morphological processing methods, wavelet transformation methods and hybrid classifier methods. For example, in fluorescein angiograms of the ocular fundus, Spencer *et al.* [4] used morphological operations to remove the vasculature, leaving the other small structures including the MAs. Quellec *et al.* [14] proposed an adaptive wavelet method which uses a local template matching in the wavelet domain to detect MAs. Mizutani *et al.* [9] proposed a double-loop filter first to extract MAs candidate regions. Then, from such the MAs candidate regions, they used a rule-based method and an artificial neural network to detect MAs. Akram *et al.* [10] proposed a hybrid classifier which is combined with the Gaussian mixture model, support vector machine and an extension of multimodel mediod.

For EXs segmentation, traditional methods have been developed, including threshold method, morphology method, region growing, support vector machine (SVM) and Naive Bayes classifier. Philips *et al.* [15] used the global threshold techniques in fundus images. Walter *et al.* [5] applied the morphological reconstruction techniques to locate candidates of EXs. Other methods based on morphological operations can be found in [6]–[8]. Li and Chutatape [16] combined the region growing and the edge detection to segment EXs. Giancardo *et al.* [11] applied a support vector machine (SVM) to detect EXs. Harangi *et al.* [12] proposed a Naive Bayes classifier to candidate pixels, achieving a F-score of 0.72 on the DIARETDB1 dataset. Zhang *et al.* [13] used a random forest algorithm to detect EXs among the candidates, obtaining an AUC value of 0.95 on the E-ophtha-EX dataset.

Different from traditional image processing methods, the deep learning method, mainly uses a deep network to complete the segmentation tasks. And such deep networks can extract the useful image features automatically and then be trained end to end. With the rapid development of deep learning, neural networks are frequently seen in MAs and EXs segmentation. For example, Mrinal [17] used a deep neural network including three convolutional layers and two fully connected layers to segment MAs automatically. Chudzik *et al.* [18] applied a fully convolutional neural network (FCNN) with the batch normalization (BN) layers and the Dice coefficient loss function to detect MAs. Kou *et al.* [19] proposed a deep residual U-Net to segment MAs, which combines a deep residual model and recurrent convolutional operations into U-Net. For EXs segmentation, Perdomo *et al.* [20] utilized LeNet [21], a convolutional neural network, to classify EXs and non-EXs accurately. Fujita *et al.* [22] used a single convolutional neural network to detect EXs, hemorrhages, and MAs simultaneously. Feng *et al.* [23] applied a fully convolutional neural network

together with a residual operation to segment both optic disc and EXs. Zheng *et al.* [24] proposed an ensemble convolutional neural network (MU-Net) to detect EXs. They used the conditional generative adversarial network (cGAN) to implement the data augmentation, which aims to alleviate the imbalanced data problem.

U-Net, as a fully convolutional neural network [25], has become a popular medical image segmentation method and has shown its excellent performance [26]–[28]. Alom *et al.* [29] applied a R2U-Net, an U-Net combining with recurrent convolutional operations, on the blood vessel segmentation, on the skin cancer segmentation and on the lung lesion segmentation. Sevastopolsky [30] proposed a modified U-Net to segment the optic disc and the cup in fundus images, obtaining the Dice value of 0.95 and 0.82, respectively, on the RIM-ONE v.3 dataset. Zeng *et al.* [26] applied the residual blocks, the multi-scale, and the channel attention mechanism on U-Net for the nuclei segmentation. This method obtains an average F1-score of 0.83 on the Cancer Genomic Atlas (TCGA) dataset and wins the third place in the computational precision medicine nuclei segmentation challenge. The U-Net-like models are also applied to the salient object detection [31]–[33]. Wang *et al.* [31] have used the RNN unit to form an iterative top-down and bottom-up saliency inference model, which can make better use of the high-level and the low-level information. Wang *et al.* [32] proposed a Pyramid Attentive and salient edGE-aware saliency model (PRET-Net) with pyramid attention structure and a salient edge detector. It efficiently increases the receptive field of the convolution layer. In [33], the authors designed an Attentive Saliency Network (ASNet) by using the recurrent architecture of the convolutional LSTM (convLSTM) and a fixation map.

III. PROPOSED METHODOLOGY

In this section, after a brief introduction of the U-Net and the residual U-Net, we introduce our new network, an enhanced residual U-Net (ERU-Net), which is obtained by adding a residual block into the modified U-Net with one downpsampling path and three upsampling paths. In order to achieve high accuracy, a preprocessing of image segmentation is also introduced in this section.

U-Net is originally proposed by Ronneberger *et al.* [34] for the biomedical image segmentation. It has a downsampling (encoding) path and a symmetric upsampling (decoding) path, as shown in Figure 2. The feature extration processes are conducted in the downsampling path, and the image expansion processes are conducted in the upsampling path. Each upsampling output is fused with the corresponding layers in the downsampling path, which can combine more feature dimensions and supplement the information lost during the downsampling process.

The residual network [35] is a type of specialized neural network that helps to handle more sophisticated deep learning models. Inspired by the residual network, a modified U-Net with residual blocks has been widely applied to segmentation
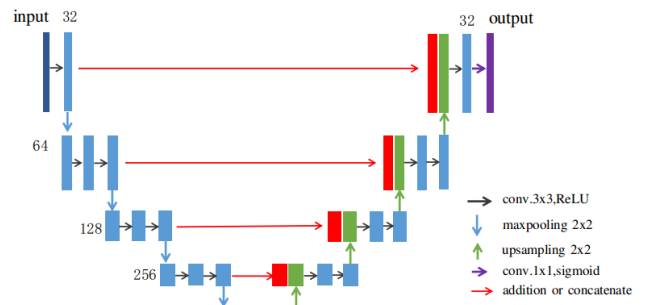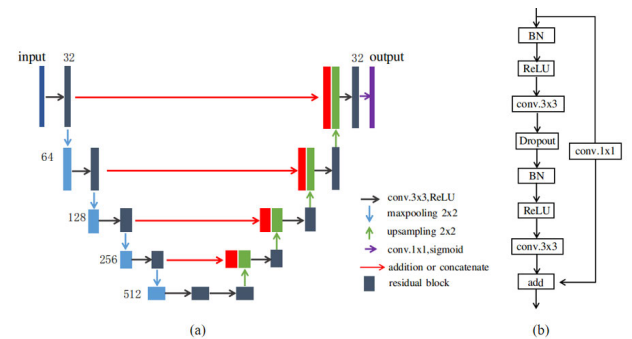

FIGURE 2. The U-Net architecture.

FIGURE 3. (a) ResU-Net architecture. (b) Residual block.

tasks ([23], [29], [36]). Experimental results show that using residual blocks can improve the performances of the U-Net type methods.

Inspired by this, we replace the regular convolutional layers of U-Net with residual blocks to obtain a ResU-Net. The residual block as shown in Figure 3(b) includes a shortcut and a few stacked layers: convolutional layers, rectified linear unit (ReLU) layers, batch normalization (BN) layers [37]. Moreover, we add a dropout [38] layer between the two convolutional layers to avoid overfitting. The residual block is expressed as follows:

$$X_{drop}(i) = D\left(F\left(A\left[B\left(X(i)\right)\right]\right)\right), \quad (1)$$
$$X(i+1) = F\left(A\left[B\left(X_{drop}(i)\right)\right]\right) + I\left(X(i)\right), \quad (2)$$

where $X(i)$, $X(i+1)$ are the input and the output of the i-th residual block. And $X_{drop}(i)$ is the output of the dropout layer in the i-th residual block. $B(\cdot)$ denotes the BN function, $A(\cdot)$ is the activation function, $D(\cdot)$ is the dropout operation, and $F(\cdot)$ is the convolutional operation. $I(\cdot)$ is an identity mapping function.

The proposed ResU-Net includes 10 above residual blocks and a sigmoid activation function as its last layer which is used to get a segmentation probability distribution map. Compared with the ResU-Net proposed by Zhang *et al.* [36], we use a dropout layer between the two convolutional layers in the residual block which can effectively avoid overfitting. In addition, we construct the ResU-Net with a 9-level architecture, a deeper network than the ResU-Net in [36].
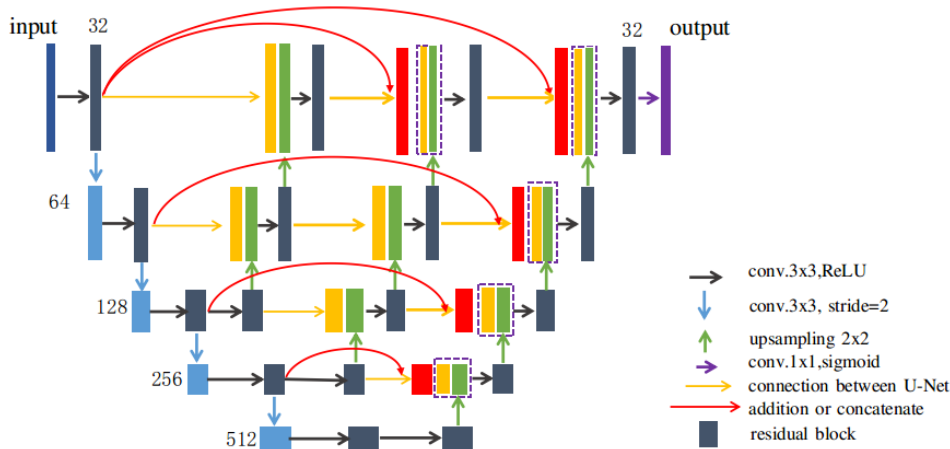
FIGURE 4. An overview of the proposed ERU-Net architecture.

TABLE 1. Configuration of the ERU-Net.

Downsampling path	Output	Upsampling path 1th	Output	Upsampling path 2th	Output	Upsampling path 3th	Output
Conv1_1 32 residual block	48x48	Conv1_2 32 residual block	48×48	Conv1_3 32 residual block	48x48	Conv1_4 32 residual block	48x48
Pool1 Conv2D 3x3 stride=2	24x24	concate(Up2_2,Conv1_1) Up2_2 Upsampling2D 2x2	48x48	add(Conv1_1,Conv2_3m) Conv2_3m 32 Conv2D 3x3 concate(Up2_3,Conv1_2) Up2_3 Upsampling2D 2x2	48x48	add(Conv1_1,Conv2_4m) Conv2_4m 32 Conv2D 3x3 concate(Up2_4,Conv1_3) Up2_4 Upsampling2D 2x2	48x48
Conv2_1 64 residual block	24x24	Conv2_2 64 residual block	24x24	Conv2_3 64 residual block	24x24	Conv2_4 64 residual block	24x24
Pool2 Conv2D 3x3 stride=2	12x12	concate(Up3_2,Conv2_1) Up3_2 Upsampling2D 2x2	24x24	concate(Up3_3,Conv2_2) Up3_3 Upsampling2D 2x2	24x24	add(Conv2_1,Conv3_4m) Conv3_4m 64 Conv2D 3x3 concate(Up3_4,Conv2_2) Up3_4 Upsampling2D 2x2	24x24
Conv3_1 128 residual block	12x12	Conv3_2 128 residual block	12x12	Conv3_3 128 residual block	12x12	Conv3_4 128 residual block	12x12
Pool3 Conv2D 3x3 stride=2	6x6			concate(Up4_3,Conv3_2) Up4_3 Upsampling2D 2x2	12x12	add(Conv3_1,Conv4_4m) Conv4_4m 128 Conv2D 3x3 concate(Up4_4,Conv3_3) Up4_4 Upsampling2D 2x2	12x12
Conv4_1 256 residual block	6x6			Conv4_3 256 residual block	6x6	Conv4_4 256 residual block	6x6
Pool4 Conv2D 3x3 stride=2	3x3					add(Conv4_1,Conv5_4m) Conv5_4m 256 Conv2D 3x3 concate(Up5_4,Conv4_3) Up5_4 Upsampling2D 2x2	12x12
Conv5_1 512 residual block	3x3					Conv5_4 512 residual block	3x3

During the training process, Cross Entropy was used as the loss function, which is defined as:

$$L = -\frac{1}{N}\sum_{1}^{N} y_i log p_i + (1 - y_i)log(1 - p_i), \quad (3)$$

where N is the pixels number of the input image and y_i denotes the ground truth of the i-th pixel. And p_i is the predicted probability value obtained by the network.

Furthermore, in order to obtain more accurate segmentation results, we propose an enhanced ResU-Net, called ERU-Net, as shown in Figure 4. The detailed configuration of the ERU-Net is listed in Table 1. Different from the U-Net and ResU-Net, ERU-Net has following three features: 1) Compared with the U-Net with one upsampling path, ERU-Net includes one downsampling path but three upsampling paths. Designing three upsampling paths aims at reducing the semantic gap between the feature maps of the downsampling paths and the upsampling paths. When the output layer of the upsampling and downsampling process shares similar feature maps, the network can achieve better feature representation capability. In addition, when high-resolution feature maps from the downsampling path are gradually enriched before fusion with the corresponding feature maps, ERU-Net with three upsampling paths can more effectively capture the details of foreground objects and better entrich the above high-resolution feature maps. The three upsampling operations start at the points when the number feature maps of the downsampling path is 128, 256 and 512, respectively. The outputs of three upsampling paths are fused with the given weight values and then this fused result is input into a residual block, a full connected layer and a sigmoid function to get the result. The connection operation between the different upsampling paths makes the downsampling path feature maps be close to the concatenation layer on the third upsampling path. 2) Instead of regular convolutional layers,

a residual block shown in Figure 3(b), is embedded ERU-Net to accumulate the features effectively. Each residual block in the upsampling path is preceded by a concatenation layer which fuses the feature channels. In more detail, this concatenation layer is the fusion between the output of the residual block on the previous upsampling path and the output of the previous step in the same upsampling path. Such feature fusion in different upsampling paths can extract more local features to help ERU-Net achieve a more accurate performance. Formally, the output of the *i*-th residual block in the last upsampling can be expressed as:

$$X_{u_3,i} = H(X_{d,i}, X_{u_2,i}, u(X_{u_3,i+1})), \quad (4)$$

where the index *i* denotes the *i*-th layer, the index *d* and *u* represent the downsampling and upsampling path, respectively, the index u_1, u_2, and u_3 represent the three unsampling paths. The function $u(\cdot)$ denotes an upsampling operation, $H(\cdot)$ is a residual learning, and *X* denotes the output of a layer. 3) In the downsampling path, we apply a convolution operation with a stride of 2 to replace the max-pooling operation in U-Net to preserve more details. After the downsampling, the number of feature maps increases to 512.

Before the original fundus images are fed into the ERU-Net, original images needs to be preprocessed to obtain the most obvious images. The fundus images with different resolutions which they are obtained from different types of fundus cameras are resized to 1024 × 1024 pixels to reduce image variation. The green channel of RGB fundus image is used for further processing because the green channel has the best contrast from the background. When the contrast of the captured image is too low, it is difficult to detect and identify the interest objects. Therefore, image enhancement is a very important step before the learning process. We apply Histogram Equalization and Contrast limited adaptive histogram equalization (CLAHE) which are two classical approaches for image enhancement, to achieve the most obvious contrast images.

IV. EXPERIMENT RESULTS
In this section, we compare the proposed ERU-Net with U-Net and ResU-Net on three public datasets: E-ophtha [13], IDRiD [39], DDR [40]. Also, we compare ERU-Net with other existing methods on the E-ophtha and IDRiD datasets.

A. DATASETS
The E-ophtha dataset [13] is a publicly accessible digital retinal fundus image dataset and is used for lesion segmentation. The dataset is divided into two parts, E-ophtha-EX and E-ophtha-MA with four different image sizes ranging from 1440 × 960 to 2544 × 1696 pixels. All the images are carefully marked by two ophthalmologists on the pixel level. On the E-ophtha-MA dataset, there are 381 images, 148 of which contain MAs. At random, 318 images (133 with and 185 without MAs) are used as the training images, and the other images are as the testing images. On the E-ophtha-EX dataset, there are 82 images in total, 47 of which contain EXs.

Also, at random, we use 60 images (35 with and 25 without EXs) as the training images, and the other 22 images (12 with and 10 without EXs) are used as the testing images.

The IDRiD dataset (Indian Diabetic Retinopathy Image Dataset) [39] is a retinal fundus image dataset which is available for the segmentation and grading challenge of DR, organized by ISBI (International Symposium on Biomedical Imaging) conference 2018. The fundus images with signs of DR in IDRiD are captured by an ophthalmic technician at an Eye Clinic using a Kowa VX-10 alpha digital fundus camera with 50-degree field of view. IDRiD provides precise pixel level annotation of abnormalities associated with DR like microaneurysms (MAs), soft exudates (SE), hard exudates (EX) and hemorrhages (HE). The pixel level annotation is done by a master's student and reviewed by two retinal specialists. This dataset consists of 81 color fundus images with their labeled ground truth which have a resolution of 4288 × 2848 pixels. The 81 color fundus images, corresponding 81 labeled ground truth of MAs and 81 labeled ground truth of EXs are applied to our experiment. For MAs and EXs segmentation, 54 images are used for training and 27 images are used for testing.

The DDR dataset [40] proposed in 2019 is a newly fundus image dataset for DR screening. DDR is the largest dataset that provides pixel level and bounding-box annotations for MAs, soft exudates, EXs, and haemorrhages. The pixel-level annotations for the four lesions are performed by six annotators using special software. This dataset consists of 757 color fundus images with their pixel level annotations. At random, 532 images are used for training and 225 images are used for testing. All the images are resized to 1024 × 1024 pixels.

B. EVALUATION METRICS
The segmentation task is evaluated on the pixel level by the indexes of the sensitivity (Se), the specificity (Sp) and the accuracy (Acc). These indexes are computed by the following equations:

$$Se = \frac{TP}{TP + FN}, \quad (5)$$

$$Sp = \frac{TN}{TN + FP}, \quad (6)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \quad (7)$$

where, TP (True positive) represents the number of positive pixels correctly classified by the classifier. TN (True negative) is the number of negative pixels which are correctly classified as negative. FP (False positive) and FN (False negative) are the number of negative pixels and positive pixels that are misclassified, respectively.

A threshold is defined by seeking the maximum value of $Se + Sp - 1$ to convert the last probability map into a binary image. Then this threshold is used to get the final binary results. Furthermore, the receiver operating characteristics (ROC) curve and the value of the area under curve (AUC) are used to evaluate the segmentation
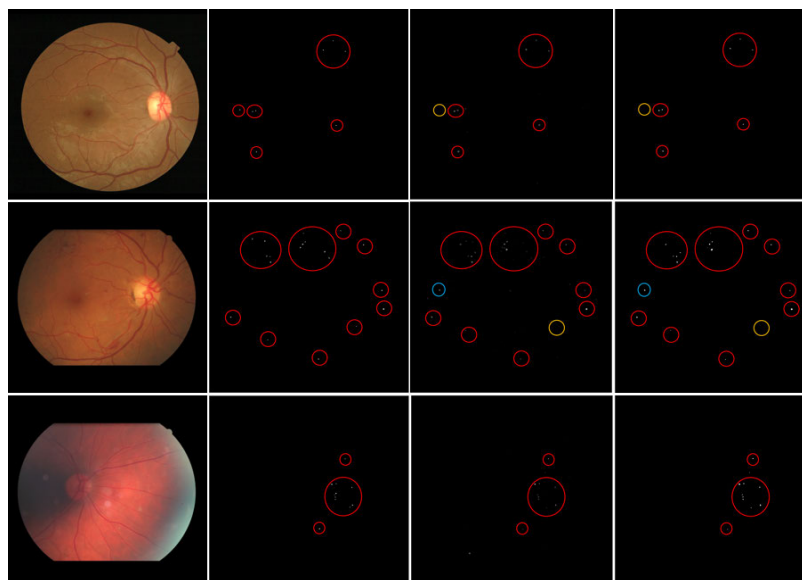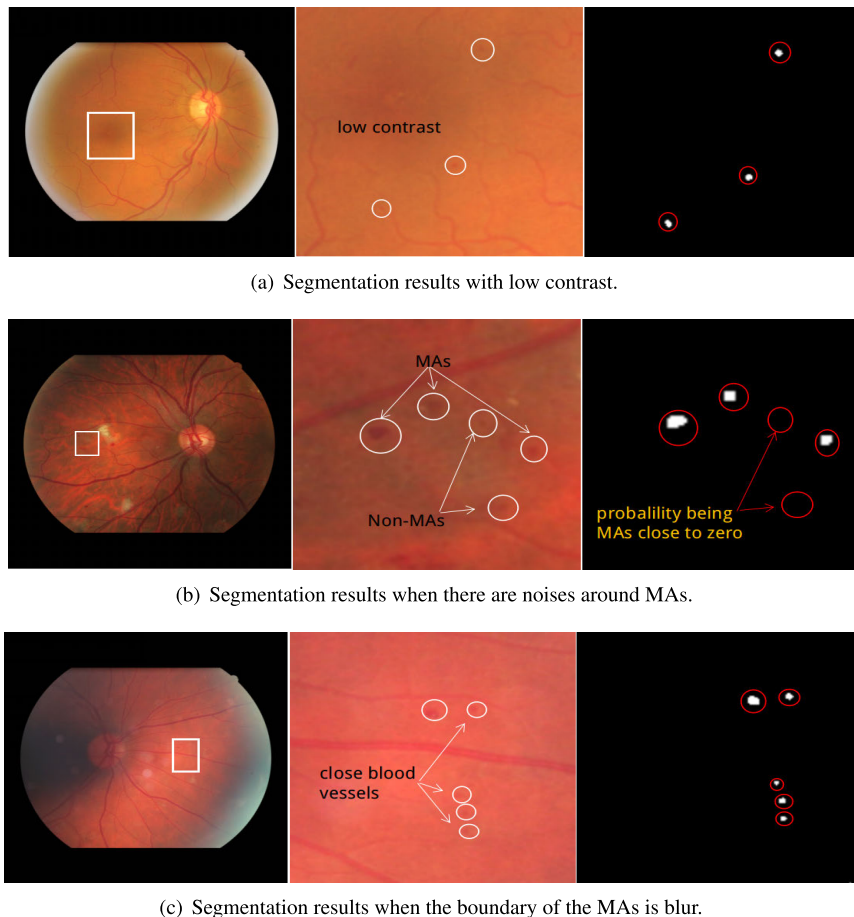
(a)
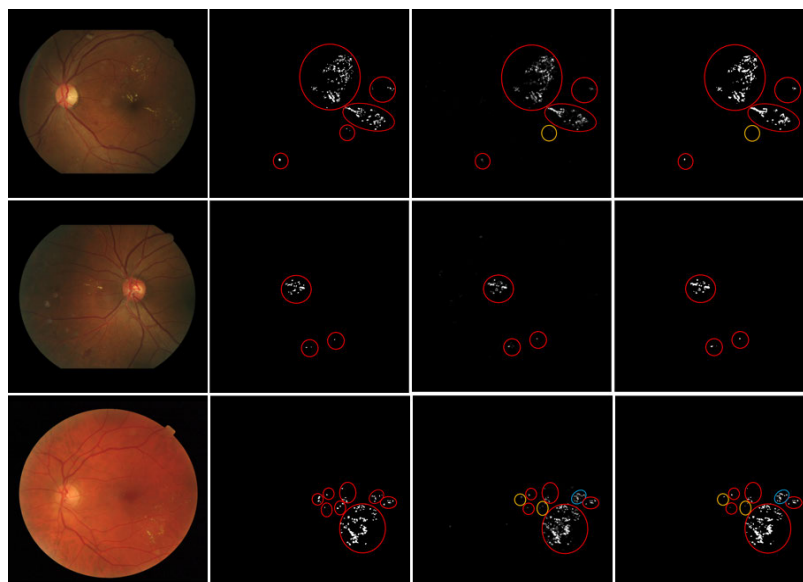
(b)

FIGURE 5. (a) and (b) show the segmentation results of ERU-Net on MAs and EXs of E-ophtha, respectively. The first to fourth columns denote the retinal image, the ground truth, the probability map of segmentation results, the binary map of segmentation results, respectively. (Red circles: true positives; yellow circles: false negatives; blue circles: false positives).

results. The AUC value is an important indicator for evaluating segmentation results because it is independent of the threshold, which is different from Se, Sp, and Acc.

C. THE PERFORMANCE OF THE PROPOSED METHOD
All the experiments are implemented using the public available Keras framework with TensorFlow on a single GPU machine with 256 G of RAM and a NIVIDIA GeForce GTX 1080Ti. For all the training, each model is trained 150 epochs, with Cross Entropy as the loss function.

We use adamoptimizer as optimizer with a momentum of 0.9, a weight decay of 0.0005, and an initial learning rate of 0.001. Training process is carried out in mini batches with the size of 96. The images in the E-ophtha training datasets are divided into 343,440 patches with 48×48 pixels for training, and 38,160 patches for validation in order to reduce the memory shortage. And all the training images on the IDRiD dataset are divided into 305,280 patches for training, and 38,160 patches for validation. On the DDR dataset, the training images are split to 340,906 patches for training, and 37,878 patches for validation.

(a) Segmentation results with low contrast.

(b) Segmentation results when there are noises around MAs.

(c) Segmentation results when the boundary of the MAs is blur.

FIGURE 6. The results of ERU-Net on the MAs segmentation. (a) Segmentation results under the low contrast condition. (b) Segmentation results when there are noises around MAs. (c) Segmentation results when the boundary of the MAs is blur.
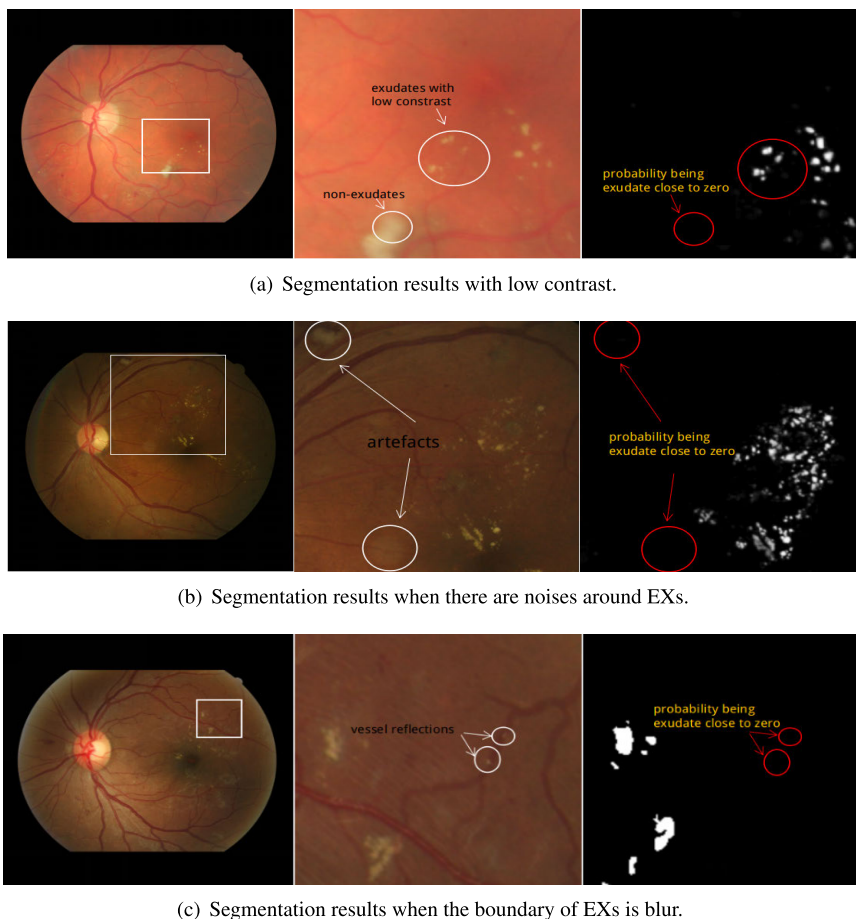
1) VISUAL RESULTS

Figure 5(a) and (b) show the visual results of ERU-Net on the E-ophtha-MA and E-ophtha-EX datasets, respectively. In both Figure 5(a) and (b), the first column denotes the retinal images, the second column is the ground truth, and the third is the probability map of segmentation results of ERU-Net. After converting this probability map into a binary map, we get the final segmentation results of ERU-Net, which is presented in the fourth column. Visually, the fourth column is very close to the second column which means that the segmentation results of ERU-Net are quite close to the ground truth.

Figure 6 and Figure 7 show the segmentation results of ERU-Net in some lesion regions with poor segmentation results, such as the regions with the low contrast, blur edge, illumination noises. In these two figures, the first column is the fundus image with MAs or EXs. The second column is the enlarged field of view images in the fundus images. The third column is the segmentation results with ERU-Net. Figure 6(a) shows the results by using ERU-Net to segment MAs in the images with low contrast. From the third column, it can be found that ERU-Net classifies the MAs and the background

correctly. Figure 6(b) presents the results by using ERU-Net to segment MAs in the images with noises in which there are some similar structures around MAs. The results of the third column illustrate that ERU-Net can segment real MAs with high accuracy even if MAs are around by some similar structures. Figure 6(c) shows that ERU-Net can also segment MAs effectively in the images where the boundary of MAs is blur. Similar performance of ERU-Net can also be found in Figure 7, which indicates that ERU-Net also segments EXs effectively in some challenging regions. From the visual results in Figure 6 and 7, ERU-Net can still achieve good performance in MAs and EXs segmentation in the fundus images with low contrast, or with blur edge, or with ilumation noises. It is mainly due to concating one downsampling path and three upsamling paths with ResU-Net, which can allow ERU-Net to capture more local feature details and surpass the noise interference.

2) COMPARISON WITH OTHER MODELS

Table 2-7 show the performance comparisons of ERU-Net, ResU-Net with U-Net for MAs and EXs segmentation on three datasets: the E-ophtha, the IDRiD, and the DDR dataset

(a) Segmentation results with low contrast.

(b) Segmentation results when there are noises around EXs.

(c) Segmentation results when the boundary of EXs is blur.

FIGURE 7. The results of ERU-Net on the EXs segmentation. (a) Segmentation results under the low contrast conditon. (b) Segmentation results when there are noises around EXs. (c) Segmentation results when the boundary of EXs is blur.

TABLE 2. Performance comparisons of U-Net, ResU-Net, and ERU-Net for MAs segmentation on E-ophtha.

Model	Se	Sp	Acc	AUC
U-Net [34]	0.766	0.792	0.99	0.8323
ResU-Net	0.970	0.966	0.99	0.9938
ERU-Net	**0.974**	**0.973**	0.99	**0.9956**

TABLE 3. Performance comparisons of U-Net, ResU-Net and ERU-Net for EXs segmentation on E-ophtha.

Model	Se	Sp	Acc	AUC
U-Net [34]	0.967	0.971	0.99	0.9931
ResU-Net	0.966	0.967	0.99	0.9941
ERU-Net	**0.983**	**0.976**	0.99	**0.9962**

TABLE 4. Performance comparisons of U-Net, ResU-Net and ERU-Net for MAs segmentation on IDRiD.

Model	Se	Sp	Acc	AUC
U-Net [34]	0.762	0.733	0.98	0.8153
ResU-Net	0.906	0.912	0.98	0.9674
ERU-Net	**0.929**	**0.935**	0.98	**0.9801**

TABLE 5. Performance comparisons of U-Net, ResU-Net and ERU-Net for EXs segmentation on IDRiD.

Model	Se	Sp	Acc	AUC
U-Net [34]	0.945	0.934	0.99	0.9814
ResU-Net	0.944	**0.940**	0.99	0.9841
ERU-Net	**0.951**	0.939	0.99	**0.9866**

in terms of four indexes: Se, Sp, Acc, and AUC. Table 2, 4, 6 are for MAs segmentation. Table 3, 5, 7 are for EXs segmentation. As shown in these six tables, the AUC value of ERU-Net for MAs and EXs segmentations on three datasets are 0.9956, 0.9962, 0.9801, 0.9866, 0.9679, 0.9801, 0.9866, 0.9609, respectively. It is clear that ERU-Net obtains the highest AUC value among these three models, ResU-Net achieves the second place, and U-Net has the lowest AUC value. In fact, on the one hand, compared with U-Net, the ResU-Net is designed by residual blocks replacing the regular convolutional layers

of U-Net. The comparison results indicate that adding residual blocks can improve U-Net's performance. On the other hand, compared with ResU-Net, ERU-Net mainly applies three upsampling in the U-path. The comparison results show such 3-upsampling-path structure in ERU-Net can capture more detail features, which indeed contributes to enhance the U-Net's performance.

To illustrate why the number of upsampling paths in ERU-Net is 3, we do an ablation experiment on such the number,

TABLE 6. Performance comparisons of U-Net, ResU-Net and ERU-Net for MAs segmentation on DDR.

Model	Se	Sp	Acc	AUC
U-Net [34]	0.724	0.772	0.77	0.8073
ResU-Net	**0.899**	0.930	0.93	0.9675
ERU-Net	0.898	**0.931**	0.93	**0.9679**

TABLE 7. Performance comparisons of U-Net, ResU-Net and ERU-Net for EXs segmentation on DDR.

Model	Se	Sp	Acc	AUC
U-Net [34]	0.848	0.910	0.91	0.9407
ResU-Net	0.864	0.919	0.92	0.9572
ERU-Net	**0.871**	**0.926**	**0.93**	**0.9609**

TABLE 8. Performance of ERU-Net with different number ($t = 2, 3, 4$) of upsampling paths on E-ophtha.

Dataset	t	Se	Sp	AUC
E-ophtha-EX	2	0.972	0.976	0.9950
	3	0.983	0.976	0.9962
	4	9.966	0.961	0.9920
E-ophtha-MA	2	0.972	0.977	0.9960
	3	0.974	0.973	0.9956
	4	0.937	0.964	0.9860

TABLE 9. The number of trainning parameters in U-Net, ResU-Net and ERU-Net.

Model	Total parameters (million)
U-Net	7.7m
ResU-Net	11.8m
ERU-Net	10.2m

denoted as t in Table 8. We adapt ERU-Net with different numbers of upsampling paths, as $t = 2, 3$ and 4. We evaluate the AUC values of ERU-Net with different number of upsampling paths on the E-ophtha dataset to segment MA and EX, respectively. As shown in Table 8, for EXs segmentation, ERU-Net with three upsampling paths ($t = 3$) achieves the highest AUC value of 0.9962. For MAs segmentation, ERU-Net with three upsampling paths achieves the second place with the AUC value of 0.9956, which is a slightly lower than the first place with the AUC value of 0.9960, corresponding to $t = 2$. Therefore, taken together, ERU-Net with three upsampling paths gets the best performance.

In Table 9, we calculate the number of the parameters which need to be trained by different models. Generally, the larger number of such training parameters, the longer training time is needed for models. Therefore, the model with as few training parameters as possible is significant to the segmentation tasks. Table 9 shows that ERU-Net needs much less training parameters than ResU-Net, which is considered as a promising feature for ERU-Net. Table 10 shows the runtime of different U-Net models. The runtime showed in Table 10 is the time required by each model to segment one image in the given dataset. It can be found that U-Net takes the least runtime while ResU-Net and ERU-Net, at the cost of some runtime, obtain higher AUC value than U-Net. The AUC values of ERU-Net, ResU-Net, and U-Net on E-ophtha and IDRiD datasets are 0.9956, 0.9938, 0.8323, 0.9801, 0.9674, 0.8153, respectively. The runtime on the IDRiD dataset takes a long time because of its high resolution.

TABLE 10. The runtime for an image in U-Net, ResU-Net and ERU-Net.

Dataset	Image Size	Model	Runtime/s
E-ophtha-MA	1024×1024	U-Net	3.6
	1024×1024	ResU-Net	4.8
	1024×1024	ERU-Net	4.7
IDRiD-MA	4288×2848	U-Net	32.9
	4288×2848	ResU-Net	42.8
	4288×2848	ERU-Net	37.3

But the ERU-Net takes less runtime and obtains higher AUC value than ResU-Net on the E-ophtha and IDRiD datasets.

TABLE 11. Performance comparisons of ERU-Net with different models for MAs and EXs segmentation on E-ophtha.

Dataset	Method	Se	Sp	AUC
E-ophtha-MA	Lam et al. [41]	-	-	0.94
	Khojasteh et al. [42]	0.85	0.96	0.94
	Costa et al. [43]	-	-	0.9712
	Kou et al. [19]	0.964	0.963	0.9943
	ERU-Net	**0.974**	**0.973**	**0.9956**
E-ophtha-EX	Imani et al. [44]	0.803	0.998	0.937
	Zhang et al. [13]	-	-	0.95
	Chudzik et al. [45]	0.846	0.999	0.967
	Chudzik et al. [46]	0.867	**0.999**	0.982
	Mo et al. [47]	0.923	-	0.9647
	Guo et al. [48]	0.8417	-	-
	ERU-Net	**0.983**	0.976	**0.9962**

In Table 11, we compare ERU-Net with other published methods, the convolutional neural networks [19], [41]–[43], [45], [47], [48] and traditional image processing methods [13], [44], [46]. From the table, it is found that on the E-ophtha-MA dataset, the ERU-Net achieves the highest AUC value of 0.9956 among all the compared methods in [19], [41]–[43]. On the E-ophtha-EX dataset, ERU-Net has the highest AUC value of 0.9962 among the ones in [13], [44]–[48]. Such numerical results illustrate that ERU-Net indeed achieves satisfactory performance. This achievement mainly benefits from its unique three U-paths structure. Such multiple U-paths obtained by one downsampling and three upsampling, help the network to capture more details. For such EXs and MAs, the fine-grained segmentation, capturing more details is indeed helpful for improving segmentation accuracy.

We also compare ERU-Net with other published deep learning methods on the IDRiD dataset for EXs segmentation. The literature [49] uses a classical U-Net model to segment EXs and optic disk, then the optic disk segmentation map is used to refine EXs segmentation results. Gupta *et al.* [50] apply the transfer learning method to segment EXs, which uses a VGG net to extract the features and apply different classifiers to obtain the segmentation results. Different from them, we employ the U-type net with three U-path directly to segment EXs, which can extract more detail features and has advantages in fine-grained segmentation task.

In order to better demonstrate the performance of the ERU-Net, many improved U-Net models are used to compare in our study. The ERU-Net can be viewed as an improved ResU-Net architecure, in which main difference from the original U-Net is that there are one downsampling path and three upsampling paths. Meanwhile, the stacked hourglass networks [27] are

TABLE 12. Performance comparisons of a single hourglass, SAT, and ERU-Net for EXs segmentation.

Dataset	Model	Se	Sp	Acc	AUC
E-ophtha-EX	Hourglass network	0	**1**	0.97	0.4494
	SAT	0.977	0.977	**0.98**	**0.9955**
	ERU-Net	**0.983**	0.973	0.97	0.9950
IDRiD-EX	Hourglass network	0.929	0.920	0.92	0.9747
	SAT	**0.948**	0.921	0.92	0.9785
	ERU-Net	0.937	**0.934**	**0.93**	**0.9789**
DDR-EX	Hourglass network	0	**1**	0.99	0.4114
	SAT	0.873	0.911	0.91	0.9557
	ERU-Net	**0.876**	0.916	**0.92**	**0.9586**

TABLE 13. The 5-fold cross validation experiment results of ERU-Net on the E-ophtha and IDRiD datasets for MAs and EXs segmentation.

Dataset	Se	Sp	Acc	AUC
E-ophtha-EX	95.96% ±1.77%	96.80% ±0.76%	96.74% ±0.77%	99.23% ±0.30%
E-ophtha-MA	94.90% ±2.10%	97.20% ±1.48%	97.00% ±1.53%	98.78% ±0.97%
IDRiD-EX	93.00% ±1.11%	93.20% ±6.11%	95.00% ±2.48%	97.13% ±2.36%
IDRiD-MA	91.78% ±2.10%	94.10% ±0.42%	94.20% ±0.56%	97.80% ±1.03%

made of several hourglass modules, in which the single hourglass module can be viewed as a modified U-Net. The stacked dense U-Nets [28] consist of two scale aggregation topology (SAT). One SAT also can be viewed as a modified U-Net. Thus, for a fair comparison, we compared ERU-Net with the single form of stacked hourglass networks and the single form of stacked dense U-Nets, or hourglass network and SAT for short, respectively.

These three networks are tested on three datasets. Because of the large memory requirement of SAT, we divided the training images into 127,200, 114,480, 126,084 patches on three datasets, respectively, which is different from the number in our manuscript. From the results in Table 12, it can be found that hourglass network failed in EXs segmentation. It is reasonable because the hourglass network is designed for human pose estimation, not for fine-grained segmentation tasks like EXs segmentation. The AUC value of SAT is similar to ERU-Net, but it consumes three times as much training time as ERU-Net. Similar results also can be found in the IDRiD and DDR datasets.

V. DISCUSSION
In order to test the reliability and stability of ERU-Net, we conduct 5-fold cross-validation experiments. For a given dataset, all the images are divided into 5 subsets with equal size at random. For each experiment condition, we conduct 5 experiments and ensure that each subset is used as a test set and the rest as the training set. The Table 13 shows the performance results of ERU-Net in 5-fold cross-validation experiments. The AUC expectation value of ERU-Net for MAs and EXs segmentation on the E-ophtha and IDRiD datasets are 0.9878, 0.9923, 0.9780, 0.9713, respectively. It can be found that in view of the average, ERU-Net also obtains a competitive performance.

In order to further test the robustness of ERU-Net for the segmentation tasks, we also apply ERU-Net to segment the optic disc(OD) on the DRISHTI-GS1 dataset.

The DRISHTI-GS1 dataset [51], as a publicly available dataset, contains 101 retinal images with a resolution of 2045 × 1752 pixels. The 101 images are divided into 50 images for training and 51 images for testing with their corresponding binary mask as the ground truth, respectively. All the images are resized to 256 × 256.

TABLE 14. Performance comparison of OD segmentation methods on the DRISHTI-GS1 dataset.

Method	Acc	Se	Sp	JACC	AUC
U-Net	0.98	0.98	0.96	0.98	0.996
ResU-Net	0.98	0.98	0.98	0.990	0.991
Zilly et al. [52]	-	-	-	0.914	-
Kumar et al. [53]	0.98	0.98	0.99	0.96	-
Al-Bander et al. [54]	**0.9969**	0.92	0.99	0.904	-
ERU-Net	0.995	**0.98**	**0.99**	**0.994**	**0.998**

Table 14 presents the comparisons of U-Net, ResU-Net, ERU-Net with other neural networks [52]–[54] on the DRISHTI-GS1 dataset in terms of Acc, Se, Sp, AUC, and Jaccard index (JACC). The JACC, calculated as $TP/(TP + FN + FP)$, is used to evaluate the similarity between the prediction and the ground truth. From Table 14, we can know that ERU-Net has the highest JACC value. It can illustrate that compared with other methods, the OD segmentation in ERU-Net is closest to the ground truth. The feature fusion between different upsampling paths makes the ERU-Net segment the edge of OD more accurately. It shows that ERU-Net is much more competitive for fine-grained segmentation.

VI. CONCLUSION
In this paper, we propose an enhanced residual U-Net (ERU-Net), which consists of one downsampling path and three upsampling paths. Different from the original U-Net, with such three upsampling paths structure, ERU-Net can enhance the corresponding fusion feature maps and capture more details of fundus images. Also, a residual block is constructed in ERU-Net to extract more representative features. The experiment results show that the performance of ERU-Net for MAs and EXs segmentation is satisfactory. Compared with other U-Net variants, ERU-Net has achieved best performance in three open public fundus image segmentation datasets. In the future work, we will apply ERU-Net into other medical image segmentation tasks, and further validate the generalization of segmentation performance.

REFERENCES
[1] K. Ogurtsova, JD. da Rocha Fernandes, Y. Huang, "IIDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040," *Diabetes Res. Clin.*, vol. 128, pp. 40–50, Jun. 2017.

[2] G. L. Ong, L. G. Ripley, R. S. Newsom, M. Cooper, and A. G. Casswell, "Screening for sight-threatening diabetic retinopathy: Comparison of fundus photography with automated color contrast threshold test," *Amer. J. Ophthalmol.*, vol. 137, no. 3, pp. 445–452, Mar. 2004.

[3] R. J. Tapp, J. E. Shaw, C. A. Harper, M. P. de Courten, B. Balkau, D. J. McCarty, H. R. Taylor, T. A. Welborn, and P. Z. Zimmet, "The prevalence of and factors associated with diabetic retinopathy in the australian population," *Diabetes Care*, vol. 26, no. 6, pp. 1731–1737, Jun. 2003.

[4] T. Spencer, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester, "An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus," *Comput. Biomed. Res.*, vol. 29, no. 4, pp. 284–302, Aug. 1996, doi: 10.1006/cbmr.1996.0021.

[5] T. Walter, J. Klein, P. Massin, and A. Erginay, "A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina," *IEEE Trans. Med. Imag.*, vol. 21, no. 10, pp. 1236–1243, Oct. 2002, doi: 10.1109/TMI.2002.806290.

[6] A. Sopharak, B. Uyyanonvara, S. Barman, and T. H. Williamson, "Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods," *Computerized Med. Imag. Graph.*, vol. 32, no. 8, pp. 720–727, Dec. 2008, doi: 10.1016/j.compmedimag.2008.08.009.

[7] D. Welfer, J. Scharcanski, and D. R. Marinho, "A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images," *Computerized Med. Imag. Graph.*, vol. 34, no. 3, pp. 228–235, Apr. 2010, doi: 10.1016/j.compmedimag.2009.10.001.

[8] M. M. Fraz, W. Jahangir, S. Zahid, M. M. Hamayun, and S. A. Barman, "Multiscale segmentation of exudates in retinal images using contextual cues and ensemble classification," *Biomed. Signal Process. Control*, vol. 35, pp. 50–62, May 2017.

[9] A. Mizutani, C. Muramatsu, Y. Hatanaka, S. Suemori, T. Hara, and H. Fujita, "Automated microaneurysm detection method based on double ring filter in retinal fundus images," *Proc. SPIE*, vol. 7260, Mar. 2009, Art. no. 72601N. [Online]. Available: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1335590

[10] M. U. Akram, S. Khalid, and S. A. Khan, "Identification and classification of microaneurysms for early detection of diabetic retinopathy," *Pattern Recognit.*, vol. 46, no. 1, pp. 107–116, Jan. 2013.

[11] L. Giancardo, F. Meriaudeau, T. P. Karnowski, Y. Li, S. Garg, K. W. Tobin, and E. Chaum, "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets," *Med. Image Anal.*, vol. 16, no. 1, pp. 216–226, Jan. 2012, doi: 10.1016/j.media.2011.07.004.

[12] B. Harangi, B. Antal, and A. Hajdu, "Automatic exudate detection with improved Naive-Bayes classifier," in *Proc. IEEE Conf.*, Jun. 2012, pp. 1–4.

[13] X. Zhang, "Exudate detection in color retinal images for mass screening of diabetic retinopathy," *Med. Image Anal.*, vol. 18, no. 7, pp. 1026–1043, 2014, doi: 10.1016/j.media.2014.05.004.

[14] G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener, and C. Roux, "Optimal wavelet transform for the detection of microaneurysms in retina photographs," *IEEE Trans. Med. Imag.*, vol. 27, no. 9, pp. 1230–1241, Sep. 2008, doi: 10.1109/TMI.2008.920619.

[15] R. Phillips, J. Forrester, and P. Sharp, "Automated detection and quantification of retinal exudates," *Graef. Arch. Clin. Exp.*, vol. 231, no. 2, pp. 90–94, Feb. 1993.

[16] H. Li and O. Chutatape, "Automated feature extraction in color retinal images by a model based approach," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 2, pp. 246–254, Feb. 2004, doi: 10.1109/TBME.2003.820400.

[17] M. Haloi, "Improved microaneurysm detection using deep neural networks," 2015, *arXiv:1505.04424*. [Online]. Available: http://arxiv.org/abs/1505.04424

[18] P. Chudzik, S. Majumdar, F. Calivá, B. Al-Diri, and A. Hunter, "Microaneurysm detection using fully convolutional neural networks," *Comput. Methods Programs Biomed.*, vol. 158, pp. 185–192, May 2018, doi: 10.1016/j.cmpb.2018.02.016.

[19] C. Kou, W. Li, Z. Yu, and J. Hao, "Microaneurysms segmentation with a U-Net based on recurrent residual convolutional neural network," *J. Med. Imag.*, vol. 6, no. 2, p. 025008, Accessed on: May 2019.

[20] O. Perdomo, J. Arevalo, and F. A. Gonzalez, "Convolutional network to detect exudates in eye fundus images of diabetic subjects," in *Proc. SIPAIM*, vol. 10160, Oct. 2017, Art. no. 101600T.

[21] B. Lenet, R. Komorowski, X. Wu, J. Huang, H. Grad, H. Lawrence, and S. Friedman, "Antimicrobial substantivity of bovine root dentin exposed to different chlorhexidine delivery vehicles," *J. Endodontics*, vol. 26, no. 11, pp. 652–655, Nov. 2000, doi: 10.1097/00004770-200011000-00005.

[22] J. H. Tan, H. Fujita, S. Sivaprasad, S. V. Bhandary, A. K. Rao, K. C. Chua, and U. R. Acharya, "Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network," *Inf. Sci.*, vol. 420, pp. 66–67, Dec. 2017, doi: 10.1016/j.ins.2017.08.050.

[23] Z. Feng, J. Yang, L. Yao, Y. Qiao, Q. Yu, and X. Xu, "Deep Retinal Image Segmentation: A FCN-Based Architecture with Short and Long Skip Connections for Retinal Image Segmentation," in *Proc. ICONIP*, 2017, pp. 713–722.

[24] R. Zheng, "Detection of exudates in fundus photographs with imbalanced learning using conditional generative adversarial network," *Biomed. Opt. Express*, vol. 9, no. 10, pp. 4863–4878, Oct. 2018.

[25] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[26] Z. Zeng, W. Xie, Y. Zhang, and Y. Lu, "RIC-unet: An improved neural network based on unet for nuclei segmentation in histology images," *IEEE Access*, vol. 7, pp. 21420–21428, 2019, doi: 10.1109/ACCESS.2019.2896920.

[27] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Proc. Lect. Notes Comput. Sci.*, 2016, pp. 483–499.

[28] J. Guo, J. Deng, N. Xue, and S. Zafeiriou, "Stacked dense U-nets with dual transformers for robust face alignment," in *Proc. BMVC*, 2018, pp. 1–6.

[29] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation,". [Online]. Available: https://arxiv.org/abs/1802.069552018

[30] A. Sevastopolsky, "Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network," *Pattern Recognit. Image Anal.*, vol. 27, no. 3, pp. 618–624, Jul. 2017.

[31] W. Wang, J. Shen, M.-M. Cheng, and L. Shao, "An iterative and cooperative top-down and bottom-up inference network for salient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5961–5970.

[32] W. Wang, S. Zhao, J. Shen, S. C. H. Hoi, and A. Borji, "Salient object detection with pyramid attention and salient edges," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1448–1457.

[33] W. Wang, J. Shen, X. Dong, A. Borji, and R. Yang, "Inferring salient objects from human fixations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 1913–1927, Aug. 2020.

[34] 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. (MICCAI)*, 2015, pp. 234–241.

[35] 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.

[36] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.

[37] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[38] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[39] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research," *Data*, vol. 3, no. 3, p. 25, Jul. 2018.

[40] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Inf. Sci.*, vol. 501, pp. 511–522, Oct. 2019, doi: 10.1016/j.ins.2019.06.011.

[41] C. Lam, C. Yu, L. Huang, and D. Rubin, "Retinal lesion detection with deep learning using image patches," *Invest Ophth Vis. Sci.*, vol. 59, no. 1, pp. 590–596, Jan. 2018.

[42] P. Khojasteh, B. Aliahmad, and D. K. Kumar, "Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms," *BMC Ophthalmol.*, vol. 18, no. 1, p. 288, Dec. 2018, doi: 10.1186/s12886-018-0954-4.

[43] P. Costa, T. Araujo, G. Aresta, A. Galdran, A. M. Mendonca, A. Smailagic, and A. Campilho, "EyeWeS: Weakly supervised pre-trained convolutional neural networks for diabetic retinopathy detection," in *Proc. 16th Int. Conf. Mach. Vis. Appl. (MVA)*, May 2019, p. 54.

[44] E. Imani and H.-R. Pourreza, "A novel method for retinal exudate segmentation using signal separation algorithm," *Comput. Meth. Prog. Bio.*, vol. 133, pp. 195–205, Sep. 2016, doi: 10.1016/j.cmpb.2016.05.016.

[45] P. Chudzik, "Exudate segmentation using fully convolutional neural networks and inception modules," *Proc. SPIE*, vol. 10547, Mar. 2018, Art. no. 1057430.

[46] P. Chudzik, B. Al-Diri, F. Caliva, G. Ometto, and A. Hunter, "Exudates segmentation using fully convolutional neural network and auxiliary codebook," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 770–773.

[47] J. Mo, L. Zhang, and Y. Feng, "Exudate-based diabetic macular edema recognition in retinal images using cascaded deep residual networks," *Neurocomputing*, vol. 290, pp. 161–171, May 2018.

[48] S. Guo, T. Li, H. Kang, N. Li, Y. Zhang, and K. Wang, "L-seg: An end-to-end unified framework for multi-lesion segmentation of fundus images," *Neurocomputing*, vol. 349, pp. 52–63, Jul. 2019.

[49] F. Zabihollahy, A. Lochbihler, and E. Ukwatta, "Deep learning based approach for fully automated detection and segmentation of hard exudate from retinal images," in *Proc. SPIE*, vol. 10953, Oct. 2019, Art. no. 1095308.

[50] S. Gupta, A. Panwar, S. Goel, A. Mittal, R. Nijhawan, and A. K. Singh, "Classification of lesions in retinal fundus images for diabetic retinopathy using transfer learning," in *Proc. Int. Conf. Inf. Technol. (ICIT)*, Dec. 2019, pp. 342–347.

[51] J. Sivaswamy, "A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis," *JSM Biomed. Imag. Data Papers*, vol. 2, no. 1, p. 1004, Mar. 2015.

[52] J. G. Zilly, J. M. Buhmann, and D. Mahapatra, "Boosting convolutional filters with entropy sampling for optic cup and disc image segmentation from fundus images," in *Proc. MLMI*, Oct. 2015, pp. 136–143.

[53] V. Kumar Singh, H. Rashwan, F. Akram, N. Pandey, M. Mostaf Kamal Sarker, A. Saleh, S. Abdulwahab, N. Maaroof, S. Romani, and D. Puig, "Retinal optic disc segmentation using conditional generative adversarial network," 2018, *arXiv:1806.03905*. [Online]. Available: http://arxiv.org/abs/1806.03905

[54] B. Al-Bander, B. Williams, W. Al-Nuaimy, M. Al-Taee, H. Pratt, and Y. Zheng, "Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis," *Symmetry*, vol. 10, no. 4, p. 87, Mar. 2018, doi: 10.3390/sym10040087.

WEI LI received the B.S. degree in mathematics from Hebei Normal University, Hebei, China, in 2017. She is currently pursuing the M.S. degree with the Beijing University of Posts and Telecommunications. Her current research interests include optimization theory and machine learning.

ZEKUAN YU received the B.S. degree in electrical engineering and automation from the China University of Mining and Technology, in 2014, and the Ph.D. degree from Peking University, China, in 2020. Since 2020, he joined the Academy for Engineering and Technology, Fudan University. His current research interests include medical image analysis and computer aided diagnosis, deep learning, and medical robot applications.

CAIXIA KOU received the B.Sc. degree from the Department of Mathematics and Applied Mathematics, Beijing Normal University, China, in 2006, and the Ph.D. degree from the State Key Laboratory of Scientific and Engineering Computing (LSEC), Chinese Academy of Sciences (CAS), Beijing, China, in 2011. In 2011, she joined the School of Sciences, Beijing University of Posts and Telecommunications. She has proposed several efficient practical optimization algorithms and released the corresponding software packages. Her current research interests include optimization algorithms and applications, software defined networking, big data analysis, and so on.
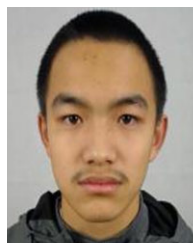
LUZHAN YUAN is currently pursuing the bachelor's degree with the Beijing University of Posts and Telecommunications.

● ● ●