



LEA U-Net: a U-Net-based deep learning framework with local feature enhancement and attention for retinal vessel segmentation

Jihong Ouyang^{1,2} · Siguang Liu^{1,2} · Hao Peng^{1,2} · Harish Garg^{3,4,5,6} · Dang N. H. Thanh⁷

Received: 15 January 2023 / Accepted: 21 April 2023 / Published online: 30 May 2023
© The Author(s) 2023

Abstract

Feature extraction of the retinal blood vessel is one of the crucial tasks in the prediction of ophthalmologic diseases. Important features are extracted based on image segmentation results. The efficiency of vessel segmentation methods could help doctors in the diagnostic of several relevant diseases as early as possible. Recently, U-Net has achieved good results in many medical image segmentation tasks, especially for images of blood vessels. However, due to the limitation of the network structure, some small features could be lost in the transmission process. As a result, there are still many research gaps for U-Net-based retinal vessel segmentation works. In this paper, we propose an improved U-Net based model to segment images of retinal vessels. The improvement focuses on U-Net from two aspects: designing a local feature enhancement module composed of dilated convolution and 1×1 convolution to enhance the feature extraction of tiny vessels; integrating an attention mechanism with skip connection of the network to highlight features related to vessel segmentation information taken from the down-sampling part to the up-sampling part. The performance of the proposed model was evaluated and compared with several published state-of-the-art approaches on the same public dataset—DRIVE, and the proposed method achieved an accuracy of 0.9563, F1-score of 0.823, TPR of 0.7983, and TNR of 0.9793. The AUC of PRC is 0.9109 and the AUC of ROC is 0.9794. The results proved the potential for clinical applications.

Keywords Deep learning · Retinal vessel segmentation · U-Net · Dilated convolution · Attention

Introduction

- ✉ Siguang Liu
lsgmliss@163.com
- ✉ Harish Garg
harishg58iitr@gmail.com
- ✉ Dang N. H. Thanh
thanhdn@ueh.edu.vn
- Jihong Ouyang
ouyj@jlu.edu.cn
- Hao Peng
93839111@qq.com

¹ College of Computer Science and Technology, Jilin University, Changchun, China

² Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, China

³ School of Mathematics, Thapar Institute of Engineering and Technology (Deemed University), Patiala 147004, Punjab, India

⁴ Department of Mathematics, Graphic Era Deemed to be University, Dehradun 248002, Uttarakhand, India

The retina is one of the most delicate parts of our body, which is composed of a very complicated structure. It might be captured by the fundus camera. The captured images provide an amount of significant information about structural changes of pathology and can be used for the prediction of some ocular diseases such as diabetic retinopathy (DR), cataract, hypertension, etc. [1, 2]. These diseases often change blood structure. For example, several abnormal blood vessels appear and start growing in the advanced stage of DR. It is also known as neovascularization [3]. As DR progresses, more pathological changes would appear: macular

⁵ Applied Science Research Center, Applied Science Private University, Amman 11931, Jordan

⁶ College of Technical Engineering, The Islamic University, Najaf, Iraq

⁷ Department of Information Technology, School of Business Information Technology, University of Economics Ho Chi Minh City, Ho Chi Minh City, Vietnam

edema caused by increased vascular permeability could damage central vision; the neovascularization and accompanying contraction of fibrous tissue can cause retinal traction detachment, leading to severe and often irreversible vision loss. The neovascularization may bleed, which can cause further complications of the retina. These pathological changes of the retina can lead to blindness [4]. Therefore, analyzing the features of blood vessels can provide an important pathological basis for the early diagnosis of eye diseases.

Before performing analysis, the ophthalmologist needs to separate the blood vessels, also known as segmentation, from the retinal image. Manual blood vessel segmentation is a very tedious task and it requires retinal ophthalmologists to spend a lot of time distinguishing between blood vessels and other areas of the fundus. This is mainly because there are many segmentation features of blood vessels, including length, width, branch angle, tortuosity, etc. Inexperienced doctors can easily make mistakes in this step. In addition, the segmentation results may also be different for various patients, and the uneven level of the segmentation personnel would also have a negative impact on the subsequent diagnosis process. Considering the above factors, with many patients, and limited medical resources, strictly manual segmentation of the vessels is impossible to perform. Therefore, automatic vessel segmentation in general, and retinal fundus vessel segmentation, in particular, should play a key role in the disease diagnosis of ophthalmology.

However, the structure of retinal vessels is extremely complicated [5]. Therefore, automatic vessel segmentation is a challenge. So far, this difficult problem has been widely studied in the literature. Machine learning methods are often used to deal with such complex problems. With achievements in deep learning studies in recent years, the problem of retinal vessel segmentation achieved important results. In particular, those methods based on U-Net can often achieve new state-of-the-art performance. This mainly benefits from the special structural design of U-Net. The U-Net [6] is designed to have a U-shape, where the down-sampling and up-sampling processes are formed a symmetrical structure. The network first extracts the detailed feature information of the image through the down-sampling process and then restores the location information of these features through the up-sampling process, so information on the context is captured and then delivered to higher resolutions to acquire a better segmentation result. Therefore, U-Net usually gives very good performance for segmenting medical images.

In summary, even though existing vessel segmentation methods, especially on U-Net architecture, usually achieve high performance, several challenges in developing processing methods still exist for small and very complex vessel structures. The fine features are easily lost in the process of multiple down-sampling and up-sampling.

In order to address the above problem, we propose a U-Net-based method named Local feature enhancement and Attention U-Net (LEA U-Net). The improvement of the proposed method over U-Net is as follows. First, a Local feature enhancement module is designed and applied to the network. The dilated convolution with different parameters is used to obtain larger local features while reducing the loss of small features caused by the down-sampling and up-sampling processes. This is mainly due to the fact that dilated convolution can easily increase the receptive field and change the size of the output feature maps by modifying its parameters. The former enables convolutional layers to capture larger ranges of vessel features. The latter allows the network to avoid using pooling during the process of obtaining feature maps with different sizes, thus preserving more detailed information. This module is a multi-output structure, and each output of the module is fused with the current feature maps at the corresponding up-sampled node in the network to supplement the up-sampled information. Second, an attention mechanism is integrated into the skip connection of the network to adjust the weight of each feature map and highlight the features related to blood vessel segmentation. This is equivalent to performing a denoising operation on the feature maps passed through the skip connection, further removing the information that is unrelated to blood vessels. The above two parts work together to improve the network's segmentation effect of tiny blood vessels.

The rest of this paper is organized as follows. The next section is for the related methods. The third section is the details of the proposed network architecture. The fourth section presents experimental results. Finally, the last section is for the conclusion.

Overview of related works

With the achievements of medical imaging technology over the past few years, there are several models developed for retinal vessel segmentation. According to the existing technologies, machine learning-based and deep learning-based methods are two primary approaches for the vessel segmentation problem.

For the traditional machine learning approach, there are two internal stages: extracting features from images and mapping the extracted features to the labels. Various feature extraction methods have been introduced, such as Gabor filter [7] and the Gaussian filter [8]. There are also many classifiers that have been proposed to handle different tasks, for instance, support vector machine (SVM) [9, 10], artificial neural networks [11], k-NN classifier [12], AdaBoost [13], etc. The algorithms composed of the above two parts were used widely in retinal vessel segmentation. Marin et al. computed a 7-D vector composed of gray-level and moment

invariants-based features for pixel representation and used a neural network (NN) scheme for pixel classification [14]. Aslani et al. proposed a new retinal vessel segmentation method. The proposed method combined features extracted by different methods, into a hybrid feature vector and trained with a Random Forest (RF) classifier [15]. The combination of different types of features resulted in increased local information with better discrimination for vessel and non-vessel pixels. Dash et al. presented a recursive method for retinal vessel segmentation [16]. This method first used adaptive thresholding technology to iteratively extract blood vessels on the pre-processed image and then applied morphological cleaning operation to generate a final vessel segmented image. For these traditional methods, the quality of the features extracted from input images has a great influence on the final result. However, features are often defined empirically by the designer, which could cause bias.

Deep learning is an advanced technique using backpropagation and a multi-layer neural network. It can automatically learn feature definitions from deep-level feature extraction methods, thereby avoiding human intervention [17]. Several deep learning architectures have been widely used and have achieved excellent results for various tasks including medical segmentation tasks [18, 19].

There are a number of studies based on deep learning that have investigated the retinal vessel segmentation problem. Wang et al. [20] combined two superior classifiers to carry out the segmentation. In this approach, Convolutional Neural Networks (CNNs) performed as a trainable feature extractor and then ensembling RFs worked as a trainable classifier. Fu et al. [21] formulated the vessel segmentation as a boundary detection problem and utilized the fully convolutional network (FCN) to generate the segmentation result. Moreover, a fully-connected Conditional Random Fields (CRFs) and a probability map of the discriminative vessel, and long-range pixel interactions are combined together. Maji et al. [22] presented a CNNs-ensemble-based framework for the detection of blood vessels in fundus color images. Ban et al. [23] presented a technique for multimode medical images based on spatial histogram. Qin et al. [24] presented a multi-focus image fusion method based on sparse decomposition. Liskowski et al. [25] proposed a segmentation method that classified multiple pixels simultaneously, which used a deep neural network trained on a large training dataset. Sappa et al. [26] presented an improved CNNs-based architecture to segment fluid abnormalities. To obtain multi-scale contextual information, the authors integrated several skip connections with atrous spatial pyramid pooling (ASPP). Tan et al. [27] used a 10-layer CNN to simultaneously segment multiple pathological features in fundus images. Similarly, another study with a 7-layer CNN proposed by Tan et al. to simultaneously segment optic disk, fovea, and blood vessels [28]. Sathananthavathi et al. proposed a parallel FCN architecture

for vessel segmentation [29]. They also studied the impact of using different levels of pre-processed images on the model. Wu et al. proposed a network architecture in which the front network converts input images into probabilistic retinal vascular maps and the subsequent network further refines these maps [30]. In addition, the model utilizes skip connections to connect two identical multi-scale backbones, allowing useful multi-scale features to be directly transmitted from shallow layers to deep layers, thereby improving segmentation performance. Sultana et al. designed an encoder-decoder model in an unconventional way [31]. First, the encoder part of the model uses up-sampling to enlarge the image and extract more detailed image features. Then, the decoder part of the model uses down-sampling to restore the feature maps to their original resolution, in order to achieve better segmentation results.

In 2015, a network architecture called U-net for medical image segmentation was proposed and achieved good performance in many tasks [6]. Recently, researchers have begun to try to use U-NET in the proposed method for retinal vessel segmentation, and continue to achieve new state-of-the-art performance. Zhang et al. utilized residual connection and cooperate with an architecture based on U-net to detect vessels [32] by adding additional labels on boundary areas and using an edge-aware mechanism to convert the original task into a multi-class task. Alom et al. utilized U-NET to propose a Recurrent Residual Convolutional Neural Network (RRCNN) [33]. In this method, recurrent convolutional layers were used to improve the effect of feature extraction, and residual units were used to help train deep architecture. Jin et al. integrated the deformable convolution into U-Net to extract contextual information and empower precise localization by combining low-level and high-level features [34]. By adaptively adjusting the receptive fields, it can recognize retinal vessels with different shapes and scales. Li et al. proposed Iter-Net [35], which is composed of U-Net-like components that are iterated multiple times, resulting in a network that is 4 times deeper than a standard U-Net. Iter-Net also employs weight sharing and skip connections to facilitate training. Zhou et al. proposed Unet++ [36], which integrates U-Net models with different depths and re-designs skip connections to obtain a highly flexible feature fusion scheme, in order to improve model performance. Li et al. proposed Res2Unet [37], which uses a multi-scale strategy to extract vessels of different widths and utilizes a channel attention mechanism to facilitate communication between channels, recalibrating the relationships between channel features. In addition, the author also proposed two post-processing methods, one for mining disconnected vessels and the other for removing false positive and false negative samples. Dong et al. proposed CRAUNet [38], which can be seen as a series of concatenated U-Net structures to obtain representations from coarse to fine. In addition, DropBlock is utilized in CRAUNet to

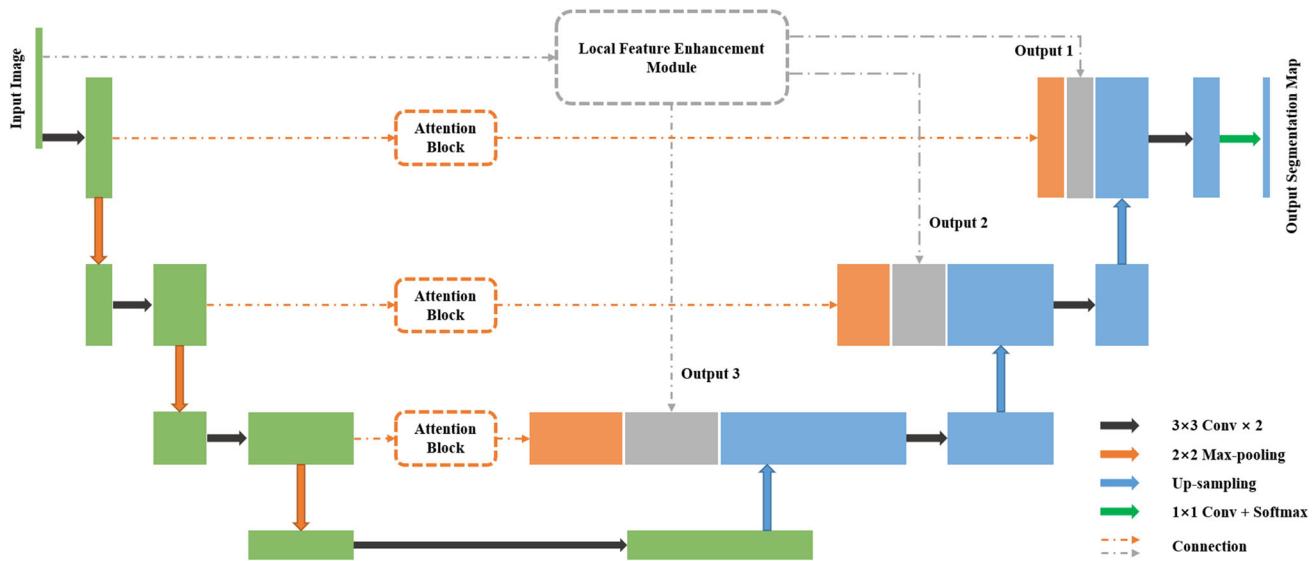


Fig. 1 Architecture of the proposed LEA U-Net

reduce overfitting. Chen et al. proposed PCAT-UNet [39], which is a U-shaped network based on Transformer with a Convolution branch. PCAT-UNet uses skip connections to fuse deep and shallow features from both sides, effectively capturing global dependency relationships and details in the low-level feature space.

In addition, due to the particularity of medical tasks, they require a high level of reliability from the models used, which needs to be supported by the models' own interpretability. However, deep learning models have a black box nature, which makes them inferior in terms of interpretability. In recent years, many scholars have been conducting related research and trying to link the internal processes of the network with the final results to enhance the interpretability of the models [40–43].

Overall, the existing deep learning methods, especially U-Net-based methods, have achieved good results in retinal blood vessel segmentation. But there still exists a research gap to improve the segmentation results of small and tiny vessels.

Proposed method

In this study, LEA U-Net is proposed with the aim of segmenting retinal fundus vessels. LEA U-Net is a deep-learning model and developed based on U-NET. We integrate U-Net with a local feature enhancement module and optimize skip connections to form attention blocks to improve performance.

Overall structure of LEA U-Net

Fig. 1 illustrates the architecture of the proposed LEA U-Net, which mainly consists of three parts, a U-shaped structure including down-sampling and up-sampling processes, a local feature enhancement module, and attention blocks. The convolutional layers in the network use ReLU as the activation function by default, except for those in attention blocks.

To balance computational complexity and efficiency, a 48×48 patch of the pre-processed grayscale image of fundus is used as the input of the network, which is the same referred to [33]. The input image is sent to the local feature enhancement module and the down-sampling part of the U-shaped structure. In the local feature enhancement module, multiple scale features are extracted to generate feature maps with different sizes, which contain many fine features that could be missed during the down-sampling process and are passed to the subsequent network. The down-sampling part contains three identical convolution-max-pooling structures, each of which is composed of two 3×3 convolutional layers and a max-pooling layer. Each max-pooling layer reduced the height and width of the feature maps by half, and then the first convolutional layer enlarges the number of channels twice. After the entire down-sampling process and two 3×3 convolutional layers, a $6 \times 6 \times 256$ size output is generated and sent to the up-sampling part. In addition, the feature maps output by the convolutional layers before each max-pooling layer are also sent to the corresponding attention block. In the attention block, the importance of the features is redistributed to highlight those that are more relevant to the task, and then

output to the up-sampling part of the U-shaped structure. The structure of the up-sampling part is symmetrical to the down-sampling part. During the entire up-sampling process, the size of feature maps is continuously enlarged, and the number of channels is continuously reduced. The output of each up-sampling layer is concatenated with the corresponding output of the local feature enhancement module and the attention block in each channel to supplement more feature information and then passed to the subsequent convolutional layer. The up-sampling part finally produces a 48*48*32 output. Subsequently, a 1×1 convolutional layer followed by softmax is added to adjust the number of channels and produce segmentation results.

To facilitate understanding, we provide pseudo code to describe the internal operation process of the model, as shown in Algorithm 1.

Algorithm 1 LEA U-Net

```

Require: Image  $X$ ;
Ensure: Segmentation map  $Y$ ;
1: Input  $X$  into the local feature enhancement module to obtain Supplementary feature maps  $\{\bar{x}_{LE}^k\}_{k=1}^3$ ;
2: set  $x_{Encoder}^0 = X$ ;
3: for  $i = 1$  to  $3$  do
4:   2 layers of  $3 \times 3$  convolution operation: set  $x_{Encoder}^i = conv_3(conv_3(x_{Encoder}^{i-1}))$ ;
5:   Using attention mechanism to process the feature maps passed through skip connection: set  $\bar{x}_{Encoder}^i = attention\_block(x_{Encoder}^i)$ ;
6:   Pooling operation: set  $x_{Encoder}^i = max\_pooling(x_{Encoder}^i)$ ;
7: end for
8: 2 layers of  $3 \times 3$  convolution operation: set  $x_{Decoder}^0 = conv_3(conv_3(x_{Encoder}^0))$ ;
9: for  $i = 3$  to  $1$  do
10:   Upsampling operation: set  $\bar{x}_{Decoder}^i = up\_sampling(x_{Decoder}^{i-1})$ ;
11:   Feature fusion operation: set  $x_{Decoder}^i = concat(\bar{x}_{Encoder}^i, \bar{x}_{LE}^i, \bar{x}_{Decoder}^i)$ ;
12:   2 layers of  $3 \times 3$  convolution operation: set  $x_{Decoder}^0 = conv_3(conv_3(x_{Decoder}^i))$ ;
13: end for
14: Using  $1 \times 1$  convolution and softmax to generate the final segmentation map:  $Y \leftarrow softmax(conv_1(x_{Decoder}^0))$ ;
15: return  $Y$ ;

```

In the following two parts, we will elaborate on the Local feature enhancement module and Attention block.

Local feature enhancement module

The U-shaped structure of U-Net could cause the loss of the features of small blood vessels. Furthermore, due to the uneven distribution of retinal blood vessels, the convolutional and pooling operations in U-Net are susceptible to the limitation of the visual field when extracting the features

of continuous large-area blood vessels. Therefore, we introduce the dilated convolution, and construct a multi-scale local feature enhancement module with it, to deliver more useful information to the up-sampling part of the network.

The dilated convolution uses the dilation parameter to define the distance between two adjacent pixels to be calculated in convolutional operation. By adjusting the dilation parameter, the receptive field can be increased at the same computational complexity. The dilated convolution with various dilation parameters is presented in Fig. 2. The colored area represents the receptive field of a filter on the input feature map, the red part means the pixels that need to be convolved, and the blue part is those pixels that do not need to be calculated. Visible, the dilated convolution is to get a more receptive field by ignoring some pixels, which means that it also has a certain loss of detail. Therefore, we did not directly replace the convolution and pooling in U-Net with dilated convolution but added a local feature enhancement module to complement the information flow in the network.

Figure 3 shows the design of the local feature enhancement module, which contains 3 blocks with similar internal structures. In each block, the information passes through a dilated convolutional layer and a 1×1 convolutional layer successively, and finally reaches the respective output ends. The input of Block 1 is the input image, and for the next two blocks, the input is the output of the dilated convolutional layer in the previous block. It can be found that the parameters of the dilated convolutional layer in the three blocks are different. The size of the filters and the dilation parameter in blocks 2 and 3 are significantly larger than that of block 1. Let us explain the reason for this design.

For Block 1, the size of the output feature maps needs to be consistent with the input image, so we need to add zero padding around the input image. In this case, if larger filters and dilation parameters are used, the proportion of the padding area in the input image needs to be increased correspondingly, which could cause the problem of sparse effective information on the edges of the image. In order to avoid this situation, we chose smaller filters and dilation parameters in Block 1. For Block 2 and Block 3, the height and width of the feature maps output by the dilated convolutional layer are both halves of the input. Zero-padding is not needed to maintain the size of the feature maps, and it could not encounter the problem of sparse effective information. So, we use larger parameters here to get better performance. In addition, the parameters of these two dilated convolutional layers are carefully designed so that the height and width of the output feature maps are exactly half of the input, and the model performance and the computational complexity are also traded off.

In order to achieve cross-channel interaction and information integration, after the dilated convolutional layer in each block, there is a 1×1 convolutional layer. The number

Fig. 2 Illustration of dilated convolution with different parameters. **a** 3×3 1-dilated convolution. **b** 3×3 2-dilated convolution. **c** 5×5 3-dilated convolution

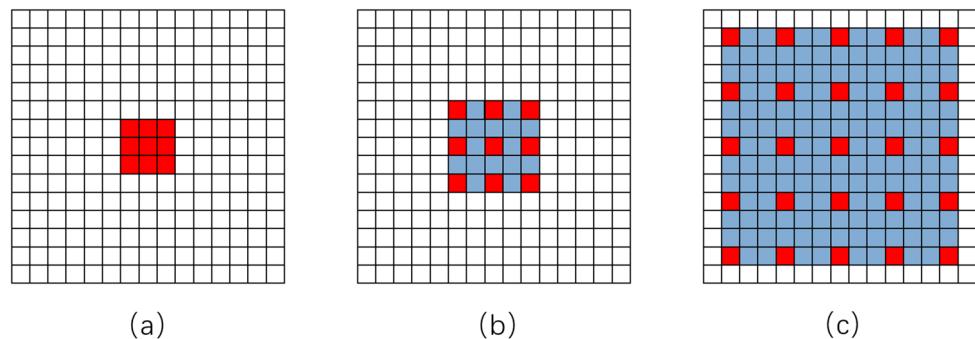
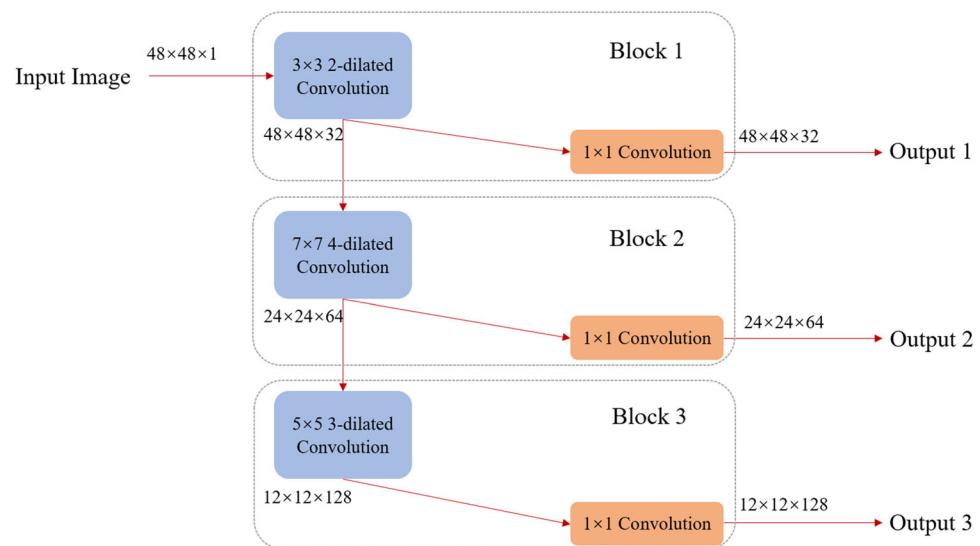


Fig. 3 Schematic of the local feature enhancement module



of channels corresponding to out1, out2, and out3 is 32, 64, and 128, respectively. The final output result of each block is fused with the information on other paths in the network to optimize the up-sampling process of the model.

Attention block

In U-Net, the partial feature maps generated by the down-sampling processes are passed to the up-sampling part through skip connections and merged with other information. We expect to use an adaptive weight redistribution module to influence the information transmitted via skip connections, highlighting those features related to vessel segmentation. Therefore, an attention module is integrated into the original skip connection to form an attention block. The structure of this block refers to the non-local block [44], with some modifications, which is shown in Fig. 4.

In the attention block, the input first passes through three 1×1 convolutional layers at the same time to generate three groups of feature maps A , B , and C , respectively. A and B are combined with element-wise addition, and the result is passed through a 1×1 convolutional layer to produce the attention weight matrix Γ . The sigmoid function is used to

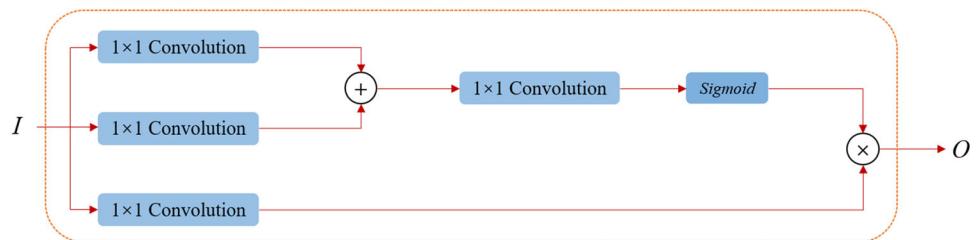
map the non-fixed value range of the output of the convolutional layer into a weight coefficient between 0 and 1. Then the attention weight matrix Γ is calculated with the linearly transformed input C , and the final output O is obtained.

The operation performed by the attention module is for each pixel of all channels. First, the attention weight matrix is generated by integrating all channel information, and then the attention weight matrix is used to allocate the importance of each pixel. In order to make the feature map passed to the subsequent network consistent with U-Net, the attention block does not change the size of the feature maps.

Computational complexity analysis

Table 1 presents the parameter and computational complexity of each part of LEA U-Net, measured in millions. The model's learning ability is to some extent related to its parameter count, while FLOPs (Floating Point Operations) is a commonly used metric to measure the computational complexity of a model, and can intuitively reflect the computational resources required for a model to run once. The growth of LEA U-Net in these two metrics is mainly due to the local feature enhancement module. But compared to

Fig. 4 Schematic of the attention block. I and O represent the input and the output of this block, respectively. \oplus denotes element-wise sum, and \otimes denotes element-wise product



U-Net backbone, the proportion of the local feature enhancement module is not significant. If regular convolutions with the same receptive field are used to compose this module, the FLOPs of it will be more than twice that of U-Net backbone. This also intuitively reflects the benefits of using dilated convolution for our model.

Experiments

To demonstrate the performance of the LEA U-Net model, we test it on the DRIVE dataset and compare the results with other recently published methods. Furthermore, in order to directly show the improvements that the local feature enhancement module and attention block bring to the model, respectively, i.e., to perform ablation experiments, we also add an LEA U-net model without the attention mechanism to the experiments, which is named LE U-net. Our models are implemented using Python 3.6, the Keras, and TensorFlow frameworks. We use cross-entropy loss, SGD with batch size 32, and an initial learning rate of 0.01. We drop the learning rate to 0.001 after 80 epochs and trained a total of 100 epochs.

Dataset and image preprocessing

DRIVE is a well-known retinal blood vessel database. It consists of forty color images (in RGB color space) of retinal fundus images with the blood vessels. The database is randomly split into a training set and a validation set with a rate of 50:50, i.e., each set contains 20 images. The validation set is only used for testing and not for model training. For each color image, there are two corresponding binary mask images in the data set, i.e., the ground truth and the binary field of view mask, as shown in Fig. 5. The former is obtained by experts' manual segmentation, and the latter indicates the

extent of the fundus area in each color image. The plane resolution of all images in DRIVE is 565×584 .

For medical image segmentation, image preprocessing is very important whether using traditional methods or deep learning-based models [45, 46]. Proper preprocessing can improve the performance of the model. Here, we use grayscale image conversion, Contrast Limited Adaptive Histogram Equalization (CLAHE) [47], and Gamma correction sequentially to preprocess the color fundus images.

First, a color image in RGB is transformed into a monochromatic grayscale image, which can reduce to a certain extent the different colors of the collected pictures due to different equipment and lighting. For the retinal vessel segmentation, this process is particularly important, which largely affects the final segmentation results [48]. We use Formula 1 to convert the grayscale image:

$$I_{Gray} = 0.3I_{Red} + 0.59I_{Green} + 0.11I_{Blue}, \quad (1)$$

where I_{Gray} , I_{Red} , I_{Green} , I_{Blue} are, respectively, pixel intensity in Grayscale mode, Red channel, Green channel, and Blue channel.

Then, CLAHE was utilized to increase the foreground-background contrast. This method can improve image contrast while reducing the noise generated in the process. We need to set a threshold first. The pixels above the threshold will be cropped on the histogram and evenly distributed on other gray levels to form a new histogram, and then perform adaptive equalization on the new histogram. In addition, it is necessary to set an equalization grid size to divide the image into non-overlapping blocks, and then process each area separately. This ensures the stability of the processing process. We set the threshold to 2 and the equalization grid size to 8.

Finally, we use Gamma correction to further adjust the contrast of the image to make the difference between light and dark around the blood vessel more obvious. Gamma cor-

Table 1 The parameter count and computational complexity of each part in LEA U-Net

| | # Parameter | FLOPs |
|----------------------------------|-----------------|-------------------|
| U-Net backbone | 2.14 M (83.59%) | 872.2 M (77.95%) |
| Local feature enhancement module | 0.33 M (12.89%) | 190.07 M (16.99%) |
| Attention block | 0.09 M (3.51%) | 56.62 M (5.06%) |
| LEA U-Net | 2.56 M (100%) | 1118.89 M (100%) |

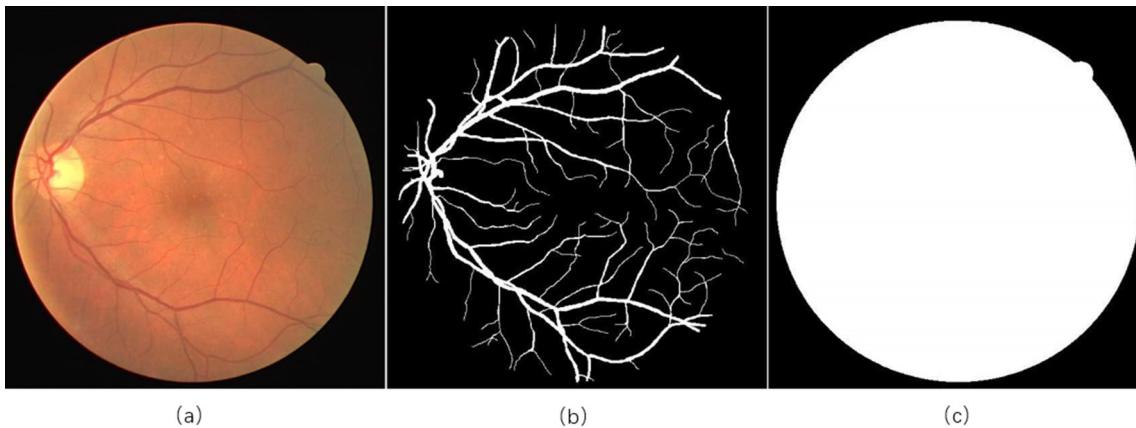


Fig. 5 Example images of DRIVE. **a** Color image; **b** corresponding ground truth; **c** corresponding mask

rection is to change the contrast between the low and high pixel intensity areas of the image by adjusting the gamma curve. Here, we need to set a parameter γ . When γ is less than 1, the contrast of the high pixel intensity area will be decreased, and the contrast of the low pixel intensity area will be increased. When γ is greater than 1, the opposite effect will be produced. Since the gray value of blood vessels in the image is generally low, a γ greater than 1 should be used. After comparison, we finally choose Gamma correction with a γ value of 1.2 to further process the images.

The step-by-step preprocessing is shown in Fig. 6. We can notice that after several preprocessing, the blood vessels of the fundus with low contrast and blurry colors have become clearer, the difference in brightness between blood vessels and surrounding non-vascular parts has become larger, and the structure of small blood vessels can be better distinguished. The images will be cropped to 48×48 patches before being used by the models, as shown in Fig. 7. This processing step increases the number of images used for model training by hundreds of times, which can effectively alleviate the overfitting problem.

Image segmentation performance evaluation metrics

To evaluate the segmentation performance, the following metrics: accuracy (ACC), F-measure (F1), true-positive rate (TPR), true-negative rate (TNR), the area under curve (AUC), and AUC of Precision–Recall curve (PRC), are used:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (2)$$

$$\text{F1} = 2 \times \frac{\text{PPV} \times \text{TPR}}{\text{PPV} + \text{TPR}} \quad (3)$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (5)$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

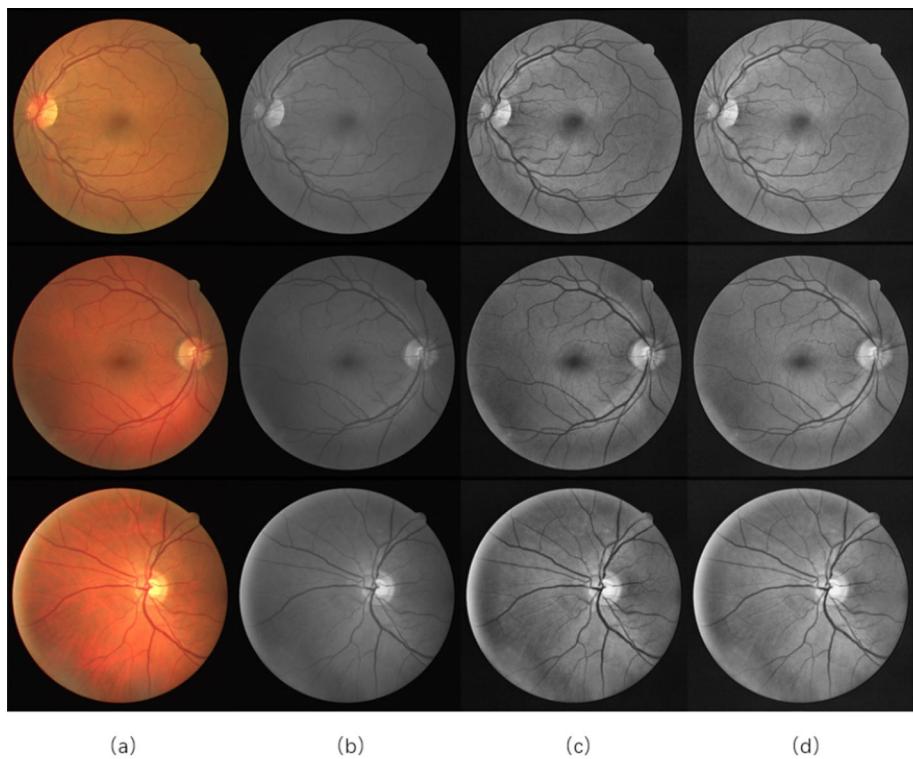
where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative.

As the main evaluation metric, ACC measures the proportion of correctly classified pixels in the image. TPR measures the proportion of blood vessel pixels that are correctly identified, and TPR measures the proportion of those background pixels that are correctly identified. F1 is used to evaluate the comprehensive performance of the model.

Results

First, we compare our method with several state-of-the-art methods as in Table 2. Generally speaking, the performance of the methods in the STA family is worse than one of the methods of DNN family. Since they rely primarily on hand-made features, and if we consider the condition of sufficient training data, the quality of these features is not good compared with the DNN-based methods. Also, the performance of several U-Net-based methods is relatively high, and the LEA U-Net achieved the best performance on the two most important metrics. It achieved the highest global accuracy of 0.9563 and the highest F1 of 0.8230. The Residual U-Net and the Recurrent U-Net have similar motivations to our method. The former uses the residual structure in U-Net to optimize the transfer of features in the network, and the latter replaces the convolutional layers before each down-sampling and up-sampling in U-Net with recurrent convolutional layers to extract more complex features. It can be seen from the experimental results that our method has more advantages in the extraction and transmission of small blood vessel features. In addition, the result of LE U-Net is better than U-Net but worse than LEA U-Net, which demonstrates that both the

Fig. 6 Example images after each preprocessing step. **a** Original images; **b** gray scaled images; **c** images after CLAHE operation; **d** images after Gamma correction



(a)

(b)

(c)

(d)

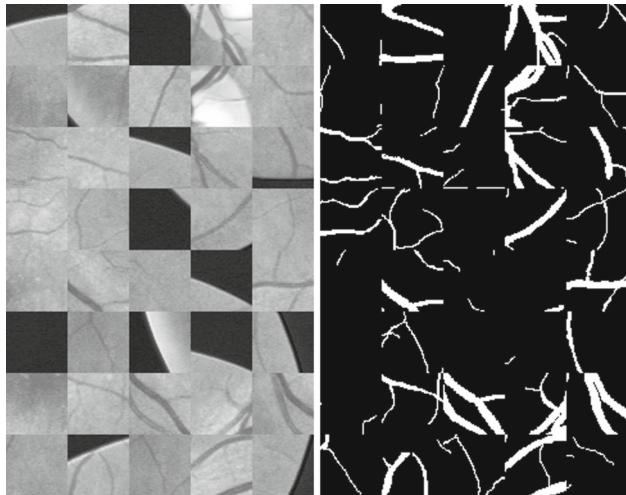


Fig. 7 Example patches in the left and corresponding ground truth are shown in the right

local feature enhancement module and the attention block in our model have an effect on performance improvement. Next, we will further analyze this by comparing other experimental results of the three models.

Figures 8 and 9, respectively, show the change of loss and ACC during the training process of the three models, where the horizontal axis indicates the number of epochs. We can see that all the blue curves corresponding to the training set are very smooth, but the red curves corresponding to the vali-

dation set are different. The red curve of U-Net still has a large range of fluctuations even after 70 epochs, while the curves of the other two models have basically stabilized. Comparing U-Net with LE U-Net, it can be observed that: at the beginning of the training phase, the loss of U-Net is lower than that of LE U-Net, but the latter has a higher accuracy, which indicates that a more stable training process will improve the performance of the model. Compared with LE U-Net, the most obvious advantage of LEA U-Net is in the convergence speed in the early stage. In short, the local feature enhancement module makes the model training process more stable, while the attention block speeds up the convergence speed of the model, which leads to better performance.

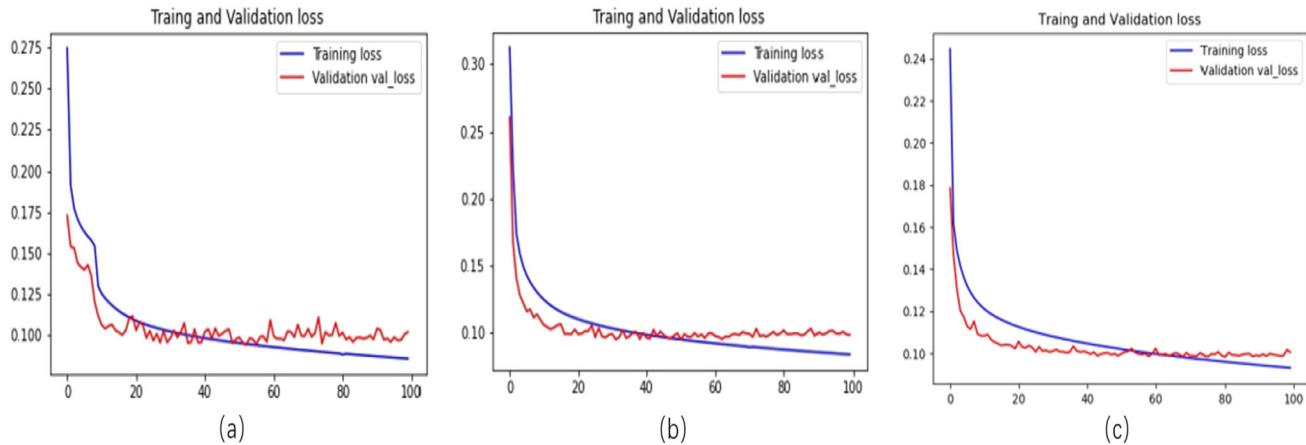
Next, we assess the model performance by ROC curves and PRC curves. The results are shown in Figs. 10 and 11. For both types of curves, if the area under the curve (AUC) is larger, the corresponding model is considered better. Since LE U-Net and LEA U-Net achieve higher values of AUC for both of the curves than one of U-Net, the performance of LE U-Net and LEA U-Net are also better than U-Net. In addition, the performance of LEA U-Net with the integration of attention mechanism is also better than one of LE U-Net.

Last but not least, the segmentation results of the three models with some details are displayed in Fig. 12. From the local magnification view, it is obvious that the three methods have different segmentation for a tiny blood vessel. Since these small features are easily missed, it is very difficult to perform segmentation precisely. Due to the limi-

Table 2 Comparisons against existing approaches on DRIVE

| Methods | Type | ACC | F1 | TPR | TNR |
|--------------------------|------|---------------|---------------|---------------|---------------|
| Azzopardi et al. [49] | STA | 0.9442 | — | 0.7655 | 0.9704 |
| Cheng et al. [50] | STA | 0.9474 | — | 0.7252 | 0.9798 |
| Fu et al. [51] | DNN | 0.9523 | — | 0.7603 | — |
| Dasgupta et al. [52] | DNN | 0.9533 | — | 0.7691 | 0.9801 |
| Roychowdhury et al. [53] | STA | 0.9520 | — | 0.7250 | 0.9830 |
| Chen et al. [54] | DNN | 0.9453 | — | 0.7426 | 0.9735 |
| Yan et al. [55] | DNN | 0.9542 | — | 0.7653 | 0.9752 |
| Li et al. [56] | DNN | 0.9517 | — | 0.7569 | 0.9749 |
| Yan et al. [57] | DNN | 0.9538 | — | 0.7631 | 0.9750 |
| U-Net [6]* | DNN | 0.9518 | 0.7891 | 0.7311 | 0.9840 |
| Residual U-Net [33] | DNN | 0.9553 | 0.8149 | 0.7726 | 0.9820 |
| Recurrent U-Net [33] | DNN | 0.9556 | 0.8155 | 0.7755 | 0.9813 |
| RIMNet [31] | DNN | 0.9478 | 0.7739 | 0.7683 | 0.9808 |
| NFN+ [38] | DNN | 0.9555 | 0.8297 | 0.8520 | — |
| Iter-Net [38] | DNN | 0.9556 | 0.8134 | 0.7607 | — |
| Unet++ [38] | DNN | 0.9541 | 0.8052 | 0.7447 | — |
| Res2Unet [37] | DNN | — | — | 0.8162 | 0.9734 |
| LE U-Net * | DNN | 0.9546 | 0.8085 | 0.7746 | 0.9809 |
| LEA U-Net * | DNN | 0.9563 | 0.8230 | 0.7983 | 0.9793 |

STA denotes the standard segmentation algorithm, and DNN denotes a method based on deep neural networks. The results of models marked with * were obtained from our own experiments, while other data were taken from the corresponding references

**Fig. 8** The change of loss during the training process of different models. **a** U-Net; **b** LE U-Net; **c** LEA U-Net

tation of networks, U-Net only extracted a few coarse pieces of information. With the help of the local feature enhancement module, LE U-Net extracted information better than U-Net. The use of the attention mechanism also makes features related to blood vessels easier to capture by the model. As a result, LEA U-Net can identify tiny vessels and hence, improve segmentation results for LE U-Net. More segmentation results of LEA U-Net are shown in Fig. 13.

Conclusions

In this paper, we proposed a U-Net-based model, named LEA U-Net, to perform retinal blood vessel segmentation in a pixel-wise manner. LEA U-Net improved U-net by two modules. (M1) A local feature enhancement module to strengthen the extraction of local features. This module uses dilated convolution with different parameters to extract

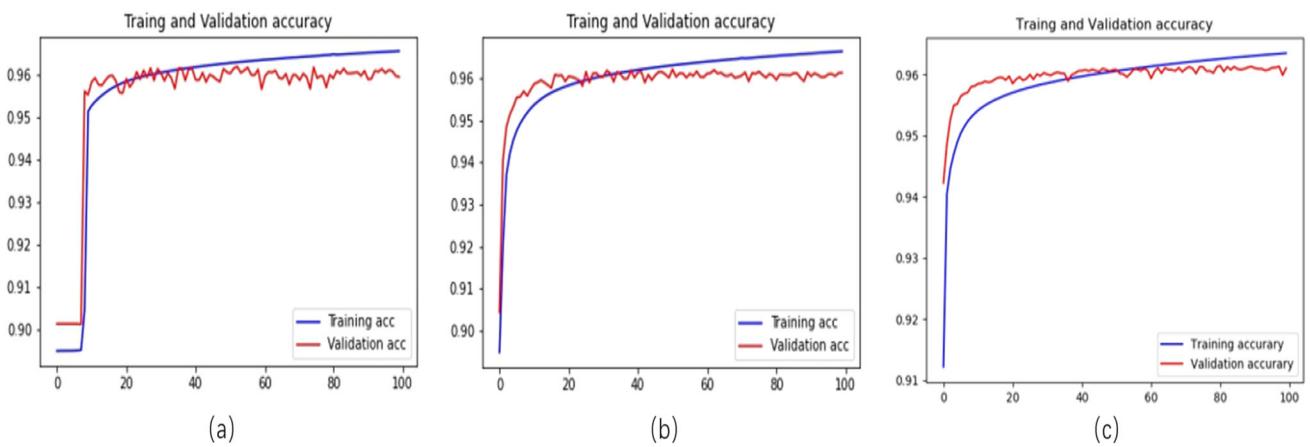


Fig. 9 The change of ACC during the training process of different models. **a** U-Net; **b** LE U-Net; **c** LEA U-Net

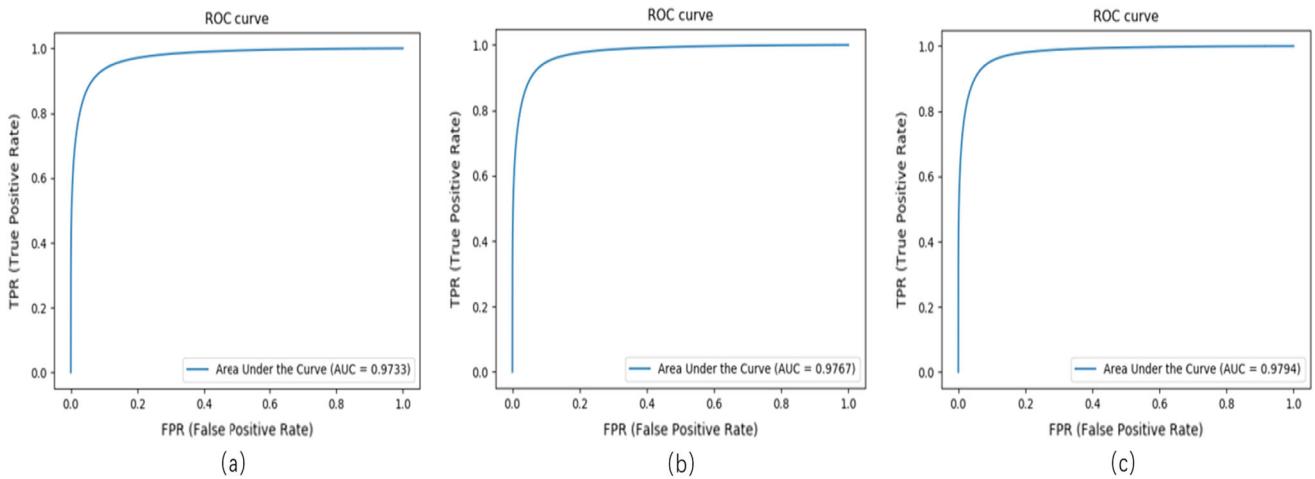


Fig. 10 ROC curves of different models. **a** U-Net; **b** LE U-Net; **c** LEA U-Net

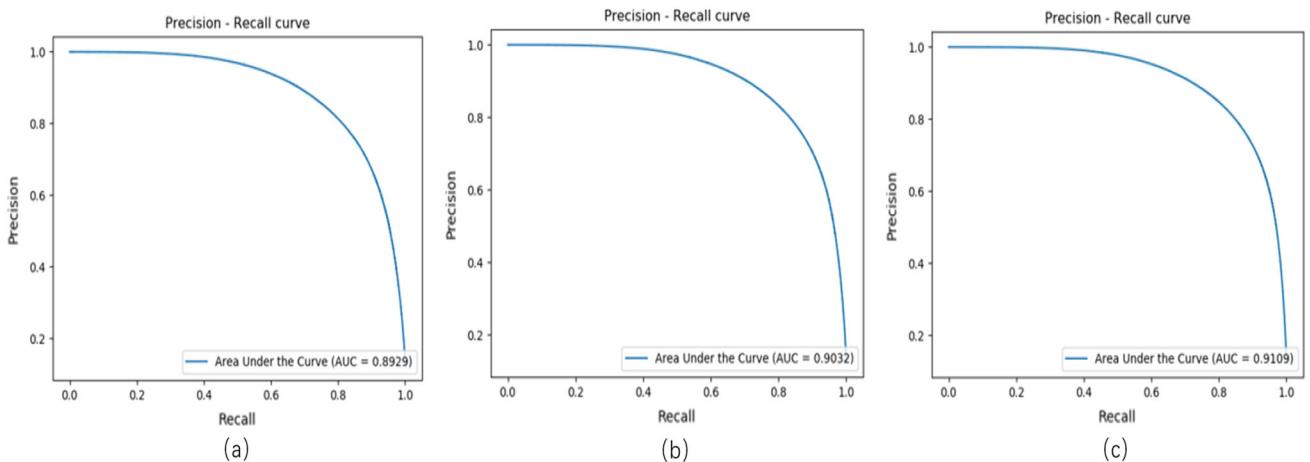


Fig. 11 PRC curves of different models. **a** U-Net; **b** LE U-Net; **c** LEA U-Net

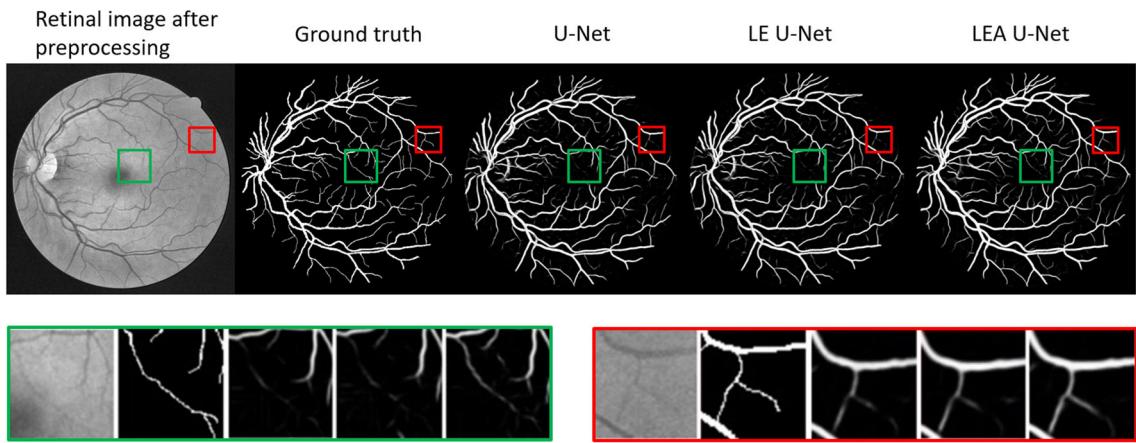
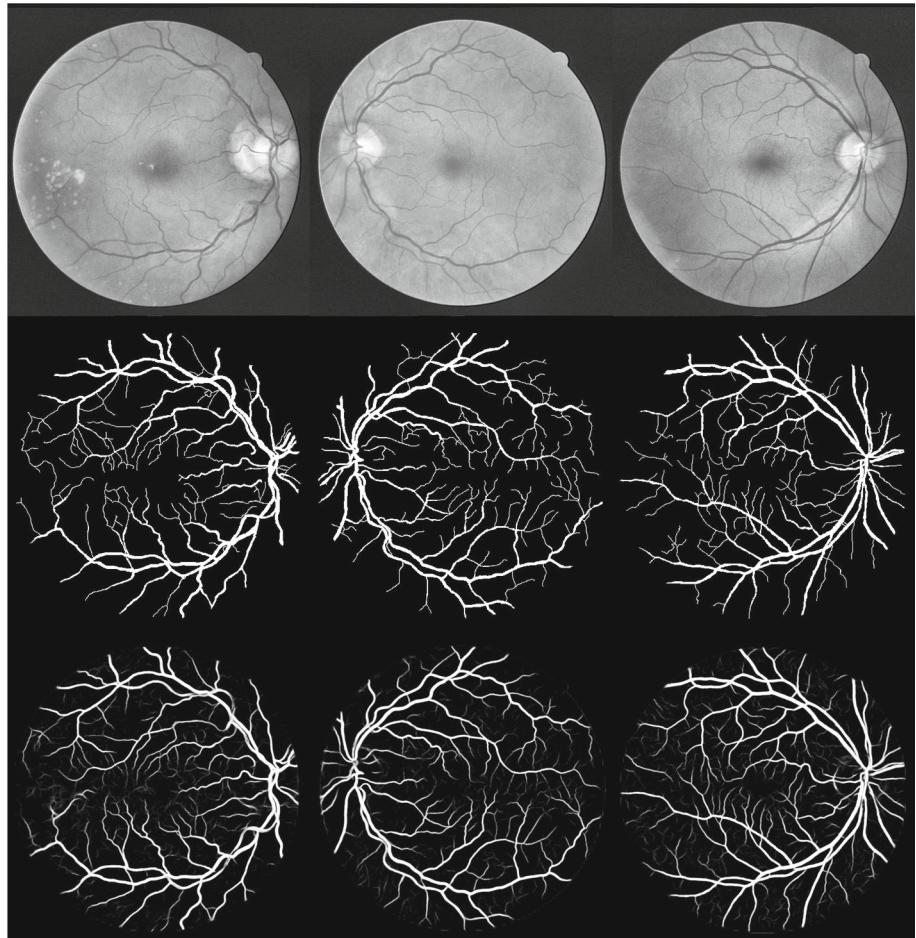


Fig. 12 Magnified view of boxed patches predicted by different models

Fig. 13 Retinal images after preprocessing (upper row), ground truth (middle row), segmentation result (bottom row)



multi-scale vessel features, in order to supplement those tiny features that are easily lost in the network down-sampling-up-sampling process. (M2) An attention mechanism to focus on highly relevant features. Experiments on the DRIVE database demonstrate the effectiveness of our model. Taking into account the requirements for timeliness in practical applications, we plan to optimize the computational com-

plexity of LEA U-Net in the future, reducing the time consumption of model training and segmentation under the premise of ensuring accuracy. Also, we shall extend our approach to deal with different applications related to face recognition [58], fundus image [59], clinical dataset [60], low-dose CT scan images [61] etc.,

Acknowledgements This study was funded by the National Natural Science Foundation of China (NSFC) (No. 61876071), the Scientific and Technological Developing Scheme of Jilin Province of China (No. 20180201003SF, No. 20190701031GH), the Energy Administration of Jilin Province (No. 3D516L921421). Grants Recipient: Dr. Jihong Ouyang. This research was funded by the University of Economics Ho Chi Minh City, Vietnam. Grant Recipient: Dr. Dang N.H. Thanh.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Mendonca AM, Campilho A (2006) Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Trans Med Imaging* 25(9):1200–1213
- Fang B, Hsu W, Lee ML (2003) On the detection of retinal vessels in fundus images
- Akram UM, Khan SA (2012) Automated detection of dark and bright lesions in retinal images for early detection of diabetic retinopathy. *J Med Syst* 36(5):3151–3162
- Shanmugam V, Banu RDW (2013) Retinal blood vessel segmentation using an extreme learning machine approach. In: IEEE Point-of-Care Healthcare Technologies (PHT). IEEE 2013:318–321
- Han Z, Yin Y, Meng X, Yang G, Yan X (2014) Blood vessel segmentation in pathological retinal image. In: 2014 IEEE international conference on data mining workshop, IEEE, 2014, pp 960–967
- Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 234–241
- Hamamoto Y, Uchimura S, Watanabe M, Yasuda T, Mitani Y, Tomita S (1998) A gabor filter-based method for recognizing handwritten numerals. *Pattern Recogn* 31(4):395–400
- Nguyen V, Blumenstein M (2011) An application of the 2d gaussian filter for enhancing feature extraction in off-line signature verification. In: 2011 international conference on document analysis and recognition, IEEE, 2011, pp 339–343
- Ricci E, Perfetti R (2007) Retinal blood vessel segmentation using line operators and support vector classification. *IEEE Trans Med Imaging* 26(10):1357–1365
- You X, Peng Q, Yuan Y, Cheung Y-M, Lei J (2011) Segmentation of retinal blood vessels using the radial projection and semi-supervised approach. *Pattern Recogn* 44(10–11):2314–2324
- Sinhanayothin C, Boyce JF, Cook HL, Williamson TH (1999) Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *Br J Ophthalmol* 83(8):902–910
- Staal J, Abràmoff MD, Niemeijer M, Viergever MA, Van Ginneken B (2004) Ridge-based vessel segmentation in color images of the retina. *IEEE Trans Med Imaging* 23(4):501–509
- Li X, Wang L, Sung E (2008) Adaboost with svm-based component classifiers. *Eng Appl Artif Intell* 21(5):785–795
- Marín D, Aquino A, Gegúndez-Arias ME, Bravo JM (2010) A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Trans Med Imaging* 30(1):146–158
- Aslani S, Sarnel H (2016) A new supervised retinal vessel segmentation method based on robust hybrid features. *Biomed Signal Process Control* 30:1–12
- Dash J, Bhoi N (2018) An unsupervised approach for extraction of blood vessels from fundus images. *J Digit Imaging* 31(6):857–868
- Song HA, Lee SY (2013) Hierarchical representation using nmf. In: International conference on neural information processing. Springer, pp 466–473
- Zhou F, Ye Y, Song Y (2021) Image segmentation of rectal tumor based on improved u-net model with deep learning. *J Signal Process Syst*
- Teodoro A, Silva D, Rosa REA (2022) A skin cancer classification approach using gan and roi-based attention mechanism. *J Signal Process Syst*
- Wang S, Yin Y, Cao G, Wei B, Zheng Y, Yang G (2015) Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing* 149:708–717
- Fu H, Xu Y, Wong DWK, Liu J (2016) Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. In: IEEE 13th international symposium on biomedical imaging (ISBI). IEEE 2016, pp 698–701
- Maji D, Santara A, Mitra P, Sheet D (2016) Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images. arXiv preprint [arXiv:1603.04833](https://arxiv.org/abs/1603.04833)
- Ban Y, Wang Y, Liu S, Yang B, Liu M, Yin L, Zheng W (2022) 2D/3D multimode medical image alignment based on spatial histograms. *Appl Sci* 12(16):8261
- Qin X, Ban Y, Wu P, Yang B, Liu S, Yin L, Liu M, Zheng W (2022) Improved image fusion method based on sparse decomposition. *Electronics* 11(15):2321
- Liskowski P, Krawiec K (2016) Segmenting retinal blood vessels with deep neural networks. *IEEE Trans Med Imaging* 35(11):2369–2380
- Sappa LB, Okuwobi IP, Li M, Zhang Y, Xie S, Yuan S, Chen Q (2021) Retfluidnet: Retinal fluid segmentation for sd-oct images using convolutional neural network. *J Digit Imaging* (2021) 1–14
- Tan JH, Fujita H, Sivaprasad S, Bhandary SV, Rao AK, Chua KC, Acharya UR (2017) Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network. *Inf Sci* 420:66–76
- Tan JH, Acharya UR, Bhandary SV, Chua KC, Sivaprasad S (2017) Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network. *J Comput Sci* 20:70–79
- Sathananthavathi V, Indumathi G, Ranjani AS (2020) Parallel architecture of fully convolved neural network for retinal vessel segmentation. *J Digit Imaging* 33(1):168–180
- Wu Y, Xia Y, Song Y, Zhang Y, Cai W (2020) NFN+: A novel network followed network for retinal vessel segmentation. *Neural Netw* 126:153–162
- Sultana F, Sufian A, Dutta P (2022) RIMNet: image magnification network with residual block for retinal blood vessel segmentation. In: IEEE Region 10 Symposium (TENSYMP). IEEE 2022, pp 1–6
- Zhang Y, Chung AC (2018) Deep supervision with additional labels for retinal vessel segmentation task. In: International conference

- on medical image computing and computer-assisted intervention. Springer, 2018, pp 83–91
- 33. Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK (2018) Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation
 - 34. Jin Q, Meng Z, Pham TD, Chen Q, Wei L, Su R (2019) Dunet: A deformable network for retinal vessel segmentation. *Knowl-Based Syst* 178:149–162
 - 35. Li L, Verma M, Nakashima Y, Nagahara H, Kawasaki R (2020) Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp 3656–3665
 - 36. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2020) Unet++: redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans Med Imaging* 39(6):1856–1867
 - 37. Li X, Ding J, Tang J, Guo F (2022) Res2Unet: a multi-scale channel attention network for retinal vessel segmentation. *Neural Comput Appl* 34(14):12001–12015
 - 38. Dong F, Wu D, Guo C, Zhang S, Yang B, Gong X (2022) CRAUNet: a cascaded residual attention U-Net for retinal vessel segmentation. *Comput Biol Med* 147:105651
 - 39. Chen D, Yang W, Wang L, Tan S, Lin J, Bu W (2022) PCAT-UNet: UNet-like network fused convolution and transformer for retinal vessel segmentation. *PLoS ONE* 17(1):e0262689
 - 40. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A (2016) Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp 2921–2929
 - 41. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision, 2017, pp 618–626
 - 42. Li X, Xiong H, Li X, Wu X, Zhang X, Liu J, Bian J, Dou D (2022) Interpretable deep learning: interpretation, interpretability, trustworthiness, and beyond. *Knowl Inf Syst* 64(12):3197–3234
 - 43. Rosso MM, Marasco G, Aiello S, Aloisio A, Chiaia B, Marano GC (2023) Convolutional networks and transformers for intelligent road tunnel investigations. *Comput Struct* 275:106918
 - 44. Wang X, Girshick R, Gupta A, He K (2018) Non-local neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp 7794–7803
 - 45. Rasta SH, Partovi ME, Seyedarabi H, Javadzadeh A (2015) A comparative study on preprocessing techniques in diabetic retinopathy retinal images: illumination correction and contrast enhancement. *J Med Signals Sensors* 5(1):40
 - 46. Thanh D, Prasath V, Hieu L, Hien N (2020) Melanoma skin cancer detection method based on adaptive principal curvature, colour normalisation and feature extraction with the abcd rule. *J Digit Imaging* 33(3):574–585
 - 47. Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, ter Haar Romeny B, Zimmerman JB, Zuiderveld K (1987) Adaptive histogram equalization and its variations. *Comput Vis Graph Image Process* 39(3):355–368
 - 48. Soares JV, Leandro JJ, Cesar RM, Jelinek HF, Cree MJ (2006) Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *IEEE Trans Med Imaging* 25(9):1214–1222
 - 49. Azzopardi G, Strisciuglio N, Vento M, Petkov N (2015) Trainable cosfire filters for vessel delineation with application to retinal images. *Med Image Anal* 19(1):46–57
 - 50. Cheng E, Du L, Wu Y, Zhu YJ, Megalooikonomou V, Ling H (2014) Discriminative vessel segmentation in retinal images by fusing context-aware hybrid features. *Mach Vis Appl* 25(7):1779–1792
 - 51. Fu H, Xu Y, Lin S, Wong DWK, Liu J (2016) Deepvessel: Retinal vessel segmentation via deep learning and conditional random field. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 132–139
 - 52. Dasgupta A, Singh S (2017) A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation. In: IEEE 14th international symposium on biomedical imaging (ISBI 2017). IEEE 2017, pp 248–251
 - 53. Roychowdhury S, Koozekanani DD, Parhi KK (2014) Blood vessel segmentation of fundus images by major vessel extraction and subimage classification. *IEEE J Biomed Health Inform* 19(3):1118–1128
 - 54. Chen Y (2017) A labeling-free approach to supervising deep neural networks for retinal blood vessel segmentation, arXiv preprint [arXiv:1704.07502](https://arxiv.org/abs/1704.07502)
 - 55. Yan Z, Yang X, Cheng K-T (2018) Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. *IEEE Trans Biomed Eng* 65(9):1912–1923
 - 56. Li Q, Feng B, Xie L, Liang P, Zhang H, Wang T (2015) A cross-modality learning approach for vessel segmentation in retinal images. *IEEE Trans Med Imaging* 35(1):109–118
 - 57. Yan Z, Yang X, Cheng K-T (2018) A three-stage deep learning model for accurate retinal vessel segmentation. *IEEE J Biomed Health Inform* 23(4):1427–1436
 - 58. Xu S, He Q, Tao S, Chen H, Chai Y, Zheng W (2022) Pig face recognition based on trapezoid normalized pixel difference feature and trimmed mean attention mechanism. *IEEE Trans Instrum Meas*. <https://doi.org/10.1109/TIM.2022.3232093>
 - 59. Jin K, Huang X, Zhou J, Li Y, Yan Y, Sun Y, Ye J (2022) FIVES: a fundus image dataset for artificial intelligence based vessel segmentation. *Sci Data* 9(1):475. <https://doi.org/10.1038/s41597-022-01564-3>
 - 60. Song X, Tong W, Lei C, Huang J, Fan X, Zhai G, Zhou H (2021) A clinical decision model based on machine learning for ptosis. *BMC Ophthalmol* 21(1):169. <https://doi.org/10.1186/s12886-021-01923-5>
 - 61. Lu S, Yang B, Xiao Y, Liu S, Liu M, Yin L, Zheng W (2023) Iterative reconstruction of low-dose CT based on differential sparse. *Biomed Signal Process Control* 79:104204. <https://doi.org/10.1016/j.bspc.2022.104204>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.