A Global and Local Enhanced Residual U-Net for Accurate Retina Vessel Segmentation

Sheng Lian, Lei Li, Guiren Lian, Xiao Xiao, Zhiming Luo, and Shaozi Li

Abstract—Retinal vessel segmentation is a critical procedure towards the accurate visualization, diagnosis, early treatment and surgery planning of ocular diseases. Recent modern approaches based on deep learning have achieved impressive performance in retinal vessel segmentation. However, these methods generally apply global image pre-processing and take the whole retinal images as input during network optimizing, which have two drawbacks for accurate retinal vessel segmentation. First, lack of full utilize of the local patch information. Second, overlook the geometric constraint that retina only occurs in a specific area on whole image or extracted patch. As a result, the global based methods still have problems at dealing fine details in retinal vessel segmentation, such as, recognize the small thin vessels, discriminate the optic disk are, etc. To address these drawbacks, this work proposes a **G**lobal and **L**ocal enhanced residual **U**-n**E**t (**GLUE**) for accurate retina vessel segmentation, which can benefit from both the globally and locally enhanced information, and focus better on only the retinal region. Experimental results on two benchmark datasets demonstrate the effectiveness of the proposed methods. Our approach consistently improves the segmentation accuracy over conventional U-Net and achieves competitive performance compared with the state of the art.

Index Terms—Retinal Vessel Segmentation, Deep Learning, Weighted Res-UNet, Global and Local Enhance

1 Introduction

According to reports in [1]–[3], Diabetic Retinopathy (DR), Glaucoma and Age-related Macular Degeneration (AMD) are the leading causes of blindness in the aging population. It's a urgent need for researchers to develop the automatic diagnosis systems for retinal pathology. The retina is the only human body part which can observe the microcirculation through a noninvasive fundus examination. With the complex fundoscope system, digital retinal images can provide a magnified view of retina area, including retinal vessel branches, optical disk and macula. The subtle changes and abnormalities in retinal vessel structures can be deemed as an important signal for diagnosing multiple diseases, including DR [4], AMD [2], cardiovascular disease [5], hypertension [6] and many chronic eye diseases [7]. The task of retinal vessel segmentation is to classify pixels in given retinal images that belong to the vessel region. Retinal blood vessels segmentation contributes to help ophthalmologists diagnosing correctly.

Manual retinal vessel segmentation is a time consuming and tedious task even for a well-trained ophthalmologist. Meanwhile, the diagnosis of retinal diseases, especially some acute diseases, require feedback as soon as possible. So, computer-aided automatic retinal vessel segmentation is essential for reducing medical costs, avoiding delayed treatment and improving efficiency. Although the quality of retinal imaging has improved significantly with the improvement of new imaging technology [8], automatic retinal vessel segmentation is still a challenging task. The main challenges can be briefly summarized as follows:

- *Indistinct small vessels:* Some small blood vessels located at the end of branches are with extremely low contrast and even indistinguishable for professional ophthalmologists.
- *Low contrast at optic disk area*: Optic disk areas are often brighter and with lower contrast, which makes it to recognize the retinal vessels.
- *Complex bifurcated structure*: Retinal vessels have bifurcated structure similar to trees. Complex vessel structures such as bifurcations, crossovers, closely parallel vessels, and junctions, are difficult to maintain.
- *Influence of abnormal area:* Some retinal images are with abnormal lesions, such as microaneurysms and exudates. Those abnormal regions increase the difficulty of segmentation task.
- *Illumination:* Poor or overexposed illumination, including light reflection and uneven background illumination caused by the light source of camera, will reduce images contrast, which results in the non-sharp boundary of retinal vessels.

To deal with these challenges, existing learning-based approaches generally apply an image-level pre-processing operation on the whole retina image and then randomly crop patches for learning a segmentation model. But only considering the image-level statistical information for pre-processing will lose sight of the local information which usually plays important role for recognizing the low-contrast local vessels. After getting the segmentation results of the learn models, some methods (e.g., [9], [10]) use a post-processing step such as Markov random field (MRF) or conditional random field (CRF) to recovery the complex bifurcated structure of the retinal blooding vessels, but the hype-parameters of MRF or CRF need to be manually fine-tuned such as kernel bandwidth.

Retina only appear in specific unitary area, in which pixels outside this area are useless for segmentation, so it's essential to get rid of the irrelevant noisy background. As well, global

- *S. Lian, L. Li, X. Xiao and S. Li are with the Cognitive Science Department, Xiamen University, Xiamen, China.*
- *Z. Luo is with the Postdoc Center of Information and Communication Engineering, Xiamen University, Xiamen, China, and the Key Laboratory of Cognitive Computing and Intelligent Information Processing of Fujian Education Institutions, Wuyi University, Wuyishan, China.*
- *G. Lian is with the College of Mathematics and Informatics, Fujian Normal University, China.*

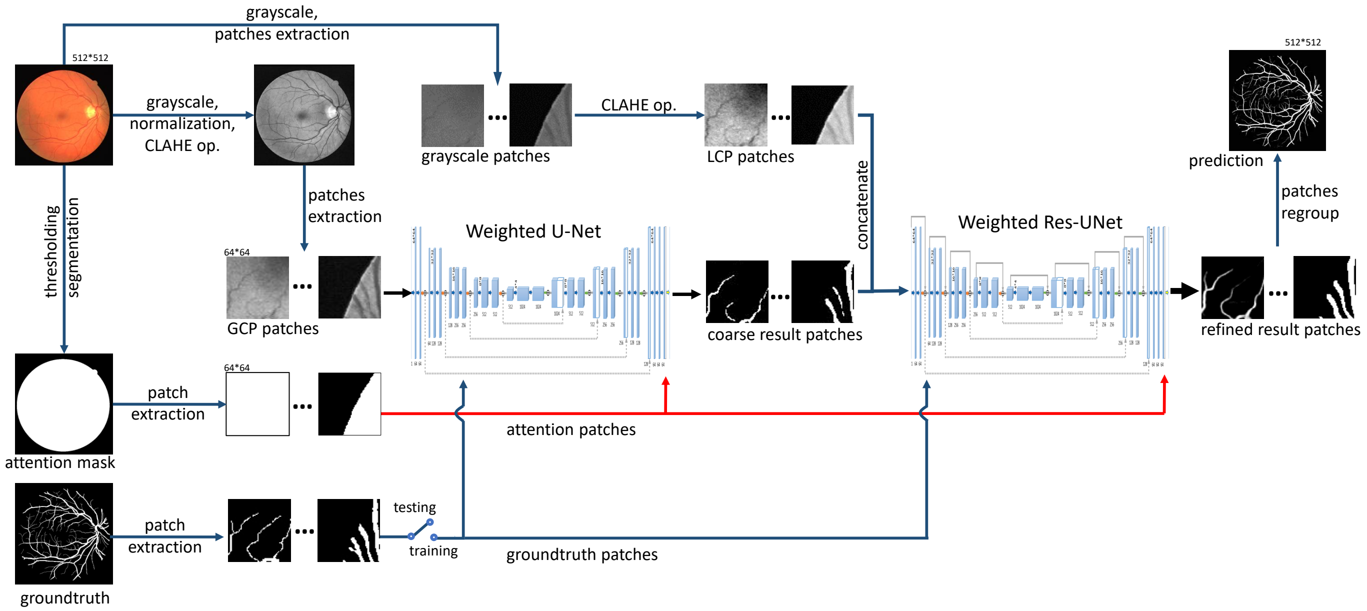

Fig. 1. The overall pipeline of our proposed model which contains two parts. A weighted U-Net takes the globally enhanced patches as input to produce a coarse segmentation, and a weighted residual U-Net (Res-UNet) with locally enhanced patches along with previous segmentation results as input to perform a refinement. GCP and LCP indicate Global CLAHE pre-processing and Local CLAHE pre-processing, respectively.

image-level and local patch-level pre-processing operations are complementary schemes with each other. In this paper, we propose a **G**lobal and **L**ocal enhanced residual **U**-n**E**t (**GLUE**) model with a cascaded refinement structure to tackle the challenging retinal vessel segmentation task. Our model as shown in Fig. 1 contains two parts, a weighted U-Net and a weighted residual U-Net (Res-UNet). The weighted U-Net takes the globally enhanced patches as input to produce a coarse segmentation. The weighted residual U-Net with locally enhanced patches along with previous segmentation results as input to perform a refinement, whose parameters are learned automatically instead of manually fine-tuned. Firstly, we generated an weighted attention mask to get rid of the irrelevant noisy background. Then, we apply CLAHE operation globally and locally to get enhanced retinal patches. Besides, we also add residual connection in the second part of our model (weighted Res-UNet) for learning more effective features. After all these improvement, our model can learn more discriminative features for distinguishing vessel and non-vessel pixels and have a better maintaining of the retinal vessel tree structure.

We conduct experiments on two representative datasets, DRIVE dataset and STARE dataset. Experimental results demonstrate that our proposed method can outperform previous representative approaches. Besides, comparison between the coarse and refined segmentation results shows the effectiveness of our proposed cascade refinement scheme. The visual example results in the zoomed-in view indicate that our model can deal with challenging situations. This paper is an extension of our previous ITME2018 conference paper [11]. In this version, we make several significant improvement by considering globally and locally enhancement on retinal images, and construct our model in a cascaded manner to do a refinement.

The rest of the paper is organized as follows. In Section 2, we review representative retinal vessel segmentation methods. In Section 3, we describe the proposed method in detail. In Section 4, we reported the experimental results. And we conclude our paper in Section 5.

2 Related Work

Realizing the importance of retinal vessel segmentation task, different segmentation methods have been proposed in various papers. Excellent survey of existing methods for retinal vessel segmentation can be referred in [12], [13]. In this section, we briefly introduce some of the most representative work. In general, the retinal blood vessel segmentation methods can mainly be divided into unsupervised and supervised.

Unsupervised methods attempt to find inherent patterns of retinal vessels without any manually annotation. Most of them are rule-based techniques, including vessel tracking [14]–[16], matched filtering [17]–[19], morphological processing [20]–[22], thresholding [23]–[25], etc.

Yin *et al.* [14] exploit the fact that retinal vessels have connected branch-like structure, and proposed a vessel tracking-based segmentation method. In this method, a Bayesian method which maximum a posterior as criterion is used to detect vessel edge points. Given these initial seed points, the entire vessel tree is tracked by following vessel centerline based on local information. Given the assumption that retinal vessel's intensity can be modeled as a Gaussian shaped curve, Wang *et al.* [18] proposed a Matched Filtering-based method with multiwavelet kernels and multiscale hierarchical decomposition. In this study, vessels are enhanced using matched filtering with multiwavelet kernels. Knowing that vessels are linear and connected structure in retina, mathematical morphology can be adopted. Fraz *et al.* [20] used the first order derivative of a Gaussian filter to identify vessel centerlines, followed by a multi-directional morphological top-hat transform to segment the vessels. Moreover, one possible way to sort out retinal vessel pixels is to segment by thresholding locally. For example, Jiang *et al.* [23] proposed an adaptive local thresholding framework on a verification-based multi-threshold probing scheme to segment retinal vessels. Generally, compared to supervised methods, unsupervised methods have a higher segmentation speed, lower computational complexity and lower segment accuracy.

Supervised methods segment retinal vessels by learning a model from a training set annotated by experienced ophthalmologists. All of the supervised methods try to select the most discriminative set of feature vectors from training images for better classifying vessel and non-vessel pixels. The supervised classifier in such methods range from kNN [26], [27], support vector machines (SVM) [28], [29], neural networks [30], [31], etc. Niemeijer *et al.* [26] extract only the green plane of retinal images and conduct k-Nearest-Neighbor (k-NN) algorithm as classifier to determine the pixels' class. Based on the assumption that image ridges coincide approximately with vessels' centerline, Staal *et al.* [27] proposed a ridge based vessel segmentation methodology with k-NN classifier. Ricci *et al.* [28] adopt line operation as feature vector and use SVM for pixel classification. In this method, a line detector based on the evaluation of the average gray level alone lines is applied for first stage unsupervised classification. Then, two orthogonal line detectors are adopted to construct feature vectors for supervised classification using a SVM. Nekovei*et al.* [30] applied a multilayer perceptron neural network, for which the inputs were derived from a principal component analysis (PCA), to identify retinal vessel pixels. To our best knowledge, this paper is the first study to use artificial neural networks-based method for retinal vessel segmentation.

Benefited from the emergence of large-scale data, the rapid development of computer/GPU power and the presentation of various artificial neural network models, deep learning-based methods boost dramatically in the field of computer vision [32]–[34]. Accompanied by these development, Convolutional Neural Networks (CNN) has become an efficient approach for analysing medical image [35]–[37], solving medical image segmentation problems [38]–[40] and can reach the state-of-the-art performance in the task of retinal vessel segmentation [41]–[43]. In this part, we will review several CNNs based retinal vessel segmentation methods. Wang et al. [44] proposed a segmentation method which uses CNN as a feature extractor and random forests as the final classifier. Wu et al. [45] firstly use the CNN to extract binary mask and then use a generalized particle filtering technique to extract retinal vessel tree under a probabilistic tracking framework. Later, Fu et al. [46] developed a multi-scale and multi-level CNN model to do the segmentation and used a Conditional Random Field (CRF) to consider the long-range interactions between pixels. Dasgupta et al. [47] utilized the fully convolutional neural networks and the multi-label inference to do structure predictions of the blood vessel. Son et al. [48] presented a method that adopts the generative adversarial training to improve the segmentation performance. Recently, Zhang et al. [49] proposed an architecture to sufficient use multi-level features and added atrous convolution to get effective multi-scale features for retinal blood vessel segmentation task. Oliveira *et al.* [43] combined the multiscale analysis provided by the Stationary Wavelet Transform with a multiscale Fully Convolutional Neural Network to cope with the varying width and direction of the vessel structure in the retina. Memari *et al.* [42] utilized a genetic algorithm enhanced spatial fuzzy c-means method for extracting an initial blood vessel network, with the segmentation further refined by using an integrated level set approach.

In a supervised method, the classification criteria are determined by the ground truth data based on given features. However,

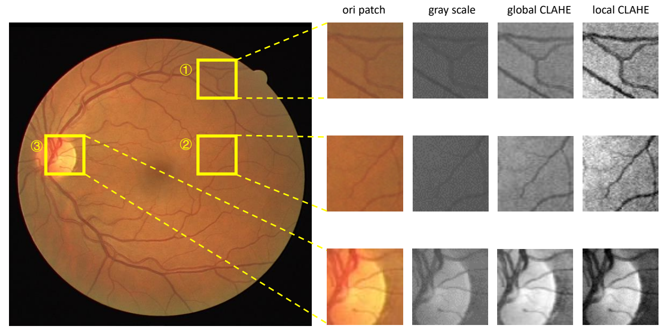

Fig. 2. Three groups of result patches after locally and globally CLAHE processing. Here, we plot examples of original patches, gray scaled patches, global CLAHE processed (GCP) patches, local CLAHE processed (LCP) patches in zoom-in view.

data annotated by professional doctors is usually expansive and difficult to obtain. As supervised methods are designed based on pre-classified data, their performance is usually better than that of unsupervised ones and can produce pretty good results for healthy retinal images. Although supervised methods, especially CNN-based methods have achieved satisfactory segmentation results in many scenarios, there are still many challenging issues of dealing with small vessels, poor illumination, etc.

3 Proposed Method

In this section, we describe the proposed **G**lobal and **L**ocal enhanced residual **U**-n**E**t (**GLUE**) model for tackling the retina vessel segmentation task in detail. The overall pipeline of the proposed method is shown in Fig. 1. The proposed model is a cascaded CNN model which contains 2 parts, a weighted U-Net and a weighted residual U-Net (Res-UNet). The weighted U-Net takes the globally enhanced patches as input to produce a coarse segmentation. The weighted residual U-Net with locally enhanced patches along with previous segmentation results as input to perform a refinement. The remainder of this section is organized as follows. Section 3.1 introduces the retinal image pre-processing steps used in our approach. Section 3.2 describe the detailed architecture of our proposed GLUE model. Finally, the scheme of overlapped prediction patches regrouping is discussed in Section 3.3.

3.1 Retinal Image Pre-processing

Globally and Locally Contrast Enhancement: As can be seen from Fig. 2, the original retinal image is with extremely low contrast. Suitable pre-processing steps can increase the contrast which alleviate the learning difficulty of CNNs and get better performance. In the following, we introduce the pre-processing steps that we applied in this study. For each retinal image, we resize it to $512 * 512$ and convert it into grayscale. To increase the retinal images' contrast, the contrast limited adaptive histogram equalization (CLAHE) operation is applied on retinal images globally and locally. CLAHE operation is a widely used pre-processing operation in medical image analysis, which has prominent enhancement effect and the computation process is as follow,

$$I = \frac{I - \mu}{\sigma}, \tag{1}$$

$$P = \frac{P - \mu}{\sigma}, \tag{2}$$

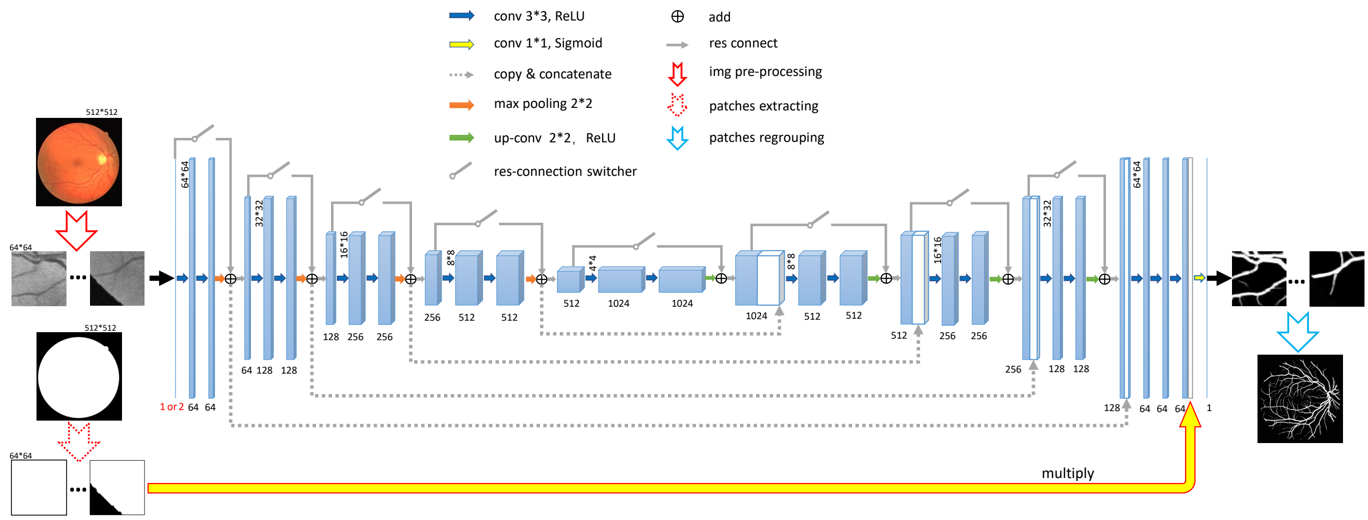

Fig. 3. The brief architecture of proposed Weighted Res-UNet with refinement. The meaning of all types of signs is marked on the top of the picture. In particular, when all the res-connection switchers are disconnected, the model indicates weighted U-Net, which acts as the first part in our approach. While when all the res-connection switchers are connected, the model indicates weighted Res-UNet, which serve as the second part in our approach. Due to the different inputs, the number of input channel for the first part and second part is 1 and 2, respectively.

where μ and σ are the mean and standard deviation of the grayscale image I or a local patch P.

Global CLAHE processing (GCP) and local CLAHE processing (LCP) are complementary with each other. Applying CLAHE operation on the entire image can maintain global information and avoid the degeneration effects of local noise like light reflection and diseased area, but GCP does not work well for some extremely low contrast areas. On the other hand, performing CLAHE operation in a local area can boost its local contrast information, especially in local areas with very thin capillary blood vessels.

As shown in Fig. 2, no matter applying CLAHE operation globally or locally, all the patches' contrast are enhanced. The two schemes have different advantages. For patch ① and ② where the local contrasts are rather low, LCP performs obviously better than GCP, and the vessels are with more contrast enhancement. While for patch ③ where light reflection or other local anomaly occurs, GCP is more robust than LCP. As such, we propose to use a cascade deep model to fully utilize the GCP and LCP as shown in Fig. 1. For the first stage, we apply CLAHE operation on entire images, then randomly crop globally enhanced image patches for training. For the second refine stage, we first extract images to patches and then apply a local CLAHE operation to increase the local contrast. By doing so, the proposed model can learn more discriminative features by considering the retinal images' global and local statistical information.

3.2 The architecture of the proposed GLUE model

In order to take full utilize of the global and local CLAHE operation, as well as model the dependencies among adjacent pixels and thus enforce complex bifurcated structure of the retina vessels, we implemented a *cascaded* CNN model. The architecture of our proposed GLUE model is shown in Fig. 1 which contains 2 parts, a weighted U-Net for coarse retinal vessel segmentation and a weighted Res-UNet to refine the segmentation results.

The overall architecture of the proposed weighted U-Net and weighted Res-UNet is shown in Fig. 3. In particular, when all the res-connection switchers (marked as gray switch symbols in the picture) are disconnected, the model indicates weighted U-Net, which serves as the first part in our approach. While when all the res-connection switchers are connected, the model indicates weighted residual U-Net, which serve as the second part in our approach. Similar to the original U-Net model [50], both the weighted U-Net and weighted Res-UNet have an encoder-decoder architecture. While beyond such architecture, we made several significant improvements by adding a weighted attention mechanism on both parts, and the skip connection scheme on the second part as introduced in [34]. All the parameters of the model are given in Fig. 3. In particular, due to the different inputs, the number of input channel for the first part and second part is 1 and 2, respectively.

3.2.1 The weighted attention mechanism

The retinal images in DRIVE [27] and STARE [51] dataset are with circular-like region of interest (ROI) and dark background. In order to model this geometric constraint, we add an attention scheme, like our former iris segmentation study [52]. A circular template ROI mask $\mathcal{M}$ is generated to estimate the potential area where the fundus most likely to appear. And the generated attention mask $\mathcal{M}$ will be used as a weighted attention scheme, which is shown in Fig 3 in the bottom as large yellow arrow. The attention mechanism is implemented by multiplying the model's second last layer's feature map with an attention mask. This operation can be expressed as

$$\mathcal{I}(x,y)=\begin{cases}\mathcal{F}(x,y)*1.0 & (x,y)\in\mathcal{M}\\ \mathcal{F}(x,y)*0.0 & (x,y)\notin\mathcal{M}\end{cases}, \tag{3}$$

in which $\mathcal{F}(x,y)$ represents the features at position (x,y). Due to the generate attention masks are very accurate, so we simply set the attention weight to be 1.0. But this does not mean that our attention mechanism is invalid. When the retinal images are extracted into patches, the attention mechanism can help our model better locate the target area in each patch.

By using this weighted attention mechanism, our model will only pay attention at the target ROI area and discard the irrelevant

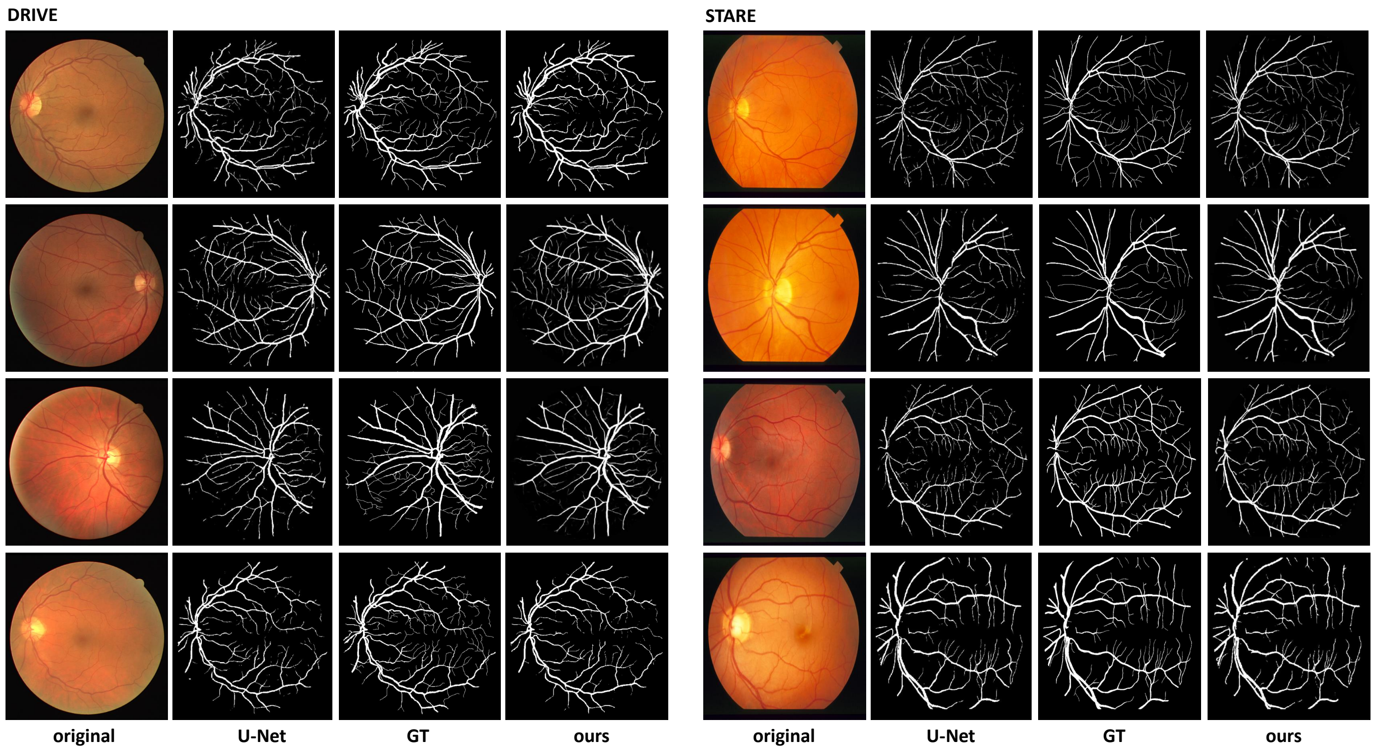

Fig. 4. Examples of retina vessels segmentation result from DRIVE dataset (listed in the left) and STARE dataset (listed in the right). Results show that our model can deal well with the problems of low illumination, hard area of the optic disk. Moreover, our model shows better blood vessel connectivity than U-Net. Detailed results in zoom-in view of extracted patches are displayed in Fig. 5.

noisy background. For the DRIVE dataset, we directly use the provided mask of fundus area as the weighted attention mask. While for STARE dataset, we computed the attention mask by a simple processing step which we convert the retinal image into grayscale, apply Gaussian filtering and then separate the fundus regions by doing binary thresholding at the value 40. As indicated in Fig. 3, both the proposed weighted U-Net (first part of our approach) and weighted Res-UNet (second part of our approach) adopt this weighted attention mechanism for better modeling the geometric constraint of fundus.

3.2.2 The skip connection scheme

As demonstrated in [34], adding skip connection can increase the depth and improve the accuracy of deep CNNs. Inspired by this work, we also added skip connections into the second part of our model as indicated by the solid gray arrow in Fig. 3. For each convolutional block in weighted Res-UNet, the skip connection scheme is implemented by using the following equation

$$\boldsymbol{y} = F(\boldsymbol{x}, \{\boldsymbol{w}_i\}) + H(\boldsymbol{x}), \tag{4}$$

where F consists of two convolution operations and one max-pooling or one up-sampling operation, H is either the identical mapping or a convolution operation to keep the input with the same feature dimensions as F.

Notice that, as introduced in Section 3.2, the first part of our model is the weighted U-Net without residual connection. In this way, in coarse segmentation step, our model can utilize the weight of network pre-trained on ImageNet [53].

3.2.3 Loss Function

In order to train the proposed model, we choose the binary cross entropy as segmentation loss function for both the weighted U-Net and the weighted Res-UNet, which goes as

$$\mathcal{L}(p,q) = -\frac{1}{n}\sum_{k=1}^{n} q_k \log p_k + (1-q_k)\log(1-p_k), \tag{5}$$

where n represents the total number of training pixels, p and q represent predicted probability and its corresponding groundtruth.

3.3 Patch Regrouping

During the testing phase, instead of doing a random overlapping patch cropping as in the training phase, we extract 64×64 image patches in sequence with an overlapping stride of 8. For each 512×512 testing retinal image, the number of overall extracted patches is $((512-64)/8+1) * ((512-64)/8+1) = 3249$. After getting the prediction of each patch, we can easily get the final segmentation result of the whole retinal image by regrouping the results from all the patches according to their positions. Specifically, for all the overlapped pixels, our computed the probability map by averaging the values of all the overlapped pixels' predictions.

4 Experiments

4.1 Datasets

We evaluate the performance of our method on two publicly benchmark datasets: DRIVE [27] and STARE [51]. Both the two datasets are representative dataset in the task of retinal vessel segmentation. The details of the two datasets are listed in Table 1.

TABLE 1
The details of the two datasets used for evaluation

Dataset	DRIVE	STARE
Sensor	Canon CR5 non-mydriatic 3CCD camera	TopCon TRV-50 fundus camera
Resolution	565*584	700*605
Format	tif	ppm
Color	RGB	RGB
# of training	20	10
# of testing	20	10

DRIVE[1] dataset contains 20 training RGB images and 20 testing RGB images with a resolutions of 768×584 pixels. There are 7 images shown signs of mild early diabetic retinopathy in this dataset.

STARE[2] dataset consists of 20 retinal fundus slides captured by a TopCon TRV-50 fundus camera. Half of the dataset comprises images of healthy subjects, and the rest contains the pathological cases which make the segmentation more challenging.

For both the DRIVE and STARE dataset, there are two groups of manual segmentation masks available annotated by two independent human experts. The manual annotations of the first expert were used as the ground truth for training and evaluation.

4.2 Evaluation Metrics

In order to compare the performance of different methods, three different metrics have been used for evaluation, including Accuracy, Sensitivity and Specificity.

Accuracy is a widely used evaluation metric for the task of binary segmentation, which computes the percentage of correctly classified pixels in the whole image. Eq (6) is used for calculating the Accuracy on test set,

$$Acc = \frac{TP + TN}{TP + FN + TN + FP}, \tag{6}$$

where TP, TN, FP and FN represent the number of true positive, true negatives, false positives and false negatives, respectively.

Sensitivity (also referred as recall and true positive rate) is another commonly used statistical measures of the performance of binary segmentation task. It measures the proportion of actual positives that are correctly classified as such. The equation goes as

$$Sen = \frac{TP}{TP + FN}. \tag{7}$$

Specificity (also called as true negative rate) measures the proportion of actual negatives that are correctly identified as such. The equation for computing the Specificity is

$$Spec = \frac{TN}{TN + FP}, \tag{8}$$

We apply these three metrics on each testing image and then report the final average value on the testing set for comparison.

1. https://www.isi.uu.nl/Research/Databases/DRIVE/
2. http://cecas.clemson.edu/ ahoover/stare/

4.3 Implementation

We implement our model in an end-to-end manner by using the Keras with TensorFlow backend. The outputs of the first part concatenate with LCP patches directly together with attention masks, will act as input of the second part. The Adam optimizer is adopted to train our model with an initial learning rate of $5 * 10^{-5}$, and we set the training epochs for both parts to be 30. The contracting paths in both parts of our proposed model have the same structure as VGG16 [33] (without fully connected layers). Then we initialize the weights of the first part (weighted U-Net) by using the pre-trained weights on ImageNet provided by Keras.

The proposed model is trained and tested following the pipeline discussed in Section 3 on a machine with Intel i7-7700K CPU and an NVIDIA 1080Ti GPU. Due to the number of training images is very limited, we randomly extracted 500 overlapped $64 * 64$ patches in each training image. Moreover, we adopt the widely used data augmentation operations in each extracted patch, such as horizontal flip, width shift range, etc. We first train the parameters for the first weighted U-Net for 30 epochs, and then training the second weighted Res-UNet for another 30 epochs by fixing the weights of the first part. The whole training procedure takes around 2 hours to finish.

4.4 Experiment Results

In this section, we evaluate the segmentation performance of the proposed GLUE model. We first implement an original U-Net model without the attention scheme and the skip connection on both datasets as the baseline model. Both our proposed model and the baseline U-Net model use the backbone of VGG16, and initialize with the weights pre-trained on ImageNet. Same data augmentation schemes are utilized on both approaches during the training step. We adopt Accuracy, Sensitivity and Specificity as evaluation metrics, as is introduced in Section 4.2. Specifically, for evaluating the performance of both our model's coarse segmentation results and the effectiveness of model refinement, we list the evaluation on the coarse results, which act as the intermediate result in our approach. We also compared with other representative state-of-the-art methods, including unsupervised methods and supervised methods.

The detailed performance comparison of our method and other representative methods are listed in Table 2. From the table, we can find that our proposed GLUE model can significantly surpass the baseline U-Net model on accuracy, sensitivity and specificity performance. And the proposed approach is close to the results of manual segmentation results by the second human observer. Also, the proposed model is in the leading position on each metric compared to other methods, except that [43] and [18] have higher accuracy on DRIVE and STARE, respectively. One thing we need to note is that many approaches (including ours) surpass the manual segmentation results by the second observer on some metrics, that is because retinal vessel segmentation is a hard task and the procedure of manual segmentation can be considered highly subjective (examples can be seen in Fig. 5).

In Fig. 4, we plot some segmentation results produced by the proposed model and the baseline U-Net model. Results show that our model can cope well with the difficulties, including low illumination, hard area of the optic disk, etc. Besides, results from our proposed model show better vessel connectivity and small vessel sensitivity than the ones from U-Net. For better displaying

TABLE 2
Accuracy, Sensitivity and Specificity comparisons on proposed Weighted Res-UNet with refinement and other promising approaches.

Method	Year	DRIVE			STARE		
		Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
2nd human expert	-	0.9725	0.7760	0.9725	0.9346	0.8956	0.9381
Unsupervised methods							
Yin *et al.* [14]	2013	0.9267	0.6522	0.9710	0.9420	0.7034	0.9668
Wang *et al.* [18]	2013	0.9461	-	-	0.9682	-	-
Fraz *et al.* [20]	2013	0.9422	0.7302	0.9742	0.9423	0.7318	0.9660
Odstrcilik *et al.* [54]	2013	0.9340	0.7060	0.9693	0.9341	0.7847	0.9512
Imani *et al.* [21]	2015	0.9523	0.7524	0.9753	0.9590	0.7502	0.9745
Mapayi *et al.* [25]	2015	0.9461	0.7632	0.9634	0.9510	0.7626	0.9657
Kumar *et al.* [55]	2016	0.9626	0.7006	-	0.9637	0.7675	-
Neto *et al.* [56]	2017	-	0.7942	0.9631	-	0.7695	0.9537
Khan *et al.* [57]	2017	0.9600	0.7470	0.980	0.9510	0.7780	0.9660
Khan *et al.* [58]	2018	0.9506	0.7696	0.9651	0.9513	0.7521	0.9812
Supervised methods							
Ricci *et al.* [28]	2007	0.9595	-	-	0.9646	-	-
Fraz *et al.* [59]	2012	0.9480	0.7406	0.9807	0.9534	0.7548	0.9763
Emary *et al.* [60]	2014	0.9376	0.6316	0.9838	0.9448	0.5864	0.9871
Wang *et al.* [44]	2015	0.9767	0.8173	0.9733	**0.9813**	0.8104	0.9791
Roychowdhury *et al.* [61]	2015	0.9519	0.7249	0.9830	0.9515	0.7719	0.9726
Liskowski *et al.* [62]	2016	0.9495	0.7763	0.9768	0.9566	0.7867	0.9754
Fu *et al.* [46]	2016	0.9523	0.7603	-	0.9585	0.7412	-
Dasgupta *et al.* [47]	2017	0.9533	0.7691	0.9801	-	-	-
Orlando *et al.* [41]	2017	0.9454	0.7897	0.9684	0.9571	0.7773	0.9789
Oliveira *et al.* [43]	2018	**0.9821**	0.8039	0.9804	0.9694	0.8315	0.9858
Memari *et al.* [42]	2018	0.961	0.761	0.981	0.951	0.782	0.965
U-Net	-	0.9594	0.7698	0.9798	0.9669	0.7536	0.9827
Ours coarse result(from 1st part)	2018	0.9625	0.7536	0.9737	0.9681	0.7461	0.9825
Ours	2018	0.9692	**0.8278**	**0.9861**	0.9740	**0.8342**	**0.9916**

the details of segmentation results, we also plot extracted patches' segmentation results in zoom-in view in Fig. 5. The examples selected indicated that compared to U-Net, our model can accurately segment tiny and indistinct vessels, and maintain the geometric connection of retinal vessels. Specificity, as is indicated by red arrows in No.4 result of STARE, our model can overcome the effects of lesions and abnormal areas in retinal images. As is mentioned above, retinal vessel segmentation is a challenging task and can be considered highly subjective. Due to the low contrast, the classification of many areas is ambiguous. In fact, the misclassified area marked out by green arrows (in No.2 6 in DRIVE and No.5 in STARE, which regarded as false positive) should actually be true positive. This shows that after a series of effective processing and training, in some cases, our model has a higher sensitivity for small and indistinguishable vessels than manual segmentation.

5 Conclusion

High-performance retinal vessel segmentation is a key step for medical diagnosing, including chronic eye disease, cardiovascular disease and diabetic retinopathy, etc. While problems such as poor illumination, missing of small vessels, complex vessel geometry, make retina vessel segmentation a challenging task. Former learning-based approaches generally apply an image-level pre-processing operation on whole retina image for segmentation model learning. However, retina region's appearance is a unitary area, and pixels outside the such area is of no use for segmentation. Moreover, as introduced in Sec 3.1, applying CLAHE operation globally and locally are complementary schemes with each other. So, in this paper, we proposed the GLUE model for addressing the challenging retinal vessel segmentation problem. Our model contains a weighted U-Net for coarse segmentation and a weighted Res-UNet for segmentation refinement. By applying CLAHE operation globally and locally, our model can benefit from both the GCP and LCP patches' information. And by adding attention scheme and residual connection scheme, our model can focus only on retinal region in each patch, avoid false segmentation caused by irrelevant noisy background, learn more discriminative features and have a better maintaining of the retinal vessel tree structure.

We adopt accuracy, sensitivity and specificity for measuring the accuracy of predicted segmentation results. We evaluate our method on the widely used DRIVE and the STARE benchmark dataset, and compare it with U-Net model as baseline model, and many other representative methods. Experimental results demonstrate that our model can achieve high-performance segmentation results. Also, we list the evaluation results on our model's coarse segmentation predictions, and the results proves the validity of our cascade scheme. Detailed segmentation results in the zoom-in view show that our model can not only accurately segment tiny and indistinct vessels, but also maintain the geometric connection of retinal vessels. Also, our model can deal well with abnormal and noisy regions like lesion areas, edge areas and optic disk areas. Since the procedure of manual segmentation can be considered highly subjective, we expect larger datasets and more accurate manual annotation for further improving the performance of our approach.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (No. 61572409, No. 61876159, No. 61806172, No. U1705286, & No. 61571188), Fujian Province 2011Collaborative Innovation Center of TCM Health Management, Collaborative Innovation Center of Chinese Oolong Tea Industry-Collaborative Innovation Center (2011) of Fujian Province, Fund for Integration of Cloud Computing and Big Data, Innovation of Science and Education (No. 2017A11032), the Open Project Program of the Key

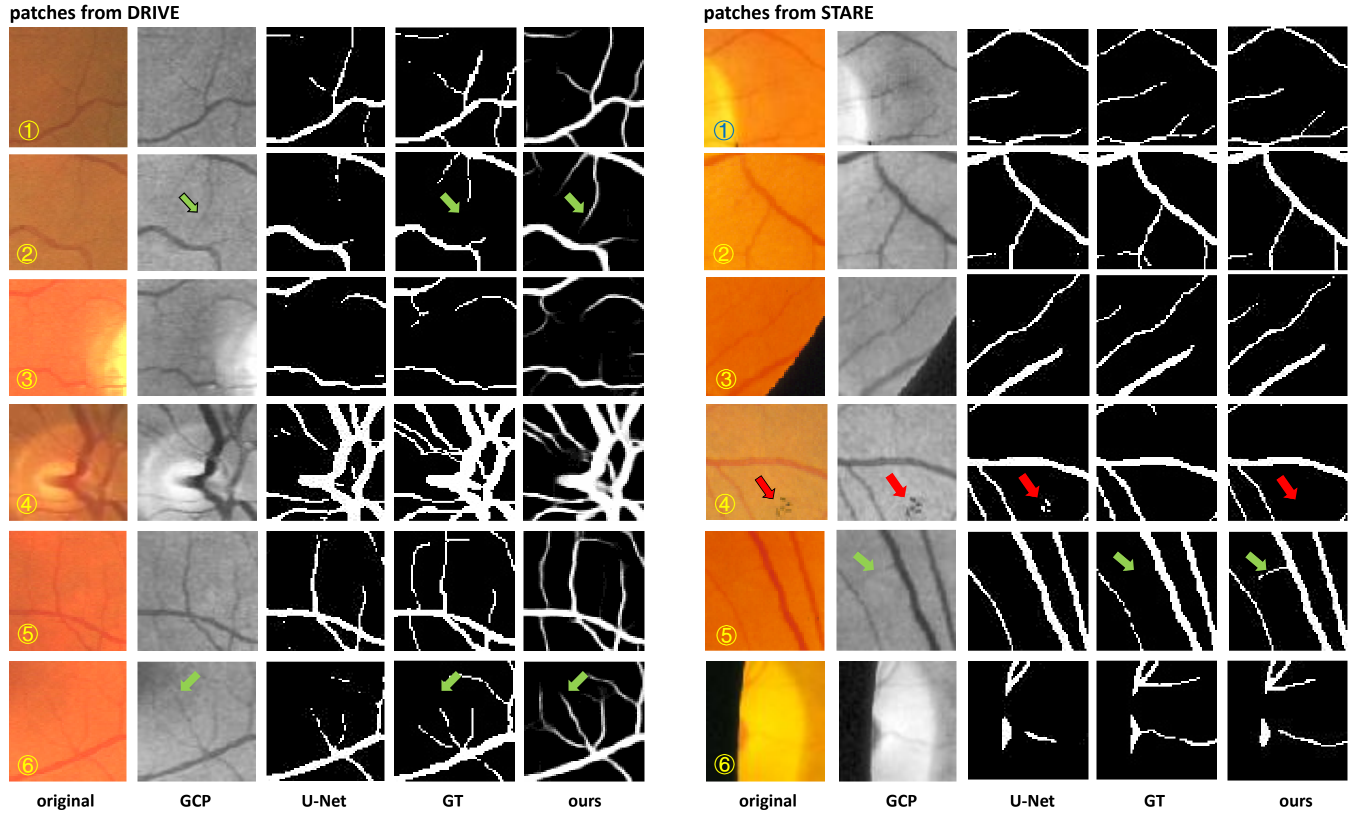

Fig. 5. Examples of extracted patches' segmentation result from DRIVE dataset (listed in the left) and STARE dataset (listed in the right). The results show that compared with U-Net, our model is able to find small blood vessels and maintain better vessel connectivity.

Laboratory of Cognitive Computing and Intelligent Information Processing of Fujian Education Institutions(Wuyi University).

REFERENCES

[1] Joanne WY Yau, Sophie L Rogers, Ryo Kawasaki, Ecosse L Lamoureux, Jonathan W Kowalski, Toke Bek, Shih-Jen Chen, Jacqueline M Dekker, Astrid Fletcher, Jakob Grauslund, et al. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes care*, page DC_111909, 2012.

[2] Wan Ling Wong, Xinyi Su, Xiang Li, Chui Ming G Cheung, Ronald Klein, Ching-Yu Cheng, and Tien Yin Wong. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *The Lancet Global Health*, 2(2):e106–e116, 2014.

[3] Yih-Chung Tham, Xiang Li, Tien Y Wong, Harry A Quigley, Tin Aung, and Ching-Yu Cheng. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*, 121(11):2081–2090, 2014.

[4] Sveinn Hakon Hardarson and Einar Stefánsson. Retinal oxygen saturation is altered in diabetic retinopathy. *British journal of ophthalmology*, 96(4):560–563, 2012.

[5] Tien Yin Wong and Rachel McIntosh. Hypertensive retinopathy signs as risk indicators of cardiovascular morbidity and mortality. *British medical bulletin*, 73(1):57–70, 2005.

[6] Alice V Stanton, Balvinder Wasan, Alessia Cerutti, Susan Ford, Ronald Marsh, Peter P Sever, Simon A Thom, and Alun D Hughes. Vascular network changes in the retina with age and hypertension. *Journal of hypertension*, 13(12 Pt 2):1724–1728, 1995.

[7] Brad Bowling. *Kanski's Clinical Ophthalmology E-Book: A Systematic Approach*. Elsevier Health Sciences, 2015.

[8] Debapriya Maji, Anirban Santara, Pabitra Mitra, and Debdoot Sheet. Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images. *arXiv: Learning*, 2016.

[9] José Ignacio Orlando and Matthew Blaschko. Learning fully-connected crfs for blood vessel segmentation in retinal images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 634–641. Springer, 2014.

[10] Huazhu Fu, Yanwu Xu, Damon Wing Kee Wong, and Jiang Liu. Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 698–701. IEEE, 2016.

[11] Xiao Xiao, Sheng Lian, Zhiming Luo, and Shaozi Li. Weighted res-unet for high-quality retina vessel segmentation. In *International Conference on Information Technology in Medicine and Education (ITME)*, pages 1–5. IEEE, 2018.

[12] Muhammad Moazam Fraz, Paolo Remagnino, Andreas Hoppe, Bunyarit Uyyanonvara, Alicja R Rudnicka, Christopher G Owen, and Sarah A Barman. Blood vessel segmentation methodologies in retinal images–a survey. *Computer methods and programs in biomedicine*, 108(1):407–433, 2012.

[13] Chetan L Srinidhi, P Aparna, and Jeny Rajan. Recent advancements in retinal vessel segmentation. *Journal of medical systems*, 41(4):70, 2017.

[14] Yi Yin, Mouloud Adel, and Salah Bourennane. Automatic segmentation and measurement of vasculature in retinal fundus images using probabilistic formulation. *Computational and mathematical methods in medicine*, 2013, 2013.

[15] Jia Zhang, Huiqi Li, Qing Nie, and Li Cheng. A retinal vessel boundary tracking method based on bayesian theory and multi-scale line detection. *Computerized Medical Imaging and Graphics*, 38(6):517–525, 2014.

[16] Jaydeep De, Huiqi Li, and Li Cheng. Tracing retinal vessel trees by transductive inference. *BMC bioinformatics*, 15(1):20, 2014.

[17] Subhasis Chaudhuri, Shankar Chatterjee, Norman Katz, Mark Nelson, and Michael Goldbaum. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Transactions on medical imaging*, 8(3):263–269, 1989.

[18] Yangfan Wang, Guangrong Ji, Ping Lin, and Emanuele Trucco. Retinal vessel segmentation using multiwavelet kernels and multiscale hierarchical decomposition. *Pattern Recognition*, 46(8):2117–2133, 2013.

[19] Michael Krause, Ralph Maria Alles, Bernhard Burgeth, and Joachim Weickert. Fast retinal vessel analysis. *Journal of Real-Time Image Processing*, 11(2):413–422, 2016.

[20] Muhammad Moazam Fraz, A Basit, and SA Barman. Application of morphological bit planes in retinal blood vessel extraction. *Journal of digital imaging*, 26(2):274–286, 2013.

[21] Elaheh Imani, Malihe Javidi, and Hamid-Reza Pourreza. Improvement of retinal blood vessel detection using morphological component analysis. *Computer methods and programs in biomedicine*, 118(3):263–279, 2015.

[22] Eysteinn Már Sigursson, Silvia Valero, Jón Atli Benediktsson, Jocelyn Chanussot, Hugues Talbot, and Einar Stefánsson. Automatic retinal vessel extraction based on directional mathematical morphology and fuzzy classification. *Pattern Recognition Letters*, 47:164–171, 2014.

[23] Xiaoyi Jiang and Daniel Mojon. Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1):131–137, 2003.

[24] M Usman Akram and Shoab A Khan. Multilayered thresholding-based blood vessel segmentation for screening of diabetic retinopathy. *Engineering with computers*, 29(2):165–173, 2013.

[25] Temitope Mapayi, Serestina Viriri, and Jules-Raymond Tapamo. Adaptive thresholding technique for retinal vessel segmentation based on glcm-energy information. *Computational and mathematical methods in medicine*, 2015, 2015.

[26] Meindert Niemeijer, Joes Staal, Bram van Ginneken, Marco Loog, and Michael D Abramoff. Comparative study of retinal vessel segmentation methods on a new publicly available database. In *Medical Imaging 2004: Image Processing*, volume 5370, pages 648–657. International Society for Optics and Photonics, 2004.

[27] Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imag.*, 23(4):501–509, 2004.

[28] Elisa Ricci and Renzo Perfetti. Retinal blood vessel segmentation using line operators and support vector classification. *IEEE transactions on medical imaging*, 26(10):1357–1365, 2007.

[29] Lili Xu and Shuqian Luo. A novel method for blood vessel detection from retinal images. *Biomedical engineering online*, 9(1):14, 2010.

[30] Reza Nekovei and Ying Sun. Back-propagation network and its configuration for blood vessel detection in angiograms. *IEEE Transactions on Neural Networks*, 6(1):64–72, 1995.

[31] Chanjira Sinthanayothin, James F Boyce, Helen L Cook, and Thomas H Williamson. Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *British Journal of Ophthalmology*, 83(8):902–910, 1999.

[32] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[33] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[34] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. CVPR*, pages 770–778, 2016.

[35] Yujie Feng, Fan Yang, Xichuan Zhou, Yanli Guo, Fang Tang, Fengbo Ren, Jishun Guo, and Shuiwang Ji. A deep learning approach for targeted contrast-enhanced ultrasound based prostate cancer detection. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2018.

[36] Ronghui Ju, Chenhui Hu, Quanzheng Li, et al. Early diagnosis of alzheimer's disease based on resting-state brain networks and deep learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2017.

[37] Yangqin Feng, Lei Zhang, and Juan Mo. Deep manifold preserving autoencoder for classifying breast cancer histopathological images. *IEEE/ACM transactions on computational biology and bioinformatics*, 2018.

[38] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35:18–31, 2017.

[39] MR Avendi, Arash Kheradvar, and Hamid Jafarkhani. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac mri. *Medical image analysis*, 30:108–119, 2016.

[40] Clement Zotti, Zhiming Luo, Alain Lalande, and Pierre-Marc Jodoin. Convolutional neural network with shape prior applied to cardiac mri segmentation. *IEEE journal of biomedical and health informatics*, 2018.

[41] José Ignacio Orlando, Elena Prokofyeva, and Matthew B Blaschko. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. *IEEE Transactions on Biomedical Engineering*, 64(1):16–27, 2017.

[42] Nogol Memari, Abd Rahman Ramli, M Iqbal Bin Saripan, Syamsiah Mashohor, and Mehrdad Moghbel. Retinal blood vessel segmentation by using matched filtering and fuzzy c-means clustering with integrated level set method for diabetic retinopathy assessment. *Journal of Medical and Biological Engineering*, pages 1–19, 2018.

[43] Américo Filipe Moreira Oliveira, Sérgio Rafael Mano Pereira, and Carlos Alberto Batista Silva. Retinal vessel segmentation based on fully convolutional neural networks. *Expert Systems with Applications*, 2018.

[44] Shuangling Wang, Yilong Yin, Guibao Cao, Benzheng Wei, Yuanjie Zheng, and Gongping Yang. Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing*, 149:708–717, 2015.

[45] Aaron Wu, Ziyue Xu, Mingchen Gao, Mario Buty, and Daniel J Mollura. Deep vessel tracking: A generalized probabilistic approach via deep learning. In *Proc. ISBI*, pages 1363–1367, 2016.

[46] Huazhu Fu, Yanwu Xu, Stephen Lin, Damon Wing Kee Wong, and Jiang Liu. Deepvessel: Retinal vessel segmentation via deep learning and conditional random field. In *Proc. MICCAI*, pages 132–139, 2016.

[47] Avijit Dasgupta and Sonam Singh. A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation. In *Proc. ISBI*, pages 248–251, 2017.

[48] Jaemin Son, Sang Jun Park, and Kyu-Hwan Jung. Retinal vessel segmentation in fundoscopic images with generative adversarial networks. *arXiv preprint arXiv:1706.09318*, 2017.

[49] Boheng Zhang, Shenglei Huang, and Shaohan Hu. Multi-scale neural networks for retinal blood vessels segmentation. *arXiv preprint arXiv:1804.04206*, 2018.

[50] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Proc. MICCAI*, pages 234–241, 2015.

[51] AD Hoover, Valentina Kouznetsova, and Michael Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Med. Imag.*, 19(3):203–210, 2000.

[52] Sheng Lian, Zhiming Luo, Zhun Zhong, Xiang Lin, Songzhi Su, and Shaozi Li. Attention guided u-net for accurate iris segmentation. *Journal of Visual Communication and Image Representation*, 56:296–304, 2018.

[53] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. Ieee, 2009.

[54] Jan Odstrcilik, Radim Kolar, Attila Budai, Joachim Hornegger, Jiri Jan, Jiri Gazarek, Tomas Kubena, Pavel Cernosek, Ondrej Svoboda, and Elli Angelopoulou. Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. *IET Image Processing*, 7(4):373–383, 2013.

[55] Debamita Kumar, Ankita Pramanik, Sudeshna Sil Kar, and Santi P Maity. Retinal blood vessel segmentation using matched filter and laplacian of gaussian. In *Proc. SPCOM*, pages 1–5, 2016.

[56] Luiz Câmara Neto, Geraldo LB Ramalho, Jeová FS Rocha Neto, Rodrigo MS Veras, and Fátima NS Medeiros. An unsupervised coarse-to-fine algorithm for blood vessel segmentation in fundus images. *Expert Systems with Applications*, 78:182–192, 2017.

[57] Khan Bahadar Khan, Amir A Khaliq, and Muhammad Shahid. A novel fast glm approach for retinal vascular segmentation and denoising. *Journal of Information Science and Engineering*, 33(6):1611–1627, 2017.

[58] Mohammad AU Khan, Tariq M Khan, DG Bailey, and Toufique A Soomro. A generalized multi-scale line-detection method to boost retinal vessel segmentation sensitivity. *Pattern Analysis and Applications*, pages 1–20, 2018.

[59] Muhammad Moazam Fraz, Paolo Remagnino, Andreas Hoppe, Bunyarit Uyyanonvara, Alicja R Rudnicka, Christopher G Owen, and Sarah A Barman. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Transactions on Biomedical Engineering*, 59(9):2538–2548, 2012.

[60] Eid Emary, Hossam M Zawbaa, Aboul Ella Hassanien, Gerald Schaefer, and Ahmad Taher Azar. Retinal vessel segmentation based on possibilistic fuzzy c-means clustering optimised with cuckoo search. In *Proc. IJCNN*, pages 1792–1796, 2014.

[61] Sohini Roychowdhury, Dara D Koozekanani, and Keshab K Parhi. Blood vessel segmentation of fundus images by major vessel extraction and subimage classification. *IEEE journal of biomedical and health informatics*, 19(3):1118–1128, 2015.

[62] Paweł Liskowski and Krzysztof Krawiec. Segmenting retinal blood vessels with deep neural networks. *IEEE Trans. Med. Imag.*, 35(11):2369–2380, 2016.

Sheng Lian is currently a PHD student at Xiamen University, China. He received his bachelor degree in Computer Science from Huazhong University of Science and Technology, China. His research interests include medical image analysis, computer vision and machine learning.

Lei Li is currently a master student at Xiamen University, China. His research interests include object detection, medical image analysis, computer vision and marching learning.

Guiren Lian is currently serve as the associate professor of the College of Mathematics and Informatics, Fujian Normal University, China. His research interests include medical imaging and computer vision.

Xiao Xiao is currently a master student at Xiamen University, China . Her research interests include computer vision and machine learning.

Zhiming Luo is currently a postdoctoral researcher at Xiamen University, China. He received his Ph.D. degree in Computer Science from Xiamen University, China and University of Sherbrooke, Canada. His research interests include traffic surveillance video analytics, medical image analysis, computer vision and machine learning.

Shaozi Li received the B.S. degree from Hunan University, and the M.S. degree from Xi'an Jiaotong University, and the Ph.D. degree from National University of Defense Technology. He currently serves as the Chair and Professor of Cognitive Science Department of Xiamen University, the Vice Director of Technical Committee on Collaborative Computing of CCF, the Vice Director of the Fujian Association of Artificial Intelligence. He is also the senior Member of IEEE, ACM and China Computer Federation (CCF). His research interests cover Artificial Intelligence and Its Applications, Moving Objects Detection and Recognition, Machine Learning, Computer Vision, Multimedia Information Retrieval, etc. He has directed and completed more than twenty research projects, including several National 863 Programs, National Nature Science Foundation of China, Ph.D. Programs Foundation of Ministry of Education of China.