Weighted Res-UNet for High-quality Retina Vessel Segmentation

Xiao Xiao*, Sheng Lian*, Zhiming Luo†‡, and Shaozi Li*
*Cognitive Science Department, Xiamen University, China
†Postdoc Center of Information and Communication Engineering, Xiamen University, China
‡The Key Laboratory of Cognitive Computing and Intelligent Information Processing of Fujian Education Institutions, Wuyishan, China

***Abstract*—Retinal vessel segmentation is a key step towards the accurate visualization, diagnosis, early treatment and surgery planning of ocular diseases. Recently, deep learning based retinal vessel segmentation methods have reached the state-of-the-art performance. Due to the extreme variations in the morphology of the vessels against the noisy background, these methods still have issues of dealing with small thin vessels, low discriminative ability at the optic disk area, etc. In this paper, we proposed a U-Net-like model with the weighted attention mechanism and the skip connection scheme for addressing these issues. Experimental results on two benchmark datasets demonstrate the effectiveness of the proposed methods.**

***Keywords*-Retinal Vessel Segmentation, Deep Learning, Weighted Res-UNet, Convolutional Neural Network**

I. Introduction

The retina is the only human body part which can observe the microcirculation through a noninvasive fundus examination, and the retinal vessel can serve as an important signal for diagnosing chronic eye diseases, cardiovascular diseases, and diabetic retinopathy [6]. The task of retinal vessel segmentation is to determine pixels in given retinal images that belong to the vessel region. A good segmentation result will make follow-up feature extraction and abnormal detection steps more efficient and accurate. Manual retinal vessel segmentation is a time consuming and tedious task even for an ophthalmologist. Computer-aided automatic retinal vessel segmentation is essential for reducing medical costs and improving efficiency. Although the quality of fundus imaging has improved significantly with the improvement of new technology [10], retinal vessel segmentation is still a challenging task. The factors that render retinal vessel segmentation a challenging task can be briefly summarized as follows:

- *Missing small vessels:* Small blood vessels located at the end of branches are sometimes indistinguishable even by human eyes.
- *Poor segmentation at optic disk area*: Optic disk areas are often brighter and with lower contrast, which makes retinal vessels hard to be segmented.
- *Failing to maintain the structure relationship*: Retinal vessels have bifurcated structure similar to trees. But such a structure is difficult to maintain when vessels are too thin to detect.
- *Illumination:* Poor or overexposed illumination, including light reflection caused by the light source of camera, will reduce images contrast, which results in the non-sharp boundary of retinal vessels.

For addressing these challenges, in this work, we propose an accurate and robust retina vessel segmentation model, namely weighted Res-UNet. Our model is built upon the original UNet model [11], while we made several significant improvements over the original architecture by adding a weighted attention mechanism and the skip connection scheme as introduced in [4]. In this way, 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.

The rest of the paper is organized as follows. Section II reviews the most recent deep learning based retinal vessel segmentation methods. In Section III, we describe the proposed method. In Section IV, we reported the experimental results. Finally, we conclude our paper in Section V.

II. Related 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 blood vessels without any labeled information, including conventional matched filtering, vessel tracking, thresholding, etc. Supervised methods segment retinal vessels by learning a model from a training set annotated by experienced Ophthalmologists. With the success of deep learning in computer vision, Convolutional Neural Networks (CNN) has become an efficient approach for solving segmentation problems and can reach the state-of-the-art performance in the retinal vessel segmentation. In this section, we will review several CNNs based retinal vessel segmentation methods.

Wang et al. [14] proposed a segmentation method which uses CNN as a feature extractor and random forests as the final classifier. Wu et al. [15] 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. [3] 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. [1] utilized the fully convolutional neural networks and the multi-label inference to do structure predictions of the blood vessel. Son et al. [12] presented a method that adopts the

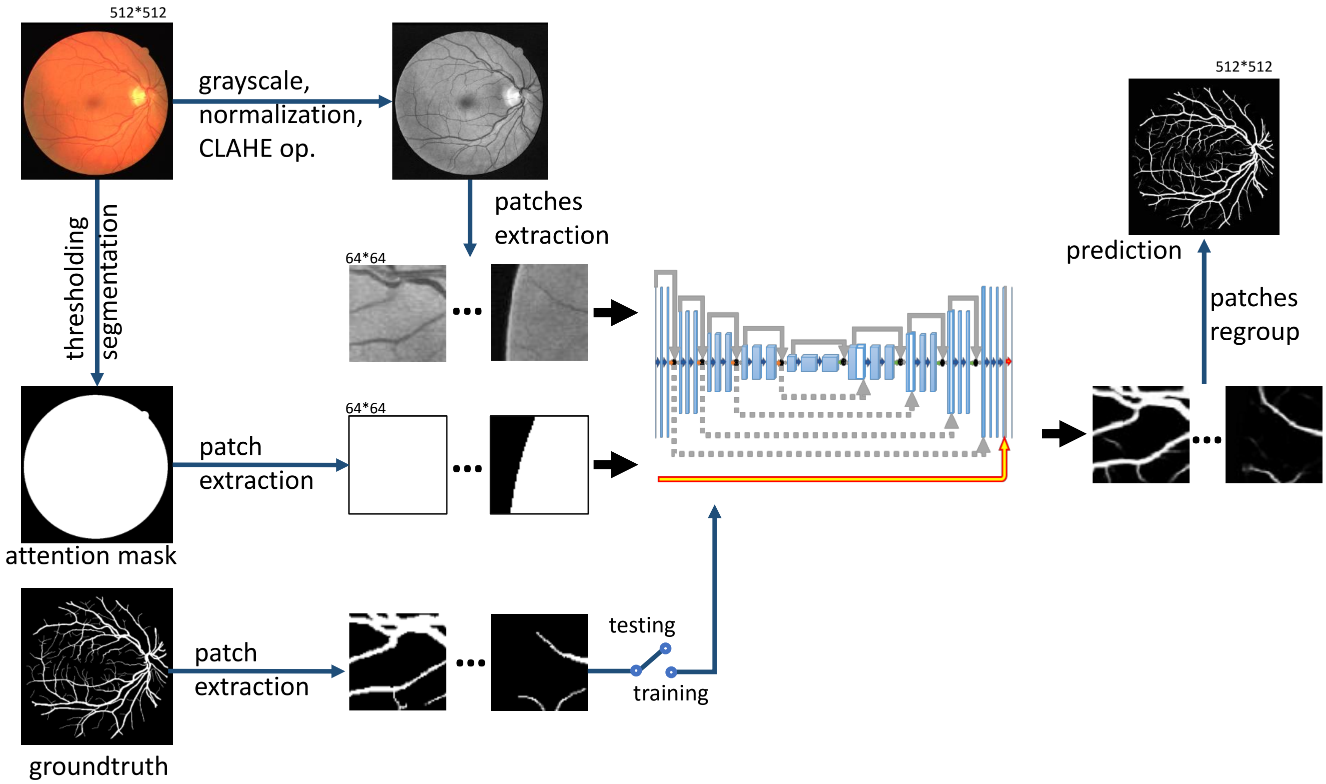

Figure 1. The overall pipeline of our approach. The thumbnail of our proposed Weighted Res-UNet(as is displayed in Fig. 2) is drawn in the middle of the picture.

generative adversarial training to improve the segmentation performance. Recently, Zhang et al. [16] 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.

III. Proposed Method

In this section, we describe the proposed weighted Res-UNet for addressing the retina vessel segmentation in detail. The overall pipeline of the proposed method is shown in Figure 1.

A. Fundus image pre-processing

As can be seen from Figure 1, the original fundus image is with extremely low contrast, and CNNs can have better performance with suitable pre-processed input. In this study, we used the contrast limited adaptive histogram equalization (CLAHE) operation as a pre-processing step to enhance the image contrast. For each fundus image, we resize it to $512 * 512$ and convert it into grayscale. Then we perform the CLAHE operation to normalize the grayscale image I by using the Equation 1,

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

where μ and σ are the mean and standard deviation of the grayscale image I.

Due to the number of training images are 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. By doing so, the amount of training images is enlarged by more than 500 times.

B. The architecture of the proposed weighted Res-UNet.

The overall architecture of the proposed weighted Res-UNet is shown in Figure 2. Similar to the original U-Net model [11], the proposed model consists of an encoder-decoder architecture. While beyond such architecture, we made several significant improvements over the original architecture by adding a weighted attention mechanism and the skip connection scheme as introduced in [4]. All the parameters of the model are given in Figure 2.

1) The weighted attention mechanism: The fundus images in DRIVE [13] and STARE [5] are with circular-like region of interest (ROI) and dark background. Then a circular template ROI mask $\mathcal{M}$ will be used as a weighted attention which is the yellow arrow as shown in Figure 2.

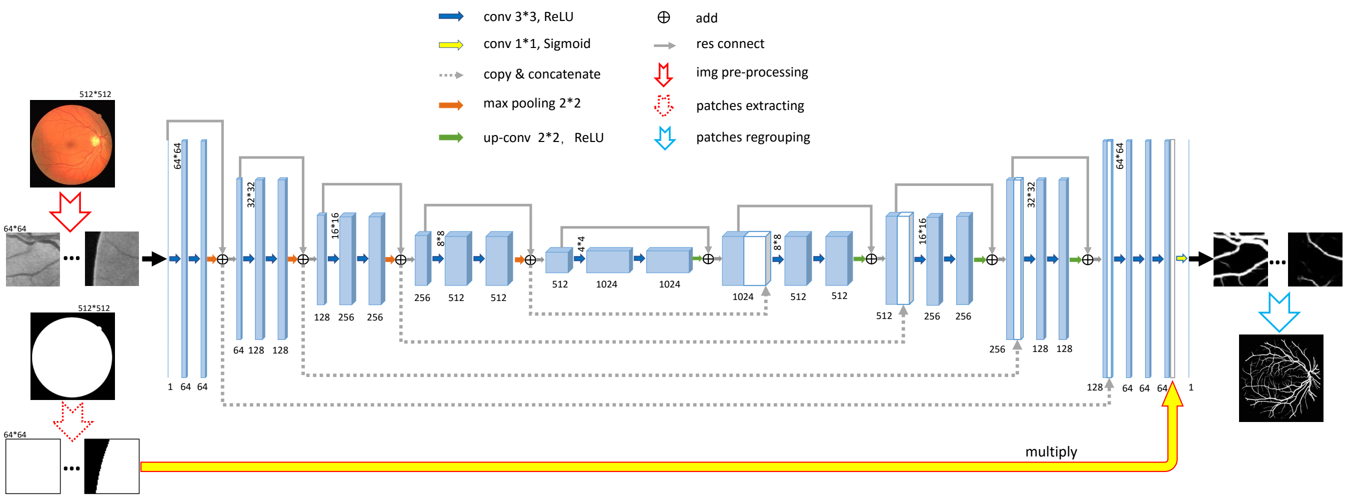

Figure 2. The brief architecture of proposed Weighted Res-UNet.The meaning of all types of signs is marked on the top of the picture.

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

$$\mathcal{I}(x,y) = \begin{cases} \mathcal{F}(x,y) * 1.0 & (x,y) \in \mathcal{M} \\ \mathcal{F}(x,y) * 0.0 & (x,y) \notin \mathcal{M} \end{cases}, \tag{2}$$

in which $\mathcal{F}(x,y)$ represents the features at position (x,y).

By using this weighted attention mechanism, our model will only pay attention at the target ROI area and discard the irrelevant 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 fundus image into grayscale, apply Gaussian filtering and then separate the fundus regions by doing binary thresholding at the value 40.

2) The skip connection scheme: As demonstrated in [4], adding skip connection can increase the depth and improve the accuracy of deep CNNs. Inspired by this work, we also added skip connections into our model as the solid gray arrow in Figure 2. For each convolutional block, this skip connection scheme is implemented by using following equation

$$\boldsymbol{y} = F(\boldsymbol{x}, \{\boldsymbol{w}_i\}) + H(\boldsymbol{x}), \tag{3}$$

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 have the same feature dimensions as F.

3) Loss Function: In order to train the proposed model, we choose the binary cross entropy as segmentation loss function, which goes as

$$\mathcal{L}(p,q) = -\frac{1}{n}\sum_{k=1}^{n} q_k \log p_k + (1-q_k)\log(1-p_k), \tag{4}$$

where n represents the number of pixels in each image, p and q represent predicted pixel value and its corresponding groundtruth.

C. Patches regroup

During the testing phase, instead of doing a random overlapping patch cropping as in the training phase, we simply tiled the 512×512 input to 8×8 no-overlapping patches of size 64×64. By doing this, after getting the prediction for each segmentation patches, we can easily get the segmentation results of the whole fundus image by regrouping the results from all the patches according to their positions.

IV. Experiments

A. Datasets

We evaluate the performance of our method on two publicly benchmark datasets: DRIVE [13] and STARE [5].

DRIVE[1] dataset contains 20 training images and 20 testing images. In this dataset, there are 7 images show signs of mild early diabetic retinopathy.

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.

B. Evaluation Metric

In order to compare the performance of different methods, two different metrics have been used for evaluation, there are Accuracy ($Acc = \frac{TP+TN}{TP+FN+TN+FP}$) and Sensitivity ($Sen = \frac{TP}{TP+FN}$), TP, TN, FP and FN represent the

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

Table I
MEAN ACCURACY AND MTPR COMPARISONS ON PROPOSED WEIGHTED RES-UNET AND OTHER PROMISING APPROACHES.

Method	DRIVE		STARE	
	Acc(%)	Sen(%)	Acc(%)	Sen(%)
Emary *et al.* [2]	0.938	0.628	0.9448	0.586
Liskowski *et al.* [9]	0.9495	-	0.9605	-
Fu *et al.* [3]	0.9523	0.7603	0.9585	0.7130
Kumar *et al.* [8]	0.9637	0.7675	0.9626	0.7006
Khan *et al.* [7]	0.9506	0.7696	0.9513	**0.7521**
U-Net	0.9585	0.6565	0.9641	0.7361
Ours	**0.9655**	**0.7715**	**0.9693**	0.7469

number of true positive, true negatives, false positives and false negatives.

C. Experiment Results

In this section, we evaluate the segmentation performance of the proposed weighted Res-UNet. We first implement an original U-Net model without the attention and the skip connection as the baseline model. We also compared with other representative state-of-the-art methods.

The performance of our method and other methods are listed in Table I. From the table, we can find that our proposed Weighted Res-UNet can surpass the baseline U-Net model on both the accuracy and sensitivity performance. Also, the proposed model can outperform other methods only except that Khan et al. [7] has a higher sensitivity on the STARE dataset.

In Figure 3, 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 than the ones from U-Net. For better display the details of segmentation results, we also plot extracted patches' segmentation results in Figure 4. 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.

V. CONCLUSION

In this paper, we proposed a U-Net-like model with weighted attention mechanism and the skip connection scheme, called Weighted Res-UNet, for addressing the challenge retinal vessel segmentation problem. We evaluate our method on the DRIVE and the STARE benchmark dataset, the accuracy and sensitivity metrics demonstrate that our model can achieve high-performance segmentation results.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No. 61572409, No. U1705286 & No. 61571188), Fujian Province 2011 Collaborative Innovation Center of TCM Health Management and 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).

REFERENCES

[1] A. Dasgupta and S. Singh, "A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation," in *Proc. ISBI*, 2017, pp. 248–251.

[2] E. Emary, H. M. Zawbaa, A. E. Hassanien, G. Schaefer, and A. T. Azar, "Retinal vessel segmentation based on possibilistic fuzzy c-means clustering optimised with cuckoo search," in *Proc. IJCNN*, 2014, pp. 1792–1796.

[3] H. Fu, Y. Xu, S. Lin, D. W. K. Wong, and J. Liu, "Deepvessel: Retinal vessel segmentation via deep learning and conditional random field," in *Proc. MICCAI*, 2016, pp. 132–139.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, 2016, pp. 770–778.

[5] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, 2000.

[6] J. J. Kanski and B. Bowling, *Clinical ophthalmology: a systematic approach*. Elsevier Health Sciences, 2011.

[7] M. A. Khan, T. M. Khan, D. Bailey, and T. A. Soomro, "A generalized multi-scale line-detection method to boost retinal vessel segmentation sensitivity," *Pattern Analysis and Applications*, pp. 1–20, 2018.

[8] D. Kumar, A. Pramanik, S. S. Kar, and S. P. Maity, "Retinal blood vessel segmentation using matched filter and laplacian of gaussian," in *Proc. SPCOM*, 2016, pp. 1–5.

[9] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 11, pp. 2369–2380, 2016.

[10] D. Maji, A. Santara, P. Mitra, and D. Sheet, "Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images," *arXiv: Learning*, 2016.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.

[12] J. Son, S. J. Park, and K.-H. Jung, "Retinal vessel segmentation in fundoscopic images with generative adversarial networks," *arXiv preprint arXiv:1706.09318*, 2017.

[13] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, 2004.

[14] S. Wang, Y. Yin, G. Cao, B. Wei, Y. Zheng, and G. Yang, "Hierarchical retinal blood vessel segmentation based on feature and ensemble learning," *Neurocomputing*, vol. 149, pp. 708–717, 2015.

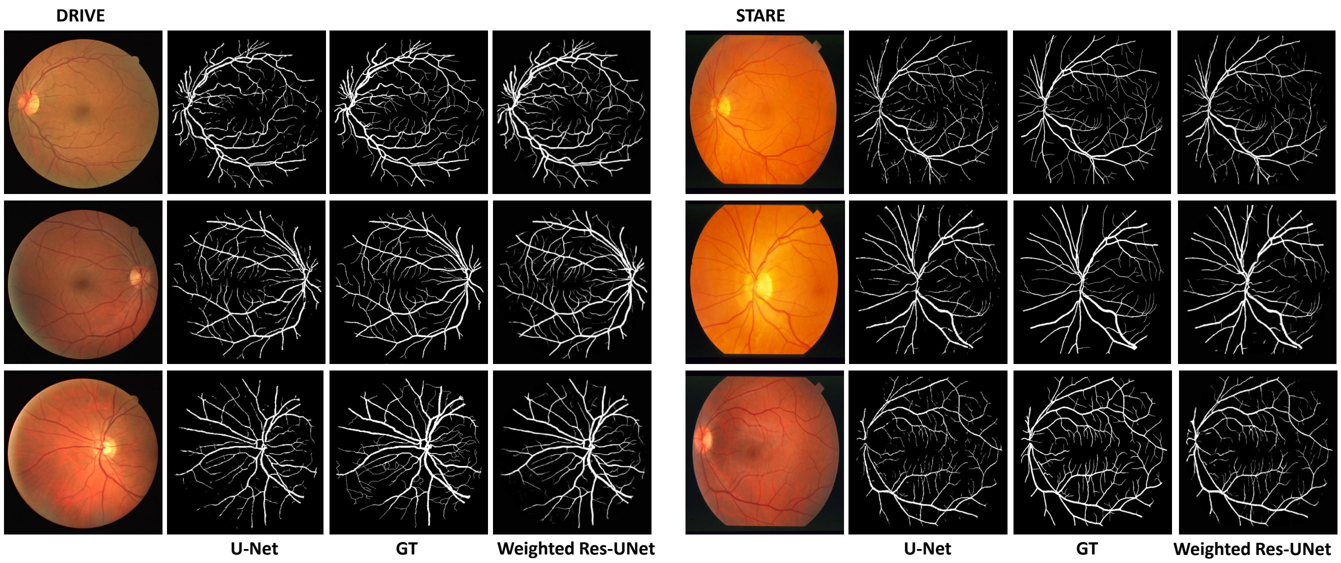

Figure 3. Examples of retina vessels segmentation result from DRIVE dataset(listed in the left) and STARE dataset(listed in the right). Results show that our model can deal well with the problems of low illumination, hard area of the optic disk. Moreover, our model shows better blood vessel connectivity than U-Net. Detailed results of extracted patches are displayed in Figure 4.

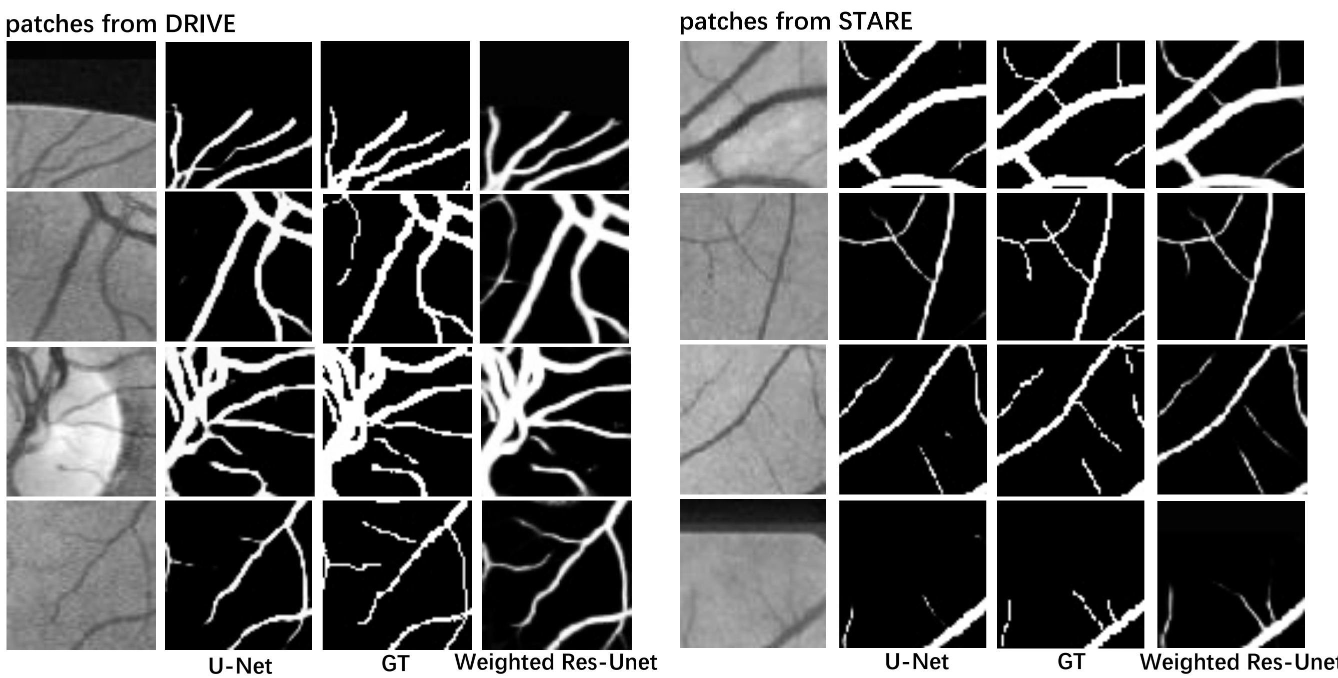

Figure 4. Examples of extracted patches' segmentation result from DRIVE dataset(listed in the left) and STARE dataset(listed in the right). The results show that compared with U-Net, our model is able to find small blood vessels and maintain better vessel connectivity.

[15] A. Wu, Z. Xu, M. Gao, M. Buty, and D. J. Mollura, "Deep vessel tracking: A generalized probabilistic approach via deep learning," in *Proc. ISBI*, 2016, pp. 1363–1367.

[16] B. Zhang, S. Huang, and S. Hu, "Multi-scale neural networks for retinal blood vessels segmentation," *arXiv preprint arXiv:1804.04206*, 2018.