IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received July 18, 2019, accepted August 16, 2019, date of publication September 17, 2019, date of current version December 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2940476

S-UNet: A Bridge-Style U-Net Framework With a Saliency Mechanism for Retinal Vessel Segmentation

JINGFEI HU[1,2,3,4], HUA WANG[1,2,3,4], SHENGBO GAO[1], MINGKUN BAO[1], TAO LIU[1,2,3,5],YAXING WANG[6], AND JICONG ZHANG[1,2,3,4,5]

[1]School of Biological Science and Medical Engineering, Beihang University, Beijing 100083, China
[2]Hefei Innovation Research Institute, Beihang University, Hefei 230013, China
[3]Beijing Advanced Innovation Centre for Biomedical Engineering, Beihang University, Beijing 102402, China
[4]School of Biomedical Engineering, Anhui Medical University, Hefei 230032, China
[5]Beijing Advanced Innovation Centre for Big Data-Based Precision Medicine, Beihang University, Beijing 100191, China
[6]Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing 100730, China

Corresponding authors: Tao Liu (tao.liu@buaa.edu.cn) and Jicong Zhang (jicongzhang@buaa.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFF0201002, in part by the National Natural Science Foundation of China under Grant 61301005 and Grant 61572055, in part by the Hefei Innovation Research Institute, Beihang University, and in part by the Thousand Young Talent Plan Station between Jicong Zhang and Jiangsu Yuwell Medical Equipment and Supply Company, Ltd.

ABSTRACT Deep learning methods have been successfully applied in medical image classification, segmentation and detection tasks. The U-Net architecture has been widely applied for these tasks. In this paper, we propose a U-Net variant for improved vessel segmentation in retinal fundus images. Firstly, we design a minimal U-Net (Mi-UNet) architecture, which drastically reduces the parameter count to 0.07M compared to 31.03M for the conventional U-Net. Moreover, based on Mi-UNet, we propose Salient U-Net (S-UNet), a bridge-style U-Net architecture with a saliency mechanism and with only 0.21M parameters. S-UNet uses a cascading technique that employs the foreground features of one net block as the foreground attention information of the next net block. This cascading leads to enhanced input images, inheritance of the learning experience of previous net blocks, and hence effective solution of the data imbalance problem. S-UNet was tested on two benchmark datasets, DRIVE and CHASE_DB1, with image sizes of 584 × 565 and 960 × 999, respectively. S-UNet was tested on the TONGREN clinical dataset with image sizes of 1880 × 2816. The experimental results show superior performance in comparison to other state-of-the-art methods. Especially, for whole-image input from the DRIVE dataset, S-UNet achieved a Matthews correlation coefficient (MCC), an area under curve (AUC), and an F1 score of 0.8055, 0.9821, and 0.8303, respectively. The corresponding scores for the CHASE_DB1 dataset were 0.8065, 0.9867, and 0.8242, respectively. Moreover, our model shows an excellent performance on the TONGREN clinical dataset. In addition, S-UNet segments images of low, medium, and high resolutions in just 33ms, 91ms and 0.49s, respectively. This shows the real-time applicability of the proposed model.

INDEX TERMS Deep learning, retinal fundus image, saliency mechanism, vessel segmentation.

I. INTRODUCTION

Early diagnosis is crucial for many diseases that lead to human vision deterioration, such as glaucoma, hypertension and diabetic retinopathy [1], [2]. Ophthalmologists typically examine retinal fundus images to assess the clinical condition of the retinal blood vessels, which is an important indicator for the diagnosis of various ophthalmic diseases. However, manual labeling of retinal vessels in these images is time-consuming, tedious and requires high clinical experience. Hence, real-time automatic segmentation of retinal blood vessels is highly needed [3], and has attracted great attention in recent decades [4].

Existing retinal vessel segmentation methods can be divided into unsupervised and supervised methods [5]. For unsupervised methods, features of given unlabeled data

The associate editor coordinating the review of this manuscript and approving it for publication was Madhu S. Nair.

samples are extracted, clustered, and used to distinguish between blood vessels and background tissues. For example, Azzopardi and Petkov [6] used a two-dimensional kernel function to fit the retinal vessel characteristics and produce a Gaussian intensity profile of the vessels. In Feng *et al.* [7], 3D directional scores were computed from retinal images, and then blood vessels were enhanced through multi-scale derivatives. Roychowdhury *et al.* [8] used fundus vascular morphology and adaptive thresholding for segmentation. A center-line detection approach for vessel segmentation was introduced by Jiang *et al.* [9]. Unsupervised methods have advantages of low sample data requirements, and low data acquisition cost. Nevertheless, features from small datasets are typically obvious individual features that do not reflect the complexity of vessel boundaries.

For supervised methods, retinal vessel segmentation is treated as a classification problem. In this problem, blood vessels and other tissues are considered to be two categories and classification is made on a pixel-by-pixel basis. For example, Strisciuglio *et al.* [10] proposed a set of COSFIRE filters, trained a support vector machine (SVM) classifier, and determined the most discriminative filter subset for vessel delineation. Orlando *et al.* [11] proposed a fully-connected conditional-random-field vessel segmentation model with structured-output SVM learning. Recently, Zhang *et al.* [12] combined vascular and wavelet features, processed 29 feature sets, and used a random-forest classifier for vessel segmentation. Compared with unsupervised methods, the results of supervised methods have high computational costs and are often strongly influenced by expert labeling and engineered features.

Deep learning achieves state-of-the-art performance in many computer vision tasks, such as image classification, image segmentation, target recognition, motion tracking and creating image subtitles [13]. In particular, the performance of deep convolutional neural networks (CNN) is close to that of the radiologists in many semantic segmentation tasks in medical image analysis. U-Net [14] is the most widely used deep learning architecture in medical image analysis, mainly because of its codec structure with jump joints which allows efficient information flow and good performance in the absence of a sufficiently large dataset. Thus, many variants of U-Net have been proposed. Alom *et al.* [15] proposed a U-Net segmentation architecture with recurrent convolutional neural network (RCNN), which is named RU-Net. Oktay *et al.* [16] proposed using an attention module with U-Net for pancreas segmentation. Jégou *et al.* [17] proposed Tiramisu, a U-Net architecture whose cascaded convolutional layers are replaced with dense blocks such that each convolutional layer is directly connected to every other layer in a feed-forward fashion. However, fundus image data is extremely unbalanced: the training dataset typically has only 20 cases, out of which the positive cases account for only 10-20% [18]. Therefore, the classic U-Net architecture cannot be applied blindly. To deal with data imbalance, earlier approaches involved extracting image patches, and randomly selecting 3,000 to 10,000 48 × 48 image patches for training [7], [19]–[22]. However, these patch-based approaches show slow convergence rates, long testing times, failure to obtain real-time results, and hence less applicability in clinical applications. Although the BTS-DSN method does not divide the fundus images into patches, the results are lower than those of a patch-based method. For small datasets, previous methods adopted diverse data augmentation methods. For instance, Bandara and Giragama [23] applied a spatially adaptive contrast enhancement technique to retinal fundus images for vessel segmentation. Oliveira *et al.* [19] used the stationary wavelet transform (SWT) to preprocess retinal fundus images. However, SWT preprocessing is complex and slow.

In this paper, we created a minimal U-Net (Mi-UNet), a simplified architecture of U-Net, with parameters whose count is only 0.23% of those of U-Net. This reduction effectively prevents overfitting on small datasets for retinal vessel segmentation. In addition, we propose Salient U-Net (S-UNet), a bridge-style architecture based on Mi-UNet. For this proposed architecture, the foreground features learned from one Mi-UNet model are taken as the foreground salient information and concatenated with the original data to be transmitted to the next Mi-UNet in a cascade manner. As shown in Table 3 and Table 4, by integrating multiple learning experiences of Mi-UNet blocks, S-UNet achieved an excellent performance in terms of the Matthews correlation coefficient (MCC), the area under curve (AUC) and the F1 score on the DRIVE and CHASE_DB1 datasets. Meanwhile, our model has also been trained and tested on actual clinical data of TONGREN, achieving an AUC of 0.9824, and a testing time of only 0.49s for a 1880 × 2816 fundus image. These results are of great significance for the practical promotion and clinical application of the proposed architecture.

The main contribution of this paper is that we propose S-UNet, a bridge-style deep learning architecture that uses a cascading approach to apply the foreground features of one Mi-UNet block as the foreground salient information of the next Mi-UNet block to enhance the input images and inherit the learning experiences of the previous Mi-UNet blocks. S-UNet uses a saliency mechanism to effectively solve the problem of data imbalance. In addition, the S-UNet parameters are only 0.7% of those of the original U-Net. This makes S-UNet one of the architectures with the fewest parameters.

The rest of the paper is organized as follows. The proposed deep learning architecture is described in Section 2. Section 3 shows the experimental setup. Results and discussion are given in Section 4, while the main conclusions are summarized in Section 5.

II. METHODS

In this section, we describe in detail the design of the bridge-style S-UNet architecture. We use a cascading approach to apply the foreground features of an earlier network block as the foreground salient information of the next network block to enhance the input images and inherit the learning

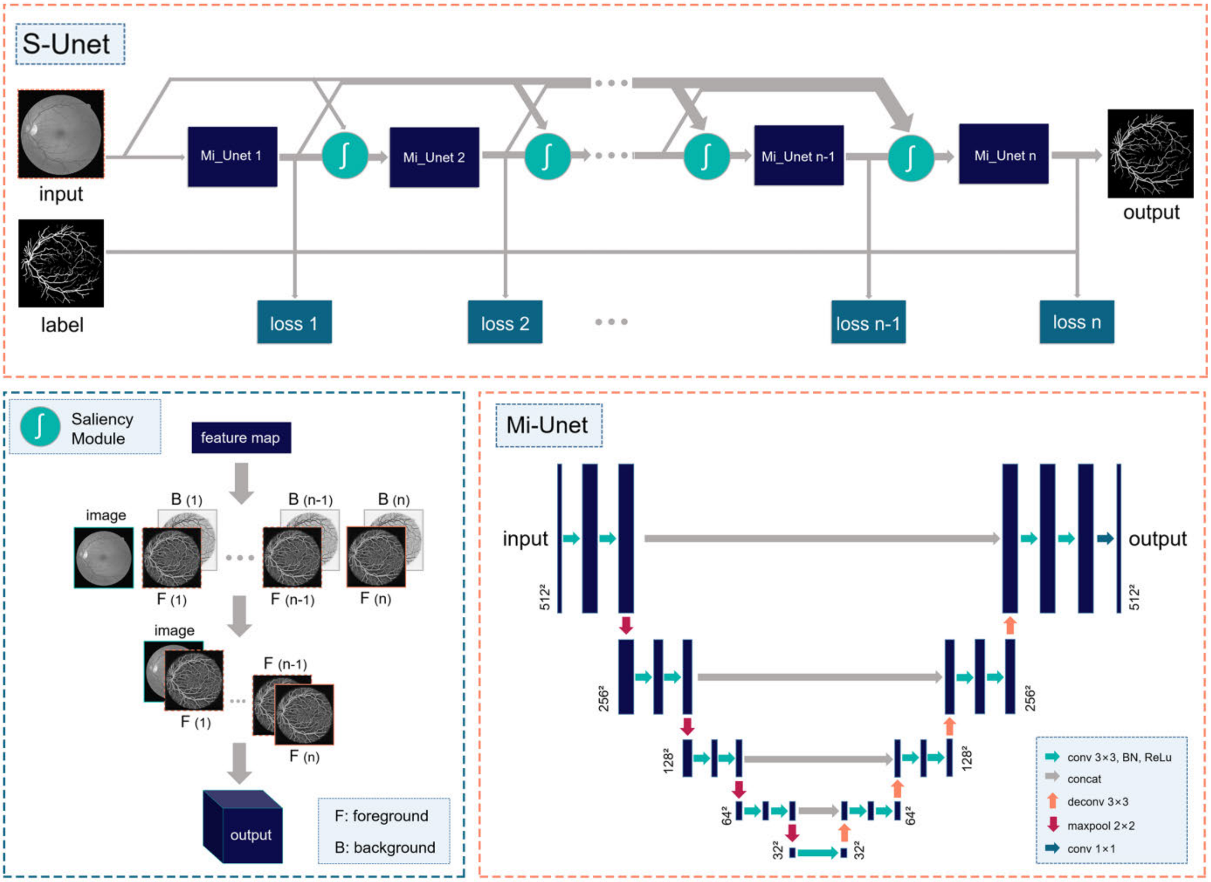


FIGURE 1. Overview of the S-UNet architecture for retinal vessel segmentation. Top: Flow diagram of the S-UNet vessel segmentation algorithm. Bottom left: The saliency module. Bottom right: The Mi-UNet block structure.

experience of the previous network blocks. We used full images and performed horizontal and vertical data augmentation. An overview of the proposed framework is shown in Fig. 1.

U-Net and its variants in the literature have encoder-decoder structures, and remarkable results have been achieved using these architectures in fundus vascular segmentation. However, for this segmentation problem, excessive downsampling can lead to loss of vascular details while too many parameters can cause overfitting. Therefore, we have simplified the classical U-Net framework to a Mi-UNet architecture which is used as the basic building block of our segmentation model. An outline of Mi-UNet is shown in Fig. 1, and the network parameters are listed in Table 1. Mi-UNet drastically reduces the parameter count to 0.07M compared to 31.03M for the baseline U-Net [14].

Let $\mathrm{S} = \{(\mathrm{X_n}, \mathrm{Y_n}), \mathrm{n} = 1, \ldots, \mathrm{N}\}$, where $\mathrm{X_n} = \{x_j(\mathrm{n}), \mathrm{i} = 1, \ldots, |\mathrm{X_n}|\}$ denotes a raw input retinal image and $\mathrm{Y_n} = \{y_j(\mathrm{n}), \mathrm{i} = 1, \ldots, |\mathrm{X_n}|\}, \mathrm{y(n)} \in \{0, 1\}$ denotes the corresponding ground-truth binary vessel segmentation map for the image $\mathrm{X_n}$. Since each image is handled separately, the subscript n is omitted for simplicity. We suppose there are N Mi-UNet blocks in the network, and we denote by $W_L(L = 1, 2, \ldots, \mathrm{N})$ the collection of all convolutional layers of the Mi-UNet blocks.

In the proposed method, we use a cascading scheme to link N Mi-UNet blocks, as shown in Fig. 1, use the foreground features of the previous Mi-UNet blocks as the foreground attention information of the next Mi-UNet block. This cascading embodies the saliency mechanism of our method (See the saliency mechanism in Fig. 1). Then, we concatenate the foreground features of the output of all of the preceding network blocks with the original input. Specifically, the output of the first block O_1 equals the product of the parameters W_1 of the first Mi-UNet block with the original input X:

$$O_1 = W_1 X \tag{1}$$

The output of the first saliency module is defined as:

$$sO_1 = (W_1 X)_f \oplus X \tag{2}$$

where $(\cdot)_f$ represents the foreground features, and $\oplus$ represents the concatenation of the foreground features and the input images, and sO_0 is X. Thus, the second Mi-UNet block gets enhanced input data as shown in Fig. 2.

TABLE 1. Mi-UNet architectural parameters for fundus vascular segmentation.

Name	Operator	Operator Repeat	Input Size	Output Features	Kernel Size	Stride Size	Concat	Params
C1	BN-ReLU-Conv	2	1×512×512	32	3×3	1×1	no	320+9248
P1	Maxpool	1	32×512×512	32	2×2	1×1	no	0
C2	BN-ReLU-Conv	2	32×256×256	20	3×3	1×1	no	5780+3620
P2	Maxpool	1	20×256×256	20	2×2	1×1	no	0
C3	BN-ReLU-Conv	2	20×128×128	12	3×3	1×1	no	2172+1308
P3	Maxpool	1	12×128×128	12	2×2	1×1	no	0
C4	BN-ReLU-Conv	2	12×64×64	12	3×3	1×1	no	1308×2
P4	Maxpool	1	12×64×64	12	2×2	1×1	no	0
C	BN-ReLU-Conv	1	12×32×32	12	3×3	1×1	no	1308
D4	Deconv	1	12×32×32	12	3×3	2×2	C4	1308
D4-C4	BN-ReLU-Conv	2	12×64×64	12	3×3	1×1	no	1308×2
D3	Deconv	1	12×64×64	12	3×3	2×2	C3	1308
D3-C3	BN-ReLU-Conv	2	12×128×128	12	3×3	1×1	no	1308×2
D2	Deconv	1	12×128×128	20	3×3	2×2	C2	2180
D2-C2	BN-ReLU-Conv	2	20×256×256	20	3×3	1×1	no	3620×2
D1	Deconv	1	20×256×256	32	3×3	2×2	C1	5792
D1-C1	BN-ReLU-Conv	2	32×512×512	32	3×3	1×1	no	9248×2
O	Conv	1	32×512×512	2	3×3	1×1	no	578
Total								68506

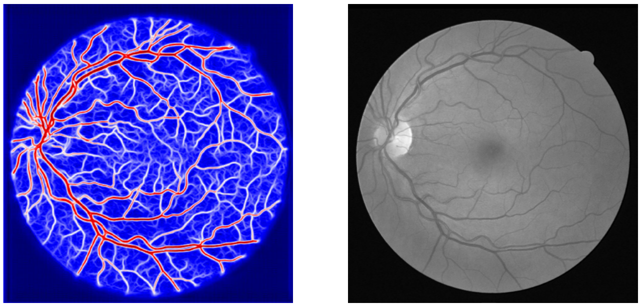


FIGURE 2. The input data of the second Mi-UNet block. a) The heat map of the output foreground features after the first Mi-UNet block (Vessels are shown in red, and the background is shown in blue). b) The input image (in gray).

Therefore, the output of the *L-1* Mi-UNet block is defined as:

$$O_{L-1} = W_{L-1} s O_{L-2}, \quad L \geq 2 \tag{3}$$

The input of the *L* Mi-UNet block is defined as:

$$I_L = sO_{L-1} = \oplus_{i=0}^{L-1} (O_{L-1})_f, \quad L \geq 2 \tag{4}$$

In this framework, in order to enforce each block to learn some new knowledge, we add an auxiliary binary cross-entropy loss function to the output of each block (See the S-UNet architecture in Fig. 1). For the output of the *L-1* Mi-UNet, the auxiliary loss function is defined as:

$$L_{BEC\,L-1} = -\frac{1}{n} \sum\nolimits_{i=1}^{n} \left(y_i log\left(y_i'\right) + (1-y_i) log(1-y_i')\right) \tag{5}$$

and the total loss is defined as:

$$Loss = \sum\nolimits_{i=1}^{N} \alpha_i * L_{BEC\,i} + \beta * \|W\|_2^2 \tag{6}$$

$$\sum\nolimits_{i=1}^{N} \alpha_i = 1 \tag{7}$$

where n denotes the pixel count in a given image, y' is the network's predicted output probability of a vessel pixel, y is the ground-truth class, and α_i is the weight of the i^{th} auxiliary loss ($\alpha_{\text{i}} = 1/\text{L}$ in our work). We use L_2 regularization with a weighting factor of $\beta = 0.0002$.

III. EXPERIMENTAL SETUP

In the following subsections, we describe the used dataset, the evaluation criteria for retinal vessel segmentation, the implementation, and the training details.

A. RETINAL IMAGE DATASETS

In this work, we evaluate our method and assess its clinical applicability on three retinal image datasets of different scales, where the first two datasets are publically available while the third one was collected by the authors. The DRIVE database [18] has forty 565 × 584 images, out of which 7 images exhibit pathological patterns. The CHASE_DB1 database [24] has twenty eight 999 × 960 images, collected from 14 children, with images from both eyes for each child. The third dataset is the TONGREN clinical database which consists of thirty 1880×2816 images, collected from 30 people at the Tongren Beijing Hospital, where five of these images show pathological patterns. Samples from the three datasets are shown in Fig. 3.

For the DRIVE and TONGREN datasets, the images were divided equally among training and testing sets. Specifically, for the DRIVE dataset, both the training and testing sets are 20 images, while for the TONGREN dataset, both the training and testing sets are 15 images. Besides, the TONGREN dataset was stratified to ensure the balance of healthy and pathological cases.

For the CHASE_DB1 dataset, no clear distinction could be made between healthy and pathological cases. Hence, a stratified *k*-fold cross-validation scheme was adopted. In this

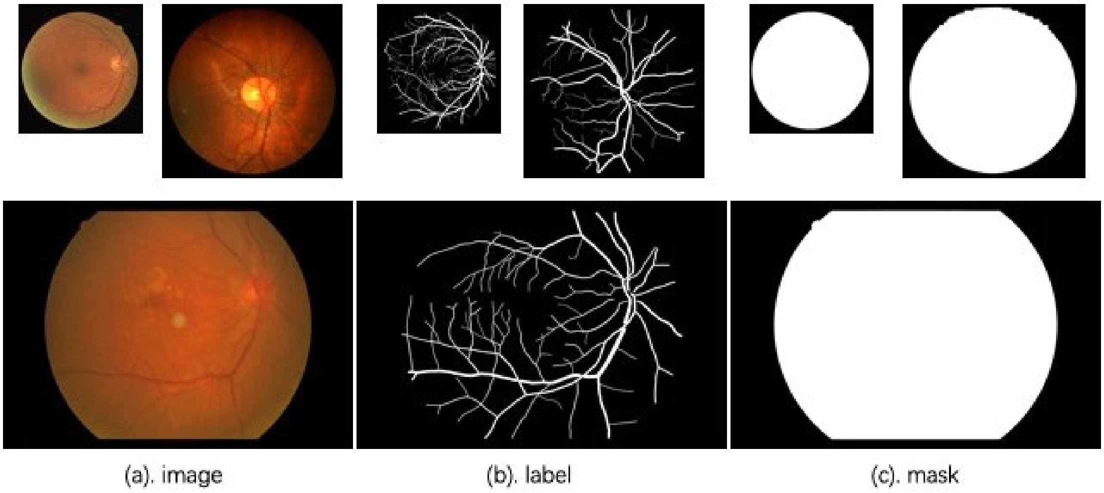


FIGURE 3. Examples of the retinal images and their pixel dimensions (H × W). (a) Input retinal images: Top Left: DRIVE (584 × 565); top right: CHASE_DB1 (960 × 999); bottom: TONGREN (1880 × 2816). (b) Vessel manual labels. (c) Fundus image masks.

scheme, the original dataset is partitioned into k folds of equal sizes, where one fold is used for testing, and the other *k-1* folds are used for training. This process is repeated k times, and the k results are averaged to produce a single performance metric estimate. For evaluation consistency, the same architecture settings and training from scratch were used for each repetition of the k-fold cross-validation. For our experiments on the CHASE_DB1 dataset, we used $k = 4$ folds, where each fold has 7 images: 3 images for one eye side and 4 images for the other.

For the DRIVE and CHASE_DB1 datasets, manual segmentation masks by two independent human observers are available. Annotations made by the first human observer were used as the ground truth for the DRIVE dataset while Hoover's annotations were used as the ground truth for the CHASE_DB1 dataset. The binary segmentation masks for the DRIVE images are publicly available. For the other datasets, we have manually created the field-of-view (FOV) masks by applying techniques similar to those in [25].

B. SEGMENTATION EVALUTION METRICS

Several metrics were used to compare the performance of the proposed method against other reference methods: sensitivity (SE), specificity (SP), accuracy (ACC), Matthews correlation coefficient (MCC), F1 score (F1), and the area under the ROC curve (AUC). The values of all of these metrics are 1 for a perfect classifier. The binary segmentation outputs were found by applying thresholding to probability maps with a threshold of 0.5. Computations were made only on the pixels within the field of view (FOV).

C. TRAINING SETUP

Since the sizes of the training sets are quite small and insufficient to deal with the model complexity, several strategies for data augmentation may be explored [7], [19]–[22], [26]–[28]. Those include image rotation by different angels, and image scaling with different factors. For example, Oliveira *et al.* [19], proposed a vessel segmentation method based on SWT [29] and extracted 48 × 48 image patches as the input data. Since we have no prior knowledge on suitable patch sizes for our method, we use full retinal images and only horizontal and vertical data augmentation. We use the gray-level images instead of the color RGB retinal images for avoiding the impact of the individual differences.

The proposed method was implemented using a TensorFlow (https://github.com/tensorflow/tensorflow) backend [30], cuDNN 9.0, an Intel(R) Xeon(R) Gold 6148 CPU with a 2.40-GHz processor, 256 GB of RAM, and an Ubuntu 16.04 operating system.

IV. RESULTS

In this section, we validate the key components of the proposed model, and compare its performance with other state-of-the-art methods on three datasets including a clinical-scale dataset of fundus images.

A. MI-UNET ASSESSMENT ON THE DRIVE DATASET

Firstly, we compare the segmentation performance of the Mi-UNet architecture with the conventional U-Net [14]. We could observe from Table 2 that the vessel segmentation results of Mi-UNet are much better than those of U-Net. In particular, the SE and F1 metrics are higher by 5.62% and 0.89%, respectively. Also, the parameter count of Mi-UNet is only 0.2% of that of U-Net. Moreover, while U-Net operates on image patches, Mi-UNet takes whole images as input. These observations validate the effectiveness and accuracy of the Mi-UNet segmentation results. A comparison of the performance and the computations of Mi-UNet and U-Net is shown in Fig. 6.

TABLE 2. Retinal vessel segmentation performance of Mi-UNet versus U-Net on the DRIVE dataset (The best results are shown in bold).

Method	MCC	SE	SP	ACC	AUC	F1	Parameters (M)	Patch/Image-based
U-Net [15]	N.A	0.7537	0.9820	0.9531	0.9755	0.8142	35	Patch-based
Mi-UNet	**0.7989**	**0.8099**	0.9772	**0.9559**	**0.9799**	**0.8231**	**0.07**	**Image-based**

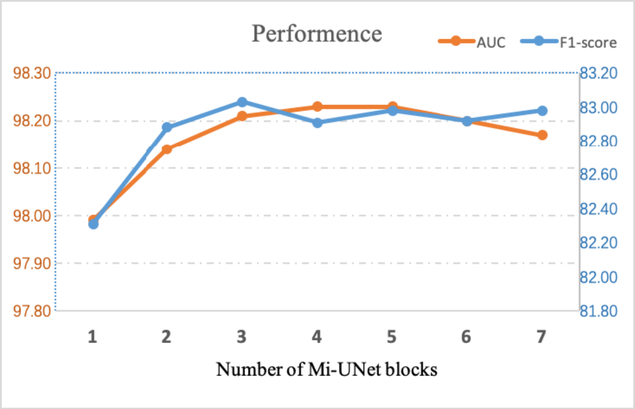


FIGURE 4. The performance of the S-UNet model with different numbers of Mi-UNet blocks.

B. S-UNET PERFORMANCE ON THE DRIVE DATASET

We assess the effectiveness of the proposed S-UNet model by performing several experiments on different numbers of Mi-UNet architectures. A comparison of the statistical measures of seven such architectures are shown in Fig. 4. When we just add one Mi-UNet block, the AUC and F1 scores are higher by 0.57% and 0.15%, respectively, compared to the case of one Mi-UNet block. For architectures with three Mi-UNet blocks, the AUC and F1 scores are significantly improved to 98.21% and 83.03%, respectively. For four or more Mi-UNet blocks, the AUC measure reaches 98.23%, the F1 score oscillates around 82.95%, and the computations increase substantially. For the cascading scheme, we let the foreground features of a previous Mi-UNet block be used as the foreground attention information of the next Mi-UNet block to enhance the input images and inherit the learning experiences of the previous blocks. Based on the performance and computations, we chose an S-UNet model with three Mi-UNet blocks as our final network.

With three Mi-UNet blocks of our S-UNet model, we visualized the segmentation results of each Mi-UNet block. As shown in Fig. 5 (b), the segmentation results of the first Mi-UNet block are relatively confusing, especially with many outliers appearing in the microvascular area. After the first application of the saliency mechanism, the vascular characteristics of the first Mi-UNet block are used as the salient information for the second Mi-UNet block. This transfer of salient information enhances the input information of the second Mi-UNet block based on the learning experience of the first Mi-UNet block. As shown in Fig. 5 (c), the results were significantly improved in the microvascular area. After intensive computations, our S-UNet model, which is equipped with three Mi-UNet blocks, gives the segmentation results shown in Fig. 5 (d), which are the closest in details to the ground-truth segmentation of Fig. 5 (e).

Earlier patch-based approaches have demonstrated reasonable performance [19], [22]. However, we just use whole retinal images as input to Mi-UNet. We quantitatively compare our method with other recent methods in Table 3. For the DRIVE dataset, S-UNet outperformed all the other methods in terms of the F1, AUC and MCC measures for both of the patch-based and image-based models. As for the AUC measure, when the number of Mi-UNet blocks is 3, our S-UNet model shows similar performance to the patch-based method of Oliveira *et al.* [19], but the latter uses various data augmentation approaches. When the number of the Mi-UNet blocks is 4 or 5, S-UNet gives the state-of-the-art performance both on the patch and image levels. In terms of the SE and ACC metrics, our S-UNet ranked as the second in terms of performance among the supervised methods. Although retinal vessel segmentation is difficult due to class imbalance, where only 10% of a retinal image corresponds to vessel pixels, we get a high SE with S-UNet for both of the patch and image levels. For the image-level variant, S-UNet improved the ACC measure by 4.21% and achieved state-of-the-art performance. The segmentation outcome of the S-UNet model on the image level is better than that of the earlier patch-level models. This reflects that our method with deeper learning can alleviate the problem of data imbalance to some extent. Though the BTS-DSN method does not also divide the fundus images into patches, S-UNet gets surpassed in all indicators apart from SP. In addition, it takes about 3 hours to train our model and only about 33ms to form the image-level vessel segmentation. Indeed, the results demonstrate that the proposed S-UNet represents an excellent approach for retinal vessel segmentation.

We also compare the performance and the computations of S-UNet with other methods, as shown in Fig. 6. In fact, our proposed S-UNet model gives the best performance with the minimum number of parameters. This is especially important for clinical application deployment.

Some S-UNet segmentation outcomes are visualized in Fig. 7. Fig. 7 (d) shows that the segmentation output includes some false-positive and false-negative results. The false-positive results are mainly reflected in the bifurcation and terminal parts of the learnt vessels, while the false-negative results are mainly reflected in the solid disk areas and the edges of the vascular diameter. In Fig. 7 (h-k),

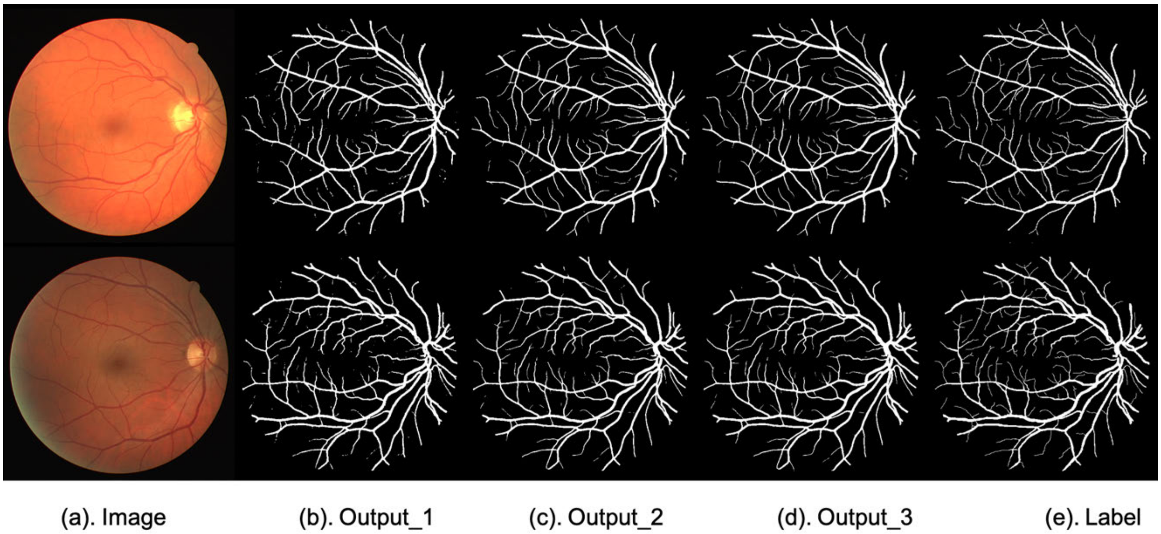


FIGURE 5. The S-UNet output: a) Raw images; b-d) The outputs of the three Mi-UNet blocks of S-UNet; e) The ground-truth retinal vessel segmentation.

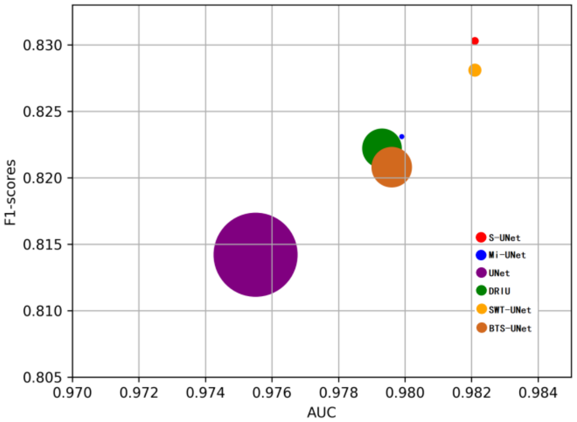


FIGURE 6. Comparison of the performance and the computations with other methods. The area of the circle represents the total of parameters. UNet is the model by Alom *et al.* [15], DRIU was proposed by Maninis *et al.* [28], SWT-UNet is the model by Oliveira *et al.* [19], and BTS-UNet is the model by Guo *et al.* [22].

the arrows point to two blood vessels that come from the bifurcation of one main blood vessel. As the retinal image is of low resolution, the small vessels are of a width of one or two pixels. The manual labeling did not mark these small vessels, but our model did capture them. In addition, as shown in Fig. 7 (e-g), there are no blood vessels in the direction indicated by the arrow, but the manual labeling depicted a straight line to the main blood vessels. The results show that our S-UNet model did not predict this area as a blood vessel. This shows the high sensitivity of S-UNet, and agrees with the quantitative results of Table 3. The quantitative performance of our model might be shown to be even better if the ground-truth annotations are more accurate.

C. S-UNET SEGMENTATION RESULTS ON THE CHASE_DB1 DATASET

For the CHASE_DB1 database, our S-UNet model gave the top results for all performance metrics with image-based network input as shown in Table 3. For the patch-based input, S-UNet gave the best performance measures except for SP, for which Oliveira *et al.* [19] took the lead. Among all methods, only our method and that of Oliveira *et al.* [19] have used cross validation for training, and reported more objective results. In additional, though the BTS-DSN method does not also divide the fundus images into patches, S-UNet gets surpassed in all indicators apart from SP. It takes about 5 hours to train our model and only about 91ms to compute the image-level vessel segmentation map.

D. S-UNET SEGMENTATION RESULTS FOR THE TONGREN CLINICAL DATASET

The DRIVE and CHASE_DB1 datasets are more than a decade old, as they were released in 2004 and 2009, respectively. The resolutions of the fundus images in these datasets is less than 1000 × 1000. However, the resolutions of clinical fundus images more recently have reached 2000 × 2000 or higher. While the earlier DRIVE and CHASE_DB1 datasets can be used to verify the effectiveness of the proposed method, higher-resolution clinical image data is needed to verify the clinical significance of the proposed method. We applied S-UNet to the TONGREN dataset, and the results are summarized in Table 3. Indeed, the proposed model achieved excellent AUC performance and needed only 0.49s to segment the vessels in a fundus image. These results are of great significance for the practical promotion and clinical application of the algorithm. It takes about 9 hours to train our model and only about 0.49s to create the vessel

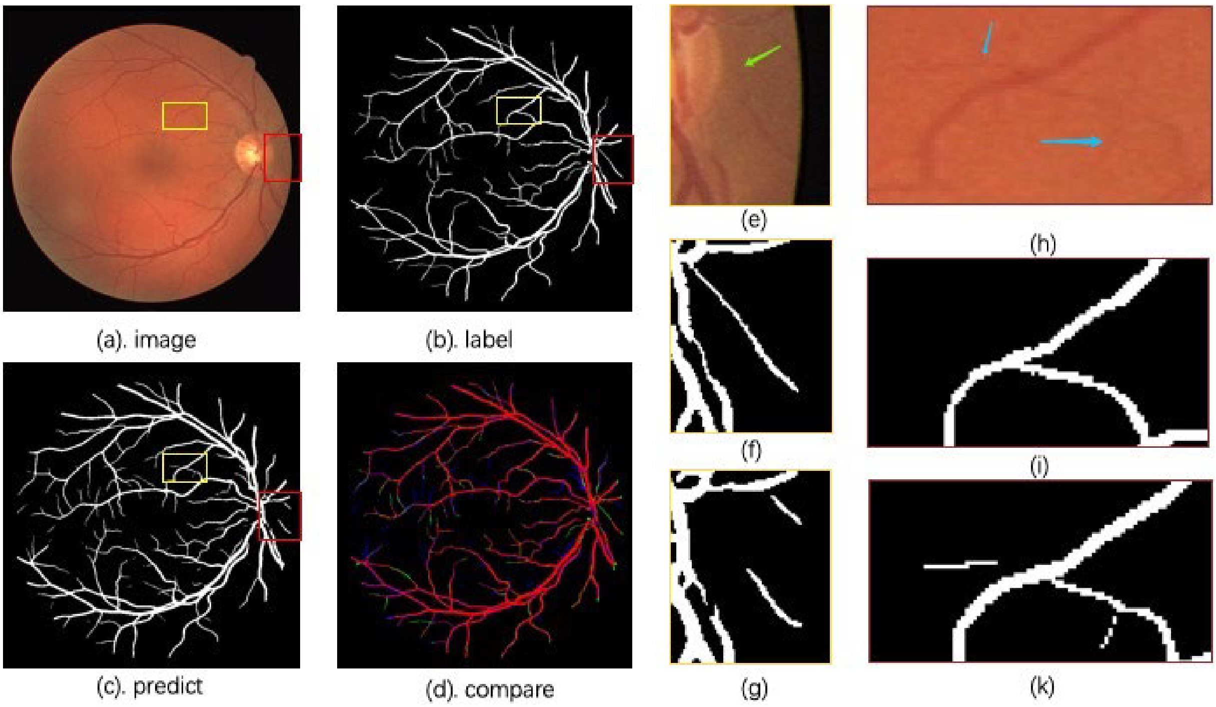


FIGURE 7. Visualization of the retinal vessel segmentation results. (a) The original image. (b) Ground-truth segmentation. (c) Segmentation with S-UNet. (d) Differences between the segmentation of (a) and (b), where red is for true positives, green is for false negatives and blue is for false positives. (e-g) One example of zoomed-in segmentation results. (h-k) Another example of zoomed-in segmentation results.

TABLE 3. Retinal vessel segmentation performance measures for different architectures on three datasets.

Dataset		Method	MCC	SE	SP	ACC	AUC	F1-scores	Patch/Image-based
DRIVE Dataset	Unsupervised	Adaptive threshold [8]	N.A	0.7395	0.9782	0.9494	0.9672	N.A	N.A
		Center-Line Detection [9]	N.A	0.8375	0.9694	0.9597	N.A	N.A	N.A
	Supervised	Skip CNN [7]	N.A	0.7743	0.9725	0.9476	0.9636	N.A	N.A
		Ensemble CNN [26]	N.A	0.7406	0.9807	0.9480	0.9747	0.7929	Patch-based
		Neural Networks [31]	N.A	0.7569	0.9816	0.9527	0.9738	N.A	Patch-based
		Deep Neural Networks [27]	N.A	0.7520	0.9806	0.9515	0.9710	N.A	Patch-based
		DRIU [28]	0.7941	0.7855	0.9799	0.9552	0.9793	0.8220	Patch-based
		Three-stage model [20]	N.A	0.7631	0.9820	0.9538	0.9750	N.A	Patch-based
		U-Net [21]	N.A	0.8723	0.9618	0.9504	0.9799	N.A	Patch-based
		RU-Net [15]	N.A	0.7792	0.9813	0.9556	0.9784	0.8171	Patch-based
		SWT-UNet [19]	0.8045	0.8039	0.9804	**0.9576**	**0.9821**	0.8281	Patch-based
		BTS-UNet [22]	0.7923	0.7800	0.9806	0.9551	0.9796	0.8208	**Image-based**
		BTS-UNet [22]	0.7964	0.7891	0.9804	0.9561	0.9806	0.8249	Patch-based
		S-UNet (ours)	**0.8055**	0.8312	0.9751	0.9567	**0.9821**	**0.8303**	**Image-based**
CHASE_DB1 Dataset	Unsupervised	Adaptive threshold [8]	N/A	0.7615	0.9575	0.9467	0.9623	N/A	N/A
	Supervised	Skip CNN [7]	N/A	0.7626	0.9661	0.9452	0.9606	N/A	N/A
		Ensemble CNN [26]	N.A	0.7224	0.9711	0.9469	0.9712	0.7566	Patch-based
		Neural networks [31]	N.A	0.7507	0.9793	0.9581	0.9716	N.A	Patch-based
		Three-stage model [20]	N.A	0.7641	0.9806	0.9607	0.9776	N.A	Patch-based
		MS-NFN[32]	N.A	0.7844	0.9819	0.9567	0.9807	N.A	Patch-based
		BTS-UNet [22]	0.7733	0.7888	0.9801	0.9627	0.9840	0.7983	**Image-based**
		RU-Net [15]	N.A	0.7756	0.9820	0.9634	0.9815	0.7928	Patch-based
		SWT-UNet [19]	0.8011	0.7779	0.9864	0.9653	0.9855	0.8188	Patch-based
		S-UNet (ours)	**0.8065**	**0.8044**	0.9841	**0.9658**	**0.9867**	**0.8242**	**Image-based**
TONGREN Dataset		**S-UNet (ours)**	0.7806	0.7822	0.9830	0.9652	0.9824	0.7994	**Image-based**

[a]N.A = Not Available, MS-NFN is the multiscale network followed network model, BTS-DSN is the multi-scale deeply supervised network with short connections.

segmentation output for the image-level input. Again, the excellent performance of our proposed method is verified.

E. ASSESSMENT OF CLINICAL DATA MODELS

Previous work on retinal vessel segmentation had trained and tested models on publicly available datasets, or, as with

TABLE 4. The performance of the S-UNet model trained under different resolutions when directly tested on datasets of other resolutions.

Training Data	Testing Data	MCC	SE	Sp	ACC	AUC	F1-score
DRIVE	DRIVE	0.8055	0.8312	0.9751	0.9567	0.9821	0.8303
(H×W,584×565)	CHASE_DB1	0.4757	0.3834	0.9814	0.9254	0.8428	0.4907
	TONGREN	0.3714	0.2431	0.9881	0.9222	0.7836	0.3562
CHASE_DB1	CHASE_DB1	0.8065	0.8044	0.9841	0.9658	0.9867	0.8242
(H×W,999×960)	DRIVE	0.6851	0.5954	0.9863	0.9365	0.9548	0.7048
	TONGREN	0.6974	0.6813	0.9798	0.9534	0.9618	0.7213
TONGREN	TONGREN	0.7806	0.7822	0.9830	0.9652	0.9824	0.7994
(H×W,1880×2816)	DRIVE	0	0	1.0	0.8727	0	0.3935
	CHASE_DB1	0	0	1.0	0.8873	0	0.3654

Oliveira *et al.* [19], cross-trained test results on datasets with similar resolutions. These approaches do not test well the robustness of the segmentation models under different resolutions and whether a model constructed with low-resolution data can be directly migrated to work on high-resolution clinical fundus data. Therefore, we tested whether the best model trained with different resolutions could get good performance directly on datasets of other resolutions. The results are shown in Table 4. We can see that larger resolution differences lead to more serious performance degradation when low-resolution models are tested on high-resolution data. For testing high-resolution models with low-resolution data, the effect is also severe because the images of the DRIVE dataset have about 60% of the resolution of the CHASE_DB1 images. If an S-UNet is trained with high-resolution TONGREN data and tested with DRIVE and CHASE_DB1 images which have only 20-30% of the training resolution, the model fails and classifies all pixels as background pixels. Therefore, publically available datasets can be used to test the performance among different methods. However, when there is a certain difference between these datasets and clinical data, it is necessary to optimize the model on actual clinical data for which segmentation algorithms will be typically and practically applied.

V. CONCLUSION

In this paper, we proposed Salient U-Net (S-UNet), a deep learning bridge-style framework that uses a cascading scheme to apply the foreground features of one Mi-UNet block as the foreground salient information of the next Mi-UNet block in order to enhance the input images and inherit the learning experiences of the previous blocks. S-UNet uses a saliency mechanism that solves the problem of data imbalance effectively. In addition, the S-UNet parameters are only 0.7% of those of the original U-Net model. So, the proposed framework is one with the fewest parameters among relevant methods in the literature. S-UNet normalizes the data, conducts horizontal and vertical data augmentation, uses full graph tests, and reaches the state-of-the-art in terms of the MCC, AUC and F1 measures on the DRIVE and CHASE_DB1 datasets. S-UNet has also been trained and tested on actual clinical data, and it achieved an AUC of 0.9824 on the TONGREN dataset. Segmenting the vessels in a 1880 × 2816 test fundus image takes only 0.49s. This real-time good performance is of great significance for the practical promotion and clinical application of the proposed algorithm. In spite of the good performance of the proposed methods, some aspects are still open for improvement. We have shown in this work that S-UNet uses a saliency mechanism to solve the problem of data imbalance effectively with the fewest parameters. So, for future research, we will optimize this saliency mechanism, test our model on a larger clinical dataset, and explore the validity of our method on other datasets.

REFERENCES

[1] P. Karpecki and B. Bowling, "Kanski's clinical ophthalmology: A systematic approach," in *Optometry and Vision Science*, 8th ed, vol. 92, no. 10. Philadelphia, PA, USA: LWW, Oct. 2015, Art. no. e386.

[2] P. Furtado, C. Travassos, R. Monteiro, S. Oliveira, C. Baptista, and F. Carrilho, "Segmentation of Eye Fundus Images by density clustering in diabetic retinopathy," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Orlando, FL, USA, Feb. 2017, pp. 25–28.

[3] C. Kirbas and F. Quek, "A review of vessel extraction techniques and algorithms," *ACM Comput. Surv.*, vol. 36, no. 2, pp. 81–121, Jun. 2014.

[4] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "Blood vessel segmentation methodologies in retinal images—A survey," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 407–433, Oct. 2012.

[5] J. Mo and L. Zhang, "Multi-level deep supervised networks for retinal vessel segmentation," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 12, pp. 2181–2193, Dec. 2017.

[6] G. Azzopardi and N. Petkov, "Automatic detection of vascular bifurcations in segmented retinal images using trainable COSFIRE filters," *Pattern Recognit. Lett.*, vol. 34, no. 8, pp. 922–933, 2013.

[7] Z. Feng, J. Yang, and L. Yao, "Patch-based fully convolutional neural network with skip connections for retinal blood vessel segmentation," in *Proc. IEEE Int. Conf. on. Image Process. (ICIP)*, Beijing, China, Sep. 2017, pp. 1742–1746.

[8] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "Iterative vessel segmentation of fundus images," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1738–1749, Jul. 2015.

[9] Z. Jiang, J. Yepez, S. An, and S. Ko, "Fast, accurate and robust retinal vessel segmentation system," *Biocybern. Biomed. Eng.*, vol. 37, no. 3, pp. 412–421, Jan. 2017.

[10] N. Strisciuglio, G. Azzopardi, and M. Vento, "Supervised vessel delineation in retinal fundus images with the automatic selection of B-COSFIRE filters," *Mach. Vis. Appl.*, vol. 27, no. 8, pp. 1137–1149, Apr. 2016.

[11] J. I. Orlando, E. Prokofyeva, and M. B. Blaschko, "A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 16–27, Jan. 2017.

[12] J. Zhang, Y. Chen, E. Bekkers, M. Wang, B. Dashtbozorg, and B. M. ter Haar Romeny, "Retinal vessel delineation using a brain-inspired wavelet transform and random forest," *Pattern Recognit.*, vol. 69, pp. 107–123, Sep. 2017.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

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

[15] M. Z. Alom, M. Hasan, C. Yakopcic, M. T. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," Feb. 2018, *arXiv:1802.06955*. [Online]. Available: https://arxiv.org/abs/1802.06955

[16] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, Y. N. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," Apr. 2018, *arXiv:1804.03999*. [Online]. Available: https://arxiv.org/abs/1804.03999

[17] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn. Workshops*, Jul. 2017, pp. 11–19.

[18] J. Staal, M. D. Abramoff, 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, Apr. 2004.

[19] A. Oliveira, S. Pereira, and C. A. Silva, "Retinal vessel segmentation based on fully convolutional neural networks," *Expert Syst. Appl.*, vol. 112, pp. 229–242, Dec. 2018.

[20] Z. Yan, X. Yang, and K.-T. Cheng, "A three-stage deep learning model for accurate retinal vessel segmentation," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 4, pp. 1427–1436, Jul. 2019.

[21] Y. Zhang and A. C. S. Chung, "Deep supervision with additional labels for retinal vessel segmentation task," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, Granada, Spain, Sep. 2018, pp. 83–91.

[22] S. Guo, K. Wang, H. Kang, Y. Zhang, Y. Gao, and T. Li, "BTS-DSN: Deeply supervised neural network with short connections for retinal vessel segmentation," *Int. J. Med. Inform.*, vol. 126, pp. 105–113, Jun. 2019.

[23] A. M. R. R. Bandara and P. W. G. R. M. P. B. Giragama , "A retinal image enhancement technique for blood vessel segmentation algorithm," in *Proc. IEEE Int. Conf. Ind. Inf. Syst. (ICIIS)*, Peradeniya, Sri Lanka, Dec. 2017, pp. 1–5.

[24] C. G. Owen, A. R. Rudnicka, R. Mullen, S. A. Barman, D. Monekosso, P. H. Whincup, J. Ng, and C. Paterson, "Measuring retinal vessel tortuosity in 10-year-old children: Validation of the computer-assisted image analysis of the retina (CAIAR) program," *Investigative Ophthalmol. Vis. Sci.*, vol. 50, no. 5, pp. 2004–2010, 2009.

[25] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1214–1222, Sep. 2006.

[26] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2538–2548, Sep. 2012.

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

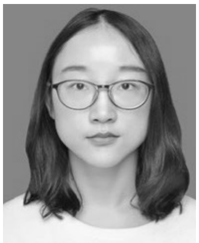
[28] K. K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, "Deep retinal image understanding," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Athens, Greece, Oct. 2016, pp. 140–148.

[29] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," *Wavelets*, J. M. Combes, A. Grossmann, P. Tchamitchian, Eds. Berlin, Germany: Springer, 1990, pp. 286–297.

[30] Tensorflow. 2019. *An Open Source Machine Learning Framework for Everyone*. [Online]. Available: https://www.github.com/tensorflow/tensorflow

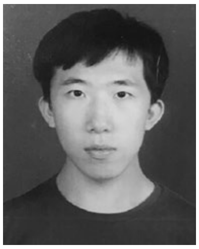
[31] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A cross-modality learning approach for vessel segmentation in retinal images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 109–118, Jan. 2016.

[32] Y. Wu, Y. Xia, Y. Song, Y. Zhang, and W. Cai, "Multiscale network followed network model for retinal vessel segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interven.*, Granada, Spain, Sep. 2018, pp. 119–126.

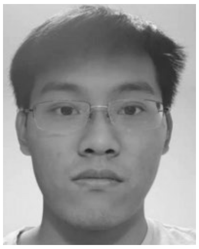


JINGFEI HU was born in Luoyang, Henan, China, in 1993. She received the M.S. degree in biomedical engineering from Beihang University, in 2019, where she is currently pursuing the Ph.D. degree with the Department of Biological Science and Medical Engineering. Her research interests include AI application in medical imaging and signal processing.

HUA WANG was born in Zhongwei, Ningxia, China, in 1993. He received the B.S. and M.S. degrees in biomedical engineering from Beihang University, in 2016 and 2019, respectively, where He is currently pursuing the Ph.D. degree. He has been engaged in medical imaging and computer vision research fields, since 2012.



SHENGBO GAO was born in Fushun, Liaoning, China, in 1996. He received the B.S. degree in biomedical engineering from the Nanjing University of Aeronautics and Astronautics, in 2019. He is currently pursuing the M.S. degree with the Department of Biological Science and Medical Engineering, Beihang University. His research interest is medical imaging.



MINGKUN BAO received the B.S. degree in biomedical engineering from Beihang University, in 2019, where he is currently pursuing the M.S. degree. His research interests include medical imaging, computer vision, and other related research fields.

TAO LIU received the Ph.D. degree from the Faculty of Engineering and Information Technology, University of Technology Sydney, Australia. He is currently an Associate Professor with the School of Biological Science and Medical Engineering, Beihang University. His research interests include neuroimaging, neuro-radiology, neurosciences, brain ageing, pattern recognition, and data mining.

YAXING WANG is currently an Ophthalmologist with the Beijing Tongren Eye Center. She is good at treating common diseases, frequently-occurring diseases, and difficult diseases in ophthalmology. She has extensive experience in the diagnosis and treatment of common diseases in ophthalmology. Her research interest includes the use of artificial intelligence to assist diagnosis of fundus-related diseases.

JICONG ZHANG received the B.S. and M.S. degrees from the Department of Electronic Engineering, Tsinghua University, in 2003 and 2006, respectively, and the Ph.D. degree in applied optimization from the University of Florida. From 2010 to 2011, he was a Research Scientist with the Research and Development Department, Cyberonics Inc. (NASDAQ: CYBX), USA. From 2011 to 2012, he was a Postdoctoral Research Fellow with the Department of Neurology and Neurosurgery, Johns Hopkins University, USA. He has been a Professor with the School of Biological Science and Medical Engineering, Beihang University, since 2014. His research interests include cognitive neuroscience, brain connectivity networks, seizure detection and prediction, electrophysiology and wearable medical devices, physiological and behavior information fusion, pattern recognition, and data mining.