Deep Guidance Network for Biomedical Image Segmentation

PENGSHUAI YIN[ID]1, **RUI YUAN**[ID]2, **YIMING CHENG**3, **AND QINGYAO WU**[ID]1, **(Member, IEEE)**
1School of Software Engineering, South China University of Technology, Guangzhou 510006, China
2School of Journalism and Communication, The Chinese University of Hong Kong, Hong Kong 999077, China
3Business School, Missouri State University, Springfield, MO 65897, USA

Corresponding author: Pengshuai Yin (pshuai.yin@gmail.com)

ABSTRACT Segmentation of 2D images is a fundamental problem for biomedical image analysis. The most widely used architecture for biomedical image segmentation is U-Net. U-Net introduces skip-connections to restore the spatial information loss caused by down-sampling operations. However, for some tasks such as the retinal vessel segmentation, the loss information of structure can not be fully recovered since the vessels is merely a curve line that can not be detected after several convolutions. In this paper, we introduce a deep guidance network to segment the biomedical image. Our proposed network consists of a guided image filter module to restore the structure information through the guidance image. Our method enables end to end training and fast inference (43ms for one image). We conduct extensive experiments for the task of vessel segmentation and optic disc and cup segmentation. The experiments on four publicly available datasets: ORIGA, REFUGE, DRIVE, and CHASEDB1 verify the effectiveness of our method.

INDEX TERMS Biomedical image segmentation, semantic segmentation, guided filter.

I. INTRODUCTION

Deep neural networks especially convolutional neural networks (CNNs) outperform the state-of-the-art in many visual recognition tasks. Such as image classification [1], image segmentation [2], object detection [3]. The success of CNNs is mainly due to the higher levels of abstraction [4]. The mid-level and high-level abstractions extracted from the image are quite effective. For biomedical image segmentation, the most popular architecture is U-Net [5]. U-Net has shown great performance for many biomedical image segmentation tasks such as optic disc and cup segmentation [6], vessel segmentation [7], tumor segmentation [8]. U-Net introduces skip-connections to fuse multi-level features and these skip connections have been shown to help recover the full spatial resolution at the network output which is essential for biomedical image segmentation [9].

In this paper, we focus on the segmentation of optic disc, optic cup and vessels in fundus image as shown in Figure.1. A fundus image mainly consists of retinal blood vessels,

The associate editor coordinating the review of this manuscript and approving it for publication was Cristian A. Linte.
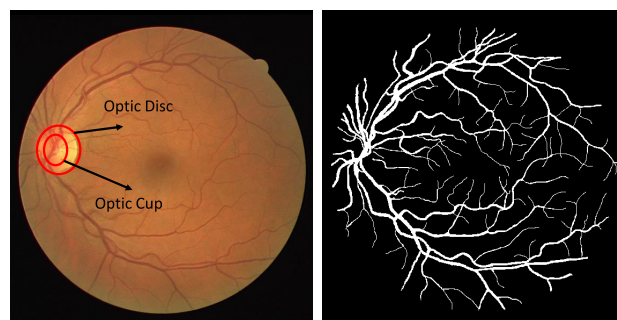

FIGURE 1. A fundus image and corresponding retinal vessel segmentation ground truth. The fundus image consists of an optic disc and an optic cup for each person.

an optic disc (OD) and an optic cup (OC). The retinal vessels in fundus images help doctors to diagnose many retinal diseases such as diabetic retinopathy, hypertensive retinopathy, solar retinopathy, retinal vein or artery occlusion [10]. The vessel widths range from one pixel to twenty pixels depending on both the width of the vessel and the image resolution. The shape of OD and OC can assist doctors to diagnose glaucoma, which is one of the leading causes of blindness [11].

The early detection and treatment for glaucoma often protect the eyes against serious vision loss. Clinically, the vertical cup to disc ratio (CDR) is a popular optic nerve head (ONH) assessment that is widely adopted by trained glaucoma specialists to screen glaucoma. The CDR is the comparison of the diameter of the optic cup to optic disc, which partially represents disease status. The normal CDR is 0.3 to 0.4. A larger CDR may indicate glaucoma or other diseases such as neuro-ophthalmic diseases. Previous studies showed that larger vertical (not horizontal) CDR is closely associated with the progression of glaucoma. Although not accurate enough, CDR is very useful in clinical practice and evaluation of glaucoma. Accurate optic disc and cup segmentations are essential for CDR measurement.

The biomedical images are collected by different types of sensors and they are contaminated by different types of noises. Such as additive white Gaussian noise (AWGN) [12], speckle noise [13], salt and pepper noise [14], rician noise [15], etc. The fundus images are also often corrupted by clouding [16]. The human eye is an optical system and the light received by the fundus camera is often attenuated along the path of the light. This can be serious when the lens of the human eye is affected by diseases such as cataracts. Since the cataract accounts for 33% of blindness worldwide [17] and its global prevalence in adults over 50 years of age was about 47.8% in 2002 [18], the clouding of retinal imaging is ineligible. Otherwise, retinal fundus images also suffer from both the additive and multiplicative noise [19]. The poor quality of the fundus image is harmful for the segmentation.

There are many challenges in fundus image segmentation due to the low image quality, for example, retinal fundus images suffer from the cataractous lens so that in optic cup segmentation, the boundary between the optic cup and optic disc may have a wide range of gradients and in hemorrhage detection or vessel segmentation, the intensity of the vessels may vary largely from one image to another and the vessel boundaries might also be obscured. The tiny vessels are obscure particularly in low contrast regions [19]. Furthermore, the down-sample operation of the U-Net results in the loss of the spatial information, some of these information such as small vessels can not restore through skip-connections or up-sampling [20].

Image filters such as bilateral filter [21] and guided image filter [22] play an important role in medical image processing [23]. Guided image filter is an efficient tool to denoise the positron emission tomography (PET) image [24] and fundus image [16]. The guided image filter (GIF) is an edge-preserving filter and computes the filtering output by considering the content of a guidance image. GIF can make the filtering output more structured and less smoothed than the input with the help of a guidance image. GIF is effective in denoising the fundus image and helps improve the performance for optic disc, optic cup segmentation and vessel segmentation [16].

In this paper, we propose a simple and efficient multiscale guided filter based network. We design a guided filter module to restore the spatial information loss of tiny thin vessels in the retinal image caused by the down-sampling operation and introduces more structure information into the neural network. Furthermore, the guided filter module also reduces the noise effect of the image. The extensive experiments show the effectiveness of our method.

The main contributions of our paper are listed below:

- We design a guided filter module to preserve the edge information, the guided filter module introduces more edge information to the feature map.
- The proposed network is simple and effective, it allows end to end training and fast inference.

II. RELATED WORK
A. OPTIC DISC AND CUP SEGMENTATION
Automatic OD and OC segmentation methods are under extensive research during the last decades [25], [26]. Early attempts for OC and OD segmentation in fundus images are based on hand-craft features such as color [27], contrast thresholding [28], level set approach [29], clustering based methods [30] and others [31]. However, manually designed features lack sufficiently discriminative power so that the imaging conditions and the complexity of pathological regions highly effect the performance. On the contrary, deep learning techniques learn comprehensively appearance feature from the labeled training set and rapidly become a popular methodology to analyze medical images [32]. In particular, convolutional neural networks show the compromising performance in many tasks such as segmentation, classification and soon become a popular method for OD and OC segmentation [6], [33]–[36].

B. RETINAL VESSEL SEGMENTATION
Recently, deep learning (DL) has been adopted to perform vessel segmentation and has reported encouraging results. U-Net [5] introduces an encoder-decoder architecture with skip connections, which demonstrates excellent performance for detecting the boundaries of neurons. Gu *et al.* [37] propose a context encoder network to capture more high-level information and preserve spatial information for 2D vessel segmentation. Yan *et al.* [38] train the U-Net simultaneously with a joint-loss including a pixel-wise and a segment-level loss. DEU-Net [39] introduces a feature fusion module to combine a spatial path with a large kernel to preserve the spatial information and a context path with a multiscale convolution block to capture more semantic information. DeepVessel [40] applies a multi-scale and multi-level Convolutional Neural Network (CNN) with a side-output layer to learn a rich hierarchical representation and model the long-range interactions between pixels by a Conditional Random Field. Compared with prior methods, our approach introduces a guided filter module to introduce more edge information from the guidance image.
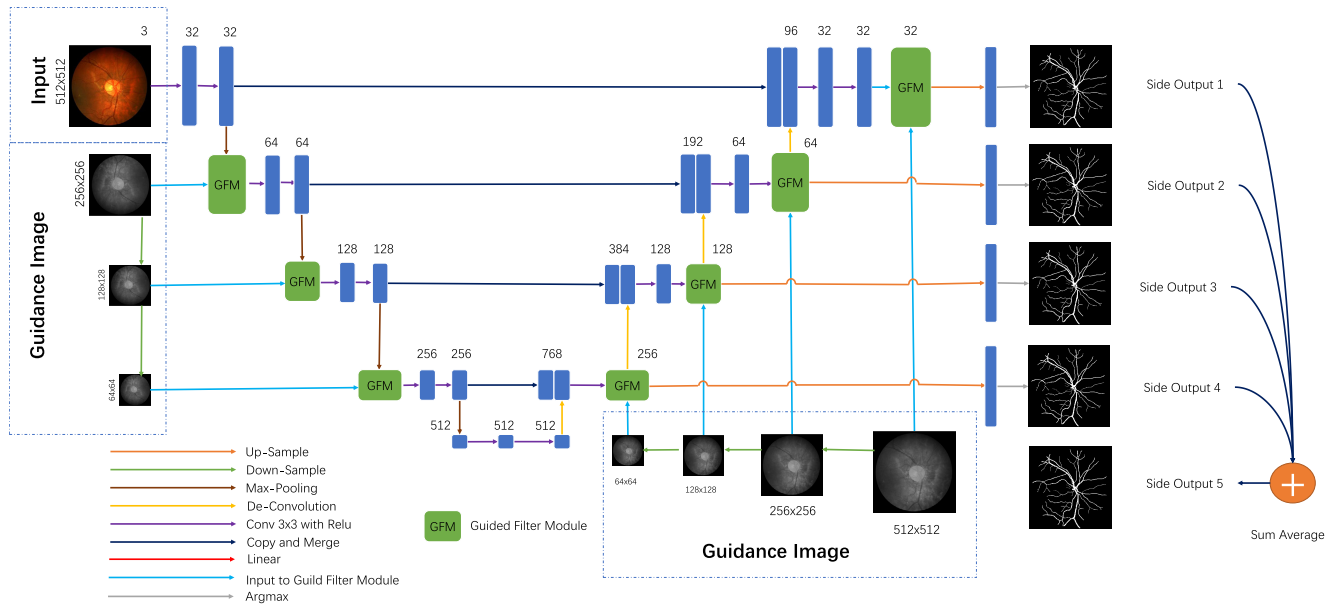
FIGURE 2. The flowchart of our proposed network. A U-shape network takes a RGB image and multiscale grey-scale guidance images as input, for each path, the output feature map is enhanced by a guided filter module. The guided filter module introduces more edge information from the guidance image into the feature map. Finally, a side-output layer predicts the vessels for each decoder path using the enhanced feature map.

III. METHOLOGY

The proposed flowchart is shown in Figure. 2. The backbone of our network is U-Net. We design a guided image filter module to preserve edge information and reduce the noise effect of the feature map. The guidance image is a grey-scale image. We also adopt a side output layer for each decoder path. Section III-A describes our flowchart and Section III-B introduces the guided filter module.

A. NETWORK ARCHITECTURE

The architecture of our network consists of a U-shape network, multiscale guided filter modules, and side-output layers. The input of the architecture is the RGB image.

1) U-SHAPE NETWORK

The original U-Net architecture is an efficient fully convolutional neural network for biomedical image segmentation. The main body of our deep architecture is similar to the U-Net that consists of the encoder path (left side) and the decoder path (right side). Each encoder path performs convolutions with the element-wise rectified linear unit (ReLU) activation function to produce encoder feature maps. The decoder path also generates a decoder feature map using the convolution layers. The skip connections transfer the corresponding feature map from the encoder path and concatenate them to up-sampled decoder feature maps.

2) MULTISCALE GUIDED FILTER MODULE

The down-sample operation of the U-Net results in the spatial information loss especially for the tiny thin vessels in the retinal image, which ultimately cannot be restored through skip-connections or up-sample operation. For optic disc and

cup segmentation, the boundary between the optic cup and optic disc is also weak. To solve this problem, we design a guided filter module to preserve the edge information from the grey-scale guidance image. The guided filter module takes the feature map and a guidance image as the input. The output is a feature map that contains the structure information from the guidance image. We add the guided filter module after every down-sample operation and up-sample operation in order to restore the spatial information loss caused by the down-sample operation. Furthermore, the guided filter module also reduces the noise effect.

3) SIDE-OUTPUT LAYERS

The side-output layer acts as a classifier that produces a companion local output map for early layers. There are M side-output layers in the network. The loss function of all the side-output layers is given as:

$$L_{side-output} = \frac{1}{M} \sum_{m=1}^{M} L_{cross-entropy}(y, y'), \quad (1)$$

$L_{cross-entropy}$ is the cross entropy loss for each side-output layer:

$$L_{cross-entropy} = -\sum_{i} (y'_i \log(y_i)), \quad (2)$$

y_i is the predicted probability value for class i and y'_i is the true probability for that class. We employ 4 side-output maps and an average layer to combine them all. The final loss function is the sum of these 5 side-output losses. The side-output layer relieves the gradient vanishing problem and helps the early layer training by backpropagating the side-output loss to the
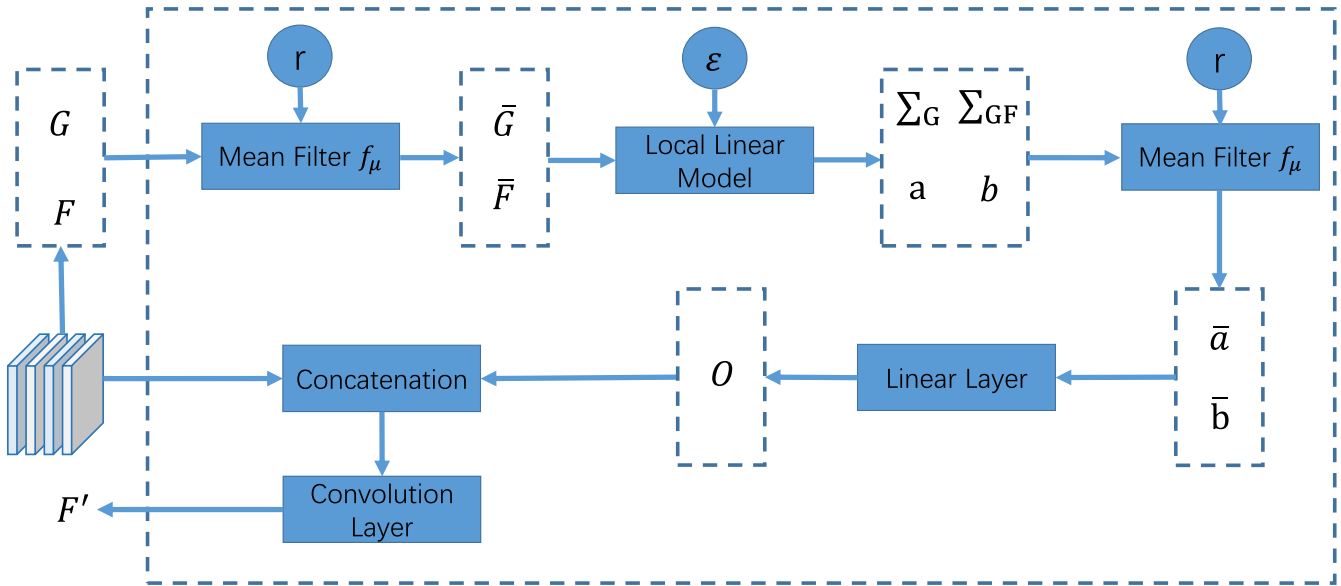
FIGURE 3. The computation graph of the guided image filter module. Guided filtering module takes feature map *F* and guidance grey-scale image *G* as inputs, generating the output *O*. The proposed layer is fully differentiable.

early layer in the decoder path. We adopt the multiscale fusion since it has been demonstrated to achieve high performance. The side-output layer also adds supervises information for each scale to output the better result. The final classifier treats the segmentation as the pixel-wise classification to produce the probability map at each pixel.

B. GUIDED FILTER MODULE

In a guided filter, the guidance image G can be identical to the input image I. The output image O is computed by a linear transform of the guidance image G in a window W_p centered at the pixel p:

$$O_i = a_p G_i + b_p, \forall i \in W_p. \quad (3)$$

Here, i is the index of pixel of G. The linear transform coefficients a_p and b_p are calculated by minimizing the objective function:

$$E = (O_i - I_i)^2 + \epsilon a_p^2. \quad (4)$$

Here, ϵ is a regularization parameter. The solution of equation (4) is:

$$a_p = \frac{\frac{1}{|W_p|} \sum_{i \in W_p} G_i I_i - \mu_p \bar{I}_p}{\sigma_p^2 + \epsilon}, \quad (5)$$

$$b_p = \bar{I}_p - a_p \mu_p. \quad (6)$$

Here, I_p represents the mean of I in W_p, $|W_p|$ denotes the cardinality of W_p. σ_p^2 and μ_p denote the variance and mean of G in W_p respectively. However, a pixel i is involved in all the overlapping windows W_p that covers i and the the output of equation (8) from different windows is not identical. The general solution is to average all the possible

values of a_p and b_p:

$$O_i = \bar{a}_{p'} G_i - \bar{b}_{p'}. \quad (7)$$

Here, $\bar{a}_{p'}$ and $\bar{b}_{p'}$ are the mean values of a_p and b_p in W_p:

$$\bar{a}_{p'} = \frac{1}{|W_p|} \sum_{p \in W_p} a_p. \quad (8)$$

$$\bar{b}_{p'} = \frac{1}{|W_p|} \sum_{p \in W_p} b_p. \quad (9)$$

With this average strategy, the abrupt intensity changes in G can be mostly preserved in O and the above averaging also reduce the noise. Motivated by this, we design a guided filter module based on the guided image filter. The guided filter module introduces more edge information from guidance image and restores the spatial information loss caused by down-sampling layers. Furthermore, the guided filter module also reduces the noise effect due to the average operation. The guided filter module can be formulated as a fully differentiable layer and we jointly training the guided filter module with backbone network to get direct supervision from side-output layers.

However, the guided filter is not optimal for retinal image processing due to it often smooth away some fine structures close to flat [16]. In our guided image filter module, the guided image filter is performed on the feature map in order to incorporate more fine structure information from the guidance image to the feature map. Specially, we concatenate the output of the guided filter module O with the origin feature map F to form a more powerful representation since the origin feature map may contain some fine structures that filtered out by the guided filter module.

Algorithm 1 Computation Graph of Guided Filtering Module

Input: Feature Map F.
Guidance Image G.
Radius r and Regularization term ϵ.
Output: Feature Map F'.

1: $\overline{F} = f_\mu(F, r)$
2: $\overline{G} = f_\mu(G, r)$
3: $\overline{G^2} = f_\mu(G.*G, r)$
4: $\overline{GF} = f_\mu(F.*G, r)$
5: $\sum_G = \overline{G^2} - \overline{G}.*\overline{G}$
6: $\sum_{GF} = \overline{GF} - \overline{G}.*\overline{F}$
7: $a = \sum_{GF} ./(\sum_G +\epsilon)$
8: $b = \overline{F} - a * \overline{G}$
9: $\overline{a} = f_\mu(a), \overline{b} = f_\mu(b)$
10: $O = \overline{a} * F + \overline{b}$
11: Concatenate F with O.
12: Perform a convolution to fuse the feature and change the channel size of the feature.
13: $F' = convolution([F|O])$
14: **return** F'

The computation graph of the guided filtering layer is shown in Figure. 3 and the details are shown in Algorithm 1. The input of the guided filter module is the feature map F and the guidance image G. For simplify, we choose the grey-scale image as the guidance image. For some biomedical image segmentation tasks, the RGB image is not available, such as computed tomography (CT) and MRI. The grey-scale image can be easily obtained. $\overline{a}$ and $\overline{b}$ are computed with mean filter f_μ with radius r. O is finally generated with a linear layer taking $\overline{a}$ and $\overline{b}$ as inputs. ϵ is the regularization term, we set r and ϵ to 2 and 1e-8 respectively. We concatenate O with F to construct a more powerful feature representation and then the feature is processed by a convolution layer to fuse the feature and change the size of channels of the feature map. The guided filter can be conveniently added into deep networks and the equations for propagating the gradients through the guided filtering module are shown in Algorithm 2. The gradient of output O could back-propagate to input F. The guided filter module restores the information loss caused by the down-sampling operation. The final feature maps are sent to side-output layers for final prediction.

IV. EXPERIMENTS

We evaluate the model for two tasks: Optic disc and cup segmentation and retinal vessel segmentation. Section IV-A describes the dataset used for experiments. Section IV-B describes the implementation details. Section IV-D introduces the evaluation criteria. Section IV-D2 evaluates the performance. Section IV-E perform ablative studies.

A. DATA PREPARATION

The experiments are performed on four public available dataset: DRIVE [41] and CHASEDB1 [42] datasets for retinal vessel segmentation. ORIGA [43] and REFUGE [44]

Algorithm 2 Gradients for Guided Filtering Module.

Input: Feature Map F.
Guidance Image G.
Radius r and Regularization Term ϵ.
Derivative for Output ∂O
Output: Gradients for the Input Feature Map ∂F.

1: $\partial b = \partial O \cdot \nabla_b f_\mu$
2: $\partial a = \partial O * F \cdot \nabla_a f_\mu - \partial b * \overline{G}$
3: $\partial \sum_{GF} = \partial a/(\sum_G +\epsilon)$
4: $\partial \sum_G = -\partial a * \sum_{GF} /(\sum_G +\epsilon)^2$
5: $\partial \overline{F} = \partial b - \partial \sum_{GF} *\overline{G}$
6: $\partial F = \partial \sum_{GF} \cdot \nabla_{G*F} f_\mu * G - \partial \overline{F} \cdot \nabla_F f_\mu$
7: **return** ∂F

---

dataset for optic disc and cup segmentation. The statistics of the dataset is shown in Table 1:

TABLE 1. Dataset statistics.

Dataset	Quantity	Resolution	Train-test split
DRIVE	40	565×584	Official train-test split
CHASEDB1	28	999×960	First 20 for train, last 8 for test
ORIGA	650	3072×2048	350 for tain, rest for test
REFUGE	800	1634×1634	Official train-test split

1) DRIVE

DRIVE consists of 40 images (7 with pathology). It is divided into 20 training images and 20 testing images along with two manual segmentations of the vessels. The first segmentations serve as ground truth while the second segmentations serve as a human observer reference for performance comparison. The images are of size 768×584 pixels and are captured in digital form from a Canon CR5 non-mydriatic 3CCD camera at 45° field of view (FOV).

2) CHASEDB1

The CHASEDB1 database [42] consists of 28 images of size 960×999 pixels. These images are captured from 14 school children by a handheld Nidek NM-200-D fundus camera at 30° field of view. The segmentation results of the first of the two observers are deployed as the ground truth. We select the first 20 images as the training set and the rest 8 images as testing set [45].

3) ORIGA

The Online Retinal Fundus Image Dataset for Glaucoma Analysis and Research (ORIGA) [43] consists of 650 images acquired through the Singapore Malay Eye Study (SiMES) and are annotated with the critical signs for glaucoma. The dataset includes 168 glaucomatous and 482 non-glaucoma images. The 650 images with manual ground truth boundaries are divided into 325 training images (including 73 glaucoma cases) and 325 testing images (including 95 glaucoma cases) as same as that in [6].

4) REFUGE

The REFUGE database [44] consists of 1200 images. The dataset includes 120 glaucomatous and 1080 non-glaucoma

images and provides pre-defined partitions into training and test.

B. IMPLEMENTATION DETAILS

In this section, we introduce the experimental settings, the implementation details and the data augmentation techniques we adopted for the paper.

1) SETTINGS

For retinal vessel segmentation, we train the network for 150 epochs, the initial learning rate is 0.001 and decays 10 times every 50 epochs. For optic disc and cup segmentation, we also train the network for 150 epochs and the initial learning rate is 0.0001 and decays 10 times every 50 epochs. The entire network is trained from scratch and do not use extra training data. We train the network on the training set and evaluated on the respective validation set for each dataset. The experiments are conducted on an NVIDIA GeForce RTX 2080 Ti GPU.

2) IMPLEMENTATION

The guided image filter has two parameters: the radius r and the regularization term ϵ. For all the experiments, we set $r = 2$ and $\epsilon = 1e - 8$. The input and output size of the network is 512×512.

3) DATA AUGMENTATION

The network takes a whole fundus image as input and the output is the same size as input. The data augmentation helps the network conquer the over-fitting issues. The augmentation techniques including random horizontally flip with probability 0.5, random rotation within the range $[0, 20°]$ and gamma contrast enhancement. We select gamma between $[0.5, 2]$. We observe that data augmentation help increase the performance.

C. OPTIC DISC AND CUP SEGMENTATION

1) EVALUATION CRITERIA

To evaluate our method for optic disc and cup segmentation, we adopt the overlapping error (E) as the evaluation metric for OD and OC segmentation:

$$E = 1 - \frac{Area(S \bigcap G)}{Area(S \bigcup G)}. \tag{10}$$

Here, S and G denote the predicted mask and corresponding ground-truth.

2) EXPERIMENT PERFORMANCE

We compare our network with several state-of-the-art models including R-Bend [46], ASM [47], LRR [48], U-Net [5], Superpixel [30], M-Net [6], Mask-RCNN [49] and PM-Net [50].

Table 2 shows the performance of OD and OC segmentation on ORIGA dataset. Our network outperforms the baseline U-Net for a large margin. Especially for the OC

TABLE 2. Performance evaluation on ORIGA dataset.

Method	E_{disc}	E_{cup}
R-Bend [46]	0.129	0.395
ASM [47]	0.148	0.313
LRR [48]	-	0.244
U-Net [5]	0.115	0.287
Superpixel [30]	0.102	0.264
M-Net [6]	0.083	0.256
Ours	**0.071**	**0.220**

segmentation. The boundary between OD and OC is weak. The guidance filter module introduces more edge information from the guidance image and improves the OC segmentation performance. We do not perform any post-processing such as a conditional random field or eclipse fitting. Figure 4 shows the segmentation result of our model compared to U-Net. Our model can accurately perform segmentation even under low image quality as shown in the last two rows of the Figure 4.

Table 3 shows the performance of OD and OC segmentation on REFUGE dataset. We train on the training dataset and evaluate the model on the validation dataset. Our method also outperforms baseline U-Net for a large margin.

TABLE 3. Performance evaluation on REFUGE validation dataset.

Method	E_{disc}	E_{cup}
U-Net [5]	0.159	0.350
Mask-RCNN [49]	0.092	0.228
PM-Net [50]	0.088	**0.223**
Ours	**0.087**	0.268

D. RETINAL VESSEL SEGMENTATION

1) EVALUATION CRITERIA

The vessel segmentation process is a pixel-based classification that any pixel is classified either as a vessel or surrounding tissue. We employ Specificity (Spe), Sensitivity (Sen), Accuracy (Acc) and Area Under ROC (AUC) as measurements. Sensitivity (Sen) reflects the ability of the algorithm to detect the vessel pixels. Specificity (Spe) is the ability to detect non-vessel pixels. The accuracy (Acc) is measured by the ratio of the total number of correctly classified pixels (sum of true positives and true negatives) to the number of pixels in the image field of view (FOV):

$$Acc = \frac{TP + TN}{TP + FN + TN + FP}. \tag{11}$$

$$Sen = \frac{TP}{TP + FN}. \tag{12}$$

$$Spe = \frac{TN}{TN + FP}. \tag{13}$$

Here, TP represents the true positive where a pixel is identified as the vessel in both the segmented image and ground truth, TN represents true negative where a non-vessel pixel of the ground truth is correctly classified in the segmented
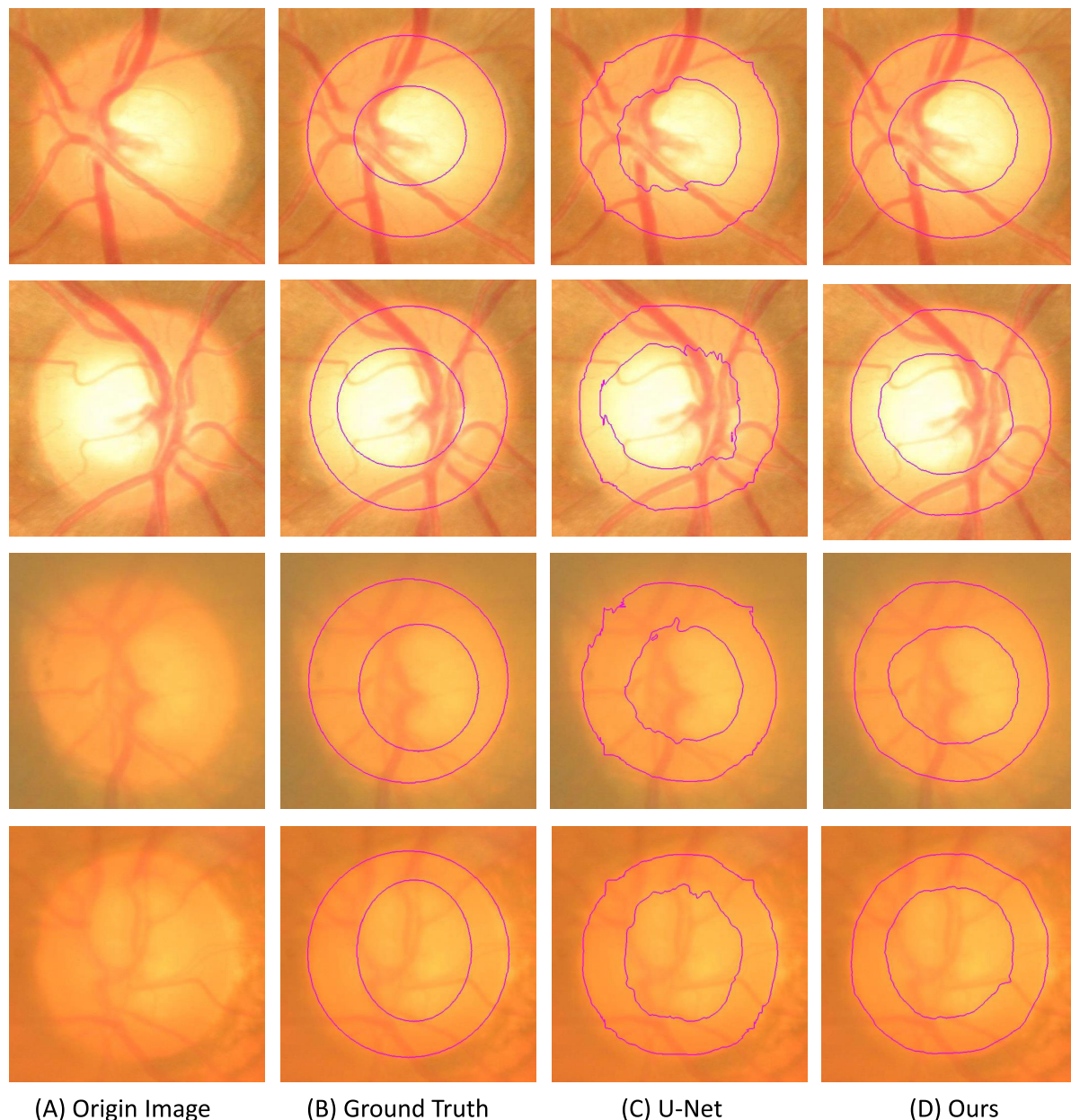
| (A) Origin Image | (B) Ground Truth | (C) U-Net | (D) Ours |

FIGURE 4. The example results of our methods for optic disc and cup segmentation, (A) is the origin image and (B) is the corresponding ground truth. (C) and (D) are the predictions using U-Net and our method respectively.

image. False negatives (FN) are the misclassifications where a vessel pixel in the ground truth image is classified as non-vessel in the segmented image and the false positives (FP) are the misclassifications where a non-vessel pixel in the ground truth image is marked as the vessel in the segmented image.

A receiver operating characteristic (ROC) curve plots the fraction of vessel pixels correctly classified as vessel versus the fraction of non-vessel pixels wrongly classified as the vessel. The closer the curve approaches the top left corner, the better is the performance of the system. The value of the area under the curve (AUC) is the most frequently used performance measure extracted from the ROC curve which is 1 for an optimal system.

2) EXPERIMENT PERFORMANCE

We compare our network with several state-of-the-art models including VesselNet [55], U-Net [5], DU-Net [39], Ladder-Net [53], Bo Liu *et al.* [54]. CE-Net [37]. Fan *et al.* [56] Our network achieves the best performance on DRIVE and CHASEDB1 dataset as shown in Table 4, and Table 5. The performance for each image on DRIVE and CHASEDB1 dataset are shown in Table 6 and Table 7, respectively. We only calculate the pixels inside the field of view for these two datasets.

For DRIVE and CHASEDB1 dataset, our network outperforms all the state-of-the-art methods and achieves the best ACC, AUC, and SPE. The experiment result shows that our method can effectively identify vessel pixels. Our method is
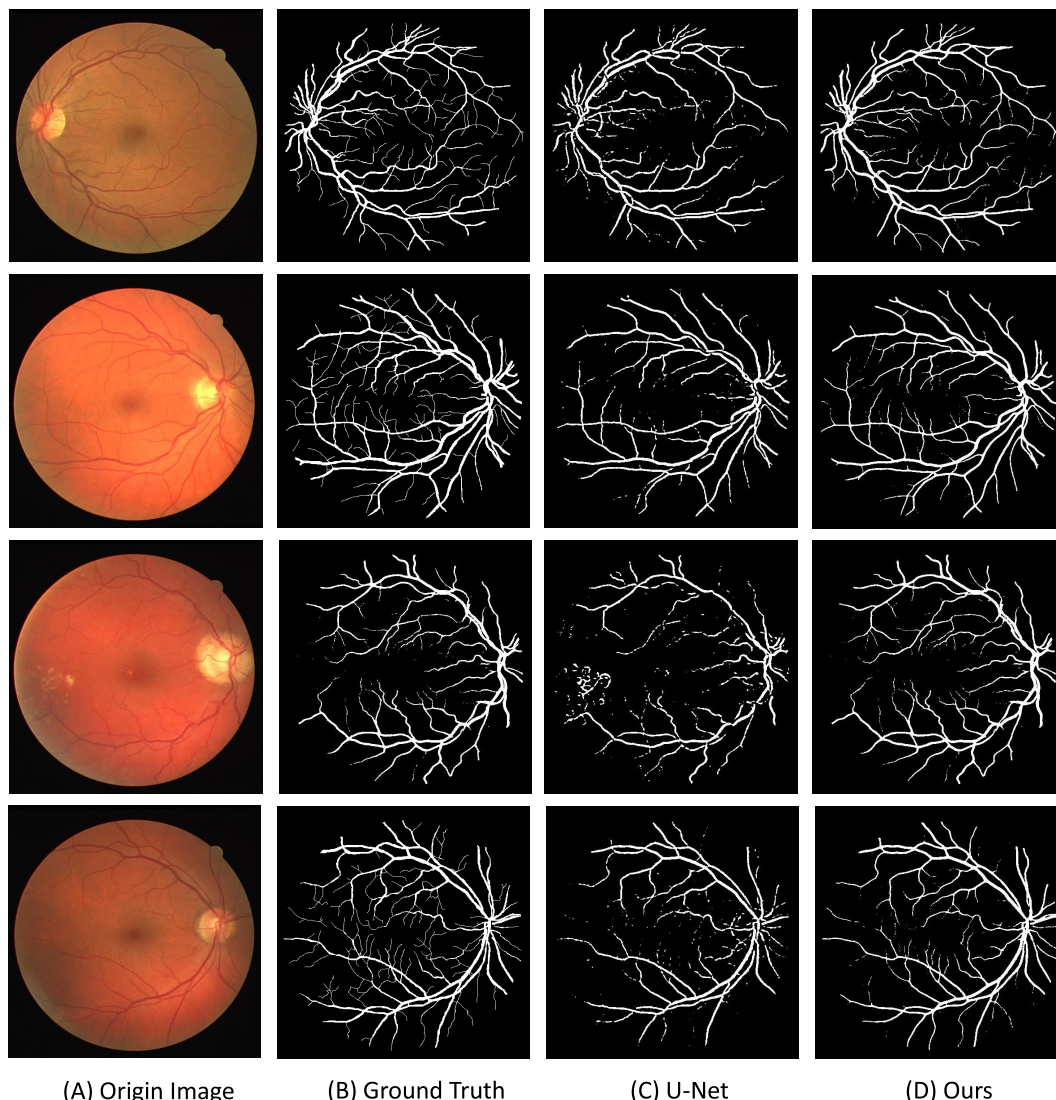
(A) Origin Image   (B) Ground Truth   (C) U-Net   (D) Ours

FIGURE 5. The example results of our methods for retinal vessel segmentation, (A) is the origin image and (B) is the corresponding ground truth. (C) and (D) are the predictions using U-Net and our method respectively.

TABLE 4. Segmentation performance for DRIVE inside FOV.

Method	Year	Acc	AUC	Sen	Spe
Human Observer	-	0.9578	N.A	0.8288	0.9701
U-Net [5]	2015	0.9531	0.9755	0.7537	0.9820
Zhang et al. [51]	2016	0.9476	0.9636	0.7743	0.9725
Liskowski et al. [52]	2016	0.9542	0.9752	0.7653	0.9818
DRIU [7]	2016	0.9541	0.9801	0.8280	0.9728
Yan et al. [38]	2018	0.9542	0.9752	0.7653	0.9818
CE-Net [37]	2019	0.9545	0.9779	**0.8309**	-
LadderNet [53]	2019	0.9561	0.9793	0.7856	0.9810
DU-Net [39]	2019	0.9567	0.9772	0.7940	0.9816
Bo Liu et al. [54]	2019	0.9559	0.9779	0.8072	0.9780
VesselNet [55]	2019	0.9578	0.9821	0.8038	0.9802
Ours	-	**0.9604**	**0.9846**	0.7614	**0.9837**

TABLE 5. Segmentation performance of CHASEDB1 inside FOV.

Method	Year	Acc	AUC	Sen	Spe
Human Observer	-	0.9545	N.A	0.8105	0.9711
U-Net [5]	2015	0.9578	0.9772	**0.8288**	0.9701
DRIU [7]	2016	0.9657	0.9746	0.7651	0.9822
Liskowski et al. [52]	2016	0.9535	0.9823	0.7816	0.9836
Yan et al. [38]	2018	0.9610	0.9781	0.7633	0.9809
LadderNet [53]	2019	0.9656	0.9839	0.7978	0.9818
DU-Net [39]	2019	0.9661	0.9812	0.8074	0.9821
VesselNet [55]	2019	0.9661	0.9860	0.8132	0.9814
Ours	-	**0.9783**	**0.9869**	0.7993	**0.9868**

built upon U-Net and the performance outperforms U-Net for a large margin. For DRIVE and CHASEDB1 dataset.

For the DRIVE dataset, ACC and AUC rank first, with a significant improvement of 0.0026 and 0.0025 compared with

the previous highest score in the work of VesselNet [55]. The SPE shows an increment of 0.0017 compared with U-Net. Table 6 shows the segmentation performance for each image. We observe that our method is robust to the pathology cases.

The CHASEDB1 dataset has nonuniform background illumination, poor contrast of blood vessels as compared with

TABLE 6. The performance for each image of DRIVE dataset inside FOV predicted by side output layer 1.

ImageID	Acc	AUC	Sen	Spe
01	0.9600	0.9883	0.8032	0.9794
02	0.9609	0.9876	0.7719	0.9889
03	0.9515	0.9808	0.7075	0.9855
04	0.9607	0.9774	0.7425	0.9916
05	0.9564	0.9783	0.7281	0.9878
06	0.9531	0.9754	0.6830	0.9896
07	0.9586	0.9802	0.6967	0.9903
08	0.9581	0.9810	0.7002	0.9876
09	0.9619	0.9821	0.7505	0.9848
10	0.9609	0.9823	0.7698	0.9823
11	0.9593	0.9796	0.7406	0.9846
12	0.9633	0.9863	0.7879	0.9829
13	0.9561	0.9795	0.7129	0.9869
14	0.9630	0.9869	0.7834	0.9820
15	0.9666	0.9868	0.8249	0.9814
16	0.9597	0.9879	0.7588	0.9858
17	0.9581	0.9825	0.7439	0.9824
18	0.9642	0.9899	0.8322	0.9781
19	0.9709	0.9917	0.8510	0.9844
20	0.9659	0.9897	0.8404	0.9776
Mean	0.9604	0.9846	0.7614	0.9837

TABLE 7. The performance for each image of CHASEDB1 dataset inside FOV predicted by side output layer 1.

ImageID	Acc	AUC	Sen	Spe
11R	0.9862	0.9907	0.8455	0.9921
11L	0.9787	0.9881	0.6398	0.9923
13L	0.9803	0.9869	0.8894	0.9850
13R	0.9766	0.9849	0.7635	0.9883
12R	0.9781	0.9875	0.7947	0.9870
14R	0.9737	0.9853	0.8264	0.9786
12L	0.9710	0.9831	0.8351	0.9792
14L	0.9819	0.9893	0.8001	0.9924
Mean	0.9802	0.9890	0.8119	0.9882

the background. Different from those of the adult retina, the morphological characteristics of choroidal vessels in retinal images taken from premature infants are more visible alongside the retinal vessels [42]. Our model achieves the highest rank in ACC and AUC increased by 0.121 and 0.0009 respectively compared to the second highest scores of VesselNet [55]. SPE is also the first place and increases by 0.0032 compared to the previous highest scores of Liskowski *et al.* [26].

The specificity values for the algorithm are the highest for all of the two image databases that indicate the low false positive rate of the methodology as compared with the other methods including the second human observer. This, in turn, indicates that the algorithm has identified less numbers of background pixels or pathological area pixels as part of a vessel than the other methods.

The best case accuracy, sensitivity, specificity, and AUC for the DRIVE database are 0.9709, 0.8510, 0.9916, and 0.9917, respectively, and the worst case measures are 0.9531, 0.6830, 0.9776, and 0.9917, respectively. The best case vessel segmentation result for the CHASEDB1 database has

TABLE 8. Ablative Studies on DRIVE inside FOV, GFM represents the guided filter module.

Method	Acc	AUC	Sen	Spe	Parameters(M)
U-Net	0.9531	0.9755	0.7537	0.9820	11.743
+Side-output	0.9575	0.9798	0.7363	0.9824	11.744
+GFM	**0.9604**	**0.9846**	**0.7614**	**0.9837**	12.162

an accuracy of 0.9862; sensitivity, specificity, and AUC are 0.8894, 0.9924, and 0.9907, respectively. The worst case accuracy is 0.9710, sensitivity and specificity are 0.6398 and 0.9786, respectively.

For the DRIVE dataset, the sensitivity of our method is 0.7614 while U-Net is 0.7537. For the CHASEDB1 dataset, the sensitivity of our method is 0.7993 while U-Net is 0.8288. For the CHASEDB1 dataset, our method achieve comparable sensitivity to other methods, such as LadderNet and DU-Net. The over-segmentation problem occurs when the network receives additional wrong edge information from the guide filter module such as the pathology area. In our model, we concatenate the feature map of the guided filter module with the origin feature map. This design eases the over-segmentation problem since the prediction is performed not only depends on the feature map of the guided filter module.

Among these methods, Yan *et al.* [38] is a multiscale method and also tackles the issue of thin versus thick vessels by employing a segment level loss function jointly with a pixel-level loss function. Our method tangle the thin versus thick vessels by adding additional edge information by multiscale guided image filter models and our method outperforms Yan *et al.* [38] on two datasets.

E. ABATIVE STUDIES

Table 8 lists the performance by adding each module to our network. Table 8 shows that the side-output layer and our guided image filter module help improve the performance. Especially, our guided image filter module improves the AUC for a large margin.

1) SIDE OUTPUT LAYERS

Different side output layer predicts vessels with different size and resolution. The side output layer increases Acc, AUC and Spe by 0.0017(p<0.05), 0.0032(p<0.05) and 0.0038(p<0.05) respectively. We also observe that our network is not affected by the pathological area.

2) GUIDED FILTER MODULE

General object edges in natural images separate two regions of different appearance, which allows the boundary to be detectable even at deeper layers. However, retinal vessels in fundus images appear merely as a curved line, which is too thin to respond in the higher stride layers. The guided filter module incorporates the edge information from guidance image to the feature map of a deep neural network, which is a benefit for vessel segmentation and optic disc and cup segmentation. Furthermore, the guided filter

module also reduces the noise effect. The guided image filter module increases Acc, AUC, Sen, Spe by 0.0029(p<0.01), 0.0058(p<0.01), 0.0251(p<0.01), 0.0013(p<0.05) respectively.

3) TIME COMSUMPTION

It only takes 44ms for our model to make a prediction with image size 512×512 using a GPU. The entire model enables end to end training and very fast inference due to the fast implementation of the guided image filter. The guided image filter module increases the model size by 0.418M

V. CONCLUSION

In this paper, we propose a multiscale network with guided filter modules for vessel segmentation and optic disc and cup segmentation. The network incorporates a guided image filter module to introduce more edge information from the guidance image to the neural network and restores the spatial information loss caused by the down-sample operation. The extensive experiments on two tasks verify the effectiveness of our method. Furthermore, our method enables end to end training and fast inference.

ACKNOWLEDGMENT

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

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

[3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[5] 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.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[6] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1597–1605, Jul. 2018.

[7] 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.* Cham, Switzerland: Springer, 2016, pp. 140–148.

[8] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 168–172.

[9] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," in *Deep Learning and Data Labeling for Medical Applications*. Cham, Switzerland: Springer, 2016, pp. 179–187.

[10] 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.

[11] H. A. Quigley, "The number of people with glaucoma worldwide in 2010 and 2020," *Brit. J. Ophthalmol.*, vol. 90, no. 3, pp. 262–267, Mar. 2006.

[12] S. Lal, M. Chandra, G. K. Upadhyay, and D. Gupta, "Removal of additive Gaussian noise by complex double density dual tree discrete wavelet transform," *MIT Int. J. Electron. Commun. Eng.*, vol. 1, no. 1, pp. 8–16, 2011.

[13] T. R. Jeyalakshmi and K. Ramar, "A modified method for speckle noise removal in ultrasound medical images," *Int. J. Comput. Elect. Eng.*, vol. 2, no. 1, p. 54, 2010.

[14] P.-Y. Chen and C.-Y. Lien, "An efficient edge-preserving algorithm for removal of salt-and-pepper noise," *IEEE Signal Process. Lett.*, vol. 15, pp. 833–836, 2008.

[15] R. D. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," *IEEE Trans. Image Process.*, vol. 8, no. 10, pp. 1408–1419, Oct. 1999.

[16] J. Cheng, Z. Li, Z. Gu, H. Fu, D. W. K. Wong, and J. Liu, "Structure-preserving guided retinal image filtering and its application for optic disk analysis," *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2536–2546, Nov. 2018.

[17] D. Pascolini and S. P. Mariotti, "Global estimates of visual impairment: 2010," *Brit. J. Ophthalmol.*, vol. 96, no. 5, pp. 614–618, May 2012.

[18] S. Resnikoff, D. Pascolini, D. Etya'Ale, I. Kocur, R. Pararajasegaram, G. P. Pokharel, and S. P. Mariotti, "Global data on visual impairment in the year 2002," *Bull. World Health Org.*, vol. 82, no. 11, pp. 844–851, 2004.

[19] P. Dai, H. Sheng, J. Zhang, L. Li, J. Wu, and M. Fan, "Retinal fundus image enhancement using the normalized convolution and noise removing," *Int. J. Biomed. Imag.*, vol. 2016, pp. 1–12, Sep. 2016.

[20] 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. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 132–139.

[21] D. Bhonsle, V. Chandra, and G. R. Sinha, "Medical image denoising using bilateral filter," *Int. J. Image, Graph. Signal Process.*, vol. 4, no. 6, p. 36, 2012.

[22] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 6, pp. 1397–1409, Jun. 2013.

[23] I. Bankman, *Handbook of Medical Image Processing and Analysis*. Amsterdam, The Netherlands: Elsevier, 2008.

[24] L. Lu, D. Hu, X. Ma, J. Ma, A. Rahmim, and W. Chen, "Dynamic PET denoising incorporating a composite image guided filter," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, Nov. 2014, pp. 1–4.

[25] A. Almazroa, R. Burman, K. Raahemifar, and V. Lakshminarayanan, "Optic disc and optic cup segmentation methodologies for glaucoma image detection: A survey," *J. Ophthalmol.*, vol. 2015, pp. 1–28, Nov. 2015.

[26] M. S. Haleem, L. Han, J. van Hemert, and B. Li, "Automatic extraction of retinal features from colour retinal images for glaucoma diagnosis: A review," *Comput. Med. Imag. Graph.*, vol. 37, nos. 7–8, pp. 581–596, Oct. 2013.

[27] C. Sinthanayothin, J. F. Boyce, H. L. Cook, and T. H. Williamson, "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images," *Brit. J. Ophthalmol.*, vol. 83, no. 8, pp. 902–910, Aug. 1999.

[28] T. Walter and J.-C. Klein, "Segmentation of color fundus images of the human retina: Detection of the optic disc and the vascular tree using morphological techniques," in *Medical Data Analysis*, J. Crespo, V. Maojo, and F. Martin, Eds. Berlin, Germany: Springer, 2001, pp. 282–287.

[29] D. W. K. Wong, J. Liu, J. H. Lim, X. Jia, F. Yin, H. Li, and T. Y. Wong, "Level-set based automatic cup-to-disc ratio determination using retinal fundus images in ARGALI," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 2266–2269.

[30] J. Cheng, J. Liu, Y. Xu, F. Yin, D. W. K. Wong, N.-M. Tan, D. Tao, C.-Y. Cheng, T. Aung, and T. Y. Wong, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 1019–1032, Jun. 2013.

[31] N. Thakur and M. Juneja, "Survey on segmentation and classification approaches of optic cup and optic disc for diagnosis of glaucoma," *Biomed. Signal Process. Control*, vol. 42, pp. 162–189, Apr. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1746809418300211

[32] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[33] X. Chen, Y. Xu, D. W. K. Wong, T. Y. Wong, and J. Liu, "Glaucoma detection based on deep convolutional neural network," in *Proc. 37th Annu. Int. Conf., IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 715–718.

[34] X. Chen, Y. Xu, S. Yan, D. W. K. Wong, T. Y. Wong, and J. Liu, "Automatic feature learning for glaucoma detection based on deep learning," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 669–677.

[35] H. Fu, J. Cheng, Y. Xu, C. Zhang, D. W. K. Wong, J. Liu, and X. Cao, "Disc-aware ensemble network for glaucoma screening from fundus image," *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2493–2501, Nov. 2018.

[36] J. Zilly, J. M. Buhmann, and D. Mahapatra, "Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation," *Comput. Med. Imag. Graph.*, vol. 55, pp. 28–41, Jan. 2017.

[37] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[38] Z. Yan, X. Yang, and K.-T. Cheng, "Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1912–1923, Sep. 2018.

[39] B. Wang, S. Qiu, and H. He, "Dual encoding U-Net for retinal vessel segmentation," in *Proc. MICCAI.* Cham, Switzerland: Springer, 2019, pp. 84–92.

[40] 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. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 132–139.

[41] 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, Apr. 2004.

[42] 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.

[43] Z. Zhang, F. S. Yin, J. Liu, W. K. Wong, N. M. Tan, B. H. Lee, J. Cheng, and T. Y. Wong, "ORIGA-light: An online retinal fundus image database for glaucoma analysis and research," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug. 2010, pp. 3065–3068.

[44] J. I. Orlando *et al.*, "REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Med. Image Anal.*, vol. 59, Jan. 2020, Art. no. 101570.

[45] 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.

[46] G. D. Joshi, J. Sivaswamy, and S. R. Krishnadas, "Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment," *IEEE Trans. Med. Imag.*, vol. 30, no. 6, pp. 1192–1205, Jun. 2011.

[47] F. Yin, J. Liu, S. H. Ong, Y. Sun, D. W. K. Wong, N. M. Tan, C. Cheung, M. Baskaran, T. Aung, and T. Y. Wong, "Model-based optic nerve head segmentation on retinal fundus images," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 2626–2629.

[48] Y. Xu, L. Duan, S. Lin, X. Chen, D. W. K. Wong, T. Y. Wong, and J. Liu, "Optic cup segmentation for glaucoma detection using low-rank superpixel representation," in *Proc. MICCAI.* Cham, Switzerland: Springer, 2014, pp. 788–795.

[49] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.

[50] P. Yin, Q. Wu, Y. Xu, H. Min, M. Yang, Y. Zhang, and M. Tan, "PM-Net: Pyramid multi-label network for joint optic disc and cup segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Oct. 2019, pp. 129–137.

[51] J. Zhang, B. Dashtbozorg, E. Bekkers, J. P. W. Pluim, R. Duits, and B. M. ter Haar Romeny, "Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores," *IEEE Trans. Med. Imag.*, vol. 35, no. 12, pp. 2631–2644, Dec. 2016.

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

[53] J. Zhuang, "LadderNet: Multi-path networks based on U-Net for medical image segmentation," 2018, *arXiv:1810.07810*. [Online]. Available: http://arxiv.org/abs/1810.07810

[54] B. Liu, L. Gu, and F. Lu, "Unsupervised ensemble strategy for retinal vessel segmentation," in *Proc. MICCAI.* Cham, Switzerland: Springer, 2019, pp. 111–119.

[55] Y. Wu, Y. Xia, Y. Song, D. Zhang, D. Liu, C. Zhang, and W. Cai, "Vessel-Net: Retinal vessel segmentation under multi-path supervision," in *Proc. MICCAI.* Cham, Switzerland: Springer, 2019, pp. 264–272.

[56] Z. Fan, J. Lu, C. Wei, H. Huang, X. Cai, and X. Chen, "A hierarchical image matting model for blood vessel segmentation in fundus images," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2367–2377, May 2019.
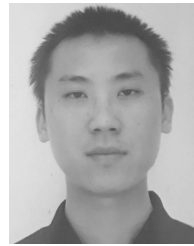
PENGSHUAI YIN received the bachelor's and master's degrees from the South China University of Technology, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree with the School of Software Engineering. His current research interests include computer vision and medical image processing.

RUI YUAN received the B.S. degree from the School of Communication & Design, Sun Yat-sen University, and the M.S. degree from the School of Journalism and Communication, The Chinese University of Hong Kong. His research interests include big data analysis and data mining.

YIMING CHENG received the master's degree in business administration from Missouri State University, in 2019. His research interests include finance and management and data mining.

QINGYAO WU (Member, IEEE) received the B.S. degree in software engineering from the South China University of Technology, China, in 2007, and the Ph.D. degree in computer science from the Harbin Institute of Technology, China, in 2013. He is currently a Professor with the School of Software Engineering, South China University of Technology. His current research interests include computer vision and data mining.

● ● ●