



# Retinal Vessel Segmentation via Multiscaled Deep-Guidance

Rui Xu<sup>1,2</sup>, Guiliang Jiang<sup>1,2</sup>, Xinchun Ye<sup>1,2(✉)</sup>, and Yen-Wei Chen<sup>3</sup>

<sup>1</sup> DUT-RU International School of Information Science and Engineering,  
Dalian University of Technology, Dalian, China  
yexch@dlut.edu.cn

<sup>2</sup> Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province,  
Dalian, China

<sup>3</sup> College of Science and Engineering, Ritsumeikan University, Kusatsu, Shiga, Japan

**Abstract.** Retinal vessel segmentation is a fundamental and crucial step to develop a computer-aided diagnosis (CAD) system for retinal images. Retinal vessels appear as multiscaled tubular structures that are variant in size, length, and intensity. Due to these vascular properties, it is difficult for prior works to extract tiny vessels, especially when ophthalmic diseases exist. In this paper, we propose a multiscaled deeply-guided neural network, which can fully exploit the underlying multiscaled property of retinal vessels to address this problem. Our network is based on an encoder-decoder architecture which performs deep supervision to guide the training of features in layers of different scales, meanwhile it fuses feature maps in consecutive scaled layer via skip-connections. Besides, a residual-based boundary refinement module is adopted to refine vessel boundaries. We evaluate our method on two public databases for retinal vessel segmentation. Experimental results show that our method can achieve better performance than the other five methods, including three state-of-the-art deep-learning based methods.

**Keywords:** Retinal vessel segmentation · Multiscaled deep-guidance  
Deep convolutional neural network

## 1 Introduction

Retinal vessel is a significant gist for diabetes, glaucoma and arteriosclerosis in clinical diagnosis. Segmentation of retinal vessels is a fundamental and crucial step for a CAD system of retinal fundus images. However, segmentation of the whole vascular trees is not easy. The retinal vasculature is composed of arteries and veins appearing as elongated features, with their tributaries visible within the retinal image [1]. Although the intensity profile of the vessel cross-section could be approximated by a Gaussian or a mixed Gaussian function, local grey level of blood vessels can vary hugely due to the effect of lighting condition and retinal pathology. Besides, blood vessel trees are dominated in the multiscaled

property. Vessel widths are variant in a wide range, from one pixel to several tens of pixels. The terminal parts of vessels are so tiny that they look similar as the surrounding background.

Previous methods of retinal vessel segmentation can be divided into two types. The first type, which we refer to as the model-based method, constructs a mathematical model for retinal vessels according to the properties including intensity, shape, gradient and contrast. Examples of this type are the Gaussian kernel-based filter-banks [2,3], the divergence of normalized gradient vector field [5], kernels based on locally adaptive derivative frames [4] and the active contour-based method [6]. The second type, which we refer to as the learning-based method, resolve the retinal vessel segmentation by using a binary classification framework, where each pixel is classified to be vessels or not. Prior works obey a classical routine, where the designation of hand-crafted features is followed by the training of a discriminative classifier. There are researches that design representative features by using ridges of vessels [7], line operators [8] and 2-D Gabor wavelet filter-banks [9]. Classical techniques, such as  $K$ -nearest-neighbor classifier [7], support vector machines [8] and Bayesian classifier [9], are utilized to train classifiers for vessel segmentation. The above-mentioned methods are highly depended on some pre-defined assumptions, and they are easily failed when noise, abnormal lighting conditions and retinal pathology exist due to the mismatching of the assumptions.

In the past several years, deep learning techniques have been extensively evolved in the field of computer vision. After the deep convolutional neural networks (DCNN) exhibit extraordinary performance on the task of image classification [10–13], they are used to deal with other vision-related tasks, such as semantic segmentation [14]. The most-widely used DCNN for segmentation is based on an encoder-decoder architecture [15,16], where feature maps in different scaled layers are learnt. Taking advantage of the inherent multiscaled feature maps can relieve the problem of spatial information loss in DCNN-based methods. For example, skip-connection is adopted to fuse high-level and low-level features in previous works [14,16]. Besides, deep-guidance, which supervises the training of features in different scales, is another technique to exploit multiscaled information in DCNN-based methods, and it is demonstrated to be very efficient to extract detailed edges of objects in natural images [19].

In this paper, we leverage the latest progress on deep learning for the segmentation of retinal vessels on fundus images. We propose a DCNN-based architecture that can fully exploit the multiscaled property of retinal vessels to ensure accurate segmentation especially for tiny vessels. Deep-guidance is adopted in our encoder-decoder based network to guarantee the training of features in specific scales, meanwhile skip-connections integrate feature maps between consecutively high- and low-scaled layers to exhaustively utilize multiscaled information. Besides, a residual-based boundary refinement module [17] is utilized to ensure clear vessel boundaries. We evaluate the propose DCNN-based method on two public datasets, Digital Retinal Images for Vessel Extraction(DRIVE) [7] and

High-Resolution Fundus (HRF) [18], and compare it with other retinal vessel segmentation methods.

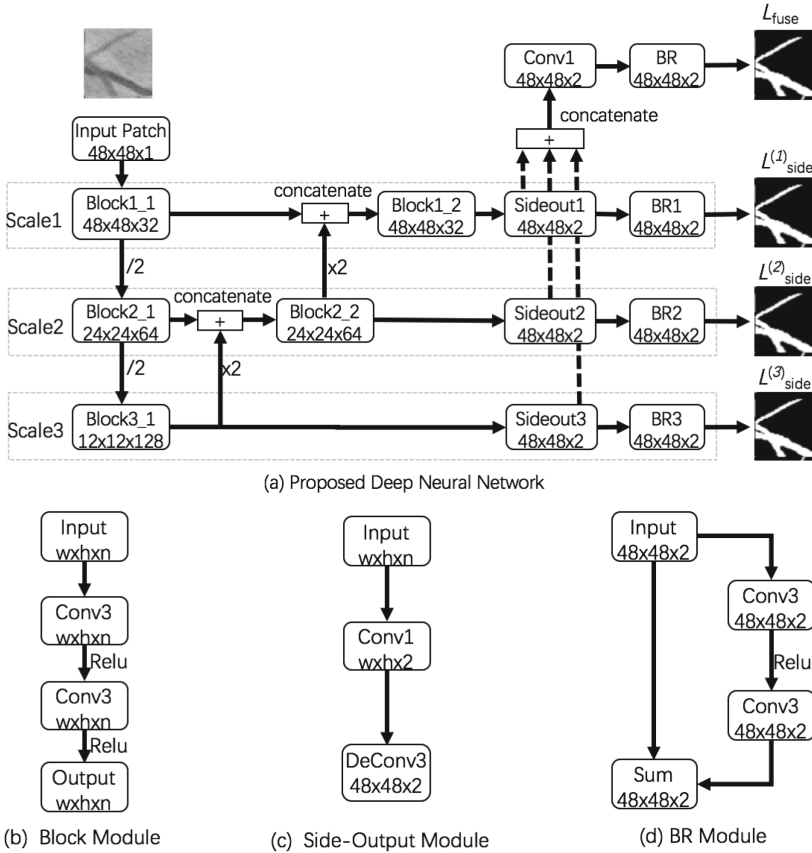


Fig. 1. Architecture of the proposed network

## 2 Related Works

Deep learning has been applied for the segmentation of retinal vessels in several previous works. Liskowski et al. firstly utilize deep learning techniques to segment vessels on retinal fundus images in [20], where several convolutional layers are followed by fully-connected layers to classify each pixel as vessels or not. Deep-guided convolutional neural networks are adopted in the works [21, 22], and a conditional random field reformulated by a recurrent neural network is added to hold interactions between pixels [22]. Recently, Maninis et al. propose a DCNN-based method to segment vessels and optic discs simultaneously [23].

Compared with these methods, the proposed network is mainly different in two aspects. First, we integrate both deep-guidance and skip connections into our network architecture in order to exhaustively exploit multiscaled information of vessels. Second, a residual-based network module is adopted to learn context information for refinement of vessel boundaries. Such carefully designed network architecture ensures our method to be efficient to extract retinal vessels, even for tiny vessels on pathological images.

### 3 Approach

Compared to databases in computer vision, medical databases usually have much fewer images. For example, there are only several tens of fundus images in the two widely-used public databases for the evaluation of retinal vessel segmentation. Considering the few amount of data, we choose to develop a patch-based DCNN method instead of an image-based DCNN method that are widely-used in computer vision and usually requires huge numbers of images for training.

As we notice that multiscale is one of the dominate features for retinal vessels on fundus images, we propose a novel network architecture that can learn and integrate multiscaled information to improve the accuracy of retinal segmentation. Inspired from the ideas in the HED network that is originally proposed for edge extraction [19], we utilize the technique of deep-guidance in our network to ensure the training of more representative feature maps on specific scales. To comprehensively take advantage of information embedded in different scales, we adopt skip connections, which is introduced in U-Net [16], to mix feature maps in consecutively high- and low-level scales. Additionally, a boundary refinement module is integrated into our network to sharpen boundaries of vessels for more accurate segmentation results. The details of the proposed patch-based DCNN method are described in this section.

#### 3.1 Network Architecture

The network architecture is illustrated in Fig. 1. The backbone of our network is based on an encoder-decoder architecture. The encoder part gradually generates feature maps in three different stages by using convolution blocks (details are shown by Fig. 1(b)) that is followed by downsampling. Each convolution block is composed by two convolutional layers, and downsampling halves spatial resolution. Feature maps in each stage have a certain spatial resolution, which correspond to a specific scale.

**Skip Connections** are adopted in the decoder part that gradually recover spatial resolution for accurate segmentation. High-level feature maps are expanded twice in spatial resolution and concatenated with consecutively low-level feature maps. Then, the concatenated features are passed into a convolution block that is composed by two convolutional layers to generate the final feature maps. By using such a kind of network structure, feature maps in two consecutive scales are integrated together.

**Deep Guidance** is adopted to supervise the training of feature maps for a specific spatial scale. Feature maps on each scale are connected outside to a classifier through a side-output and boundary refinement module. The side-output module is composed by a convolutional and deconvolutional layer (Fig. 1(c)) to calculate score maps that have the same spatial resolution with the input. Though such a network structure, feature maps on each scale are trained with the guidance from a specific loss function.

**Boundary Refinement** is a residual model based structure whose details are shown by Fig. 1(d). In our network, this module is connected to score maps in side-output paths to ensure vessel boundaries on each scale to be sharper. Besides, it is also utilized in the fusion path that integrate score maps of all scales to achieve optimal vessel segmentation results.

### 3.2 Network Training

Our network are directly trained from a set of patch-pairs, which are denoted by  $S = \{(X_n, Y_n), n = 1, \dots, N\}$ , where  $X_n = \{x_j^{(n)}, j = 1, \dots, |X_n|\}$  is an image patch and  $Y_n = \{y_j^{(n)}, j = 1, \dots, |X_n|\}, y_j^{(n)} \in \{0, 1\}$  is the corresponding ground truth of segmentation. For simplicity, we denote all parameters of convolution blocks in the backbone as  $\mathbf{W}$ . In this paper, our network has three side-output paths, each of which is associated with a side-output module, boundary refinement module and a classifier. If parameters of all modules are denoted as  $\mathbf{w} = (\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \mathbf{w}^{(3)})$ , the objective function of side-output paths can be represented by Eq. 1.

$$L_{side}(\mathbf{W}, \mathbf{w}) = \sum_{i=1}^3 \alpha_i L_{side}^{(i)}(\mathbf{W}, \mathbf{w}^{(i)}) \quad (1)$$

where  $L_{side}$  denotes the loss function for side-output paths. As the distributions of vessel or non-vessel pixels in patches are highly unbalanced, we adopt the following class-balanced cross-entropy loss function to calculate Eq. 1.

$$\begin{aligned} L_{side}^{(i)}(\mathbf{W}, \mathbf{w}^{(i)}) = & -\beta \sum_{j \in Y_+} \log Pr(y_j = 1|X; \mathbf{W}, \mathbf{w}^{(i)}) \\ & - (1 - \beta) \sum_{j \in Y_-} \log Pr(y_j = 0|X; \mathbf{W}, \mathbf{w}^{(i)}) \end{aligned} \quad (2)$$

where  $\beta = |Y_-|/|Y_+|$  and  $1 - \beta = |Y_+|/|Y_-|$ .  $|Y_+|$  and  $|Y_-|$  denote the vessel and non-vessel ground truth label sets, respectively.

The predicted score maps on each scale is fused together by a fusion-path that includes a convolutional layer, boundary refinement module and a classifier. We denote parameters of all these modules as  $\mathbf{h}$ , and calculate fusion-path loss function  $L_{fuse}(\mathbf{W}, \mathbf{w}, \mathbf{h})$  by using class-balanced cross-entropy. Putting everything together, we minimize the following objective function via standard stochastic gradient descent.

$$L = L_{side}(\mathbf{W}, \mathbf{w}) + L_{fuse}(\mathbf{W}, \mathbf{w}, \mathbf{h}) \quad (3)$$

### 3.3 Implementation

We implement the proposed network by using public Keras library with the back-end of Tensor-Flow. The network is initialized randomly and directly trained on image-patches that are cropped from retinal images. Hyper-parameters include learning-rate (0.01), mini-batch size (32), drop-out rate (0.2) and side-output loss parameter  $\alpha_i$  (1).

Our network takes  $48 \times 48$  image-patches as input and predict segmentation probability maps of the same resolution. When the network is used to segment a retinal image, a sliding window manner is operated on the image with a stride of 8 pixels, and average probability is calculated in overlapped regions.

## 4 Experiments

### 4.1 Data, Preprocessing and Evaluation Protocol

We evaluate our proposed method on two publicly available datasets, which are DRIVE [7] and HRF [18]. The DRIVE database includes 40 fundus images, 33 of which have nearly no-sign of diabetic retinopathy. The rest 7 images only show mild diabetic retinopathy. The HRF database contains 45 images including 15 for healthy patients, 15 for diabetic retinopathy and 15 for glaucomatous patients.

In experiments, all fundus images are processed by the following preprocessing: (1) color images are converted to grey-scale images. (2) a histogram equalization routine called CLAHE [24] is operated on all gray-scale images. (3) gamma correction is performed.

We choose 20 images in DRIVE and 30 images in HRF as the training set, and the rest images in two databases as testing set. In order to get image patch pairs for training,  $48 \times 48$  image patches are randomly cropped from the training set to train deep neural networks. Finally, we obtain 200,000 patches in DRIVE and 240,000 patches in HRF for the training. We evaluate different methods on the testing set by calculating the area under curve (AUC) for receiver operating characteristic (ROC) and precision and recall curve (PR).

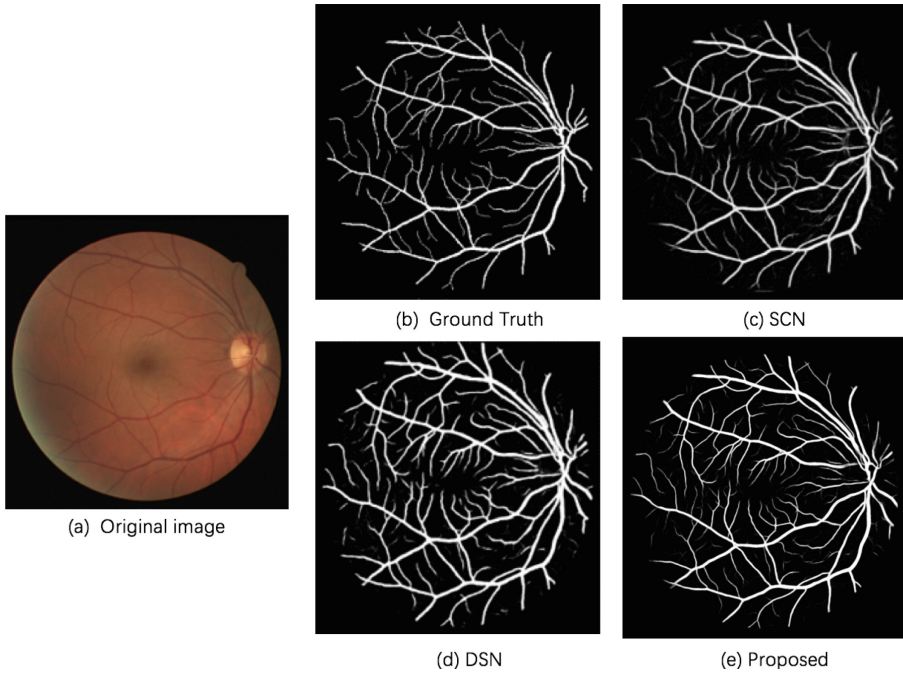
### 4.2 Network Ablation

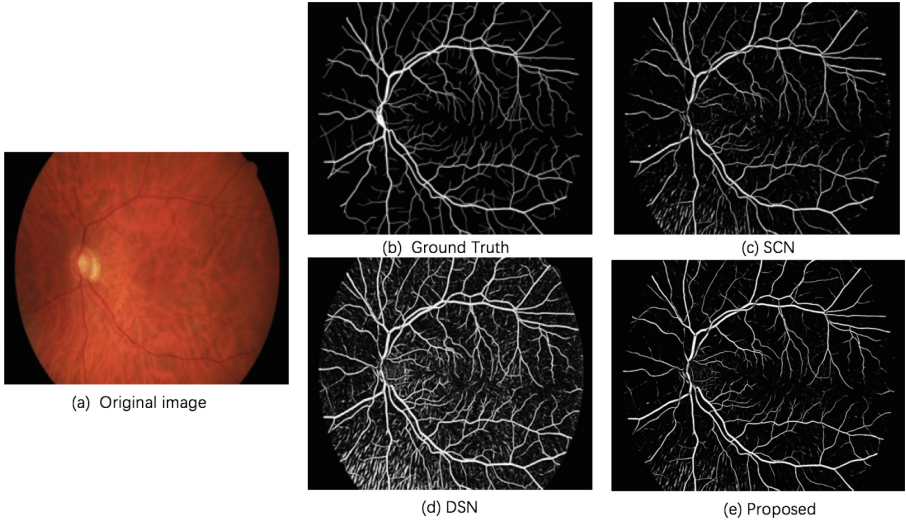
In this subsection, we evaluate the proposed method by gradually adding the network modules mentioned in Sect. 2. Here, we compared 4 different kinds of network architecture by evaluating the AUC of ROC and PR on DRIVE and HRF databases. The comparison results are given in Table 1. The four kinds of networks are denoted as (1) SCN, which only includes the skip-connections but no deep guidance or boundary refinement, (2) DSN, which only has the deep guidance without the other two network modules (3) MDGN, which exploits multiscaled information by skip-connections and deep guidance but boundary refinement is not used (4) MDGN-BR, which is the proposed network architecture using all techniques mentioned in Sect. 2.

**Table 1.** Experiments of network ablation.

Methods	DRIVE		HRF	
	ROC	PR	ROC	PR
SCN	0.9717	0.8895	0.9727	0.8577
DSN	0.9747	0.8999	0.9707	0.8609
MDGN	0.9786	0.9090	0.9766	0.8802
MDGN-BR	<b>0.9793</b>	<b>0.9104</b>	<b>0.9770</b>	<b>0.8805</b>

From Table 1, it can be seen that the AUC of ROC and PR for both SCN and DSN are relatively low. SCN includes skip-connections for integrating feature maps in different scales, however there is no network structure to ensure that feature maps in each scale can be learnt optimally. Thus, it is not able to fully exploit multiscaled information by only using skip-connections. A similar situation happens when only deep guidance is adopted in DSN, which only has network structures to supervise the learning of feature maps in different scales but no structures to fuse them. By utilizing both skip-connections and deep

**Fig. 2.** Examples of probability map (or score map) for vessel segmentation on DRIVE database



**Fig. 3.** Examples of probability map (score map) for vessel segmentation on HRF database

guidance, network performance can be highly improved, as shown by MDGN in Table 1. AUC values of both ROC and PR can be further improved when boundary refinement module is adopted, which demonstrates the efficiency of such a network module utilized in our proposed method.

Figures 2 and 3 give an example of vessel-like probability map for the databases of DRIVE and HRF respectively. It can be seen that the proposed method can give higher probability scores for tiny vessels, which visually demonstrates that our network can exhaustively exploit multiscaled information to improve the performance of vessel segmentation. Besides, we also notice that our method is robust to retinal pathology. Compared with DRIVE, HRF includes more cases with severe retinal pathology, which usually degrades performance of vessel segmentation methods. Due to pathological effect, severe short-tubular artifacts exist in the results of SCN and DCN, as shown by the left bottom part in Fig. 3(c) and (d). However, our method is not much affected by retinal pathology and less artifacts exist in our result (Fig. 3(e)).

### 4.3 Comparison with Other Methods

We compare the proposed network with other methods in this subsection. Table 2 gives the comparison results. Here, we implement two state-of-the-art deep learning-based methods, which are widely used in computer vision and denoted as HED [19] and FCN [14] respectively. We adopt the pre-trained VGG network [11] as the backbones for both of them, and then finely tune network parameters on retinal images. From Table 2, it can be seen that the two deep-learning based method achieve relatively low performances. Both of them take the whole retinal images as the input, however, DRIVE and HRF include only several tens



**Table 2.** Comparison with different kinds of methods.

Methods	DRIVE		HRF	
	ROC	PR	ROC	PR
HED [19]	0.9662	0.8728	0.9511	0.8022
FCN [14]	0.9558	0.8494	0.9464	0.7815
LADF [4]	0.9636	–	0.9608	–
DCNN-FC [20]	0.9720	–	–	–
Line [8]	0.9633	–	–	–
DRIU [23]	–	0.822	–	–
Proposed	<b>0.9793</b>	<b>0.9104</b>	<b>0.9770</b>	<b>0.8805</b>

of retinal images, which are not enough for training. Although fine-tuning of a pre-trained network on few images can give relative good segmentation, that can not relieve the full power of the network. This is also the reason why we choose patch-based DCNN method in this paper.

The evaluation quantities for the other 4 methods in Table 2 are directly copied from the original papers. DCNN-FC is the first work that utilizes deep convolutional neural network for retinal segmentation [20], which adopts stacked convolutional and fully-connected layers to classify pixels. LADF [4] takes advantage of locally adaptive derivative frames to design an optimal filter to extract retinal vessels. Line [8] extracts features by using line operator and trains  $K$ -NN classifier for pixel classification. DRIU [23] uses a base network architecture on which two set of specialized layers are trained to solve both the retinal vessel and optic disc segmentation. Compared with these methods, the proposed method achieves the highest AUC values for both ROC and PR in the two public databases. These results demonstrate the efficiency of the proposed method for retinal vessel segmentation.

## 5 Conclusion

In this paper, we propose a novel DCNN-based network to segment retinal vessels from fundus images. In order to exhaustively exploit multiscaled information on retinal images, skip connections and deep guidance are utilized in our network to ensure better learning of features in different scales and fusion of them for vessel segmentation. Besides, boundary refinement module is adopted to make sharper vessel boundaries. By using these techniques, the proposed method outperformed other retinal segmentation methods on the public DRIVE and HRF databases.

**Acknowledgement.** This work was supported by National Natural Science Foundation of China (NSFC) under Grant 61772106 and Grant 61702078, and by the Fundamental Research Funds for the Central Universities.

## References

1. Fraz, M.M., Remagnino, P., et al.: Blood vessel segmentation methodologies in retinal images—a survey. *Comput. Methods Programs Biomed.* **108**(1), 407–433 (2012)
2. Chaudhuri, S., Chatterjee, S., et al.: Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Trans. Med. Imaging* **8**(3), 263–269 (1989)
3. Hoover, A., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Med. Imaging* **19**(3), 203–210 (2000)
4. Zhang, J., Dashtbozorg, B., et al.: Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. *IEEE Trans. Med. Imaging* **35**(12), 2631–2644 (2016)
5. Lam, B.Y., Yan, H.: A novel vessel segmentation algorithm for pathological retina images based on the divergence of vector fields. *IEEE Trans. Med. Imaging* **27**(2), 237–246 (2008)
6. Al-Diri, B., Hunter, A., Steel, D.: An active contour model for segmenting and measuring retinal vessels. *IEEE Trans. Med. Imaging* **28**(9), 1488–1497 (2009)
7. Staal, J., Abramoff, M.D., et al.: Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **23**(4), 501–509 (2004)
8. Ricci, E., Perfetti, R.: Retinal blood vessel segmentation using line operators and support vector classification. *IEEE Trans. Med. Imaging* **26**(10), 1357–1365 (2007)
9. Soares, J.V., Leandro, J.J., et al.: Retinal vessel segmentation using the 2-D gabor wavelet and supervised classification. *IEEE Trans. Med. Imaging* **25**(9), 1214–1222 (2006)
10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: *NIPS* (2012)
11. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: *ICLR* (2015)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *CVPR* (2015)
13. Szegedy, C., Liu, W., et al.: Going deeper with convolutions. In: *CVPR* (2015)
14. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(4), 640–651 (2017)
15. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017)
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
17. Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J.: Large kernel matters - improve semantic segmentation by global convolutional network. In: *CVPR* (2017)
18. Budai, A., Bock, R., Maier, A., Hornegger, J., Michelson, G.: Robust vessel segmentation in fundus images. *Int. J. Biomed. Imaging* **2013**, 11 (2013)
19. Xie, S., Tu, Z.: Holistically-nested edge detection. In: *ICCV* (2015)
20. Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels with deep neural networks. *IEEE Trans. Med. Imaging* **35**(11), 2369–2380 (2016)

21. Mo, J., Zhang, L.: Multi-level deep supervised networks for retinal vessel segmentation. *Int. J. Comput. Assist. Radiol. Surg.* **12**(9), 1–13 (2017)
22. Fu, H., Xu, Y., Lin, S., Kee Wong, D.W., Liu, J.: DeepVessel: retinal vessel segmentation via deep learning and conditional random field. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) *MICCAI 2016*. LNCS, vol. 9901, pp. 132–139. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46723-8\\_16](https://doi.org/10.1007/978-3-319-46723-8_16)
23. Maninis, K.-K., Pont-Tuset, J., Arbeláez, P., Van Gool, L.: Deep retinal image understanding. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) *MICCAI 2016*. LNCS, vol. 9901, pp. 140–148. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46723-8\\_17](https://doi.org/10.1007/978-3-319-46723-8_17)
24. Pizer, S.M., Amburn, E.P.: Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Image Process.* **39**(3), 355–368 (1987)