



Residual Multiscale Full Convolutional Network (RM-FCN) for High Resolution Semantic Segmentation of Retinal Vasculature

Tariq M. Khan¹(✉), Antonio Robles-Kelly¹, Syed S. Naqvi²,
and Muhammad Arsalan³

¹ The School of Information Technology, Faculty of Science Engineering and Built Environment, Deakin University, Waurn Ponds, VIC 3216, Australia

tariq.khan@deakin.edu.au

² Department of ECE, COMSATS University Islamabad, Islamabad Campus, Islamabad, Pakistan

³ Division of Electronics and Electrical Engineering, Dongguk University, 30 Pildong-ro, 1-gil, Jung-gu, Seoul 100-715, Korea

Abstract. In a fundus image, Vessel local characteristics like direction, illumination and noise vary considerably, making vessel segmentation a challenging task. Methods based upon deep convolutional networks have consistently yield state of the art performance. Despite effective, of the drawbacks of these methods is their computational complexity, whereby testing and training of these networks require substantial computational resources and can be time consuming. Here we present a multi-scale kernel based on fully convolutional layers that is quite lightweight and can effectively segment large, medium, and thin vessels over a wide variations of contrast, position and size of the optic disk. Moreover, the architecture presented here makes use of these multi-scale kernels, reduced application of pooling operations and skip connections to achieve faster training. We illustrate the utility of our method for retinal vessel segmentation on the DRIVE, CHASE.DB and STARE data sets. We also compare the results delivered by our method with a number of alternatives elsewhere in the literature. In our experiments, our method always provides a margin of improvement on specificity, accuracy, AUC and sensitivity with respect to the alternative.

Keywords: Retinal vessel segmentation · Convolutional neural networks · Diabetic retinopathy

1 Introduction

Retinal fundus images contain important features often used to diagnose eye-related illnesses such as diabetic retinopathy (DR), glaucoma, age-related macular degeneration (AMD) and systemic illnesses such as arteriosclerosis and hypertension. Among these diseases, DR and AMD are the major causes of blindness

[1, 2]. Fundus images, acquired during an ophthalmic exam, are used to inspect and monitor DR and AMD disease progression. As a result, a computer-aided diagnosis system that can significantly reduce the burden on the ophthalmologists and alleviate the inter and intra observer variability is highly desired.

Here, we focus on the segmentation of retinal blood vessels. These originate from the centre of optic disc and spread over the other regions of the retina. The blood vessels are responsible for supplying blood to the entire region of the retina, whereby microaneurysm, hemorrhages and exudate lesions are formed in the retinal image due to leakages taking place and appear as bright spots in the fundus image. Recently, convolutional neural networks (CNNs) have gained significant importance in semantic segmentation [3]. Methods such as those presented in [4–6] have yielded state of the art performance. Moreover, approaches such as that in [7] are able to address the pixel-wise classification problem by mapping low resolution features produced by the encoder back to the input resolution through a decoder. The advantage of such mapping resides in the fact that they can preserve fine-grained information, which is of capital importance for effective boundary detection.

As related to retinal vessel segmentation, the authors in [8] explore a deep learning approach that focuses on the thickness of the retinal vasculature. In [9], the authors present a skip connection encoder-decoder architecture that is quite effective detecting vessel boundaries. Gu et al. [10] present a context encoder for vessel segmentation network. Yan et al. [8] introduced a joint-loss including both a pixel-wise and a segmentation-level cost. Despite the higher accuracy of these deep learning methods, there are still many problems that demand significant attention from researchers. One of the drawbacks of these methods is their computational complexity, whereby both the pre-processing and post-processing tasks needed for deep learning approaches require substantial computational resources, training and testing times.

This paper presents a residual multiscale full convolutional network (RM-FCN) for retinal vessel segmentation. The proposed method is quite lightweight compared to other methods elsewhere in the literature, with only 6 convolutional layers with 3 multi-scale fully convolutional kernels per layer. The proposed model not only is able to accurately detect thick vessels but, when applied to the thin ones, these are also segmented due to the use of our multi-scale architecture. In our networks only two max-pooling operations are required and these are paired with external skip-connections. This yields an architecture that makes use of reduced convolutional layers, multi-scale kernels and reduced application of pooling operations so as to achieve a faster training. The rest of the paper is organized as follows. Our architecture is in Sect. 2. We then present results for retinal image segmentation and compare to alternatives in Sect. 3. Finally, in Sect. 4, we conclude on the developments presented here.

2 Residual Multiscale Network

Recall that, in retinal vessel segmentation applications, the vessel size may vary considerably across patients with a variety of medical conditions. Diabetic

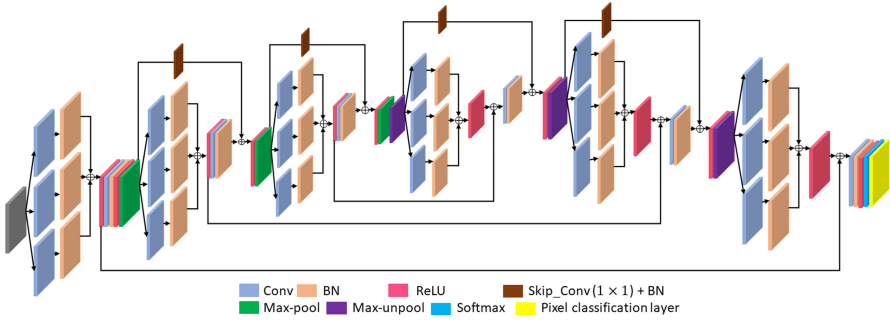


Fig. 1. Block diagram of the proposed method

retinopathy can cause the swelling of the retinal vessels and can also encourage the development of smaller, newer ones. Hypertensive retinopathy, in the other hand, can cause the shrinkage of retinal vessels. As mentioned above, here we employ multi-scale kernels to develop a neural network architecture that can cope with large size variations.

Neural networks elsewhere in the literature often employ a single sized convolutional kernel which often focuses in larger vessels and, therefore, is not quite effective for the segmentation of smaller vascular structures. This accounts for the notion that very thin vessels may not affect overly affect the overall performance in terms. This is debatable since several diagnosis in medical applications heavily rely upon small-sized vessels. Our multiscale kernels are based on 3×3 , 5×5 , and 7×7 convolutions for large, medium, and very small vessels, respectively. The architecture of our RM-FCN is illustrated in Fig. 1.

To construct our network, we have used multiscale convolutional blocks with important design concerns. The first of these is to keep to a minimum the use of pooling layers which are used to reduce the dimension of the feature maps. This is since these pooling operations also cause the loss of spatial information. Secondly, we employ multi-scale kernels so as to account for the large variation in retinal vessel sizes. Thirdly, we reduce the overall number of convolutions in the network. These can also be responsible for spatial information loss. Finally, we employ fine-grained information and residual skip paths to improve the segmentation results and make training more computationally efficient. Figure 2 shows the overall architecture of our proposed multi-scale convolutional blocks within the network. The network has six multi-scale convolutional blocks, where the first block is an input one, followed by two down multi-scale blocks. There is an intermediate block which connects down and up blocks. This is followed by the two up-multiscale convolutional blocks with a final output one which is equipped with a softmax loss layer.

In Fig. 2 presents the example up multi-scale convolutional block, which receives the feature map F from the pooling layer and distributes them to the convolutions C_3^A , C_5^A , C_7^A and C_1^A . Note that C_1^A is, in fact, part of the skip connection. These kernels have sizes 3×3 , 5×5 , 7×7 , 1×1 , respectively. Each of the multi-scale convolutional kernels C_3^A , C_5^A , C_7^A outputs the features F_a , F_b , F_c , respectively. These are given by

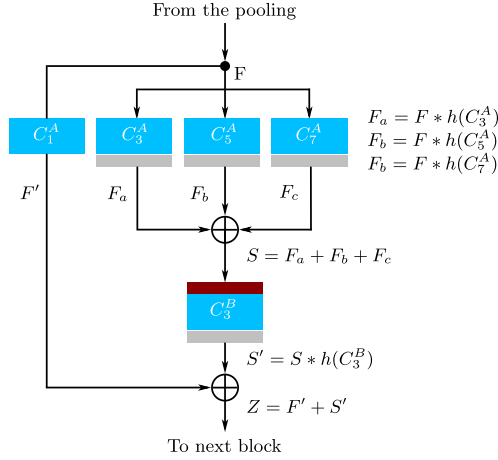


Fig. 2. Block diagram of the proposed multiscale convolutional block.

$$\begin{aligned}
 F_a &= F * C_3^A \\
 F_b &= F * C_5^A \\
 F_c &= F * C_7^A
 \end{aligned} \tag{1}$$

which are then used to obtain S , which is given by

$$S = F_a + F_b + F_c \tag{2}$$

Thus, S can be viewed as a combined feature map which can later be fed into a ReLU and batch normalized. This is done after an additional convolution C_3^B is applied so as to obtain the feature map S' given by

$$S' = S * C_3^B \tag{3}$$

where S' is the multi-scale feature map. To further improve the feature map quality S' is combined with F' , which arises from the skip path comprising C_1^A (a 1×1 convolutional kernel). This yield the feature map Z given by $Z = F' + S'$.

As shown it the figure, the encoder blocks generate the respective feature maps using convolutions between the input image and a multi-scale filter bank. Here, we have followed [11] and applied batch normalisation on the features followed by a ReLU. For the down sampling blocks, the resulting feature maps are fed to the a 2×2 , non-overlapping max-pooling with a stride of size 2. In this manner, the down-sampled feature maps created from the final down-sampling block can be used for the up-sampling procedure. This is carried out by using the indices of the max-pooling information. In our architecture, the feature maps yielded by the down-sampling blocks are unpooled. These maps, which are sparse in nature, are augmented in the up-sampling blocks by the multi-scale filter banks. These dense feature maps are then normalized by using batch normalization. The size of the feature maps yielded by the up-sampling

blocks are identical to those obtained by the respective down-sampling blocks. The only difference is in the final layer of the decoder, where a multi-channel function map is obtained as an output compared to the three-channel RGB data of the first encoder. At output, our network yields a final map where pixels are labelled as vessels or not on the basis of a soft-max classifier.

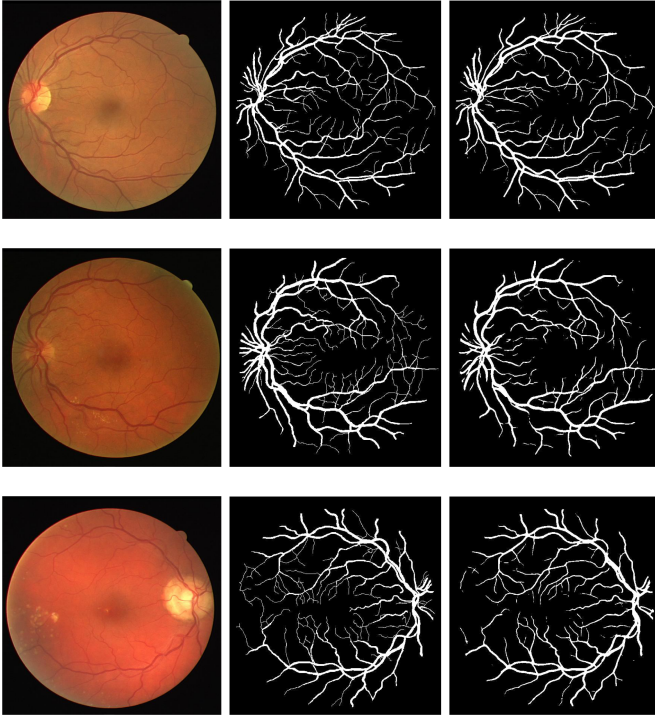


Fig. 3. Segmentation results of the RM-FCN model on three noisy test images i.e. image number 1, 3 and 8 from the DRIVE dataset. From left-to-right, we show the input images, the ground truth segmentation map and that yielded by our method.

3 Experiments

3.1 Datasets

We now turn our attention to the evaluation of our method on three publicly available retinal image databases. These are the CHASE [12]¹, DRIVE [13]² and STARE [14]³ data sets. The DRIVE dataset covers a wide age range of diabetic

¹ The dataset can be found at <https://blogs.kingston.ac.uk/retinal/chasedb1/>.

² The dataset is widely available at <https://drive.grand-challenge.org/>.

³ More information regarding the STARE project can be found at <https://cecas.clemson.edu/~ahoover/stare/>.

patients and consists of 20 color images for training and 20 color images for testing. The STARE dataset is a collection of 20 color retinal fundus images captured at 35° FOV with an image size of 700×605 pixels. Out of these 20 images, 10 images contain pathologies. Two different manual segmentation as ground truth are available. Here we employ the first experts segmentation as ground truth where available. There is no dedicated test dataset available for STARE. The CHASE dataset consists of 28 color images of 14 school children in England. Two different manual segmentation maps are available as ground truth. Again, here we employ the first experts segmentation for our experiments. The CHASE dataset doesn't contain any dedicated training or testing sets. Here we have used the first 20 images for training and the last 8 images for testing.

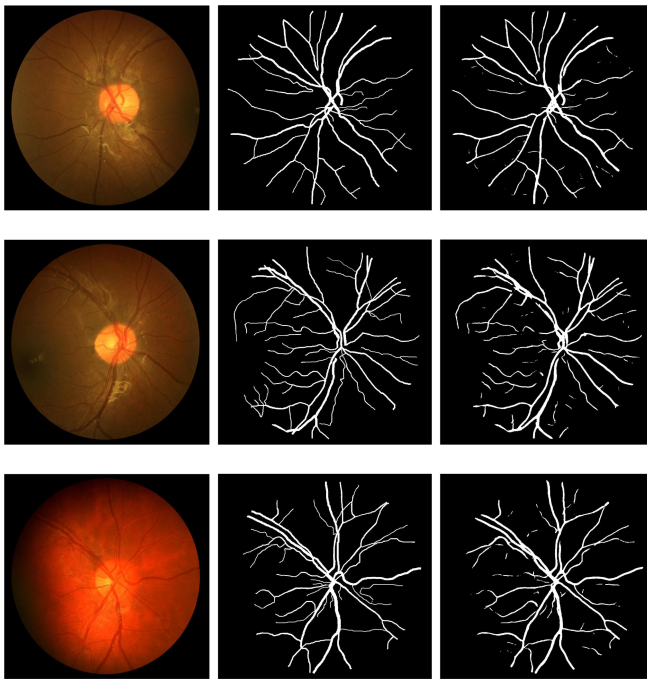


Fig. 4. Segmentation results of the RM-FCN model on images number 5, 6 and 8 from the CHASE_DB dataset. From left-to-right, we show the input images, the ground truth segmentation map and that yielded by our method.

3.2 Results and Comparison

Here we compare the results obtained by our approach on the three data sets above with those yielded by a number of alternatives. For all the methods under consideration we have used four common performance parameters. These are Sensitivity (Se), Specificity (Sp), Accuracy (Acc) and AUC. These results are shown in Tables 1, 2 and 3.

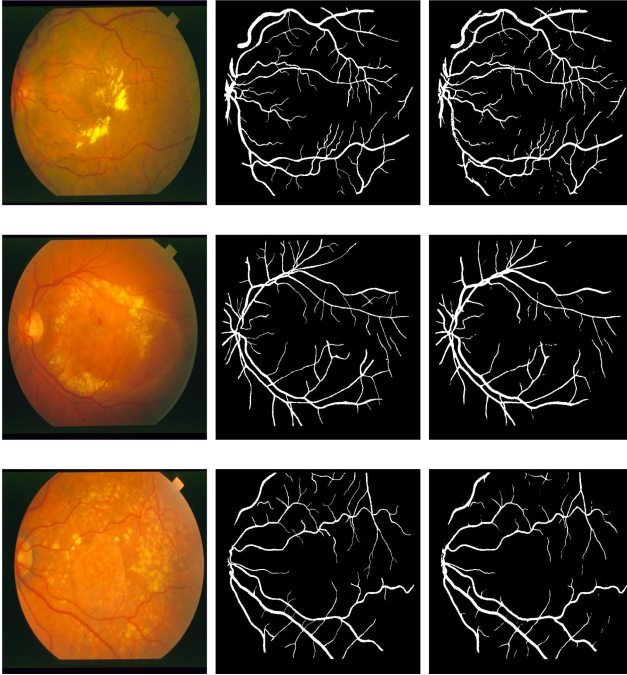


Fig. 5. The segmentation results of the RM-FCN model on the two noisy test images i.e. image number 1, 2 and 3 from the STARE dataset. From left-to-right, we show the input images, the ground truth segmentation map and that yielded by our method.

We also show qualitative results in Figs. 3, 4 and 5 for the three data sets under consideration. In all figures we show, from left-to-right the input imagery, the segmentation ground truth provided by the hand-labeled vessel maps and

Table 1. Performance comparison of our RM-FCN on DRIVE data set with respect to other methods elsewhere in the literature

Method	Year	Se	Sp	Acc	AUC
Guo <i>et al.</i> [15]	2019	0.7891	0.9804	0.9561	0.9806
Ma <i>et al.</i> [16]	2019	0.7916	0.9811	0.9570	0.9810
Wang <i>et al.</i> [17] DU-Net	2019	0.7940	0.9816	0.9567	0.9772
Wu <i>et al.</i> [18]	2019	0.8038	0.9802	0.9578	0.9821
Gu <i>et al.</i> [10] CE-Net	2019	0.8309	–	0.9545	0.9779
Arsalan <i>et al.</i> [19] VessNet	2019	0.8022	0.9810	0.9655	0.9820
Wang <i>et al.</i> [20]	2020	0.7991	0.9813	0.9581	0.9823
Yin <i>et al.</i> [21]	2020	0.8038	0.9837	0.9578	0.9846
Segnet-Basic	2020	0.7949	0.9738	0.9579	0.9720
Our method	2020	0.8342	0.9825	0.9695	0.9830

the results yielded by our method. From the figures, we can see that our method can cope well with thinner vessels, preserving well the fine-grained detail while being quite robust to different conditions, variations in contrast and optic disk position and size.

From Table 1, it is clear that our method’s accuracy is the highest amongst the alternatives for the DRIVE data set. The second best accuracy on the Drive data set is that delivered by the method of Arsalan et al. [19]. In terms of sensitivity, on the DRIVE dataset, our method also achieve the highest value. The second best sensitivity on DRIVE dataset is that of the method in [10] (CE-Net). Similarly, the results presented in Table 2 indicate that method proposed here has the best overall performance on the CHASE data set across all the measures used. The sensitivity achieved by the Arsalan et al. [19] is the second highest in Table 2. The accuracy of Yin et al. [21] is the best among all the approaches under consideration. Finally, Table 3 shows that is also the best performing method on the STARE data set. The sensitivity achieved by the Arsalan et al. [19] is again the second highest.

Table 2. Performance comparison of our RM-FCN on CHASE_DB1 data set with respect to other methods elsewhere in the literature

Method	Year	Se	Sp	Acc	AUC
Zhang <i>et al.</i> [22]	2016	0.7626	0.9661	0.9452	0.9606
Khawaja <i>et al.</i> [23]	2019	0.7974	0.9697	0.9528	NA
Jin <i>et al.</i> [24]	2019	0.7595	0.9878	0.9641	0.9832
Arsalan <i>et al.</i> [19] VessNet	2019	0.8206	0.9800	0.9726	0.9800
Wang <i>et al.</i> [20]	2020	0.8186	0.9844	0.9673	0.9881
Yin <i>et al.</i> [21]	2020	0.7993	0.9868	0.9783	0.9869
Segnet-basic	2020	0.8190	0.9735	0.9638	0.9780
Our method	2020	0.8463	0.9828	0.9735	0.9810

Table 3. Performance comparison of our RM-FCN on STARE database with respect to other methods elsewhere in the literature

Method	Year	Se	Sp	Acc	AUC
Chen <i>et al.</i> [25] Deeplab v3++	2018	0.8320	0.9760	0.9650	0.9735
Jin <i>et al.</i> [24]	2019	0.8155	0.9752	0.9610	0.9804
Guo <i>et al.</i> [15]	2019	0.7888	0.9801	0.9627	0.9840
Wang <i>et al.</i> [17]	2019	0.8074	0.9821	0.9661	0.9812
Wu <i>et al.</i> [18]	2019	0.8132	0.9814	0.9661	0.9860
Arsalan <i>et al.</i> [19] VessNet	2019	0.8526	0.9791	0.9697	0.9883
Wang <i>et al.</i> [20]	2020	0.8239	0.9813	0.9670	0.9871
SegNet-Basic	2020	0.8118	0.9738	0.9543	0.9728
Our method	2020	0.8565	0.9834	0.9739	0.9890

4 Conclusions

In this paper we have presented a residual multi-scale network for retinal vessel segmentation that employs skip connections, multiscale filters and a reduced number of pooling operations so as to segment large, medium and thin vasculature under large variations of contrast, optic disk position and size. We have illustrated the utility of the method for the task in hand by performing experiments on three publicly accessible databases, namely CHASE DB1, STARE and DRIVE. In our experiments, our network outperformed a number of state-of-the-art alternatives. For our comparison, we have used well-known measurement parameters, namely sensitivity, balanced accuracy and accuracy.

References

1. Khan, T.M., Alhusein, M., Aurangzeb, K., Arsalan, M., Naqvi, S.S., Nawaz, S.J.: Residual connection-based encoder decoder network (RCED-net) for retinal vessel segmentation. *IEEE Access* **8**, 131257–131272 (2020)
2. Khan, T.M., Naqvi, S.S., Arsalan, M., Khan, M.A., Khan, H.A., Haider, A.: Exploiting residual edge information in deep fully convolutional neural networks for retinal vessel segmentation. In: 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2020)
3. Khan, T.M., Abdullah, F., Naqvi, S.S., Arsalan, M., Khan, M.A., Shallow vessel segmentation network for automatic retinal vessel segmentation. In: 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–7. IEEE (2020)
4. Khan, T.M., Robles-Kelly, A., Naqvi, S.S.: A semantically flexible feature fusion network for retinal vessel segmentation. In: Yang, H., Pasupa, K., Leung, A.C.-S., Kwok, J.T., Chan, J.H., King, I. (eds.) *ICONIP 2020. CCIS*, vol. 1332, pp. 159–167. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-63820-7_18
5. Khawaja, A., Khan, T.M., Naveed, K., Naqvi, S.S., Rehman, N.U., Nawaz, S.J.: An improved retinal vessel segmentation framework using Frangi filter coupled with the probabilistic patch based denoiser. *IEEE Access* **7**, 164344–164361 (2019)
6. Khan, M.A.U., Khan, T.M., Bailey, D.G., Soomro, T.A.: A generalized multi-scale line-detection method to boost retinal vessel segmentation sensitivity. *Pattern Anal. Appl.* **22**(3), 1177–1196 (2018). <https://doi.org/10.1007/s10044-018-0696-1>
7. 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)
8. Yan, Z., Yang, X., Cheng, K.T.: Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. *IEEE Trans. Biomed. Eng.* **65**, 1912–1923 (2018)
9. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention* (2015)
10. Gu, Z., et al.: CE-net: context encoder network for 2D medical image segmentation. *IEEE Trans. Med. Imaging* **38**(10), 2281–2292 (2019)
11. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *International Conference on Machine Learning*, pp. 448–456 (2015)

12. Fraz, M.M., et al.: An approach to localize the retinal blood vessels using bit planes and centerline detection. *Comput. Methods Programs Biomed.* **108**(2), 600–616 (2012c)
13. Staal, J., Abramoff, M.D., Niemeijer, M., Viergever, M.A., van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **23**(4), 501–509 (2004)
14. Hoover, A.D., 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)
15. Guo, S., Wang, K., Kang, H., Zhang, Y., Gao, Y., Li, T.: BTS-DSN: deeply supervised neural network with short connections for retinal vessel segmentation. *Int. J. Med. Inf.* **126**, 105–113 (2019)
16. Ma, W., Yu, S., Ma, K., Wang, J., Ding, X., Zheng, Y.: Multi-task neural networks with spatial activation for retinal vessel segmentation and artery/vein classification. In: *Medical Image Computing and Computer Assisted Intervention* (2019)
17. Wang, B., Qiu, S., He, H.: Dual encoding U-net for retinal vessel segmentation. In: Shen, D., et al. (eds.) *MICCAI 2019*. LNCS, vol. 11764, pp. 84–92. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_10
18. Wu, Y., et al.: Vessel-Net: retinal vessel segmentation under multi-path supervision. In: *Medical Image Computing and Computer Assisted Intervention* (2019)
19. Arsalan, M., Oqais, M., Mahmood, T., Cho, S.W., Park, K.R.: Aiding the diagnosis of diabetic and hypertensive retinopathy using artificial intelligence-based semantic segmentation. *J. Clin. Med.* **8**(9), 1446 (2019)
20. Wang, D., Haytham, A., Pottenburgh, J., Saeedi, O., Tao, Y.: Hard attention net for automatic retinal vessel segmentation. *IEEE J. Biomed. Health Inf.* **24**, 3384–3396 (2020)
21. Yin, P., Yuan, R., Cheng, Y., Wu, Q.: Deep guidance network for biomedical image segmentation. *IEEE Access* **8**, 116106–116116 (2020)
22. Zhang, J., Dashtbozorg, B., Bekkers, E., Pluim, J.P.W., Duits, R., Romeny, B.M.: Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. *IEEE Trans. Med. Imaging* **35**(12), 2631–2644 (2016)
23. Khawaja, A., Khan, T.M., Khan, M.A.U., Nawaz, S.J.: A multi-scale directional line detector for retinal vessel segmentation. *Sensors* **19**(22), 4949 (2019)
24. Jin, Q., Meng, Z., Pham, T.D., Chen, Q., Wei, L., Su, R.: DUNet: a deformable network for retinal vessel segmentation. *Knowl. Based Syst.* **178**, 149–162 (2019)
25. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with Atrous separable convolution for semantic image segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) *ECCV 2018*. LNCS, vol. 11211, pp. 833–851. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_49