Vessel-Net: Retinal Vessel Segmentation Under Multi-path Supervision

Yicheng Wu[1], Yong Xia[1(✉)], Yang Song[3], Donghao Zhang[2], Dongnan Liu[2], Chaoyi Zhang[2], and Weidong Cai[2]

[1] National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China
yxia@nwpu.edu.cn

[2] School of Computer Science, University of Sydney, Sydney, NSW 2006, Australia

[3] School of Computer Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia

Abstract. Due to the complex morphology of fine vessels, it remains challenging for most of existing models to accurately segment them, particularly the capillaries in color fundus retinal images. In this paper, we propose a novel and lightweight deep learning model called Vessel-Net for retinal vessel segmentation. First, we design an efficient inception-residual convolutional block to combine the advantages of the Inception model and residual module for improved feature representation. Next, we embed the inception-residual blocks inside a U-like encoder-decoder architecture for vessel segmentation. Then, we introduce four supervision paths, including the traditional supervision path, a richer feature supervision path, and two multi-scale supervision paths to preserve the rich and multi-scale deep features during model optimization. We evaluated our Vessel-Net against several recent methods on two benchmark retinal databases and achieved the new state-of-the-art performance (i.e. AUC of 98.21%/98.60% on the DRIVE and CHASE databases, respectively). Our ablation studies also demonstrate that the proposed inception-residual block and the multi-path supervision both can produce impressive performance gains for retinal vessel segmentation.

Keywords: Retinal vessel segmentation · Vessel-Net · Inception-residual block · Multi-path supervision

1 Introduction

Retinal vessel segmentation in color fundus images has been widely used for quantitative analysis of ophthalmologic diseases including diabetic retinopathy (DR) and glaucoma [1]. However, it remains challenging to achieve accurate segmentation of retinal vessels, especially the capillaries and other fine structures, largely due to the complex vessel morphology (e.g. the thin and curved vessel).

D. Shen et al. (Eds.): MICCAI 2019, LNCS 11764, pp. 264–272, 2019.
https://doi.org/10.1007/978-3-030-32239-7_30

Traditionally, retinal vessel segmentation has been conducted by designing filter-based features to capture the unique morphological characteristics of vessels. For instance, various types of filters were used to extract the 41-dimensional visual features to describe retinal vessels [2], and a combination of shifted filter responses (COSFIRE) [3] was designed to segment the retinal vessels in fundus images. Recently, deep learning (DL) has been adopted to perform vessel segmentation with promising results. Several data augmentation algorithms were introduced to augment limited training data [4]. In addition, an unsupervised model with image matting was proposed to segment retinal vessels [5]. The additional labels of thick and thin vessels were introduced explicitly and an edge-based mechanism was incorporated into U-Net to achieve a better result [6]. Furthermore, the conditional random field (CRF) method was used for post-processing [7]. A cascaded architecture with multi-scale refinement was also proposed to further improve the segmentation performance [8]. While these DL-based methods have reported encouraging results, we hypothesize that the vessel segmentation performance can be further improved by more effective modeling the multi-scale visual information associated with vessels with varying thickness.

In this paper, we propose a novel and highly effective deep learning model called Vessel-Net for retinal vessel segmentation in color fundus images. The Vessel-Net contains five inception-residual (IR) blocks for better feature representation, and each IR block contains three parallel paths including one residual convolutional layer and two enhancement paths. To the best of our knowledge, we are the first to combine the advantages of Inception and residual methods for retinal vessel segmentation, without introducing too many additional parameters. Furthermore, we design four supervision paths called multi-path supervision to train the proposed Vessel-Net, in which the richer feature supervision combines all feature maps of our Vessel-Net and multi-scale supervision further preserves multi-scale features. With this multi-path supervision, multi-scale complementary information can be better preserved, which is critical to the segmentation of fine structures.

We evaluated our Vessel-Net against several recent retinal vessel segmentation algorithms on two benchmark databases: the digital retinal images for vessel extraction (DRIVE) database [9] and the child heart and health study (CHASE) database [10]. Our results show that the proposed Vessel-Net, even without adjusting any hyper-parameters for each experiment, achieves the area under the ROC curve (AUC) of 98.21% on DRIVE and 98.60% on CHASE and sets the new state of the art.

2 Method

The proposed Vessel-Net has a U-like encoder-decoder architecture, which contains five IR blocks, 2×2 convolutional layers, up-sampling layers, and four supervision paths (see Fig. 1). The convolutional layers with 2×2 kernels and a stride of 2 are designed to contract the feature maps, whereas the corresponding 2×2 up-sampling layers are used to expand the feature maps. Hence, the Vessel-Net takes 48×48 patches extracted from pre-processed retinal images as input

and produces the vessel probability maps with the same size. We now delve into the details of the IR block and multi-path supervision.

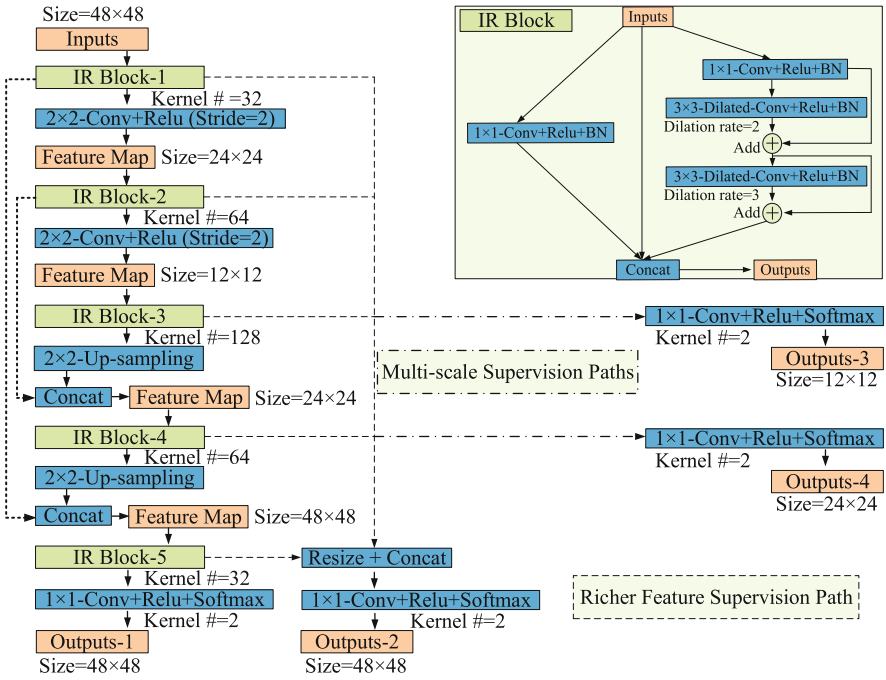

Fig. 1. Diagram of the proposed Vessel-Net.

2.1 Inception-Residual Block

The U-Net and its variants, such as the recurrent residual U-Net (R2U-Net) [11], are popular semantic segmentation tools, which have shown promising performance in many biomedical image applications [12]. The convolutional block in U-Net contains, sequentially, a 3×3 convolutional layer, a dropout layer, and another 3×3 convolutional layer (see Fig. 2). Nevertheless, convolutional block in R2U-Net consists of a 1×1 convolutional layer and two 3×3 convolutional layers, and replaces the dropout layer with skip connections and backward connections to achieve the recurrent learning (see Fig. 2).

To widen the model while avoiding superabundant parameters, we design the IR block (see Fig. 2) to replace the convolutional blocks in U-like networks for a higher feature representation capability. Specifically, we remove backward connections in the convolutional block of R2U-Net, then introduce two parallel enhancement paths (one includes a 1×1 convolutional layer, and the other is a skip connection from the input to output), and finally use a concatenation layer to combine the outputs of three parallel paths. Another difference between the IR

block and the convolutional blocks in U-Net and R2U-Net is that we use dilated convolutions with a dilation rate of 2 and 3 in two 3×3 adjacent convolutional layers, respectively, to realize a large field of view. Compared to Inception block [13], the lightweight IR block does not leverage the convolutional layer of large kernels to extract features since it would cause over-fitting when the training data of retinal vessels is highly limited. Note that, each convolutional layer, with a stride of 1, is followed by the ReLU activation and a batch normalization layer to reduce over-fitting as much as possible.

As shown in Fig. 1, the channels of five IR blocks in our Vessel-Net are 32, 64, 128, 64 and 32 from top to bottom.

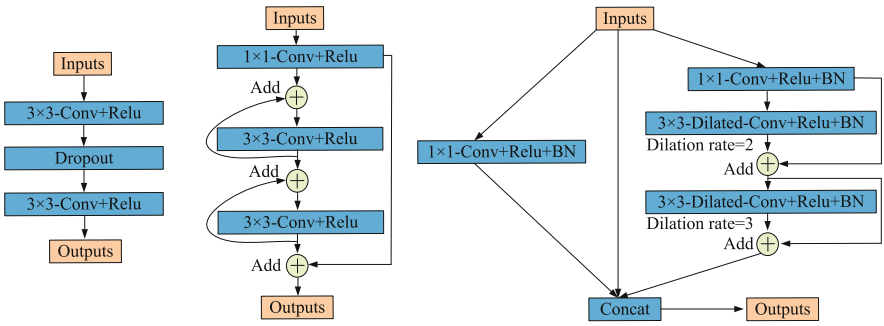

Fig. 2. U-Net block (left), R2U block (middle) and the proposed IR block (right).

2.2 Multi-path Supervision

It has been acknowledged that all feature maps contain essential image details and well deserve to be preserved [14,15]. This is especially true for vessel images, which contain complex structures at different scales. Therefore, we resize the feature maps produced by all IR blocks to 48×48 using the corresponding up-sampling layers, concatenate them, and thus feed them to a 1×1 convolutional layer followed by the ReLU activation and a Softmax layer to generate an auxiliary output rP (see Fig. 1). This output path enables the feature maps obtained at different depths to be preserved and further processed to provide the richer feature supervision to model training. In addition, we feed the outputs of the 3rd and 4th IR block to a 1×1 convolutional layer followed by the ReLU activation and a Softmax layer, respectively, resulting in a 12×12 output msP_3 and a 24×24 output msP_4. These two paths enable the feature maps at two scales to provide the multi-scale supervision to model training. Therefore, Vessel-Net has four outputs, and hence the total loss can be defined as follows:

$$Loss = CE(P, GT) + \lambda_2 \times CE(rP, GT) + \sum_{i=3}^{4} \lambda_i \times CE(msP_i, msGT_i) \quad (1)$$

where CE represents the categorical cross-entropy function, P is the vessel probability map generated by the decoder, GT is the ground truth, $msGT_3$ and $msGT_4$ denote the ground truth of size 12×12 and 24×24, respectively, and the weight parameters $\lambda_2 \sim \lambda_4$ represent the balance among four outputs. The $msGT_3$ and $msGT_4$ were obtained by down-sampling GT via 2×2 and 4×4 max-pooling, respectively.

Consequently, the proposed Vessel-Net is trained under the main supervision provided by the decoder, the richer feature supervision, and two types of multi-scale supervision. Thus, the complementary information acquired by five IR blocks can be combined to explore the effective representations of vessels of variable scales/thicknesses.

3 Experiments

3.1 Database

We evaluated the proposed Vessel-Net on the DRIVE database and CHASE database, which contain 40 color fundus retinal images of size 584×565 and 28 color fundus retinal images of size 999×960, respectively. The DRIVE database is officially split into two equal subsets for training and testing. In both databases, each retinal image is equipped with two manual annotations. To make fair comparative evaluation, we adopted the following settings in the literature: (1) using the first manual annotation as the ground truth and the second one as a human observer's segmentation [9]; (2) splitting the CHASE database into a training set of 20 images and a testing set of 8 images [8,16]; and (3) generating manually the field of view (FOV) mask for each image in the CHASE database [8,16,17].

3.2 Implementation Details

Fundus images were pre-processed by using the CLAHE [18], gamma adjusting and database normalization algorithms to reduce noise and improve contrast. In the training stage, we first randomly extracted 24.5K partly overlapped 48×48 patches in each pre-processed fundus image, resulting in a training set of 490K patches on each database. Then, we adopted the mini-batch stochastic gradient descent (mini-batch SGD) with a batch size of 32 as the optimizer, and set empirically the weight parameters $\lambda_2 \sim \lambda_4$ to 1, 1/3 and 2/3, respectively, the learning rate to 0.01, and the maximum epoch number to 150. Note that we used the same settings of hyper-parameters on both databases to demonstrate the robustness of our proposed Vessel-Net.

In the testing stage, we extracted 48×48 patches with a stride of 5 along both horizontal and vertical directions, and then fed each of them to the trained Vessel-Net. To recompose the entire images, we averaged the obtained probability maps of partly overlapped patches. Finally, we applied the threshold of 0.5 to the recomposed vessel probability map to generate the binary segmentation result.

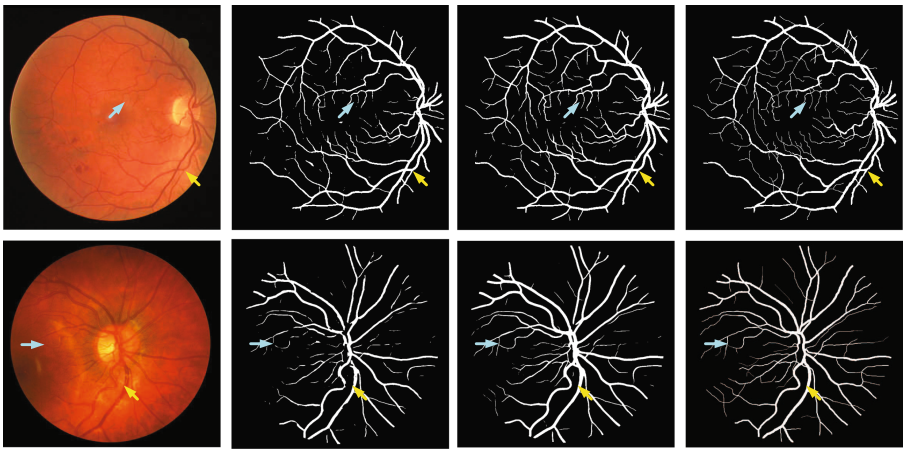

Fig. 3. Two examples of retinal images (1st column) from the DRIVE (top row) and CHASE (bottom row) databases, the segmentation results of U-Net (2nd column) and the proposed Vessel-Net (3rd column), and the ground truth (4th column). The samples of thin and thick vessels are pointed by blue and yellow arrows, respectively. (Color figure online)

4 Results

Figure 3 shows two retinal image examples from the DRIVE and CHASE databases, respectively, the corresponding segmentation results obtained by U-Net and the proposed Vessel-Net, and the ground truth. It reveals that Vessel-Net is able to delineate most thin and thick vessels and preserve more spatial structures of retinal vessels than U-Net (pointed by blue and yellow arrows). Such ability is essential for further topology estimation and reconstructions.

The retinal vessel segmentation results can be evaluated quantitatively by accuracy (ACC), specificity (SP), sensitivity (SE), and AUC. Tables 1 and 2 give the performance of the second human observer, several state-of-the-art methods, and the proposed Vessel-Net on the DRIVE and CHASE databases. Inter-observer variation is also included by assessing the second observer's annotations against the ground truth (first observer). It shows that our Vessel-Net achieves the highest AUC (i.e. 0.14%/0.35% higher than the second best), the highest accuracy (0.11%/0.24% higher than the second best), top-two sensitivity, and comparable specificity on both databases. It also shows that, compared to the inter-observer variation, our Vessel-Net produces a lower variation from the ground truth. This implies that our method can be used to provide more standardized segmentation of vessel images than manual processing. Furthermore, we suggest that, for potential clinical applications, our Vessel-Net can be combined with other post-processing operators [7,8] to further improve the segmentation of fine structures.

Table 1. Comparison with state-of-the-art methods and 2nd observer on DRIVE.

Method	AUC (%)	Accuracy (%)	Specificity (%)	Sensitivity (%)
2nd observer	N.A	94.72	97.24	77.60
Fu et al. [7]	N.A	95.23	N.A	76.03
Liskowski et al. [4]	97.90	95.35	98.07	78.11
Li et al. [16]	97.38	95.27	98.16	75.69
Orlando et al. [19]	95.07	N.A	96.84	78.97
Yan et al. [20]	97.52	95.42	98.18	76.53
Zhang et al. [6]	97.99	95.04	96.18	**87.23**
Wu et al. [8]	98.07	95.67	**98.19**	78.44
Vessel-Net (ours)	***98.21***	***95.78***	*98.02*	*80.38*

Table 2. Comparison with state-of-the-art methods and 2nd observer on CHASE.

Method	AUC (%)	Accuracy (%)	Specificity (%)	Sensitivity (%)
2nd observer	N.A	95.45	97.11	81.05
Li et al. [16]	97.16	95.81	97.93	75.07
Orlando et al. [19]	95.24	N.A	97.12	72.77
Yan et al. [20]	97.81	96.10	98.09	76.33
Wu et al. [8]	98.25	96.37	**98.47**	75.38
Vessel-Net (ours)	***98.60***	***96.61***	*98.14*	***81.32***

The ablation experiments were conducted on both databases to demonstrate the performance gain caused by each component of the proposed Vessel-Net. Table 3 shows, from top to bottom, the performance of baseline U-Net, R2U-Net, the U-Net with traditional Inception block, the Vessel-Net without using multi-path supervision (i.e. w/o MP), the Vessel-Net without IR blocks (w/o IR), and

Table 3. Ablation studies using the same experiment settings on both databases. (*The results of U-Net are obtained from [11])

Database	DRIVE(%)				CHASE(%)			
Indicator	AUC	ACC	SP	SE	AUC	ACC	SP	SE
U-Net [12]*	97.55	95.31	98.20	75.37	97.72	95.78	97.01	**82.88**
R2U-Net [11]	97.84	95.56	98.13	77.92	98.15	96.34	**98.20**	77.56
U-Net with Inception	97.76	95.50	97.94	78.73	98.14	96.23	97.87	79.82
Vessel-Net w/o MP	98.15	95.74	98.24	78.65	98.52	96.55	98.20	80.09
Vessel-Net w/o IR	98.18	95.74	**98.37**	77.73	98.49	96.51	98.13	80.32
Vessel-Net (ours)	**98.21**	**95.78**	98.02	**80.38**	**98.60**	**96.61**	98.14	81.32

the proposed Vessel-Net. It reveals that, on both databases, (1) using both IR blocks and multi-path supervision results in an AUC gain of 0.66% and 0.88%; (2) using IR blocks to replace inception blocks results in an AUC gain of 0.39% and 0.38%; (3) using multi-path supervision (IR blocks are available) further results in an AUC gain of 0.06% and 0.08%; and (4) using IR blocks (multi-path supervision is available) further results in an AUC gain of 0.03% and 0.11%.

5 Conclusion

In this paper, we present the Vessel-Net, a novel U-like deep convolutional neural network, for retinal vessel segmentation in color fundus images. The newly designed IR blocks and multi-path supervision are highly effective in capturing rich multi-scale information. Our results on the DRIVE and CHASE databases suggest that the proposed Vessel-Net achieved the new state-of-the-art performance.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China under Grants 61771397, in part by the Science and Technology Innovation Committee of Shenzhen Municipality, China, under Grants JCYJ20180306171334997, in part by the Seed Foundation of Innovation and Creation for Graduate Students in NPU under Grants ZZ2019029, and in part by the Project for Graduate Innovation team of NPU.

References

1. Fraz, M.M., et al.: Blood vessel segmentation methodologies in retinal images - a survey. Comput. Methods Progr. Biomed. **108**(1), 407–433 (2012)
2. Lupascu, C.A., Tegolo, D., Trucco, E.: FABC: retinal vessel segmentation using AdaBoost. IEEE Trans. Inf Technol. Biomed. **14**(5), 1267–1274 (2010)
3. Azzopardi, G., Strisciuglio, N., Vento, M., Petkov, N.: Trainable COSFIRE filters for vessel delineation with application to retinal images. Med. Image Anal. **19**(1), 46–57 (2015)
4. Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels with deep neural networks. IEEE Trans. Med. Imaging **35**(11), 2369–2380 (2016)
5. Fan, Z., et al.: A hierarchical image matting model for blood vessel segmentation in fundus images. IEEE Trans. Image Process. **28**(5), 2367–2377 (2019)
6. Zhang, Y., Chung, A.C.S.: Deep supervision with additional labels for retinal vessel segmentation task. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 83–91. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_10
7. 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
8. Wu, Y., Xia, Y., Song, Y., Zhang, Y., Cai, W.: Multiscale network followed network model for retinal vessel segmentation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 119–126. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_14

9. Staal, J., Abr'amoff, 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)
10. Fraz, M.M., et al.: An ensemble classification-based approach applied to retinal blood vessel segmentation. IEEE Trans. Biomed. Eng. **59**(9), 2538–2548 (2012)
11. Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K.: Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation. arXiv preprint arXiv:1802.06955 (2018)
12. 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
13. Szegedy, C., et al.: Going deeper with convolutions. In: CVPR 2015, pp. 1–9 (2015)
14. Liu, Y., Cheng, M.M., Hu, X., Wang, K., Bai, X.: Richer convolutional features for edge detection. In: CVPR 2017, pp. 3000–3009 (2017)
15. Zhang, J., Tao, D.: FAMED-Net: a fast and accurate multi-scale end-to-end dehazing network. arXiv preprint arXiv:1906.04334 (2019)
16. Li, Q., et al.: A cross-modality learning approach for vessel segmentation in retinal images. IEEE Trans. Med. Imaging **35**(1), 109–118 (2016)
17. Soares, J.V.B., Leandro, J.J.G., Cesar, R.M., Jelinek, H.F., Cree, M.J.: Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. IEEE Trans. Med. Imaging **25**(9), 1214–1222 (2006)
18. Setiawan, A.W., Mengko, T.R., Santoso, O.S., Suksmono, A.B.: Color retinal image enhancement using CLAHE. In: ICISS, pp. 1–3 (2013)
19. Orlando, J., Prokofyeva, E., Blaschko, M.: A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. IEEE Trans. Biomed. Eng. **64**(1), 16–27 (2017)
20. 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**(9), 1912–1923 (2018)