M2E-Net: Multiscale Morphological Enhancement Network for Retinal Vessel Segmentation

Le Geng[1], Panming Li[1], Weifang Zhu[1], and Xinjian Chen[1,2(✉)]

[1] School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China
xjchen@suda.edu.cn

[2] State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou 215123, China

Abstract. Accurate retinal vessel segmentation plays a critical role in the early diagnosis of diabetic retinopathy. The uneven background distribution and complex morphological structure are still the challenges for segmenting the vessels, especially the capillaries. In this paper, we propose a novel deep network called Multiscale Morphological Enhancement Network (M2E-Net). First, we develop a multiscale morphological enhancement block, specifically the Top-hat/Bottom-hat module. Then, the block is embedded in the shallow layers of the U-shape fully convolutional network, which aims to reduce the influence of the uneven background distribution and enhance the texture contrast of fundus images. It is the first time introducing explainable morphological operation into CNN. Furthermore, the Structural Similarity (SSIM) loss guides the network to pay more attention to predicting the fine structures and clear boundaries during training. The proposed M2E-Net is evaluated on two digital retinal databases, DRIVE and CHASEDB1, and achieves the state-of-the-art performance with AUC of 98.40% and 98.94% respectively. The experimental results demonstrate the effectiveness of our proposed method.

Keywords: Retinal vessel segmentation · Multiscale morphological enhancement · Structural similarity loss

1 Introduction

Retinal vessel segmentation in fundus images is of great value in the diagnosis, screening, treatment, and evaluation of various cardiovascular and ophthalmic diseases, such as diabetes, hypertension, arteriosclerosis, and choroidal neovascularization [18]. However, due to the uneven background distribution and complex morphological structure, accurate vessel segmentation is a remaining challenge.

Y. Peng et al. (Eds.): PRCV 2020, LNCS 12305, pp. 493–502, 2020.
https://doi.org/10.1007/978-3-030-60633-6_41

Recent retinal vessel segmentation algorithms published in the literature can be divided into two categories. The first category of methods is generally following traditional image processing approaches, including filter-based and model-based techniques [15]. For example, different types of filters are jointly employed to extract 41-dimensional visual features to represent retinal vessels for subsequent classification [7]. To suppress background structure and image noise, various filtering methods have been proposed, such as Hessian matrix-based filters [4] and symmetry filter [17]. Another category of methods is based on deep learning (DL), which has shown its advantages in many computer vision tasks, including retinal vessel segmentation. Fu et al. [6] proposed a fully convolutional neural network combined with a fully-connected Conditional Random Fields (CRFs). Wu et al. [13] designed an efficient inception-residual convolutional block for improved feature representation. Lei et al. [8] combined spatial attention and channel attention to further integrate local features with their global dependencies adaptively.

Recently, attention mechanisms are becoming prevailing in deep learning. In medical image segmentation, structural information is the most valuable and of great significance, such as edges, shapes, etc. In ET-Net, the authors embedded edge-attention representations to guide the segmentation network to focus on edge information [16]. Yan et al. [15] make the network increase the attention to the edge by polishing the loss function.

Standing on the shoulders of these studies, in this paper, we propose the Multiscale Morphological Enhancement network (M2E-Net) which aims to improve the performance of retinal vessel segmentation. To reduce the effect of unevenly distributed background and improve the contrast of capillaries in fundus images, we designed a multiscale morphological enhancement block. It follows the principle of top-hat operation, which is widely used in traditional algorithms. To the best of our knowledge, it is the first time that the theory of morphological operation is introduced into the convolutional neural network. We embed three multiscale morphological enhancement (M2E) blocks into the shallow layers of the baseline modified from LinkNet [2]. To predict fine structures and clear boundaries, we introduce the multiscale SSIM loss into the hybrid loss function to capture the structural characteristics of vessels. Experimental results on two digital retinal databases, DRIVE and CHASEDB1, show significant performance improvements with AUC of 98.40% and 98.94% respectively.

2 Method

The overview of the proposed retinal vessel segmentation framework M2E-Net is shown in Fig. 1, which is an improved U-shape encoder-decoder architecture. Specifically, the residual network (ResNet) block is employed as the backbone for each encoder block. The multiscale morphological enhancement blocks are inserted in the shallow layers of the encoder to reduce the influence of the uneven background distribution and enhance the texture contrast. Then the features from the encoder are added to the decoder path via skip connections to achieve

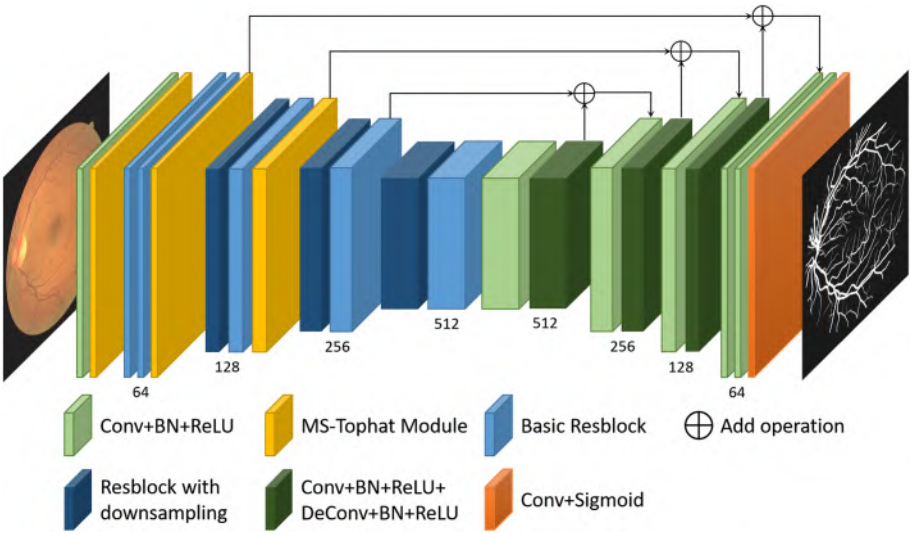

Fig. 1. The architecture of the multiscale morphological enhancement network: M2E-Net.

rich representation for better segmentation. Finally, the hybrid loss consisting of binary cross-entropy, dice loss, and multiscale structural similarity loss is employed to guarantee better performance.

2.1 Multiscale Morphological Enhancement Block

Motivation. When performing retinal vessel segmentation based on traditional algorithms, for the problems of inconspicuous contrast (such as capillaries) and uneven illumination and background, top-hat or bottom-hat transformation is a fast and efficient enhancement method. Inspired by this, we hope to embed such a trainable module in a network to achieve a similar effect during the training. In addition, with the research on the interpretability of deep learning, we can easily treat shallow convolution operations as some elementary filtering operations, such as low-pass filtering, high-pass filtering, and edge detection. However, the mathematical morphology operations which are common and efficient in image processing have not been applied or found in trained models. We consider this to be one of the gaps between traditional image processing algorithms and data-driven deep learning algorithms.

Implementation. We design a module based on top-hat transformation, which can extract the bright detail features in the image. The top-hat transformation mainly includes dilation, erosion, and subtraction. According to the definition of morphological operations based on the maximum and minimum values, we use maxpooling with a stride of 1 to implement dilation. This operation is equivalent

to dilating the input image with a square structure element, whose size is the same as the kernel size of maxpooling. Similarly, we should use minpooling with a stride of 1 to achieve erosion. However, there is no minpooling in common deep learning frameworks, such as tensorflow, pytorch, and mxnet. We implement the equivalent minpooling for erosion operation by following steps: (1) reverse the original image by multiplying -1; (2) perform dilation; (3) reverse the dilation result. Then, we can easily implement a top-hat module with a specified kernel size of N using the module designed above.

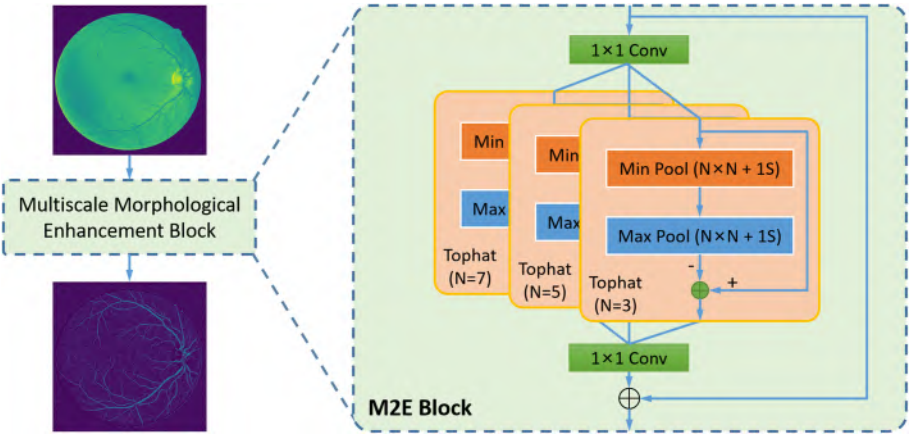

Fig. 2. The illustrations of proposed Multiscale Morphological Enhancement Block.

The top-hat module can only be used to extract bright details, but dark details are also significant features. We add the control factor before the top-hat module, which is implemented by a 1×1 convolution. It can automatically adjust and extract valuable light or dark details and suppress unnecessary details. To highlight the details of different scales, the outputs of different kernel size top-hat modules are fused via a 1×1 convolution, which also adjusts the fused result to the same size as the input. Finally, the input and the result after adaptive fusion are added as the overall output of the enhancement module. The whole structure of the M2E block is shown in Fig. 2. In order to ensure compatibility, the M2E block is entirely implemented by existing operators in the deep learning framework.

Contribution. Figure 2 shows that after the image or feature map is processed by the M2E block, the contrast between capillaries and the background is enhanced. It can also learn adaptive features which is more advanced than traditional morphological operations. The detailed features on the image or feature map mainly exist in the shallow layers of the neural network, so we embed it in the shallow layers of the convolutional network, specifically after the first

three blocks of Resnet in Fig. 1. We also tried to embed the block deeper, and the experimental results show that the performance of the network will be degraded. To the best of our knowledge, this is the first time introducing an explainable morphological operation into the CNN.

2.2 Structural Similarity Loss

When designing a medical image segmentation algorithm based on deep learning, we pay more attention to pixel-wise segmentation, which can guarantee the convergence of the model. There are still some limitations, for example, models trained with BCE loss usually have low confidence in classifying boundary pixels, leading to blurry boundaries. In [3,16], edge loss is utilized to increase attention to edges, but it is still calculated at pixel-level, and the structural relevance is not fully considered. In the area of image quality assessment, structure similarity is a common evaluation index, which can reflect the structural similarity between two images [12].

Inspired by the design of dice loss, we convert the structural similarity, originally proposed for image quality assessment, into a loss function. We integrate it into our hybrid loss to learn the structural information of the vessel. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two corresponding patches, with the same size $N \times N$, cropped from the predicted probability map S and the binary ground truth mask G respectively. The SSIM loss is defined as:

$$L_{ssim} = 1 - \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{1}$$

where μ_x, μ_y and σ_x, σ_y are the mean and standard deviations of $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively, σ_{xy} is their covariance, C_1 and C_2 are used to avoid dividing by zero. The patch size N is a hyper-parameter for SSIM and it determines the receptive field. We apply multiscale SSIM loss here which is a combination of three SSIM losses with the N fixed to 5, 7, 11. BCE loss, multiscale SSIM loss, and dice loss are fused together equally as the final employed hybrid loss, which calculates loss from pixel-level, patch-level, and image-level respectively. Hence, the trained model could obtain high-quality regional segmentation and clear boundaries.

$$L_{hybrid} = L_{bce} + L_{ssim} + L_{dice} \tag{2}$$

In addition, SSIM loss has a friendly side effect that it introduces edge attention, which is especially suitable for the segmentation of slender targets such as vessels. It considers a local neighborhood of each pixel and assigns higher weights to the boundary, i.e., the loss is higher around the boundary, even when the predicted probabilities on the boundary and the rest of the foreground are the same.

As shown in Fig. 3, we enumerate three different situations in the training process to compare the changes in the loss of BCE and SSIM at different pixel positions. The first row is the heatmap of BCE loss, and the second row is that of SSIM loss. These three columns represent three different situations. The

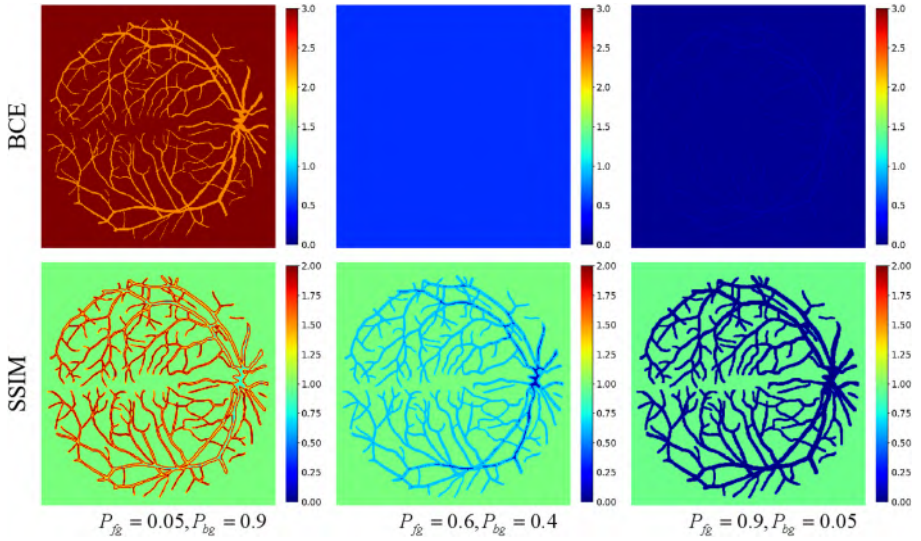

Fig. 3. The illustrations of the impact of losses. P_{fg} and P_{bg} denote the predicted probability of the foreground and background, respectively

heatmaps indicate that the BCE loss treats pixels at any position equally. For SSIM, even when the predicted probability on the boundary is the same as the probability of other pixel positions in the foreground, the loss of the boundary pixel position is higher. This is equivalent to automatically adding the attention mechanism to the boundary, which helps to optimize the boundary. Compared with a simple convex target, the edge pixels in the slender vessels occupy a larger proportion of the target pixels. Therefore, the characteristics of SSIM make it very suitable as a loss function for complex structures.

3 Experiments

3.1 Database

Two public databases DRIVE [10] and CHASEDB1 [5] are used to evaluate the proposed M2E-Net. The DRIVE database includes 40 color fundus images divided into two parts: 20 training images and 20 testing images. All fundus images have the same resolution of 565×584. Each image in the testing set has two manual labels, while each image in the training set has only one manual label. We use the labels provided by the first observer as the ground truth for performance evaluation.

The CHASEDB1 database has 28 color images of the retina and the resolution of each image is 999×960, which are taken from both eyes of 14 school children. Usually, the first 20 images are used for training, and the rest are used for testing, as described in [15]. The binary field of view (FOV) mask and segmentation ground truth is obtained by manual methods.

3.2 Implementation Details

Our proposed M2E-Net was implemented based on the PyTorch library (version 1.1.0). The network was trained on one NVIDIA K400 GPU which has a memory about 12 GB. Adam optimizer was applied to optimize the whole network and the learning rate was fixed to 2e−4. To improve the generalization capabilities of the model, we applied random data augmentation before training, including adjusting brightness, color, contrast, sharpness, and rotating the images. The enhancement factors were all following the log-normal distribution. After data enhancement, z-score processing was performed for normalizing each channel of images. In the testing stage, the data augmentation was removed while the z-score processing was still retained. During the training period, we used a hybrid loss function that is a weighted sum of three items to optimize the model together.

4 Results

Figure 4 shows two retinal fundus images and their ground truth, together with the segmentation results obtained by UNet and the proposed M2E-Net. The example from the DRIVE database is shown in the first row, and the one from the CHASEDB1 is shown in the second row. Due to the proposed enhancement block and SSIM loss, it can be observed that M2E-Net produces more accurate and detailed results than UNet. Vessels at low contrast can also be segmented, which is helpful for medical analysis and diagnosis.

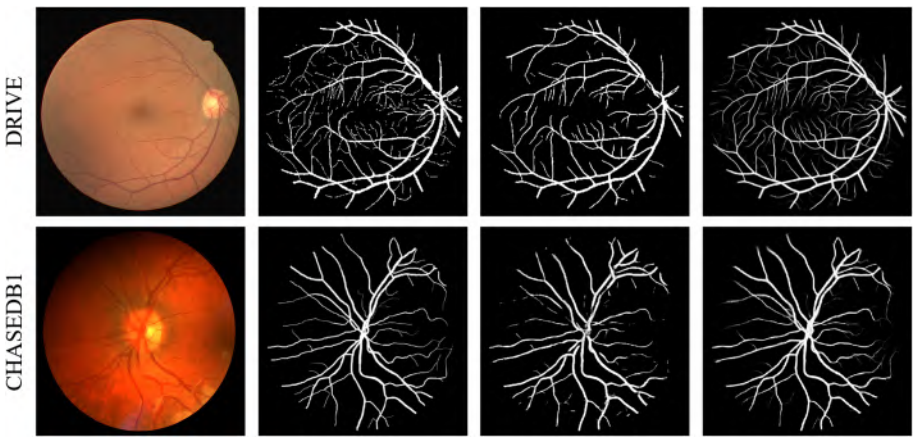

Fig. 4. Two test images from the DRIVE (1st row) and CHASEDB1 (2nd row) databases, ground truth (2nd column), the U-Net (3rd column), and the proposed M2E-Net (4th column).

Table 1. Results of M2E-Net and other methods on DRIVE database.

Methods	Acc	AUC	Sens	Spec
U-Net (2015) [9]	0.9531	0.9601	0.7537	0.9639
R2U-Net (2018) [1]	0.9556	0.9784	0.7792	0.9813
DE U-Net (2019) [11]	0.9567	0.9772	0.7940	0.9816
ACE-Net (2019) [18]	0.9569	0.9742	0.7725	0.9842
Vessel-Net (2019) [13]	0.9578	0.9821	0.8038	0.9802
Proposed method	**0.9682**	**0.9840**	**0.8149**	**0.9846**

Table 2. Results of M2E-Net and other methods on CHASEDB1 database.

Methods	Acc	AUC	Sens	Spec
R2U-Net (2018) [1]	0.9634	0.9815	0.7756	0.9712
MS-NFN (2018) [14]	0.9637	0.9825	0.7538	**0.9847**
LadderNet (2018) [19]	0.9656	0.9839	0.7978	0.9818
DE U-Net (2019) [11]	0.9661	0.9812	0.8074	0.9821
Vessel-Net (2019) [13]	0.9661	0.9860	0.8132	0.9814
Proposed method	**0.9751**	**0.9894**	**0.8510**	0.9835

As for the evaluation metrics, accuracy, sensitivity, and specificity are used to evaluate the performance of M2E-Net. To further evaluate the performance of different neural networks, we also calculated the area under the receiver operating characteristics curve (AUC). Table 1 and 2 give the quantitative evaluation results of several recently proposed state-of-the-art methods, and the proposed M2E-Net on the DRIVE and CHASEDB1 database respectively. The experimental result shows that the proposed model outperforms other models by achieving the highest AUC, accuracy, and sensitivity for each database.

Table 3. Ablation studies using the same experiment settings on DRIVE database.

Methods	Acc	AUC	Sens	Spec
Baseline	0.9518	0.9684	0.8133	0.9651
Baseline + M2E (S)	0.9665	0.9830	**0.8568**	0.9813
Baseline + M2E (E)	0.9615	0.9819	0.8136	0.9732
Baseline + MS-SSIM	0.9677	0.9834	0.8290	0.9832
M2E-Net	**0.9682**	**0.9840**	0.8149	**0.9846**

In order to further understand the performance gain of the proposed block and the effect of the SSIM loss, we conducted the ablation experiments on the

DRIVE database. Table 3 shows, from top to bottom, the performance of baseline, baseline with M2E blocks embedded in shallow layers (M2E-Net without MS-SSIM loss), baseline with M2E blocks embedded entirely, baseline trained with multiscale SSIM loss and M2E-Net trained with multiscale SSIM loss (proposed method). It reveals that: (1) the multiscale morphological enhancement block leads to improvement; (2) the block is more suitable for shallow layers; (3) multiscale SSIM loss helps boost the performance. When the M2E block is embedded entirely from shallow to deep layers, the performance of the model declines instead. This is because the M2E module is originally designed to extract detailed features from images or feature map, which only exist in the shallow layer of the model. The application of morphological operation in deep layers will destroy the semantic information and lead to poor performance.

5 Discussion and Conclusions

In this paper, we propose a Multiscale Morphological Enhancement Network (M2E-Net) for retinal vessel segmentation. The newly designed M2E block is effective in extracting useful detailed information and reducing the influence of the uneven background distribution. The multiscale SSIM loss, increasing edge attention during training, also boosts the segmentation performance. Experimental results on DRIVE and CHASEDB1 indicate that the proposed method has significantly outperformed the state-of-the-art for retinal vessel segmentation.

References

1. 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)
2. Chaurasia, A., Culurciello, E.: LinkNet: exploiting encoder representations for efficient semantic segmentation. In: 2017 IEEE Visual Communications and Image Processing (VCIP), pp. 1–4. IEEE (2017)
3. Fang, Y., Chen, C., Yuan, Y., Tong, K.: Selective feature aggregation network with area-boundary constraints for polyp segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 302–310. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_34
4. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A., Delp, S. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998). https://doi.org/10.1007/BFb0056195
5. 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)
6. 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
7. Lupascu, C.A., Tegolo, D., Trucco, E.: FABC: retinal vessel segmentation using AdaBoost. IEEE Trans. Inf Technol. Biomed. **14**(5), 1267–1274 (2010)

8. Mou, L., et al.: CS-Net: channel and spatial attention network for curvilinear structure segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 721–730. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_80
9. 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
10. Staal, J., Abràmoff, 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)
11. 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
12. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
13. Wu, Y., et al.: Vessel-Net: retinal vessel segmentation under multi-path supervision. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 264–272. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_30
14. 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
15. 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)
16. Zhang, Z., Fu, H., Dai, H., Shen, J., Pang, Y., Shao, L.: ET-Net: a generic edge-attention guidance network for medical image segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 442–450. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_49
17. Zhao, Y., et al.: Automatic 2-D/3-D vessel enhancement in multiple modality images using a weighted symmetry filter. IEEE Trans. Med. Imaging **37**(2), 438–450 (2017)
18. Zhu, Y., Chen, Z., Zhao, S., Xie, H., Guo, W., Zhang, Y.: ACE-Net: biomedical image segmentation with augmented contracting and expansive paths. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 712–720. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_79
19. Zhuang, J.: LadderNet: multi-path networks based on U-Net for medical image segmentation. arXiv preprint arXiv:1810.07810 (2018)