Dual Attention Multiscale Network for Vessel Segmentation in Fundus Photography

Pengshuai Yin
School of Future Technology, South China University of Technology

Yupeng Fang
School of Software Engineering, South China University of Technology

Qingyao Wu
School of Software Engineering, South China University of Technology

QiLin Wan (wanqilin@gbamwf.com)
Guangdong-Hong Kong-Macao Greater Bay Area Weather Research Center for Monitoring Warning and Forecasting

Research Article

Keywords: vessel segmentation, fundus image processing, attention network

Posted Date: December 16th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1170254/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Dual Attention Multiscale Network for Vessel Segmentation in Fundus Photography

Pengshuai Yin¹, Yupeng Fang², Qingyao Wu² and QiLin Wan³*

Abstract

Background: Automatic vessel structure segmentation is an essential step towards an automatic disease diagnosis system. The task is challenging due to the variance shapes and sizes of vessels across populations.

Methods: A multiscale network with dual attention is proposed to segment vessels in different sizes. The network injects spatial attention module and channel attention module on feature map which size is \( \frac{1}{8} \) of the input size. The network also uses multiscale input to receive multi-level information, and the network uses the multiscale output to gain more supervision.

Results: The proposed method is tested on two publicly available datasets: DRIVE and CHASEDB1. The accuracy, AUC, sensitivity, specificity on DRIVE dataset is 0.9615, 0.9866, 0.7693, and 0.9851, respectively. On the CHASEDB1 dataset, the metrics are 0.9797, 0.9895, 0.8432, and 0.9863 respectively. The ablative study further shows effectiveness for each part of the network.

Conclusions: Multiscale and dual attention mechanism both improves the performance. The proposed architecture is simple and effective. The inference time is 12ms on a GPU and has potential for real-world applications. The code will be made publicly available.

Keywords: vessel segmentation; fundus image processing; attention network

Background

The segmentation of vasculature in retinal images is important in aiding the management of many diseases, such as diabetes and hypertension. Diabetic Retinopathy (DR) is caused by high blood sugar levels and results in the swelling of the retinal vessels [1]. Hypertensive Retinopathy (HR) is caused by high blood pressure and results in the narrowing of vessels or increased vascular tortuosity [2]. The early diagnosis of pathological diseases often helps patients to receive timely treatment. However, manually labeling vessel structures is time-consuming, tedious, and subject to human error. Automated segmentation of retinal vessels is in high demand and can release the intense burden of skilled staff.

The retinal blood vessels structure is extremely complicated together with high tortuosity and various shapes, such as angles, branching patterns, length, and width [3]. The high anatomical variability and varying vessel scales across populations make the task challenging. Furthermore, the noise and poor contrast company by the low resolutions further increase the difficulty. Traditional vessel segmentation methods often cannot robustly segment all vessels of interest.

Deep learning methods show impressive performance on image segmentation. The most widely used architecture is U-Net [4]. The coarse-to-fine feature representation
learned by U-Net is suitable to gain satisfactory performance on a small dataset. The attention U-Net can further improve the performance [5]. The attention module automatically learns to focus on interest vessels with varying shapes while preserving computational efficiency. DA-Net [6] proposes spatial attention module and channel attention module for natural scene parsing. The spatial attention module and the channel attention module utilize a self-attention mechanism to capture features dependencies in the spatial and channel dimensions respectively. The spatial attention module aggregates feature at all positions with weighted summation and the channel attention module captures the channel dependencies between any two channel maps. Motivated by the DA-Net, this paper design a dual attention multiscale network for vessel segmentation.

Related Works
In this section, we explain some most commonly used attention mechanisms and multiscale networks for vessel segmentation.

Attention Network
Attention U-Net [5] captures a sufficiently large receptive field to collect semantic contextual information and integrate attention gates to reduce false-positive predictions for small objects that show large shape variability. Ni et.al [7] proposes a global channel attention module for vessel segmentation that emphasizes the inheritance relationship of the entire network feature. CS-Net [8] integrates channel attention and spatial attention into U-Net for 2D and 3D vessel segmentation. Hao et.al [9] exploits contextual frames of sequential images in a sliding window centered at the current frame and equipped with a channel attention mechanism in the decoder stage. Li et.al [10] proposes an attention gate to highlight salient features that are passed through the skip connections. HANet [11] automatically focuses the network’s attention on regions which are “hard” to segment. The vessel regions which are “hard” or “easy” are based on the coarse segmentation probabilistic map.

Multiscale Network
Yue et.al [12] utilizes different scale image patches as inputs to learn richer multiscale information. Roberto et.al [13] proposes a multiple-scale Hessian approach to enhance the vessels followed by thresholding. Wu et.al [14] generate multiscale feature maps by max-pooling layers and up-sampling layers. The first multiscale network converts an image patch into a probabilistic retinal vessel map, and the following multiscale network further refines the map. Yin et.al [15] proposed to utilize multiscale input to fuse multi-level information.

Different from all the above methods, the proposed method takes advantage of both multiscale and attention mechanisms.

Results
Data Preparation
We conduct experiments on two datasets: DRIVE and CHASEDB1.

**DRIVE:** The Digital Retinal Images for Vessel Extraction (DRIVE) is a dataset for retinal vessel segmentation, which consists of 40 color fundus images of size
768 × 584 pixels, including 7 abnormal pathology cases. It was equally divided into 20 images for training and 20 images for testing along with two manual segmentations of the vessels. The first segmentation is accepted as the ground-truth for performance evaluation while the second segmentation is accepted as a human observer reference for performance comparison. The images are captured in digital form from a Canon CR5 non-mydriatic 3CCD camera at 45° field of view (FOV).

CHASEDB1: The CHASEDB1 dataset [16] for retinal vessel segmentation which consists of 28 color retina images of size 960 × 999 pixels are collected from both left and right eyes of 14 school children. These images are captured by a handheld Nidek NM-200-D fundus camera at 30° field of view and each image is annotated by two independent human experts. We select the first 20 images for training and the rest 8 images for testing [17].

Evaluation Criteria
The vessel segmentation process is a pixel-based classification, with each pixel being classified as a vessel or surrounding tissue. We employ 4 indicators (Specificity (Spe), Sensitivity (Sen), Accuracy (Acc) and Area Under ROC (AUC)) to measure model performance. Specificity (Spe) is the ability to detect non-vessel pixels. Accuracy (Acc) is measured by the ratio of the total number of correctly classified pixels (the sum of true positives and true negatives) to the number of pixels in the image field of view (FOV):

\[
Acc = \frac{TP + TN}{TP + FN + TN + FP}.
\]

\[
Sen = \frac{TP}{TP + FN}.
\]

\[
Spe = \frac{TN}{TN + FP}.
\]

Here, TP(true positive) is where a pixel is identified as the vessel in both the segmented image and ground truth. TN(true negative) is where a non-vessel pixel of the ground truth is correctly classified in the segmented image.

Implementation Details
We set the learning rate at 0.001 decayed by a factor of 10 every 50 epochs. The network was trained for 300 epochs from scratch on an NVIDIA GeForce RTX 3090 Ti GPU. The input images of the neural network are resized to 512 × 512. In order to improve the generalization ability of the network, we also use several data enhancement techniques, including random horizontal flip with a probability of 0.5, random rotation in [−20°, 20°], and gamma contrast enhancement in [0.5, 2].

Performance Evaluation
In this section, we compared our method with other state-of-the-art methods on DRIVE and CHASEDB1 datasets. The methods include U-Net [4], Zhang et al. [18], Liskowski et al. [19], DRIU [20], Yan et al. [21], CE-Net [22], LadderNet [23], DUNet [24], Bo Liu et al. [25], VesselNet [26], DA-Net [6], Yin et al. [15] and CS-Net [8].
Table 1 Segmentation performance for DRIVE inside FOV.

| Method            | Year | Acc   | AUC   | Sen   | Spe   |
|-------------------|------|-------|-------|-------|-------|
| Human Observer    | -    | 0.9578| N.A   | 0.8288| 0.9701|
| U-Net [4]         | 2015 | 0.9531| 0.9755| 0.7537| 0.9820|
| Zhang et al. [18] | 2016 | 0.9476| 0.9636| 0.7743| 0.9725|
| Liskowski et al. [19]| 2016 | 0.9542| 0.9752| 0.7653| 0.9818|
| DRU [20]          | 2016 | 0.9541| 0.9801| 0.8280| 0.9728|
| Yan et al. [21]   | 2018 | 0.9542| 0.9752| 0.7653| 0.9818|
| Wu et al. [14]    | 2018 | 0.9578| 0.9821| 0.8038| 0.9802|
| CE-Net [22]       | 2019 | 0.9545| 0.9779| 0.8309| 0.9816|
| Liskowski et al. [19]| 2016 | 0.9535| 0.9823| 0.7816| 0.9836|
| YU-Net [24]       | 2019 | 0.9615| 0.9808| 0.8075| 0.9841|
| Bo Liu et al. [25]| 2019 | 0.9559| 0.9779| 0.8072| 0.9780|
| VesselNet [26]    | 2019 | 0.9578| 0.9821| 0.8038| 0.9802|
| Yue et al. [12]   | 2019 | 0.9561| 0.9796| 0.8199| 0.9762|
| DA-Net [6]        | 2019 | 0.9615| 0.9808| 0.8075| 0.9841|
| Yin et al. [15]   | 2020 | 0.9604| 0.9846| 0.7614| 0.9837|
| CS-Net [8]        | 2021 | 0.9632| 0.9825| 0.8216| 0.9863|
| Ours              | -    | 0.9615| 0.9866| 0.7693| 0.9851|

Table 1 shows the performance on the DRIVE dataset. Figure 1 shows the prediction of the proposed method on DRIVE dataset. The proposed method achieves the highest AUC compared to other methods. CS-Net inserts the attention module into the branch which the size of the feature map equal to the $\frac{1}{16}$ of the original size. Our proposed attention module gathers the information from the feature map with $\frac{1}{8}$ of the original image size. The input image size of CS-Net is $384 \times 384$ while the input image size of the proposed network is $512 \times 512$, our proposed method still shows higher AUC compared to CS-Net. Our method achieves higher performance compared to DA-Net, since DA-Net directly upamples the attention map as the output. The attention module used in our method is also different from CS-Net. We concatenate the spatial attention map, the channel attention map and the sum of these two maps together to form a more discriminate feature representation.

Table 2 shows the performance evaluation on CHASEDB1 dataset. Figure 2 shows the correspondense prediction. Our method surpasses all the other methods. The experimental results verify the effectiveness of the proposed method.

Table 2 Segmentation performance of CHASEDB1 inside FOV.

| Method            | Year | Acc   | AUC   | Sen   | Spe   |
|-------------------|------|-------|-------|-------|-------|
| Human Observer    | -    | 0.9545| N.A   | 0.8105| 0.9711|
| U-Net [4]         | 2015 | 0.9578| 0.9772| 0.8288| 0.9701|
| DRU [20]          | 2016 | 0.9657| 0.9746| 0.7651| 0.9822|
| Liskowski et al. [19]| 2016 | 0.9535| 0.9823| 0.7816| 0.9836|
| Yan et al. [21]   | 2018 | 0.9610| 0.9781| 0.7633| 0.9809|
| LadderNet [23]    | 2019 | 0.9656| 0.9839| 0.7978| 0.9818|
| DU-Net [24]       | 2019 | 0.9661| 0.9812| 0.8074| 0.9821|
| VesselNet [26]    | 2019 | 0.9661| 0.9860| 0.8132| 0.9814|
| Yin et al. [15]   | 2020 | 0.9783| 0.9869| 0.7993| 0.9868|
| Ours              | -    | 0.9797| 0.9895| 0.8432| 0.9863|

Ablative Studies
This section evaluates the performance of each part of the network. Table 3 and Table 4 shows the performance of the proposed network using different modules on DRIVE dataset and CHASEDB1 dataset respectively. The evaluation metric
Table 3  The performance for each part of DRIVE dataset inside FOV.

| Method                 | Acc   | AUC   | Sen   | Spe   | MIOU |
|------------------------|-------|-------|-------|-------|------|
| U-Net                  | 0.9531| 0.9755| 0.7537| 0.9820| -    |
| Multiscale Network     | 0.9614| 0.9863| 0.7668| 0.9851| 0.679|
| + Attention Module     | 0.9615| 0.9866| 0.7693| 0.9851| 0.681|

Table 4  The performance for each part of CHASEDB1 dataset inside FOV.

| Method                 | Acc   | AUC   | Sen   | Spe   | MIOU |
|------------------------|-------|-------|-------|-------|------|
| U-Net                  | 0.9578| 0.9772| 0.8288| 0.9701| -    |
| Multiscale Network     | 0.9796| 0.9893| 0.8340| 0.9862| 0.6527|
| + Attention Module     | 0.9797| 0.9895| 0.8432| 0.9863| 0.6563|

includes mean IOU commonly used in the semantic segmentation. The multiscale architecture significantly improves the performance compared to U-Net and the attention mechanism further improves the performance.

Discussion

The dual attention network combines spatial attention, channel attention with a multiscale network. In this section, We analyze the time efficiency and the parameters of the proposed model.

Time Efficiency  The inference time for AM-Net is 12ms on a 3090Ti Nvidia GPU for one image with a size of $512 \times 512$. The simple and effective architecture can be easily applied to smart AI applications.

Model Parameters  The proposed AM-Net has 9.95M parameters and 75.438G flops. The multiscale backbone needs 9.415M parameter and 73.163G flops. The dual attention module only needs 0.54M parameters.

The proposed network can perform fast inference and does not have many parameters. The simple and effective multiscale dual attention network has the potential to deploy to real-world applications.

Conclusion

In the paper, we propose a dual attention multiscale network. The network contains multiscale input, multiscale output and dual attention module. The dual attention module contains spatial attention and channel attention. The spatial attention gathers information from different positions and the channel attention module models the relationship between different channels. The experiments verify the effectiveness of the proposed method. The proposed framework enables fast inference and can be deployed to real-world applications.

Figure 1  The segmentation example of the proposed method on DRIVE dataset.

Figure 2  The segmentation example of the proposed method on CHASEDB1 dataset.

Methods

The network structure consists of multiscale input, dual attention module and multiscale output.
Multiscale Input
The multiscale input consists of an image with multiscale. The input image is down-scaled to \( \frac{1}{4} \), \( \frac{1}{8} \), and \( \frac{1}{8} \) of the origin size. Each image is sent to the corresponding encoder path. The multiscale input can gather different level information.

U-shape Architecture
The architecture of the proposed method is shown in Figure 3. Our network is constructed based on U-Net, and the input of the encoder path is an image pyramid. Two \( 3 \times 3 \) convolution layers are applied to the encoder path, then is a \( 2 \times 2 \) max-pooling operation with the element-wise rectified linear unit (ReLU) activation function to generate encoder feature maps. The feature map of the down-sample is connected to the feature map of the down-scaled input image. The number of feature maps is also doubled after down-sampling, enabling the architecture to learn complex structures efficiently. Like the encoder path, the decoder path produces a decoder feature map by using two \( 3 \times 3 \) convolution layers and a \( 2 \times 2 \) up-sampling layer, with the number of feature maps halved to preserve symmetry. The feature map of the encoder path is appended to the input of the corresponding decoder path by using skip connections. Finally, the high-dimensional feature representation of the output of the last decoder layer is fed to the dual-attention module to learn the relationship between position and channel, so that the feature representation with more discernible power can be sent to the multiscale output layer for final prediction.

Dual Attention Mechanism
The dual attention mechanism contains a spatial attention module and a channel attention module as shown in Figure 4.

Spatial Attention Module: The spatial attention module models rich contextual dependencies over feature maps by learning a spatial attention matrix, which represents the spatial relationships between the features of any two pixels. Different from DA-Net [6], we put the attention module in the branch with the feature map size equal to the \( \frac{1}{8} \) of the original image rather than directly resampling the attention map as the output. The design retains more detailed information without adding many parameters. Furthermore, the vessel segmentation requires skip-connection operations to fuse low-level information and recover the spatial information loss caused by down-sampling operations. The input feature representation \( S \in \mathbb{R}^{C \times H \times W} \) is feed into three convolution layers to generate three feature maps \( A, B \) and \( C \), where \( A, B, C \in \mathbb{R}^{C \times H \times W} \). The three feature maps are reshaped into \( C \times N \), where \( N = H \times W \) is the total pixel number. After that, the transpose of \( A \) and \( B \) is multiplied and followed by a softmax layer to form the spatial attention \( SA \in \mathbb{R}^{N \times N} \):
\[ SA_{ij} = \frac{\exp(A_i, B_j)}{\sum_{i=1}^{N} \exp(A_i, B_j)}. \]  

(4)

The spatial attention module \( SA \) represents the impact of position \( i \) on position \( j \). A similar feature representation will have a greater correlation. The transposition of \( SA \) is multiplied by \( C \) to form a feature representation and reshaped to \( R^{C \times H \times W} \). The result is multiplied by a scale parameter \( \alpha \) and followed by an element-wise summation with input feature \( F \) to generate the final spatial attention feature map \( SAO \):

\[ SAO_i = \alpha \sum_{j=1}^{N} (SA_{ij}C_j) + S_i. \]  

(5)

Here, \( \alpha \) is a learnable parameter and it is initialized to 0. The spatial attention module calculates the weighted sum across all positions of the feature map. The relationship between vessel pixels at different locations will be fully learned. The similar vessel pixels promote each other and the spatial attention module improves the semantic consistency.

**Channel Attention Module**: The relationship between different feature map channels of high-level features can be learned by the channel attention module. The long-range contextual information in the channel dimension helps improve the vessel segmentation performance since different vessel responses are associated with each other. The original feature representation \( S \in R^{C \times H \times W} \) is reshaped to \( A' \in R^{C \times N} \), where \( N = H \times W \) is the total pixel number. \( A' \) and the transpose of \( A' \) is multiplied and followed by a softmax layer to form the channel-wise attention map:

\[ CA_{ij} = \frac{\exp(A'_i, A'_j)}{\sum_{i=1}^{C} \exp(A'_i, A'_j)}. \]  

(6)

The channel attention map \( CA \) calculates the impact of channel \( i \) on channel \( j \). The transpose of \( CA \) and the input feature map \( S \) is multiplied and reshape the result to \( R^{C \times H \times W} \). The reshaped result is scaled by parameter \( \beta \) and element-wise sum with \( S \) to form the final channel attention \( CAO \in R^{C \times H \times W} \):

\[ CAO_i = \beta \sum_{j=1}^{C} (CA_{ij}A'_j) + A'_i. \]  

(7)

Different from other attention modules, we concatenate the spatial attention map, the channel attention map, the summation of the two maps together to form a more capable feature representation.

**Multiscale Output**

Multiscale outputs provide more supervision in network training. There are \( M \) side-output layers in the network, and each side-output layer can be considered as
a classifier to generate a matching local output map for the earlier layers. Here are the loss functions of all the side-output layers:

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

(8)

\[ L_{cross-entropy} \text{ is the cross entropy loss for each side-output layer:} \]

\[ L_{cross-entropy} = -\sum_i (y'_i \log(y_i)), \]  

(9)

\( y_i \) is the predicted probability value for class \( i \) and \( y'_i \) is the true probability for that class. We compute 4 side-output maps and an average layer to combine them all while the final optimization function is the sum of these 5 side-output losses. The side-output layer alleviates the gradient vanishing problem by back-propagating the side-output loss to the early layer in the decoder path, which is helpful for the training of the early layer. We use multiscale fusion because it has been proven to achieve high performance. The side-output layer also adds supervises information for each scale so that it could output a better results. The final layer which is considered as classifier treats the vessel segmentation as a pixel-wise classification to produce the probability map of each pixel.

Acknowledgements
Not applicable

Funding
Not applicable

Abbreviations
HR: Hypertensive Retinopathy; DR: Diabetic Retinopathy; Spe: Specificity; Sen: Sensitivity; Acc: Accuracy; AUC: Area Under ROC; ROC: Receiver Operating Characteristic Curve;

Availability of data and materials
The dataset is freely available, The source code will be released freely available. All dataset and source code are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
This article does not contain any studies with human participants or animals performed by any of the authors.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
All authors consent.

Authors’ contributions
Pengshuai Yin and Yupeng Fang wrote the main manuscript text and prepared all figures. All authors reviewed the manuscript.

Author details
1School of Future Technology, South China University of Technology, Guangzhou, China.  2School of Software Engineering, South China University of Technology, Guangzhou, China.  3Guangdong-Hong Kong-Macao Greater Bay Area Weather Research Center for Monitoring Warning and Forecasting, , Shenzhen, China.

References
1. Smart, T.J., Richards, C.J., Bhatnagar, R., Pavesio, C., Agrawal, R., Jones, P.H.: A study of red blood cell deformability in diabetic retinopathy using optical tweezers. In: Optical Trapping and Optical Micromanipulation XII, vol. 9548, p. 954825 (2015). International Society for Optics and Photonics
2. Irshad, S., Akram, M.U.: Classification of retinal vessels into arteries and veins for detection of hypertensive retinopathy. In: 2014 Cairo International Biomedical Engineering Conference (CIBEC), pp. 133–136 (2014). IEEE
3. Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A.: Blood vessel segmentation methodologies in retinal images—a survey. Computer methods and programs in biomedicine 108(1), 407–433 (2012)
4. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234–241 (2015). Springer
5. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
6. Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H.: Dual attention network for scene segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3146–3154 (2019)
7. Ni, J., Wu, J., Wang, H., Tong, J., Chen, Z., Wong, K.K., Abbott, D.: Global channel attention networks for intracranial vessel segmentation. Computers in biology and medicine 118, 103639 (2020)
8. Mou, L., Zhao, Y., Fu, H., Liu, Y., Cheng, J., Zheng, Y., Su, P., Yang, J., Chen, L., Frangi, A.F., et al.: Cs2-net: Deep learning segmentation of curvilinear structures in medical imaging. Medical image analysis 67, 101874 (2021)
9. Hao, D., Ding, S., Qiu, L., Lv, Y., Fei, B., Zhu, Y., Qin, B.: Sequential vessel segmentation via deep channel attention network. Neural Networks 128, 172–187 (2020)
10. Li, R., Li, M., Li, J., Zhou, Y.: Connection sensitive attention u-net for accurate retinal vessel segmentation. arXiv preprint arXiv:1903.05558 (2019)
11. Wang, D., Haytham, A., Pottenburgh, J., Saeedi, O., Tao, Y.: Hard attention net for automatic retinal vessel segmentation. IEEE Journal of Biomedical and Health Informatics 24(12), 3384–3396 (2020)
12. Yue, K., Zou, B., Chen, Z., Liu, Q.: Retinal vessel segmentation using dense u-net with multiscale inputs. Journal of Medical Imaging 6(3), 034004 (2019)
13. Annunziata, R., Gazzelli, A., Ballerini, L., Mecocci, A., Trucco, E.: Leveraging multiscale hessian-based enhancement with a novel exude inpainting technique for retinal vessel segmentation. IEEE journal of biomedical and health informatics 20(4), 1129–1138 (2015)
14. Wu, Y., Xia, Y., Song, Y., Zhang, Y., Cai, W.: Multiscale network followed network model for retinal vessel segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 119–126 (2018). Springer
15. Yin, P., Yuan, R., Cheng, Y., Wu, Q.: Deep guidance network for biomedical image segmentation. IEEE Access 8, 116106–116116 (2020)
16. Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A.: An ensemble classification-based approach applied to retinal blood vessel segmentation. IEEE Transactions on Biomedical Engineering 59(9), 2538–2548 (2012). doi:10.1109/TBME.2012.2205687
17. Li, Q., Feng, B., Xie, L., Liang, P., Zhang, H., Wang, T.: A cross-modality learning approach for vessel segmentation in retinal images. IEEE Transactions on Medical Imaging 35(1), 109–118 (2015)
18. Zhang, J., Dashbozorg, B., Bekkers, E., Pluim, J.P., Duits, R., ter Haar Romeny, B.M.: Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. IEEE transactions on medical imaging 35(12), 2631–2644 (2016)
19. Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels with deep neural networks. IEEE transactions on medical imaging 35(11), 2399–2380 (2016)
20. Maninis, K.-K., Pont-Tuset, J., Arbeláez, P., Van Gool, L.: Deep retinal image understanding. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 140–148 (2016). Springer
21. Yan, Z., Yang, X., Cheng, K.-T.: Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. IEEE Transactions on Biomedical Engineering 65(9), 1912–1923 (2018)
22. Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J.: Ce-net: Context encoder network for 2d medical image segmentation. IEEE Transactions on Medical Imaging (2019)
23. Zhuang, J.: Laddernet: Multi-path networks based on u-net for medical image segmentation. arXiv preprint arXiv:1810.07810 (2018)
24. Wang, B., Qiu, S., He, H.: Dual encoding u-net for retinal vessel segmentation. In: MICCAI, pp. 84–92 (2019). Springer
25. Liu, B., Gu, L., Lu, F.: Unsupervised ensemble strategy for retinal vessel segmentation. In: MICCAI, pp. 111–119 (2019). Springer
26. Wu, Y., Xia, Y., Song, Y., Zhang, D., Liu, D., Zhang, C., Cai, W.: Vessel-net: retinal vessel segmentation under multi-path supervision. In: MICCAI, pp. 264–272 (2019). Springer
Figures

Figure 1

<p>The segmentation example of the proposed method on DRIVE dataset.</p>
Figure 2

The segmentation example of the proposed method on CHASEDB1 dataset.

Figure 3

The Framework of the proposed framework.
Figure 4

<p>The proposed dual attention module</p>