LETTER
# A Retinal Vessel Segmentation Network Fusing Cross-Modal Features

**Xiaosheng YU**[†a)]**, Jianning CHI**[†]**,** *Nonmembers***, and Ming XU**[††]**,** *Member*

**SUMMARY**    Accurate segmentation of fundus vessel structure can effectively assist doctors in diagnosing eye diseases. In this paper, we propose a fundus blood vessel segmentation network combined with cross-modal features and verify our method on the public data set OCTA-500. Experimental results show that our method has high accuracy and robustness.
*key words: vessel segmentation, CNN, feature fusion, cross-modal*

## 1. Introduction

Eye diseases are often harmful to human health. Common eye diseases include glaucoma, cataracts, and macular degeneration. Early detection and diagnosis are important for the treatment of eye diseases [1]. The retinal vessel system is an important structure of the fundus, and its morphological changes can infer the severity of many neurological and hematological diseases and help to understand disease progression and evaluate treatment effects [2]. Currently, there are two main types of retinal vascular imaging techniques: color fundus imaging and optical coherence laminar angiography. Figure 1(a) shows the fundus image generated by the color fundus imaging technique, in which the vessel structure is less obvious and it is difficult to present richer vascular information.

OCTA is a high-resolution, non-invasive 3D imaging technology for living organisms, which can capture 3D vessel information of the retina at micrometer resolution using coherent light in clinical ophthalmology [3], [4], as shown in Fig. 1(b). Its vertical projection map is shown in Fig. 1(c). Compared with color fundus imaging technology, OCTA technology can capture more abundant fundus blood vessel information, and has become a major detection tool for fundus blood vessel structure.

Traditionally, vessel markings are mainly done by medical practitioners. However, manual drawing of vessel masks is time-consuming and laborious, and susceptible to personal experience [5]. Therefore, using computer vision technology to accurately segment retinal vessels to assist doctors in diagnosing related diseases has become an important topic.

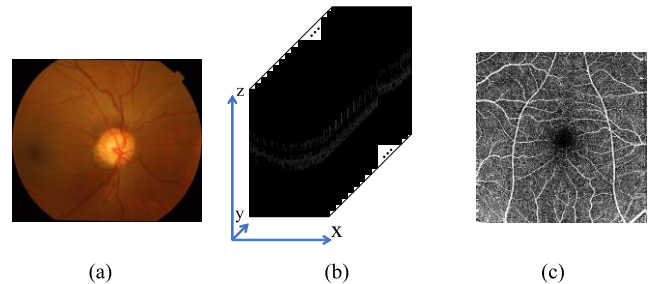Traditional fundus vessel segmentation methods mainly



**Fig. 1** Fundus images generated by different retinal imaging techniques. (a) Retinal color images; (b) 3D OCTA volume image; (c) OCTA projection map.

include adaptive threshold [6]–[8], edge detection [9], [10], and matching filtering [11]–[13]. The structure of the fundus image is complicated and fuzzy, so the traditional segmentation algorithm has low segmentation accuracy and poor adaptability. In recent years, with the development of the field of artificial intelligence, more and more researchers use deep learning methods to segment blood vessels.

Compared with traditional methods, the method based on convolutional neural network (CNN) has been successfully applied in the field of medical segmentation because of its strong feature extraction ability. Classic CNN models include Unet [14], FCN [15], and ResNet [16]. On this basis, researchers further improved the classical network model to improve the accuracy and robustness of the vessel segmentation task. Mou et al. [17] proposed the CS-Net, which takes Unet as the basic structure and adds a spatial attention module and a channel attention module for extracting vessel structures in OCTA projection maps. Li et al. [18] proposed an image projection network (IPN), which combined with rich vessel information provided by 3D OCTA data and used continuous projection learning modules to output vessel structure 2D prediction results. Li et al. [19] further improved IPN and proposed IPN-V2, which enhanced the horizontal direction perception of the original network. Although CNN-based methods have the strong feature extraction ability, the segmentation results of retinal blood vessel walls and small blood vessels are still blurred or lost due to the low signal-to-noise ratio of OCTA images and the limited field of view of the CNN module.

To solve the above problems, in order to make full use of 3D OCTA volume data and 2D vascular projection map provided by dataset OCTA-500, we proposed an end-to-end fundus vascular segmentation network (RVS-Net) combin-

---

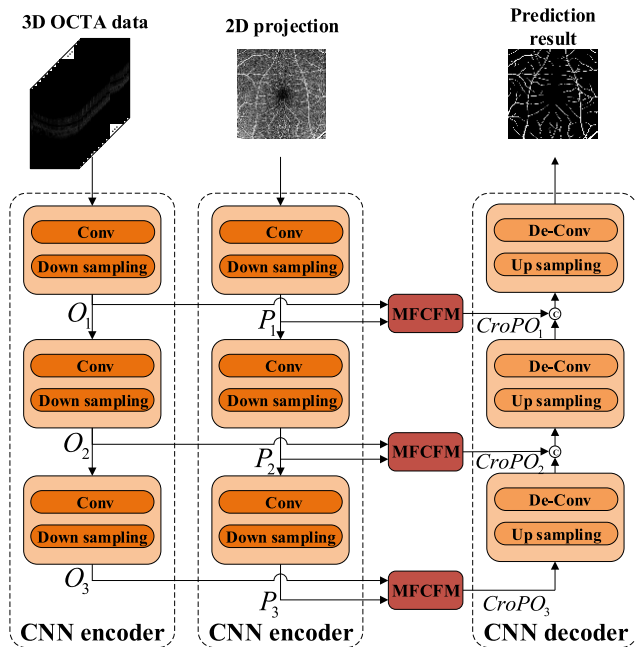**Fig. 2**    The architecture of the proposed RVS-Net.



**Fig. 3**    The architecture of the proposed MFCFM.

ing cross-modal features. The experimental results show that the proposed network can segment fundus blood vessels with high accuracy, can effectively reduce lesion and capillary interference, and has strong robustness.

## 2. Methodologies

### 2.1 Model Architecture

In this paper, the RVS-Net is proposed to extract the vessel structure of 3D OCTA images. As shown in Figure 2, the RVS-Net as a whole is an encoder-decoder network structure consisting of three parts: the CNN encoder module, the multimodal feature cross fuse module (MFCFM), and the CNN decoder module.

### 2.2 CNN Encoder Module

To better utilize CNN to extract features of retinal images and retain spatial information, we design the CNN encoder using the residual convolution (Conv) and down sampling modules in ResNet-50 [16]. Inspired by Li et al. [18], in the 3D OCTA data CNN encoder module, we perform vertical maximum pooling on the extracted features $O_1$, $O_2$, $O_3$ to ensure that the dimensions of $O_1$, $O_2$, $O_3$ and $P_1$, $P_2$, $P_3$ are consistent.

### 2.3 MFCFM Module

Blood vessels in retinal images are mostly elongated structures, and the deep CNN structure is easy to cause local information loss, which makes the decoder prone to problems such as blood vessel loss or rupture during feature recovery [3]. Therefore, we further designed MFCFM to fuse
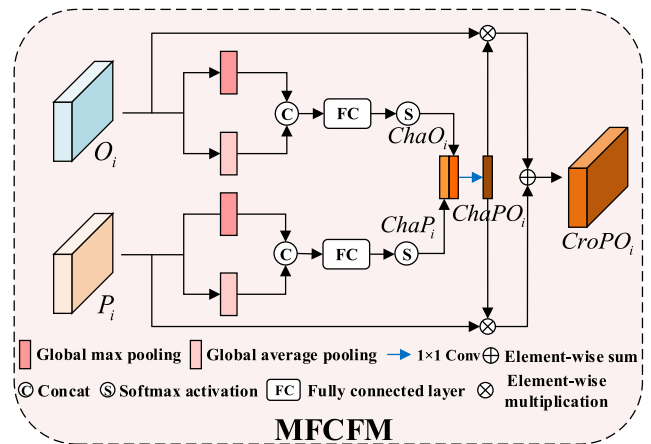
cross-modal features in order to extract more valuable features while suppressing useless features, thereby improving the accuracy of vessel segmentation. The specific structure of MFCFM is shown in Fig. 3.

In the MFCFM module, we first use the global average pooling (GAP) and global max pooling (GMP) layers to obtain a rich global information matrix for $O_i$ and $P_i$, and combine the full connection layer and Softmax activation function to preserve valuable features on the channel. Taking $O_i$ as an example, the above specific formula is as follows:

$$ChaO_i = SoftMax\left(FC\left(Concat\left(Avg\left(O_i\right), Max\left(O_i\right)\right)\right)\right)$$
(1)

where $Avg$ and $Max$ are the GMP and GAP layer, respectively, $Concat$ is the matrix concatenate, $FC$ is the full connection layer, $SoftMax$ is the Softmax activation function. $ChaO_i$ is the channel attention feature map of the 3D OCTA data. In the same way, $ChaP_i$ is the channel attention feature map of the 2D projection map.

Further, we use $1 \times 1$ Conv to fuse cross-modal channel attention feature maps $ChaO_i$ and $ChaP_i$. The specific formula is as follows:

$$ChaPO_i = 1 \times 1Conv\left(ChaP_i, ChaO_i\right)$$
(2)

where $ChaPO_i$ is the fused cross-modal channel attention feature map.

Finally, we use element-wise multiplication and element-wise sum to enhance and fuse features $O_i$ and $P_i$ to generate cross-modal fusion feature $CroPO_i$. The specific formula is as follows:

$$CroPO_i = \left(O_i \otimes ChaPO_i\right) \oplus \left(P_i \otimes ChaPO_i\right)$$
(3)

### 2.4 CNN Decoder Module

In the CNN decoder module, we still use the residual convolution structure in ResNet-50 to design deconvolution, so

as to maintain sufficient spatial information in the feature recovery stage and establish long-distance pixel dependencies.

## 2.5 Loss Function

The proposed RVS-Net network is trained in an end-to-end manner, and the loss function consists of two main parts: the Dice and the Cross Entropy (CE) loss function. The specific formula is as follows:

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^{N} G_i Y_i}{\sum_{i=1}^{N} (G_i + Y_i)} \tag{4}$$

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^{N} (G_i \log (Y_i) + (1 - G_i) \log (1 - Y_i)) \tag{5}$$

$$L_{total} = L_{Dice} + L_{CE} \tag{6}$$

## 3. Experimental Results and Analysis

We used a publicly available dataset OCTA-500 [19] to evaluate the performance of the proposed vessel segmentation method. Published by the School of Computer Science and Engineering, Nanjing University of Science and Technology, OCTA-500 provides 3D OCTA volume data and projection maps retinal images of 500 subjects, which can be divided into two subsets according to the fields of view: OCTA-6M and OCTA-3M. Among them, OCTA6M is mainly from the fundus data of patients with retinal diseases (macular degeneration, diabetic retinopathy), and OCTA3M is mainly from the normal population.

We compare the proposed algorithm with existing OCTA vessel segmentation algorithms: CS-Net [17], IPN [18], and IPN-V2 [19]. The following indexes are used for quantitative analysis: average similarity coefficient (DICE), Jaccard coefficient (JAC), balance accuracy (BACC), precision (PRE), and recall (REC), which are specifically defined as follows:

$$DICE = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{7}$$

$$JAC = \frac{TP}{TP + FP + FN} \tag{8}$$

$$BACC = \frac{TPR + TNR}{2} \tag{9}$$

$$PRE = \frac{TP}{TP + FP} \tag{10}$$

$$REC = TPR = \frac{TP}{TP + FN} \tag{11}$$

$$TNR = \frac{TN}{TN + FP} \tag{12}$$

where TP and FP represent true positive and false positive, respectively, TN and FN represent true negative and false negative, respectively, TPR is the true positive rate, and TNR is the true negative rate.

The experiments are implemented on the PyTorch and trained on NVIDIA RTX3090 GPU with 24 GB memory. In the experiment, we set the Adam optimizer with weight attenuation coefficient of 0.9, batch size of 4, number of iterations of 1000, and learning rate of 1e-4.

To visually verify the feasibility of the proposed vessel segmentation method, Fig. 4 shows the experimental results of the proposed algorithm compared with the above segmentation algorithms under the conditions of the healthy fundus, age-related macular degeneration, and diabetic retinopathy, where rows 1, 3, and 5 correspond to the healthy fundus, age-related macular degeneration fundus and diabetic retinopathy fundus, and rows 2, 4, and 6 are the corresponding local magnification images.

Among the above vessel segmentation methods, CS-Net has poor vessel connectivity and seriously less segmentation. Compared with IPN, IPNV2 has improved the segmentation accuracy, but there is some over-segmentation of blood vessels. When blood vessels in medical images are over-segmented, doctors may be disturbed by false information, leading to misdiagnosis, which in turn leads to some unnecessary treatments and delays in the condition. The proposed method effectively integrates the vascular features of OCTA volume data and projection map through the cross-modal feature fusion module, which enhances the vascular connectivity and segmentation performance of small vessels, and effectively overcomes retinopathy and capillary interference by combining 3D OCTA volume data. The overall segmentation is more accurate, which is helpful to assist doctors in diagnosing eye diseases.

In order to further quantitatively verify the effectiveness of the proposed vessel segmentation method, we combined the above indicators with a series of OCTA vessel segmentation methods for experimental comparison. Table 1 and Table 2 show the experimental comparison results of DICE, JAC, and BACC on the OCTA-6M dataset and OCTA-3M dataset. Experimental results show that the proposed algorithm's DICE, JAC, BACC, PRE, and REC are superior to other segmentation algorithms. The experimental results show that the proposed segmentation network combining OCTA volume data and vessel projection has high accuracy and robustness.

In order to measure the effectiveness of the designed

**Table 1**  Experimental comparison results on the OCTA-6M dataset.

| Method | DICE | JAC | BACC | PRE | REC |
|--------|------|-----|------|-----|-----|
| CS-Net | 87.85 | 78.47 | 92.71 | 97.06 | 86.54 |
| IPN | 88.64 | 79.64 | 93.38 | 97.68 | 88.08 |
| IPN-V2 | 89.08 | 80.41 | 93.77 | 98.02 | 88.62 |
| Ours | **89.93** | **81.42** | **94.21** | **98.89** | **89.58** |

**Table 2**  Experimental comparison results on the OCTA-3M dataset.

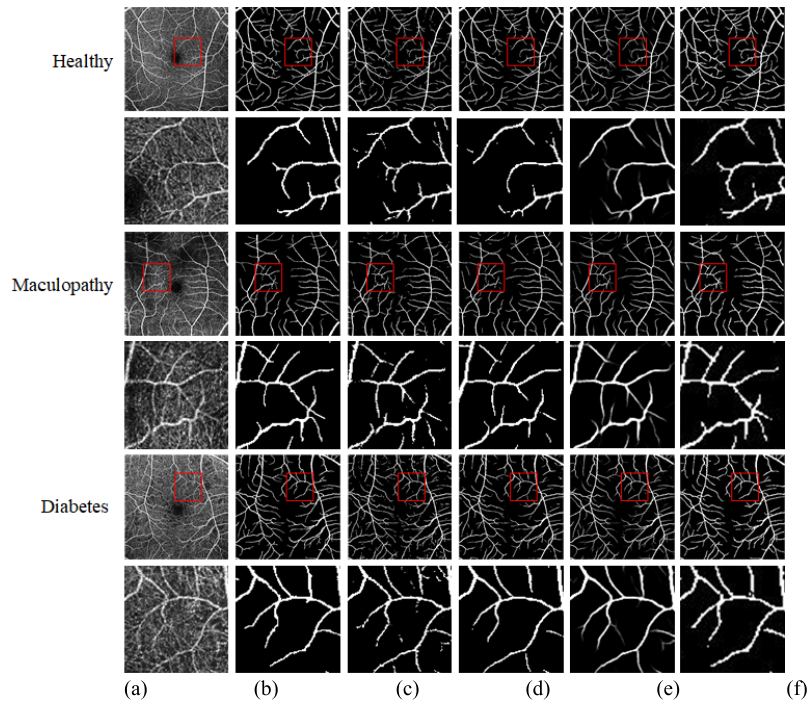| Method | DICE | JAC | BACC | PRE | REC |
|--------|------|-----|------|-----|-----|
| CS-Net | 90.38 | 82.53 | 94.30 | 97.53 | 89.17 |
| IPN | 90.71 | 83.11 | 95.08 | 98.05 | 90.84 |
| IPN-V2 | 92.32 | 85.81 | 95.30 | 98.39 | 91.04 |
| Ours | **92.57** | **86.32** | **95.41** | **99.02** | **91.25** |

**Fig. 4** Experimental results of vessel segmentation under different conditions. (a) Test images. (b) Ground truth. (c) CS-Net. (d) IPN. (e) IPN-V2. (f) Our method.

**Table 3** Ablation experimental strategy.

| Strategy | 3D data | 2D data | MFCFM |
|---|---|---|---|
| A | √ | | |
| B | √ | | √ |
| C | | √ | |
| D | | √ | √ |
| E | √ | √ | |
| F | √ | √ | √ |

**Table 4** Ablation experimental results of OCTA-6M dataset.

| Method | DICE | JAC | BACC | PRE | REC |
|---|---|---|---|---|---|
| A | 87.24 | 77.53 | 89.64 | 98.41 | 84.26 |
| B | 88.63 | 79.52 | 92.80 | 98.50 | 86.68 |
| C | 86.22 | 76.13 | 89.13 | 98.22 | 83.85 |
| D | 88.84 | 79.72 | 92.36 | 98.58 | 86.25 |
| E | 87.89 | 78.68 | 90.58 | 98.65 | 85.98 |
| F | **89.93** | **81.42** | **94.21** | **98.89** | **89.58** |

**Table 5** Ablation experimental results of OCTA-3M dataset.

| Method | DICE | JAC | BACC | PRE | REC |
|---|---|---|---|---|---|
| A | 88.56 | 79.21 | 91.51 | 98.52 | 86.25 |
| B | 89.14 | 80.63 | 93.29 | 98.68 | 87.28 |
| C | 87.82 | 78.36 | 90.45 | 98.54 | 86.14 |
| D | 89.48 | 80.82 | 92.90 | 98.66 | 87.12 |
| E | 90.02 | 82.63 | 93.82 | 98.72 | 88.25 |
| F | **92.57** | **86.32** | **95.41** | **99.02** | **91.25** |

dual-flow network structure and the proposed MFCFM module, we designed a set of ablation experiments, the experimental strategies are shown in Table 3, and the experimental results are shown in Tables 4 and 5.

## 4. Conclusion

In this paper, we propose an end-to-end retinal vessel segmentation network: RVS-Net. Firstly, we propose to segment the vessel structure of OCTA fundus images by combining OCTA volume data with the projection map. In the feature fusion stage, we further propose a feature fusion module to fuse the cross-modal OCTA retinal vessel features to effectively improve the vessel segmentation performance. The proposed algorithm is experimentally validated on the OCTA-500 dataset, and compared with a series of vessel segmentation algorithms, our algorithm has higher overall segmentation accuracy and performs better on retinal images containing lesions and clinical applicability.

## Acknowledgments

## References

[1] M.S. Haleem, L. Han, J. van Hemert, and B. Li, "Automatic extraction of retinal features from colour retinal images for glaucoma diagnosis: A review," Comput. Med. Imaging Graph., vol.37, no.7–8, pp.581–596, Oct. 2013.

[2] Y. Ma, H. Hao, J. Xie, H. Fu, J. Zhang, J. Yang, Z. Wang, J. Liu, Y. Zheng, and Y. Zhao, "ROSE: A retinal OCT-angiography vessel segmentation dataset and new model," IEEE Trans. Med. Imag., vol.40, no.3, pp.928–939, 2021.

[3] R.F. Spaide, J.G. Fujimoto, N.K. Waheed, S.R. Sadda, and G. Staurenghi, "Optical coherence tomography angiography," Prog. Retin.

Eye Res., vol.64, pp.1–55, 2018.

[4] W. Geitzenauer, C.K. Hitzenberger, and U.M. Schmidt-Erfurth, "Retinal optical coherence tomography: Past, present and future perspectives," Br. J. Ophthalmol., vol.95, no.2, p.171, 2011.

[5] Y. Zhao, Y. Zheng, Y. Liu, Y. Zhao, L. Luo, S. Yang, T. Na, Y. Wang, and J. Liu, "Automatic 2-D/3-D vessel enhancement in multiple modality images using a weighted symmetry filter," IEEE Trans. Med. Imag., vol.37, no.2, pp.438–450, Feb. 2018.

[6] X. Jiang and D. Mojon, "Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images," IEEE Trans. Pattern Anal. Mach. Intell., vol.25, no.1, pp.131–137, 2003.

[7] T. Mapayi, S. Viriri, and J.R. Tapamo, "Adaptive thresholding technique for retinal vessel segmentation based on GLCM-energy information," Comput. Math. Methods Med., vol.2015, 2015.

[8] K. Rezaee, J. Haddadnia, and A. Tashk, "Optimized clinical segmentation of retinal blood vessels by using combination of adaptive filtering, fuzzy entropy and skeletonization," Appl. Soft Comput., vol.52, pp.937–951, 2017.

[9] D. Koozekanani, K. Boyer, and C. Roberts, "Retinal thickness measurements from optical coherence tomography using a Markov boundary model," IEEE Trans. Med. Imag., vol.20, no.9, pp.900–916, 2001.

[10] M.W.K. Law and A.C.S. Chung, "Weighted local variance-based edge detection and its application to vascular segmentation in magnetic resonance angiography," IEEE Trans. Med. Imag., vol.26, no.9, pp.1224–1241, 2007.

[11] M. Mirzafam and N. Aghazadeh, "A three-stage shearlet-based algorithm for vessel segmentation in medical imaging," Pattern. Anal. Appl., vol.24, no.2, pp.591–610, 2021.

[12] D. Gou, Y. Wei, H. Fu, and N. Yan, "Retinal vessel extraction using dynamic multi-scale matched filtering and dynamic threshold processing based on histogram fitting," Mach. Vision Appl., vol.29, no.4, pp.655–666, 2018.

[13] Q. Li, J. You, and D. Zhang, "Vessel segmentation and width estimation in retinal images using multiscale production of matched filter responses," Expert. Syst. Appl., vol.39, no.9, pp.7600–7610, 2012.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pp.234–241, 2015.

[15] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol.39, no.4, pp.640–651, April 2017.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp.770–778, 2016.

[17] L. Mou, Y.T. Zhao, L. Chen, J. Cheng, Z. Gu, H. Hao, H. Qi, Y. Zheng, A. Frangi, and J. Liu, "CS-Net: Channel and spatial attention network for curvilinear structure segmentation," Medical Image Computing and Computer-Assisted Intervention – MICCAI 2019, pp.721–730, 2019.

[18] M. Li, Y. Chen, Z. Ji, K. Xie, S.T. Yuan, Q. Chen, and S. Li, "Image projection network: 3D to 2D image segmentation in OCTA images," IEEE Trans. Med. Imag., vol.39, no.11, pp.3343–3354, 2020.

[19] M. Li, K. Huang, Q. Xu, J. Yang, Y. Zhang, Z. Ji, K. Xie, S. Yuan, Q. Liu, and Q. Chen, "OCTA-500: A retinal dataset for optical coherence tomography angiography study," 2022, [Online]. Available: https://ieee-dataport.org/open-access/octa-500