Automatic segmentation of deep foveal avascular zone in optical coherence tomography angiography using deep learning method

Jingmin Luan*
School of Computer and Communication Engineering,
Northeastern University, Qinhuangdao City, Hebei Province, China
*e-mail: luanjingmin@neuq.edu.cn

Ruijie Gan
School of Computer and Communication Engineering,
Northeastern University, Qinhuangdao City, Hebei Province, China
e-mail: lancegan0405@outlook.com

Jing Yu
School of Computer and Communication Engineering,
Northeastern University, Qinhuangdao City, Hebei Province, China
e-mail: glint.yu@outlook.com

Zehao Wei
School of Computer and Communication Engineering,
Northeastern University, Qinhuangdao City, Hebei Province, China
e-mail: nyanti@163.com

***Abstract*: Optical coherence tomography angiography (OCTA) is widely used for visualizing the retinal vascular network, allowing convenient quantification of the deep foveal avascular zone (dFAZ). Accurate segmentation of the dFAZ is essential for diagnosing many retinal-related diseases, but achieving precise segmentation results with clear boundaries poses challenges due to various factors. In this study, we propose a hierarchical and parallel dual-branch deep learning model with attention mechanisms for dFAZ ssegmentation. We tested and compared our method with other mainstream models using our dataset. The results obtained using our proposed model demonstrate clearer and smoother boundaries, with a Dice coefficient of 0.8573, an IoU of 0.8117, and a 95% Hausdorff distance of 19.7116, indicating accurate segmentation of the dFAZ and providing significant clinical utility.**

Keywords: Deep learning, OCTA, dFAZ segmentation

I. INTRODUCTION

The foveal avascular zone (FAZ) is a region located at the center of the retina devoid of capillaries, consisting of the deep FAZ (dFAZ) and superficial FAZ (sFAZ), which are respectively located in two parallel vascular networks within the retina [11]. It is closely associated with retinal diseases such as Diabetic Retinopathy (DR) [12]and Retinal Vein Occlusion (RVO) [13]. Moreover, some studies have found that changes in area due to retinal diseases are more pronounced in the dFAZ compared to the sFAZ. Tang et al. [14] reported a more significant relationship between the severity of DR and dFAZ compared to sFAZ. Dimitrova et al. [15] found that changes in the dFAZ were most common in patients with diabetic macular edema. Therefore, the dFAZ holds even greater clinical research significance.

Optical coherence tomography angiography (OCTA) is an emerging non-invasive imaging technology for visualizing retinal vasculature, capable of obtaining high-resolution images of the retinal vascular network [4]. With OCTA images, we can effectively observe the dFAZ and conduct quantitative segmentation to determine its area size, thereby facilitating the diagnosis of retinal-related diseases [18].

In recent years, several methods have been proposed for segmenting the dFAZ in OCTA images, broadly categorized into three types. The first type involves manual delineation using image annotation software, which is time-consuming, labor-intensive, and often susceptible to noise in OCTA images. The second type includes traditional image morphological and statistical methods, such as watershed algorithms [2], region-growing algorithms [7], and binary segmentation [8], among others. However, they are typically designed based on specific datasets, which introduces a dependency on dataset-specific hyperparameters, resulting in poor generalization performance. The third type is based on deep learning methods, with many leveraging the classical model Unet [3]. For example, Nguyen et al. [16] utilized the Hessian Filter for vessel enhancement in original OCTA images, combined with Unet for precise dFAZ segmentation. Liang et al. [17] integrated style transfer networks with Unet, first obtaining binarized vessel images from original OCTA images through style transfer networks, and then training Unet for simultaneous vessel and dFAZ segmentation.

However, the methods mentioned above have limitations as they all overlook the importance of boundary information in dFAZ segmentation, often resulting in segmentation outcomes with unclear boundaries. To fully leverage boundary information and address the issue of blurry dFAZ boundaries, we propose a hierarchical and parallel dual-branch network. In the main branch network, we employ convolutional blocks with residual connections to learn and infer the main mask of the dFAZ. Additionally, we introduce the Channel and Dilated Spatial (CDS) attention module at the network's lower layers to focus on more critical feature maps and pixels. In the auxiliary branch network, which shares the same encoder with the main branch, the extracted semantically rich feature maps are input into our proposed boundary extraction module (BEM) to learn and extract boundary information. Subsequently, the results from the auxiliary branch are fused with those from the main branch network through convolutional fusion in the final stage,

2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT) | 979-8-3503-8555-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/AINIT61980.2024.10581503

yielding dFAZ segmentation results refined by boundary information.

II. METHODOLOGY

A. Network Architecture

The specific architecture of our proposed model can be seen in Fig. 1. The model consists of two parallel branches. Both branches are trained simultaneously, and semantic information from them is coupled at the end. The main branch is an encoder-decoder-style symmetric U-shaped network, where the encoder extracts multi-scale features from the input image, and the decoder refines the extracted features element-wise to obtain the segmentation result of the dFAZ. The network has a total of five layers. The first two layers of the encoder and decoder are ordinary consecutive convolution blocks, which consist of two basic convolution blocks C_{conv} . The basic convolution block includes a convolutional layer, a batch normalization layer, and a non-linear activation layer, represented as follows:

$$C_{conv} = ReLu\left(BN(C_{onv}^{3\times3})\right) \tag{1}$$

In the middle two layers, to prevent network degradation due to the deepening of layers and to enhance the network's inference capabilities [5], we employed the residual convolution block Res_{conv} , which consists of a residual connection and C_{conv} . At the bottom of the network, we employed the proposed CDS attention module. Additionally, except for the bottom layer, we incorporated skip connections from the encoder to the decoder at each layer, providing the corresponding multi-scale feature maps to the decoder for richer contextual information.

The auxiliary branch is composed of the proposed Boundary Extraction Module (BEM). It extracts boundary information from the feature maps extracted by the encoder of the main branch and conducts supervised learning for the dFAZ's boundary. In the final result fusion stage, it refines the boundary part of the segmentation result from the main branch.

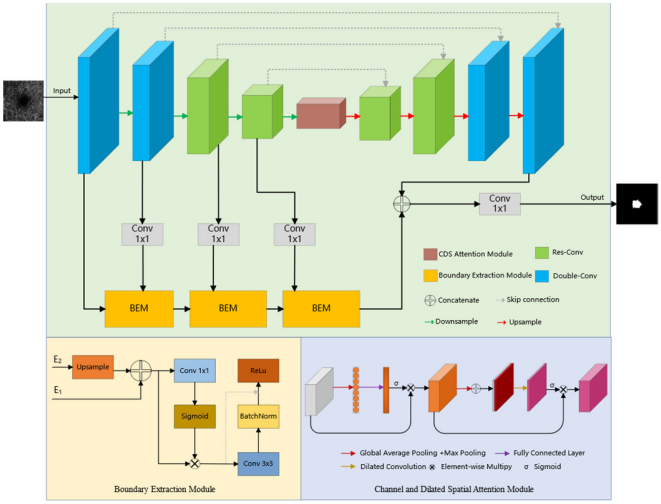

Fig1.Construction of our proposed model

B. Channel and Dilated Spatial Attention Module

Attention mechanisms have emerged as a technology aimed at mimicking the strong focus on regions of interest in human perception. Hu et al. [20] first proposed channel attention, enabling the model to derive weights for each channel of the feature map during the learning process. Woo et al. [19] introduced the spatial attention mechanism, allowing the model to learn the weights of each pixel in the spatial dimension. They combined it with channel attention to form the CBAM module, enabling the network to learn weights simultaneously from both channel and spatial dimensions. As depicted in Fig. 1, we enhanced the spatial attention mechanism proposed by Woo et al, incorporating dilated convolution to calculate spatial weights for a larger receptive field, extracting more contextual information. We then combined this with the channel attention mechanism proposed by Hu et al to create our CDS attention module.

For the calculation of channel attention, we first apply spatial dimension global average pooling P_g and max pooling P_m to the input feature map $X^{C\times H\times W}$. Then we perform pixel-wise addition and apply a fully connected layer F_c to obtain the encoded channel weight feature map. Finally, after applying the Sigmoid activation function σ, we obtain the channel attention $W_c^{C\times1\times1}$, the process is shown below:

$$W_c = \sigma\left(F_c(P_g(X) + P_m(X)\right) \tag{2}$$

After multiplying the X by W_c, we obtain the feature map X'. Next, we proceed to calculate spatial attention. Firstly, we employ the channel dimension global average pooling P_g' and max pooling P_m' to the input feature map X'. Then, the results are concatenated along the channel dimension and passed by a dilated convolutional kernel with a size of 3, dilation of 2, and padding of 2. Finally, after applying the Sigmoid activation functionσ, the spatial attention W_s is obtained.

$$W_s = \sigma(Dconv(P_g'(X') \oplus P_m'(X'))) \tag{3}$$

Here, $\oplus$ means the concatenate operation and $Dconv$ means the dilated convolution operation. Then, we multiply X' with W_s, get the final feature map that weighted in channel and spatial dimension.

C. Boundary Extraction Module

In recent years, a plenty of studies have noticed the importance of the boundary information for semantic segmentation. Towaki et al. [9] proposed a CNN composed of Regular Stream and Shape Stream, where shape information is learned as a separate branch in Shape Stream. Zhen et al. [10] considered semantic segmentation and semantic boundary detection as a pair of dual tasks. They introduced an iterative pyramid context module to couple the two tasks, simultaneously storing and interacting with shared latent semantic information from both tasks.

Inspired by these studies, we propose the Boundary Extraction Module (BEM) as shown in Fig1, and utilize it to construct our auxiliary branch. Here, E_1 is the output of the previous BEM (in the first BEM, E_1 is the feature map of the first layer output from the encoder in main branch, resized to the original input image size), and E_2 is the feature map of different layers' output from the encoder via $C_{onv}^{1\times1}$. Upon entering the BEM module, E_2 is upsampled to the size of E_1

and both feature maps are concatenated along channel dimension to obtain the feature map E .Then E undergoes $C_{onv}^{1\times1}$ followed by Sigmoid activation to get the boundary map M_e. After multiplying E with M_e, the boundary information of both feature maps is aligned. The process is shown below:

$$E = E_1 \oplus Up(E_2) \tag{4}$$

$$M_e = \sigma(C_{onv}^{1\times1}(E) \tag{5}$$

Finally, supervised boundary learning is performed through the basic convolution module $\mathrm{C_{conv}}$ with a residual connection, resulting in the final output

III. EXPERIMENT

A. Dataset and Implementation Detail

Our dataset comprises 94 OCTA images sourced from the Department of Ophthalmology, First Hospital of Qinhuangdao City, Hebei Province, China. Acquisition was facilitated using the Spectralis OCTA system by Heidelberg Engineering, Germany. The focus of data collection centered on the optic disc region of patients diagnosed with conditions such as diabetic retinopathy (DR) and diabetic macular edema (DME). Measurements were conducted at a wavelength of 870nm, with an A-scan scanning speed of 85,000 Hz, covering a scanning area of 4.5mm × 4.5mm.

In our experiments, we divided the dataset into training, validation, test sets at a ratio of 6:1:3. the batch size is 4 and the Adam optimizer with a learning rate of 0.001 is used for gradient updates. The entire project was based on the open source library PyTorch (version: 1.12.0+cu116), and ran on a personal computer equipped with Windows 11, an RTX 3060 GPU (12GB) and an Intel Core i5-12500 CPU.

B. Total Loss Function

The total loss function is shown below:

$$L = L_{mask} - \alpha L_{edge} \tag{6}$$

Here, the L_{edge} and L_{mask} represents the loss of the auxiliary branch and the whole network respectively, and the binary loss function is applied on both of them. The α is a balancing factor between two loss and calculated as $\alpha = \frac{GT_{edge}}{GT}$, there GT_{edge} is the boundary of ground truth and defined as below:

$$GT_{edge} = GT_{mask} - Erode(GT_{mask} - SE) \tag{7}$$

In this operation, $Erode$ denotes the morphological erosion operation, and SE represents the erosion kernel with a size of 9 by 9[21]

C. Evaluation Metrics and Results

To evaluate our proposed model's result in pixel-wise, we employed Accuracy (Acc), Recall, Intersection over Union (IoU), Dice coefficient. And the area under ROC(AUC) was used to assess the performance of model, the 95% Hausdorff Distance (HD95) was utilized to evaluate the boundary clarity of segmentation result. The pixel threshold for binarization was set to 10.

Table 1. Results of proposed model and other mainstream models

Model	Acc↑	Recall↑	IoU↑	Dice↑	AUC↑	HD95↓
Unet	0.8997	0.9013	0.7246	0.7999	0.9004	40.5984
Unet++	0.9023	0.9046	0.7630	0.8271	0.9034	72.9487
U2net	0.9025	**0.9071**	0.7711	0.8335	0.9033	23.5134
Ours	**0.9045**	0.9019	**0.8117**	**0.8573**	**0.9046**	**19.7116**

To highlight the competitiveness of our model, we conducted comparisons with several mainstream models Unet [3], Unet++ [6], U2net [1]. All models were trained and tested following the implementation details described. The final results are shown in Table1, demonstrating our model's superior competitiveness. Our model reached the highest accuracy of 0.9045, IoU of 0.8117, Dice coefficient of 0.8573, AUC of 0.9046, and HD95 of 19.7116

Fig2 shows the comparison of segmentation results between our model and other mainstream models. Furthermore, we compare the boundary clarity of their segmentation results as shown in Fig3, it can be seen that our model not only correctly predicts the segmentation boundary, but also is smoother than other models.

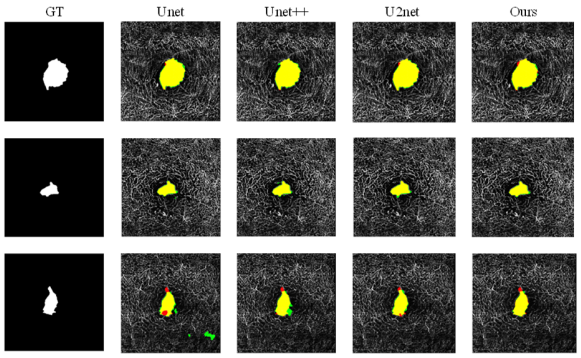

Fig2.Comparison of segmentation results. The green, red, yellow regions represent the predicted result, ground truth, overlap of two areas, respectively.

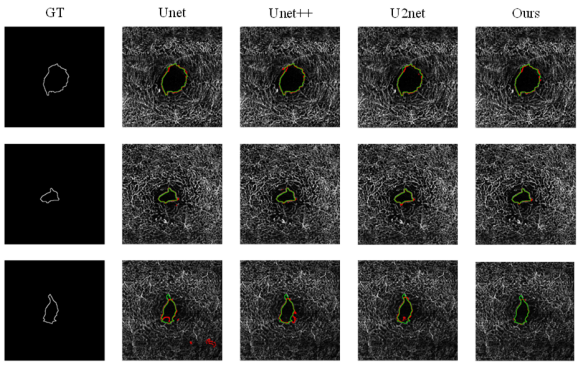

Fig3.Boundaries of segmentation results. The red, green regions represent the boundary of predicted mask, boundary of ground truth

D. Ablation Experiment

To validate the effectiveness of our proposed boundary auxiliary branch and CDS attention module, we conducted ablation experiments on the dataset by: (1) removing the CDS module from the network, (2) removing the auxiliary branch. The results are shown in Table 2, all models were trained and

tested according to the implementation details described, and final results are shown in Table2.

Table 2. Ablation experiment results on our dataset

Model	Acc	Recall	IoU	Dice	Precision
Ours	**0.9041**	0.9013	**0.8107**	**0.8566**	**0.8182**
w/o CDS	0.9015	0.9038	0.7688	0.8252	0.7705
w/o BEM	0.8963	**0.9087**	0.6845	0.7785	0.6848

From Table 2, it is evident that removing the CDS module results in a decrease in performance metrics across all indicators. This provides evidence that our proposed CDS module assists the network in better comprehending crucial information within the multi-scale feature maps, thereby enhancing the refinement of segmentation results. Furthermore, when the boundary auxiliary branch is removed, there is a significant drop in model performance. This underscores the crucial coupling between boundary information and semantic segmentation tasks. It also highlights how semantic segmentation tasks can benefit significantly from the assistance of boundary information.

IV. CONCLUSION

In this paper, we propose a medical image segmentation model for automatic quantification of dFAZ in OCTA images. By introducing attention mechanisms and a boundary information auxiliary branch, we get the segmentation results of dFAZ with clear boundaries, effectively addressing the issue of unclear dFAZ boundaries in previous studies. Furthermore, compared to three mainstream semantic segmentation models, our model achieves the best performance across multiple evaluation metrics, it not only accurately segments the main body of dFAZ but also produces clear and smooth boundaries, attributed to the CDS attention module and the boundary information auxiliary branch composed of BEM.

In future work, we plan to design a model capable of simultaneously segmenting both sFAZ and dFAZ, providing more useful information for the diagnosis of retinal diseases. Our work enhances the segmentation capability of deep learning methods for dFAZ, especially in boundary regions. We hope our proposed method contributes to clinical medical diagnosis.

REFERENCES

[1] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R. Zaiane, Martin Jagersand, U2-Net: Going deeper with nested U-structure for salient object detection, Pattern Recognition, vol. 106, p. 107404, 2020.

[2] Jian Liu, Shixin Yan, Nan Lu, Dongni Yang, Chunhui Fan, Hongyu Lv, Shuanglian Wang, Xin Zhu, YuqianZhao, Yi Wang, Zhenhe Ma, Yao Yu, Automatic segmentation of foveal avascular zone based on adaptivewatershed algorithm in retinal optical coherence tomography angiography images, Journal of InnovativeOptical Health Sciences, 15(01), 2242001, 2022.

[3] Ronneberger, Olaf, Fischer, Philipp, Brox, Thomas, U-Net: Convolutional Networks for Biomedical ImageSegmentation, Springer International Publishing, 2015.

[4] Javed, A., Khanna, A., Palmer, E., Wilde, C., Zaman, A., Orr, G., Kumudhan, D., Lakshmanan, A.,Panos, G. D., Optical coherence tomography angiography: a review of the current literature, J Int Med Res, 2023, 51(7), 3000605231187933, Jul.

[5] He, K., Zhang, X., Ren, S., Sun, J., Deep Residual Learning for Image Recognition, 2016 IEEE Conferenceon Computer Vision and Pattern Recognition (CVPR), 2016, 770-778, doi: 10.1109/CVPR.2016.90.

[6] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, Jianming Liang, UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation, 2020, arXiv:1912.05074.

[7] M. Az, J. Novo, P. N., F. Gomez-Ulla, M. G. Penedo, M. Ortega, Automatic segmentation of the fovealavascular zone in ophthalmological OCT-A images, PLoS One, 2019, Volume 14, Number 2, Page e0212364.

[8] Enrique J. Carmona, Macarena Díaz, Jorge Novo, Marcos Ortega, Modeling, Localization, and Segmentation of the Foveal Avascular Zone on Retinal OCT-Angiography Images, IEEE Access, 2020, Volume 8, Pages 152223-152238.

[9] Towaki Takikawa, David Acuna, Varun Jampani, Sanja Fidler, Gated-SCNN: Gated Shape CNNs for Semantic Segmentation, 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, Pages 5228-5237.

[10] Mingmin Zhen, Jinglu Wang, Lei Zhou, Shiwei Li, Tianwei Shen, Jiaxiang Shang, Tian Fang, Long Quan, Joint Semantic Segmentation and Boundary Detection Using Iterative Pyramid Contexts, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, Pages 13663-13672

[11] Hogan MJ, Alvarado JA, Weddell JE (1971) Histology of the human eye: anatlas and textbook. W.B. Saunders, Philadelphia

[12] J Conrath, Roch Giorgi, David Raccah, and B Ridings, “Foveal avascular zone in diabetic retinopathy: Quantitative vs qualitative assessment,” Eye (London, England), vol. 19, pp. 322–6,04 2005.

[13] Balaratnasingam C, Inoue M, Ahn S, McCann J, Dhrami-Gavazi E, Yannuzzi LA, Freund KB. Visual Acuity Is Correlated with the Area of theFoveal Avascular Zone in Diabetic Retinopathy and Retinal Vein Occlusion. American Academy of Ophthalmology. 2016 Aug (123).

[14] F.Y. Tang, E.O. Chan, Z.H. Sun, R. Wong, J. Lok, S. Szeto, et al.., Clinically relevant factors associated with quantitative optical coherence tomography angiography metrics in deep capillary plexus in patients with diabetes, Eye Vis. 7 (1) (2020)https://doi.org/10.1186/s40662-019-0173-y.

[15] G. Dimitrova, E. Chihara, Implication of deep-vascular-Layer alteration detected by optical coherence tomography angiography for the pathogenesis of diabetic retinopathy, Ophthalmologica 241 (4) (2019) 179–182, https://doi.org/10.1159/ 000495624.

[16] Nguyen, Q.N., Nguyen, V.TQ., Hsu, T., Vajzovic, L., Ngo, H.T. (2022). Automatic Foveal Avascular Zone Segmentation Using Hessian-Based Filter and U-Net Deep Learning Network. In: Van Toi, V., Nguyen, TH., Long, V.B., Huong, H.T.T. (eds) 8th International Conference on the Development of Biomedical Engineering in Vietnam. BME 2020. IFMBE. Proceedings,vol85.Springer,Cham.https://doi.org/10.1007/978-3-030-75506-5_68

[17] Zhijin Liang, Junkang Zhang, Cheolhong An, Foveal Avascular Zone Segmentation of Octa ImagesUsing Deep Learning Approach with Unsupervised Vessel Segmentation, in ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1200-1204, 2021.

[18] Y. Ma *et al.*., "ROSE: A Retinal OCT-Angiography Vessel Segmentation Dataset and New Model," in *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 928-939, March 2021, doi: 10.1109/TMI.2020.3042802

[19] Woo, S., Park, J., Lee, JY., Kweon, I.S. (2018). CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science(), vol 11211. Springer, Cham. https://doi.org/10.1007/978-3-030-01234-2_1

[20] J. Hu, L. Shen and G. Sun, "Squeeze-and-Excitation Networks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 7132-7141, doi: 10.1109/CVPR.2018.0074

[21] Menglin Guo, Mei Zhao, Allen MY Cheong, Federico Corvi, Xin Chen, Siping Chen, Yongjin Zhou, AndrewKC Lam, Can deep learning improve the automatic segmentation of deep foveal avascular zone in opticalcoherence tomography angiography?, Biomedical Signal Processing and Control, 2021, Volume 66, Page102456