frontiers in Neuroscience

ORIGINAL RESEARCH
published: 13 October 2021
doi: 10.3389/fnins.2021.758887

Optimized-Unet: Novel Algorithm for Parapapillary Atrophy Segmentation

Cheng Wan[1], Jiasheng Wu[2], Han Li[1], Zhipeng Yan[3], Chenghu Wang[3], Qin Jiang[3], Guofan Cao[3], Yanwu Xu[4]* and Weihua Yang[3]**

[1] College of Electronic and Information Engineering/College of Integrated Circuits, Nanjing University of Aeronautics and Astronautics, Nanjing, China, [2] School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China, [3] The Affiliated Eye Hospital of Nanjing Medical University, Nanjing, China, [4] Intelligent Healthcare Unit, Baidu, Beijing, China

In recent years, an increasing number of people have myopia in China, especially the younger generation. Common myopia may develop into high myopia. High myopia causes visual impairment and blindness. Parapapillary atrophy (PPA) is a typical retinal pathology related to high myopia, which is also a basic clue for diagnosing high myopia. Therefore, accurate segmentation of the PPA is essential for high myopia diagnosis and treatment. In this study, we propose an optimized Unet (OT-Unet) to solve this important task. OT-Unet uses one of the pre-trained models: Visual Geometry Group (VGG), ResNet, and Res2Net, as a backbone and is combined with edge attention, parallel partial decoder, and reverse attention modules to improve the segmentation accuracy. In general, using the pre-trained models can improve the accuracy with fewer samples. The edge attention module extracts contour information, the parallel partial decoder module combines the multi-scale features, and the reverse attention module integrates high- and low-level features. We also propose an augmented loss function to increase the weight of complex pixels to enable the network to segment more complex lesion areas. Based on a dataset containing 360 images (Including 26 pictures provided by PALM), the proposed OT-Unet achieves a high AUC (Area Under Curve) of 0.9235, indicating a significant improvement over the original Unet (0.7917).

Keywords: medical image segmentation, high myopia, parapapillary atrophy, convolutional neural network, fundus image

OPEN ACCESS

Edited by:
Jian Zheng,
Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences (CAS), China

Reviewed by:
Xi Jiang,
University of Electronic Science and Technology of China, China
Weifang Zhu,
Soochow University, China

****Correspondence:***
Guofan Cao
caoguofan587@163.com
Yanwu Xu
ywxu@ieee.org
Weihua Yang
benben0606@139.com

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

Received: *15 August 2021*
Accepted: *23 September 2021*
Published: *13 October 2021*

Citation:
Wan C, Wu J, Li H, Yan Z, Wang C, Jiang Q, Cao G, Xu Y and Yang W (2021) Optimized-Unet: Novel Algorithm for Parapapillary Atrophy Segmentation. Front. Neurosci. 15:758887. doi: 10.3389/fnins.2021.758887

INTRODUCTION

Myopia is a common eye disease, which refers to the blur of vision when light enters the eye and gathers in front of the human retina (Saw et al., 1996). Some patients experience symptoms such as headache and eye fatigue. Myopia is the main cause of vision loss worldwide (Fredrick, 2002). Vision can be corrected using glasses, contact lenses, and refractive surgery; however, they do not solve the potential defects (Morgan et al., 2012; Dolgin, 2015). In recent years, the global proportion of patients with myopia has been increasing (Holden et al., 2016). Adolescent myopia has become a common phenomenon in East and Southeast Asia (Morgan et al., 2018). The degree of myopia is usually divided by the size of the diopter D, which is divided into mild myopia (−3 D or below), moderate myopia (−3 D to −6 D), and high myopia (−6 D or above)

(Saw et al., 2005). Patients with high myopia are more likely to have retinal detachment, and the probability of suffering from glaucoma is higher. Floating objects and shadows appear in the field of vision in many highly myopic patients. The medical burden of high myopia includes pathological complications such as myopic macular degeneration, choroidal neovascularization, cataracts, and glaucoma (Pan et al., 2012).

Parapapillary atrophy (PPA) often occurs in the fundus of patients with high myopia. The extent and development of PPA are useful medical assessment tools because they are closely related to the severity of several eye diseases and conditions, including glaucomatous optic nerve damage, visual field defects, and myopia (Heijl and Samander, 1985; Park et al., 1996; Uchida et al., 1998; Dai et al., 2013). The size and position of PPA are not fixed. If it progresses to the macular area, patients will find it difficult to see objects close to them. Generally, it is determined whether it is still expanding according to the shrinking edge. A clear edge indicates that the PPA has probably stopped progressing; the fuzzy and irregular edges indicate that it is still progressing.

Currently, convolutional neural networks are widely used in the field of medical diagnosis (Fang et al., 2020; Xia et al., 2020; Yang et al., 2021). This study attempts to use a new type of convolutional neural network (OT-Unet) to automatically segment PPA. The OT-Unet is based on Unet, using three pre-training models: VGG, ResNet, and Res2Net in the convolutional feature extraction stage. The edge attention, parallel partial decoder, and reverse attention modules were added to the network, and the loss function was improved simultaneously. Considering these improvements, the accuracy of the network segmentation has significantly improved.

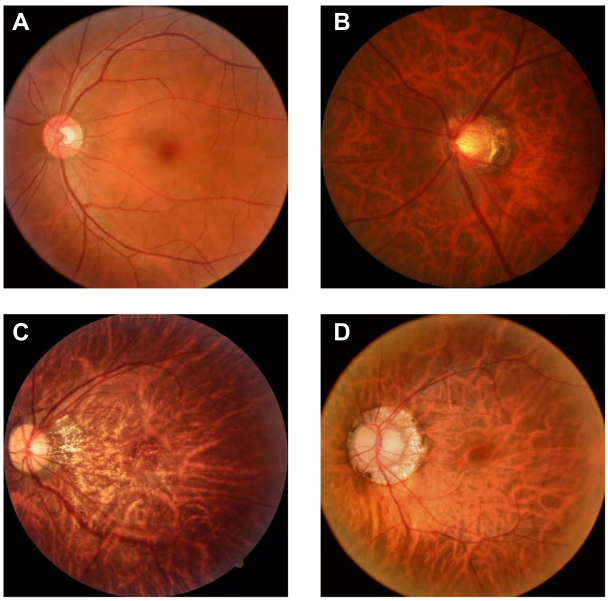

FIGURE 1 | Color photo of fundus: **(A)** fundus of normal eye: It has clear boundary and clear blood vessels, and the overall color is rosy; **(B)** fundus of high myopia: It has blurred boundary, parapapillary atrophy and leopard eye fundus; **(C)** progressive parapapillary atrophy: It has blurred boundary; **(D)** unprogressive parapapillary atrophy: It has clear borders.

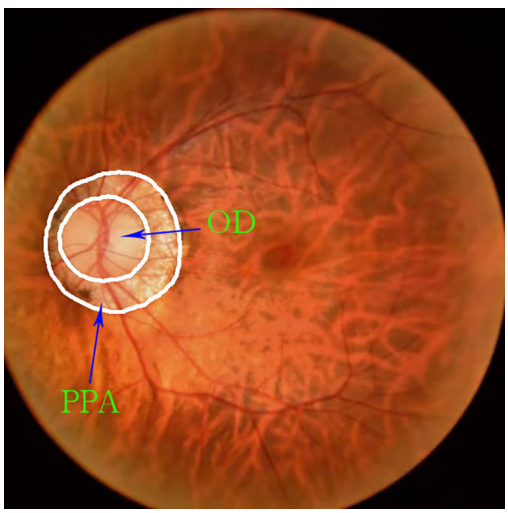

FIGURE 2 | Relationship between the position of the parapapillary atrophy and optic disc: The part inside the white circle in the middle is optic disc, and the ring part between the two white circles is parapapillary atrophy.

MATERIALS AND METHODS

Data Acquisition and Processing

There are few datasets of color fundus photos for high myopia, and only 26 images on iChallenge-PALM can be retrieved on the Internet. This cannot meet the needs of segmentation network training. The research team obtained more than 400 datasets from the Affiliated Eye Hospital of Nanjing Medical University. The shooting equipment was a Topcon TRC-NW300 non-mydriatic fundus camera. Preliminary processing of the data was performed. The blurred pictures and pictures with severely deformed fundus were removed, the rectangular pictures were cut into squares, and the size was unified. Finally, 360 color fundus photos of better quality were obtained. **Figure 1** shows color photo of fundus. The labelme tool was used to label the PPA, and the labeling was performed under the guidance of a professional doctor. **Figure 2** shows the relationship between the position of the parapapillary atrophy and optic disc.

The resolution of the color fundus photos used in the experiment was 352 × 352, and the edge truth map was obtained from the real label map using Adobe Illustrator 2019. The dataset was certified by a professional doctor. According to the ratio of 4:1, we divided the data set into 288 training pictures and 72 test pictures. To highlight the PPA while reducing the size of the image, the grayscale fundus photos were obtained. This operation was implemented using Adobe PhotoShop 2019. **Figure 3** shows the original image and its corresponding grayscale map, PPA truth map and PPA edge map.

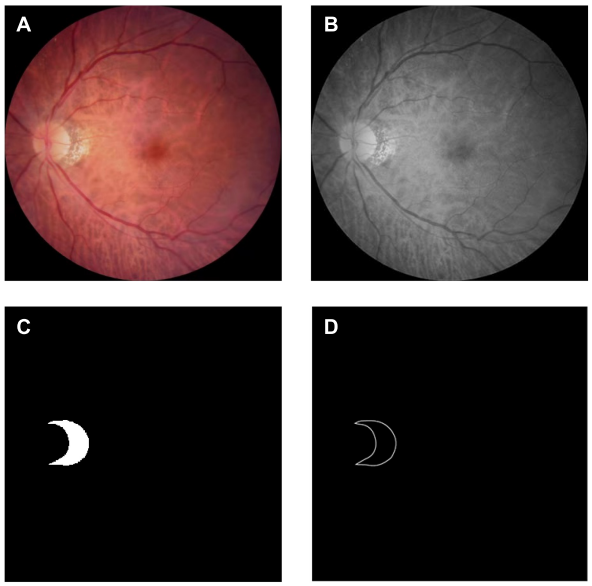

FIGURE 3 | Dataset: **(A)** color photo of the fundus of high myopia: Obtained using Topcon TRC-NW300 non-mydriatic fundus camera; **(B)** grayscale image of the fundus of high myopia: Apply Adobe PhotoShop 2019 to grayscale images; **(C)** gorund truth mask of the parapapillary atrophy: Obtained by labeling the grayscale image using labelme; **(D)** edge map of the parapapillary atrophy: Obtained by processing the real label image using Adobe Illustrator 2019.

Optimized-Unet Overview

The network block diagram of the OT-Unet algorithm proposed in this study is shown in **Figure 4**. The network uses Unet as the backbone network and introduces a pre-training model to generate five convolutional layers. The first two layers $\{f_i, i = 1, 2\}$ are used to extract low-level feature maps that are rich in contextual information, and the high-level feature maps extracted from the last three layers $\{f_i, i = 3, 4, 5\}$ include more local information. An edge attention module is added between the low- and high-level feature convolutional layers to extract the edge feature information of the lesion area. Using the edge feature enhancement module in the second convolutional layer can extract richer local information. Compared with the first convolutional layer, the resolution of the image is lower, which can speed up the calculation. Simultaneously, parallel partial decoders are used to aggregate multi-scale high-level feature information to generate a global map. Since the aggregation of low-level features does not significantly improve the segmentation performance of the network, the network chooses to aggregate three high-level features to obtain richer joint feature information. In addition, the low-level feature map is input to the reverse attention module at all levels under the effect of the global map. These reverse attention modules are cascaded with each other to aggregate the low- and high-level feature information. It can be seen from the figure that the second convolutional layer information, the high-level convolutional layer information, and the aggregated information output by the parallel partial decoder are connected and processed in the reverse attention module. In addition, the use of three reverse attention modules ensures that the network generates sufficiently rich aggregate information. The feature information generated by the last-level inverse attention module is activated by the sigmoid activation function to generate the final lesion area segmentation prediction map. The following will be introduced in detail: the backbone network (Unet), key modules, and improved loss functions.

Backbone Network—Unet

In 2015, Ronneberger et al. (2015) proposed the Unet structure. Unet is a semantic segmentation network based on Fully Convolutional Networks (FCN), which is currently widely used in the field of biomedical image segmentation. The segmentation network system includes contraction (also known as an encoder) and expansion (also known a decoder) paths.

ImageNet's Pre-trained Model

Visual Geometry Group

In 2014, Simonyan and Zisserman (2014) proposed a new network known as the VGG. The image passes through the convolutional layer, and the filter uses a very small receptive field of 3×3 (the minimum size to capture the concepts of left/right, up/down, and center).

ResNet

From experience, the depth of the network is critical to the performance of the model. When the number of network layers is increased, the network can extract more complex feature patterns, so theoretically better results can be achieved when the model is deeper. However, when the network reaches a certain level, the depth of the network increases, and the accuracy of the network becomes saturated or even declined. Considering the ResNet (He et al., 2016), the author uses a residual block to avoid this situation.

Res2Net

Most of the existing methods use a hierarchical method to represent multi-scale features; nonetheless, Gao et al. (2021) have constructed hierarchical residual connections in a single residual block in a different way, proposing the Res2Net neural network building block. This module can express particle-level multiscale features and expand the range of receptive fields.

Feature Expression Enhancement Module

Edge Attention Module

The enhancement effect of edge information in segmentation features has been verified in many studies (Zhang et al., 2019; Xing, 2020; Zhao et al., 2020; Zhou, 2020). The resolution of the low-level feature map was higher, and the edge information was richer. The network inputs the feature map obtained by the second- and low-level convolution to the edge attention module, extracts the edge information feature, and obtains the corresponding map. The difference between the generated edge map and edge truth map G_e corresponding to the

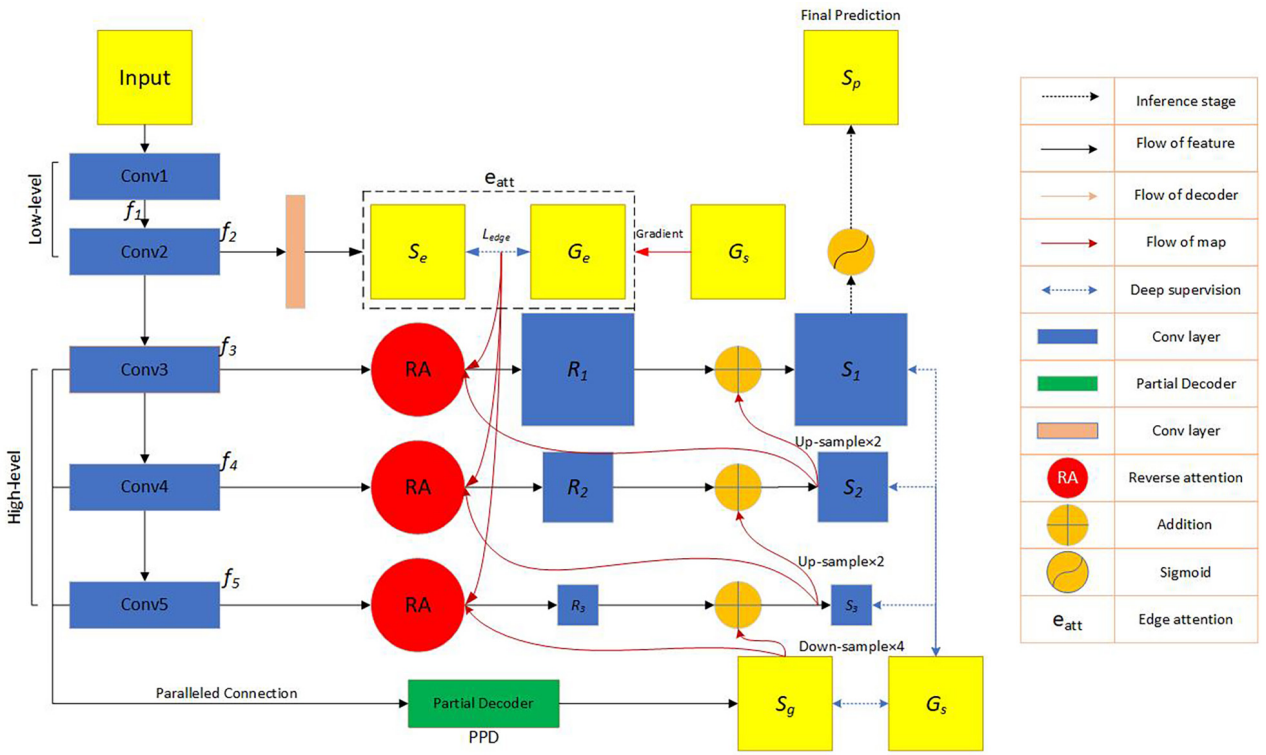

FIGURE 4 | Flow chart of the automatic network segmentation: In the convolutional feature extraction stage, three pre-training models of VGG, ResNet and Res2Net are used, respectively, and the edge attention module, parallel partial decoder module and reverse attention module are added to enhance the feature expression ability. The model also improves the loss function. f_i represents the feature information obtained by the image through the i-th convolutional layer. S_e represents the feature information generated by the second convolutional layer information through the edge attention module. G_e represents the edge map obtained by the ground truth map. L_{edge} represents the standard binary cross-entropy loss function. G_s represents the ground truth map. S_g represents the feature information obtained by the parallel partial decoder aggregating high-level convolutional layer information. R_i(i = 1, 2, 3), respectively, represent the feature information output by the three reverse attention modules. S_i(i = 1, 2, 3), respectively, represent the information obtained by the combination of the reverse attention module information and the higher-level information.

true label is calculated using the binary cross entropy (BCE) loss function:

$$L_{edge} = -\sum_{x=1}^{w}\sum_{y=1}^{h}\left[G_e log\left(S_e\right) + \left(1-G_e\right) log\left(1-S_e\right)\right] \quad (1)$$

where (x, y) represent the coordinates of the pixel points in the predicted edge map S_e and the edge truth map G_e. G_e is derived from the real label G_s. w and h represent the width and height of the feature map, respectively.

Parallel Partial Decoder Module

Segmentation through the combined use of high and low feature maps is a common method of medical segmentation (Qian et al., 2015; Zhou et al., 2018; Gu et al., 2019; Fan et al., 2020). However, because of their large size, low-level feature maps require more resources, and the effect on performance improvement is not obvious (Wu et al., 2019). The parallel partial decoder $Pd(\cdot)$ is used to fuse the high-level features to form a prediction map $S_g = Pd\left(f_3, f_4, f_5\right)$, guiding the input of the inverse attention module. **Figure 5** shows parallel partial decoder module frame diagram.

Reverse Attention Module

Inspired by the study of Chen et al. (2018), the network uses the reverse attention module to extract richer information. The inverse attention module uses a progressive framework. Its information comes from the same convolutional layer, and it includes low-level features f_2 and aggregated features from a higher level. This method can obtain more complex feature information and optimize the segmentation performance of the network.

The predicted feature map of the upper layer was expanded after the sigmoid activation, inversion, and smoothing. We multiplied the expanded feature by the high-level output feature (dot multiplication ·) and concatenated it with the edge attention feature $e_{att}\left(f_2\right)$ to obtain the corresponding reverse attention feature output as demonstrated below:

$$R_i = C\left(f_i \cdot A_i, Dow\left(e_{att}\right)\right) \quad (2)$$

where $Dow(\cdot)$ and $C(\cdot)$ are the down-sampling concatenation operations, respectively.

Figure 6 shows frame diagram of the reverse attention module.

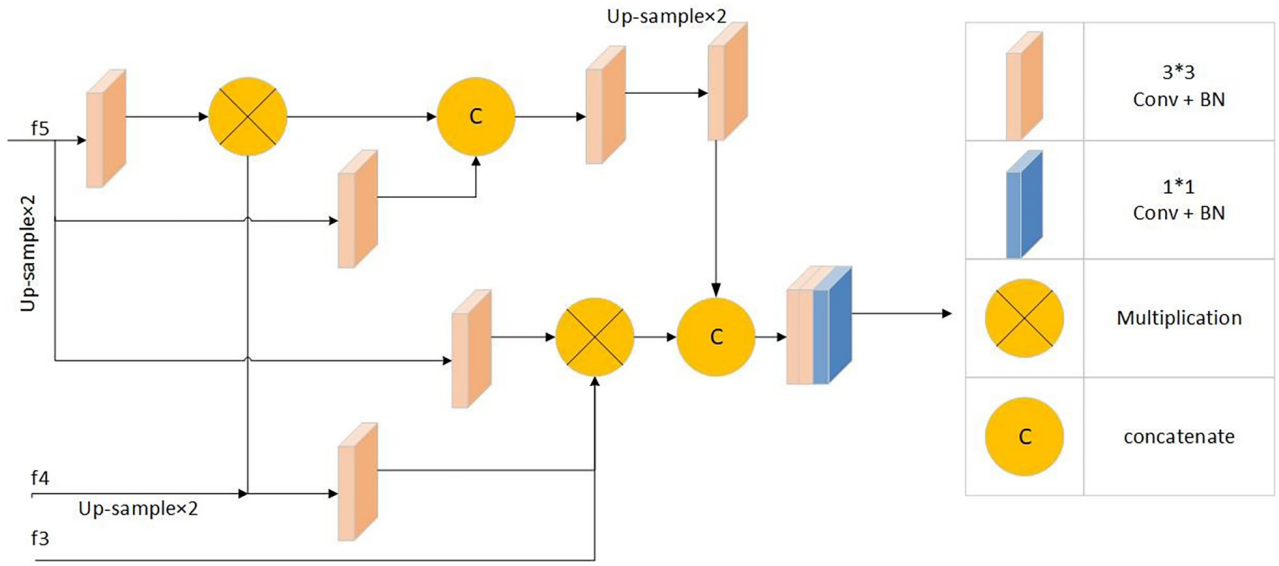

FIGURE 5 | Parallel partial decoder module frame diagram: Including 1*1 convolution + batch normalization processing module, 3*3 convolution + batch normalization processing modules, multiplication operation and concatenate operation.

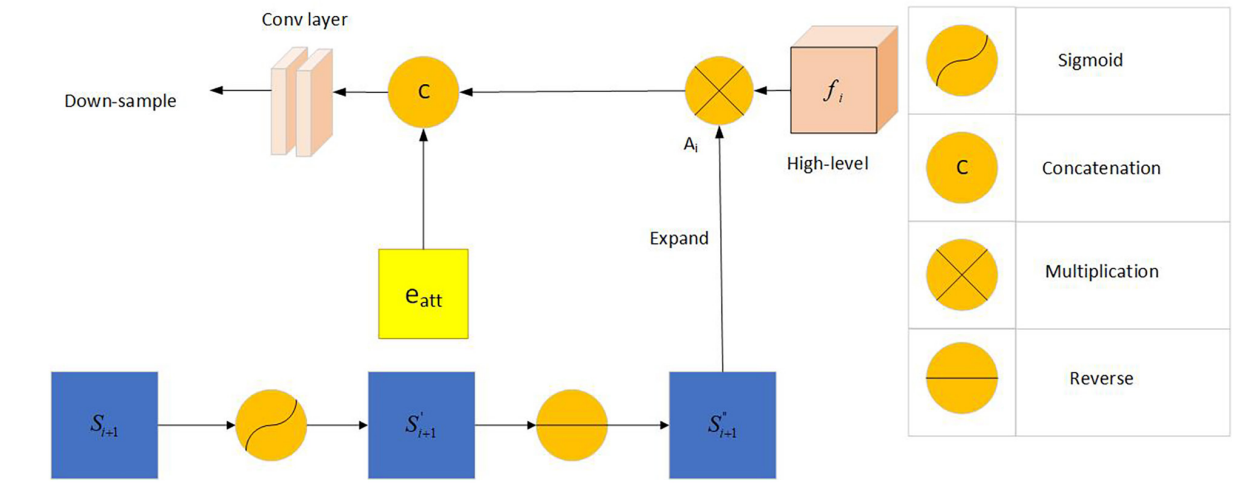

FIGURE 6 | Frame diagram of the reverse attention module: Including Sigmoid activation function, concatenation operation, multiplication operation and reverse operation. f_i (i = 3,4,5) represents the feature information obtained by the picture through the i-th convolutional layer. S_{i+1}(i = 1, 2) represents the information obtained by the combination of the reverse attention module information and the higher-level information. A_i represents the output after S_{i+1} expand.

Loss Function

The standard IoU loss and binary cross-entropy loss functions are the commonly used loss functions. The calculation formulas are as follows:

$$L_{IoU} = 1 - \frac{|A \cap B|}{|A \cup B|} \tag{3}$$

$$L_{BCE} = -\left(ylog\left(p\right) + \left(1-y\right) log\left(1-p\right)\right) \tag{4}$$

$$weight = 1 + 5\left|AVG(B) - B\right| \tag{5}$$

$$L^w_{IoU} = 1 - \frac{(A \cap B) \cdot weight + 1}{|A \cup B| \cdot weight - |A \cap B| \cdot weight + 1} \tag{6}$$

$$L^w_{BCE} = \frac{L_{BCE} \cdot weight}{weight} \tag{7}$$

where A represents the predicted map, B represents the true label, y is the true category, and p is the probability of the predicted category. AVG() is a function in the torch library, specifically torch.nn.AvgPool2d(kernel_size, stride = None, padding = 0), kernel size is the size of the window, stride is the stride of the

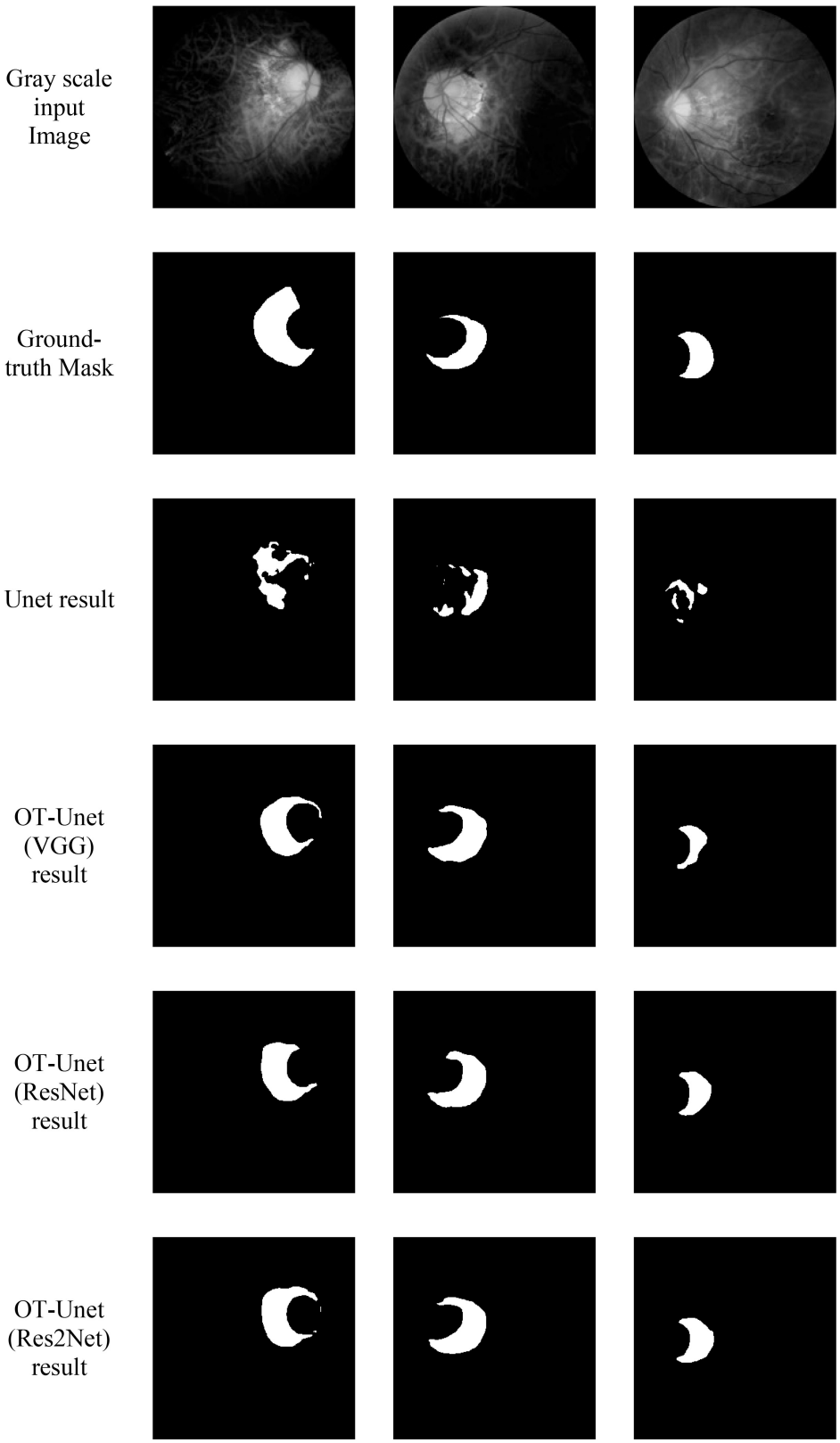

FIGURE 7 | Visualized results of the segmentation of parapapillary atrophy in high myopia: It shows the Gray scale input Image, the Ground-truth Mask, and the segmentation results of Unet, OT-Unet (VGG), OT-Unet (ResNet), and OT-Unet (Res2Net).

TABLE 1 | Comparison of segmentation results of different models.

Methods	Precision	Sensitivity	Specificity	AUC	IoU	DSC
Unet	0.7226	0.8303	0.9879	0.7917	0.4731	0.6413
OT-Unet (VGG)	**0.8227**	0.8225	**0.9963**	0.9134	0.7004	0.8086
OT-Unet (ResNet)	0.7980	0.8398	0.9957	0.9171	0.6877	0.8022
OT-Unet (Res2Net)	0.8020	**0.8450**	0.9958	**0.9235**	**0.7034**	**0.8101**

DSC means Dice similarity coefficient. Bold values indicate that the value is the largest in the same indicator.

TABLE 2 | Comparison of the segmentation results of OT-Unet (Res2Net) in different sizes of lesion areas.

Type of lesion	Precision	Sensitivity	Specificity	AUC	IoU	DSC
Small lesion	0.7479	0.8287	0.9956	0.9076	0.6517	0.7712
Large lesion	**0.9251**	**0.8820**	**0.9962**	**0.9352**	**0.8209**	**0.8986**

Bold values indicate that the value is the largest in the same indicator.

window, and padding is implicit zero padding to be added on both sides.Input(*N,C,Hin,Win*), output(*N,C,Hout,Wout*), where

$$H_{out} = \left\lfloor \frac{H_{in} + 2 \times \text{padding}\,[0] - \text{kernel_size}\,[0]}{\text{stride}\,[0]} + 1 \right\rfloor \quad (8)$$

$$W_{out} = \left\lfloor \frac{W_{in} + 2 \times \text{padding}\,[1] - \text{kernel_size}\,[1]}{\text{stride}\,[1]} + 1 \right\rfloor \quad (9)$$

Where *N* stands for quantity, *C* stands for channel, *Hin* stands for input height, *Win* stands for input width, *Hout* stands for output height, and *Wout* stands for output width.

However, the weights assigned to each pixel by the above two loss functions are the same, and they do not focus on the extraction of complex pixel samples. This study combines the weighted IoU and weighted BCE loss functions to obtain a new loss function:

$$L_{seg} = L_{IoU}^{w} + L_{BCE}^{w} \quad (10)$$

To facilitate the calculation, each predicted feature map is restored to its original size through an up-sampling operation (for example, S_3^{up}). Therefore, we rewrite the total loss function as:

$$L_{total} = L_{seg}\left(G_s, S_g^{up}\right) + L_{edge} + \sum_{i=3}^{5} L_{seg}\left(G_s, S_i^{up}\right) \quad (11)$$

RESULTS

Visualization of Segmentation Results

Figure 7 shows the visualized results of high myopia grayscale images, real labels, and the visual segmentation results of the PPA. It can be observed that the OT-Unet segmentation algorithm has a better segmentation effect than the original Unet.

Comparison of Segmentation Results

The following are the experimental segmentation results of the four networks.

Considering **Table 1**, it can be observed that the segmentation network of this study has improved for all the indicators. It can be seen that OT-Unet is significantly better than Unet in all indicators. The three pre-training models of OT-Unet have different performance in various indicators. OT-Unet (VGG) has the best performance on the Precision and Specificity indicators; on the other indicators, OT-Unet (Res2Net) has the best performance. **Figure 8** shows experimental ROC curve diagram of different segmented image networks.

Comparison of the Segmentation Results of Large and Small Lesions

Based on the relationship between the PPA and papilla diameter (PD), we divided the PPA that was less than or equal to one-third of the PD into small lesion areas (50 images), and the rest were large lesion areas (22 images). The best-performing Res2Net pre-training model was used for the segmentation to obtain the visualization and quantification results.

The quantification results of the segmentation of the large and small lesions are shown in **Table 2**. When OT-Unet segmented large lesions, its Precision, Sensitivity, Specificity and AUC scores were higher than those of small lesions. When segmenting a small lesion area, the lesion area is smaller than the optic disc and is more disturbed by it. Other non-lesion regions around the optic disc also interfere with the segmentation.

According to the segmentation results, the segmentation performance of the OT-Unet on a large lesion area is better than that of a small lesion area in all indicators. When segmenting a small lesion area, the model is more susceptible to the influence of other areas around the optic disc, and even an extreme situation where the segmentation area does not match the real label at all occurs as shown in **Figure 9**.

DISCUSSION

As a method of auxiliary diagnosis, automatic segmentation of ophthalmic medical images can help ophthalmologists to understand a patient's fundus more conveniently and clearly, indicating that this study is very valuable. The width of the PPA is positively correlated with the degree of myopia; therefore, early diagnosis is very important for patients with high myopia. Automatic segmentation of medical images can effectively extract and express image features with less preprocessing and reduce labor costs. Considering the introduction of many excellent segmentation network models and rapid improvement in image processor performance, deep learning can achieve higher segmentation accuracy.

Currently, there are few studies on the automatic segmentation of the PPA. The datasets on the Internet are very limited, manual labeling is time-consuming and laborious, and labeling accuracy cannot be guaranteed. In addition, the early PPA was crescent-shaped, which occurred near the optic disc with a small area. The difference between PPA and the brightness of the optic disc is not obvious, making it easy to be affected by the optic disc when splitting the PPA. Meanwhile, patients with high myopia often have a leopard-shaped fundus,

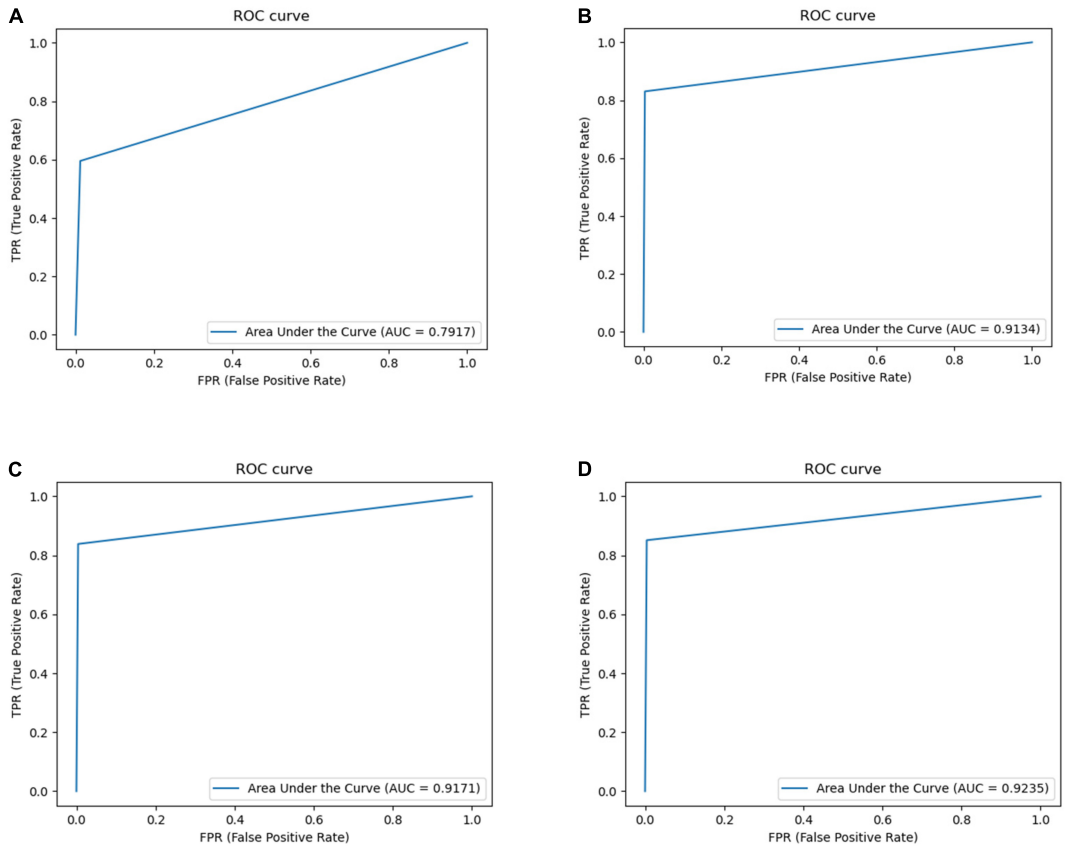

FIGURE 8 | Experimental ROC curve diagram: **(A)** Unet (AUC = 0.7917); **(B)** VGG network (AUC = 0.9134); **(C)** ResNet (AUC = 0.9171); **(D)** Res2Net (AUC = 0.9235).

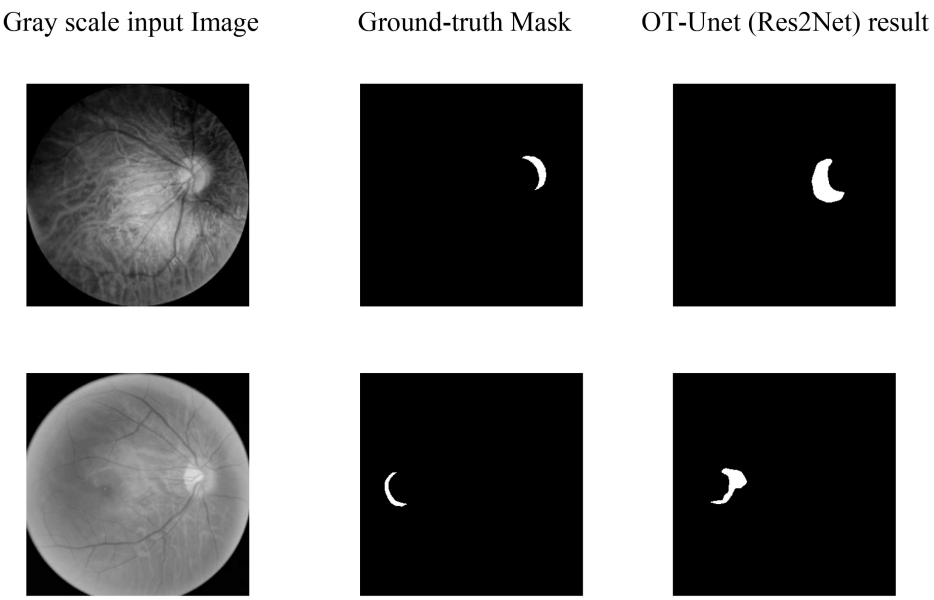

FIGURE 9 | Segmentation of extremely small lesions: It can be seen that the segmentation prediction map does not correspond to the Ground-truth Mask at all.

which is a result of stretching the retina. Visible blood vessels affect the recognition and segmentation of the lesion area and reduce the accuracy of the segmentation. There is no obvious rule for the expansion of the PPA, and the shape and size of the PPA of the fundus in different patients are quite different. There was obvious pigmentation in the PPA, which will also affect the segmentation of PPA. It can be observed that various factors restrict the study of the PPA segmentation network.

The experimental results show that Unet can only segment the approximate outline of the lesion area, which is greatly affected by the optic disc. Also, the prediction map is irregular and has many noises. In the feature extraction stage, OT-Unet uses VGG, ResNet and Res2Net three pre-training models to extract richer feature information. Considering the problem of irregular contours of the lesion, this study adopts the method of adding the edge attention module to extract the contour features of the lesion area during the training phase of feature learning. The trained model can generate a clearer boundary prediction map. Solving the problem of the Unet being severely interfered by the optic disc, this study uses a parallel partial decoder and reverse attention modules to obtain more high-level and low-level fusion features. This helps the network learnt to distinguish between the PPA and optic disc, avoiding splitting the disc. OT-Unet also improved the loss function to get more accurate segmentation results.

However, according to the visualization results, the segmentation result still cannot completely avoid the interference of the optic disc, and the effect is not sufficient when segmenting the small-sized PPA. In the future, the learning ability of the low-level features of the network will be further enhanced to make the network perform better in the segmentation of small lesion areas.

Compared to the original segmentation network (Unet), the improved network has a better effect on lesion segmentation; however, there are still some areas to be improved. Only 288 training images are used in this study, which weakens the generalization ability of the network. Considering high myopia, the size and shape of the PPA at different stages of development are very different, and the use of the same segmentation strategy will reduce the segmentation effect. The improved network makes it difficult to segment small-sized PPA. In the future, it will be necessary to use larger datasets and use data enhancement methods for expansion. Before segmentation, a classification network can be used to classify the PPA according to early, middle, and late stages. Subsequently, based on the characteristics of the para-optical atrophy in the different stages, targeted segmentation strategies can be formulated for segmentation. Regarding the loss function, this study does not systematically study the weight distribution of the weighted IoU and binary cross-entropy loss functions; nonetheless, it simply adds them. In the future, the relationship between these two loss functions can be explored, and a more appropriate weight distribution of the loss functions can be found to improve the segmentation performance of the network.

CONCLUSION

Considering the segmentation task of PPA for high myopia, this study proposes an OT-Unet algorithm network. In this study, three pre-training models (VGG, ResNet, and Res2Net) are used in the Unet for the extraction of convolutional features. Between the high- and low-level convolutions, this study introduces an edge attention module to extract edge feature maps and enrich the network information. Multi-scale high-level feature maps use a parallel partial decoder module to perform feature fusion and obtain global information. The network also uses a reverse attention module which uses a progressive framework to extract high- and low-level feature information. Considering the loss function, the network combines the weighted IoU and binary cross-entropy loss functions to increase the weight of the complex pixels. This shows that the improvement of the network structure and loss function significantly improves the segmentation performance of the network and obtains a better segmentation effect than the Unet in the segmentation of the lesion area. Compared to the Unet, the improved OT-Unet is superior for all the evaluation criteria.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

CW contributed to the conception of the study and performed the experiments. JW performed the data analyses and wrote the manuscript. HL helped to perform the analysis with constructive discussions. ZY, CHW, and QJ collected the data and directed the writing of the manuscript. GC contributed significantly to analysis and manuscript preparation. YX revised the manuscript. WY supervised the whole study. All authors contributed to the study and approved the submitted version.

FUNDING

This study was supported by the Chinese Postdoctoral Science Foundation (2019M661832), Jiangsu Planned Projects for Postdoctoral Research Funds (2019K226), the Jiangsu Province Advantageous Subject Construction Project, and the Nanjing Enterprise Expert Team Project.

ACKNOWLEDGMENTS

We thank the Affiliated Eye Hospital of Nanjing Medical University for providing the fundus data set.

REFERENCES

Chen, S., Tan, X., Wang, B., Lu, H., Hu, X., Fu, Y., et al. (2018). "Reverse attention for salient object detection," in *Lecture Notes in Computer Science*, eds V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss (Cham: Springer), 236–252. doi: 10.1007/978-3-030-01240-3_15

Dai, Y., Jonas, J. B., Huang, H., Wang, M., and Sun, X. (2013). Microstructure of parapapillary atrophy: beta zone and gamma zone. *Invest. Ophthalmol. Vis. Sci.* 54, 2013–2018. doi: 10.1167/iovs.12-11255

Dolgin, E. (2015). The myopia boom. *Nature* 519, 276–278. doi: 10.1038/519276a

Fan, D. P., Zhou, T., Ji, G.-P., Zhou, Y., Chen, G., Fu, H., et al. (2020). Inf-net: automatic covid-19 lung infection segmentation from ct images. *IEEE Trans. Med. Imaging* 39, 2626–2637. doi: 10.1109/TMI.2020.2996645

Fang, J., Xu, Y., Zhang, X., Hu, Y., and Liu, J. (2020). "Attention-based saliency hashing for ophthalmic image retrieval," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, (Piscataway, NJ: Institute of Electrical and Electronics Engineers), 990–995. doi: 10.1109/BIBM49941.2020.9313536

Fredrick, D. R. (2002). Myopia. *BMJ* 324, 1195–1199. doi: 10.1136/bmj.324.7347.1195

Gao, S. H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., and Torr, P. (2021). Res2net: a new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 652–662. doi: 10.1109/TPAMI.2019.2938758

Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., et al. (2019). Ce-net: context encoder network for 2d medical image segmentation. *IEEE Trans. Med. Imaging* 38, 2281–2292. doi: 10.1109/TMI.2019.2903562

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Las Vegas, NV: Institute of Electrical and Electronics Engineers), 770–778. doi: 10.1109/CVPR.2016.90

Heijl, A., and Samander, C. (1985). "Peripapillary atrophy and glaucomatous visual field defects," in *Proceedings of the 6th International Visual Field Symposium. Documenta Ophthalmologica Proceedings Series*, eds A. Heijl and E. L. Greve (Dordrecht: Springer), 403–407. doi: 10.1007/978-94-009-5512-7_58

Holden, B. A., Fricke, T. R., Wilson, D. A., Jong, M., Naidoo, K. S., Sankaridurg, P., et al. (2016). Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology* 123, 1036–1042. doi: 10.1016/j.ophtha.2016.01.006

Morgan, I. G., French, A. N., Ashby, R. S., Guo, X., Ding, X., He, M., et al. (2018). The epidemics of myopia: aetiology and prevention. *Prog. Retin. Eye Res.* 62, 134–149. doi: 10.1016/j.preteyeres.2017.09.004

Morgan, I. G., Ohno-Matsui, K., and Saw, S.-M. (2012). Myopia. *Lancet* 379, 1739–1748. doi: 10.1016/S0140-6736(12)60272-4

Pan, C. W., Ramamurthy, D., and Saw, S.-M. (2012). Worldwide prevalence and risk factors for myopia. *Ophthalmic Physiol. Opt.* 32, 3–16. doi: 10.1111/j.1475-1313.2011.00884.x

Park, K. H., Tomita, G., Liou, S. Y., and Kitazawa, Y. (1996). Correlation between peripapillary atrophy and optic nerve damage in normal-tension glaucoma. *Ophthalmology* 103, 1899–1906. doi: 10.1016/S0161-6420(96)30409-0

Qian, S., Hai, C. Z., Lin, M., and Zhang, C. (2015). Saliency detection based on conditional random field and image segmentation. *Acta Autom. Sin.* 41, 711–724.

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Cham: Springer), 234–241. doi: 10.1007/978-3-319-24574-4_28

Saw, S. M., Gazzard, G., Shih-Yen, E. C., and Chua, W.-H. (2005). Myopia and associated pathological complications. *Ophthalmic Physiol. Opt.* 25, 381–391. doi: 10.1111/j.1475-1313.2005.00298.x

Saw, S. M., Katz, J., Schein, O. D., Chew, S. J., and Chan, T. K. (1996). Epidemiology of myopia. *Epidemiol. Rev.* 18, 175–187. doi: 10.1093/oxfordjournals.epirev.a017924

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *Arxiv* [Preprint] ArXiv: 1409.1556,

Uchida, H., Ugurlu, S., and Caprioli, J. (1998). Increasing peripapillary atrophy is associated with progressive glaucoma. *Ophthalmology* 105, 1541–1545. doi: 10.1016/S0161-6420(98)98044-7

Wu, Z., Su, L., and Huang, Q. (2019). "Cascaded partial decoder for fast and accurate salient object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Piscataway, NJ: Institute of Electrical and Electronics Engineers), 3907–3916. doi: 10.1109/CVPR.2019.00403

Xia, Z., Yue, G., Xu, Y., Feng, C., Yang, M., Wang, T., et al. (2020). "A novel end-to-end hybrid network for Alzheimer's disease detection using 3D CNN and 3D CLSTM," in *Proceedings of the 17th International Symposium on Biomedical Imaging (ISBI)*, (Iowa City, IA: Institute of Electrical and Electronics Engineers), 1–4. doi: 10.1109/ISBI45749.2020.9098621

Xing, C. S. (2020). *Research on Medical Image Segmentation Method Based on Dual Attention Mechanism*. Changchun: Jilin University.

Yang, P., Zhou, F., Ni, D., Xu, Y., Chen, S., Wang, T., et al. (2021). "Fused sparse network learning for longitudinal analysis of mild cognitive impairment," in *Proceedings of the IEEE Transactions on Cybernetics*, (Piscataway, NJ: Institute of Electrical and Electronics Engineers), 233–246. doi: 10.1109/TCYB.2019.2940526

Zhang, Z., Zhang, Z., Fu, H., Dai, H., Shen, J., Pang, Y., et al. (2019). "Et-net: a generic edge-attention guidance network for medical image segmentation," in *Proceedings of the International Conference on Medical Image Computingand Computer-Assisted Intervention*, (Cham: Springer), 442–450.

Zhao, J., Liu, J. J., Fan, D. P., Cao, Y., Yang, J., and Cheng, M.-M. (2020). "EGNet: edge guidance network for salient object detection," in *Proceedings of the CVF International Conference on Computer Vision (ICCV)*, (Piscataway, NJ: Institute of Electrical and Electronics Engineers).

Zhou, Y. (2020). Saliency detection based on edge effect and attention mechanism. *Comput. Eng. Softw.* 41, 111–116.

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., and Liang, J. (2018). "Unet++: a nested u-net architecture for medical image segmentation," in Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, (Cham: Springer), 3–11.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Wan, Wu, Li, Yan, Wang, Jiang, Cao, Xu and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.