

Dear reviewer #1:

Thank you for your comments concerning our manuscript. We are grateful to you for giving us a precious opportunity to revise the submitted manuscript. Those comments are all valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our research. We have studied the comments carefully and made a correction that meets with approval.

The corrections in the paper and the responses to your comments are as follows:

Responds to your comments:

Reviewer Comment 1:

Abstract: "It is very time-consuming for ophthalmologists to directly analyze OCT images because the data collected by OCT equipment is enormous, so there is an urgent need for automatic segmentation technology." The need of automatic segmentation technology is not due to difficulty in OCT image analysis, but due to the potential segmentation error using OCT built-in software. Besides, the more important thing is to accurately quantify the segmented layers and provide metrics to evaluate the structural changes.

Response 1:

Thank you for your valuable comment. We have revised the abstract to clarify the content of the article. In addition, metrics are added to quantify the model segmentation results and edge structure changes.

Line 3: "Accurate retinal layer segmentation assists ophthalmologists in improving the efficiency of disease diagnosis. Deep learning technology is one of the most effective methods for processing OCT retinal layer images, which can segment different retinal layers and effectively obtain the topological structure of the boundary. This paper proposes a neural network method for retinal layer segmentation based on CSWin Transformer(CTS-Net), which can achieve pixel-level segmentation and obtain smooth boundaries. A Dice loss function based on boundary areas (BADice Loss) is proposed to make CTS-Net learn more features of edge regions and improve the accuracy of boundary segmentation. We applied the model to the publicly available data set of glaucoma retina, and the test results showed that mean absolute distance(MAD), root mean square error(RMSE), and Dice-Similarity Coefficient(DSC) metrics were 1.79 pixels, 2.15 pixels, and 92.79%, respectively, which were better than the model compared. In the cross-validation experiment, the range of MAD, RMSE, and DSC are 0.05 pixels, 0.03 pixels, and 0.33%, respectively, with a slight difference, which further verifies the generalization ability of CTS-Net."

Reviewer Comment 2:

Same comment on the Introduction as on the Abstract. The authors didn't describe the importance of OCT segmentation and the limitation of existing methods.

Response 2:

Thank you for your valuable comment. We modified the introduction, described the importance of OCT segmentation, and analyzed the limitations of existing methods.

Line 30: "OCT retinal layer image segmentation results are significant for thickness[4,9] and surface shape analysis[10]. However, images collected by OCT technology are usually rough due to noise, and the layer boundary is unclear. Ophthalmologists must carefully analyze retinal OCT images to identify the retinal layer and its edge, which usually takes plenty of time. From the above, there is an urgent need for OCT automatic segmentation technology."

Line 38: "Graph-based methods [11-13], which only use hand-designed features, are vulnerable to the noise and distortion of OCT images. In recent years, deep learning has developed rapidly and is widely used in the medical field [14]. The U-shaped[15] neural network framework based on convolutional neural network (CNN) is the most widely used in ophthalmic retinal layer segmentation [16], such as [17] uses a fully convolutional neural network to predict each OCT image pixel and then extract the edge, [18] divides the pixel into ten categories and applies the shortest path to obtain the edge of each layer,[19]use ResUnet to directly predict the retinal layer category of each OCT image's pixel and each layer's edge position. Although the CNN-based model achieves good performance, it cannot learn the interaction of semantic information between global and long-range due to the limitation of the convolution operation. On the contrary, the model constructed based on Transformer has the ability of global modeling, and its performance surpasses the model based on CNN [20]. Among them, the modified model based on Transformer such as SwinTransformer [21], and CSWin Transformer [22] have outstanding performance in terms of accuracy. The model based on the Transformer achieves satisfactory performance, but there are fewer applications in OCT images in retinal layer segmentation."

Reviewer Comment 3:

Line 86: "The underlying retinal images are generally grayscale images collected through X-ray, CT, MRI, and OCT technologies." What do the authors mean? Retinal images are definitely not collected through X-ray, CT, MRI.

Response 3:

Thank you for your valuable comment. We apologize for this omission, we have corrected them in the paper.

Line 83: "In the images of the retina layer collected by OCT technology[5], the change of gray intensity depicts the tissue characteristics of different retina layers..."

Reviewer Comment 4:

Line 232: The authors should give more information on the OCT dataset and the ground truth labelling as I mentioned before.

Response 4:

Thank you for your valuable comment. We added a description of the number of patients with peripapillary OCT and glaucoma in the data set.

Line 229: "We utilized 100 sample data obtained by peripapillary circular scanning, half of which were glaucoma patients. The label image size is 800×1000, with segmentation marks of the retinal nerve fiber layer (RNFL), ganglion cell plexiform layer(GCIPL), and choroid layer (CL)."

Reviewer Comment 5:

Line 270 and Line 342: Only 100 images were used for the training, without comparison with a model trained with large dataset or testing on unseen dataset. It is not scientific enough to claim that it can achieve outstanding performance on small datasets. It could be overfitted or maybe a model with large sample size can perform even better.

Response 5:

Thank you for your valuable comment. We apologize for this omission , We have added a 5-fold cross-validation technology experiment in the paper to further evaluate the robustness of our method.

Line 332: "4.5.3.Cross Validation

A stable and reliable model is essential, so we use the cross-validation method to evaluate the robustness of CTS-Net. We did a 5-fold cross-validation experiment. Based on the data set divided in 4.3, we split the 80 samples of the training set into five parts, one of which is the validation set, and the test set remains unchanged. Five models are obtained by training CTS-Net without repeated selection of the validation set, and the final results are shown in Tables 5 and 6."

Reviewer Comment 6:

The discussion part should have more in-depth discussion of comparison with other studies/methods, the strengths and limitations of the proposed method.

Response 6:

Thank you for your valuable comment. We modified the discussion section, compared our method with other models, and pointed out the advantages and disadvantages of our method.

Line 345: "5.Discussion

Retinal layer segmentation for glaucoma is challenging due to the high-resolution of the images involved and the high requirement for retinal layer edge segmentation. We designed CTS-Net based on the analysis of the cross-attention mechanism of the CSWinTransformer and the hierarchical characteristics of OCT retinal layer images. The proposed method is superior to the comparison model in MAD and RMSE metrics and less than the comparison model in Std, which shows that the test results of the model are relatively stable. Although the results of the BADL+DL loss function are similar to those of the Diceloss function (Table 4), when BADL is added, the MAD and RMSE metrics

of the test results of RelayNet, Swin-Net and CTS-Net models are reduced, which indicates that BADL can guide the model to improve the performance of edge segmentation. It also shows that BADL can be used as a general loss function for image segmentation.

Our work also has some limitations. The proposed method must preprocess the rectangular image into a square before it can be used. However, [19,35,36] can directly input the rectangular image into the neural network. This problem can be solved by fine-tuning the CSWin Transformer structure to make it available for rectangular images. Our method is limited to 2D images and cannot process 3D data with rich information, which may be our future research work."

We hope the corrections could meet your approval. If there is any progress, please do not hesitate to contact us. Thank you again to your review.

Best regards,

The authors