Peripapillary Atrophy Segmentation in Fundus Images via Multi-task Learning

Xiao Wei[a,b], Bo Jiang[a,*], Yuye Ling[b,*], Peiyao Jin[c], Yifan Wang[b], Xinbing Wang[b], and Chenghu Zhou[d]

[a]John Hopcroft Center for Computer Science, Shanghai Jiao Tong University, Shanghai, China
[b]Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China
[c]Department of Ophthalmology, Shanghai General Hospital, Shanghai Jiao Tong University, Shanghai, China
[d]Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

ABSTRACT

Peripapillary atrophy (PPA), a type of aberrant retinal symptom frequently present in older individuals or people with myopia, might indicate the severity of glaucoma or myopia. It is particularly beneficial for diagnosis when PPA is segmented effectively in fundus images. Deep learning is now frequently used for PPA segmentation. However, previous segmentation algorithms frequently mix up PPA with its neighboring tissue, the optic disc (OD), and generate the incorrect PPA area even though PPA is not present in the fundus image. To address these problems, we propose an improved segmentation network based on multi-task learning by combining detection and segmentation of PPA. We analyze the shortcomings of widely used loss functions and define a modified one to guide the training process of the network. We design a three-class segmentation task by introducing the information of OD, forcing the network to learn the difference of characteristics between OD and PPA. Evaluation on a clinical dataset shows that our method achieves an average Dice coefficient of 0.8854 in PPA segmentation, outperforming UNet and TransUNet, two state-of-the-art methods, by 24.4% and 10.6%, respectively.

Keywords: Peripapillary atrophy segmentation, Optic disc, Dice loss, Multi-task learning

1. INTRODUCTION

Peripapillary atrophy (PPA) is the atrophy of the retina and choroid. Fundus images are ocular records that can reflect the structure of the retina. The macula, vessels, optic disc (OD), and optic cup are among the principal structures that can be seen. Many eye diseases present abnormal symptoms on fundus images. For example, the shape and size of PPA in fundus images can indicate the progression of eye diseases such as glaucoma and pathological myopia. However, manual extraction of PPA areas in fundus images is time-consuming, so an automatic and accurate segmentation of PPA can assist doctors in diagnosing diseases.

In recent years, deep learning models have shown excellent performance in medical image segmentation, such as UNet.[1] The follow-up works are primarily separated into three groups: a better model structure, combining complemental data, and appropriate optimization objectives that fit the actual tasks. Various studies have designed novel structures to extract richer and more comprehensive features. Due to the inherent locality assumption in convolution operations, CNN-based algorithms do not effectively model explicit long-range relations. SwinUNet[2] and TransUNet[3] merged Transformer[4] into UNet to address this issue. HBA-UNet[5] proposed a hierarchical bottleneck attention to highlight retinal abnormalities. Some studies consider introducing additional data to enhance the feature learning process of the model. He et al.[6] and Li et al.[7] introduced boundary masks to smooth continuous surfaces. By combining a refined network trained by manually corrupted ground-truth mask, Batra et al.[8] improved the connectivity of segmentation. Most methods use Dice loss[5] instead of cross-entropy (CE) loss, or a hybrid loss[3] to solve the problem of class imbalance in segmentation. To solve the problem of data imbalance, Salehi et al.[9] proposed Tversky loss. However, the current segmentation performance still suffers from irregular borders, since OD is adjacent to PPA and their border is blurred. Furthermore, in non-diseased fundus images, PPA might not be present. Another issue in PPA segmentation is how to prevent forecasting the PPA area in healthy instances.

*Corresponding authors: Bo Jiang (bjiang@sjtu.edu.cn) and Yuye Ling (yuye.ling@sjtu.edu.cn).

Medical Imaging 2023: Image Processing, edited by Olivier Colliot, Ivana Išgum,
Proc. of SPIE Vol. 12464, 124641C · © 2023 SPIE
1605-7422 · doi: 10.1117/12.2652897

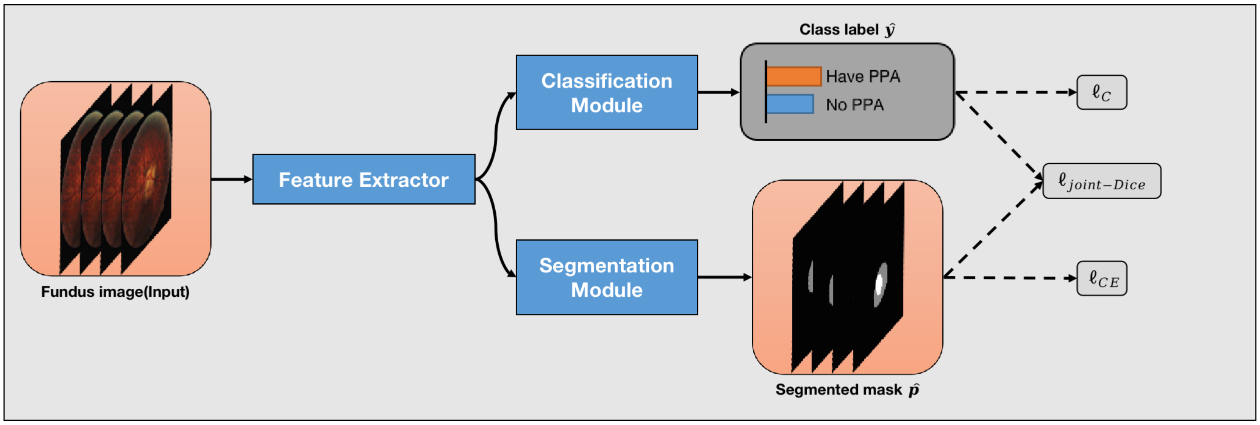

Figure 1. The schematic of the proposed method. The input is a fundus image, then the features obtained from the feature extractor are sent into the classification and segmentation module separately. The whole network output class label and segmented mask to compute the loss of different tasks.

To address the problem above, we analyze the loss function in the training stage and find that Dice loss implicitly expects that all samples have the segmentation target. This results in incorrect segmentation output for non-target images. For samples without PPA, Dice loss cannot accurately assess the segmentation effect of the model. This phenomenon does not occur in OD segmentation because OD is the fundamental tissue and always exists in fundus images. The blurred edge of the segmentation target often appears in medical images, since the tissues or organs are highly similar. This leads to inaccurate segmentation close to the border.

In this paper, we design a new training method for PPA segmentation based on multi-task learning. And we propose a novel loss function based on Dice loss and enable it to converge effectively through a classification loss. Therefore, we can improve the over-segmentation problem of non-target samples while maintaining the existing segmentation performance among samples with the segmentation target. The information of OD is also introduced to ameliorate the effect of the blurred boundaries. The contributions of our approach can be summarized as follows: (1) We tackle the barriers that existing methods produce over-segmentation in non-PPA samples and segmentation errors in boundary pixels. (2) We theoretically analyze the problems of existing methods and propose a novel multi-task learning method. (3) We show state-of-the-art performance on PPA segmentation on a clinical dataset, especially with at least a 10.6% improvement over other methods.

2. METHOD

Current segmentation methods produce over-segmentation in non-PPA samples and are prone to segmentation errors of pixels at the border. To address these issues, we propose a new training method based on joint-Dice loss and multi-task learning, and adopt OD information to avoid the prediction error of pixels at the edge. Among them, the joint-Dice loss can improve the segmentation performance, especially for samples without PPA, and multi-task learning is the basis for ensuring the effectiveness of joint-Dice loss. The model learns to categorize every pixel into the background, OD, and PPA instead of explicitly segmenting background and target, which is advantageous for learning discriminable characteristics. In the sections that follow, we first introduce the network used, then we clarify the Dice loss constraint, and then we propose a brand-new end-to-end training technique.

2.1 Framework

The schematic of the proposed method is illustrated in Fig. 1. Specifically, it consists of a feature extraction network, a classification module, and a segmentation module. Given a fundus image as input, the feature extractor can extract the high-level features for classification and the multi-resolution features for segmentation. After passing the high-level features to the classification module, it can decide if an image contains PPA or not by generating a classification probability. The classification probability is used to calculate the loss ℓ_C for the

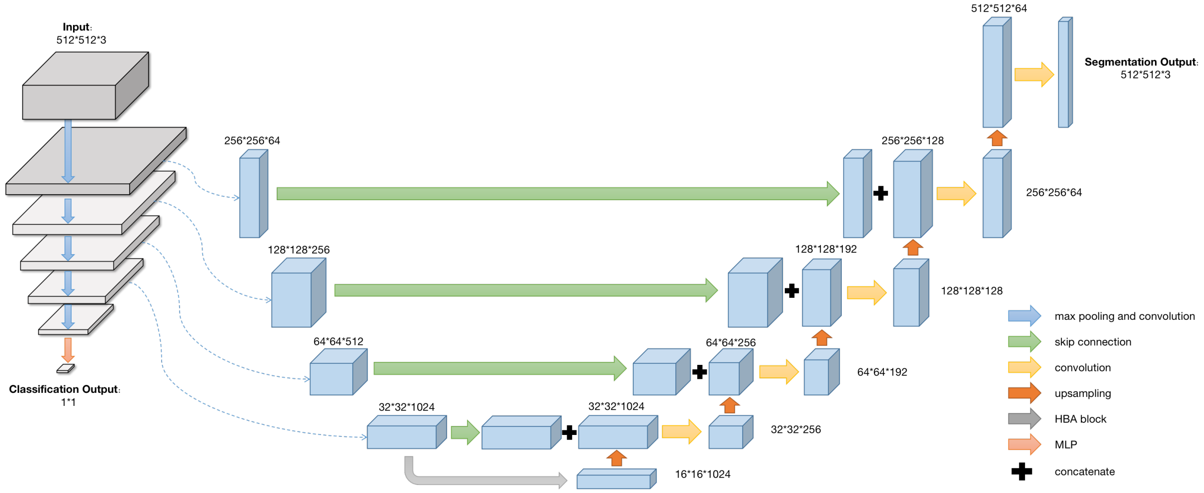

Figure 2. Detailed network structure. The network is a residual U-Net with ResNet50 as encoder and five upsampling blocks which consist of the residual blocks and convolution layers. The final upsampling block removes residual operation and is connected with a softmax layer to generate segmentation probabilities.

classification task. The multi-resolution features are sent into the segmentation module to obtain the predicted PPA area. Based on the classification probabilities and segmentation labels output by the model, we can calculate the proposed loss $\ell_{joint-Dice}$ for the segmentation task. Simultaneously, we adopt Cross-Entropy loss ℓ_{CE} to guide the segmentation task.

Fig. 2 indicates the detailed structure of our network. HBA-UNet[5] is the foundation upon which the network architecture is developed. We employ a mechanism of multi-task learning and modify the output layer to introduce OD information. The feature extraction network adopts ResNet50[10] as the backbone. Fully connected layers are connected with the feature extraction network to generate category probabilities, as the classifier. Multi-resolution feature maps are fed into the segmentation module, which is a typical UNet-like decoder[1] . To obtain a richer feature representation, the model can integrate shallow characteristics with deep features through skip connections. And the deepest feature map is passed into hierarchical bottleneck attention (HBA) blocks to highlight retinal abnormalities that may be beneficial for PPA segmentation. The HBA block adopts three attention mechanisms: channel, content, and relative-position attention. Then, by a series of upsampling processes, the hidden features from different scales are decoded, and various resolution features are aggregated. The convolution process is then employed to merge the concatenated features and alter the channel size. Finally, the output of the final upsampling layer is sent into the softmax layer to generate three-category segmentation probabilities for PPA and OD segmentation.

2.2 Loss Function

Without loss of generality, we perform the analysis under binary segmentation, and the formula can be transferred to multi-class segmentation. The Dice coefficient is a statistic proposed to measure the similarity of two sets. Dice loss is inspired by the Dice coefficient. We write the Dice loss for binary classification below:

$$\ell_{Dice} = 1 - \frac{2\sum_i^N p_i g_i + \varepsilon}{\sum_i^N p_i + \sum_i^N g_i + \varepsilon}, \qquad (1)$$

In the above, $g_i \in \{0,1\}$ specifies the label of ground-truth of pixel i and $p_i \in [0,1]$ denotes the model's predicted probability, where N indicates the total number of pixels in the image. In practice, ε is employed to guarantee the denominator is not equal to 0. Since in the training stage, each g_i equals 0 for samples without segmentation target, and p_i belongs to $[0,1]$. Considering this term $\sum_i^N p_i$, the order of magnitude difference between correctly and incorrectly segmented results is small. The value of ε is often small, which leads to the

Table 1. PPA segmentation for different training methods tested on the clinical dataset. $DICE_{ppa}$: Dice coefficient of PPA for whole samples in testing data, Excluded $DICE_{ppa}$: Dice coefficient of PPA excluded non-PPA samples, ACC: the accuracy whether the PPA is detection correctly, Sen: the sensitivity of the detection results, Spe: the specificity of the detection results. Values in parentheses represent standard deviation.

Network	$DICE_{ppa}$	Excluded $DICE_{ppa}$	Acc	Sen	Spe
UNet[1]	0.7115 (0.3363)	0.8456 (0.1319)	0.8372	**1.0**	0.1250
R50 UNet[1]	0.7959 (0.3056)	0.8473 (0.1401)	0.8837	**1.0**	0.3750
Att-UNet[13]	0.6988 (0.3123)	0.8157 (0.1075)	0.8488	**1.0**	0.1875
HBA-UNet[5]	0.7763 (0.2969)	0.8394 (0.1724)	0.9070	**1.0**	0.5000
TransUNet[3]	0.8002 (0.3104)	0.8974 (0.1186)	0.8721	0.9857	0.3750
Swin-UNet[2]	0.7236 (0.3319)	0.8604 (0.0986)	0.8372	**1.0**	0.1250
Transfuse[14]	0.6572 (0.3085)	0.7646 (0.1420)	0.8488	**1.0**	0.1875
Ours	**0.8854 (0.1860)**	**0.9020 (0.0786)**	**0.9651**	**1.0**	**0.8125**

Dice loss approaching 1 under both correct segmentation and under-segmentation, and it is difficult to separate the loss in these two cases by just adjusting the value of ε.

We propose an improved Dice loss that combines classification labels and probabilities. Based on the new objective function, the model can be forced to discriminate whether non-target samples are correctly segmented. We define the joint-Dice loss as:

$$\ell_{joint-Dice} = (y + \hat{y} - y\hat{y})\,\ell_{Dice}, \tag{2}$$

where $y \in \{0,1\}$ is the label reflecting the absence or presence of the segmentation target, and $\hat{y} \in [0,1]$ is the classification probability. For samples without segmentation target, $\ell_{joint-Dice}$ will be reducible to $\hat{y}$ since $\ell_{Dice} = 1$ and $y = 0$. Therefore, under the new definition, $\ell_{joint-Dice} > 0.5$ when correctly segmented, $\ell_{joint-Dice} < 0.5$ when incorrectly segmented. For samples with segmentation targets, $\ell_{joint-Dice}$ will degenerate to ℓ_{Dice}, so the model will be forced to improve segmentation performance.

To obtain the $\hat{y}$ above, we introduce a CE loss for another classification task, ℓ_C, which turns our optimization goal to multi-task learning. To accelerate model learning and maintain robustness, we employ a hybrid loss of CE loss, ℓ_{CE} , and the proposed loss, $\ell_{joint-Dice}$, in the segmentation task. Finally, we optimize the loss function as follows:

$$\ell_{total} = \lambda_1 \ell_{CE}\,(p,g) + \lambda_2 \ell_{joint-Dice}\,(p,g,\hat{y},y) + \lambda_3 \ell_C\,(\hat{y},y) \tag{3}$$

3. EXPERIMENTS AND RESULTS

3.1 Dataset and Implementation Details

The dataset we use is provided by Shanghai General Hospital, which contains 851 clinical data. This dataset is collected from an epidemiology survey, and OD and PPA areas are pre-labeled by experts.

The input images were resized to 512×512 pixels. We used CLAHE[11] to enhance the texture and contrast of fundus images. We perform flip and rotation augmentation to the input during training. We divided the dataset with 680 samples in the training set, 85 samples for validation and 86 samples for testing. The hyper-parameters in loss function in Equation (3) were set as $\lambda_1 = 0.25$, $\lambda_2 = 0.25$, $\lambda_3 = 0.5$. The ResNet50[10] backbone in HBA-UNet[5] was pre-trained on ImageNet[12] . Models were trained with Adam optimizer with a learning rate of 0.0001, a momentum of 0.9, and a weight decay of 0.0001. The batch size was set as 4 and the default number of training epochs was 500. The model was implemented using Keras and all experiments were conducted using a single NVIDIA Geforce RTX 3090 GPU.

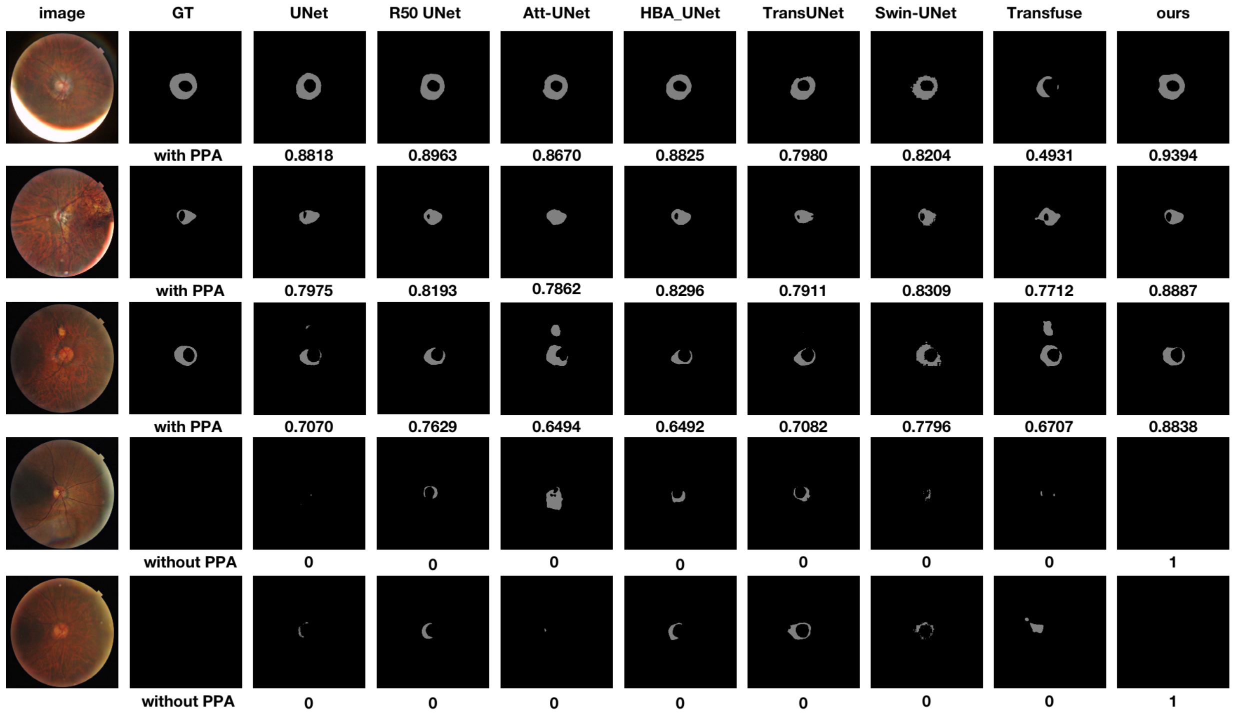

Figure 3. Visualization of several methods in PPA segmentation. The original fundus images and ground truth are in the first and second columns. The baselines and our method are in the next columns. We annotate the figure with the presence of PPA in fundus images and Dice score for each result map.

3.2 Results

To evaluate the performance of our method and compare it against prior segmentation algorithms, we compare the segmentation results from five compartments: Dice similarity coefficient of PPA, Dice similarity coefficient of PPA on testing data that samples without PPA are removed, the accuracy of whether the PPA is detected correctly, and the corresponding sensitivity and specificity. The state-of-the-art baselines we choose include: UNet[1] , R50 UNet,[1] Att-UNet[13] , HBA-UNet[5] , TransUNet[3] , Swin-UNet[2] , and Transfuse[14] . Among them, the first four methods are based on pure convolutional neural networks, and the last three fuse Transformers[4] and CNNs. All of them are trained on two-class segmentation tasks. The evaluation results are shown in Tab. 1. From observation, our method outperforms state-of-the-art algorithms. We obtain a Dice coefficient of PPA of 0.8854, by the improvement of 10.6% on TransUNet. The sensitivity is equal to 1 means that those methods segment the PPA region for all fundus images with PPA. Our method obtains the best score in specificity. This shows the applicability of the presented training method to the datasets including samples without segmentation targets. Additionally, we take the samples without PPA out of the test set, the Dice score demonstrates that our approach is continually improving. Though the loss function we designed primarily solves the segmentation problem of non-PPA samples, the suggested technique benefits from the addition of the OD label information.

We also show a visualization comparison of the segmentation results of the different training methods, as shown in Fig. 3. The top three rows in this figure show examples of images with noise, low-contrast images, and images with blurred edges respectively. Results suggest that our technique, when compared to other methods, can better segment PPA because the pixels near the boundary between PPA and OD are segmented more accurately. In order to illustrate the boost of our algorithm for non-PPA samples, we show the examples in the last two rows. Our approach produces accurate segmentation results, whereas other baselines typically provide erroneous segmentation masks even when PPA is absent from fundus images.

We conduct the ablation investigation to assess the effects of OD information, joint-Dice loss, and multi-task learning on PPA segmentation (see Tab. 2). The addition of OD information can enhance both the segmentation

Table 2. Ablation results. OD: Whether OD segmentation is combined, MTL: whether multi-task learning is adopted, Joint-Dice: whether modified Dice loss is used.

OD	MTL	Joint-Dice	$DICE_{ppa}$	Excluded $DICE_{ppa}$	Acc	Sen	Spe
×	×	×	0.7763 (0.2969)	0.8394 (0.1724)	0.9070	**1.0**	0.5000
√	×	×	0.7853 (0.3186)	0.8934 (0.1133)	0.8721	**1.0**	0.3125
√	√	×	0.8406 (0.2517)	0.8899 (0.1058)	0.9302	**1.0**	0.6250
×	√	√	0.8556 (0.2423)	0.8940 (0.1229)	0.9302	0.9857	0.6875
√	√	√	**0.8854 (0.1860)**	**0.9020 (0.0786)**	**0.9651**	**1.0**	**0.8125**

accuracy of samples with PPA as well as the model's ability to discriminate between samples with and without PPA. The multi-task learning technique mainly brings improvement in the correct segmentation of samples without PPA. The joint-dice loss is associated with multi-task learning, which can enhance the model's accuracy in segmenting data without PPA. In a word, the ablation experiments show that each part we employ improves the segmentation of PPA.

In addition, we provide further experiments to demonstrate the benefits of multi-task learning-based segmentation approaches over multi-step segmentation methods. We first train a ResNet50 network to detect the presence of PPA in fundus images. Then, for samples with PPA, the fundus image is submitted to the segmentation network to extract the PPA area, whereas for data without PPA, the segmentation map is a matrix of all zeros. TransUNet[3] which has the best performance among baseline. From Tab. 3, we observe that our proposed one-step method is better than the two-step method. This is because errors are passed cumulatively in two-step algorithms. The errors in the first step will further affect the segmentation results in the second step.

3.3 Discussion

We explore how the algorithm has improved based on the variation in training loss. There will be false negatives and positives when we directly train ResNet50, a classification network utilizing cross-entropy loss, to detect PPA on the fundus images. This is a result of the restricted amount of data and the fact that classification labels offer less information than segmentation labels. In contrast, the model will strive to strike a compromise between those tasks if we use multi-task learning, using Dice loss for segmentation and cross-entropy loss for classification. Although it can prevent the issue of false negatives, it will result in the over-segmentation of samples with no PPA. We first analyze the reduction of false negatives. These samples will be misjudged when only the classification task is adopted. With the addition of the segmentation task, the loss under the segmentation task will guide the model to learn more correct features. For some samples without PPA, the classifier can classify them correctly, but their Dice loss approaches 1. It will mislead the model and cause over-segmentation. This results in a drop in specificity from 0.8125 to 0.6250. In our training methods, we propose a new loss function to avoid the problem of Dice loss, so both segmentation and classification are improved.

Table 3. Comparison between one-step and two-step methods.

Description	Network	$DICE_{ppa}$	Acc	Sen	Spe
one-step	TransUNet	0.8002 (0.3104)	0.8721	0.9857	0.3750
two-step	TransUNet+ResNet	0.8719 (0.2245)	0.9419	0.9714	**0.8125**
one-step	**Ours**	**0.8854 (0.1860)**	**0.9651**	**1.0**	**0.8125**

4. CONCLUSIONS

In this work, we introduce a novel PPA segmentation method based on multi-task learning that incorporates PPA detection and segmentation, to address the challenge with non-target samples in the dataset for PPA segmentation. We also introduce OD information to force the network to distinguish between the PPA and OD. Our training method is simple and highly effective. We have confirmed its effectiveness through theoretical analysis and experimental illustration. The results show that it achieves state-of-the-art performance on a clinical dataset.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (No. 42050105, 62072302, 61905141, 61960206002).

REFERENCES

[1] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*International Conference on Medical image computing and computer-assisted intervention*], 234–241, Springer (2015).

[2] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., and Wang, M., "Swin-unet: Unet-like pure transformer for medical image segmentation," *arXiv preprint arXiv:2105.05537* (2021).

[3] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., and Zhou, Y., "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306* (2021).

[4] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I., "Attention is all you need," *Advances in neural information processing systems* **30** (2017).

[5] Tang, S., Qi, Z., Granley, J., and Beyeler, M., "U-net with hierarchical bottleneck attention for landmark detection in fundus images of the degenerated retina," in [*International Workshop on Ophthalmic Medical Image Analysis*], 62–71, Springer (2021).

[6] He, Y., Carass, A., Liu, Y., Jedynak, B. M., Solomon, S. D., Saidha, S., Calabresi, P. A., and Prince, J. L., "Structured layer surface segmentation for retina oct using fully convolutional regression networks," *Medical image analysis* **68**, 101856 (2021).

[7] Li, M., Zhao, H., Xu, J., and Li, H., "Peripapillary atrophy segmentation with boundary guidance," in [*International Workshop on Ophthalmic Medical Image Analysis*], 101–108, Springer (2021).

[8] Batra, A., Singh, S., Pang, G., Basu, S., Jawahar, C., and Paluri, M., "Improved road connectivity by joint learning of orientation and segmentation," in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 10385–10393 (2019).

[9] Salehi, S. S. M., Erdogmus, D., and Gholipour, A., "Tversky loss function for image segmentation using 3d fully convolutional deep networks," in [*International workshop on machine learning in medical imaging*], 379–387, Springer (2017).

[10] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 770–778 (2016).

[11] Reza, A. M., "Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement," *Journal of VLSI signal processing systems for signal, image and video technology* **38**(1), 35–44 (2004).

[12] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems* **25** (2012).

[13] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., et al., "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999* (2018).

[14] Zhang, Y., Liu, H., and Hu, Q., "Transfuse: Fusing transformers and cnns for medical image segmentation," in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], 14–24, Springer (2021).