A well-generalized deep learning framework towards glaucoma assessment

Chenglang Yuan[1,2], Cheng Bian[1], Munan Ning[1], and Yefeng Zheng[1]

[1] Tencent Jarvis Lab
{tronbian, masonning, yefengzheng}@tencent.com
[2] School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China
yuanchenglang@email.szu.edu.cn

Abstract. Glaucoma is the main cause of irreversible blindness worldwide, which is not easy to be noticed in the early state. Early diagnosis and treatment of glaucoma can prevent patients from potential vision loss. In this paper, we construct a united framework to implement the glaucoma classification, Optic Cup (OC)/Optic Disc (OD) segmentation and fovea localization. To ensure our model can be generalized to different devices, test time training (TTT) and unsupervised domain adaptation (UDA) strategies are utilized for the classification and segmentation frameworks. Experiments on the REFUGE2 [5] testset show that the proposed framework achieves 0.9738 in AUC of glaucoma classification, 87.49%/96.21% in Dice of OC/OD segmentation and gets the average Euclidean distance error of 8.67 pixels in fovea localization in the single model.

1 Introduction

Being the leading cause of irreversible vision loss, glaucoma is a serve chronic ocular disease with minimum symptoms which will cause progressively damage to the optic nerve. As the report in [6], more than 80 million people are affected by glaucoma in 2020 and such number will be expected up to 110 million people by the year 2040. The Cup-to-Disc Ratio (CDR) is a quantitative clinical measurement for ophthalmologists to evaluate the damage progress of glaucoma, which is calculated by the ratio of the vertical diameter of the Optic Cup (OC) to the vertical diameter of Optic Disc (OD). The higher the CDR is, the higher the risk of glaucoma. Yet, the measurement of CDR is principally based on the empirical estimation of ophthalmologists, which will lead to biased decisions by different experts. Moreover, performing manual CDR measurement case by case is both labor intensive and time-consuming that makes the large-scale diagnosis scenario impractical. Therefore, there is an urgent need for an automatic and efficient solution for OC and OD segmentation or direct CDR prediction.

With the superiority of performance compared with traditional methods, deep learning techniques have dominated in most medical tasks. Recent works on computer-aided diagnosis of glaucoma can be simply classify into three deep learning methods, which are: glaucoma classification [1,4], OC and OD segmentation [12,13] and CDR estimation [3,16]. Although the proposed deep models from those works have addressed specific tasks on glaucoma properly, those well-trained models' performances are prone to collapse when being evaluated on multiple centers or devices. Thus, more and more researchers are paying attention to the importance of the generalization of the deep models.

As the second competition of REFUGE, REFUGE2 [5] provides three tasks of challenges to raise the researcher's interest in the study of glaucoma Computer-Aided Diagnosis (CAD) system, which includes: the classification of glaucoma, the segmentation of OC/OD and the localization of fovea. Meanwhile, REFUGE2 additionally requires challengers to present generalized models to address multiple devices scenario. To this end, this paper proposes a united architecture to achieve glaucoma classification, OC/OD segmentation and fovea localization with the capability of device adaptation.

2 Methodology

An overview of the proposed glaucoma CAD system is shown in Fig. 1. Following the rules of REFUGE2, our system can be divided into three frameworks (segmentation, localization and classification). For the classification and segmentation frameworks, we apply a two-stage approach to achieve better performance. Specifically, we first acquire the coarse segmentation mask with a standard UNet [8]. Secondly, a region of interest (ROI) on the optical nerve head of the original images is cropped namely as stage1. Then, Those ROIs are selected as inputs of the classification and segmentation frameworks namely as stage2. To ensure our framework can be well generalized to multiple devices, two strategies are employed to the proposed classification and segmentation frameworks respectively, including test time training (TTT), and unsupervised domain adaptation (UDA). More details will be demonstrated in the following sections.

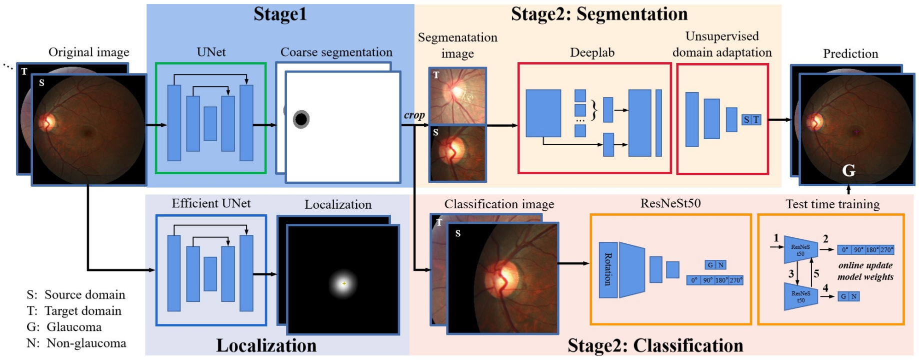

Fig. 1: An overview of the proposed framework.

2.1 Glaucoma Classification

For our classification network, Resnest50 [15] is chosen as our baseline. To increase the generalization capability of the classification network for different devices, we propose to integrate the test-time training (TTT) strategy to our framework, which was first proposed by Sun *et al.* [9]. Unlike current transfer learning methods employing adversarial training or fine-tuning strategies in the training stage, the concept of TTT is to enforce the framework to optimize itself with test data in the inference stage which is served as a plug-to-play strategy for model generalization. The key step for TTT is to construct a self-supervised auxiliary task for our original L layers baseline. Here, we select the rotation prediction as the auxiliary task, which had been evaluated the efficacy on improving the performance in the previous works. Suppose we have N samples training dataset $\mathcal{D}_s = \{(x_s^i, y_s^i, y_p^i)\}_{i=1}^N$, where x_s, $y_s \in [0,1]$ and y_p denote the input image, original task annotation and the auxiliary task annotation, respectively. We set the annotation of the proxy task $y_p \in [0,1,2,3]$ to represent the rotation of [0°,90°,180°,270°]. Denote that the proposed classification framework with l shared layers. Then, the parameter of the main branch can be written as $\theta_m = (\theta_1, \theta_2, \ldots, \theta_l)$, and the parameter of the task-specific branch can be regarded as $\theta_s = (\theta_{l+1}, \theta_{l+2}, \ldots, \theta_L)$. Notably, the auxiliary branch utilizes independent parameters, which is $\theta_p = (\theta'_{l+1}, \theta'_{l+2}, \ldots, \theta'_L)$. Hence, the objective loss function can be formulated as:

$$\min_{\theta_m, \theta_s, \theta_p} \frac{1}{N} \sum_{i=1}^{N} \left(l_b(x_s^i, y_s^i; \theta_m, \theta_s) + l_p(x_s^i, y_p^i; \theta_m, \theta_p) \right), \tag{1}$$

where l_p and l_s are the binary cross-entropy and softmax cross-entropy loss functions of the main and auxiliary tasks, which can be viewed as the multi-task learning in the training stage. Once the training stage is finished, the TTT strategy is applied to the framework for each single test sample x. We use the same learning rate and optimizer of the training stage to fine-tune the main and auxiliary branches, which can be formulated as:

$$\min_{\theta_m} \frac{1}{4} \sum_{y'_p=0}^{3} l_p(rotate(x, y'_p * 90°), y'_p; \theta_m, \theta_p). \tag{2}$$

Four degrees rotation $rotate(\cdot)$ will be performed for each x. The auxiliary task loss function is optimized by the augmented inputs together with the auxiliary annotations y'_p. The weights of the auxiliary branch should be fixed to ensure the main branch being fine-tuned properly. Finally, we can make the predictions from the task-specific branch after applying the TTT strategy. It is worth noting that θ_m must be re-initiated to θ_m^* before the TTT strategy, which is the optimal weights in the training stage.

2.2 Optical Cup and Disc Segmentation

As shown in Fig. 1, having acquired the ROIs from stage1, we employ DeeplabV3+ [2] framework to achieve the precise segmentation of OC/OD. Noticed that two independent heads are employed on both stages frameworks to estimate OC and OD individually, which can be viewed as two independent binary tasks. A hybrid loss is chosen as the segmentation loss to supervise the OC/OD prediction from these heads:

$$\begin{aligned}\mathcal{L}_{hybrid} =& -\frac{1}{K}\sum_{k=1}^{K}(y_s(k)\log p_s(k) + (1-y_s(k))\log(1-p_s(k))) \\ &+ (1-\sum_{k=1}^{K}\frac{2y_s(k)p_s(k)}{y_s^2(k)+p_s^2(k)})\end{aligned} \tag{3}$$

where k iterates over all locations and over all ground truths y_s of OC/OD and prediction masks p_s from the training set, and K is the total number of iterations. In order to maintain the segmentation performance on different devices, a classical UDA strategy [11] has been adopted. The underlying principle of this strategy is to propose a discriminator $\mathbf{D}$ to identify what datasets the predictions are derived from, which can be formulated as:

$$\mathcal{L}_D = -\frac{1}{N}\sum_{n=1}^{N} z\log\mathbf{D}(p_s) - \frac{1}{M}\sum_{m=1}^{M}(1-z)\log(1-\mathbf{D}(p_t)), \tag{4}$$

$$\mathcal{L}_{adv} = -\frac{1}{M}\sum_{m=1}^{M}\log\mathbf{D}(p_t), \tag{5}$$

where M is the total size of the testset; p_t is the prediction mask of testset sample; z is domain indicator, when $z = 1$ denotes the prediction from the training dataset whereas $z = 0$ denotes the prediction from the testing dataset.

2.3 Fovea Localization

In the fovea localization task, we decide to utilize the classic keypoint detection technique [14, 17], which is frequently used in medical detection challenges in recent years with a potential to capture tiny regions accurately. In this regard, we perform the fovea localization based on an EfficientUNet [10], where EfficientNet-b5 is utilized as the feature extractor. Formally, the objective function of the localization task can be defined as:

$$\mathcal{L}_{KP} = \frac{1}{K}\sum_{k=1}^{K}(y_s(k) - p_s(k))^2, \tag{6}$$

where y_s is the standard keypoint Gaussian map of fovea; p_s is the the framework's prediction; k iterates over all locations and over all y_s and p_s; K is the total number of iterations. Afterwards, the predicted location of fovea can be calculated by an *argmax* operation of p_s in the inference stage.

Table 1: Implementation datails of the proposed frameworks.

Configuration	**Classification**	**Segmentation**	**Localization**
batch size	144	24	7
learning rate	0.01	0.001	0.001
optimizer	SGD	Adam	Adam
cross-validation	10 folds	10 folds	10 folds
augmentation*	1,2	3,4,5	3,4,5
test time aug.	Yes	Yes	Yes

augmentation modes: **1**/random crop; **2**/horizontal flip; **3**/random flip; **4**/random rotation; **5**/random scale.

3 Experiments

3.1 Implementation details and data processing

We adopt different pre-processing operations on the three frameworks. For the segmentation framework, we crop the ROIs from stage 1 predictions with the size of 2 times that of the disc's diameter and resize the ROIs to 512×512. While the classification framework, the ROIs are calibrated as the size of 2.5 times that of the disc's diameter and resized to 256×256. As to the localization framework, all samples will be directly resized down to 512×512. The radius kernel of the keypoint Gaussian map is set to 20 pixels. Both ROIs are applied to the data normalization operation via subtracting minimum value and dividing the difference values of maximum and minimum. All proposed framework is implemented in PyTorch using an NVIDIA Tesla V100 GPU. The experimental configuration of three frameworks are listed in Table 1.

Table 2: Quantitative experiment of the proposed classification framework.

Method	**AUC**
ResNeSt200	0.9533
ResNeSt101	0.9589
ResNeSt50	0.9655
ResNeSt50+test time aug.	0.9708
ResNeSt50+test time aug.+TTT	**0.9738**

3.2 Quantitative and Qualitative Analysis:

We evaluate the effectiveness of the proposed frameworks on the REFUGE2 online test dataset. To compare the classification framework's performance quantitatively, Area Under Curve (AUC) is adopted as our evaluation metric. For the segmentation framework, we utilize Dice coefficient (Dice) and Cup-to-Disc Ratio (CDR) in the evaluation. While the localization framework, mean Euclidean

Table 3: Quantitative experiment of the proposed segmentation framework.

Method	OC DSC	OD DSC	CDR
UNet [8] (stage 1 Only)	0.8397	0.9540	0.043
Ours (stage 1+2)	0.8619	0.9510	0.040
Ours +test time aug.	0.8699	0.9585	0.039
Ours +test time aug.+UDA	**0.8749**	**0.9621**	**0.037**

distances (MED) is chosen as our criteria. All evaluation results are reported from the online leaderboard. A quantitative experiment on classification is listed in Table 2. Results of different baselines show that the RestNeSt50 performs better than other deeper architectures. When applying the test time augmentation together with TTT strategy, we achieve around 0.83% improvement in AUC. From Table 3, it can be observed that the UDA strategy effectively boosts the online testset performance with 87.49% in Dice score of OC and 96.21% Dice score of OD, respectively, and obtains the lowest CDR error, achieving the best performance across all the other compared methods. We also conduct another experiment to compare the proposed localization with the state-of-the-art detection baselines. Specific details are presented in Table 4. Obviously, the results verify that our localization framework surpasses the frontier baselines with a large margin. Fig. 2 visualize several representative predictions by the proposed framework.

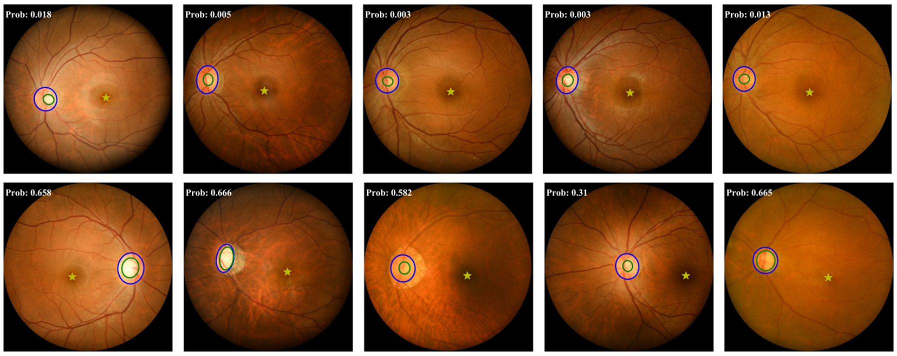

Fig. 2: Examples of framework's prediction of glaucoma assessment. The classification probabilities are presented at the top-left. Blue and green contours represent the boundaries of OD and OC, while the yellow stars denote the fovea location.

Table 4: Quantitative experiment of the proposed localization framework.

Method	**MED**
YoLoV3 [7]	16.43
UNet [7]	12.55
EfficientUNet [10]	10.67
EfficientUNet+test time aug.	**8.67**

4 Conclusions

In this study, we proposed a united framework to realize glaucoma assessment, which consists of the glaucoma classification, OC/OD segmentation and fovea localization. Specifically, we employed TTT and UDA strategies for the classification and segmentation tasks to improve the model generalization. Experiments on the online testset varied the efficacy of the proposed framework.

References

1. Chakravarty, A., Sivaswamy, J.: Glaucoma classification with a fusion of segmentation and image-based features. In: 2016 IEEE 13th international symposium on biomedical imaging (ISBI). pp. 689–692. IEEE (2016)
2. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 801–818 (2018)
3. Ferreira, M.M., Esteve, G.P., Junior, G.B., de Almeida, J.D.S., de Paiva, A.C., Veras, R.: Multilevel cnn for angle closure glaucoma detection using as-oct images. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP). pp. 105–110. IEEE (2020)
4. Gómez-Valverde, J.J., Antón, A., Fatti, G., Liefers, B., Herranz, A., Santos, A., Sánchez, C.I., Ledesma-Carbayo, M.J.: Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning. Biomedical optics express 10(2), 892–913 (2019)
5. Orlando, J.I., Fu, H., Breda, J.B., van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P.A., Kim, J., Lee, J., et al.: Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Medical image analysis 59, 101570 (2020)
6. Quigley, H.A., Broman, A.T.: The number of people with glaucoma worldwide in 2010 and 2020. British journal of ophthalmology 90(3), 262–267 (2006)
7. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
8. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
9. Sun, Y., Wang, X., Liu, Z., Miller, J., Efros, A.A., Hardt, M.: Test-time training with self-supervision for generalization under distribution shifts. In: International Conference on Machine Learning (ICML) (2020)

10. Tan, M., Le, Q.V.: Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946 (2019)
11. Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7472–7481 (2018)
12. Wang, S., Yu, L., Li, K., Yang, X., Fu, C.W., Heng, P.A.: Boundary and entropy-driven adversarial learning for fundus image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 102–110. Springer (2019)
13. Wang, S., Yu, L., Yang, X., Fu, C.W., Heng, P.A.: Patch-based output space adversarial learning for joint optic disc and cup segmentation. IEEE transactions on medical imaging 38(11), 2485–2495 (2019)
14. Yuan, C., Bian, C., Kang, H., Liang, S., Ma, K., Zheng, Y.: Identification of primary angle-closure on as-oct images with convolutional neural networks. arXiv preprint arXiv:1910.10414 (2019)
15. Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Zhang, Z., Lin, H., Sun, Y., He, T., Mueller, J., Manmatha, R., et al.: Resnest: Split-attention networks. arXiv preprint arXiv:2004.08955 (2020)
16. Zhao, R., Chen, X., Liu, X., Chen, Z., Guo, F., Li, S.: Direct cup-to-disc ratio estimation for glaucoma screening via semi-supervised learning. IEEE Journal of Biomedical and Health Informatics 24(4), 1104–1113 (2019)
17. Zhou, X., Wang, D., Krähenbühl, P.: Objects as points. arXiv preprint arXiv:1904.07850 (2019)