Layer Segmentation and Glaucoma Diagnosis in Optical Coherence Tomography Images: A Technical Report for GOALS Challenge

Juntao Jiang, Qi Wei

Zhejiang University, Shandong University

Abstract. Optical coherence tomography (OCT) is widely used in ophthalmology for diagnosis and treatment of eye diseases. OCT images can provide cross-sectional information of the retina structures to achieve more accurate analysis compared with color fundus images. Segmentation and thickness quantification of layers are helpful in the diagnosis of many retinal and optic nerve diseases. This paper is a technical report for the MICCAI 2022 Challenge: GOALS, introducing our approach to the task of layer segmentation of images and the task of automatic diagnosis of glaucoma. We applied ResNet34 as the model for glaucoma automatic diagnosis task, and HR-Net18+OCRNet as the model for layer segmentation task. Our final score was 8.8044, ranking 15^{th} in the preliminary round of the competition.

Keywords: OCT · layer segmentation · glaucoma automatic diagnosis · ResNet34 · HR-Net18 · OCRNet · GOALS.

1 Introduction

Optical coherence tomography (OCT) is a non-contact, non-invasive and high-resolution biomedical imaging technology, which can collects optical backscattering signal for cross-sectional and volumetric imaging of the biological tissues [1], such as blood flow, polarization status, structure data, elastic parameters, and molecular content [2] and has been widely used in ophthalmology [3, 4], cardiology [5, 6], endoscopy [7, 8], dermatology [9, 10] and oncology [9, 10]. In OCT-scanned eye images, the retina contains several layers of varying thickness and intensity. Segmenting and measuring the thickness of each layer can be used to be utilized as an indicator retinal health [11].

Glaucoma is the second leading cause of blindness [12]. It can be indicated by both thickness of NFL layer in OCT and the Euclidian distance between NFL and the Inner Limiting Membrane (ILM). Deep learning-based classification methods for OCT analysis has been a powerful tool for glaucoma diagnosis [13–16]. Meanwhile, deep learning-based layer segmentation [17–22] also facilitates more accurate diagnosis.

GOALS Challenge [23] is held as a part of the 9th Ophthalmic Medical Image Analysis Workshop (OMIA9) at MICCAI 2022 and the two subtasks

are: 1) A segmentation task to determine the retinal nerve fiber layer, ganglion cellinner plexiform layer, and choroidal layer, which are helpful for diagnosis and differentiation of glaucoma; 2) an automatic diagnosis task of glaucoma.

In this technical report, we will give specific details of our methods used in these two tasks. For the segmentation task, we applied OCRNet [24, 25] with HRNet-18 backbone [26]. To solve sample imbalance problems, we applied OHEM cross entropy loss [27], Dice loss [28], Lovász-Softmax loss [29] and Relax Boundary Loss [30]. For the classification task, the ResNet34 [31] model was used.

2 Challenge Tasks

The challenge includes 2 tasks: 1) OCT Layer Segmentation task and (segmentation task) 2) Glaucoma Detection (classification task).

2.1 Dataset

The classification task and the segmentation task share a joint datset. The dataset [23] released by GOALS was provided by Sun Yat-sen Ophthalmic Center, Sun Yat-sen University, Guangzhou, China, which contains 300 Circumpapillary Optical Coherence Tomography (Circumpapillary OCT). The data was split into three equal groups for the training process, the preliminary competition process and the final process respectively. Groudtruths for both tasks are provided in this dataset.

For the segmentation task, after initial annotation by 5 clinical ophthalmologists, rechecking and fusion by more senior ophthalmologist from Zhongshan Ophthalmic Center of Sun Yat-sen University, China, different pixel values were assigned to the pixels inside the edges of RNFL, GCIPL and choroid layer region.

For the classification task, the label of Glaucoma and non-Glaucoma was assigned, which was determined by clinical records and the results of all clinical examinations.

2.2 The Segmentation Task

The challenge task is designed to segment three layers, which have positive significance for the diagnosis of glaucoma, including RNFL, GCIPL, and choroid layer in OCT images. The golden standard of OCT layer segmentation was saved in PNG image format with an mask image of segmentation result. In the segmented image, the pixel value of 0 represents the RNFL, the pixel value of 80 represents the GCIPL, the pixel value of 160 represents the choroid, and the pixel value of 255 represents other areas.

Evaluation Metrics The segmentation task of the challenge was evaluated by DICE coefficient evaluation and mean Euclidean distance (MED).

The DICE coefficient is defined as:

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|}$$

where, X represents the segmented target pixel point set in the ground truth; Y represents the segmented pixel point set in the prediction result; $|X \cap Y|$ represents the intersection between X and Y; $|X|$ and $|Y|$ represent the number of the elements of X and Y.

The mean Euclidean distance (MED) is defined as:

$$MED = \frac{1}{N} \sum_{i=1}^{N} \sqrt{(x_i - x_i^0)^2 + (y_i - y_i^0)^2}$$

where, N is the number of pixels on the edge of the segmentation result, (x_i, y_i) is the i th pixel on the edge of the segmentation, and $\left(x_i^0, y_i^0\right)$ is the nearest pixel on the gold standard edge to (x_i, y_i)

The final segmentation score $Score_{task1}$ will be given from a weighted score from three regions:

$$Score_{task1} = 0.4 \times Score_{RNFL} + 0.3 \times Score_{GCIPL} + 0.3 \times Score_{Choroid}$$

$$Score_{region} = 0.5 \times Score_{Dice_{region}} + 0.5 \times Score_{MEDregion}$$

$$Score_{Dice_{region}} = Dice \times 10$$

$$Score_{MED_{region}} = (MED + 1)^{-0.3}$$

where, $region \in (RNFL, GCIPL, Choroid)$

2.3 The Classification Task

The classification task was evaluated by F1 score, accuracy, sensitivity, and specificity. They are defined as follows respectively:

$$F1 = \frac{2 \times P \times R}{P + R} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sen = \frac{TP}{TP + FN}$$

$$Spe = \frac{TN}{TN + FP}$$

where, P and R represent precision and recall of the detection results among the testing images. TP, TN, FP and FN represent the numbers of true positive, true negative, false positive, and false negative detection of the glaucoma.

The final score of classification task is defined as:

$$Score_{task2} = (0.1 \times AUC + 0.2 \times F1 + 0.2 \times Acc + 0.25 \times Sen + 0.25 \times Spe) \times 10$$

2.4 Final Score of the Preliminary Round Challenge

The final score of the preliminary round challengeare calculated by the scores of two subtasks:

$$Score = 0.8 \times Score_{task1} + 0.2 \times Sorce_{task2}$$

3 Segmentation Methods

3.1 Data Splitting

We split the training set by the ratio of 9:1 into the training set and the validation set randomly.

3.2 Models

OCRNet OCRNet [24, 25] firstly construct object regions based on a feature map supervised by the ground-truth segmentation, then compute the object region representations. Secondly, the representation similarity between each pixel and each object region is caculated. And the weighted aggregation of all the object region representations according to the caculated similarities with the pixel is used to augment the representation of each pixel.

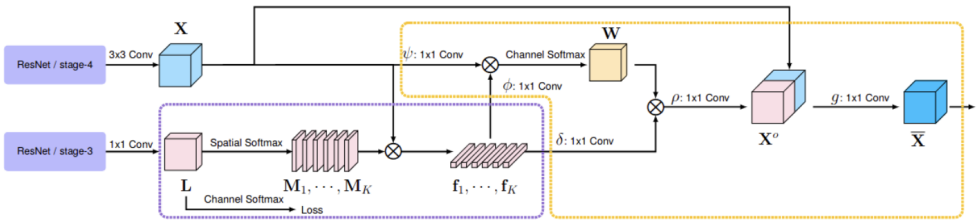

Fig. 1. The pipeline of OCRNet: the figure is from [24]

HRNet High-resolution representations are critical for position-sensitive vision problems. A high-resolution network (HRNet) can connect high- to low-resolution convolutional streams and repeatedly exchange information across resolutions. The benefit is that the generated representation is semantically richer and more spatially accurate.

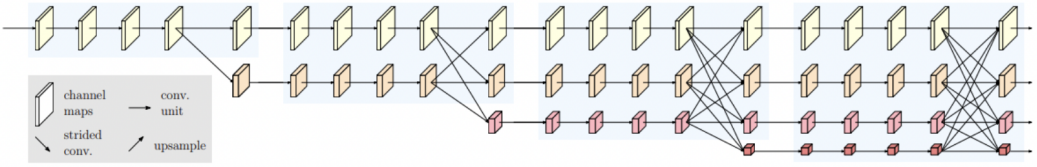

Fig. 2. The pipeline of HRNet: the figure is from [26]

3.3 Implementation Details

For all masks, we map pixel with value of 80 to 1, 160 to 2 and 255 to .3

In the training process, the input data format is the cropped patches. We used RandomPaddingCrop augmentation method in Paddleseg [32] to crop the image as 288*288 patches. We used RandomVerticalFlip, RandomHorizontalFlip, Normalize methods to do augmentation and preprocess. The probability and parameters are all default in Paddleseg. In the validation process, we used Normalize method to do preprocess. The parameters are all default in Paddleseg.

We applied OhemCrossEntropyLoss, DiceLoss, RelaxBoundaryLoss and LovaszSoftmaxLoss with ratio of 3:3:3:1 in our implementation.

The optimizer is sgd, the momentum is 0.9 and the weight decay is 4.0e-5. The learning rate scheduler is PolynomialDecay.In the first training. The initial learning rate is 0.01 and the end learning rate is 0. The power is 0.9. The total interations are 6000.

After the first training, we used the model that achieved the best result in validation set as the pretrained model then did the second training. Similarly, we did the third, fourth and fifth training. We adjusted the initial learning to 0.0000005, 0.0000001, 0.00000001 and 0.000000001 respectively. The batch size was 8. The validation was done in every 10 iteration and the evaluation metric was mIOU (including the background class) in the first and the second training. In the training after the second one, the evaluation metric was weighted mIOU(0.4*RNFL, 0.3*GCIPL, 0.3*choroid).

The deep learning framework used was paddlepaddle [33]. The programming language is python.

4 Classification Methods

4.1 Data splitting

We split the training set by the ratio of 9:1 into the training set and the validation set randomly.

4.2 Models

Resnet34 Resnet34 is a 34 layer convolutional neural network that has a residual architecture.

4.3 Implementation Details

The short side of the input image was size was resized to 600 and the longer one was also resized to keep ratio. Then a 512*512 image was cropped from the resized image. The total epochs were 200. The validation was done in every epoch. The mean of normalizing was [0.485, 0.456, 0.406] and the std was [0.229, 0.224, 0.225]. The batch size is 64. The validation was done in every epcoh and the evaluation metric was top1 accuracy.

The deep learning framework used was paddlepaddle. The programming language is python.

5 Results

Our final results are shown as below.

Table 1. Results of preliminary round

score	AUC	F1	ACC	SEN	SPE	RNFL DICE	RNFL ED	GCIPL DICE	GCIPL ED	choroid DICE	choroid ED
8.8044	1	1	1	1	1	0.9475	1.1068	0.8864	1.3118	0.9438	1.8178

References

1. Ran, A. R., Tham, C. C., Chan, P. P., Cheng, C. Y., Tham, Y. C., Rim, T. H., & Cheung, C. Y. (2021). Deep learning in glaucoma with optical coherence tomography: a review. Eye, 35(1), 188-201.
2. Huang, D., Swanson, E. A., Lin, C. P., Schuman, J. S., Stinson, W. G., Chang, W., ... & Fujimoto, J. G. (1991). Optical coherence tomography. science, 254(5035), 1178-1181.
3. Potsaid, B., Gorczynska, I., Srinivasan, V. J., Chen, Y., Jiang, J., Cable, A., & Fujimoto, J. G. (2008). Ultrahigh speed spectral/Fourier domain OCT ophthalmic imaging at 70,000 to 312,500 axial scans per second. Optics express, 16(19), 15149-15169.
4. van Velthoven, M. E., Faber, D. J., Verbraak, F. D., van Leeuwen, T. G., & de Smet, M. D. (2007). Recent developments in optical coherence tomography for imaging the retina. Progress in retinal and eye research, 26(1), 57-77.
5. Boppart, S. A., Tearney, G. J., Bouma, B. E., Southern, J. F., Brezinski, M. E., & Fujimoto, J. G. (1997). Noninvasive assessment of the developing Xenopus cardiovascular system using optical coherence tomography. Proceedings of the National Academy of Sciences, 94(9), 4256-4261.
6. Suter, M. J., Nadkarni, S. K., Weisz, G., Tanaka, A., Jaffer, F. A., Bouma, B. E., & Tearney, G. J. (2011). Intravascular optical imaging technology for investigating the coronary artery. JACC: Cardiovascular Imaging, 4(9), 1022-1039.

7. Tearney, G. J., Boppart, S. A., Bouma, B. E., Brezinski, M. E., Weissman, N. J., Southern, J. F., & Fujimoto, J. G. (1996). Scanning single-mode fiber optic catheter–endoscope for optical coherence tomography. Optics letters, 21(7), 543-545.
8. Tearney, G. J., Brezinski, M. E., Bouma, B. E., Boppart, S. A., Pitris, C., Southern, J. F., & Fujimoto, J. G. (1997). In vivo endoscopic optical biopsy with optical coherence tomography. Science, 276(5321), 2037-2039.
9. Gambichler, T., Moussa, G., Sand, M., Sand, D., Altmeyer, P., & Hoffmann, K. (2005). Applications of optical coherence tomography in dermatology. Journal of dermatological science, 40(2), 85-94.
10. Schmitt, J. M., Yadlowsky, M. J., & Bonner, R. F. (1995). Subsurface imaging of living skin with optical coherence microscopy. Dermatology, 191(2), 93-98.
11. Eladawi, N., Elmogy, M., Ghazal, M., Helmy, O., Aboelfetouh, A., Riad, A., ... & El-Baz, A. (2018). Classification of retinal diseases based on OCT images. Frontiers in Bioscience-Landmark, 23(2), 247-264.
12. Roychowdhury, S., Koozekanani, D. D., Reinsbach, M., & Parhi, K. K. (2015, August). 3-D localization of Diabetic Macular Edema using OCT thickness maps. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 4334-4337). IEEE.
13. Maetschke, S., Antony, B., Ishikawa, H., Wollstein, G., Schuman, J., & Garnavi, R. (2019). A feature agnostic approach for glaucoma detection in OCT volumes. PloS one, 14(7), e0219126.
14. Russakoff, D. B., Mannil, S. S., Oakley, J. D., Ran, A. R., Cheung, C. Y., Dasari, S., ... & Chang, R. T. (2020). A 3D deep learning system for detecting referable glaucoma using full OCT macular cube scans. Translational Vision Science Technology, 9(2), 12-12.
15. Asaoka, R., Murata, H., Iwase, A., & Araie, M. (2016). Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. Ophthalmology, 123(9), 1974-1980.
16. Muhammad, H., Fuchs, T. J., De Cuir, N., De Moraes, C. G., Blumberg, D. M., Liebmann, J. M., ... & Hood, D. C. (2017). Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. Journal of glaucoma, 26(12), 1086.
17. Kafieh, R., Rabbani, H., Abramoff, M. D., & Sonka, M. (2013). Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map. Medical image analysis, 17(8), 907-928.
18. Bogunović, H., Sonka, M., Kwon, Y. H., Kemp, P., Abràmoff, M. D., & Wu, X. (2014). Multi-surface and multi-field co-segmentation of 3-D retinal optical coherence tomography. IEEE transactions on medical imaging, 33(12), 2242-2253.
19. Fang, L., Cunefare, D., Wang, C., Guymer, R. H., Li, S., & Farsiu, S. (2017). Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search. Biomedical optics express, 8(5), 2732-2744.
20. Gopinath, K., Rangrej, S. B., & Sivaswamy, J. (2017, November). A deep learning framework for segmentation of retinal layers from OCT images. In 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR) (pp. 888-893). IEEE.
21. Shah, A., Zhou, L., Abrámoff, M. D., & Wu, X. (2018). Multiple surface segmentation using convolution neural nets: application to retinal layer segmentation in OCT images. Biomedical optics express, 9(9), 4509-4526.

22. Li, Q., Li, S., He, Z., Guan, H., Chen, R., Xu, Y., ... & Wang, W. (2020). DeepRetina: layer segmentation of retina in OCT images using deep learning. Translational Vision Science & Technology, 9(2), 61-61.
23. Fang, H., Li, F., Fu, H., Wu, J., Zhang, X., & Xu, Y. (2022). Dataset and Evaluation algorithm design for GOALS Challenge. arXiv preprint arXiv:2207.14447.
24. Yuan, Y., Huang, L., Guo, J., Zhang, C., Chen, X., Wang, J. (2018). Ocnet: Object context network for scene parsing. arXiv preprint arXiv:1809.00916.
25. Yuan, Y., Chen, X., & Wang, J. (2020, August). Object-contextual representations for semantic segmentation. In European conference on computer vision (pp. 173-190). Springer, Cham.
26. Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., ... & Xiao, B. (2020). Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 43(10), 3349-3364.
27. Shrivastava, A., Gupta, A., & Girshick, R. (2016). Training region-based object detectors with online hard example mining. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 761-769).
28. Milletari, F., Navab, N., & Ahmadi, S. A. (2016, October). V-net: Fully convolutional neural networks for volumetric medical image segmentation. In 2016 fourth international conference on 3D vision (3DV) (pp. 565-571). IEEE.
29. Berman M, Triki A R, Blaschko M B. The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4413-4421.
30. Li X, Li X, Zhang L, et al. Improving semantic segmentation via decoupled body and edge supervision[C]//European Conference on Computer Vision. Springer, Cham, 2020: 435-452.
31. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
32. Liu, Y., Chu, L., Chen, G., Wu, Z., Chen, Z., Lai, B., Hao, Y. (2021). Paddleseg: A high-efficient development toolkit for image segmentation. arXiv preprint arXiv:2101.06175.
33. Ma, Y., Yu, D., Wu, T., Wang, H. (2019). PaddlePaddle: An open-source deep learning platform from industrial practice. Frontiers of Data and Domputing, 1(1), 105-115.