PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Multi-view-based automatic method for multiple diseases screening in retinal OCT images

Wang, Ting, Zhu, Weifang, Wang, Meng, Wang, Lianyu, Chen, Zhongyue, et al.

Ting Wang, Weifang Zhu, Meng Wang, Lianyu Wang, Zhongyue Chen, Tian Lin, Haoyu Chen, Xinjian Chen, "Multi-view-based automatic method for multiple diseases screening in retinal OCT images," Proc. SPIE 12032, Medical Imaging 2022: Image Processing, 120322A (4 April 2022); doi: 10.1117/12.2611478

Event: SPIE Medical Imaging, 2022, San Diego, California, United States

Multi-View-Based Automatic Aided Diagnosis Method for Screening Multiple Diseases in Retinal OCT Images

Ting Wang[1], Weifang Zhu[1], Meng Wang[1], Lianyu Wang[1], Zhongyue Chen[1], Tian Lin[3], Haoyu Chen[3], Xinjian Chen[1,2,*]

[1]School of Electronics and Information Engineering, Soochow University, Suzhou, Jiangsu Province, 215006, China

[2]State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou, 215123, China

[3]Joint Shantou International Eye Center, Shantou University and the Chinese University of Hong Kong, Shantou, 515000, China

*indicates the corresponding author

ABSTRACT

Optical coherence tomography (OCT), a non-invasive high-resolution imaging technology of retinal tissues, has been widely used in the diagnosis of retinal diseases. However, the shortage of ophthalmologists and the overloaded work have caused great difficulties in screening for retinal diseases. Therefore, developing an accurate automatic diagnosis system for screening retinal diseases in OCT images is essential for the prevention and treatment of retinal diseases. To this end, we propose a novel multi-view-based automatic aided diagnosis method for simultaneously screening multiple diseases in retinal OCT images. First, we collected 11,211 cases of 11 common retinal diseases from the ophthalmology clinic, and each case included two OCTs acquired from different views. Then, to automatically and accurately screen diseases in retinal OCT images, a novel multi-view attention network is proposed for screening retinal diseases based on the collected data. Finally, we conduct experiments based on the collected clinical data to evaluate the performance of the proposed method. The AUC of the proposed method achieves 0.9023, which indicates the effectiveness of the proposed method.

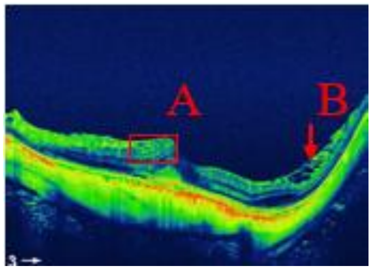
Keywords: Retinopathy, Multi-label classification, Deep Learning, Attention Mechanism, OCT image

1. INTRODUCTION

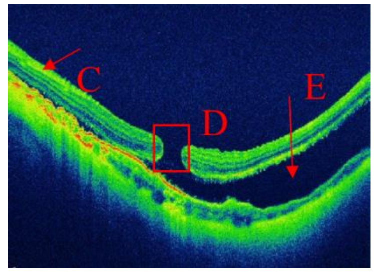
Optical coherence tomography (OCT) is a non-invasive high-resolution imaging technology for revealing the tissue details of retina. The OCT images can reveal significant diseases related to many retinal diseases such as macular hole, retinoschisis, etc. Fig. 1 shows two OCT B-scan images with five common retinal diseases. According to a report from the World Health Organization (WHO) in 2020, more than 1.106 billion people suffer from visual impairment worldwide [1]. However, the shortage of ophthalmologists and the overloaded work have brought great difficulties to efficient and accurate screening of retinal diseases. Therefore, to develop an accurate automatic retinal diseases screening method in OCT images is essential for preventing and treating retinal diseases, which can greatly reduce the load of ophthalmologists.

Recently, many deep learning based methods have been proposed for automatic medical image analysis and achieved excellent performances [2] [3]. Although previous works for retinal diseases screening have achieved promising results, these studies mainly focused on one single specific disease, and some diseases require imaging from multi-view to be diagnosed. Furthermore, in clinical practice, most patients suffer from more than one disease. Therefore, it is of great significance to develop a multi-view-based automatic multiple disease screening method. To this end, we propose a novel multi-view-based automatic method for simultaneously screening multiple diseases in retinal OCT images. Additionally, there are many studies focused on attention mechanism [4], which can guide the model focus on the semantic-related region and improve the performance of the network. Inspired by their success, we also propose a novel attention module, which can adaptively guide the model to learn the feature information of different views and improve the performance.

Medical Imaging 2022: Image Processing, edited by Olivier Colliot, Ivana Išgum, Proc. of SPIE Vol. 12032, 120322A · © 2022 SPIE
1605-7422 · doi: 10.1117/12.2611478



(a)



(b)

Figure 1. OCT images with multiple diseases. (A) edema, (B) seperation, (C) epimacular membrane, (D) macular hole, (E) retinal detachment.

2. FORMATTING OF MANUSCRIPT COMPONENTS

2.1 Overview

The architecture of multi-view-based automatic method is shown in Fig.2. To follow the clinical practice that ophthalmologists make diagnosis commonly based on OCT images collected from different views, two retinal OCT B-Scans acquired from different views are employed as the input of the proposed multi-view attention network. In order to fully extract rich features from different views and avoid over-fitting, a pre-trained ResNext-50 [7] is adopted as the backbone to capture rich feature information from two input OCT images with weight sharing strategy. A novel channel-spatial joint attention module (C-SJA) is proposed and applied to adaptively learn the correlation between features from different views. A fully connected layer is used to classify the multi-view features from C-SJA module to obtain the final screening results.

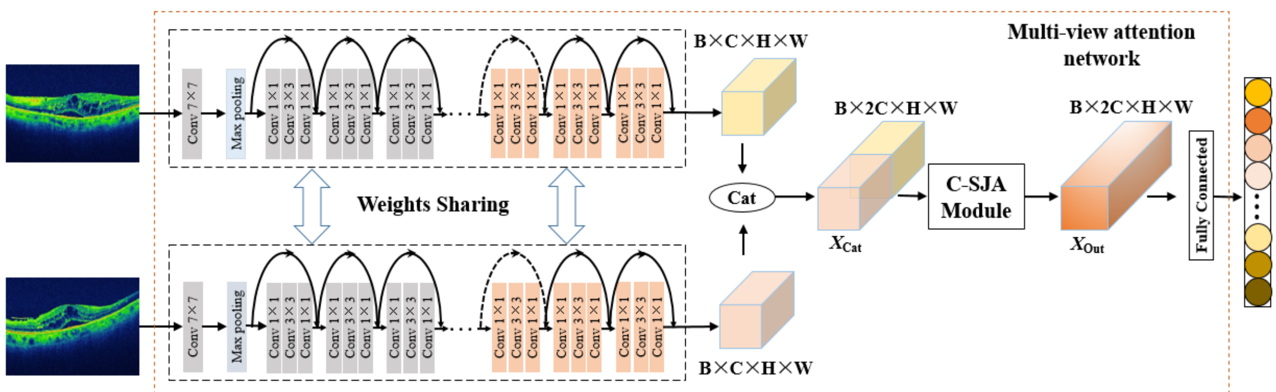


Figure 2. Overview of the proposed multi-view-based automatic aided diagnosis method. We use a pre-trained resnet-50 is adopted as the backbone to capture rich feature information from two input OCT images based on weight sharing.

2.2 Multi-view feature learning method based on weight sharing backbone network

In clinical practice, ophthalmologists commonly need OCT images acquired from multi-views to make an accurate diagnosis. However, it is still a challenge to efficiently and accurately integrate the features of different view in the design of computer automatic aided systems. To tackle this problem, we propose a novel multi-view feature learning strategy based on weight sharing backbone network (MvWSharing) for extracting rich features of different views. As shown in Fig.2, the pre-trained ResNext-50[7] without fully connected layer is adopted as our backbone network to extract the feature information in OCT images. The backbone weights between different views are shared, which aims to ensure effective feature extraction and avoid overfitting due to the introduction of more parameters. The effectiveness of multi-view feature learning method based on weight sharing backbone network will also be demonstrated in Section RESULTS.

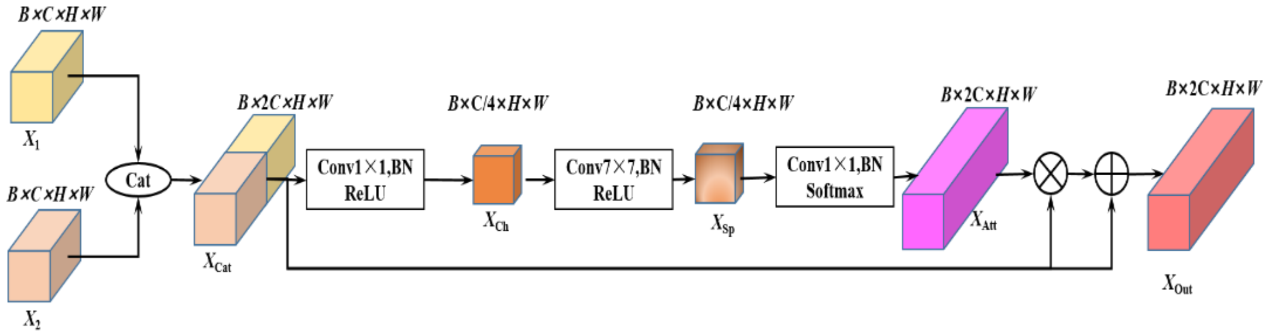


Figure 3. The structure of the proposed C-SJA module.

2.3 Channel-Space Joint Attention Module (C-SJA)

In the multi-view automatic screening system, it is crucial to effectively learn the weights of multi-view feature information to improve the prediction accuracy.Therefore, we propose a novel C-SJA module to adaptively learn the correlation between channel and spatial dimension in multi-view feature information, as shown in Fig. 3.

It can be seen from Fig. 3 that the feature maps from the two views are concatenated as the input (XCat) of C-SJA. Then, a convolutional layer with a kernel size of 1×1 followed by batch normalization (BN) and ReLU is adopted to squeeze the channel dimension of Xcat and learn the correlation of the channel dimension in different view features, as follows:

$$X_{Cat} = \mathrm{Concat}\left(X_1, X_2\right) \in R^{B,2*C,H,W} \tag{1}$$

$$X_{Ch} = \mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{Conv}1\times1\left(X_{Cat}\right)\right)\right) \in R^{B,C/4,H,W} \tag{2}$$

where B, C, H, W represent the batch size, channel, height, and width of the feature map from the backbone, respectively. And then, a convolutional layer with kernel size of 7×7 followed by BN and ReLU is applied on XCh to capture the global feature correlation in spatial dimension:

$$X_{Sp} = \mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{Conv}7\times7\left(X_{Ch}\right)\right)\right) \in R^{B,C/4,H,W} \tag{3}$$

Furthermore, to obtain the feature response map that matches the dimension of the input, we first use Conv1×1 to unsqueeze the channel of XSp, and then use the softmax normalize the feature response in channel dimensions, as follows:

$$X_{Att} = \mathrm{Sigmoid}\left(BN\left(\mathrm{Conv}1\times1(X_{Sp})\right)\right) \in R^{B,2*C,H,W} \tag{4}$$

Finally, the feature map Xout with channel and spatial correlation response is obtained as follows:

$$X_{out} = X_{cat} + X_{cat} \times X_{Att} \tag{5}$$

As shown in Eq.5 that the addition of Xcat and Xcat * Xatt constructs the residual architecture, which ensures that the network can adaptively integrate different view features and avoid the problem of gradient disappearance caused by weak response information.

3. RESULTS

3.1 Datasets

In this paper, we establish a big database with multiple retinal diseases, which contains 11,211 cases acquired from the ophthalmology clinic and each case includes two OCTs obtained from different views. These retinal OCT B-scans are all pseudo-color images, including 11 common retinal diseases. We randomly divide 11,211 cases into training set, validation set and test set according to the ratio of 0.6, 0.2 and 0.2, respectively. All OCT B-scans are resized to (224,224) before fed into the network. The label of each case is encoded as a one-hot vector with 11 dimensions and only values of 0 or 1, where 0 and 1 indicate the presence and absence of a certain retinal diseases, respectively. TABLE I lists the disease corresponding to each dimension in the one-hot vector.

TABLE 1. THE DISEASE CORRESPONDING TO EACH DIMENSION IN THE ONE-HOT VECTOR

One-hot vector	***Diseases***
10000000000	Macular hole
01000000000	Retinoschisis
00100000000	Macular edema
00010000000	Epiretinal membranes
00001000000	Posterior detachment of the vitreous body
00000100000	Abnormal reflection of internal limiting membrane
00000010000	Abnormal reflection of IS/OS
00000001000	Abnormalities of photoreceptor layer
00000000100	Retinal detachment
00000000010	Pigment epithelium detachmentr
00000000001	Choroidal atrophy

3.2 Parameter settings

The proposed multi-view attention network for multiple diseases screening in retinal OCT images is implemented on the public Pytorch platform and trained on an NVIDIA GPU RTX 3090 with 24G memory. The optimizer is Adam with an initial learning rate of 0.0001, and the batch size is set to 16. To ensure the fairness and objectiveness, all methods involved in this paper are implemented on the same environment and with the same super-parameters.

3.3 Results

TABLE II The Comparison Results Of MvConcat and MvWSharing

	Network	AUC	mAP
MvConcat	Resnext-50 [7]	0.7131	0.5073
	Densenet-121 [8]	0.7811	0.5178
	Resnet-50 [5]	0.7453	0.5046
	Inception-v4 [9]	0.7744	0.5180
	Inceptionresnet-v2 [9]	0.7627	0.5020
	Se-resnet50 [6]	0.7230	0.4944
MvWSharing	Resnext-50	0.8619	0.6833
	Densenet-121	0.8582	0.6760
	Resnet-50	0.8541	0.6382
	Inception-v4	0.8416	0.6598
	Inceptionresnet-v2	0.8465	0.6561
	Se-resnet50	0.8329	0.5993
	Resnext-50+C-SJA (Proposed)	**0.8713**	**0.7027**

To objectively validate the effectiveness of the proposed method, the area under the ROC curve (AUC) and the mean average precision (mAP) are adopted as the indicators.

We first explore the performance of multi-view feature concatenation learning method (MvConcat), which directly concatenates OCT images acquired from different views as input to screen for retinal diseases. The results are listed in TABLE II. As shown in TABLE II, compared with the networks based on MvConcat, the proposed multi-view feature learning method based on MvWSharing achieves better performance. In particular, compared with the ResNext-50 based MvConcat, the AUC and mAP of the ResNext-50 based MvWSharing improve by 14.88% and 21.95% respectively. The results prove the effectiveness of the proposed MvWSharing.

In addition, we also conduct experiments to demonstrate the effectiveness of the proposed C-SJA module. The results also be listed in TABLE II. It can be seen from TABLE II that the proposed method (ResNext-50+C-SJA) achieves the best performance. Compared with ResNext-50 based MvWSharing, the AUC and mAP of the proposed method increase

0.94% and 1.94% respectively. These results prove the effectiveness of the proposed C-SJA module. The experimental results demonstrate the effectiveness of the proposed method.

4. CONCLUSION

In this paper, we have developed a novel multi-view-based automatic aided diagnosis method that can simultaneously screen multiple retinal diseases in OCT images. To validate the performance of the proposed method, comprehensive experiments have been conducted on our newly established databased. The experimental results demonstrate the effectiveness of the proposed method. To the best of our knowledge, it is the first time to develop a computer-aided diagnosis method that can simultaneously diagnose 11 retinal diseases based on multi-view retinal OCT images, and have achieves excellent results.

5. ACKNOWLEDGEMENTS

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFA0701700, in part by the National Nature Science Foundation of China under Grant 61622114, and in part by the National Basic Research Program of China under Grant 2014CB748600.

REFERENCES

[1] K. Ohno-Matsui, T. Yoshida, S. Futagami, K. Yasuzumi, N. Shimada and A. Kojima, "Patchy atrophy andlacquer cracks predispose to the development of choroidal neovascularization in pathological myopia," BritishJournal of Ophthalmology 87(5), 570-573 (2003).

[2] Q. Wang, Y. Zheng, G. Yang, W. Jin, X. Chen, and Y. Yin. "Multiscale rotation-invariant convolutional neural networks for lung texture classification". IEEE Journal of Biomedical and Health Informatics, vol.22, no.1, pp.184-195, 2017.

[3] Z. Zhang, H. Fu, H. Dai, J. Shen, Y. Pang, and L. Shao. "ET-Net: A Generic Edge-aTtention Guidance Network for Medical Image Segmentation". International Conference on Medical Image Computing and Computer-Assisted Intervention 2019. Springer, 2019, pp. 442-450.

[4] Sanghyun Woo, Jongchan Park, Joon-Young Lee, In So Kweon. "Cbam: Convolutional block attention module." Proceedings of the European conference on computer vision (ECCV). 2018.

[5] He, K., Zhang, X., Ren, S., &Sun, J. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[6] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[7] Xie, S., Girshick, R., Dollar, P., Tu Z., &He, K. "Aggregated residual transformations for deep neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[8] Huang, G., Liu, Z., Van Der Maaten, L., &Weinberger, K. Q. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[9]Szegedy, C., loffe, S., Vanhoucke, V., &Alemi, A. "Inception-v4, inception-resnet and the impact of residual connections on learning." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 31. No. 1. 2017.