PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Multi-modality network based on CGAN and attention mechanism for glaucoma grading

Ling Liu, Yuanyuan Peng, Dehui Xiang, Fei Shi, Xinjian Chen

Ling Liu, Yuanyuan Peng, Dehui Xiang, Fei Shi, Xinjian Chen, "Multi-modality network based on CGAN and attention mechanism for glaucoma grading," Proc. SPIE 12464, Medical Imaging 2023: Image Processing, 124643L (3 April 2023); doi: 10.1117/12.2654113

SPIE.

Event: SPIE Medical Imaging, 2023, San Diego, California, United States

Multi-modality Network Based on CGAN and Attention Mechanism for Glaucoma Grading

Ling Liu [1], Yuanyuan Peng [1], Dehui Xiang [3], Fei Shi[3], Xinjian Chen[1,2]*

[1]School of Electronics and Information Engineering, Soochow University, Suzhou, 215006, China
[2]State Key Laboratory of Radiation Medicine and Protection, Soochow University,Suzhou, 215123, China
[3]Guangzhou Women and Children Medical Center, Guangzhou, 510623, China

ABSTRACT

Glaucoma is a progressive optic neuropathy characterized by changes in the structure of the optic nerve head and visual field，which is one of the major irreversible blinding eye diseases worldwide. Early screening and timely diagnosis of glaucoma is of significant importance. Fundus color photography and optical coherence tomography (OCT) are the two most effective imaging modalities for glaucoma screening, where significant ocular structural changes, such as vertical cup-to-disc ratio (vCDR) on fundus images and retinal nerve fiber layer (RNFL) thickness on OCT volumes, can be present with both imaging modalities. In recent years, multi-modal deep learning methods have shown great advantages in image classification and segmentation tasks. In this paper, we propose a multi-modal glaucoma grading network with two main contributions: (1) To address the inherent shortage of multi-modal training data, conditional generative adversarial network (CGAN) is used to generate more synthetic images, extending the dataset over the only available dataset. (2) A multi-modality cross-attention (MMCA) module is proposed to further improve the classification accuracy.

KEYWORDS: Glaucoma Grading, Fundus Color Photography, OCT, Neural Network, Multi-modal, CGAN, Cross-attention

1. INTRODUCTION

Glaucoma is an eye disease in which intraocular pressure rises intermittently or continuously. Glaucoma has affected more than 70 million people worldwide, making it one of the leading causes of irreversible blindness in the world[1-2]. Accurate early screening and treatment of glaucoma are very important because there are most likely no warning signs in the early stages of glaucoma. The optic disc region and retinal nerve fiber layer are the primary sites of severe damage to pre-blinding glaucoma lesion, as shown in Figure 1. Therefore, accurate early detection and monitoring of changes in the optic disc region and RNFL has become the primary focus of glaucoma treatment.

Medical Imaging 2023: Image Processing, edited by Olivier Colliot, Ivana Išgum,
Proc. of SPIE Vol. 12464, 124643L · © 2023 SPIE
1605-7422 · doi: 10.1117/12.2654113

Many previous studies on automated or semi-automated methods for Glaucoma diagnosis are mainly focused on Fundus color photography or OCT. For example, Raghavendra et al. proposes a novel tool for the accurate detection of glaucoma using deep learning technique[3], where an eighteen layer convolutional neural network (CNN) is effectively trained to extract robust features for Classification. Similarly, Raja et al. proposes a framework based on VGG-16 architecture for feature extraction and classification of retinal layer pixels[4].
However, as far as we know, there are still very few studies using multi-modality information to diagnose glaucoma. Multi-modality fusion is a popular research topic in the field of medical imaging, because the fusion of two modal information can provide more accurate results than single modality. In this paper, a simple and novel deep learning based method is proposed for glaucoma grading, which exploits multi-modality information though dual-branch network, Conditional-GAN (CGAN) to expand the dataset and a cross-attention module to fully exploit complementary information, respectively.

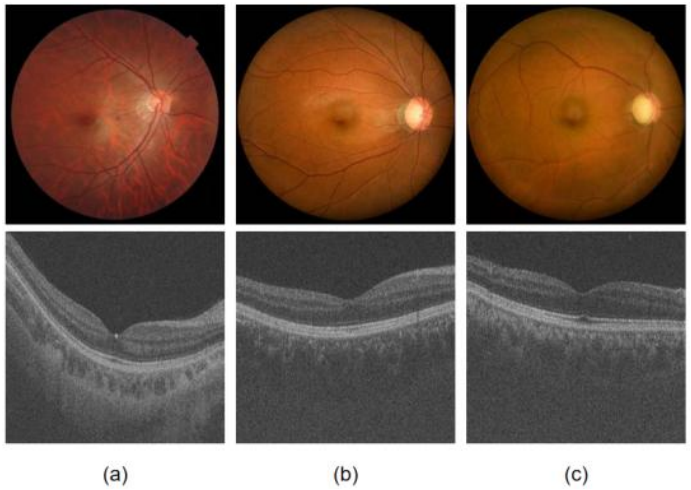

Figure 1. Fundus and OCT images of no glaucoma, early glaucoma, and moderate or advanced glaucoma. (a) no glaucoma. (b)early glaucoma, (c) moderate or advanced glaucoma

2. METHODS

In this section, the proposed method is described as two parts: structure of the proposed CGAN network，structure of cross-attention classification network and the two-stage training strategy.

2.1 Structure of the proposed CGAN network

The GAMMA dataset consists of 100 pairs of fundus and OCT images labeled by ophthalmologists with three grades of glaucoma (50 "normal", 26 "early", 24 "moderate or advanced"), where each OCT volume contains 256 2D slices from the B channel.
Inspired by pixtopixHD[5]，we design a CGAN to generate more training data and the overall architecture of our network is shown in Figure 2. We first trained a classification network with a single branch of fundus color photographs to get CAM images with pathological data[6]. Together with the original images, they are

fed into the CGAN network as data pairs. Finally, we can obtain a large number of synthetic fundus color images by flipping and translating the CAM images. We only apply this strategy on fundus color photos because the number of OCT slices is large enough. By this data enhancement method, our bimodal data pairs are expanded nearly three times, which can greatly improve the generalization ability of the model and get better results.

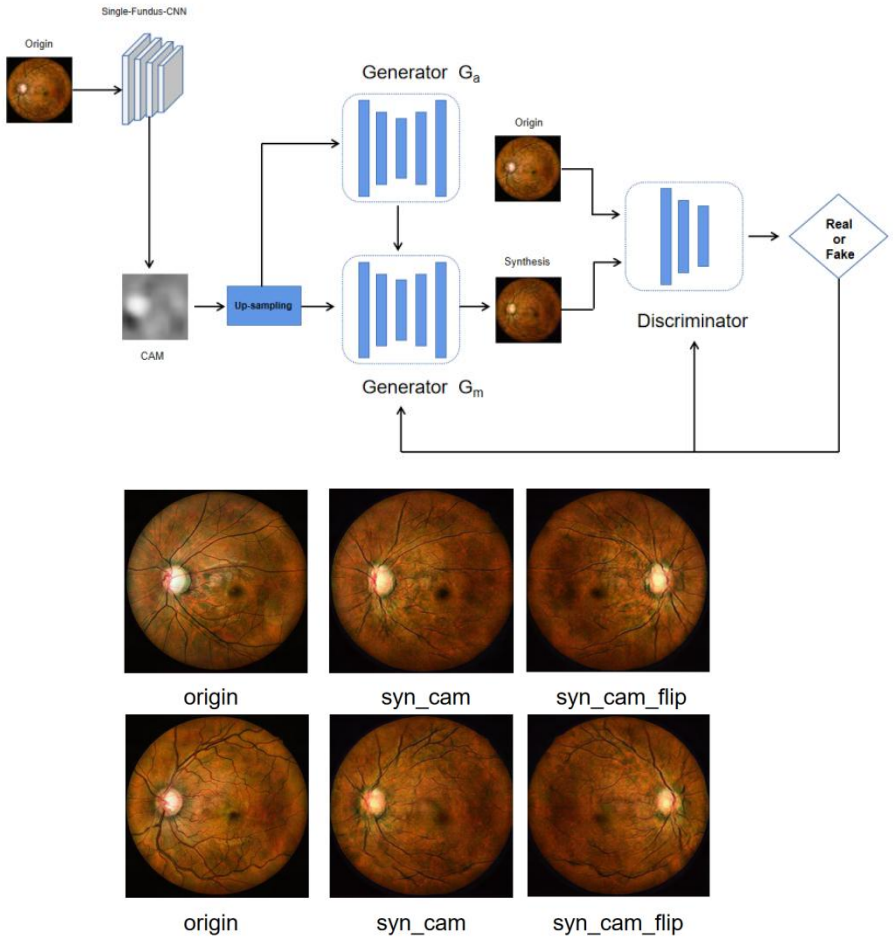

Figure 2. GAN image synthesis by CAM

2.2 Structure of cross-attention classification network

To promote a better integration of the dual-branch features, we design a novel multi-modality cross-attention network based on attention mechanism and the overall architecture of our network is shown in Figure 3. For

each branch of the two-stream CNN, we adopt ResNet-34 as the backbone of each branch, where one branch is responsible for extracting the features of the 3D fundus images, and the other branch takes three single-channel OCT images as input to achieve symmetry with the first branch. In addition, the baseline method takes the concatenation operation as fusion operation of dual-branch features. However, the direct concatenation operation may lead to information redundancy,which is not conducive to the improvement of performance. Therefore, we perform feature selection by cross-attention module before feature fusion to fully exploit complementary information. Finally, we fuse the dual-branch features through addition operation and feed it into a fully connected layer for final classification.

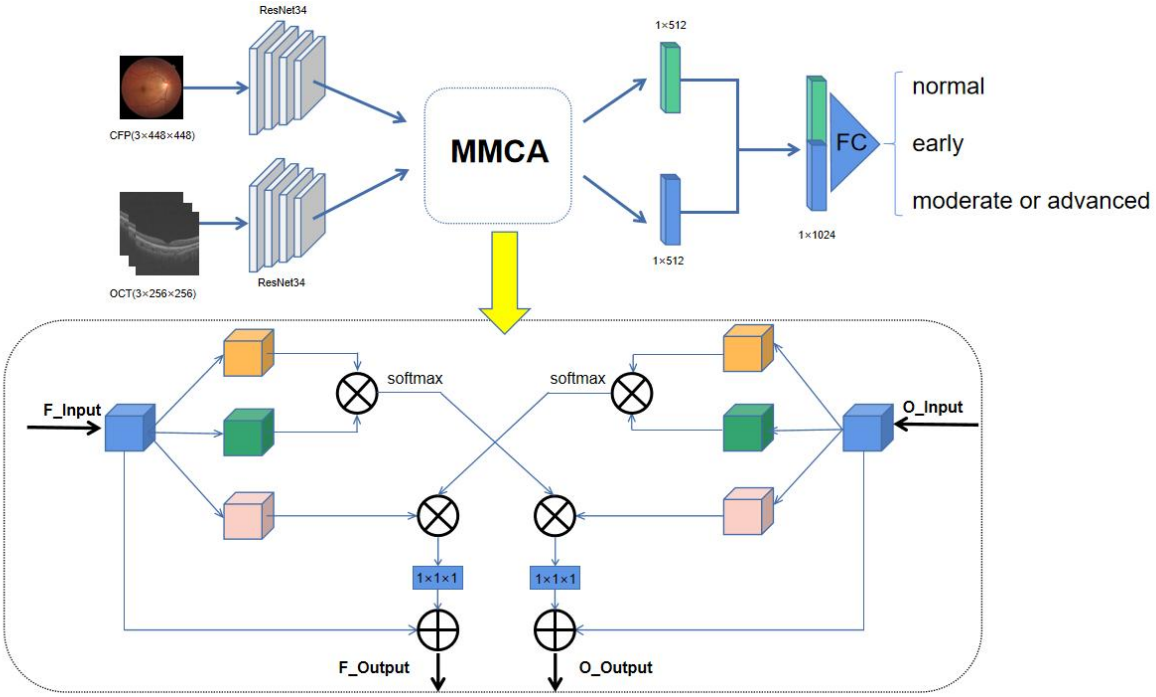

Figure 3. The overall architecture of our network.

2.3 Two-stage training strategy

Given a small number of real data sets, using the proposed image synthesis strategy under cam conditions enables us to construct a large number of multi-modal training examples. However, working directly with real datasets and synthetic datasets is problematic, as the latter can easily outnumber the former. To do this, we use a two-stage training strategy, namely pre-training and fine-tuning. In the first stage, synthetic images were used to train. We used the real data set to fine-tune in the second stage. Although simple, this two-stage training strategy is very effective.

3. RESULTS

3.1 Datasets and preprocessing

The fundus images and oct images released by GAMMA were provided by Sun Yat-sen Ophthalmic Center, Sun Yat-sen University, Guangzhou, China, which were labeled as non-glaucoma, early-glaucoma and Moderate or advanced glaucoma.The ground truth of glaucoma grading task for each data was determined by clinical records and was based on the results of all clinical examinations.
100 data pairs of two clinical modality images are used to train and evaluate the performance of model. In order to reduce the computational cost, all fundus images and oct images are downsampled to 256 ×256 by using bilinear interpolation. To prevent over-fitting and enhance the generalization ability of the model, online data augmentation has been performed, including random rotation 30° , horizontal flipping and vertical flipping.

3.2 Parameter settings

We train the model with back-propagation algorithm by minimizing the cross-entropy cost function. Adam is used as the optimizer, where initial learning rate is set to 0.00001. The batch size and epoch are set to 4 and 50, respectively. During training, all networks are trained with identical optimization schemes and the best model is saved on validation set with 4-fold validation strategy.

3.3 Results

To visually explain how individual modes contribute, we extend the class-activation mapping technique to multimodal scenarios, and to visualize their contributions,thereby revealing which part of the input image contributes the most. As shown in Figure 4, the visualization results demonstrate that our model can extract the potential features for cup-to-disc ratio (vCDR) on fundus images and thickness information on OCT volume.
In order to evaluate the effectiveness of dual-branch network, CGAN and MMCA module in the pre-trained ResNet34, a series of ablation studies are conducted. We call the fine-tuned ResNet34 as Backbone. As shown in Table 1, compared to Backbone, the performance of multi-modal network is better than that of single-modal network. In addition, compared to multi-modal network, multi-modal with MMCA module (Multi-modal+MMCA) improves the accuracy and kappa by 5.1% and4.6%, respectively. Specially, multi-modal with CGAN (Multi-modal+CGAN) improves the accuracy and kappa by 2.87% and 0.89%, respectively. The proposed network (Multi-modal+All) performs well in terms of all quantitative metrics, which demonstrates the effectiveness of the proposed method in the glaucoma grading.

Table 1：Classification Results

Method	Fundus	OCT	ACC	KAPPA
Backbone	✔		0.7500±0.0437	0.6872±0.0130
Backbone		✔	0.6606±0.0347	0.5653±0.0012
Multi-modal	✔	✔	0.7923±0.0292	0.7370±0.0230
Multi-modal+MMCA	✔	✔	0.8435±0.0293	0.7833±0.0178
Multi-modal+CGAN	✔	✔	0.8210±0.0179	0.7459±0.0212
Multi-modal+ALL	✔	✔	**0.9001±0.0106**	**0.8280±0.0325**

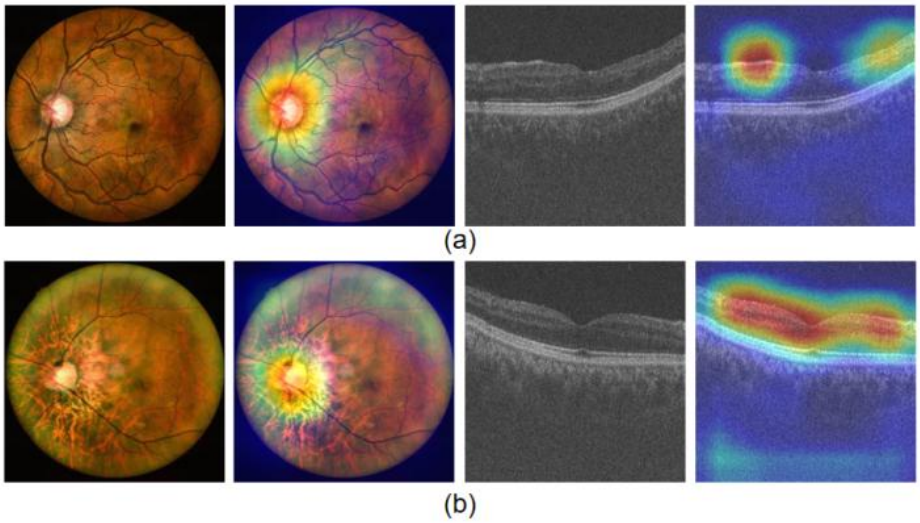

Figure 4. Heat maps of class activation. (a)Fundus and OCT images of early glaucoma and Corresponding heat maps, (b)Fundus and OCT images of moderate or advanced glaucoma and Corresponding heat maps

4. CONCLUSIONS

In this paper, we propose a Multi-modality Network Based on CGAN and Attention Mechanism for Glaucoma Grading. Firstly, we introduce the CGAN for data augmentation and generate more training data. Secondly, we propose a cross-attention classification network inspired by non-local Network [7], which promote a better integration of the dual-branch features. In this way, the network can focus on the key information of the two modality images in our task. Finally, we also introduce the two-stage training strategy learning to further improve the classification accuracy, which can make full use of shallow information.The experimental results demonstrate the effectiveness and feasibility of the proposed method. The proposed method provides a promising technology which can assist pediatric ophthalmologists in glaucoma diagnosis. In the near future, we focus on the further study of the segmentation of the two modality.

5. ACKNOWLEDGEMENTS

This study was supported in part by the National Key R&D Program of China (2018YFA0701700) and part by the National Nature Science Foundation of China (U20A20170，61622114，62271337 and 61971298).

6. REFERENCE

[1]Allingham R R, Moroi S, Shields M B, et al. Shields' textbook of glaucoma[M]. Lippincott Williams & Wilkins, 2020.
[2]Stamper R L, Lieberman M F, Drake M V. Becker-Shaffer's diagnosis and therapy of the glaucomas E-Book[M]. Elsevier Health Sciences, 2009.
[3]Raghavendra U, Fujita H, Bhandary S V, et al. Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images[J]. Information Sciences, 2018, 441: 41-49.
[4]Raja H, Akram M U, Shaukat A, et al. Extraction of retinal layers through convolution neural network (CNN) in an OCT image for glaucoma diagnosis[J]. Journal of Digital Imaging, 2020, 33(6): 1428-1442.
[5]Wang T C, Liu M Y, Zhu J Y, et al. High-resolution image synthesis and semantic manipulation with conditional gans[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8798-8807.
[6]Wang W, Li X, Xu Z, et al. Learning two-stream CNN for multi-modal age-related macular degeneration categorization[J]. IEEE Journal of Biomedical and Health Informatics, 2022.
[7]Wang X, Girshick R, Gupta A, et al. Non-local neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7794-7803.