LOCALIZATION OF RETINAL NERVE FIBER LAYER DEFECT IN FUNDUS IMAGE BY VISUAL FIELD GUIDED LEARNING NETWORK

[1] Chun-Fu Yeh (葉淳輔), *[1] Guan-An Chen* (陳冠安), *[2] Min-Yu Huang* (黃敏祐) ,
[2] Kwou-Yeung Wu (吳國揚)

[1] Industrial Technology Research Institute, Hsinchu, Taiwan
[2] Chung-Ho Memorial Hospital Kaohsiung Medical University, Kaohsiung, Taiwan
E-mail: crackereidolon@gmail.com.tw

ABSTRACT

Retinal Nerve Fiber Layer Defect (RNFLD) can be an earliest sign to detect the ongoing glaucomatous damage. However, existing measurements, including visual field test and optic cup/disc ratio, fail to reflect RNFLD. Although optical coherence tomography (OCT) may provide information about RNFLD, the field of view (FOV) of OCT is smaller than that of fundus camera. This means early RNFLD may be undetected by OCT. In order to screen out patients with early-stage glaucoma, we propose to build a deep neural network to both predict glaucoma and locate RNFLD in fundus image by constraining its latent space with visual field map (VFM), which has wider FOV than fundus image and indicates visual field loss led by RNFLD. Since VFM does not match fundus image at pixel level, the challenge of this net-work would be to learn the spatial relationship between fundus image and VFM in addition to the prediction of glaucoma. To tackle this challenge, we compared three encoder-decoder convolutional neural network (CNN) with distinctive architectures in this study: (i) encoder-decoder CNN, (ii) encoder-decoder CNN with spatial transformer network (STN) and (iii) generative adversarial network (GAN), whose generator is the same as (i). The main evaluation metrics in this study was the correlation coefficient between predicted VFM and real VFM. Be-sides, accuracy and AUC of each network for the prediction of glaucoma were measured to make sure the predicted VFMs were closely related to glaucoma. The study was conducted on the dataset we collected from a medical center. Our results demonstrated that the correlation coefficient produced from model (iii) was the highest and it also did well in the prediction of glaucoma. This proposed network would be the first one to predict glaucoma and locate RNFLD simultaneously to provide explainable results for ophthalmologists and address the pixel-level mismatch between fundus images and VFM.

Keywords: *Retinal Nerve Fiber Layer Defect, Glaucoma, Generative Adversarial Network.*

1. INTRODUCTION

Glaucoma, which is characterized by the progressive optic neuropathy, is the second leading cause of irreversible blindness, and Asia would become the top three areas affected the most [1, 2]. Although the blindness resulting from glaucoma is preventable by early detection and treatments [3], existing devices and measurements were not feasible to detect the early indicator of Glaucoma, namely retinal nerve fiber layer defect (RNFLD), particularly for large-scale population screening. To deal with this problem, a feasible measurement for RNFLD in large-scale screening is imperative.

Fundus camera (FC) is considered as a more economical device for large-scale screening [4], and optic cup-to-disc ratio (C/D ratio) is the most common indicator derived from fundus image by an ophthalmologist. During screening, a suspect with C/D ratio more than 0.8 would be referred to ophthalmology clinic for in-depth evaluation by optical coherence tomography (OCT) [5] and visual field test (VFT) [6]. However, C/D ratio may be an indirect indicator for RNFLD and its modest inter-rater and intra-rater variability may affect its reliability [7, 8]. This could lead to delayed diagnosis of glaucoma. A new measurement derived from fundus image for RNFLD should be developed for the need of early detection.

While OCT and VFT provide valuable information about RNFLD and the consequent visual field loss, their cost and test duration are high in comparison with those of FC. Specifically, OCT is used to measure the thickness of retinal layers of fundus [5]. This reflects RNFLD directly but it is not cost-effective in large-scale screening. On the other hand, VFT is used to detect visual field loss led by RNFLD. The visual field map (VFM) generated by VFT is the direct revelation of whether specific part of optic nerve reacts to incoming lights [6], indicating RNFLD as well. Since it takes al-most thirty minutes to complete VFT, it is not feasible for large-scale screening. To locate RNFLD in fundus image reliably, a possible method would be to establish the mapping from fundus image to VFM or thickness map from OCT. As VFM has larger field-of-view (FOV) than fundus image (Fig. 1), our proposed method aimed to map fundus image to VFM by a deep neural network (DNN). As a result, the location of RNFLD in fundus image could be inferred from the predicted VFM of DNN.

The major challenge in establishing the mapping between fundus image and VFM is the mismatch at pixel level between these two images. VFM is a machine-generated grid with equal space between vertices, while fundus image is the one which the dis-tance between pixels do not correspond to the real distance in fundus. That is, fundus image could be considered a distorted image from VFM. Due to this mismatch, the mapping from fundus image to its corresponding VFM may be difficult to learn.

In this study, we compared three distinctive encoder-decoder convolutional neural network (CNN) to tackle the aforementioned challenge. First, an encoder-decoder CNN was constructed as the baseline to learn such mapping. Second, in order to learn the mapping explicitly, a spatial transformer network (STN) is added to the encoder-decoder CNN to transform predicted VFMs. At last, a generative adversarial network (GAN) is proposed, whose generator is the same as the encoder-decoder CNN mentioned in the above two and discriminator is a common CNN, to tackle the challenge.

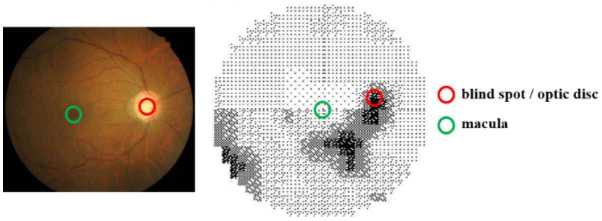


Fig. 1. Examples of fundus images and their corresponding visual field maps (VFM). The field of view (FOV) of VFM is 60 degree while the FOV of fundus image is 45 degree. Note that VFM was vertically flipped from its original VFM to match the orientation of fundus image.

1.2 Related Work

In order to detect suspects with glaucoma via fundus camera in large-scale population screening, a CNN based on inception-v3 architecture was developed to predict glaucoma with fundus image as input [9]. The result showed the area under receiver operator characteristic curve (AUC) achieved 0.986 with sensitivity of 95.6% and specificity of 92.0%. Despite of its effectiveness in screening suspects of glaucoma, the model did not provide explainable results for why each suspect was predicted as glaucoma. Furthermore, similar research, which focused on the analysis of the heatmaps generated from CNN, indicated that the optic disc area was the most important area for the prediction of glaucoma [10]. Inferred from these results, CNN models trained only on fundus images for glaucoma may predict suspects with early RNFLD as non-referable glaucoma because early RNFLD usually appears at regions near macula and away from optic disc.

Recently, a study demonstrated the performance of a model predicting the thickness of RNFL around optic disc with fundus images as input [11], which was trained on fundus images paired with thickness maps from spectral-domain OCT (SD-OCT). Although it was promising for the quantification of RNFLD in fundus image, the FOV of SD-OCT was too small to detect suspects with RNFLD away from optic disc. To screen suspects with early RNFLD with fundus images, the FOV from the ground truth (either VFT or OCT) of RNFLD should be larger than fundus images.

The main objective of this study was to develop an encoder-decoder CNN which would be trained on fundus images paired with corresponding VFMs to locate RNFLD in addition to the prediction of glaucoma. With this model, the screening for early-stage glaucoma in large-scale population would be achieved.

1.3 Contribution

The contributions of this study can be summarized as follows: (i) Proposing a deep learning model to locate RNFLD in fundus images with VFMs for the first time. (ii) Addressing the pixel-level mismatch between fundus images and VFMs by using GAN. (iii) Reducing the burden of annotation on ophthalmologists by using VFMs as annotations for RNFLD in fundus images.

2. Method

2.1 Data Collection and Preprocessing

Currently, there is no available open dataset containing fundus images paired with their corresponding VFMs. To collect such pairs, this study was first approved by the local institutional review board (IRB) of Chung-Ho Memorial Hospital Kaohsiung Medical University in Taiwan (KMUHIRB-E(I)-20180241). Then, there were 740 fundus images (produced by KOWA Nonmyd 7) paired with their VFMs (produced by Carl Zeiss Meditec HFA II, 2007) collected retrospectively from 351 subjects. The included subject was either the individual diagnosed with glaucoma or the one with healthy fundus and visual field. Among 740 pairs, there were 445 pairs from the glaucoma subjects and 295 pairs from subject in healthy condition.

To evaluate the generalizability of the proposed models, 5-fold cross validation was applied during modeling. That is, each fold included 20% of 351 subjects. In each training, four folds of data were treated as training dataset and the rest of data was validation dataset.

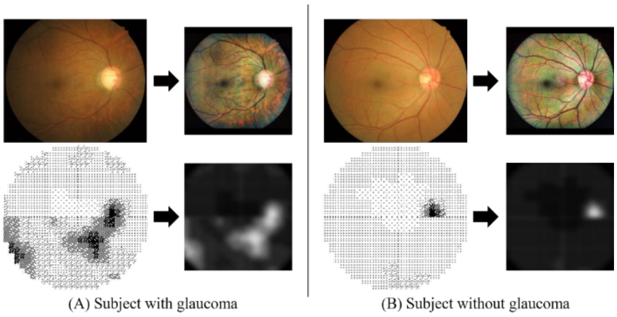


Fig. 2. Preprocessed fundus images and VFMs. The pair (A) on the right is from the subject with glaucoma. The other pair (B) on the left is from the subject without any significant visual field deficit. The size of original fundus image were 2144 x 1424. The preprocessed fundus images and VFMs were resized to 512 x 512.

To deal with the color variability in fundus image and make retinal nerve fiber more perceivable, three contrast limited adaptive histogram equalization (CLAHE) [12] with in-between Gaussian smoothing were applied on every fundus image in the dataset. Two examples were illustrated in **Fig. 2.**

As for VFM, following steps were conducted in sequence to get a VFM in gray scale: (i) Gaussian blur with kernel size of 30 x 30. (ii) Dilation with kernel size of 5 x 5 following a Gaussian blur with kernel size of 30 x 30. (iii) Erosion with kernel size of 5 x 5 following a Gaussian blur with kernel size of 10 x 10. (iv) Reversal of black and white. These steps were to remove single black dots in VFM, which were not related to visual field deficits, and to ensure that gray regions related to RNFLD were kept. The final step was to represent regions related to RNFLD with brighter pixel values.

2.2 Model Architectures

The architectures of the three proposed models were demonstrated in **Fig. 3**. The encoder part of the model (i) had same structure as part of the inception-v3 architecture (from first layer to 6e layer) described in this paper [13]. Then, four consecutive convolution modules, which constituted the decoder, were applied to the feature maps generated from the encoder. Each module was mainly comprised of a dropout layer, a convolutional layer with kernel size of 7 x 7, a batch normalization layer, an activation layer of leaky Rectified linear unit (leaky ReLU) and an upsampling layer. The depths of the feature maps from these four modules were 144, 72, 36 and 4 respectively, and the sizes of these maps were 64, 128, 256 and 512. With this decoder, a gray image was generated by its last layer (an additional convolution layer with kernel size of 7 x 7) and a sigmoid function was further applied to the generated image to map its pixel values to values between 0 and 1. The final output was regarded as the predicted VFM.

In addition to the prediction of VFM, the prediction of glaucoma was done by the same model. That is, a fully connected layer with a sigmoid function was concatenated on the feature vector produced by an average pooling layer applied on the feature maps from the decoder. By training this model to predict glaucoma and VFM simultaneously, the feature vector would be the representative of RNFLD, which is the cause of visual field loss and consequent glaucoma.

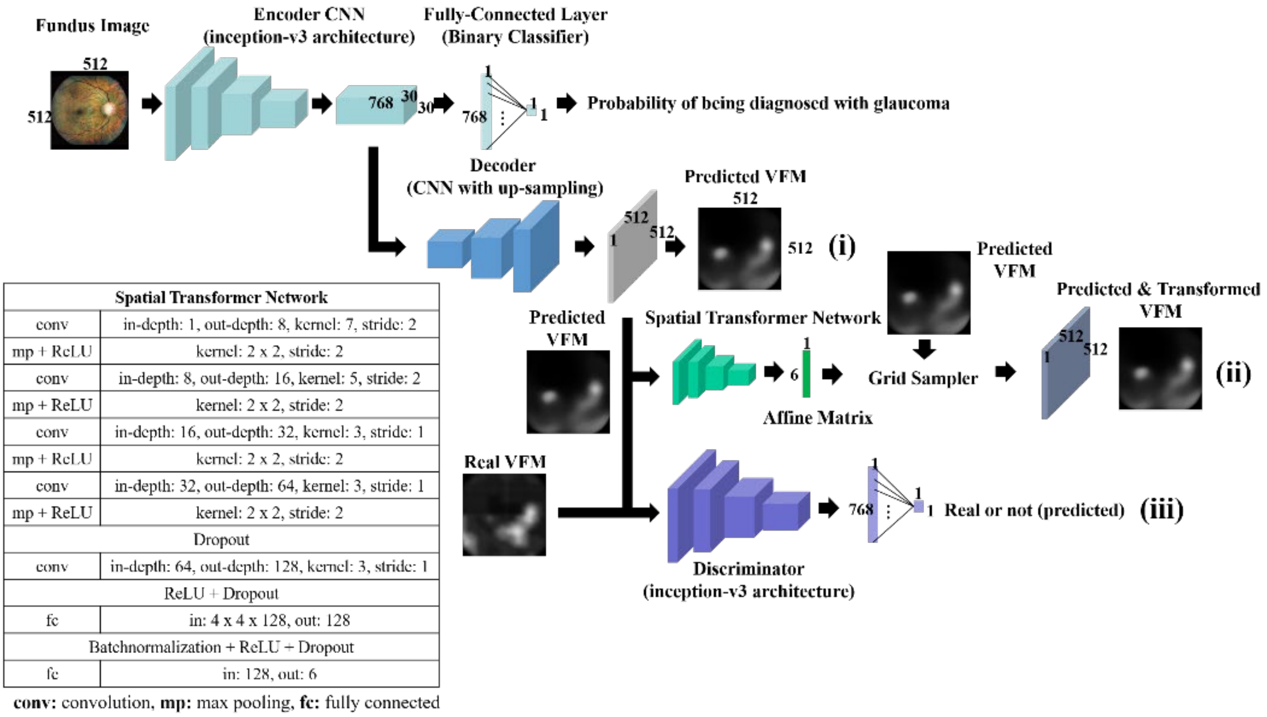
Building on model (i), model (ii) included a STN [14], which aimed to learn the spatial relationship between fundus images and VFMs. The detail of this STN is described in **Fig. 3**. Following this design, the affine matrix was derived from the predicted VFM. Then, a grid sampler was applied to sample the predicted VFM with its corresponding affine matrix to get the transformed map, which was the final output.

In model (iii), a discriminator, which followed inception-v3 architecture, was used to facilitate the generator to produce VFMs as similar as the real VFMs. This generator was the same model as model (i).

2.3 Evaluation Metrics

To evaluate the similarity between predicted VFMs and real VFMs, Pearson correlation coefficient (r) and dice coefficient (Dice) were calculated for each predicted VFM and real VFM. For the calculation of Dice, a binary thresholding with 0.5 was first applied to both predicted VFMs and real VFMs. Then, equation (1) for Dice was calculated for each predicted VFM and real VFM. Moreover, the evaluation metrics for the prediction of glaucoma included accuracy and AUC score.

$$Dice(X,Y) = \frac{2\,|X \cap Y|}{|X||Y|} \quad (1)$$



Spatial Transformer Network	
conv	in-depth: 1, out-depth: 8, kernel: 7, stride: 2
mp + ReLU	kernel: 2 x 2, stride: 2
conv	in-depth: 8, out-depth: 16, kernel: 5, stride: 2
mp + ReLU	kernel: 2 x 2, stride: 2
conv	in-depth: 16, out-depth: 32, kernel: 3, stride: 1
mp + ReLU	kernel: 2 x 2, stride: 2
conv	in-depth: 32, out-depth: 64, kernel: 3, stride: 1
mp + ReLU	kernel: 2 x 2, stride: 2
Dropout	
conv	in-depth: 64, out-depth: 128, kernel: 3, stride: 1
ReLU + Dropout	
fc	in: 4 x 4 x 128, out: 128
Batchnormalization + ReLU + Dropout	
fc	in: 128, out: 6

conv: convolution, **mp:** max pooling, **fc:** fully connected

Fig. 3. The proposed model architectures.

2.4 Implementation Details

To optimize both the prediction of VFMs and that of glaucoma, the losses produced by correlation coefficient (1 – r) and dice coefficient (1 – Dice) were computed and combined with the binary cross entropy (BCE) loss calculated for the prediction of glaucoma. During training, the loss from dice coefficient would be weighted by 0.05 in order to facilitate these models to learn more about the spatial correlation than the exact match at pixel level.

As for model (iii), the loss from the generator was calculated with BCE, indicating how well the predicted VFMs from the generator could cheat the discriminator. This loss was added to the decoder loss described above. On the other hand, the discriminator loss was the sum of the two binary cross entropy, one for how well it made right judgement and the other for how well it classified predicted VFMs as fake VFMs. During training, the generator loss with decoder loss was optimized following the discriminator loss.

In each training for cross validation, hyper-parameters were set as following: (i) Number of epoch: 3000 (ii) Learning rate: 0.0005 (iii) Batch size: 32 (iv) Dropout rate: 0.3 (v) Adam optimizer with beta1 set to 0.9 and beta2 set to 0.999. The results from 5 validation sets would be averaged. These experiments were conducted using Pytorch (version 1.0.0) with two NVIDIA GTX 1080Ti GPU.

3. Results and Discussions

Table 1 shows the performance of these three models on validation dataset. The model (iii) had the highest r (0.7283, moderate-to-high correlation) and Dice over the other two models. This may indicate that using GAN to tackle the mismatch at pixel level between fundus images and VFMs would be better than using STN to learn the affine transformation between them. Moreover, despite of similar accuracy in these models, the model (iii) achieved the highest AUC, suggesting that model (iii) might be more capable of detecting glaucoma suspects out of general population.

Table 1. Averaged cross validation results.

Model	r	Dice	Accuracy (%)	AUC
(i)	0.6940	0.5473	88.52	0.8852
(ii)	0.6931	0.5438	90.98	0.9165
(iii)	0.7283	0.5772	89.34	0.9593

The spatial relationship between fundus images and VFMs may be a non-linear relationship. Compared model (ii) with model (i), STN seemed to have no effect on either r or Dice. This suggested that affine transformation, which is also a linear transformation, was insufficient to describe the spatial relationship between fundus images and VFMs. In contrast, model (iii) achieved better performance than model (i), indicating that such spatial relationship could be addressed by facilitating the model, particularly the generator, to simulate the pixel distribution in VFMs. The spatial relationship might be a non-linear spatial relationship.

With encoder-decoder CNN or even GAN, the model could learn to predict glaucoma based on the features related to RNFLD. Inferred from the results, it seemed that model would have higher ability (by AUC) to discriminate between individuals with glaucoma and those without glaucoma as it could better locate RNFLD (by r and Dice). This further suggested that VFMs could serve as a guide for the model to focus on the features related to RNFLD quickly with comparably limited dataset. Via this training method, the model could provide explainable results to ophthalmologists for large-scale screening, reduce the burden on annotating RNFLD manually on fundus images and possibly reduce the need to collect hundreds of thousands pairs of data.

4. Conclusion

In this study, we demonstrated the potency of GAN to tackle the mismatch at pixel level between fundus images and VFMs, and showed its ability to guide the model to learn features related to RNFLD. With GAN, it may indicate that unsupervised learning on these pairs of data would be possible in the future. This would further reduce the burden both on ophthalmologists and researchers who need to pair each fundus image with its VFM. In addition, this explainable model could be applied in large-scale population screening to detect suspects with early RNFLD which is not located near optic disc. This would ultimately increase the screening rate of suspects with early-stage glaucoma and prevent them from blindness. More data would be included to validate these results.

ACKNOWLEDGEMENT

This work was supported by the Industrial Technology Research Institute (ITRI) of Taiwan with Project Grant No. J301ARB300.

REFERENCES

[1] Kingman, S., “Glaucoma is second leading cause of blindness globally,” *Bulletin of the World Health Organization* vol.82, 11, pp. 887-888, 2004.

[2] Tham, Y.C., Li X., et al., “Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis,” *Ophthalmology* vol.121, 11, pp. 2081-2090, 2014.

[3] Peeters, A., Schouten, J. S., et al., “Cost-effectiveness of early detection and treatment of ocular hypertension and primary open-angle glaucoma by the ophthalmologist,” *Eye* vol.22, 3, pp. 353-362, 2008.

[4] Abramoff, M. D., Garvin M. K., Sonka, M., “Retinal Imaging and Image Analysis,” *IEEE Reviews in Biomedical Engineering* vol.3, pp. 169-208, 2010.

[5] Bussel, I. I., Wollstein, G., Schuman, J. S. “OCT for glaucoma diagnosis, screening and detection of glaucoma progression,” *The British journal of ophthalmology 98(Suppl 2)*, pp. ii15-ii19, 2013.

[6] Alencar, L. M., Medeiros, F. A., “The role of standard automated perimetry and newer functional methods for glaucoma diagnosis and follow-up,” *Indian journal of ophthalmology 59(Suppl1)*, pp. S53-S58, 2011.

[7] Jampel, H. D., Friedman, D., et al., “Agreement among glaucoma specialists in assessing progressive disc changes from photographs in open-angle glaucoma patients,” *American journal of ophthalmology* vol.147, 1, pp. 39-44, 2008.

[8] Dagdelen, K., Dirican, E., “The assessment of structural changes on optic nerve head and macula in primary open angle glaucoma and ocular hypertension,” *International journal of ophthalmology* vol.11, 10, pp. 1631-1637, 2018

[9] Li, Z., He, Y., et al., “Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photography,” *Ophthalmology* vol.125, 8, pp. 1199-1206, 2018.

[10] Phan, S., Satoh, S., et al., “Evaluation of deep convolutional neural networks for glaucoma detection,” *Japanese Journal of Ophthalmology (online)*, pp.1613-2246, 2019.

[11] Medeiros, F. A., Jamal, A. A., Thompson, A. C., “From Machine to Machine: An OCT-Trained Deep Learning Algorithm for Objective Quantification of Glaucomatous Damage in Fundus Photographs, *Ophthalmology*, vol.126,4, pp. 512-521, 2019.

[12] Setiawan, A. W., Mengko, T. R., Santoso, O. S., Suksmono, A.B., “Color retinal image enhancement using CLAHE,” *International Conference on ICT for Smart Society*, pp. 1-3, 2013.

[13] Szegedy, C., Vanhoucke, V., et al., “Rethinking the Inception Architecture for Computer Vision,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.2818-2826, 2016.

[14] Jaderberg, M., Simonyan, K., et al., “Spatial Transformer Networks,” *In Advances in neural information processing systems*, pp. 2017-2025, 2015.