Automated detection of nerve fiber layer defects on retinal fundus images using fully convolutional network for early diagnosis of glaucoma

Ryusuke Watanabe[a], Chisako Muramatsu*[b], Kyoko Ishida[c], Akira Sawada[d], Yuji Hatanaka[e], Tetsuya Yamamoto[d], Hiroshi Fujita[b]
[a]Department of Electrical, Electronic and Computer Engineering, Gifu University, 1-1 Yanagido, Gifu 501-1193, Japan
[b]Department of Intelligent Image Information, Graduate School of Medicine, Gifu University, 1-1 Yanagido, Gifu 501-1194, Japan
[c]Department of Ophthalmology, Ohashi Medical Center, Toho University, 2-17-6 Ohashi, Meguro-ku, Tokyo 153-8515, Japan
[d]Department of Ophthalmology, Graduate School of Medicine, Gifu University, 1-1 Yanagido, Gifu 501-1194, Japan
[e]School of Engineering, The University of Shiga Prefecture, 2500 Hassaka-cho, Hikone, Shiga 522-8533, Japan

ABSTRACT

Early detection of glaucoma is important to slow down progression of the disease and to prevent total vision loss. We have been studying an automated scheme for detection of a retinal nerve fiber layer defect (NFLD), which is one of the earliest signs of glaucoma on retinal fundus images. In our previous study, we proposed a multi-step detection scheme which consists of Gabor filtering, clustering and adaptive thresholding. The problems of the previous method were that the number of false positives (FPs) was still large and that the method included too many rules. In attempt to solve these problems, we investigated the end-to-end learning system without pre-specified features. A deep convolutional neural network (DCNN) with deconvolutional layers was trained to detect NFLD regions. In this preliminary investigation, we investigated effective ways of preparing the input images and compared the detection results. The optimal result was then compared with the result obtained by the previous method. DCNN training was carried out using original images of abnormal cases, original images of both normal and abnormal cases, ellipse-based polar transformed images, and transformed half images. The result showed that use of both normal and abnormal cases increased the sensitivity as well as the number of FPs. Although NFLDs are visualized with the highest contrast in green plane, the use of color images provided higher sensitivity than the use of green image only. The free response receiver operating characteristic curve using the transformed color images, which was the best among seven different sets studied, was comparable to that of the previous method. Use of DCNN has a potential to improve the generalizability of automated detection method of NFLDs and may be useful in assisting glaucoma diagnosis on retinal fundus images.

Keywords: Retinal fundus images, glaucoma, nerve fiber layer defect, deep convolutional neural network, fully convolutional network, computer-aided detection

1. INTRODUCTION

Glaucoma is the second leading cause of blindness in the world, and the number of affected patients is expected to increase in the aging society [1]. Once the retinal nerve is damaged and the loss of visual field occurs, the recovery is very difficult. Therefore, early detection of glaucoma is important for slowing down progression and preventing total vision loss. However, due to the slow progression and the lack of symptom in its early stages, most patients are unaware of having the disease, and diagnosis is delayed [2].

*chisa@fjt.info.gifu-u.ac.jp; phone +81-58-230-6518; fax +81-58-230-6514

Medical Imaging 2017: Computer-Aided Diagnosis, edited by Samuel G. Armato III, Nicholas A. Petrick,
Proc. of SPIE Vol. 10134, 1013438 · © 2017 SPIE · CCC code: 1605-7422/17/$18 · doi: 10.1117/12.2254574

Retinal fundus photographs are often obtained for eye examination, clinical record, and follow-up of disease condition in general ophthalmology visits. It is sometimes obtained in general health checkups as it can noninvasively show vessel conditions which reflect the whole body health. Fundus photographic examination is relatively inexpensive and handy so that it may be suitable for screening.

Early signs of glaucoma that can be found on retinal fundus images include a retinal nerve fiber layer defect (NFLD) and retinal cup enlargement, and NFLDs generally appear before the structural change of optic disc. However, detection of early subtle NFLDs can be difficult, and the number of ophthalmologists specialized in glaucoma is limited. Our earlier study indicated that the variability in diagnostic findings between ophthalmologists exist [3]. Therefore, computer-aided detection of NFLDs can be useful for the efficient and consistent diagnosis of glaucoma.

Odstrcilik et al. [4] conducted the texture analysis based on Gaussian Markov random fields and local binary patterns on retinal fundus images. They found the significant correlation between the predicted output from regression models and retinal nerve fiber layer thickness measured by optical coherence tomography.

In our previous study, we proposed a multi-step NFLD detection scheme which combines the outputs of Gabor filtering, k-means clustering, and adaptive thresholding [5]. Although the result indicated its usefulness for automated NFLD detection, the potential problem was that the method included too many rules with some parameters tuned for the test cases. The classifier was trained with the independent dataset; however, the generalizability was still a concern.

For solving the problem and further reducing the false positives, our purpose in this study was to investigate an end-to-end learning system without pre-specified features and complicated rules. In this preliminary investigation, a deep convolutional neural network (DCNN) was trained for automated detection of NFLD regions, and the results by using different input data were compared.

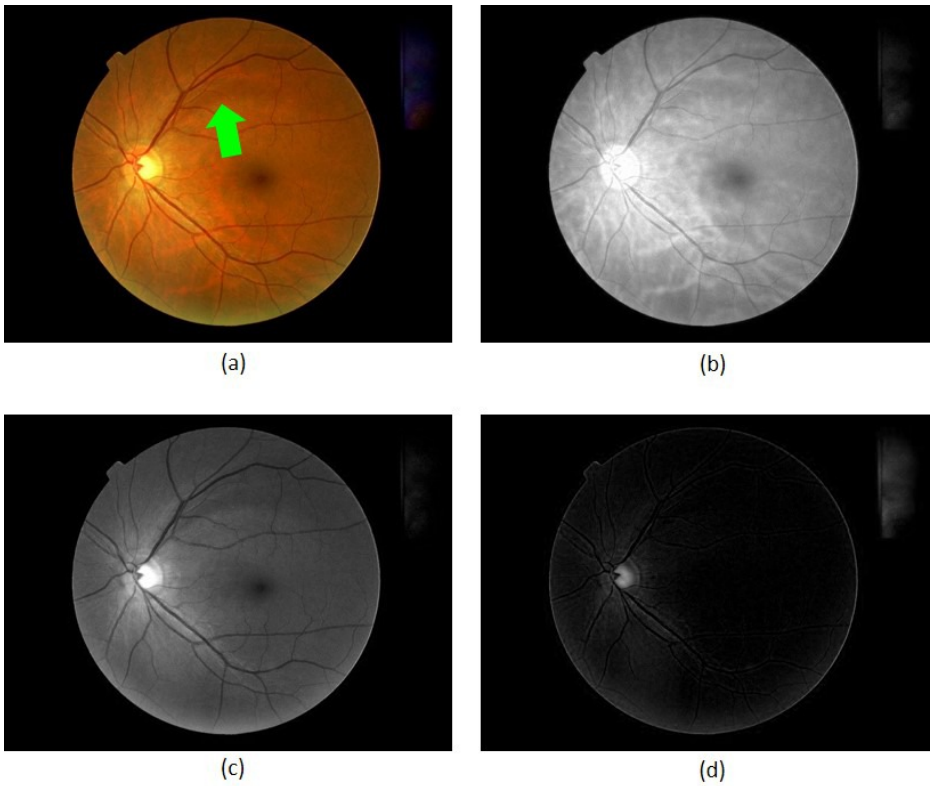

Figure 1. Fundus image with NFLD. (a) Original image and its (b) red component, (c) green component, and (d) blue component

2. METHODS

2.1 Databases

The retinal fundus images used in this study were obtained from the dataset collected as a part of the Eye Health Care Project in Tajimi, Japan [6], which includes clinical data and eye examination results, such as visual field test, intraocular pressure, and retinal fundus images. The cases used in the randomized study, the Tajimi Glaucoma Epidemiological Study [2], were not included in this dataset. The fundus photographs were obtained with IMAGEnet digital fundus camera system (TRC-NW6S, Topcon, Tokyo, Japan) and have a matrix size of 768 x 576 pixels in JPEG format.

The training dataset consists of 162 cases, including 81 cases with NFLDs and 81 age and gender-matched non-glaucomatous cases selected from the original database by a stratified randomization. The positive cases were examined by two ophthalmologists independently, and 99 regions identified as NFLDs by both ophthalmologists that corresponded the original clinical reports were employed as the gold standard of NFLDs in this study [3]. Figure 1 shows a color retinal fundus image with NFLD region and its red, green and blue component images.

The test dataset consists of 261 cases, including 130 cases with NFLDs and 131 age and gender-matched cases without NFLDs by the stratified random selection; one NFLD case was excluded because of the image replication. One ophthalmologist identified the NFLD regions without any other clinical information in the same way as the training dataset. The test dataset included 203 NFLDs that were concordant with the original clinical reports.

2.2 Overview of the previous method

The detailed process of our previous detection method was described elsewhere [5]. Briefly, an original color fundus image was converted to a grayscale image by selecting the green channel in which NFLDs are depicted with the highest contrast as shown in Fig.1(c). In order to exclude blood vessel regions from potential candidates, a "vessel-erased image" [7] was created. The major blood vessels were detected by black top-hat transformation, and the pixels corresponding to the detected vessel regions were interpolated by the surrounding non-vessel pixels.

Subsequently, the images were quasi polar transformed using ellipse functions so that the directions of the retinal nerves are roughly vertically aligned as shown in Fig. 2. The preprocessing, including background correction, linear contrast enhancement, unsharp masking, and median filtering, was applied to improve visibility of NFLDs. The candidates of NFLDs were detected by applying Gabor filters of four sizes. The second and third candidates were detected by applying adaptive thresholding and k-means clustering. The Gabor filter candidates that overlapped with both the second and third detection results were remained. After rule-based false positive reduction of blood vessel regions, 11 features were used for additional false positive reduction by linear discriminant analysis (LDA).

2.3 Deep convolutional neural network with fully convolutional layers

In this study, we investigated the use of a fully convolutional network (FCN) proposed by Long et al. [8] which was customized from VGG 16-layer net. It includes three rounds of training, called FCN-32s, FCN-16s, and FCN-8s, which provide increasingly more detailed outputs. The detailed network architecture is described in [8]. Figure 3 shows the architecture of FCN-8s. It consists of 7 sets of convolutional layers, 5 max pooling layers, and 3 deconvolutional layers.

2.4 Input training data

We investigated the detection performance of the FCN trained with different input image sets: (a) Original color images of abnormal cases, (b) original color images of both normal and abnormal cases, (c) ellipse-based polar transformed color images, (d) transformed G plane images, (e) transformed Gabor filtered color images, (f) transformed color halved images, and (g) transformed color halved images with different data augmentation. For (f) and (g), transformed image was divided into half, each of which corresponds to upper temporal and lower temporal regions, and those including at least one NFLD were only employed to include a larger fraction of positive pixels in training. For all transformed images, vessel-erased images were used.

For each data set, data augmentation by rotation and intensity transformation was performed. For (a) and (b), images were rotated from -30 to +30 degrees by step of 2 degrees, which range covers the frequent site of NFLDs. For (c) to (g), images were rotated by ±2 degrees and flipped horizontally. In (a), and (b), intensity was shifted by ±5 and ±10, whereas in (e) and (g), intensity was shifted from -10 to +10 by step of 2. In the remainings, intensity was transformed by using a gamma function, with the gamma parameter changed from 0.8 to 1.4 by step of 0.05.

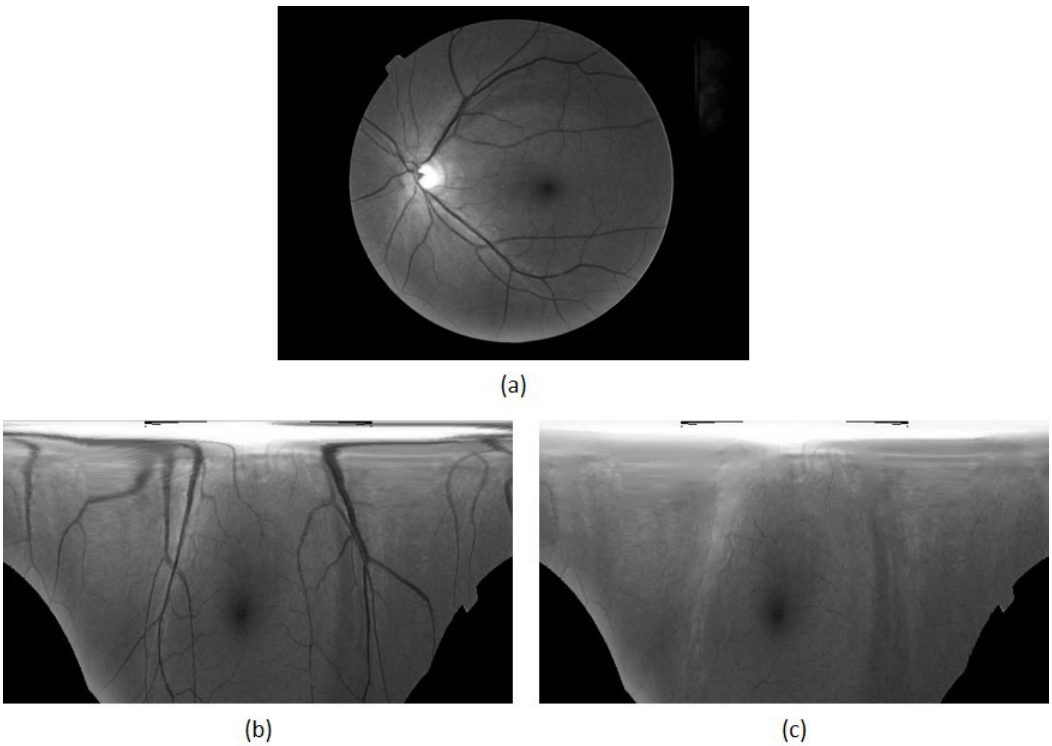

Figure 2. Modified polar transformation based on ellipses. (a) Original green component image, (b) transformed image, and (c) transformed "vessel erased" image

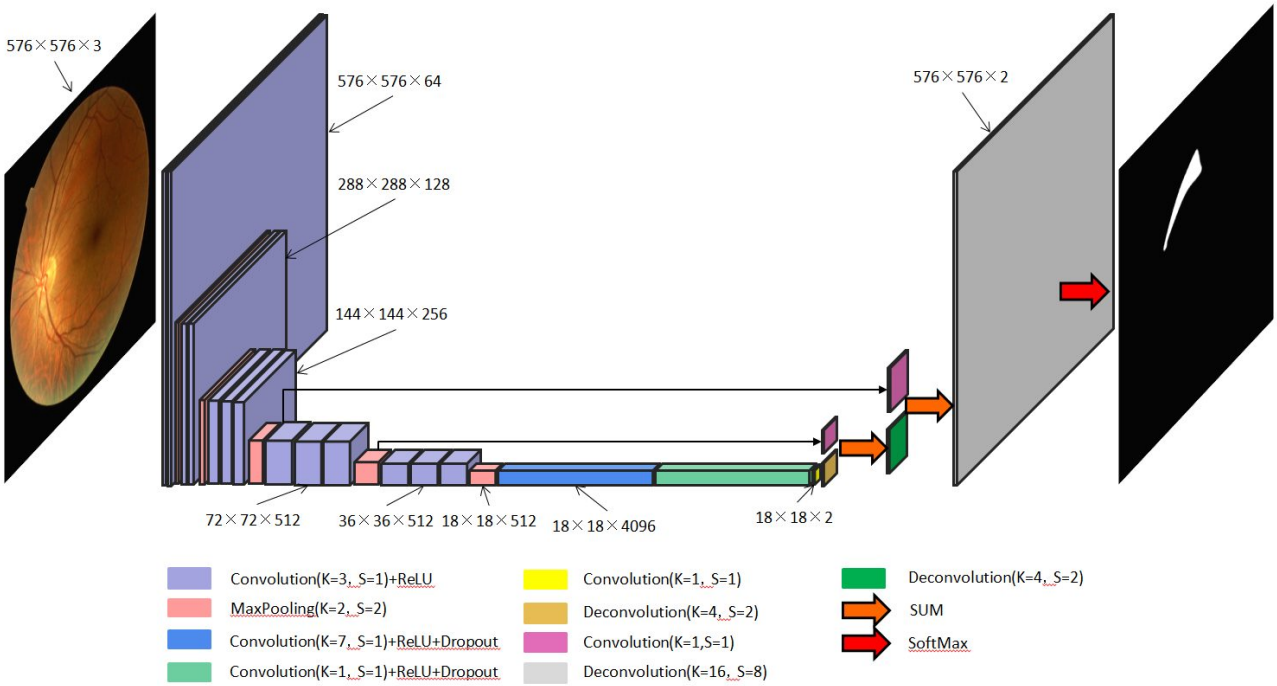

Figure 3. Fully convolutional network architecture (FCN-8s)

As a result, the numbers of training images became 12555, 25110, 6318, 6318, 5346, 7098, and 6006 from (a) to (g), respectively. The sizes of input images were 576 x 576 pixels for the original images, 540 x 300 pixels for the transformed images, and 270 x 300 pixels for the transformed half images.

2.5 Probability output

The original FCN by Long et al. [8] outputs the score of 0 or 1, corresponding to background or foreground, for each pixel. In this study, we added a softmax layer for testing to retrieve probability of a pixel being in NFLD regions. In this way, we are able to set an arbitral threshold for detecting NFLD regions.

2.6 Evaluation

The detected region was considered as a true positive if the region was overlapped with the gold standard NFLD regions, as in the previous studies [3, 5]. If multiple regions were detected in the same gold standard NFLD region, only one was counted as the true positive. The results were evaluated using a free response receiver operating characteristic (FROC) curve by varying the threshold value from 0 to 100% by step of 5. The best result was compared with that of the previous study.

3. RESULTS

Table 1 shows the sensitivities and the numbers of false positives (FPs) per image when the threshold was set to 50%, as in the original FCN. When using both the normal and abnormal cases for training, the number of false positives was decreased at the cost of reduced sensitivity. Using the transformed images, the sensitivity and false positive rate were generally improved. In attempt to improve sensitivity by balancing the positive and negative pixels, the transformed images were halved and only the upper and/or lower sides with NFLDs were used for training. This seemed reasonable as NFLDs are frequently found in upper temporal and lower temporal regions almost symmetrically. However, although the sensitivity was slightly improved, the number of FPs was largely increased. NFLDs are most clearly observed in G component image and barely noticeable in B component image. Nonetheless, using color images provided slightly better performance than using G plane images only. The use of Gabor filtered images was not useful in this study. The result suggests that such filter could effectively be learned by CNN. Among 7 sets of input data, set (c) provided the best performance.

The result was compared with that by the previous method. Figure 4 shows FROC curves of the previous and proposed methods. The highest sensitivity by both methods was 98%, while the number of FPs at this sensitivity was slightly lower in the current study (5.42 vs 6.20). The FROC curves by the two methods were comparable. We were able to obtain similar performance yet with more simpler and generalized method.

Table 1. Sensitivities and false positive rates for different input data

Dataset	Input	Number of training images	Sensitivity (%)	Number of false positives / image
(a)	Original (abnormal only)	12555	68	1.20
(b)	Original (all)	25110	62	0.47
(c)	Transformed	6318	71	0.59
(d)	Transformed (G plane only)	6318	66	0.28
(e)	Transformed (Gabor filtered)	5346	61	0.52
(f)	Transformed and halved	7098	74	2.04
(g)	Transformed and halved	6006	73	1.83

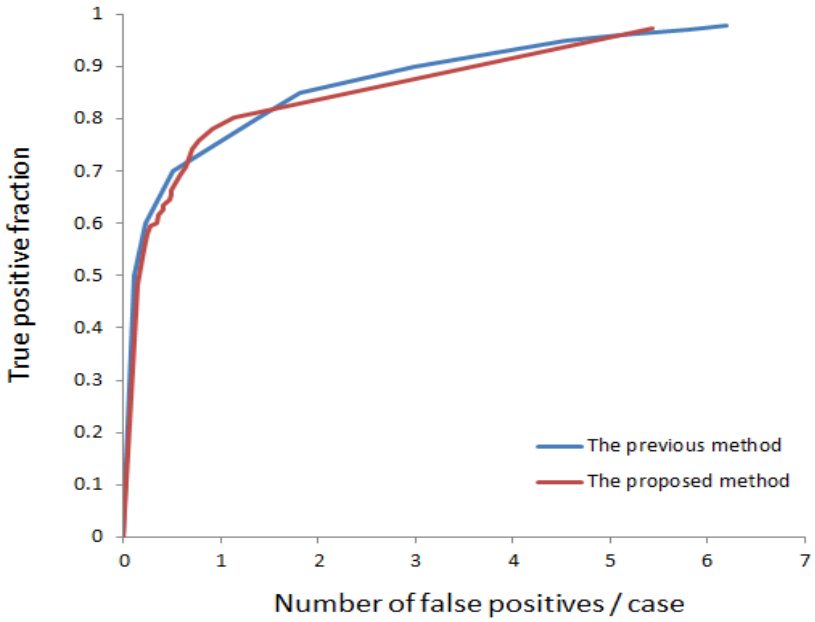

Figure 4. FROC curves for detection of NFLDs by the previous method and proposed method

4. CONCLUSION

We investigated the use of FCN for automatic detection of NFLDs on retinal fundus images. Among different types of input images studied, the best performance in terms of the sensitivity at a reasonable number of false positives was obtained using unfiltered color images with quasi-polar transform. The results indicate that by inputting unprocessed images, CNN was able to learn features that are effective for NFLD detection. The result in this study was comparable to that in the previous study suggested by the FROC curves. In the previous study, however, many rules and parameters, such as the parameters of preprocessing methods, Gabor filter parameters, adaptive thresholds, thresholds for k-means clustering outputs, and thresholds for the feature values in rule-based scheme, were experimentally determined on the basis of the test results. Using the proposed study, such parameters could be automatically learned. The method could be further improved by optimizing the network architecture and adding post-processing method.

ACKNOWLEDGEMENTS

This study was partly supported by a research grant by the Mazda Foundation and a Grant-in-Aid for Scientific Research on Innovative Areas (Multidisciplinary Computational Anatomy, No. 26108005) by Ministry of Education, Culture, Sports, Science and Technology, Japan.

REFERENCES

[1] Quigley, H.A. and Broman, A. T., "The number of people with glaucoma worldwide in 2010 and 2020," Br. J. Ophthalmol. 90, 262-267 (2006).
[2] Iwase, A., Suzuki, Y., Araie, M., Yamamoto, T., Abe, H., Shirato, S., Kuwayama, Y., Mishima, H. K., Shimizu, H., Tomita, G., Inoue, Y. and Kitazawa, Y., "The prevalence of primary open-angle glaucoma in Japanese; the Tajimi Study," Ophthalmol. 111(9), 1641-1648 (2004).

[3] Muramatsu, C., Hayashi, Y., Sawada, A., Hatanaka, Y., Hara, T., Yamamoto, T. and Fujita, H., "Detection of retinal nerve fiber layer defects on retinal fundus images for early diagnosis of glaucoma," J. Biomed. Opt. 15(1), 0160211-016021-7 (2010).
[4] Odstrcilik, J., Kolar, R., Tornow, R. P., Jan, J., Budai, A., Mayer, M., Vodakova, M., Laemmer, R., Lamos, M., Kuna, Z., Gazarek, J., Kubena, T., Cernosek, P., and Ronzhina, M., "Thickness related textureal properties of retinal nerve fiber layer in color fundus images," Comput Med Imaging Graph. 38, 508-516 (2014).
[5] Muramatsu, C., Ishida, K., Sawada, A., Hatanaka, Y., Yamamoto, T., and Fujita, H., "Automated detection of retinal nerve fiber layer defects on fundus images: false positive reduction based on vessel likelihood," Proc. SPIE Med Imaging 9785, 97852L1-97852L6 (2016).
[6] Yamamoto, T., Iwase, A., Araie, M., Suzuki, Y., Abe, H., Shirato, S., Kuwayama, Y., Mishima, H. K., Shimizu, H., Tomita, G., Inoue, Y. and Kitazawa, Y., "The Tajimi Study report 2: prevalence of primary angle closure and secondary glaucoma in a Japanese population," Ophthalmol. 112(19), 1661-1669 (2005).
[7] Nakagawa, T., Suzuki, T., Hayashi, Y., Mizukusa, Y., Hatanaka, Y., Ishida, K., Hara, T., Fujita, H. and Yamamoto, T., "Quantitative depth analysis of optic nerve head using stereo retinal fundus image pair," J. Biomed. Opt. 13(6), 064026 (2008).
[8] Long, J., Shelhamer, E., and Darrel, T., "Fully convolutional networks for semantic segmentation," ArXiv 1411.4038 (2015).