Development of an End-to-End Deep Learning System for Glaucoma Screening Using Color Fundus Images

Joonseok Lee[1], JoonHo Lee[1], Hyoseob Song[1], and Chihoon Lee[1, 2]

[1] AI Research Center, Technology Research, Samsung SDS, Seoul, Korea
[2] Artificial Intelligence Team, SDSRA, Samsung SDS, San Jose, CA USA

{js1985.lee, joonholee, hyoseob.song, lee.chihoon}@samsung.com

Abstract. Deep learning approaches are being applied to detect glaucoma from retinal fundus images. Previous studies, however, have shown less than enough performance to be used for automated glaucoma screening. In the paper, we proposed a novel end-to-end glaucoma detection system and associated methods. Optic disc (OD) and optic cup (OC) segmentation method is proposed, which is developed from M-Net but defines OD/OC segmentation as cascaded single-labeled segmentation tasks and has many improvements. A coarse OD detector is trained to take advantage of the knowledge of OD position and then used to prepare and enhance training data. A fine OD segmentation is conducted from the prepared data and then OC segmentation is conducted from the tightly cropped area resulting from the OD segmentation. In addition to this 'divide-and-conquer' methods, our OD/OC segmentation methods comprise decent pipeline including preparation, pre-processing, deep learning and post-processing. Glaucoma screening system is then proposed as combination of OD/OC segmentation method and a glaucoma classification method where separate convolutional neural networks are trained to learn features from decently arranged crop image including features around OD/OC and retinal nerve fiber layer defect (RNFLD). When trained and validated using REFUGE dataset, dice coefficient was 0.9426 for OD, 0.8565 for OC and mean absolute error of CDR was 0.0507 in OD/OC segmentation task. AUC was 0.9637 and sensitivity was 0.9 in classification task. Further improvement is expected via applying more training datasets and fine-tuning.

Keywords: OD/OC segmentation, glaucoma detection, deep learning.

1 Introduction

Glaucoma is one of the leading causes of irreversible blindness. Early screening and detection methods for glaucoma have been researched for that reason. In general, intraocular pressure leads to torsion of the optic nerve and retinal nerve fiber layer defect (RNFLD) and results in glaucoma. Glaucoma suspect can be detected from retinal fundus images whose key features are torsion of optic disc (OD), high ratio of the vertical cup to disc (CDR), pale disc, and disc hemorrhage.

In the whole fundus photograph, the area to be observed for classification of glaucoma is the region where OD and RNFLD can appear with very high probability in case. RNFLD is mainly a pattern starting from the OD and extending inward to the vessels, and it may appear microscopically difficult to read unless it is an ophthalmologist.

Optic cup (OC) is the cup-shaped part which is located at the center of the OD. A Lager CDR is a well-accepted indicator of glaucoma but determining CDR is affected by each clinician's experience and opinion. Though it can be observed when there are other neuro-ophthalmic diseases than glaucoma, previous studies showed that a larger CDR is closely related with glaucoma progression and thus useful in evaluation of glaucoma.

Most of the previous studies for large-scale screening of glaucoma were based on hand-crafted features to measure CDR, but deep learning approaches especially using convolutional neural networks (CNNs) have been successful to demonstrate better performance than previous approaches such as U-Net [6], M-Net [7], DE-Net [8]. However, those state-of-the-art approaches even need to be improved in that the CDR is not accurate enough and area under curve (AUC) of receiver operating characteristics (ROC) is not high enough to be used for large-scale screening of glaucoma.

In this paper, we proposed an improved OD and OC segmentation method which is fully automated and defined as a cascaded single-label tasks while M-Net uses the multi-label setting which is the baseline method of ours and thus the same segmentation performance measures such as dice coefficient and vertical CDR are used for comparison. In the meantime, a separate CNN is trained to observe more features including disc hemorrhage and RNFLD and predict glaucoma suspect, which shows high accuracy in terms of AUC. Placing the above methods into a well-arranged pipeline results in a state-of-the-art performance in segmentation and classification tasks on glaucoma using only 400 training images from REFUGE

The remainder of this paper consists of the followings. In section 2, previous works related with OD/OC segmentation and glaucoma classification are reviewed. In section 3, the details of our methodology are presented. Experiments to measure the performance of our methodology are described in section 4, and then we concluded this paper in section 5.

2 Related Works

Since manual annotation is labor intensive, researchers have sought automatic methods for OD/OC segmentation. A number of methods have been presented in the literature for automatic segmentation of OD and OC [9], which are recently followed by deep learning methods such as U-Net [6], M-Net [7], DE-Net [8].

Research in this area has primarily focused on segmentation of OD, using various techniques such as intensity gradient analysis, Hough transforms [10], template matching, pixel feature classification [6], vessel geometry analysis, deformable models [3] and level sets [4,13]. Some OD segmentation techniques based on active contours [11, 12] and morphological feature [12] are capable of producing reliable OD boundary.

However, their performance still depends on the initialization and the ability to identify weak edges of neuro retinal rim in fundus images.

OC segmentation is more challenging because the depth information is not available in 2D retinal fundus images. As a result, OC boundary is ill defined and in-homogeneous which makes the segmentation more difficult.

Previous OC detection techniques are based either on classifying pixels as part of the cup or rim (the disc area outside the cup) [2, 5] or on an analysis of sliding windows [1]. Other OC segmentation methods are based on level sets [13], super-pixels classification [14] and sparse dictionary learning [16]. In another method [15], fusion of cup segmentations from multi-view fundus images was performed to improve the performance. In some recent work, the presence of glaucoma in fundus images is predicted by classification using SVM [18] and deep feature learning [17], thereby bypassing the OD/OC segmentation.

3 Proposed Method

Our end-to-end glaucoma screening system is illustrated in Fig. 1. Each component of the framework is elaborated on in the following sections.

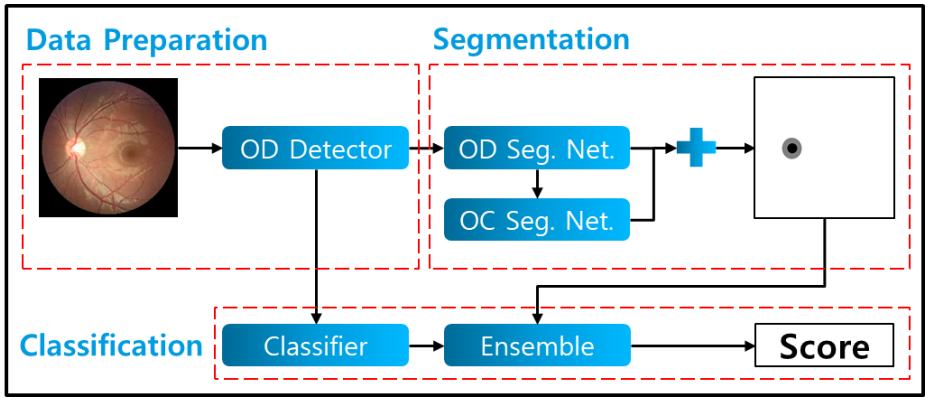

Fig. 1. Illustration of our segmentation and classification framework, which mainly includes disc detector, CNN based OD/OC segmentation, and Glaucoma classifier.

3.1 OD Detection

OD detection is conducted automatically using CNN which has the almost same architecture as M-Net but it has differently prepared input images and changed the number of output channel to one. M-Net is a variation of U-Net and has introduced multi-label setting with polar transformation to 400×400 input images. In our method, 400×400 entire fundus images are used as input after resizing from the original images and applying polar transformation to them. In addition, histogram matching technique is applied during the test. Then the contours of OD area are extracted to specify the center

of OD. By doing so, the position of OD is properly determined and that information is repeatedly used to decide how to preprocess the given fundus image. This helps for the performance of the other methods to be improved.

3.2 OD/OC Segmentation

OD segmentation is conducted automatically using a CNN which has similar architecture to OD detection and trained with REFUGE Training 400 dataset from the scratch. In order to enhance the performance compared with M-Net, novel methods are used through the whole of training-test pipeline. The originality in our OD/OC segmentation method is as follows.

A. Divide and Conquer

We separate this task into OD segmentation out of the whole fundus image and OC segmentation out of the bounding box which is tightly cropped to contain whole of the segmented OD. They can be merged if necessary after both task is completed.

M-Net is successful to avoid OD/OC class imbalance problem via multi-label setting, but side effects are often observed such that OC is classified as OD and vice versa. This makes the accurate segmentation still difficult to achieve. Similarly, multi-label loss function can cause relatively biased optimization result to one of OD and OC depending on the weight setting.

To tackle these issues, we introduce divide-and-conquer strategy to OD/OC segmentation resulting in relatively two easy tasks. Predefined OD detector is used to specify the center of OD as described in section 3.1 and then each 480×480 image is prepared where the OD is located at the center of the crop and the polar transformation is applied to the cropped image. This decently prepared image is used as input to the OD detection CNN to produce refined OD segmentation. Thanks to the well- prepared input, the result of OD segmentation outperforms M-Net under the REFUGE validation dataset.

Second, OC segmentation is conducted out of the bounding box which is tightly cropped to contain whole of the detected OD as mentioned. As most of the time the bounding box is not a square, is resized to 320×320. Polar transformation is not applied in this case because OC is probably not placed right in the middle of the bounding box tightly-cropped on OC and sometimes OC can span most of the area in the square after polar transformation applied if CDR is rather large. In this setting, the region of interest (ROI) to conduct OC segmentation is drastically decreased and the task is just a single-labeled problem, thus false positives are decreased accordingly. Finally, the resulting OD mask and OC mask are merged to complete OD/OC segmentation.

B. Pre-processing

The quality and characteristic vary depending on the vendor of the camera. For instance, the REFUGE Training 400 and REFUGE Validation 400 dataset used in this paper are taken by Zeiss camera and Canon camera for each. This can increase the variance between training dataset and test dataset which may result in lower performance. To overcome this problem, histogram matching technique is applied during testing. The average of the histogram in the training dataset as a template was generated

and validation dataset was matched to it as shown in the Fig.2. Also, although the resolution of the training set and the validation set are different, the size of the ROI crop was adjusted to be consistent with the overall image size.

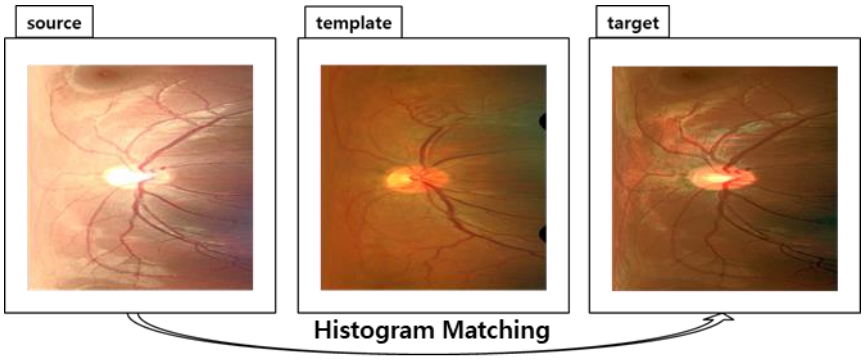

Fig. 2. A sample image of the REFUGE validation dataset was matched to the average histogram template of the REFUGE training dataset.

C. Post-processing

The following post-processing is applied to improve performance for the binary-color mask image obtained from OD segmentation and OC segmentation each. Assuming OD and OC are in the form of elliptic curve, the largest connected area is chosen in the mask and the ellipse fitting is used so as to make a result in final segmentation. Through this method, partially inaccurate predictions occurred during segmentation are corrected.

D. Data augmentation

During training of OD segmentation CNN, random horizontal shifting and random horizontal flip are applied to the polar transformed images in order to delay over-fitting. In case of OC segmentation CNN, horizontal flip and rotation by 90°, 180° and 270° are applied to delay over-fitting while polar transformation is not applied.

3.3 Glaucoma Classification

Well-known ResNet50 is separately trained and used to screen glaucoma combined with OD/OC segmentation CNN. To keep consistency with the image preparation method described above, there are some augmentations in the network architecture in that the size of the input is increased to 600 × 600 and the final output class is adjusted to be either glaucoma or non-glaucoma. This network is fine-tuned with REFUGE Training 400 from the model pre-trained with Image-Net except for fc layer. Although the resolution of the training set and the validation set are different, the size of the ROI crop was adjusted to be consistent with the overall image size. The originality of our methods are as follows.

A. Pre-processing

As mentioned above, RNFLD can appear in a fine pattern, thus scaling down input image can cause all of the features to disappear or to be hardly observed. For that reason, it is helpful to increase performance that high-resolution images are fed as input of the network. It is, however, very expensive to conduct convolutional operations to a high-resolution image. Therefore, our method first conducts finding the key part of the image and next cropping them to make an ROI which should be fed as an input to the network. To find ROI from the whole fundus image, pre-trained OD detector is used to specify the center of OD as described in Section 3.1. To make RNFLD well included in the receptive fields of the network, the entire fundus image is rotated by 45° clockwise. To locate the OD in the upper left corner of the image, it is cropped by 600 × 600 size image. Also, as mentioned in Section 3.2.*B*, histogram matching is applied during the test to compensate image variance per camera vendor.

B. Data augmentation

Glaucoma data accounts for only 10% of total data REFUGE Training 400. Therefore data augmentation is conducted to avoid possible problems caused from the class imbalance. When an ROI is created in the pre-processing phase, the glaucoma entire fundus images rotate at 40° or 50° as well as 45°. Horizontal flipping or vertical flipping is not used and instead only used is flipping perpendicular to the line from the upper left corner to the lower right corner to keep the position where OD and RNFLD can appear in the tightly arranged regions.

4 Experiments

4.1 Dataset

In this paper, M-Net is trained with REFUGE Training 400 dataset from the scratch while M-Net is trained with ORIGA dataset. REFUGE dataset consists of 1200 color fundus images, and training, validation, and test set, each of it includes 400 color fundus images. Training set include additionally OD and OC mask images. Glaucoma images account for 10% for each set. As training set is taken by Zeiss camera and set of validation and test are taken by Canon camera, quality and characteristic of images have significant differences. In this paper, experiments has been done using REFUGE dataset: Training 400 for training Validation 400 for validation.

4.2 Segmentation

To evaluate segmentation performance, dice coefficients of each OD and OC, and mean absolute error of vertical CDR of OD and OC are calculated. Proposed method enhances 0.16 of OC dice coefficient and 0.05 of OD dice coefficient compared to M-Net implementation.

Table 2. Segmentation performance

Performance	M-Net	Proposed method
OC Dice coefficient	0.69543592182	0.856473323
OD Dice coefficient	0.893317627973	0.942646310
MAE CDR	0.062101458873	0.050673643

4.3 Classification

AUC and sensitivity are calculated to evaluate classification performance according to convention of the REFUGE 2018 glaucoma challenge. Through the proposed method, AUC and sensitivity are 0.963680556 and 0.9 respectively. The ensemble classifier's performance is still in the process of validation using REFUGE 2018 system and it is expected to show better performances.

5 Conclusion

In this paper, we have proposed an end-to-end glaucoma screening system with its component-wise methods. It consists of a high-performance OD/OC segmentation CNN which is based on M-net, but composes a novel pipeline with divide-and-conquer and ensemble methods and a glaucoma detector with a higher accuracy, which is built by combining a separately trained CNN and the OD/OC segmentation CNN.

Our methods have achieved a good performance using only REFUGE dataset. It will be a next step for us to measure performance using as many available datasets as possible including ORIGA dataset which is very popular for glaucoma screening. Better performance and more clear comparison to state of the arts can be achieved.

References

1. Abramoff M., Alward W., Greenlee E., Shuba L., Kim C., Fingert J., Kwon Y.: Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features. Invest Ophthalmol Vis Sci 48(4), 1665–73 (2007)
2. Liu J.,Wong D.W.K., Lim J.H., Li H., Tan N.M., Zhang Z.,Wong T.Y., Lavanya R.: ARGALI: an automatic cup-to-disc ratio measurement system for glaucoma analysis using level-set image processing. In: Int. Conf. Biomed. Eng. (2008)
3. Li C., Xu C., Gui C., Fox M.: Level set evolution without re-initialization: A new variational formulation. In: CVPR 430–6 (2005)
4. Wong D.W.K., Lim J.H., Tan N.M., Zhang Z., Lu S., Li H., Teo M., Chan K., Wong T.Y.: Intelligent fusion of cup-to-disc ratio determination methods for glaucoma detection in ARGALI. In: Int. Conf. Engin. in Med. and Biol. Soc. 5777–80 (2009)
5. O. Ronneberger, P. Fischer, and T. Brox: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: MICCAI 2015, pp. 234–241, Springer, Heidelberg (2015).
6. Huazhu Fu, Jun Cheng, Yanwu Xu, Damon Wing Kee Wong, Jiang Liu, and Xiaochun Cao: Joint Optic Disc and Cup Segmentation Based on Multi-label Deep Network and Polar Transformation, IEEE Trans. Med. Imag, vol. 37, no. 7, pp. 1597–1605, (2018).

7. Huazhu Fu, Jun Cheng, Yanwu Xu, Changqing Zhang, Damon Wing Kee Wong, Jiang Liu, and Xiaochun Cao: Disc-aware Ensemble Network for Glaucoma Screening from Fundus Image, IEEE Trans. Med. Imag, (2018). DOI: 10.1109/TMI.2018.2837012
8. Almazroa, Ahmed, et al. Optic Disc and Optic Cup Segmentation Methodologies for Glaucoma Image Detection: A Survey, J. Ophthalmol, (2015).
9. X. Zhu, R. M. Rangayyan, and A. L. Ells, Detection of the Optic Nerve Head in Fundus Images of the Retina Using the Hough Transform for Circles, Journal of Digital Imaging, vol. 23, no. 3, pp. 332-341, (2010).
10. Mittapalli, P.S. and Kande, G.B. Segmentation of optic disk and optic cup from digital fundus images for the assessment of glaucoma. Biomedical Signal Processing and Control, vol. 24, pp. 34-46, (2016).
11. A.Aquino, M. Gegundez-Arias, and D. Marin., Detecting the optic disc boundary in digital fundus images using morphological edge detection and feature extraction techniques, IEEE Trans. Med. Imag, vol. 20, no. 11, pp. 1860-1869, (2010).
12. D.W.K.Wong and et. al., Level set based automatic cup to disc ratio determination using retinal fundus images in argali, in Proc. IEEE EMBC, pp. 2266-2269, (2008).
13. Y. Xu, L. Duan, and et. al., Optic cup segmentation for glaucoma detection using low rank superpixel representation, in Proc. MICCAI Part 1, pp. 788-795, (2014).
14. Joshi, G.; Sivaswamy, J.; Krishnadas, S.R., Depth Discontinuity-Based Cup Segmentation from Multiview Color Retinal Images, IEEE Trans. Biomed. Eng., vol.59, no.6, pp. 1523-1531, (2012).
15. Chakravarty. A, Sivaswamy. J, Coupled Sparse Dictionary for Depth-Based Cup Segmentation from Single Color Fundus Image, In: MICCAI 2014, pp. 745-754, (2014).
16. Chen, Xiangyu, et al. Automatic Feature Learning for Glaucoma Detection Based on Deep Learning. In: MICCAI 2015, pp. 669-677, (2015).
17. Kohler, Thomas, et al. Computer-Aided Diagnostics and Pattern Recognition: Automated Glaucoma Detection. Teleophthalmology in Preventive Medicine. Springer Berlin Heidelberg, pp. 93-104, (2015).