COMBINING MULTIPLE DEEP FEATURES FOR GLAUCOMA CLASSIFICATION

Annan Li   Yunhong Wang

Beijing Advanced Innovation Center
for Big Data and Brain Computing,
Beihang University Beijing, China 100191

Jun Cheng   Jiang Liu

Ningbo Institute of Industrial Technology,
Chinese Academy of Sciences,
Zhengjiang 315201, China.

ABSTRACT

Glaucoma is one of the leading cause of blindness. Although there is still no cure, early detection can prevent serious vision loss. Therefore automated glaucoma detection/classification is an important issue. In the past decade, segmentation based approach such as those based on cup-to-disc-ratio are popular, but single indicator limit its performance. Recently, convolutional neural network based image classification approaches that can use more image cues achieve good performance. In this paper, we propose a new glaucoma classification by combining multiple features extracted by different convolutional neural networks. Its effectiveness is clearly demonstrated on the publicly available Origa [1] dataset. It achieves an area under the receiver operating characteristic curve of 0.8483, which better than the 0.838 given by on manual marked cup-to-disc-ratio. To our knowledge, it is the first approach surpass human in glaucoma classification.

Index Terms— Glaucoma classification, convolutional neural network, feature fusion

1. INTRODUCTION

Glaucoma is a group of diseases that damage the eyes optic nerve and can result in vision loss. It is reported that glaucoma is the second-leading cause of blindness after cataracts [2]. There is no cure for glaucoma. Vision lost from the disease cannot be restored. However, with early detection and treatment, serious vision loss can be avoided. Therefore, screening the glaucoma is very important. Examination of the optic nerve via funduscopy is a necessary method in glaucoma screening. The risk of glaucoma can be assessed by analyzing the visible damage to optic nerve or change in the cup-to-disc ratio and also rim appearance and vascular change. Besides manual examination, automated glaucoma screening or detection is emerging in the past decade.

Optic disc, optic cup, peripapillary atrophy (PPA) and retinal nerve fibre layer (RNFL) are the four kinds of structures considered relevant to glaucoma [3]. One of the most important indicators for glaucoma screening is the cup-to-disc-ratio (CDR). Due to the high pressure, the CDR in glaucomatous eyes is larger than in normal eyes. Consequently, diagnosis can be automated achieved by segmenting optic disc and optic cup [4, 5]. RNFL lies on top of the retina. Its defect is another important indicator. Two kinds of abnormal phenomenons can be observed in color fundus images: a wedge-shaped dark area with its tip touching the border of optic disc and the disappearance of bright bundle striations around optic disc. Studies on automated RNFL defect detection can be found in [6, 7]. There are two kinds of PPA:α-PPA and β-PPA. The former lies in the peripheral zone while the latter is usually in more central zone. In glaucomatous eye is that the β-PPA appears more frequently in the temporal side of the optic disc. Therefore, automated PPA detection [8, 9] can be also applied to glaucoma screening.

CDR based glaucoma detection is popular. However single indicator based approach is limited. An alternative approach is treating glaucoma detection as a general image classification problem. Recent works [10, 11] have proved that image classification, especially based on convolutional neural networks (ConvNets), is feasible for glaucoma detection. The performance can be very close to those given by manual CDR. Although the pathology behind classifier is difficult to explain, an obvious advantage is that general image classification utilizes more information. It is not surprise that it can outperform single indicator based approach.

ConvNets applied to glaucoma detection are pre-trained on large-scale datasets for general image classification. Began with the *AlexNet* [12], architecture of ConvNets has evolved four generations [13, 14, 15], which provide multiple choice of deep features. We observe that different ConvNets can capture different information of fundus appearance, which implies they may be complementary. In this paper, we propose a simple yet effective method for automated glaucoma classification by combining the features of multiple ConvNets. Experimental results on the Origa dataset [1] show that the proposed method outperforms manual marked CDR. To our knowledge, it is the first automated glaucoma classification approach that surpasses human.

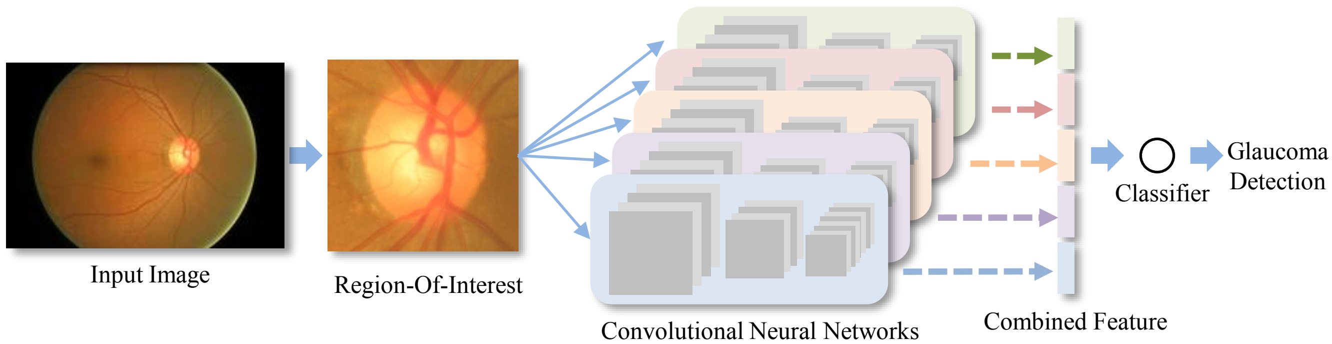


Fig. 1: Pipeline of glaucoma detection.

2. METHODOLOGY

As shown in Figure 1, the proposed glaucoma classification method consists of three steps: region-of-interest (ROI) extraction, deep feature extraction and the glaucoma classification.

2.1. Region of Interest

The visible damage or change caused by glaucoma are mainly at the optic disc area. Therefore, to enhance the performance, locating a region-of-interest centered at the optic disc is a necessary step. In this paper we adopt the ROI extraction method used in [11]. In fundus or retinal image analysis, optic disc detection is usually formulated as a single point estimation problem, in which the detection is considered successful if the point falls into the mask of optic disc. As can be seen, the accuracy of such detection is insufficient. To deal with this problem, we refine the disc center by detecting the disc boundary. As shown in Figure 2 (a) and (d), we first detect the

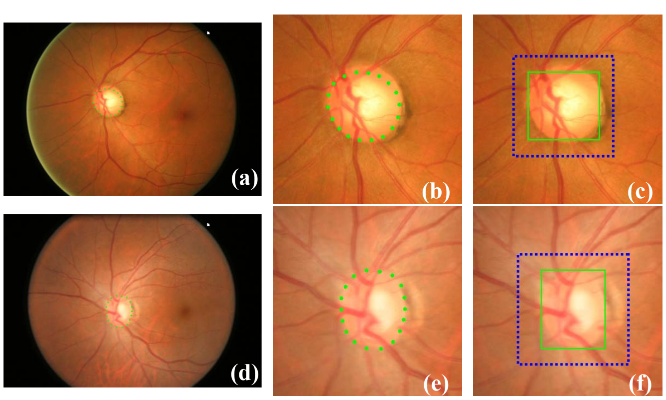


Fig. 2: Region-of-interest (ROI) extraction. From input images (a,d), the optic disc (b,d) is localized by a deformable shape model [16]. ROI (c,f) is a square (dashed square) centered at the optic disc.

disc boundary using a deformable shape model [16]. The outputs are 18 landmarks that can effectively represent the shape of optic disc (see Figure 2 (b) and (e)). With these landmarks, a rectangle bounding box of the disc can be obtained. The ROI is defined as a square patch centered at the bounding box center. The side length of the ROI is determined by 1.4 times of the longest side of the bounding box. All ROIs are normalized to 256×256. Exemplar bounding boxes (blue dashed squares) are shown in Figure 2 (c) and (f).

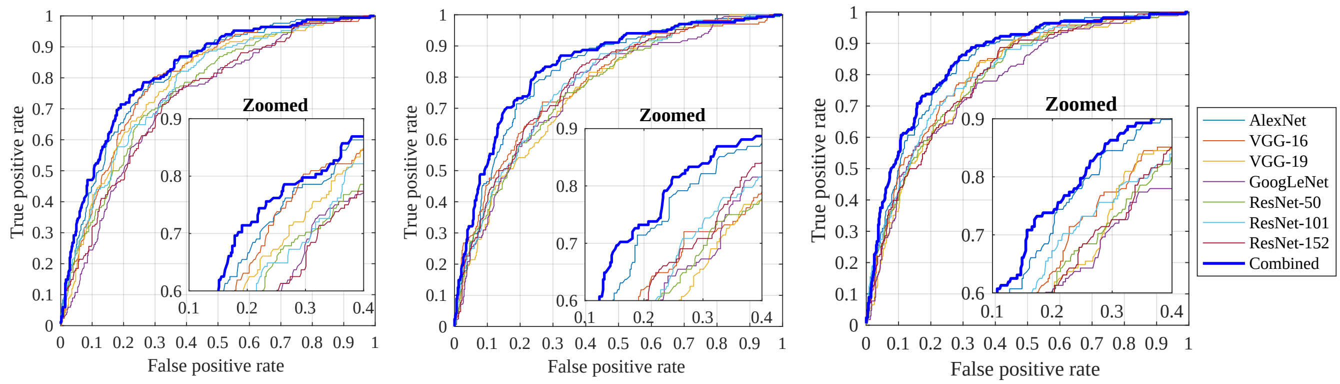
2.2. Deep Feature Extraction

With the normalized ROI, visual features can be extracted using ConvNets. The layers of a typical ConvNet can be divided into three categories, i.e. convolutional layer, fully connected layer and softmax layer respectively. A convolutional layer convolve the input image into multiple channels with reduced two dimensional size, which form a three dimensional tensor. Stacking several convolutional layers, the tensor size is increased and then reduced layer by layer. After the convolution operation, the three dimensional tensor is reshaped to an one dimensional vector and fed to the fully connected layer. The softmax layer maps the output of fully connected layer to class labels. In this paper, we use the output of convolutional layers as the representation of input image. Specifically, it is the output of last pooling operation.

In this work, we investigate seven ConvNets: *AlexNet* [12], *VGG-16*, *VGG-19* [13], *GooLeNet* [14], *ResNet-50*, *ResNet-50* and *ResNet-152* [15]. AlexNet and VGG have more fully connected layers but less convolutional layers, while GoogLeNet and ResNet have very deep convolutional layers but only one fully connected layer. The output feature dimension of AlexNet and VGG is 4,096. In GoogLeNet and ResNet, they are 1,024 and 2,048 respectively. As shown in Figure 1, we concatenate the features extracted by the above-mentioned ConvNets into a long vector. Consequently, a 19,456 dimensional vector is obtained.

2.3. Glaucoma Classification

Glaucoma detection is formulated as a binary classification problem and performed by a linear support vector machine (SVM) classifier [17]. To enhance the performance, we also



(a) Holistic feature (b) Local feature (c) Holistic+local feature

Fig. 3: Glaucoma classification results in receiver operating characteristic (ROC) curves. Combining multiple deep features outperforms single feature (Best viewed in color).

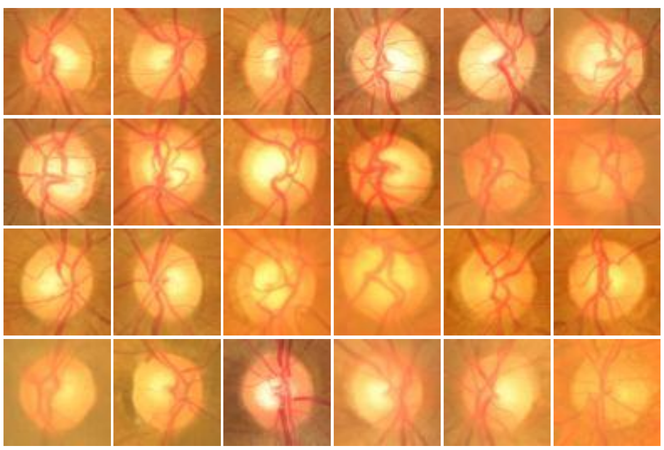


Fig. 4: Example ROIs of Origa [1] dataset.

adopt the "holistic+local" strategy in [11]. Denote the decision values of holistic and local classifiers as $dec^{holistic}$ and dec^{local} respectively, the final decision dec^{fusion} of glaucoma classification is given by

$$dec^{fusion} = dec^{holistic} w + dec^{local} (1 - w), \qquad (1)$$

where $0 \leq w \leq 1$ is a balancing weight.

3. EXPERIMENTS

We conduct experiments on the publicly available Origa [1] dataset. It consists of 650 color fundus images. The image size is uniform 3072×2048. Figure 4 shows some examples of ROI extracted using the method described in Section 2.1. Glaucoma classification experiments are achieved by a 10-fold cross validation. In each fold, 90% images are used for training and the rest 10% are used for test. We use the LibSVM library [17]. The C value is fixed to 10^{-5} in all experiments.

Experimental results of single and combined feature are shown in Figure 3 (a). As can be seen, AlexNet get the best performance among the deep features. Because of relative low feature dimension, GoogLeNet is not as good as other ConvNets. Although it has deeper architecture, inadequate feature dimension limits its representation capability. Similar phenomenon occurs in ResNet. Besides the feature dimension, another possible explanation is AlexNet and VGG have more fully connected layers. To some extent, these layers play similar role of SVM. Such class specific modeling is fused into the convolutional layers in GoogLeNet and ResNet. However, discriminant information of the 1,000 general visual objects is irrelevant to glaucoma. No matter what performance of single feature, they show complementary properties. The combined feature clearly outperforms any of the single feature.

Besides holistic features, we also investigate local features and "holistic+local" feature. Corresponding ROCs are shown in Figure 3 (b) and (c) respectively. Like on holistic feature, the combined feature also get best results on local and "holistic+local" feature, which provides additional evidence for the effectiveness of the proposed feature combination. To find the proper w in Equation (1), we explore the weight by a step of 0.05. Corresponding area under the curves (AUC) for each w are shown in Figure 5. As can be seen, the "holistic+local" strategy is very effective. It consistently gets better AUC than single local or holistic feature in all the features. Figure 5 also clearly show that the combined feature get a higher AUC than single features.

We compare proposed method with the super-pixel based glaucoma screening method by Cheng et al. [4], reconstruction based approach by Xu et al. [5] and the ConvNets based method by Chen et al. [10]. The former two are the leading CDR based approach on Origa dataset, while the latter one is the leading image classification based approach. The best AUCs are shown in Table 1. Xu et al. [5] achieves an AUC of 0.823. Chen et al. [10], which is a single ConvNet

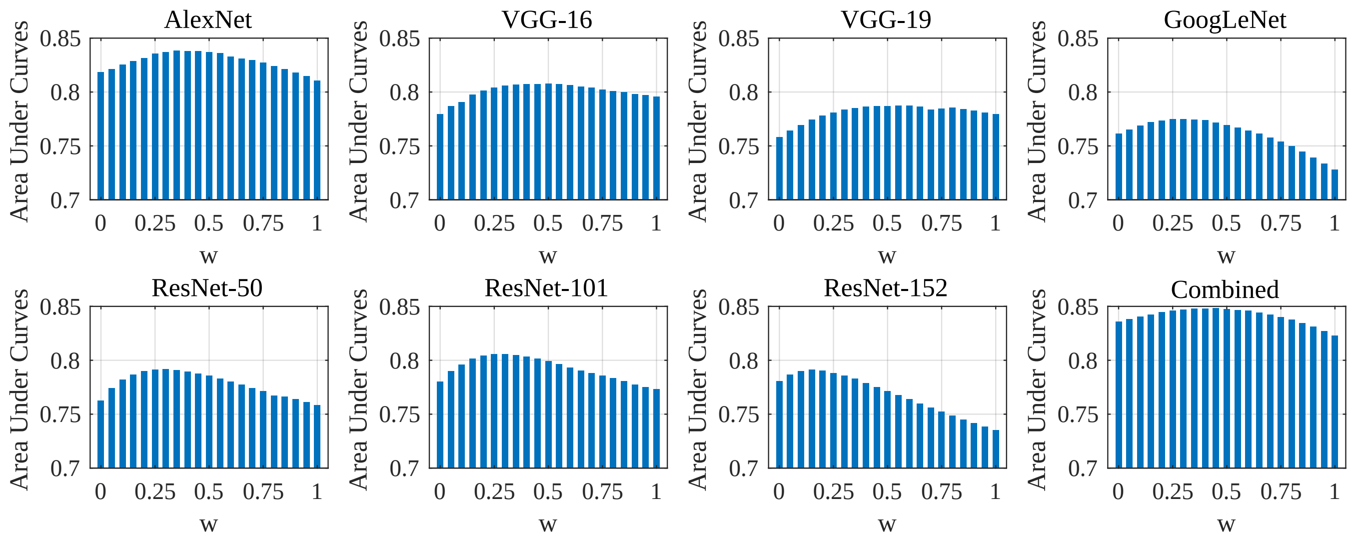


Fig. 5: Area under curves (AUC). Integrating holistic and local features consistently obtains better results on all investigated feature.

Table 1: Best area under curves.

Method	Local	Holistic	Holistic+Local
AlexNet	0.8184	0.8108	0.8384
VGG-16	0.7796	0.7956	0.8078
VGG-19	0.7583	0.7795	0.7875
GoogleNet	0.7613	0.7282	0.7749
ResNet-50	0.7627	0.7584	0.7916
ResNet-101	0.7801	0.7732	0.8058
ResNet-152	0.7807	0.7352	0.7911
Combined	**0.8359**	**0.8229**	**0.8483**
Cheng et al. [4]	0.800		
Xu et al. [5]	0.823		
Chen et al. [10]	0.838		
Manual CDR	0.839		

based approach, get an AUC of 0.838. In our method, the best AUC of single feature is achieved using AlexNet. With "holistic+local" strategy it can get an AUC of 0.8384, which is the best performance reported in [11]. The combined feature with "holistic+local" strategy can get an AUC of 0.8483, which is clearly above all the compared methods. It is notable that, the AUC of manual marked CDR is 0.839. To our knowledge, our method is the first one surpasses manual CDR on Origa dataset. It demonstrates the promising future of classification based approach on glaucoma screening.

4. CONCLUSIONS AND DISCUSSIONS

In this paper, we propose a classification based glaucoma detection method that combining multiple deep convolutional features. The optic disc is located by a deformable shape model. A square region-of-interest is sampled and normalized according to the bounding box of optic disc. The features of ROI are represented by seven popular convolutional neural networks. Glaucoma detection is carried out by applying linear support vector machine classifier on the concatenated deep features. The proposed method achieves an AUC of 0.8483 on the publicly available Origa dataset, which surpasses the performance of manual marked cup-to-disc-ratio.

In our experiments, AlexNet gets the best performance among the investigated convolutional neural networks. However, new network architectures such as ResNet are better in general image classification. The results show the feature extracted by such network may be not that suitable for glaucoma classification, but it also implies that the potential of new network is not fully explored. As discussed in previous section, using the output of last convolutional layer may be not the best choice. But, using the output of early layer will result in very high feature dimension. Besides, it can be expected that combining more feature could further improves the performance. It will increases the feature dimension too. In future work, we will explore applying feature selection to tackle this problem.

5. ACKNOWLEDGMENT

This work was supported by the Foundation for Innovative Research Groups through the National Natural Science Foundation of China under Grant 61421003.

6. REFERENCES

[1] Zhuo Zhang, Feng Shou Yin, Jiang Liu, Wing Kee Wong, Ngan Meng Tan, Beng Hai Lee, Jun Cheng, and Tien Yin Wong, "Origa-light: An Online Retinal Fundus Image Database for Glaucoma Analysis and Research," in *International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2010, pp. 3065–3068.

[2] "Glaucoma," `https://en.wikipedia.org/wiki/Glaucoma/`, [Online; accessed 27-October-2017].

[3] Muhammad Salman Haleem, Liangxiu Han, Jano van Hemert, and Baihua Li, "Automatic Extraction of Retinal Features from Colour Retinal Images for Glaucoma Diagnosis: A Review," *Computerized Medical Imaging and Graphics*, vol. 37, no. 7, pp. 581–596, 2013.

[4] Jun Cheng, Jiang Liu, Yanwu Xu, Fengshou Yin, Damon Wing Kee Wong, Ngan-Meng Tan, Dacheng Tao, Ching-Yu Cheng, Tin Aung, and Tien Yin Wong, "Superpixel Classification Based Optic Disc and Optic Cup Segmentation for Glaucoma Screening," *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 1019–1032, 2013.

[5] Yanwu Xu, Stephen Lin, Damon Wing Kee Wong, Jiang Liu, and Dong Xu, "Efficient Reconstruction-Based Optic Cup Localization for Glaucoma Screening," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2013, vol. 8151, pp. 445–452.

[6] Yuji Hatanaka Akira Aoyama Masakatsu Kakogawa Takeshi Hara Hiroshi Fujita Tetsuya Yamamoto Yoshinori Hayashi, Toshiaki Nakagawa, "Detection of Retinal Nerve Fiber Layer Defects in Retinal Fundus Images using Gabor Filtering," in *Proceeding of SPIE Medical Imaging*, 2007, vol. 6514, pp. 6514 – 6514 – 8.

[7] P. G. Prageeth, D. J, and A. Sukesh Kumar, "Early Detection of Retinal Nerve Fiber Layer Defects using Fundus Image Processing," in *IEEE Recent Advances in Intelligent Computational Systems*, 2011, pp. 930–936.

[8] Ngan Meng Tan, Jiang Liu, Damon Wing Kee Wong, Joo Hwee Lim, Zhuo Zhang, S. Lu, H. Li, Seang Mei Saw, Louis Tong, and Tien Yin Wong, "Automatic Detection of Pathological Myopia using Variational Level Set," in *International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2009, pp. 3609–3612.

[9] Benghai Lee, D. W. K. Wong, Ngan Meng Tan, Zhuo Zhang, Joo Hwee Lim, H. Li, Fengshou Yin, Jiang Liu, Weimin Huang, Seang Mei Saw, Louis Tong, and Tien Yin Wong, "Fusion of Pixel and Texture Features to Detect Pathological Myopia," in *IEEE Conference on Industrial Electronics and Applications*. IEEE, 2010, pp. 2039–2042.

[10] Yanwu Xu, Shuicheng Yan2and Damon Wing Kee Wong, Tien Yin Wong, and Jiang Liu, "Automatic Feature Learning for Glaucoma Detection Based on Deep Learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2013, vol. 8151, pp. 445–452.

[11] Annan Li, Jun Cheng, Damon Wing Kee Wong, and Jiang Liu, "Automatic Detection of Pathological Myopia using Variational Level Set," in *International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2016, pp. 1328–1331.

[12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[13] Karen Simonyan and Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[14] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going Deeper with Convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 770–778.

[16] Annan Li, Zhiheng Niu, Jun Cheng, Fengshou Yin, Damon Wing Kee Wong, Shuicheng Yan, and Jiang Liu, "Learning Supervised Descent Directions for Optic Disc Segmentation," *Neurocomputing*, vol. 275, pp. 350–357, 2018.

[17] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27, 2011.