Automatic Feature Learning for Glaucoma Detection Based on Deep Learning

Xiangyu Chen[1], Yanwu Xu[1], Shuicheng Yan[2], Damon Wing Kee Wong[1], Tien Yin Wong[3], and Jiang Liu[1]

[1] Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore
[2] Department of Electrical and Computer Engineering, National University of Singapore
[3] Department of Ophthalmology, National University of Singapore

Abstract. Glaucoma is a chronic and irreversible eye disease in which the optic nerve is progressively damaged, leading to deterioration in vision and quality of life. In this paper, we present an Automatic feature Learning for glAucoma Detection based on Deep LearnINg (ALADDIN), with deep convolutional neural network (CNN) for feature learning. Different from the traditional convolutional layer that uses linear filters followed by a non-linear activation function to scan the input, the adopted network embeds micro neural networks (multilayer perceptron) with more complex structures to abstract the data within the receptive field. Moreover, a contextualizing deep learning structure is proposed in order to obtain a hierarchical representation of fundus images to discriminate between glaucoma and non-glaucoma pattern, where the network takes the outputs from other CNN as the context information to boost the performance. Extensive experiments are performed on the *ORIGA* and *SCES* datasets. The results show area under curve (AUC) of the receiver operating characteristic curve in glaucoma detection at 0.838 and 0.898 in the two databases, much better than state-of-the-art algorithms. The method could be used for glaucoma diagnosis.

1 Introduction

Glaucoma is a chronic eye disease that leads to vision loss, in which the optic nerve is progressively damaged. It is one of the common causes of blindness, and is predicted to affect around 80 million people by 2020 [8]. Glaucoma is characterized by the progressive degeneration of optic nerve fibres, which leads to structural changes of the optic nerve head, the nerve fibre layer and a simultaneous functional failure of the visual field. As the symptoms only occur when the disease is quite advanced, glaucoma is called the silent thief of sight. Although glaucoma cannot be cured, its progression can be slowed down by treatment. Therefore, timely diagnosis of this disease is important.

Glaucoma diagnosis is typically based on the medical history, intra-ocular pressure and visual field loss tests together with a manual assessment of the Optic

N. Navab et al. (Eds.): MICCAI 2015, Part III, LNCS 9351, pp. 669–677, 2015.
DOI: 10.1007/978-3-319-24574-4_80

Disc (OD) through ophthalmoscopy. OD or optic nerve head is the location where ganglion cell axons exit the eye to form the optic nerve, through which visual information of the photo-receptors is transmitted to the brain. In 2D images, the OD can be divided into two distinct zones; namely, a central bright zone called the optic cup (in short, cup) and a peripheral region called the neuroretinal rim. The loss in optic nerve fibres leads to a change in the structural appearance of the OD, namely, the enlargement of cup region (thinning of neuroretinal rim) called cupping. Since one of the important indicators is the enlargement of the cup with respect to OD, various parameters are considered and estimated to detect the glaucoma, such as the vertical cup to disc ratio (CDR) [9], disc diameter [10], ISNT rule [11], and peripapillary atrophy (PPA) [12]. Since all these measurements focus on the study of OD and most of them only reflect one aspect of the glaucoma disease, effectively capturing the hierarchical deep features of OD to boost the glaucoma detection is our main interest in this paper.

Unlike natural scene images, where typical analysis tasks are related to object detection of regions that has an obvious visual appearance (e.g. texture, shape or color), glaucoma fundus images reveal a complex mixture of visual hidden patterns. These patterns could be only observed by the training and expertise of the examiner. Deep learning (DL) architectures are formed by the composition of multiple linear and non-linear transformations of the data, with the goal of yielding more abstract and ultimately more useful representations [13]. Convolutional neural networks (CNNs) are deep learning architectures, are recently been employed successfully for image segmentation and classification tasks [14,13,15]. DL architectures are an evolution of multilayer neural networks (NN), involving different design and training strategies to make them competitive. These strategies include spatial invariance, hierarchical feature learning and scalability [13][17].

In this paper, we develop a novel deep learning architecture to capture the discriminative features that better characterize the important hidden patterns related to glaucoma. The adopted DL structure consists of convolutional layers which use multilayer perceptrons to convolve the input. This kind of layers could model the local patches better [4]. Unlike conventional CNN, we develop a contextualizing training strategy, which is employed to learn deep hidden features of glaucoma. In the proposed deep CNN, the context takes the responsibility of dynamically adjusting the model learning of CNN, which exploit to effectively boost glaucoma detection by taking the outputs from one CNN as the context of the other one. In addition, to reduce the overfitting problem, we adopt response-normalization layers and overlapping-pooling layers . In order to further boost the performance, dropout and data augmentation strategies are also adopted in the proposed DL architecture.

2 Method

2.1 Feature Learning Based on Deep Convolutional Neural Network

The overview of our proposed automatic feature learning for glaucoma detection is shown in Fig.1, the net of CNN contains 6 layers: five multilayer perceptron

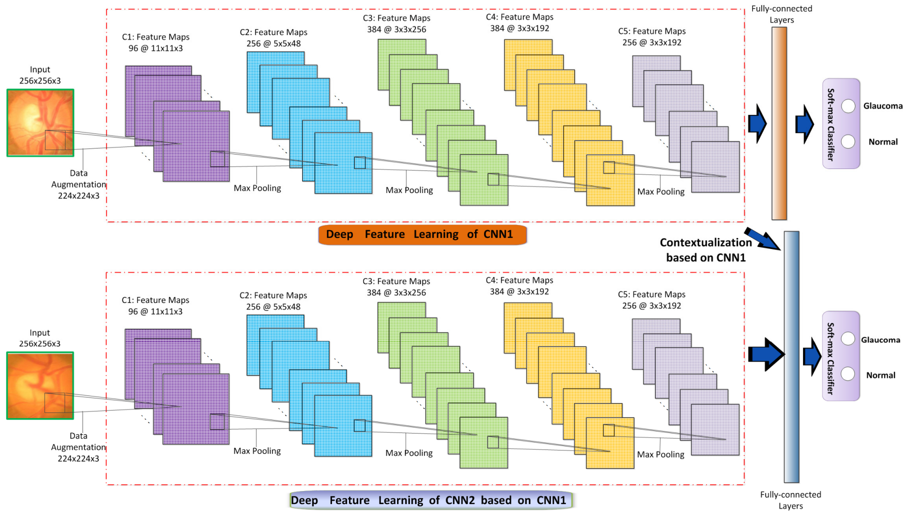

Fig. 1. System overview of our proposed automatic feature learning for glaucoma detection. The net of CNN contains 6 layers: five multilayer perceptron convolutional layers and one fully-connected layer. Data Augmentation is done by extracting random 224 × 224 patches from the input images. 256 × 256 × 3 denotes the dimension of input image. 96@11 × 11 × 3 denote 96 kernels of size 11 × 11 × 3.

convolutional layers [4] and one fully-connected layer. Response-normalization layers and overlapping layers are also employed in our proposed learning architecture as in [14].

Convolutional Layers. Convolutional layers are used to learn small feature detectors based on patches randomly sampled from a large image. A feature in the image at some location can be calculated by convolving the feature detector and the image at that location. We denote $\mathbf{L}^{(n-1)}$ and $\mathbf{L}^{(n)}$ as the input and output for the n-th layer of CNN. Let $\mathbf{L}^{(0)}$ be the 2D input image patch and $\mathbf{L}^{(N)}$ be the output of the last layer N. Let $M_I^{(n)} \times M_I^{(n)}$ and $M_O^{(n)} \times M_O^{(n)}$ be the size of the input and the output map, respectively, for layer n. Denote $P_I^{(n)}$ and $P_O^{(n)}$ as the number of input and output maps respectively for the n-th layer. Since the input to n-th layer is the output of $(n-1)$-th layer, $P_I^{(n)} = P_O^{(n-1)}$ and $M_I^{(n)} = M_O^{(n-1)}$. Let $\mathbf{L}_j^{(n)}$ be the j-th output-feature-map of n-th layer. As the input-feature-maps of the n-th layer are actually the output-feature-maps of the $(n-1)$-th layer, $\mathbf{L}_i^{(n-1)}$ is the i-th input-feature-map of layer n. Finally, we could get the output of a convolutional layer:

$$\mathbf{L}_j^{(n)} = f(\varSigma_i \mathbf{L}_i^{(n-1)} * \mathbf{w}_{ij}^{(n)} + b_j^{(n)} \mathbf{1}_{M_O^{(n)}}), \tag{1}$$

where $\mathbf{w}_{ij}^{(n)}$ is the kernel linking i-th input map to j-th output map, $*$ denotes the convolution, $0 \leq i \leq P_I^{(n-1)}, 0 \leq j \leq P_O^{(n-1)}$, and $b_j^{(n)}$ is the bias element for j-th output-feature-map of n-th layer.

Multilayer Perceptron Convolution Layers. The convolution filter in traditional CNN is a generalized linear model (GLM) for the underlying data patch, and it is observed that the level of abstraction is low with GLM [4]. In this section, we replace the GLM with a more potent nonlinear function approximator, which is able to enhance the abstraction ability of the local model. Given no priors about the distributions of the latent concepts, it is desirable to use a universal function approximator for feature extraction of the local patches, as it is capable of approximating more abstract representations of the latent concepts. Radial basis network and multilayer perceptron are two well known universal function approximators.

There two reasons for employing multilayer perceptron in the proposed DL architecture: First, multilayer perceptron is compatible with the structure of convolutional neural networks, which is trained using back-propagation; Second, multilayer perceptron can be a deep model itself, which is consistent with the spirit of feature re-use. This type of layer is denoted as mlpconv in this paper, in which MLP replaces the GLM to convolve over the input. Here, Rectified linear unit is used as the activation function in the multilayer perceptron. Then, the calculation performed by mlpconv layer is shown as follows:

$$f_{i,j,k_1}^1 = \max(w_{k_1}^1 x_{i,j} + b_{k_1}, 0), \tag{2}$$

$$f_{i,j,k_1}^n = \max(w_{k_n}^n f_{i,j}^{n-1} + b_{k_n}, 0), \tag{3}$$

where n is the number of layers in the multilayer perceptron. (i, j) is the pixel index in the feature map, x_{ij} stands for the input patch centered at location (i, j), and k is used to index the channels of the feature map.

Contextualizing Training Strategy. Different from the traditional CNNs, training neural network independently, we adopt a contextualizing training for our proposed DL architecture, where the whole deep CNN is called Contextualized Convolutional Neural Network (C-CNN). For CNN training, we takes the outputs from one learned CNN as the context input of its own fully-connected layer. The context takes the responsibility of adjusting the model learning of CNN, and thus the contextualized convolutional neural network is achieved. Fig.1 gives an illustration of the algorithmic pipeline. As shown in Fig.1, CNN2 is trained by above mentioned C-CNN strategy, which takes the output of convolutional layers of CNN1 as a contextualized input for its own fully-connected layer. The C-CNN comprises the five multilayer perceptron convolution layers of CNN1 and whole network of CNN2. Then the prediction result of glaucoma is from the soft-max classifier of CNN2.

2.2 Glaucoma Classification

Disc Segmentation. Since optic disc is the main area for glaucoma diagnosis, disc images are the input images of our proposed C-CNN. To segment the optic disc from a retinal fundus image, we employ the method of Template Matching as adopted in [19]. The adopted disc segmentation method is based on peripapillary atrophy elimination, where the elimination is done through edge filtering, constraint elliptical Hough transform and peripapillary atrophy detection. Then each segmented disc image is down-sampled to a fixed resolution of 256 × 256. Finally, the mean value over all the pixels in the disc image is subtracted from each pixel to remove the influence of illumination variation among images.

Dropout and Data Augmentation. To reduce overfitting on image data, we employ data augmentation to artificially enlarge the dataset using label-preserving transformations, and dropout for model combination. Dropout consists of setting to zero the output of each hidden neuron with probability 0.5 [16]. If the neurons in CNN are dropped out, they do not contribute to the forward pass and do not participate in back propagation. During testing, we use all the neurons but multiply their outputs by 0.5. We use dropout in the fully-connected layer in our proposed deep learning architecture.

Data augmentation consists of generating image translations and horizontal reflections [14]. At training time, we perform the data augmentation by extracting random 224 × 224 patches including their horizontal reflections from the 256×256 images, and training our network on these extracted patches. The size of our training dataset will be increased by a factor of 2048. If we do not adopt this scheme, our network will suffer from substantial overfitting. At test time, the CNN makes a prediction by extracting five 224 × 224 patches including the four corner patches and the center patch, as well as their horizontal reflections, and averaging the predictions made by the network soft-max layer on these ten patches. We refer to this test strategy as multi-view test (MVT).

Automatic Classification by Softmax Regression. A softmax regression is a generalization of a logistic regression classifier, which considers as input the condensed feature maps of the pooling layer. For the binary classification setting, the classifier is trained by minimizing the following cost function:

$$J(\Omega) = 1/k[\sum_{i=1}^{k} y_i \log h_{\Omega}(v_i) + (1 + y_i) \log(1 - h_{\Omega}(v_i))], \tag{4}$$

where $(v_1, y_1), ..., (v_k, y_k)$ is the training set containing k images, and $h_{\Omega}(v_i) = 1/(1 + \exp(-\Omega^T v_i))$. For the i-th image, $v_i \in \Re^q$ is the image representation obtained from the output of the pooling layer and $y_i \in \{0, 1\}$ is class label. Ω is a weight vector of $q \times z$ (z is the pool dimension).

3 Experiments

To evaluate the glaucoma diagnosis performance of our proposed C-CNN method, we perform experiments on two glaucoma fundus image datasets ORIGA[5] and

Table 1. The AUCs of different CNN architectures on the ORIGA and SCES datasets. $ORIGA^m$ means that all results in this row are obtained based on various CNN structures without multi-view test strategy on ORIGA. $ORIGA^d$ means that all results in this row are obtained based on various CNN structures without dropout on ORIGA.

Methods	CNN	3 CNN	5 CNN	7 CNN	C-CNN	3 C-CNN	5 C-CNN	7 C-CNN
$ORIGA$	82.4 %	82.4 %	82.6%	82.5%	82.9%	83.0%	**83.8%**	83.7%
$SCES$	86.9 %	87.0%	87.7%	87.6%	88.6%	88.9%	**89.8%**	89.8%
$ORIGA^m$	82.0 %	82.0 %	82.3%	82.1%	82.6%	82.6%	**83.2%**	83.2%
$SCES^m$	86.4 %	86.5%	87.2%	87.1%	88.0%	88.1%	**89.0%**	89.0%
$ORIGA^d$	81.6 %	81.7 %	82.0%	82.0%	82.3%	82.4%	**83.0%**	82.9%
$SCES^d$	86.0 %	86.1%	86.7%	86.6%	87.5%	87.5%	**88.4%**	88.3%

SCES[6]. We compare our algorithm to state-of-the-art reconstruction-based [7], pixel [1], sliding window [2] and superpixel [3][18] based methods. In addition, we compare our system against the current clinical standard for glaucoma detection using intra-ocular pressure (IOP) and to CDR values from expert graders. To validate the effectiveness of our proposed C-CNN architecture, we also perform extensive experiments for glaucoma prediction utilizing different types of CNN architectures and testing strategies.

3.1 Evaluation Criteria

In this work, we utilize the area under the curve (AUC) of receiver operation characteristic curve (ROC) to evaluate the performance of glaucoma diagnosis. The ROC is plotted as a curve which shows the tradeoff between sensitivity TPR (true positive rate) and specificity TNR (true negative rate), defined as

$$TPR = \frac{TP}{TP + FN}, \; TNR = \frac{TN}{TN + FP}, \tag{5}$$

where TP and TN are the number of true positives and true negatives, respectively, and FP and FN are the number of false positives and false negatives, respectively.

3.2 Experimental Setup

We adopt the same settings of the experiments for glaucoma diagnosis in [7] in this work to facilitate comparisons. The ORIGA dataset with clinical glaucoma diagnoses, is comprised of 168 glaucoma and 482 normal fundus images. The SCES dataset contains 1676 fundus images, and 46 images are glaucoma cases.

3.3 Comparison of Different Types of CNN Architectures

We systematically compare our proposed C-CNN with different types of CNN architectures and testing strategies as listed in Table 1. Amongst them,

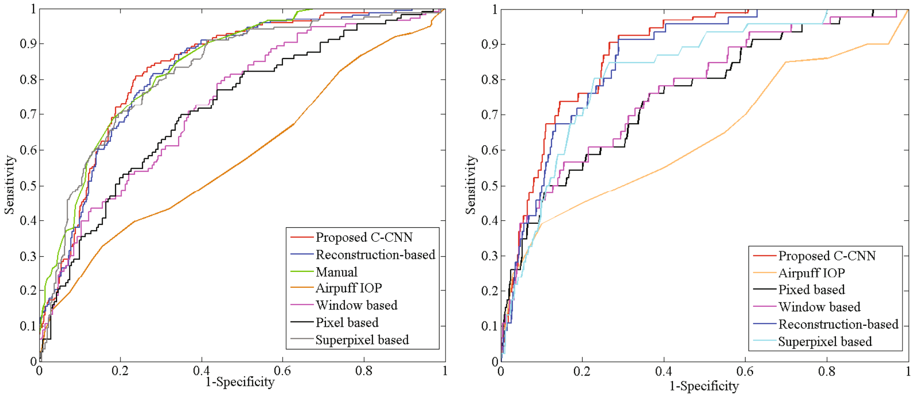

Fig. 2. Glaucoma diagnosis performance on *ORIGA* dataset (left) and *SCES* dataset (right).

- *CNN* means that the final prediction is just from one CNN.
- *3 CNN* means that the final prediction of glaucoma is obtained by averaging three similar CNNs.
- *C-CNN* means the net has 2 CNNs (the five multilayer perceptron convolution layers of CNN1 and whole network of CNN2) as shown in Fig. 1. The prediction result is from CNN2.
- *5 C-CNN* means that we have 5 CNNs and they have the concatenated structure as shown in Fig. 1.
- *7 C-CNN* means that we have 7 CNNs and they have the similar concatenated structure as shown in Fig. 1.

From the quantitative results from Table 1, we are able to observe that: 1) the method of *5 C-CNN* outperforms all the competing methods on *ORIGA* and *SCES* datasets, 2) under six different setting ($ORIGA$, $ORIGA^m$, $ORIGA^d$, $SCES$, $SCES^m$, $SCES^d$), *5 C-CNN* has the best performance. In this work, we use the *5 C-CNN* strategy to validate the glaucoma diagnosis.

3.4 Glaucoma Diagnosis

To validate the effectiveness of our method C-CNN on glaucoma diagnosis accuracy, we compare the predictions of C-CNN (here we use the 5 C-CNN) to state-of-the-art algorithms. In addition, we compare to the current standard of care for glaucoma detection using IOP, as well as CDR grading results from an expert grader. For *ORIGA* dataset, we adopt the same setting of [7]. The training set contains a random selection of 99 images from the whole 650 images, and the remaining 551 images are used for testing. For *SCES* dataset, we use the 650 images from *ORIGA* for training, and the whole 1676 images of *SCES* are the test data.

As shown in Fig. 2, the proposed C-CNN method outperforms previous automatic methods and IOP on *ORIGA* and *SCES* datasets. The AUC values of our method on *ORIGA* and *SCES* are 0.838 and 0.898, respectively. For the state-of-the-art reconstruction-based method, the AUC values are 0.823 and 0.860. On *SCES* dataset, our proposed algorithm has better boost performance than that of *ORIGA*. The reason is that the size of training set on *ORIGA* is only 99.

4 Conclusion

In this paper, we present an automatic feature learning scheme for glaucoma detection based on deep convolutional neural network, which is able to capture the discriminative features that better characterize the hidden patterns related to glaucoma. The adopted DL structure consists of convolutional layers which use multilayer perceptrons to convolve the input. Moreover, we develop a contextualizing training strategy, which is employed to learn deep features of glaucoma. In the proposed deep CNN, the context takes the responsibility of dynamically adjusting the model learning of CNN, which exploit to effectively boost glaucoma detection by taking the outputs from one CNN as the context of the other one. In future work, we plan to extend our study of deep leaning architecture based on C-CNN to multiple ocular diseases detection.

References

1. 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., pp. 5777–5780 (2009)
2. Xu, Y., Xu, D., Lin, S., Liu, J., Cheng, J., Cheung, C.Y., Aung, T., Wong, T.Y.: Sliding Window and Regression based Cup Detection in Digital Fundus Images for Glaucoma Diagnosis. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 1–8. Springer, Heidelberg (2011)
3. Xu, Y., Liu, J., Lin, S., Xu, D., Cheung, C.Y., Aung, T., Wong, T.Y.: Efficient Optic Cup Detection from Intra-image Learning with Retinal Structure Priors. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 58–65. Springer, Heidelberg (2012)
4. Lin, M., Chen, Q., Yan, S.: Network In Network. In: International Conference on Learning Representations 2014 (2014)
5. Zhang, Z., Yin, F., Liu, J., Wong, D.W.K., Tan, N.M., Lee, B.H., Cheng, J., Wong, T.Y.: Origa-Light: An Online Retinal Fundus Image Database for Glaucoma Analysis and Research. In: IEEE Int. Conf. Engin. in Med. and Biol. Soc., pp. 3065–3068 (2010)
6. Sng, C.C., Foo, L.L., Cheng, C.Y., Allen, J.C., He, M., Krishnaswamy, G., Nongpiur, M.E., Friedman, D.S., Wong, T.Y., Aung, T.: Determinants of Anterior Chamber Depth: the Singapore Chinese Eye Study. Opthalmology 119(6), 1143–1150 (2012)
7. Xu, Y., Lin, S., Wong, T.Y., Liu, J., Xu, D.: Efficient Reconstruction-Based Optic Cup Localization for Glaucoma Screening. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part III. LNCS, vol. 8151, pp. 445–452. Springer, Heidelberg (2013)

8. Quigley, H.A., Broman, A.T.: The number of people with glaucoma worldwide in 2010 and 2020. In: Ophthalmol 2006 (2006)
9. Damms, T., Dannheim, F.: Sensitivity and specificity of optic disc parameters in chronic glaucoma. In: Invest. Ophth. Vis. Sci. (1993)
10. Michael, D., Hancox, O.D.: Optic disc size, an important consideration in the glaucoma evaluation. In: Clinical Eye and Vision Care 1999 (1999)
11. Harizman, N., Oliveira, C., Chiang, A., Tello, C., Marmor, M., Ritch, R., Liebmann, J.M.: The isnt rule and differentiation of normal from glaucomatous eyes. In: Arch Ophthalmol 2006 (2006)
12. Jonas, J.B., Fernandez, M.C., Naumann, G.O.: Glaucomatous parapapillary atrophy occurrence and correlations. In: Arch Ophthalmol (1992)
13. Bengio, Y., et al.: Representation learning: A review and new perspectives. In: Arxiv 2012 (2012)
14. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Neural Information Processing Systems, NIPS (2012)
15. Le, Q.V., et al.: Building high-level features using large scale unsupervised learning. In: International Conference on Machine Learning, ICML (2011)
16. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. In: Arxiv 2012 (2012)
17. Tu, Z.: Auto-context and its application to high-level vision tasks. In: Computer Vision and Pattern Recognition, CVPR (2008)
18. Xu, Y., Duan, L., Lin, S., Chen, X., Wong, D.W.K., Wong, T.Y., Liu, J.: Optic Cup Segmentation for Glaucoma Detection Using Low-Rank Superpixel Representation. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part I. LNCS, vol. 8673, pp. 788–795. Springer, Heidelberg (2014)
19. Cheng, J., Liu, J., Wong, D.W.K., Yin, F., Cheung, C.Y., Baskaran, M., Aung, T., Wong, T.Y.: Automatic Optic Disc Segmentation with Peripapillary Atrophy Elimination. In: IEEE Int. Conf. Engin. in Med. and Biol. Soc., pp. 6224–6227 (2011)