Improved Microaneurysm Detection using Deep Neural Networks

Mrinal Haloi
Indian Institute of Technology, Guwahati
Email: h.mrinal@iitg.ernet.in/mrinal.haloi11@gmail.com

***Abstract*—In this work, we propose a novel microaneurysm (MA) detection for early dieabetic ratinopathy screening using color fundus images. Since MA usually the first lesions to appear as a indicator of diabetic retinopathy, accurate detection of MA is necessary for treatment. Each pixel of the image is classified as either MA or non-MA using deep neural network with dropout training procedure using maxout activation function. No preprocessing step or manual feature extraction is required. Substantial improvements over standard MA detection method based on pipeline of preprocessing, feature extraction, classification followed by postprocessing is achieved. The presented method is evaluated in publicly available Retinopathy Online Challenge (ROC) and Diaretdb1v2 database and achieved state-of-the-art accuracy.**

***Keywords:* Diabetic Ratinopathy, deep neural network, microaneurysms.**

I. Introduction

In recent days diabetic retinopathy (DR) is one of the most common severe eye diseases causing blindness in developing and developed countries. Accoding to WHO [1] DR is the primary pathology for 4.8% of the 37 milion blindness cases around the world. Since DR is progressive disease, early stage detection and treatment can save patient from losing sight. For analyzing progress in disease fundus image of patient need be checked regularly. Fast and reliable automatic computer aided diagnosis system will reduce burden on specialists and will give better perfomance for DR mass screening. In most of the DR screening system sensitivity and specificity is used as efficiency measurement.

In general MA appears as the first lesson for diabetic retinopathy. Reliable detection of MA has major importnace for diabetics screening purpose. In color fundus images MA appears as small red dots with very small radius less than that of the major optic vain. In reality these are tiny swollen capillaries in the retina, can discharge blood leading to other pathological symtoms such as exudates, heamorrhages etc. Various challenges such as vessels bifurcations and crossing, illumination and contrast changes, artifacts, degradation of image due to imageing device setup etc. appears in automatic fundus image based DR screening system. MA detection is a well investigated research area for DR mass screening system. Our motivation of these works is to present a new method to detect MA under different challenging situation and to achieve high sensitivity and specificity. Fig. 1 depicts typical retinal image with pathological features such as microaneurysms, exudates etc and non-pathological features such as the optic disc, the macula etc. Complexity of MA detection can be observed from Fig. 1.

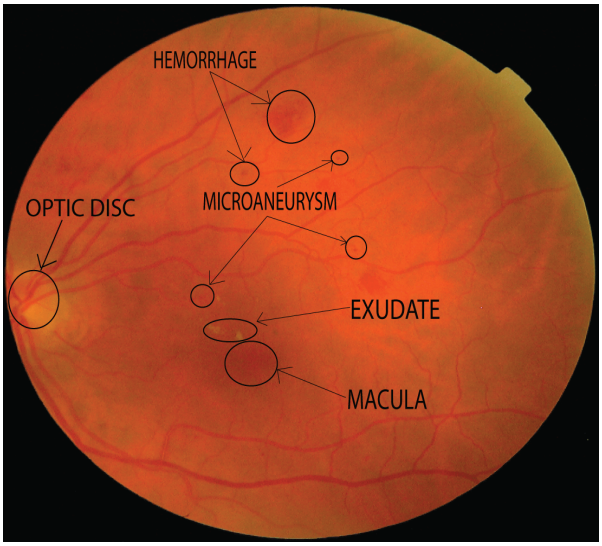

Fig. 1: Typical Pathological Retinal Image

Most of the MA detection works presented till now have common pipeline of three to four stages; first prepocessing the image, secondly manual feature extraction followed by classification and a postprocessing step. Also the use of high contrast green channel of the fundus image is very common in MA detection research. In general exisiting methods used morphological method, filtering based and supervised classification using hand crafted features etc. Antal et al. [2] had developed ensemble-based microaneurysms detection system and claimed first prize in ROC online challege and also achieved good result on other dataset.

They have ensembled several preprocessing and cadidate extraction method to develope their final model. But they didn't addressed the problems of degradation and illuminace changes. Also Quellec et al. [8] Proposed template matching based method for MA detection in wavelet domain and developed optimally adapted wavelet family. Their method prone false detection and true rejection due to heamorrhages and big vessels respectively. Neijmer et al. [10] combined priviously existed method for candidate extraction and used pixel wise classification using manually designed features. In another works by pereira et al. [11] exploited multi-agent

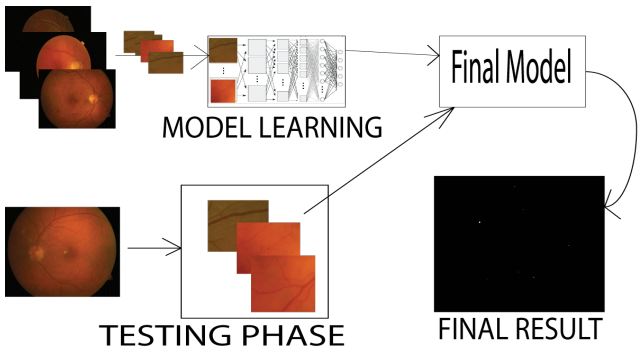

Fig. 2: Method Overview

system for MA segmentation precedded by gaussian and kirsch filter based preprocessing. Final MA candidates evolves from multi agent interation with preprocessed image. Lazer et al. [9] microanurysms was detected using rotating cross section profiles based method which depends on cicularity and diameter of MAs. For each profiles peak was detected and features such as size, height and shape was calculated.

In this work, we propose a deep learning based pixel wise MA classification method invariant to luminane, contrast changes and artifacts. No image based preprocessing or feature extraction stages is required. In addition to that this method perforamnce independent on vessel structures, the optic disc and the fovea. Hence extraction or detection of these features are not required. To increase the accuracy of the method dropout training with maxout activation function is used. Training of this network is time consuming but testing phase is very fast and suitable for real time applications. We have achieved state-of-the-art performance with very low flase positive rate on publicly available datasets.

II. Methods

Microanerysms (MA) usually follow a guassian like intensity distribution and have isolated structures from neighbours. To detect this tiny structures a pixel based deep neural network (DNN) [19] is developed. Pixel based classification is useful for this type of complex detection. Every pixel of a image is classified as MA or non-MA. For any given pixel, class label is predicted using three color channel RGB values in a square window centered on that pixel of size w. The window around the given pixel may contain other MA. An overview of the method has been depicted in Fig. 2.

A. Data Manipulation

Because of local maximum structures of MA; rest of pixels of the window centered on that pixel need to be processed efficiently to get high classification probability. For account of this effect input data was modified by using two techniques for neighbouring data supression, specifically foveation and nonuniform sampling. The concept of foveation orginated from uneven size and organization of photo-receptive cells and ganglions in the human eye. Visual acuity is maximum at the middle of the retina termed as fovea and decreases towards periphery of the retina. Foveation proved to ve very effective in nonlocal means denoising algorithms [20]. In foveation central section of the window is focused, while the peripheral pixels are defoucused using linear space invariant gaussian blur. Standard deivation of gaussian blur kernels increases with distance from central section.

It has been observed that increasing input window size in DNN improves performance significantly, but at the same time computational complexity also increases. Nonuniform sampling was used to selectively depreciate window pixels towards periphery. Only central section of input window is smapled at full resolution, while smapling resolution decreases towards periphery. Using this method large window can be trained with relatively fewer neurons.

B. Network Architecture

DNNs are hierarchical neural networks, inspired by the simple and complex cells in human primary visual cortex. A DNN comprised of convolutional layer alternate wth max-pooling layer [21] followed by fully connected layers and a final classification layer. DNN very definite power of learning discriminative features from raw image patches make it efficient for computer vision tasks, in comparisons to traditional hand crafting features. The newrok used in this work contains five layers including the classification layer; the first three are comprised of convolutional layers each followed by max-pooling. The convolutional layers are followed by one fully connected hidden layers and the softmax classification layer is fully connected with two neurons for MA and non-MA. In this work, we have incorporated dropout[] training algorithm for three convolutional layers and one fully connected hidden layer. And maxout activation function is used for all layers in the network except the softmax layer.

1) Convolutional Layer: The convolutional layer [18] is the core building block of a deep neural network parameterized by the input volume size $M_i \times M_i \times D_i$, the receptive filed or filter size F, the depth of conv layer K and the stride or skipping factor S. If input border is zero padded with size of P_i then number of neurons in the output volume $M_o \times M_o \times D_o$ in is calculated as follows.

$$M_o = \frac{M_i - F + 2P}{S} + 1; D_o = K \tag{1}$$

Stride should be chosen such that M_o is an integer.

2) Max-Pooling Layer: Max-pooling layer ensure fast convergence in comparison to traditional neural networks. In addition to that max -poling provides translation invariance. Input image is partitioned into set of non-overlapping rectangles and maximum value of each subregions is chosen for output. If $W(k,l)$ are subregions then output is obtained as follows.

$$y_{k,l} = \max_{ij \in W(k,l)} x_{i,j} \tag{2}$$

Suppose the input volume of size $M_i \times M_i \times D_i$ for max-pooling layer with spatial extent F and skipping factor S; then output volume of size $M_o \times M_o \times D_o$ is calculated as follows.

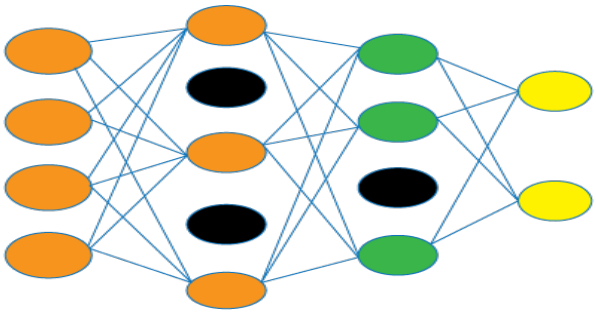

Fig. 3: Dropout training illustration

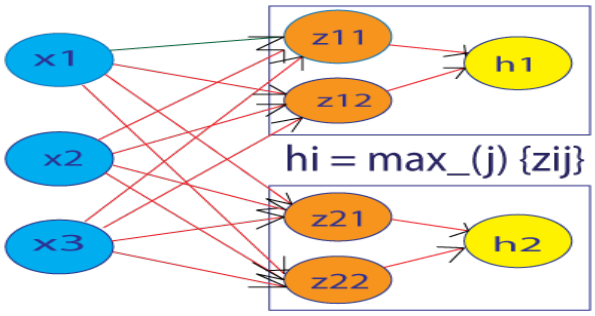

Fig. 4: Maxout activation function illustration

$$M_o = \frac{M_i - F}{S} + 1; D_o = D_i \tag{3}$$

If value of $F > S$, then the process is called overlapping pooling, in general model with overlapping pooling is less prone to overfit.

3) Dropout and maxout: Dropout [17] is one of the most important improvment in machine learning, proved to be sucessfull in many application. It has been observed that combining output of many models improve accuracy significantly, nut in case of deep neural networks training many models more than computationaly costly. Nitish et al. [17] intoduced dropout training for deep neural networks, means to reduce overfitting by randomly omitting the output of each hidden neuron with a probability of 0.5. Training is similar to standard neural network using stochastic gradient descent. The only difference is that dropped out neurons don't take part in forward pass and backpropagation. Suppose a neural network model with L hidden layers and W, b are weights and biases matrix of the network. If $l \in 1,2,...,L$ is hidden layer index; z^l and y^l denote vector of inputs and ouputs respectively at layer l. The following equation described feed forward operation

$$\begin{aligned} z^{l+1} &= W^{l+1} y^l + b^{l+1} \\ y^{l+1} &= f(z^{l+1}) \end{aligned} \tag{4}$$

Using dropout training, the feed forward equation becomes

$$\begin{aligned} (r_i)^l &= Bernoulli(p) \\ \check{y}^l &= r^l * y^l \\ z^{l+1} &= W^{l+1} \check{y}^l + b^{l+1} \\ y^{l+1} &= f(z^{l+1}) \end{aligned} \tag{5}$$

where f any activation fucntion, in our case f is maxout activation function.

A typical situation of dropout training has been explined in Fig. 3, the black circle denotes dropped out node from the network. Dropped out nodes do not participate in training and testing.

The conventional way to represent a neuron's output f as a function of its input x with $f(x) = (1 + e^{-x})^{-1})$ or $f(x) = tanh(x)$. Problems arises with this type of function in gradient descent training, as these function saturate early with positive and negative x values. Gradient descent stuck in this type of fucntion, but lots better improvement can be achieved slightly modifying the activation function as proposed by Goodfellow et al. [16], the maxout network. Maxout is a new kind of activation function for deep neural network with dropout training procedure. In maxout algorithm the input is divided to the activation function into k unit groups and maximum response is recorded. Fig. 4 depicts typical situation of maxout activation function. Given a input $x \in R^d$, a maxout hidden uint h_i implements the following function

$$\begin{aligned} z_{i,j} &= x^T W_{...ij} + b_{i,j} \\ h_i(x) &= \max_{j \in [1,k]} z_{i,j} \end{aligned} \tag{6}$$

where $W \in R^{d \times m \times k}$ and $b \in R^{m \times k}$.

C. Learning nets

From pixel level expert annotated groundtruth each pixel is considered either as MA or non-MA. Training set consists of windows centered on image pixels. If a window lies party outside of the image border, rest of the pixels are derived by horizontal reflections. Windows with a MA pixel at the center is consided as MA smaples and that of with non-MA consided for non-MA samples for training. Moreover to reduce overfitting and to ensure rotational invariance the most common method is to enlarge the dataset using random rotations and using horizontal reflections for border pixels.

Training procedure for dropout neural networks using maxout activation function resemblance with traditional neural networks except a few things. In case of dropout network learning each neuron is dropped with a probability of $Bernouli(p)$, resulting a thinned network. In addition to that forward and backpropagation are done only on this thinned network. Convergence of stochastic gradient descent have got much better improvement in this network using maxout activation function. Also one particular form of regularization specifically constraining the norm of the incoming weight vector at each hidden unit found to be especially useful for dropout training. This is termed as max-norm regularization inspired from the previous use in the context of collaborative filtering [22].

$$\Delta w^t = m^t \Delta w^{t-1} - (1-m^t)\varepsilon^t < \nabla_w L >$$
$$w^t = w^{t-1} + \Delta w^t \qquad (7)$$
$$\|w^t\| \leq c$$

the weight norm constraint only for fully connected layers.

$$\varepsilon^t = \varepsilon_0 f^t$$
$$m^t = \begin{cases} \frac{t}{T} m_i + (1-\frac{t}{T}) m_f & t < T \\ m_f, & t \geq T \end{cases} \qquad (8)$$

where c is a fixed constant, t is the iteration index, ε is the learning rate,m is the momentum variable.

III. Experiments, Results and Discussions

This method have been tested on publicly available ROC [12], Messidor [13] and Diaretdb1v2 [14] dataset. Both of these well annonated with pixel wise labelling, whcih facillitate the design of our pixel wise classification model. ROC contain 50 training image of 768*576 pixels, Messidor consists of 1200 losslessly compressed images with 45 degree filed of view and Diaretdb1v2 includes 89 images of 1500*1152 pixels. The images of messidor were captured using 8 bits per color plane at 1440*960, 2240*1488 or 2304*1536 pixels. Each image is provided with a grading score of R0 to R3. R0 and R1 corresponds to no DR and mild DR respectively; where as R2 and R3 are sever DR and proliferate DR respectively. The grading based on number of MAs and Haemorrhages with presence or absence of neovascularization. No grading scheme avialable for ROC and Diaretdb1v2 datasets. Not all the images of have MA as pathological features. Pixels with other label such as haemorrhages, blood vessels crossings (between two different vessels) and bifurcations (one vessel orginated from another one) and end point of disconnected vessels are considered as non-MA samples. For each pixel input network receive six different windows using data augmentation. For each pixel three windows is obtained using vertical and horizontal mirroring. And then, each window was modified using foveation and nonuniform sampling producing two final windows, this setting emphasize the central pixel and efficient use of bigger window size. Images were taken at different conditions with different cameras with their native resolution and compression settings. The retinal specialist annotations were obtained from a combination of three ophthalmologists with retinal fellowship training. All experiments are conducted on a Ubuntu machine with 12GB RAM, intel i7 3.10GHz processor and NVIDIA GTX 590 graphics card with 1024 CUDA cores. We use Pylearn2 [15] machine learning library built on the top of Theano. Pylearn2 is come with efficient implementation of dropout training with maxout activation fucntion. Total 90000 MA and 1.5 million non-MA windows were used to train the network. While constructing the non-MA windows it have been emphaised to include extensive number of possible false postives. And small number of trivial non-MA windows were included. This setting help the network to learn proper distinctive features.

For accuracy analysis for exudates detection we will compute true positive (TP) a number of exudates pixels correctly detected, false positive (FP) a number of non-exudate pixels which are detected wrongly as exudate pixels, false negative (FN) number of exudate pixels that were not detected and true negative (TN) a number of no exudates pixels which were correctly identified as non-exudate pixels. For better representation of accuracy sensitivity and specificity at pixel level was used as our measurement. Thus the global sensivity SE and the global specificty SP and accuracy AC for each image is defined as follows.

$$SE = \frac{TP}{TP+FN}$$
$$PRED = \frac{TP}{TP+FP}$$
$$SP = \frac{TN}{TN+FP} \qquad (9)$$
$$AC = \frac{TP+TN}{TP+TN+FP+FN}$$

The best network architecture is depicted in Table. I with three convolutional layers each followed by a max-pooling layer and one fully connected layer. And a softmax layer on the top of network with two neurons for MA and non-MA probability values. Probability of droping a uint on each layer is shown in $Bernouli(p)$ column. Each convolutional layer has size 5×5 with a stide of 2 pixels for first layer and 1 pixel for next two layers. Overlapping pooling is used in each of the max-pooling layer with a stride of 2 pixels and pooling regions size 3×3

TABLE I: Network Architecture

Layer	Type	Maps & Neurons	Size	Stride	Bern(p)
0	input	$3 \times 129 \times 129$	...	...	0.1
1	Conv	$64 \times 63 \times 63$	5×5	2	0.2
2	MP	$64 \times 31 \times 31$	3×3	2	..
3	Conv	$64 \times 27 \times 27$	5×5	1	0.2
4	MP	$64 \times 13 \times 13$	3×3	2	..
5	Conv	$64 \times 9 \times 9$	5×5	1	0.5
6	MP	$64 \times 4 \times 4$	3×3	2	..
7	FC	290	1×1	...	0.5
8	FC	2	1×1	...	..

To detect MA in an unseen image, we first apply a mask to get all pixels of interest removing the usual black region apperas during fundus photography. Also a color threshold was defined to left out trivial non-MA pixels to reduce computation time. Window of size 129×129 centered at each image pixel is extracted, for pixels nearby boundaries windows were extracted using horizontal mirroring. Each window have R, G, B color channel. Detector will assign a probability value of being MA and non-MA to each pixel in the image. Finally a probability map of being MA is generated for the testing image.

To remove possible false detection reach connected region of proability map is processed using the concept of convexity

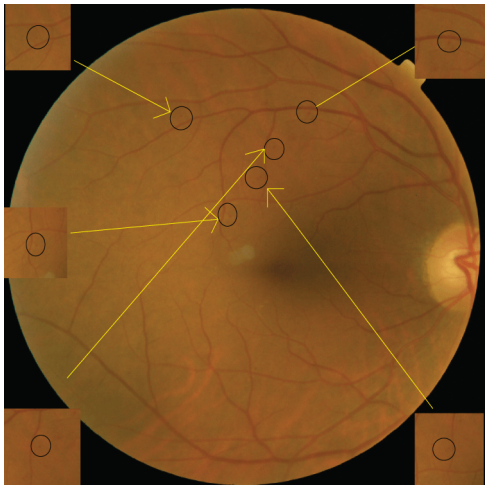

Fig. 5: Detected MA pixels at the center of the windows

and area of the region. Lets consider M is the set of all MA pixels then

- N: Number of connected region in probaility Map.
- For each region update $M = M \cup P_R | Area_{P_R} \leq 21 \cap Convexity_{P_R} \geq 0.8$
- $Area_{P_R}$ is the area of the region P_R and $Convexity_{P_R}$ is the convexity

This will ensure that no vessels crossing, bifurcations and haemorrhages are included in MA detection.

For a typical image pixel based detection output is shown in Fig. 5. We have observed that this method can reliably detect MA candidates.

A comparisons of this method with existing DR screening system is shown in Table II. Eventhough this comparions is not done on same ground, since dataset and the proportion of images having DR symtoms is different. But Sensitivty (Sens), Specificity (Spec) and area under the curve (AUC) value can be accepted for mutual comparions. Our method perform significantly better than the existing methods.

TABLE II: Comparison of automatic DR screening systems.

Method	DR(%)	Sens	Spec	AUC
Proposed Method	46	97%	95%	0.988
Antal et al. [2]	46	90%	91%	0.989
Agurto et al. [3]	76.26	NA	NA	0.89
Abramoff et al. [4]	4.8	84%	64%	0.84
Jelinek et al. [5]	30	85%	90%	NA

This method achieved lower false positve rate than other existing systems. A comparions of sensitivity vs average number of false positive pixels per image is shown in Fig. 6. Variations of sensitivity with 1-specificity is shown in Fig. 7. Also for comparison purpose result of one existing method also ploted in this figure.

Table III shows a comparions of this method on Messidor dataset with recent state-of-the-art system for the scenario R0 vs R1. It can be observed that our method achieve a

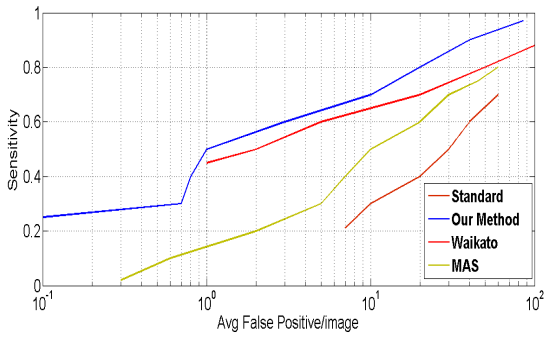

Fig. 6: Comparions of Sensitivity vs Average Number False positive pixels per image

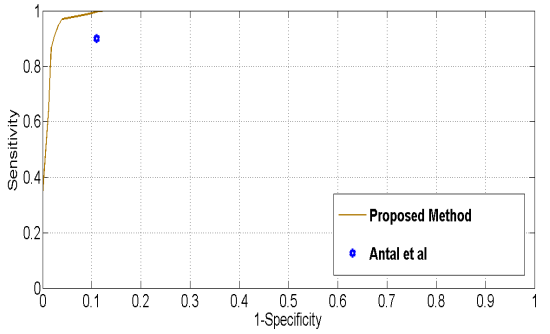

Fig. 7: Comparions of Sensitivity vs 1-Specificity

accuracy of 95% with sensitivity and specificity of 97% and 94% respectively.

TABLE III: Comparions of result on the Messidor Dataset for the scenario R0 vs R1

Method	Sensitivity	Specificity	Acc	AUC
Proposed Method	97%	95%	95.4%	0.982
Antal et al. [2]	94%	90%	90%	0.942

Also for the scenario No DR vs DR Table IV shows result comparions with existing state-of-the-art method on same dataset.

TABLE IV: Comparions of result on the Messidor Dataset for the scenario No DR/DR

Method	Sensitivity	Specificity	Acc	AUC
Proposed Method	97%	96%	96%	0.988
Antal et al. [2]	90%	91%	90%	0.989

A extensive evalution was also carried out on ROC dataset. Due to lack of testing data label we have used a part training data (not used in the training of this method) to evaluate accuracy of our method. Table V depicts comparions of AUC values with other methods on the same dataset.

TABLE V: Comparions of result on the ROC Dataset (Our result only on subset of train Data, Since test data label not available)

Method	AUC
Proposed Method	0.98
Human Expert	0.96
OK Medical [6]	0.89
Fujita Lab [7]	0.88
LaTIM [8]	0.87

IV. Conclusion

In this work, we have presented a deep learning based computer-aided system for microaneurysm detection. The deep network consists of 5 layers including sofmax output layer and dropout training with maxout activation function is used to improve accuracy. In comparions to other existing method, this sytem does not require additional blood vessels extraction step, preprocessing and feature design. This method has been tested in publicly available datasets and achieved state-of-the-art performance for MA candudates extraction with low false positive rate, hence useful for diabetic mass screening purpose.

Acknowledgment

References

[1] Who DR report http://www.who.int/blindness/causes/priority/en/index5.html

[2] Antal, Blint, and Andrs Hajdu. "An ensemble-based system for automatic screening of diabetic retinopathy." Knowledge-Based Systems 60 (2014): 20-27.

[3] C. Agurto, E.S. Barriga, V. Murray, S. Nemeth, R. Crammer, W. Bauman, G. Zamora, M.S. Pattichis, P. Soliz, Automatic detection of diabetic retinopathyand age-related macular degeneration in digital fundus images, Invest. Ophthalmol. Vis. Sci. 52 (8) (2011) 58625871.

[4] M. Abramoff, J. Reinhardt, S. Russell, J. Folk, V. Mahajan, M. Niemeijer, G. Quellec, Automated early detection of diabetic retinopathy, Ophthalmology 117 (6) (2010) 11471154.

[5] H.J. Jelinek, M.J. Cree, D. Worsley, A. Luckie, P. Nixon, An automated microaneurysm detector as a tool for identification of diabetic retinopathy in rural optometric practice, Clinical Exp. Optometry 89 (5) (2006) 299305.

[6] B. Zhang, X. Wu, J. You, Q. Li, and F. Karray, Hierarchical detection of red lesions in retinal images by multiscale correlation filtering, in SPIE Medical Imaging 2009: Computer-Aided Diagnosis, N. Karssemeijer and M. L. Giger, Eds., vol. 7260. SPIE, 2009, p. 72601L.

[7] A. Mizutani, C. Muramatsu, Y. Hatanaka, S. Suemori, T. Hara, and H. Fujita, Automated microaneurysm detection method based on double ring filter in retinal fundus images, in SPIE Medical Imaging 2009: Computer-Aided Diagnosis, N. Karssemeijer and M. L. Giger, Eds., vol. 7260, no. 1. SPIE, 2009, p. 72601N.

[8] G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener, and C. Roux, Optimal wavelet transform for the detection of microaneurysms in retina photographs, IEEE Transactions on Medical Imaging, vol. 27, no. 9, pp. 12301241, 2008.

[9] Lazar, Istvan, and Andras Hajdu. "Microaneurysm detection in retinal images using a rotating cross-section based model." Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on. IEEE, 2011.

[10] Niemeijer, Meindert, et al. "Automatic detection of red lesions in digital color fundus photographs." Medical Imaging, IEEE Transactions on 24.5 (2005): 584-592.

[11] Pereira, Carla, et al. "Using a multi-agent system approach for microaneurysm detection in fundus images." Artificial intelligence in medicine 60.3 (2014): 179-188.

[12] L. Giancardo, F. Meriaudeau, T. Karnowski, Y. Li, K. Tobin, and E. Chaum, Microaneurysm detection with radon transform-based classification on retina images, in Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE. IEEE, 2011, pp. 59395942.

[13] Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology. Available http://messidor.crihan.fr

[14] Kauppi, T., Kalesnykiene, V., Kamarainen, J.-K., Lensu, L., Sorri, I., Raninen A., Voutilainen R., Uusitalo, H., Klviinen, H., Pietil, J., DIARETDB1 diabetic retinopathy database and evaluation protocol, In Proc of the 11th Conf. on Medical Image Understanding and Analysis (Aberystwyth, Wales, 2007)

[15] Ian J. Goodfellow, David Warde-Farley, Pascal Lamblin, Vincent Dumoulin, Mehdi Mirza, Razvan Pascanu, James Bergstra, Frdric Bastien, and Yoshua Bengio. "Pylearn2: a machine learning research library". arXiv preprint arXiv:1308.4214

[16] Goodfellow, Ian J., et al. "Maxout networks." arXiv preprint arXiv:1302.4389 (2013).

[17] Srivastava, Nitish, et al. "Dropout: A simple way to prevent neural networks from overfitting." The Journal of Machine Learning Research 15.1 (2014): 1929-1958.

[18] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.

[19] Dan Claudiu Ciresan, Ueli Meier, and Jurgen Schmidhuber. Multi-column deep neural networks for image classification. In Computer Vision and Pattern Recognition, pages 36423649, 2012.

[20] A. Foi and G. Boracchi, "Anisotropically Foveated Nonlocal Image Denoising", Proc. IEEE Int. Conf. Image Process. (ICIP 2013), pp. 464-468, Melbourne, Australia, Sep. 2013.

[21] Scherer, Dominik, Andreas Mller, and Sven Behnke. "Evaluation of pooling operations in convolutional architectures for object recognition." Artificial Neural NetworksICANN 2010. Springer Berlin Heidelberg, 2010. 92-101.

[22] Nathan Srebro and Adi Shraibman. Rank, trace-norm and max-norm. In Proceedings of the 18th annual conference on Learning Theory, COLT'05, pages 545-560, Berlin, Heidelberg, 2005. Springer-Verlag.