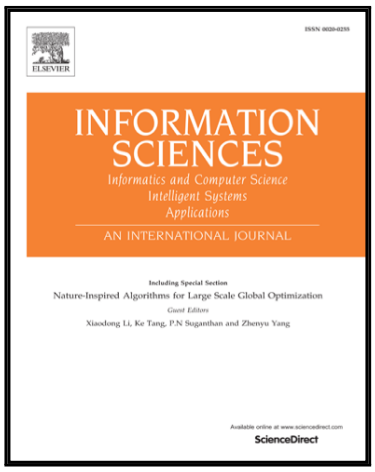


Automated Segmentation of Exudates, Haemorrhages, Microaneurysms using Single Convolutional Neural Network

Jen Hong Tan , Hamido Fujita , Sobha Sivaprasad , Sulatha V. Bhandary , A. Krishna Rao , Kuang Chua Chua , U. Rajendra Acharya

PII: S0020-0255(17)30892-7
DOI: 10.1016/j.ins.2017.08.050
Reference: INS 13054

To appear in: *Information Sciences*

Received date: 30 June 2017
Revised date: 10 August 2017
Accepted date: 13 August 2017

Please cite this article as: Jen Hong Tan , Hamido Fujita , Sobha Sivaprasad , Sulatha V. Bhandary , A. Krishna Rao , Kuang Chua Chua , U. Rajendra Acharya , Automated Segmentation of Exudates, Haemorrhages, Microaneurysms using Single Convolutional Neural Network, *Information Sciences* (2017), doi: 10.1016/j.ins.2017.08.050

Highlights

- Automated segmentation of exudates, haemorrhages and micro-aneurysms
- A 10-layer convolutional neural network is employed.
- Trained and tested on CLEOPATRA database
- 149 images were used for training; another 149 images were used for testing
- Achieved a sensitivity of 0.8758 and 0.7158 for exudates and dark lesions respectively

Automated Segmentation of Exudates, Haemorrhages, Microaneurysms using Single Convolutional Neural Network

Jen Hong Tan[1], Hamido Fujita[2], Sobha Sivaprasad[3] , Sulatha V. Bhandary[4], A. Krishna Rao[4], Kuang Chua Chua[1], U. Rajendra Acharya [1,5,6],

[1]Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore.
[2]Iwate Prefectural University, Faculty of Software and Information Science, Iwate 020-0693, Japan.
[3]NIHR Moorfields Biomedical Research Centre, London, UK
[4]Department of Ophthalmology, Kasturba Medical College, Manipal India 576104
[5]Department of Biomedical Engineering, School of Science and Technology, SIM University, Singapore.
[6]Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, Malaysia

Abstract

Screening for vision threatening diabetic retinopathy by grading digital retinal images reduces the risk of blindness in people with diabetes. Computer-aided diagnosis can aid human graders to cope with this mounting problem. We propose to use a 10-layer convolutional neural network to automatically, simultaneously segment and discriminate exudates, haemorrhages and micro-aneurysms. Input image is normalized before segmentation. The net is trained in two stages to improve performance. On average, our net on **30,275,903** effective points achieved a sensitivity of 0.8758 and 0.7158 for exudates and dark lesions on the CLEOPATRA database. It also achieved a sensitivity of 0.6257 and 0.4606 for haemorrhages and micro-aneurysms. This study shows that it is possible to get a single convolutional neural network to segment these pathological features on a wide range of fundus images with reasonable accuracy.

Keywords – exudates, microaneurysms, haemorrhages, convolutional neural network, fundus image, segmentation, diabetic retinopathy

dx.doi.org/xx.xxxx/xxxxxxxxxxxxxxx

1. Introduction

Diabetic retinopathy (DR) is the leading cause to blindness in adults aged between 20 and 74 years old [4]. It affects more than one third of people with diabetes, estimated at more than 95 million people globally [16,37]. Screening for vision threatening retinopathy by grading colour retinal photographs for disease severity reduces the risk of blindness. Currently, significant manpower resources are required for human graders to grade these images at regular intervals. Alternate solutions are required to cope with screening the increasing numbers of people with diabetes over the coming decades.

Computer-aided diagnosis can aid human graders and play a significant role in the future healthcare system globally. There is already a deep learning based algorithm that, has been trained and tested on a large dataset that grade retinal images with reasonable accuracy [9]. These types of algorithms could aid graders as a preliminary screener to discriminate eyes with and without referable retinopathy so that graders can then concentrate on eyes with referable retinopathy only.

However, the current grade output provided by a computer is not sufficiently useful for human graders because the grade is not lesion specific. An algorithm that marks out the location of these retinopathy lesions will provide more information to graders and save time and resources.

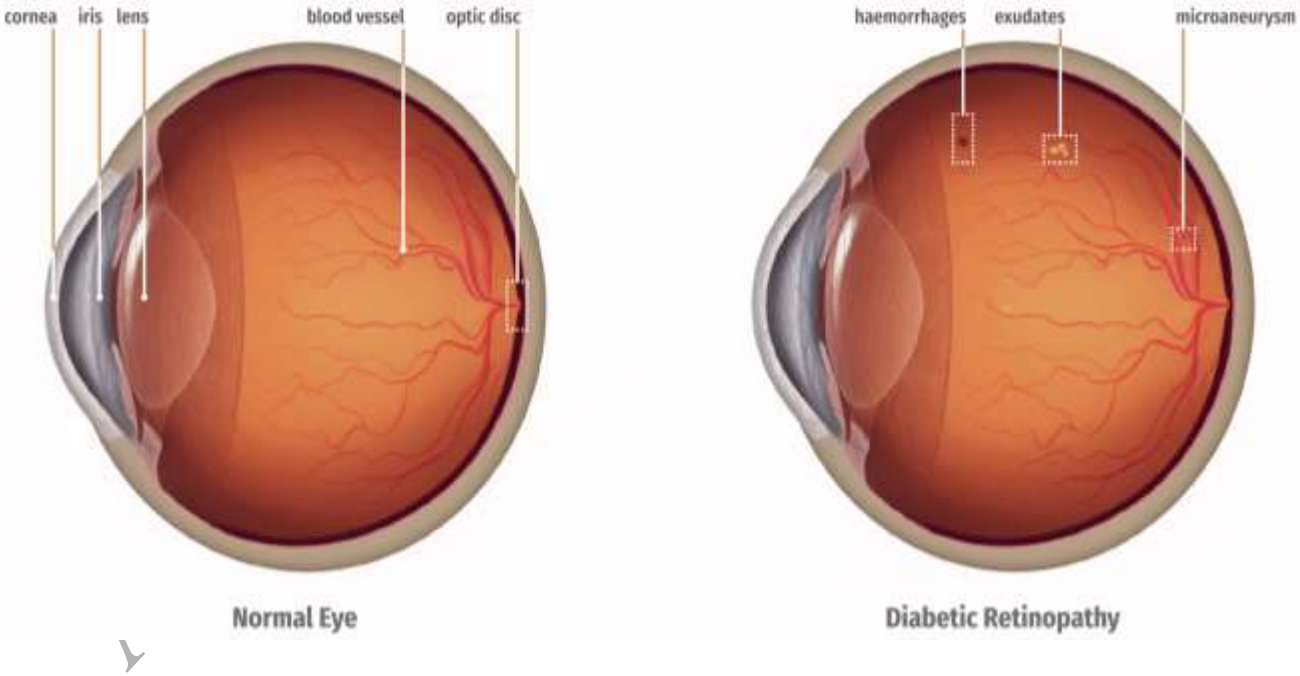


Figure 1. Illustration of normal eye and diabetic retinopathy.

So far there is no single algorithm that could ***automatically*** and ***simultaneously*** segment and discriminate exudates, haemorrhages and micro-aneurysms (see **Figure 1**), the three most common pathological features in early DR. Most of the relevant works in literature solely segment either exudates, micro-aneurysms, or

haemorrhages and micro-aneurysms under a single label (usually denoted as red lesions or dark lesions). When there were solutions proposed to segment both exudates and dark lesions, separate algorithms were adopted to detect each pathological feature.

Automated segmentation of exudates in fundus image are well studied in literature with the first report in 1993, where Philips et al. [22] used a combination of image sharpening, shade correction, local and global thresholding to segment a digitized fundus image. Subsequently, Gardner et al. [8] published a work that sliced a 700 x 700 pixels fundus image into squares of 20 x 20 pixels (non-overlapping), and neural network was applied on each square to report the presence/absence of blood vessels, exudates or haemorrhages in the square.

In the 21st century, Sinthanayothin et al. [26] reported using recursive region growing segmentation and 'Moat Operator' to automatically segment exudates and haemorrhages respectively. They split a fundus image into non-overlapping squares of 10 x 10 pixels (each square was called a segment in the paper), and ophthalmologists marked the presence of exudates and haemorrhages. They evaluated their segmentations on the basis of segment instead of pixel.

Walter et al. [32] then evaluated automated segmentation of exudates pixel by pixel. Morphological processes were used in their work to locate candidates for exudates. They applied threshold to remove false positives; boundaries were delineated with the use of watershed transformation. Several similar solutions have since been proposed. Sopharak et al. [28] used a set of adjusted morphological operation to detect exudates; Welfer et al. [36] suggested a two-stage strategy to perform the task.

Osareh et al. [21] on the other hand, developed a solution that first grouped pixels into segments using fuzzy C-means clustering. From each segment their algorithm extracted 18 features and fed these 18 values into a three layer neural network. Although they located exudates at pixel resolution, their segmentations were evaluated on the basis of segment, not pixel. Fuzzy C-means was used again by Sopharak et al. [27]. Unlike Osareh et al. [21], the evaluation was done on the pixel level.

For the segmentation of haemorrhages or micro-aneurysms in fundus images, Gardner et al. [8] was probably the first group to work on this problem, followed by Sinthanayothin et al. [26]. Larsen et al.[14] reported employed a commercial product to automatically detect red lesions, but the method used was not described. Niemeijer et al. [20], Fleming et al. [6] and Tang et al. [30] employed *k*-Nearest Neighbour to classify red lesions, while support vector machine was used by Kande et al [12]. García et al. [7] assessed the segmentation performance of four classifiers (multilayer perceptron, support vector machine, radial basis function, majority voting), and found that radial basis function gave the best result.

Walter et al. [33] on the other hand, used morphological processes and kernel density estimation to automatically segment micro-aneurysms. Quellec et al. [23]

and Bae et al. [2] adopted template matching to perform the automated segmentation; Antal et al. [1] enlisted an ensemble of micro-aneurysms detectors to achieve the task. Almost all the above studies evaluated their output on the basis of lesion or image. They counted the number of lesions/images (instead of pixel) correctly or incorrectly identified, and then reported the sensitivity and specificity for their works, even though their segmentations were delineated at pixel level.

In this study, we evaluated our segmentations on per pixel basis, not lesion or image, on a set of 149 images collected from 11 clinical sites. We think this is the most honest assessment on the performance of a segmentation algorithm .Our method used a single convolutional neural network to segment and discriminate exudates, haemorrhages and micro-aneurysms. To the best of our knowledge, we believe this is the first report that ***does all the above segmentations not just automatically, but also simultaneously, with the outputs evaluated on the basis of pixel.***

Convolutional neural network is basically an extension of the conventional neural network. It has more layers (compared to neural network), and it uses shared weights as filters to capture local connectivity in image. Before deep learning, machine learning generally needs human to handcraft features for it to work well. In convolutional neural network, however, the net learns by itself to formulate the filters that work best for the task at hand, and thus saves the manual effort to design features or pre-processes.

Convolutional neural network has advanced a lot in the past decade. It has achieved almost perfect accuracy in the discrimination of Arabic numerals [34], and it almost outperformed human observer in the classification of images [10]. Variants of convolutional neural network even defeated the world's elite Go players [25].

When it comes to segmentation in fundus image, so far convolutional neural network has only been used for blood vessels [18,19,29,35], optic disc [29] and fovea [29]. As for pathological feature, it has been only utilized for the detection of haemorrhages [31]. In that study [31], a neighborhood of 41 x 41 around the pixel of interest was extracted and fed into a 10-layer (input is considered as a layer) convolutional neural network. The output consisted only of a single neuron, indicating the presence/absence of haemorrhages. The segmentation was delineated at pixel level, but the evaluation was done on per image basis.

Our algorithm on the other hand, extracts a neighbourhood of 51 x 51 around the pixel of interest, and feeds the input to a 10-layer convolutional neural network. The final layer of our network has four neurons, each for background, exudates, haemorrhages and micro-aneurysms. It delineates the segmentation of all the three pathological features at pixel level, and segmentations output are compared against ground truth pixel by pixel. In the coming sections, we illustrate the details of the materials used in section 2, the method in section 3, and the training and testing procedures in section 4.

dx.doi.org/xx.xxxx/xxxxxxxxxxxxxxx

2. Materials

We have developed and have tested our algorithm on the CLEOPATRA database. The CLEOPATRA study is a multicentre, prospective, two-arm, parallel-group, randomised, clinical trial investigating the role of light mask as a treatment for early diabetic retinopathy and macular oedema. Patients were recruited from 15 National Health Service hospitals in the United Kingdom. This clinical trial was chosen as it represents an enriched dataset of lesions required for our study. Only the baseline retinal images were used for this study, the details of which are tabulated in **Table 1**.

Two ophthalmologists were arranged to mark the ground truth of exudates, haemorrhages and micro-aneurysms. The first ophthalmologists marked all the 298 images, and the second ophthalmologists marked the first 135 images. We used the ground truths provided by the first ophthalmologists as the ultimate truth, and compared the markings of the second ophthalmologists against the ultimate truth to perform inter-grader assessment.

Table 1. Details on the images used in this study.

Clinical site number	Original image size		Resized image size		Number of images	Camera
	Column	Row	Column	Row		
1	3008	2000	872	580	4	Topcon NW6
2	3008	2000	872	580	10	Topcon TRC 50
3	2912	2912	570	570	10	RS-3000 (Nidek)
4	2588	1958	529	400	11	3D OCT-2000 Topcon
5	3872	2592	630	422	35	Topcon
6	3216	2136	635	422	17	
7	2896	1944	626	420	2	
8	2896	1944	644	432	11	Topcon TRC50IX
9	2588	1958	529	400	33	Zeiss FF450
10	2392	2048	584	500	25	Zeiss FF450
11	3216	3216	647	430	23	Topcon 50DX
12	4928	3264	634	420	56	Topcon TRC 50DX
13	4288	2848	640	425	34	Topcon TRC 50DX
14	2144	1424	647	430	14	HRA OCT Spectralis SN Spec-Cam 01479-S3300
	3456	2404	848	590	6	
	2144	1424	647	430	1	
15	2000	1312	655	435	6	Topcon TRC 50IX
Total					298	

3. Methods

3.1 Normalization

Before our algorithm performs any segmentation on the pathological features, the input image must be normalized. We used the method stated in the Section 3.1 from [29] to do the normalization. In this method, a colour fundus image is first converted from RGB colour space to LUV colour space. Procedures to remove varying local contrast and uneven illumination are then applied on the luminance channel (L channel). With the adjusted L channel, together with the original U and V channel, the image is converted back to RGB colour space and the normalization completes.

When an image is captured under low light condition, the image records not only weak luminance, but also little information on hue. When we perform the above normalization on fundus image, areas that receive no light or little light will appear grey in colour. This is because in the normalization the removal of uneven illumination brightens any dark areas (by adjustment on the L channel), but dark areas generally are not just dark, they also contain little or no hue. When these areas are brightened up, they have no hue to show but grey (in colorimetry, grey is a colour that shows level of brightness but no hue and colourfulness [11]).

Since there is no anatomical feature on the retinal layer that is basically grey in colour, we can use this property to assess the quality of an original fundus image based on the normalized image. Let (r, g, b) denote a colour vector where each element in the vector takes a whole number from 0~255. We define a colour as grey if

$$95 \leq g \leq 159 \tag{1}$$

and

$$|r - g| \leq 16$$

$$|g - b| \leq 16 \tag{2}$$

Furthermore, we use the below metric

$$A\ score = \frac{\text{the number of pixels that are not grey}}{\text{total amount of pixels}} \times 100 \tag{3}$$

The higher the *A score*, the more information captured in the image, the better the quality of the image (in terms of the amount of information captured in the image). Several examples are illustrated in **Figure 2**.

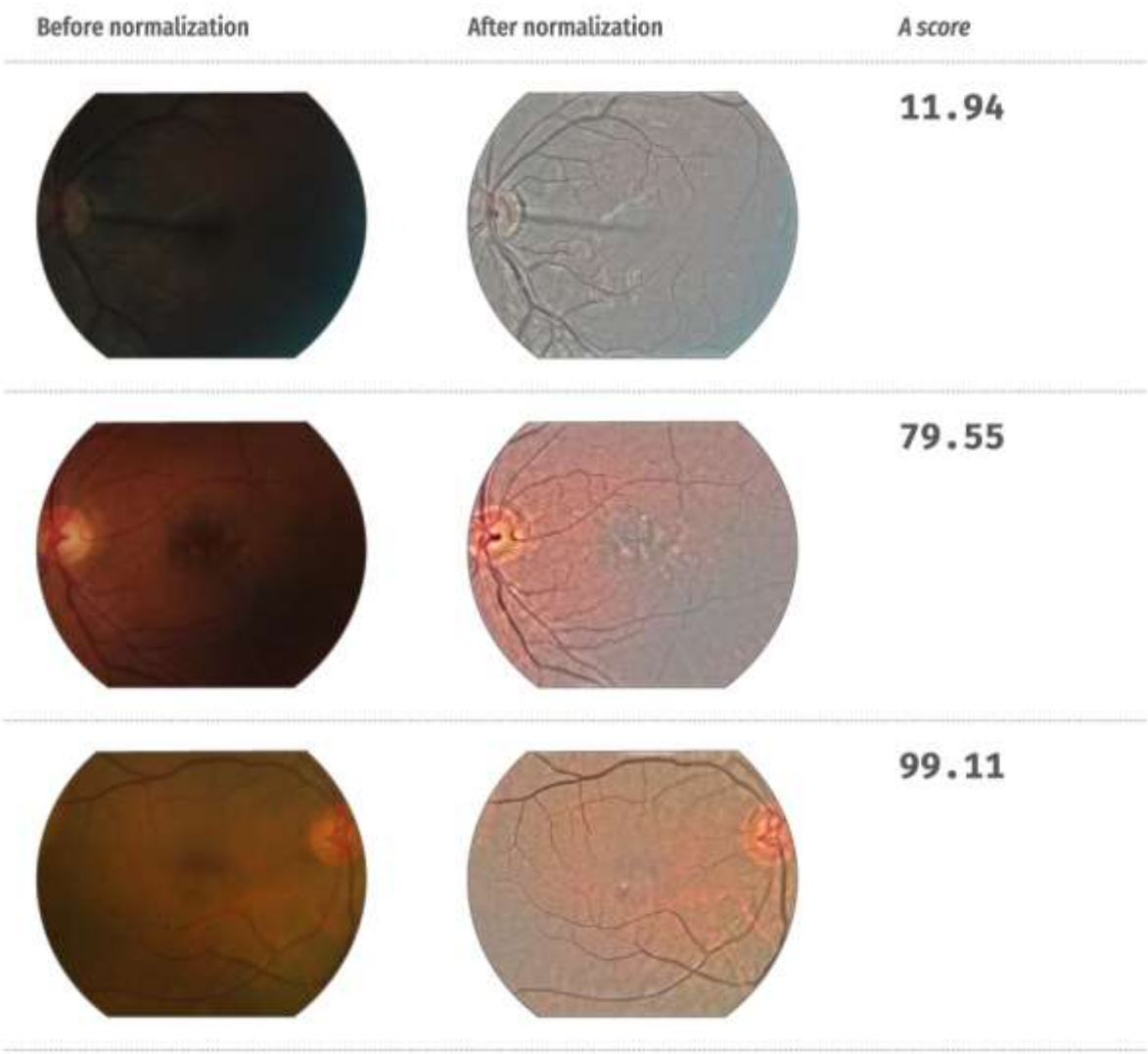


Figure 2. Before and after normalization on three fundus images collected in this study.

dx.doi.org/xx.xxxx/xxxxxxxxxxxxxxx

3.2 The architecture

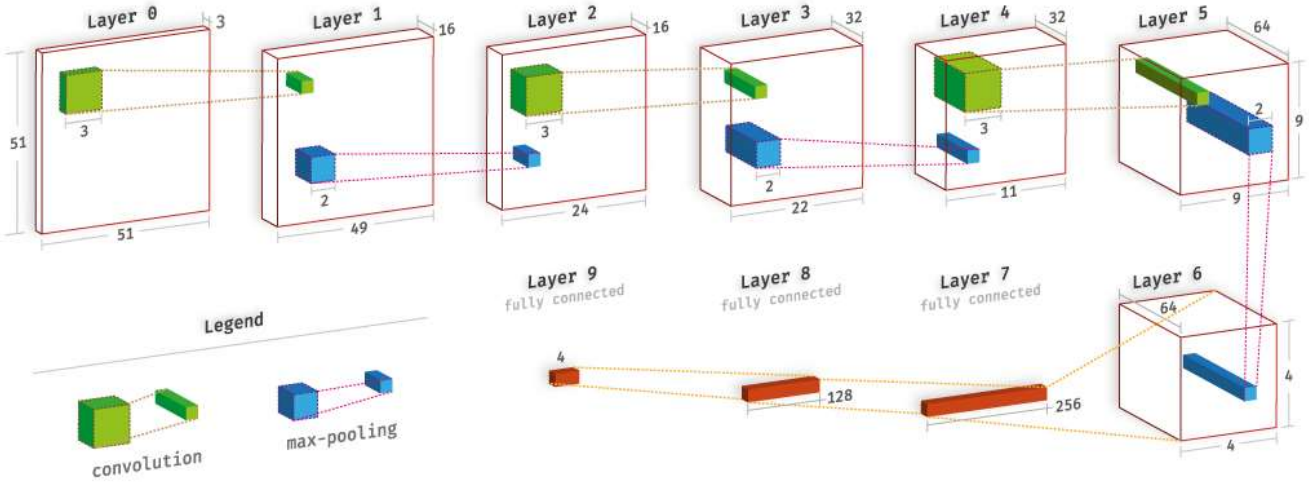


Figure 3. The structure of our convolutional neural network.

Figure 3 illustrates the architecture of our convolutional neural network and the corresponding information is tabulated in **Table 2**. For every pixel of interest (x_i, y_i), our algorithm extracts a neighbourhood of 51 x 51 (with (x_i, y_i) as center) on all three channels (the red, green, blue channel of the fundus image) and forms an input of size 51 x 51 x 3. This input is then convoluted with sixteen 3 x 3 x 3 kernels, followed by a max-pooling sizing 2 x 2. After that, another round of convolution is done between the sixteen feature maps in layer 2 and thirty-two 3 x 3 x 16 kernels, followed by a 2 x 2 max-pooling. Convolution is applied again, this time on the thirty-two feature maps in layer 4 with sixty-four 3 x 3 x 32 kernels. Max-pooling of 2 x 2 is applied one last time to form layer 6.

All the neurons in layer 6 are then connected fully to the 256 neurons in layer 7, which are also fully connected to the 128 neurons in layer 8. The output layer (layer 9), which receives inputs from the 128 neurons in layer 8, has four neurons, representing respectively background, exudate, haemorrhage and micro-aneurysm. The classification is concluded in the last layer by identifying the class of the neuron that produces the largest probability value.

Layer 1, 3, 5, 7 and 8 uses leaky rectifier linear unit as their activation function; softmax function is used for layer 9. The weights and the biases of these layers are initialized using the method laid out by He et al. [10]. For layer 9 the biases are randomly generated on a Gaussian distribution.

Table 2. The architecture of our convolutional neural network.

Layers	Type	No. of feature maps	No. of neurons in the layer	The size of the kernel involves to form each feature map	stride	No. of trainable parameters
0	Input	3	51 x 51 x 3	-	-	-
1	Convolution	16	49 x 49 x 16	3 x 3 x 3	1	448
2	Max-pooling	16	24 x 24 x 16	2 x 2	2	-
3	Convolution	32	22 x 22 x 32	3 x 3 x 16	1	4,640
4	Max-pooling	32	11 x 11 x 32	2 x 2	2	-
5	Convolution	64	9 x 9 x 64	3 x 3 x 32	1	18,496
6	Max-pooling	64	4 x 4 x 64	2 x 2	2	-
7	Fully-connected	-	256	-	-	262,400
8	Fully-connected	-	128	-	-	32,896
9	Fully-connected	-	4	-	-	516
Total						319,396

4. Training and testing procedures

We renamed the images collected from the study in numerical order. Images with odd number were used for training, even number for testing. We defined a point as effective if it falls on the fundus area in a fundus image. There are in total **30,204,027** effective points available from all the 149 images selected for training, of which **29,901,865** are identified as background, and an amount of **146,711**, **90,639** and **64,812** are marked as exudates, haemorrhages and micro-aneurysms respectively (see **Table 3**). As the data set itself is highly skewed (the amount of background point is overwhelming), we performed the training in two stages.

Table 3. The total training samples.

	No. of points	Percentage
Background	29,901,865	99.00%
Exudates	146,711	0.49%
Haemorrhages	90,639	0.30%
Micro-aneurysms	64,812	0.21%
Total	30,204,027	100.00%

In stage one training, on each of the 149 images, we randomly selected 50 background points, 5% of the points marked as exudates, 10% of the points identified as haemorrhages and another 10% of points determined as micro-aneurysms to form the training set (see **Table 4**). This gave in total **30,346** training points. Furthermore, from the testing images we randomly picked 4000 effective points in each category for the use of validation after every epoch of training.

Table 4. The sample selection rules for stage one training.

	Selection rule on each image	Total
Background	50 points	7,450
Exudates	5%	7,336
Haemorrhages	10%	9,073
Micro-aneurysms	10%	6,487
Total		30,346

We ran the training for 60 epochs and used the final net to classify the odd rows and the odd columns of all the 149 images selected for ***training***. We term this step as quarter-full classification. Furthermore, on each of the 149 training images, we randomly selected 1000 background points and another 1000 background points that were misclassified in the quarter-full classification to form the sample set for

stage two training (see **Table 5** and **Table 6**). Each sample was randomly rotated and left-right flipped. We ran the training again for 60 epochs, and the net that gave the best performance was picked to run a complete classification on all the 149 images selected for testing.

Table 5. The sample selection rules for stage 2 learning.

	Original amount	**Selected for training**	
Background	29,901,865	298,000	(On each image, (1000 points) + (1000 points misclassified in quarter-full classification); both randomly selected)
Exudates	146,711	120,000	(randomly selected)
Haemorrhages	90,639	120,000	(90,639+randomly select 29,361)
Micro-aneurysms	64,812	120,000	(64,812+randomly select 55,188)
Total	30,204,027	658,000	

Table 6. The validation set for stage two training. Samples were randomly selected from testing images.

	Total
Background	10,000
Exudates	1,000
Haemorrhages	1,000
Micro-aneurysms	1,000
Total	13,000

Our net used standard backpropagation and stochastic gradient in both stage of training. Regularization was used to prevent overfitting; momentum was not adopted in this study. Batch size was set to 10, and the weights were updated by

$$\mathbf{w}^l \leftarrow \left(1 - \frac{\eta\lambda}{\varphi}\right)\mathbf{w}^l - \frac{\eta}{\kappa}\frac{\partial c}{\partial \mathbf{w}^l} \quad (4)$$

where l denotes the layer number, η the learning rate, φ the total number of training samples, λ the regularization parameter and κ the batch size. c represents the cost function, which in our case is log-likelihood function. Our training updated biases using

$$\mathbf{b}^l \leftarrow \mathbf{b}^l - \frac{\eta}{\kappa}\frac{\partial c}{\partial \mathbf{b}^l} \quad (5)$$

The learning rate and the regularization parameter was set to 0.01 and 0.1 respectively in equations (4) and (5) to facilitate moderate pace of learning.

5. Results

In total there were **30,275,903** effective points available for testing (see **Table 7** for the breakdown). Our net achieved a sensitivity of 0.9572, 0.8758, 0.6257 and 0.4606 for background, exudates, haemorrhages and micro-aneurysms respectively (see **Table 8** for detail and **Table 9** for the corresponding confusion matrix). If pixels of haemorrhages and micro-aneurysms are counted under a single label named dark lesion, our net achieved a sensitivity of 0.7158 for this pathological feature (see **Table 10** and **Table 11**).

Table 7. The total testing samples.

	No. of points	Percentage
Background	30,053,990	99.27%
Exudates	78,970	0.26%
Haemorrhages	78,759	0.26%
Micro-aneurysms	64,184	0.21%
Total	30,275,903	100.00%

Table 8. Sensitivity and specificity in respective classes (four categories).

Class	Sensitivity	Specificity
Background	0.9572	0.7875
Exudates	0.8758	0.9873
Haemorrhages	0.6257	0.9893
Micro-aneurysms	0.4606	0.9799

Table 9. Confusion matrix (four categories).

	Classification				Total
	Background	**Exudates**	**Haemorrhages**	**Micro-aneurysms**	
True background	28,982,239	385,338	316,109	594,759	30,053,990
True exudates	8,777	69,463	535	540	78,970
True haemorrhages	12,299	1,582	49,357	15,641	78,759
True micro-aneurysms	26,332	683	8,023	29,882	64,184
					30,275,903

Table 10. Sensitivity and specificity in respective classes (three categories; haemorrhages and micro-aneurysms combined under 'dark lesions').

Class	Sensitivity	Specificity
Background	0.9572	0.7875
Exudates	0.8758	0.9873
Dark lesions	0.7158	0.9700

Table 11. Confusion matrix (three categories; haemorrhages and micro-aneurysms combined under 'dark lesions').

	Classification			**Total**
	Background	**Exudates**	**Dark lesions**	
True background	28,982,239	385,338	910,868	30,053,990
True exudates	8,777	69,463	1,075	78,970
True dark lesions	38,631	2,220	102,903	142,943
				30,275,903

We have also tabulated the comparison of the performance of our algorithm against previous studies and the inter-human grader assessment done in our study in **Table 11**. Our net had the second highest sensitivity on the segmentation of exudates among studies that evaluated their output on per pixel basis. It also had the second worst sensitivity on the segmentation of haemorrhages and micro-aneurysms among the comparable works listed in the table. However, our study evaluated the segmentation of those dark lesions on the basis of pixel, not lesion. **Figure 4** shows some of the segmentations delineated by our net.

Table 12. Comparison of performance among similar studies. We include only studies that evaluated segmentation on either the basis of pixel, square, segment or lesion and published exact values for sensitivity or specificity. Studies marked with * combined haemorrhages and micro-aneurysms under single label when they were published.

		Exudates		**Haemorrhages**		**Micro-aneurysms**	
	Evaluation basis	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Our method	**Per pixel**	**0.8758**	**0.9873**	**0.6257**	**0.9893**	**0.4606**	**0.9799**
2nd human grader from our study	**Per pixel**	**0.4590**	**0.9962**	**0.3842**	**0.9974**	**0.0873**	**0.9977**
Gardner et al.**[8]**	Per square	0.9410	-	0.8960	-	-	-
*Sinthanayothin et al. **[26]**	Per square	0.8850	0.9970	0.7750	0.8870	0.7750	0.8870
Walter et al. **[32]**	Per pixel	0.9280	-	-	-	-	-
Sopharak et al. **[28]**	Per pixel	0.8000	0.9946	-	-	-	-
Welfer et al. **[36]**	Per pixel	0.7048	0.9884	-	-	-	-
Osareh et al.**[21]**	Per segment	0.9300	0.9410	-	-	-	-

Sopharak et al. **[27]**	Per pixel	0.8728	0.9924	-	-	-	-
*Niemeijer et al. **[20]**	Per lesion	-	-	0.3100	-	0.3100	-
Fleming et al. **[6]**	Per lesion	-	-	-	-	0.6370	-
*García et al. **[7]**	Per lesion	-	-	0.8601	-	0.8601	-
Walter et al. **[33]**	Per lesion	-	-	-	-	0.8847	-
Quellec et al. **[23]**	Per lesion	-	-	-	-	0.8962	-
Bae et al. **[2]**	Per lesion	-	-	0.8500	-	-	-

We wrote all our code (including convolutional neural network) in MATLAB. We used no any external library or toolbox except MATLAB's image processing toolbox. Part of the processes in deep learning was written in C MEX file to speed up calculation. All the algorithms were run on CPU. We trained and tested our net on a workstation that came with two Intel Xeon 2.20 GHz (E5-2650 v4) processor and a 512GB RAM. Generally a complete segmentation on an image could take **3,433** to **4,131** seconds, depending on the size of the image. On the other hand, an epoch of training needed at least **29,491** seconds.

dx.doi.org/xx.xxxx/xxxxxxxxxxxxxxx

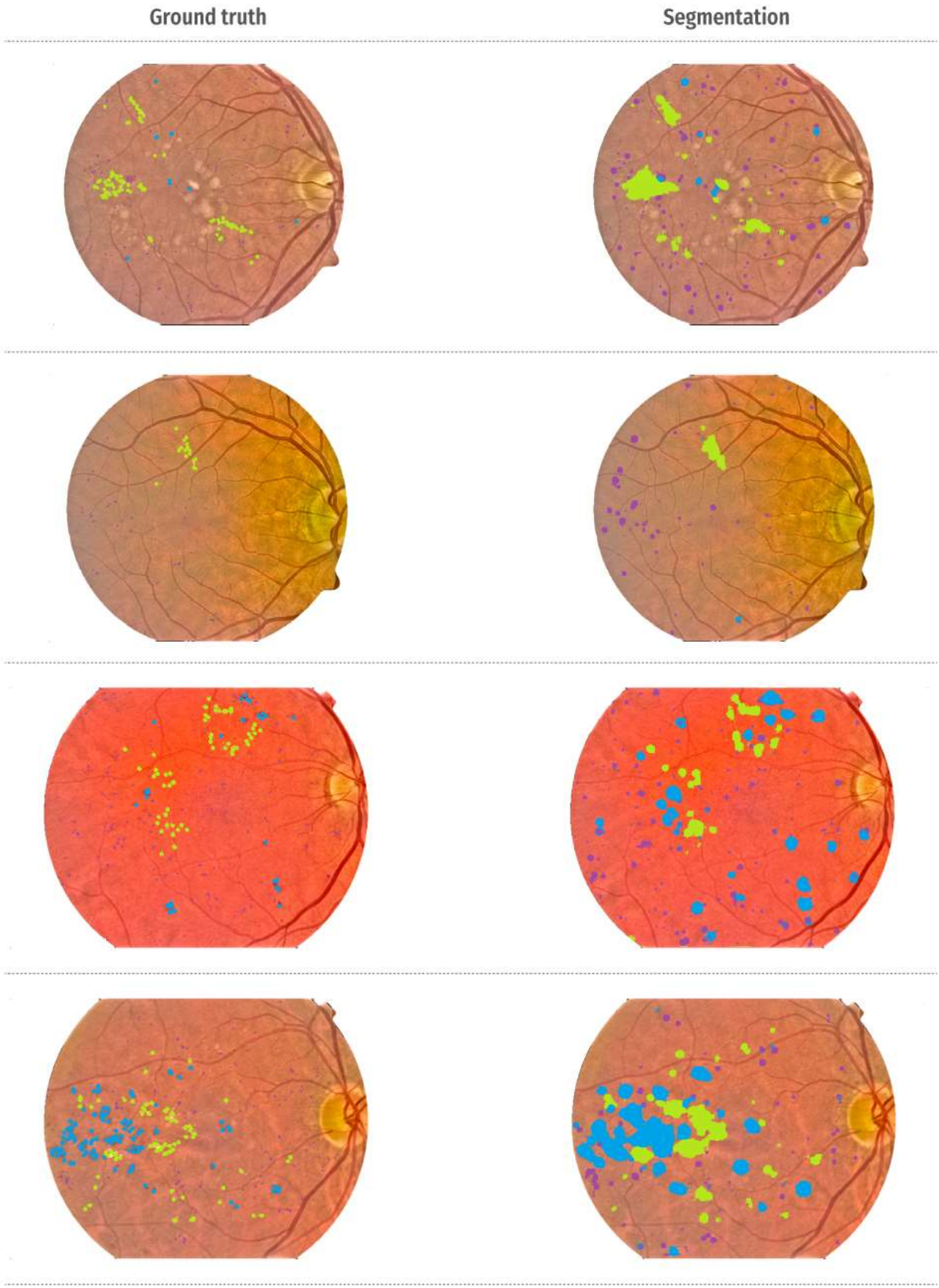


Figure 4. Segmentation outputs. Light green markings denote exudates; light blue markings denote haemorrhages, and purple markings denote micro-aneurysms.

6. Discussion

Segmentation of exudates, haemorrhages and micro-aneurysms in each retinal image is challenging as evidenced by the huge variation between human graders in our study. The second human grader from this study could only achieve a sensitivity of 0.4590, 0.3842, 0.0873 for exudates, haemorrhages and micro-aneurysms when his delineations were compared to first grader. Such poor agreements among graders on delineations of the same pathological features were already previously reported in the ground truths provided in DIARETDB1 [13] database.

This study also faced a further challenge of diversity of our datasets that mirrors real-life situations. The images were collected from 15 clinical sites, using various fundus cameras. The images came in 15 different sizes, their resolutions ranged from 2000 x 1312 to 4928 x 3264. The lowest *A score* in our dataset is 10.67. For MESSIDOR [5] and DIARETDB1, these scores were 94.05 and 66.60, respectively (see **Table 13** for more information).

Table 13. Image quality in various databases.

	The number of images with *A score*			
	>= 99.5	>= 90 but < 99.5	>= 60 but < 90	< 60
CLEOPATRA	228	39	22	9
MESSIDOR	1,178	28	0	0
DIARETDB1	44	27	18	0

The quality of images may partly explain the lower sensitivity and specificity in the segmentation of haemorrhages and micro-aneurysms by our net. ***However, this is the first study to evaluate such diverse quality.*** This is important especially when the capture technology is changing to mobile hand-held cameras in low income countries. Any new computer algorithms developed should be able to perform accurately in such scenarios especially if the images are taken through an undilated pupil.

Although we have arranged 149 images with lesions for training, the number of samples in each pathological feature available for learning was still insufficient. There were only **146,711**, **90,639** and **64,812** points of exudates, haemorrhages and micro-aneurysms in the training set. As a comparison, in our previous study we had **569,415**, **79,321** and **75,221** points of blood vessels, optic disc and fovea available for training, extracted from ***only 20*** fundus images sizing 565 x 584. One limitation of our system is that we have not identified new vessels in this study, because it was an exclusion criteria for the CLEOPATRA study. This will constitute part of our future work on this problem.

There is definitely a need for more expert-delineated ground truths to improve the performance of our segmentation, but clear discrimination between haemorrhages and micro-aneurysms is sometimes impossible. Some micro-aneurysms may appear large enough to be classified as haemorrhages, and some parts of a haemorrhage may have the shape of a micro-aneurysm. One may argue that it is not essential to discriminate these lesions. However, micro-aneurysm turnover has been shown to predict disease progression, and therefore identifying these lesions are important. [3,17,24].

Our net could achieve only a sensitivity of 0.6257 and 0.4606 for haemorrhages and micro-aneurysms, but if these two pathological features are counted as one, the sensitivity improves to 0.7158. This shows how confusing these two pathological features can be to the net.

And our net was not just confused about haemorrhages and micro-aneurysms, it also had trouble delineating the exact boundary of each pathological feature. Our net generally marked a larger pathological area than it should. Micro-aneurysms appeared larger and rounder; several sites of exudates were joined as a single chunk; haemorrhages spread wide and far. The use of max-pooling layers has made the net unable to exactly pin down the boundary of these features. However, according to our observations, without max-pooling layers the accuracy is poorer, and the processing time is longer. We think this issue can be avoided if the net has an autoencoder-like structure, which is definitely a direction worth studying in future work.

Despite these challenges, we believe our method is better in two aspects. First, our solution needs no pre-processing that removes optic disc or blood vessels. The elimination of these anatomical features was essential for almost all the methods proposed to date. In fact, the performance of their algorithm is largely a reflection of the performance of their removal of optic disc or blood vessels.

Secondly, our solution uses no threshold in the process of producing segmentation. The value that serves as a threshold is usually studied carefully after observing its effect on segmentation output. But these observations are generally done on a specific dataset, the size of which is usually small. A threshold value that works on a dataset may not do so for another dataset, and it is not clear if it would work at all. A method that uses no threshold to produce segmentation is more robust, and it also has less hassle when it is applied to a new dataset.

Convolutional neural network takes humongous amount of time to run. In our study, it needs at least 3,000 seconds to complete segmentation. But it can significantly be reduced to about 37.5 seconds if we run the code on GPU (convolutional neural network runs about 20 times faster on a GPU [15]) and use the net to classify only the odd columns and the odd rows of an image. For points that are located on even column and even row, its membership can be easily derived from its neighbourhood without having significant detrimental effect on the segmentation performance.

7. Conclusion

The segmentation of exudates, haemorrhages and micro-aneurysms is a challenging problem. Most of the methods proposed tackled only one or two of the pathological features. We have developed an algorithm that used a 10-layer convolutional neural network to **automatically and simultaneously** segment and discriminate exudates, haemorrhages and micro-aneurysms. The input image was normalized before it was fed into the net. Our method was trained and tested on the CLEOPATRA database. Our net has achieved a sensitivity of 0.9572, 0.8758, 0.6257 and 0.4606 for background, exudates, haemorrhages and micro-aneurysms. In future, it is possible to improve the sensitivities of the segmentation on haemorrhages and micro-aneurysms by using autoencoder-like net structure. The segmentation can also be improved by having more training samples of haemorrhages and micro-aneurysms.

References

[1] B. Antal, A. Hajdu, Improving microaneurysm detection in color fundus images by using context-aware approaches, Computerized Medical Imaging and Graphics. 37 (2013) 403–408.

[2] J.P. Bae, K.G. Kim, H.C. Kang, C.B. Jeong, K.H. Park, J.-M. Hwang, A Study on Hemorrhage Detection Using Hybrid Method in Fundus Images, J Digit Imaging. 24 (2010) 394–404.

[3] M. Bhaskaranand, C. Ramachandra, S. Bhat, J. Cuadros, M.G. Nittala, S. Sadda, et al., Automated Diabetic Retinopathy Screening and Monitoring Using Retinal Fundus Image Analysis, Journal of Diabetes Science and Technology. 10 (2016) 254–261.

[4] N. Cheung, P. Mitchell, T.Y. Wong, Diabetic retinopathy, The Lancet. 376 (2010) 124–136.

[5] E. Decencière, X. Zhang, G. Cazuguel, B. Laÿ, Feedback on a publicly distributed image database: the Messidor database, Image Analysis & …. (2014).

[6] A.D. Fleming, S. Philip, K.A. Goatman, J.A. Olson, P.F. Sharp, Automated microaneurysm detection using local contrast normalization and local vessel detection, IEEE Trans. Med. Imaging. 25 (2006) 1223–1232.

[7] M. García, M.I. López, D. Álvarez, R. Hornero, Assessment of four neural network based classifiers to automatically detect red lesions in retinal images, Medical Engineering & Physics. 32 (2010) 1085–1093.

[8] G.G. Gardner, D. Keating, T.H. Williamson, A.T. Elliott, Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool, British Journal of Ophthalmology. 80 (1996) 940–944.

[9] V. Gulshan, L. Peng, M. Coram, M.C. Stumpe, D. Wu, A. Narayanaswamy, et al., Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs, Jama. 316 (2016) 2402–9.

[10] K. He, X. Zhang, S. Ren, J. Sun, Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, (2015) 1026–1034.

[11] R.W.G. Hunt, The reproduction of colour, John Wiley & Sons, Chichester, 2005.

[12] G.B. Kande, T.S. Savithri, P.V. Subbaiah, Automatic Detection of Microaneurysms and Hemorrhages in Digital Fundus Images, J Digit Imaging. 23 (2009) 430–437.

[13] T. Kauppi, V. Kalesnykiene, J.K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, et al., DIARETDB1 diabetic retinopathy database and evaluation protocol, in: 11th Conference on Medical Image Understanding and Analysis,

Aberystwyth, Wales, 2007.

[14] M. Larsen, J. Godt, N. Larsen, H. Lund-Andersen, A.K. Sjølie, E. Agardh, et al., Automated Detection of Fundus Photographic Red Lesions in Diabetic Retinopathy, Investigative Ophthalmology & Visual Science. 44 (2003) 761–6.

[15] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature. 521 (2015) 436–444.

[16] R. Lee, T.Y. Wong, C. Sabanayagam, Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss, Eye and Vision. (2015) 1–25.

[17] S.F. Leicht, M. Kernt, A. Neubauer, A. Wolf, C.M. Oliveira, M. Ulbig, et al., Microaneurysm Turnover in Diabetic Retinopathy Assessed by Automated RetmarkerDR Image Analysis - Potential Role as Biomarker of Response to Ranibizumab Treatment, Ophthalmologica. 231 (2014) 198–203.

[18] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, T. Wang, A Cross-Modality Learning Approach for Vessel Segmentation in Retinal Images, IEEE Trans. Med. Imaging. 35 (2015) 109–118.

[19] P. Liskowski, K. Krawiec, Segmenting Retinal Blood Vessels With Deep Neural Networks, IEEE Trans. Med. Imaging. 35 (2016) 2369–2380.

[20] M. Niemeijer, B. van Ginneken, J. Staal, M.S.A. Suttorp-Schulten, M.D. Abramoff, Automatic detection of red lesions in digital color fundus photographs, IEEE Trans. Med. Imaging. 24 (2005) 584–592.

[21] A. Osareh, M. Mirmehdi, B. Thomas, R. Markham, Automated identification of diabetic retinal exudates in digital colour images, British Journal of Ophthalmology. 87 (2003) 1220–1223.

[22] R. Phillips, J. Forrester, P. Sharp, Automated detection and quantification of retinal exudates, Graefes Arch Clin Exp Ophthalmol. 231 (1993) 90–94.

[23] G. Quellec, M. Lamard, P.M. Josselin, G. Cazuguel, B. Cochener, C. Roux, Optimal wavelet transform for the detection of microaneurysms in retina photographs, IEEE Trans. Med. Imaging. 27 (2008) 1230–1241.

[24] M.L. Rasmussen, R. Broe, U. Frydkjaer-Olsen, B.S. Olsen, H.B. Mortensen, T. Peto, et al., Microaneurysm count as a predictor of long-term progression in diabetic retinopathy in young patients with type 1 diabetes: the Danish Cohort of Pediatric Diabetes 1987 (DCPD1987), Graefes Arch Clin Exp Ophthalmol. 253 (2014) 199–205.

[25] D. Silver, A. Huang, C.J. Maddison, A. Guez, L. Sifre, G. van den Driessche, et al., Mastering the game of Go with deep neural networks and tree search, Nature. 529 (2016) 484–489.

[26] C. Sinthanayothin, J.F. Boyce, T.H. Williamson, H.L. Cook, E. Mensah, S. Lal, et al., Automated detection of diabetic retinopathy on digital fundus images, Diabet. Med. 19 (2002) 105–112.

[27] A. Sopharak, B. Uyyanonvara, S. Barman, Automatic Exudate Detection from Non-dilated Diabetic Retinopathy Retinal Images Using Fuzzy C-means Clustering, Sensors. 9 (2009) 2148–2161.

[28] A. Sopharak, B. Uyyanonvara, S. Barman, T.H. Williamson, Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods, Computerized Medical Imaging and Graphics. 32 (2008) 720–727.

[29] J.H. Tan, R. Acharya U, S.V. Bhandary, K.C. Chua, S. Sivaprasad, Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network, Journal of Computational Science. 20 (2017) 70–79.

[30] L. Tang, M. Niemeijer, J.M. Reinhardt, M.K. Garvin, M.D. Abràmoff, Splat Feature Classification With Application to Retinal Hemorrhage Detection in Fundus Images, IEEE Trans. Med. Imaging. 32 (2013) 364–375.

[31] M.J.J.P. van Grinsven, B. van Ginneken, C.B. Hoyng, T. Theelen, C.I. Sanchez, Fast Convolutional Neural Network Training Using Selective Data Sampling: Application to Hemorrhage Detection in Color Fundus Images, IEEE Trans. Med. Imaging. 35 (2016) 1273–1284.

[32] T. Walter, J. Klein, P. Massin, A. Erginay, A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina, IEEE Trans. Med. Imaging. 21 (2002) 1236–1243.

[33] T. Walter, P. Massin, A. Erginay, R. Ordonez, C. Jeulin, J.-C. Klein, Automatic detection of microaneurysms in color fundus images, Medical Image Analysis. 11 (2007) 555–566.

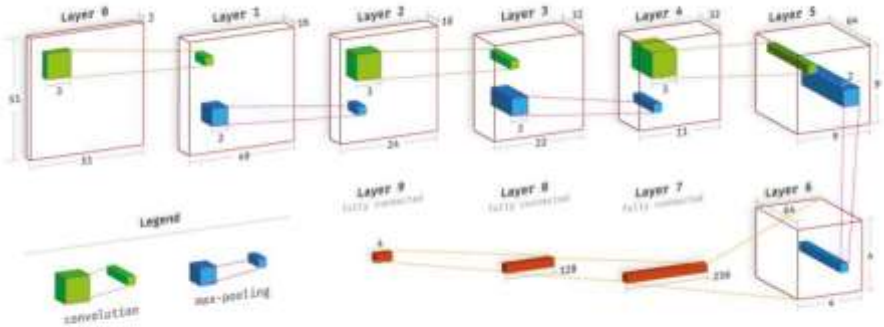
[34] L. Wan, M. Zeiler, S. Zhang, Y.L. Cun, R. Fergus, Regularization of Neural Networks using DropConnect, (2013) 1058–1066.

[35] S. Wang, Y. Yin, G. Cao, B. Wei, Y. Zheng, G. Yang, Hierarchical retinal blood vessel segmentation based on feature and ensemble learning, Neurocomputing. 149 (2014) 1–11.

[36] D. Welfer, J. Scharcanski, D.R. Marinho, A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images, Computerized Medical Imaging and Graphics. 34 (2010) 228–235.

[37] J. Yau, S.L. Rogers, R. Kawasaki, Global prevalence and major risk factors of diabetic retinopathy, Diabetes Care 35 (2012) 556–564.

Graphical Abstract



dx.doi.org/xx.xxxx/xxxxxxxxxxxxxxx