Deep Learning-Based Approach for the Semantic Segmentation of Bright Retinal Damage

Cristiana Silva[1], Adrián Colomer[2(✉)], and Valery Naranjo[2]

[1] Campus Gualtar, University of Minho, 4710 Braga, Portugal
[2] Instituto de Investigación e Innovación en Bioingeniería (I3B), Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain
adcogra@i3b.upv.es

Abstract. Regular screening for the development of diabetic retinopathy is imperative for an early diagnosis and a timely treatment, thus preventing further progression of the disease. The conventional screening techniques based on manual observation by qualified physicians can be very time consuming and prone to error. In this paper, a novel automated screening model based on deep learning for the semantic segmentation of exudates in color fundus images is proposed with the implementation of an end-to-end convolutional neural network built upon U-Net architecture. This encoder-decoder network is characterized by the combination of a contracting path and a symmetrical expansive path to obtain precise localization with the use of context information. The proposed method was validated on E-OPHTHA and DIARETDB1 public databases achieving promising results compared to current state-of-the-art methods.

Keywords: Semantic segmentation · Deep learning · Fundus images Exudates · U-Net

1 Introduction

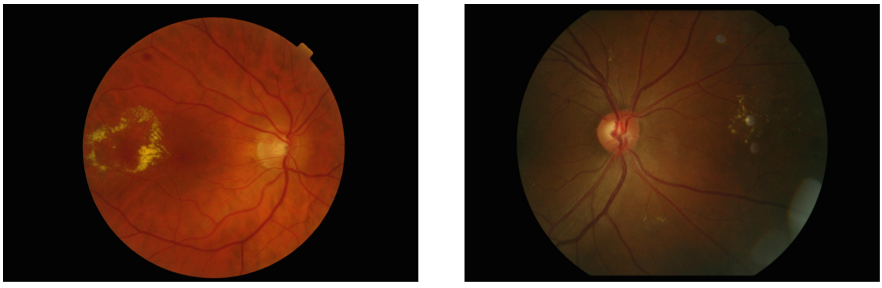
According to the World Health Organization (WHO), diabetic retinopathy (DR), a complication of diabetes manifested in the retina, is a major cause of blindness within the working age population in the developed world. It occurs as a result of accumulated damage to the retinal small blood vessels [1]. Due to a common absence of symptoms in early stages, this disease can go unnoticed until the changes in the retina have progressed to a level where treatment is nearly impossible or irreversible vision loss has occurred.

The risk of blindness in diabetic patients could be significantly reduced through regular screening by suitably trained observers for the development of DR, since an early detection and timely treatment can halt or reverse the

H. Yin et al. (Eds.): IDEAL 2018, LNCS 11314, pp. 164–173, 2018.
https://doi.org/10.1007/978-3-030-03493-1_18

progression of the disease [2]. However, the number of qualified physicians available for direct examinations of the population at risk is limited in most countries. Moreover, the conventional retina examination techniques based on manual observation can be highly subjective, very time consuming and prone to error. These facts highlight the need for automated DR diagnosis techniques based on color fundus retinal photography with high accuracy and quick convergence rate for them to be suitable for real-time applications.

One of the primary signs of DR is the development of retinal exudates (Fig. 1(a)) which consist of lipid and protein accumulations in the retina of various shapes, locations, and sizes, according to the stage of the disease. Its accurate detection can be seriously affected by several factors related to the acquisition process with fundus cameras as well as retina's anatomy. The presence of other bright elements (drusen, optic disk, and optic nerve fibers), dust spots, random brightness, noise presence, uneven illumination, low contrast, and color variation represent a challenge for the task at hand, as it can be seen in Fig. 1(b), making this an extensively studied topic.



(a) (b)

Fig. 1. Fundus images. (a) Image with the presence of exudates, (b) Noisy image with artefacts and uneven illumination.

Most of the classical approaches developed so far involve an image preprocessing stage, followed by a candidate extraction step where structures with similar characteristics as the lesion are selected. Finally, several features are extracted for each lesion candidate and a classification algorithm (usually, a machine learning classifier [3]) is applied to eliminate false positives. The main implemented techniques for extracting lesion candidates can be categorized into dynamic thresholding [4], mathematical morphology [5–7], and clustering [8]. A hand-crafted feature extraction requires domain expertise and effort for it to be optimized to specific problems. This process is deemed unobjective since the researcher has to manually decide on the features to be used in a classifier with knowledge obtained through specialized clinicians. This motivates the development of a novel framework capable of automatically learn the most relevant features and accurately segment exudates.

Recently, convolutional neural networks (CNNs), a branch of deep learning, have emerged as a powerful tool for making automatic image recognition tasks more successful. Its great performance in biomedical applications [9,10] can be explained through its capability of hierarchically extract features from raw image pixel intensities by learning and formulating the appropriate filters for the task at hand. The first attempts in CNN-based approaches for DR evaluation emerged in a Kaggle[1] competition [11,12] where images were classified by the severity of the disease. To the best of the author's knowledge, only a single work has been developed towards the segmentation of exudates by using deep neural networks. In [13] a patch-based CNN architecture is proposed with the aim of providing a pixel-wise classification by returning the probability of each pixel belonging to one of two classes: exudate or non-exudate. Two fully-connected layers are responsible for the binary classification of image pixels. The resulting map is then combined with the output of optic disc and vessel detection procedures. Despite the fact that the network's input includes optic disk pixels, potentially affecting its results, this architecture is not best suited for a pixel-level classification.

The main contribution of this paper is a novel deep learning-based approach for the automatic semantic segmentation of exudates in color fundus images. An encoder-decoder CNN built on top of the U-Net architecture [14] is implemented for this application. The model uses labelled pixels to learn the connection between local features and the associated specific classes, and then classify each pixel based on which class presents the highest probability for that pixel. Compared to most exudate segmentation methods, the algorithm undertaken in this work is more robust due to the fact that it is end-to-end, in other words, it is almost entirely trainable and free of hand designed and fixed modules. To the best of the author's knowledge, this is the first attempt of adapting an encoder-decoder CNN architecture to retinal images' semantic segmentation.

2 Methods

CNNs have been successfully applied to semantic segmentation [15], specifically fully convolutional networks (FCNs) which follow the encoder-decoder architecture. This image-to-image structure emerged as a solution for pixel-wise predictions since it outputs high resolution segmentation maps with localization as well as semantics information, performing very well in biomedical image segmentations [16,17]. Following these nets, a novel neural network structure was introduced in [14], the so-called "U-Net" for its U-shaped architecture. It differs from FCNs in its extended decoding branch by taking into account useful global context information in higher resolution layers, being able to work with small sets of training images and still provide more precise segmentations. It has proven its effectiveness in biomedical image segmentations [14,18,19] as it outperforms existing methods on biomedical challenges. In this paper, a CNN is built on top of U-Net for the semantic segmentation of retinal images, as shown in Fig. 2.

[1] https://www.kaggle.com/c/diabetic-retinopathy-detection.

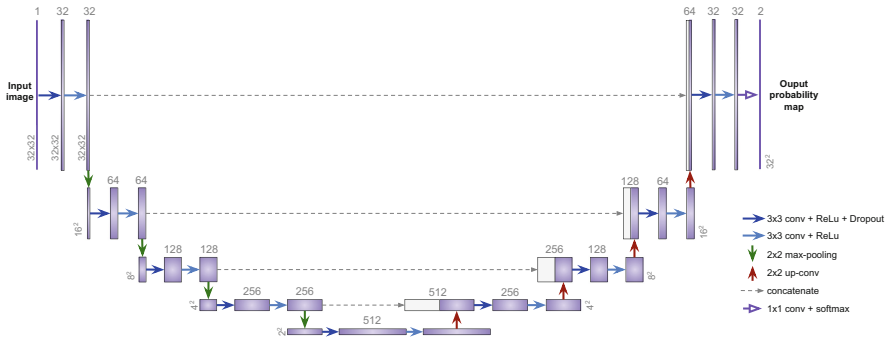


Fig. 2. Architecture of the proposed network built upon U-Net.

2.1 Image Pre-processing

Due to fundus images' heterogeneity, they aren't adequate to be used directly as input to the network. First of all, images belonging to the same dataset often present different resolutions. A standardized image resolution is required by the developed framework. Thus, a spatial normalization is performed using a reference image as a size invariant to resize all images in the dataset.

To reduce memory consumption and training time, a conversion from RGB to grayscale images is performed. This is done by extracting the green channel which is commonly used to segment the lesions [20]. While the red channel is often saturated and with low contrast and the blue channel usually very noisy and with poor dynamic range, the green channel shows the maximum contrast between lesions and background.

Fundus images commonly suffer from non-uniform illumination and poor contrast caused by different lightning conditions in the acquisition rooms as well as retina's anatomical variability. To solve this problem, an image contrast enhancement is carried out by performing a contrast limited adaptive histogram equalization (CLAHE). This window-based technique provides a uniform distribution of grey values across an established 8×8 pixels-sized window, improving local contrast and, thus, raising the visibility of some hidden features.

To prevent the network from learning retinal images' inherent background noise, a 5×5 median filter is applied. This filter smooths image data by performing a spatial filtering on each pixel using the grey level values present in a square window surrounding that pixel. Finally, the intensity values of the images are scaled to [0,1].

2.2 Network Architecture

The core element of this method is the convolutional neural network built upon U-Net architecture [14]. Unlike the usual CNN architectures with only contracting layers for image classification, U-Net is an image-to-image framework as it takes an image as input and returns a probability map as output. This is possible

thanks to the addition of an expansive path (decoder) symmetrical to the typical CNN's contracting path (encoder) to obtain pixel-wise labeling. Precise localization with the use of context is achieved in this model through the combination of high resolution feature maps from the contracting path with upsampled outputs from the expansive path.

In similarity to the original U-Net, each contracting block of the architecture implemented in this work (see Fig. 2) is composed by two 3×3 convolution, each followed by a rectified linear unit (ReLu) activation function ($f(x) = \max(0,x)$). Afterwards, 2×2 max pooling is applied, reducing image resolution by 2. At each block, the number of filters is doubled. On the other hand, in the expansive path, the opposite happens. The same pair of convolutional layers are applied in each block, preceded by an "up-convolution", that is, an up-sampling of the feature map (increasing image resolution by 2) followed by a 2×2 convolution. Then, the resulting feature map is concatenated with the corresponding feature map from the contracting path, size-wise. Finally, a 1×1 convolution and a pixel-wise softmax activation function are applied to obtain the desired number of classes and final probabilities for each pixel.

The modifications carried out in this implementation involve the addition of a dropout layer between two consecutive convolutional layers to avoid overfitting and the reduction of filters for all convolutional layers along the network to simplify the architecture and reduce training time while maintaining the same level of performance. Moreover, all convolutions are implemented with zero-padding to preserve the spatial size of the input image. Therefore, it becomes unnecessary to crop the feature maps from the contracting path for them to be concatenated with the feature maps from the expansive path, as established in [14], since they already present the same resolution.

2.3 Training and Testing

Like any other deep learning approach, this work involves training the network first, and then test the resulting model on new images. In most cases, lesions compose less than one percent of the total number of pixels in a retinal image. For this reason, the network computes the probability of a pixel being an exudate using local features in a square window centered on the pixel itself. Moreover, this patch-based approach is carried out to substantially increase the amount of training data, improving model's performance.

Each time the network is trained, input images are subjected to the aforementioned preprocessing techniques and split into patches using a square sliding window with overlap. While the sliding window goes through the full images, patches partially or completely outside the FOV or containing optic disk pixels, previously detected by means of [21], are excluded. Still, the resulting patches present unbalanced classes, that is, the number of patches classified as healthy is substantially higher than the ones classified as pathological in a retinal image, which can overwhelm the net and result in overfitting to the majority class. Given N pathological patches and M healthy patches where $M >> N$, a random selection of N healthy patches is applied to balance the classes.

At testing phase, to improve performance and obtain smoother predictions, consecutive overlapping patches with a stride of 8 pixels are used to obtain the lesion probability of each pixel by averaging probabilities over all predicted patches covering that pixel. Once again, patches partially or completely outside the FOV or containing optic disk pixels are excluded. Hence, once the predictions are generated, the resulting overlapped patches are recomposed considering the missing patches and the overlapping technique, and the final probability maps are obtained as images in the original resolution. These images can then be used for further performance evaluation.

3 Experiments

The proposed model was trained and validated on E-OPHTHA [22] public database which contains two subsets, depending on the lesion type. The subset selected to be used in this implementation contains forty-seven retinal images with the presence of exudates. These lesions are manually annotated by ophthalmologists at a pixel level. The dataset presents four different image resolutions, ranging from 1440×960 pixels to 2544×1696 pixels. After the first pre-processing technique where a spatial normalization is performed, the images were scaled down to a final resolution of 1440×960 pixels.

3.1 Implementation Details

The framework was developed using Python 3.5 and OpenCV 3.0, from the pre-processing techniques to the attainment of output probability maps. The resizing of the images as well as the model evaluation were performed using Matlab®R2016a. An Intel Core i7-7700K@4.20 GHz processor with 32 GB of RAM and Ubuntu 16.04 LTS as operating system was used throughout the all process. A NVIDIA GeForce®TITAN Xp with 12 GB of GDDR5X RAM was the GPU used. The CNN model was designed recurring to Keras framework with Theano as backend.

3.2 Training Parameters

The framework's design allows a fast and easy adjustment of parameters and datasets to be used for training and testing. Several tests were performed in which tunable parameters were adjusted according to its impact in the model's performance. In order to obtain predictions for all the images in the dataset and, thus, provide robustness to the proposed method, cross-validation was applied. For this purpose, the forty-seven images were randomly partitioned into $k = 5$ folds. In each fold iteration, out of the k partitions, a single partition is retained to test the model, while the remaining $k-1$ partitions are used as training data.

The retinal input images and their corresponding labelled segmentations were used to train the network with the employment of stochastic gradient descent for optimization and a cross-entropy loss function. The network was trained with

a momentum of 0.9, a weight decay of $1e^{-6}$ and a fixed learning rate of $1e^{-3}$. It was set to continue its training for 2000 epochs with a batch size of 32. The input of the net were 32×32 pixels patches which were extracted with a stride of 16 pixels for both width and height.

3.3 Results

In order to obtain binary segmentation maps from the probability maps, the optimal threshold for each fold was determined by computing the best trade-off between sensitivity and specificity. At this stage, five performance evaluation metrics - accuracy (Acc), sensitivity (Sen), specificity (Spe), area under the ROC curve (AUC), and Standard Deviation (Std) - based on the number of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) pixels, were used to quantify segmentation results taking into account the labelled pixels from the ground-truth provided by experts.

A pixel-level segmentation of exudates could only be accurately validated on E-OPHTHA, since it is the first public database to provide pixel-level annotations by experts. For this reason, a comparison between the proposed method and existing classical methods which carried out a validation of their pixel-level classification on this database is shown in Table 1.

Table 1. Comparative exudate segmentation results for the validation of different methods at pixel-level on E-OPHTHA database.

	Accuracy	Sensitivity	Specificity	AUC
Haloi et al. [5]	-	0.9582	-	0.9620
Imani et al. [4]	-	0.8032	0.9983	0.9370
Proposed	**0.9936**	**0.8941**	**0.9931**	**0.9927**

To further evaluate the model's ability to generalize to heterogeneous fundus images with different acquisition methods, the proposed model was also validated on DIARETDB1 public database. This database consists of 89 retinal images acquired with the same 50°FOV digital fundus camera and, consequently, with a fixed resolution of 1500×1152 pixels. A spatial normalization revealed to be unnecessary since retinal structures are fairly comparable. Because this database contains images with the presence or absence of exudates, a subset of 42 images containing this type of lesion was selected for this experiment. Exudates are, once again, manually annotated by experts but not at a pixel-level which is equivalent to a wide amount of false positives around the lesions. For this reason, this database isn't suitable for a semantic segmentation validation. However, this test was performed specifically to validate the model's performance in different databases and a wider set of images.

In Table 2, the proposed method (tested in both DIARETDB1 (dtdb) and E-OPHTHA (eoph) databases) is compared with several algorithms which present

measurements for their performance at the pixel-level on private datasets or recurring to private ground-truth annotations for public databases. Even though these works don't use a common dataset or segmentation approach, this comparison is made to show the advantages of the proposed method in terms of the aforementioned measurements.

Table 2. Comparative exudate segmentation results for the validation at pixel-level of different methods on distinct datasets.

	Accuracy	Sensitivity	Specificity	AUC
Welfer et al. [6]	-	0.7048	0.9884	-
Sopharak et al. [8]	0.9910	0.8720	0.9920	-
Sopharak et al. [3]	0.9841	0.9228	0.9852	-
Harangi et al. [7]	-	0.86	-	-
Prentasiċ et al. [13]	-	0.78	-	-
Proposed (dtdb)	**0.9701**	**0.8451**	**0.9809**	**0.9535**
Proposed (eoph)	**0.9936**	**0.8941**	**0.9931**	**0.9927**

As it can be seen in Tables 1 and 2, the proposed method outperforms existing algorithms in most evaluation metrics. Haloi et al. [5] and Imani et al. [4] present higher values for sensitivity and specificity, respectively, on E-OPHTHA database, while Sopharak et al. [3] exceeds sensitivity values on a private dataset. Even though the proposed model's performance on DIARETDB1 isn't higher than most methods, it is still a great outcome taking into consideration the nature of this database, as it was previously explained.

Figure 3 illustrates the qualitative segmentation results on E-OPHTHA. The validation approach presents some drawbacks due to the nature of its performance evaluation method. Manual segmentation performed by humans at a

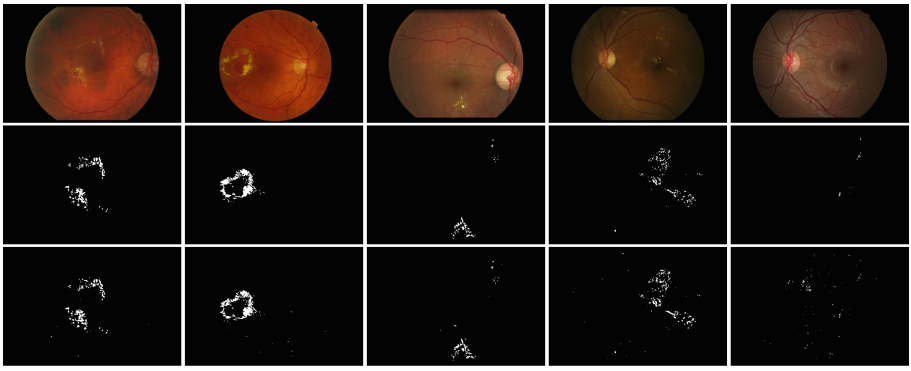


Fig. 3. Qualitative results of the exudate semantic segmentation. First row: original images; Second row: ground-truth annotations from experts; Third row: resulting segmentation maps from the proposed method.

pixel-level is prone to small-scale errors. In Fig. 3(e) it is noticeable that ground-truth annotations can be slightly misleading, originating a considerable amount of FP pixels which are, in fact, TP pixels. The remaining pixels misclassified as exudates are caused by the presence of noise and bright reflections along the main retinal vessels. There might be also some ambiguous regions where faint exudates are not considered by experts but accurately identified by the network. Furthermore, quantitative results are severely penalized in a pixel-level classification due to the small amount of pixels that are labelled as exudates in a retinal image. This means that the ratio between FN and TP pixels is inevitably lower, decreasing sensitivity values significantly. Nevertheless, the resulting segmentation maps are overall extremely similar to the expert's annotations, demonstrating the model's ability to accurately segment exudates, even in challenging situations such as in Fig. 3(e), where the image presents a lot of noise and uneven illumination. Prediction time is alongside segmentation accuracy when it comes to the major requirements for automated screening methods. Note that segmentation map takes around 36 s to be computed, allowing real-time feedback in clinical use.

4 Conclusions

In this work, a novel end-to-end network built on top of U-Net for semantic segmentation of exudates in fundus images is proposed. The preliminary experimental results show clear advantages of the proposed method over classical exudate segmentation algorithms.

In future work, bright reflections along the main retinal vessels will be the subject of post-processing techniques to reduce noise in segmentation results. In addition, grayscale images will be replaced by RGB images as input to the model. Finally, the proposed method will be applied to the detection of other kinds of DR related lesions.

Acknowledgements. This paper was supported by the European Union's Horizon 2020 research and innovation programme under the Project GALAHAD [H2020-ICT-2016-2017, 732613]. The work of Adrián Colomer has been supported by the Spanish Government under a FPI Grant [BES-2014-067889]. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

1. World Health Organization: Diabetes fact sheet. Sci. Total Environ. **20**, 1–88 (2011)
2. Verma, L., Prakash, G., Tewari, H.K.: Diabetic retinopathy: time for action. No complacency please! Bull. World Health Organ. **80**(5), 419–419 (2002)
3. Sopharak, A.: Machine learning approach to automatic exudate detection in retinal images from diabetic patients. J. Mod. Opt. **57**(2), 124–135 (2010)

4. Imani, E., Pourreza, H.R.: A novel method for retinal exudate segmentation using signal separation algorithm. Comput. Methods Programs Biomed. **133**, 195–205 (2016)
5. Haloi, M., Dandapat, S., Sinha, R.: A Gaussian scale space approach for exudates detection, classification and severity prediction. arXiv preprint arXiv:1505.00737 (2015)
6. Welfer, D., Scharcanski, J., Marinho, D.R.: A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images. Comput. Med. Imaging Graph. **34**(3), 228–235 (2010)
7. Harangi, B., Hajdu, A.: Automatic exudate detection by fusing multiple active contours and regionwise classification. Comput. Biol. Med. **54**, 156–171 (2014)
8. Sopharak, A., Uyyanonvara, B., Barman, S.: Automatic exudate detection from non-dilated diabetic retinopathy retinal images using fuzzy C-means clustering. Sensors **9**(3), 2148–2161 (2009)
9. Havaei, M., Davy, A., Warde-Farley, D.: Brain tumor segmentation with deep neural networks. Med. Image Anal. **35**, 18–31 (2017)
10. Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels with deep neural networks. IEEE Trans. Med. Imag. **35**(11), 2369–2380 (2016)
11. Pratt, H., Coenen, F., Broadbent, D.M., Harding, S.P.: Convolutional neural networks for diabetic retinopathy. Procedia Comput. Sci. **90**, 200–205 (2016)
12. Gulshan, V., Peng, L., Coram, M.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA **316**(22), 2402–2410 (2016)
13. Prentašić, P., Lončarić, S.: Detection of exudates in fundus photographs using deep neural networks and anatomical landmark detection fusion. Comput. Methods Programs Biomed. **137**, 281–292 (2016)
14. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
15. Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Garcia-Rodriguez, J.: A review on deep learning techniques applied to semantic segmentation, pp. 1–23. arXiv preprint arXiv:1704.06857 (2017)
16. Deng, Z., Fan, H., Xie, F., Cui, Y., Liu, J.: Segmentation of dermoscopy images based on fully convolutional neural network. In: IEEE International Conference on Image Processing (ICIP 2017), pp. 1732–1736. IEEE (2017)
17. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440. IEEE (2014)
18. Li, W., Qian, X., Ji, J.: Noise-tolerant deep learning for histopathological image segmentation, vol. 510 (2017)
19. Chen, H., Qi, X., Yu, L.: DCAN: deep contour-aware networks for object instance segmentation from histology images. Med. Image Anal. **36**, 135–146 (2017)
20. Walter, T., Klein, J.C., Massin, P., Erginay, A.: 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**(10), 1236–1243 (2002)
21. Morales, S., Naranjo, V., Angulo, U., Alcaniz, M.: Automatic detection of optic disc based on PCA and mathematical morphology. IEEE Trans. Med. Imaging **32**(4), 786–796 (2013)
22. Zhang, X., Thibault, G., Decencière, E.: Exudate detection in color retinal images for mass screening of diabetic retinopathy. Med. Image Anal. **18**(7), 1026–1043 (2014)