Boosted Exudate Segmentation in Retinal Images Using Residual Nets

Samaneh Abbasi-Sureshjani[1], Behdad Dashtbozorg[1], Bart M ter Haar Romeny[1,2], and François Fleuret[3]

[1] Eindhoven University of Technology, P.O. Box 513, 5600 MB, The Netherlands
[2] Sino-Dutch Biomedical and Information Engineering School, Northeastern University P.O. Box 129, 500, Zhihui Street, 110167 Shenyang, China
[3] Idiap Research Institute, Rue Marconi 19, 1920 Martigny, Switzerland
{s.abbasi,B.Dasht.Bozorg,B.M.terhaarRomeny}@tue.nl
francois.fleuret@idiap.ch

Abstract. Exudates in retinal images are one of the early signs of the vision-threatening diabetic retinopathy and diabetic macular edema. Early diagnosis is very helpful in preventing the progression of the disease. In this work, we propose a fully automatic exudate segmentation method based on the state-of-the-art residual learning framework. With our proposed end-to-end architecture the training is done on small patches, but at the test time, the full sized segmentation is obtained at once. The small number of exudates in the training set and the presence of other bright regions are the limiting factors, which are tackled by our proposed importance sampling approach. This technique selects the misleading normal patches with a higher priority, and at the same time avoids the network to overfit to those samples. Thus, no additional post-processing is needed. The method was evaluated on three public datasets for both detecting and segmenting the exudates and outperformed the state-of-the-art techniques.

Keywords: Exudate segmentation, retinal images, residual nets, importance sampling, diabetic retinopathy, diabetic macular edema

1 Introduction

Diabetes is threatening the health of many people in the world, and it normally remains undiagnosed unless its symptoms and complications appear. Among its different ophthalmic complications, diabetic retinopathy (DR) is one of the most common and the most vision-threatening complication. DR is classified to non-proliferative DR (NPDR) and proliferative diabetic retinopathy (PDR), which determines the severity level and the need for further treatments. On the other hand, diabetic macular edema (DME) defined as retinal thickening, is another important complication that might happen in eyes at any DR severity level. Diabetic eyes are typically categorized into three groups based on the DME severity (healthy, moderate or severe DME). There are several signs and

symptoms associated with DR and DME such as microaneurysms, hemorrhages, hard exudates and cotton-wool spots. Among these signs, exudates are the early signs of both moderate NPDR and moderate DME. They are largely made up of extracellular lipid, which has leaked from abnormal retinal capillaries, and appear as white, yellowish or waxy lesions situated mainly in the outer plexiform layer of the retina [6, 13].

In this work, we propose a fully automatic method for segmenting the exudates in retinal images. Retinal images are one of the cost-effective and non-invasive sources of medical information, which are widely used for diagnosis purposes and studying the progression of the diseases. There are several works in the literature proposed for segmenting the exudates, which are mainly based on a series of preceding handcrafted feature extraction and landmark classification steps (e.g. [3, 14]). The CNNs are taking over the conventional image processing approaches in various computer vision tasks, and medical applications are not excluded from this rapid change. Recently, new methods using the state-of-the-art convolutional neural networks (CNNs) have been proposed for automatic lesion and landmark detection, e.g., [1, 9, 11]. In the work proposed by [11], the exudate probability map created by a 10 layer CNN is combined with the output of other methods segmenting the optic disc, vessels, and bright borders to create the final masks for the exudates. While in [9], a 7-layer CNN is trained to localize the exudates and they are later used in another network for the automatic classification of DME.

In this article, we present a novel end-to-end segmentation framework using residual nets (ResNets) [5]. The depth of the neural networks is the key factor in enhancing the performance of visual recognitions tasks. A residual learning framework solves the difficulty of training very deep networks by introducing a reference to the input layer. Using this framework, it is possible to gain higher performance, while keeping the complexity the same. Since limited images with their annotations are available, we train the segmentation network with small patches and then after the kernels are learned well, we test the method on full-sized images. Bright reflections are always misleading for the network. In order to help the network to learn better from non-exudate, but exudate-looking samples, we propose a new sampling approach. This sampling technique presents the misleading negative samples to the network more often. This leads to a faster convergence compared to the uniform sampling. The paper is structured as follows: in Sec. 2, all the steps of our proposed network, sampling technique, and data preparations are explained in detail. In Sec. 3, the evaluation results are presented. At the end, the paper is concluded in Sec. 4.

2 Methodology

In this section, all the steps of our proposed technique are presented. An overview of our approach is depicted in Fig. 1. Each element in this figure will be explained in the following.

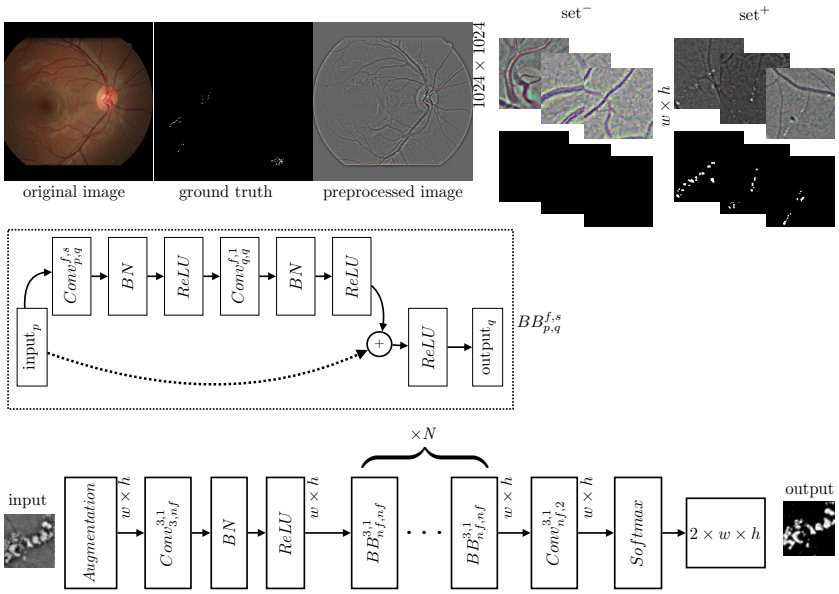

Fig. 1: An overview of our proposed architecture.

2.1 Material

There are very few public datasets with manual segmentations of exudates available. We used the e-ophtha-EX dataset [14], which contains 82 images with different image sizes ranging from 1440×960 px to 2544×1696 px. All the images are taken with the same field-of-view of 45°. Among these 82 images, 47 images contain in total 2278 annotated exudate components, and 35 images are normal images, which might contain misleading structures such as optical artifacts and vessel reflections. Since the images have different resolutions, after removing the black boundaries so that we have a square bounding box, each image, and its corresponding manual segmentation are scaled to have the resolution of 1024×1024 px. Then the images are enhanced by removing the local mean values using a Gaussian kernel (G_σ) with the scale of $\sigma = 1024/30$ as $I_{\text{enh}} = I - I * G_{\sigma=1024/30}$, where $*$ is the convolution operator. The top row, the left side of Fig. 1 shows a sample image before and after pre-processing together with its corresponding manual segmentation.

We split the data into the training/validation sets (80/20% split) at the image level. This splitting is done for both healthy and pathological images randomly. In the next step, for each of the training and validation sets, we define two different categories called set$^-$ and set$^+$. The set$^-$ includes patches that have been selected at random locations from the normal images; while the patches in set$^+$ have been selected around the centers of exudate components in the pathological images. All patches have a fixed size of $w \times h$. We selected $w = h = 128$ to have large enough patches so that different structures (blood

vessels, exudates, reflections, etc.) are easily differentiable from each other. Some sample patches for both set^- and set^+ are shown in the top row, the right side of Fig. 1. By selecting maximum 800 patches per image, there are in total 24035 patches in the training set ($|\text{set}^-| = 22400$ and $|\text{set}^+| = 1635$), and 6184 patches in the validation set ($|\text{set}^-| = 5600$ and $|\text{set}^+| = 584$).

The e-ophtha-EX set was mainly used for training the network. For evaluating the method, two other public datasets were used, the DiaRetDB1 [7] and DR2 [10] public datasets. The DiaRetDB1 includes 89 color fundus images captured using a 50° field-of-view digital fundus camera. Because the annotations have been done by different experts, we only consider the exudate annotations with the agreement higher than 75%. These annotations do not show the exact contours of the lesions and only an approximate mask of the lesion is provided. The same pre-processing as e-ophtha-EX dataset is applied on these images. The DR2 dataset includes in total 529 images and only the presence (not the location) of different types of lesions are provided. We only used 379 images including 300 normal images and 79 images with exudates. The images have the resolution of 867×575 px. We rescaled all the images to 512×512 px and pre-processed with $\sigma = 512/30$ after cropping the black boundaries.

2.2 Network Details

We use ResNets as the main blocks of our architecture. The basic ResNet block ($BB^{f,s}_{p,q}$) used in our model is shown in the middle row of Fig. 1. In this figure, $Conv^{f,s}_{p,q}$ is a 2D convolutional layer, where f, p, q, and s represent the filter size ($f \times f$), the number of input planes, the number of output planes, and the convolution step size, respectively. The batch normalization layer and the rectified linear unit are represented by BN and $ReLU$ respectively. The full model is depicted in the bottom row of Fig. 1. In this architecture, the cascade of N ResNet blocks ($BB^{3,1}_{nf,nf}$) in combination with two convolutional layers, one $ReLU$, one BN and one softmax layer is used. In this work, nf and N are set to 64 and 9 respectively, and the cross entropy criterion is used for the loss measurements. Using the filter sizes of $f = 3$ and stride $s = 1$ in all convolutional layers, the spatial resolution of the output is similar to the input, i.e. the input patch has the size of $3 \times w \times h$ and the output of the network has the size of $2 \times w \times h$. The first and second channels of the output represent for each pixel the probability of being a non-exudate or exudate pixel.

In each epoch, the input patches are augmented before feeding them to the network. The transformations include horizontal and vertical flipping with the probability of 0.5, rotation between $-10°$ to $10°$ with uniform probability and elastic transformations. The weights are initialized as in [4] and training was done from scratch. We used the stochastic gradient descent optimization technique with Nesterov momentum updates [8]. At the beginning, the learning rate was set to 0.1 and momentum to 0.9. We decreased the learning rate every time the loss stopped decreasing, and the training was done until there were no significant changes in the loss value.

2.3 Importance Sampling

One of the main difficulties in exudate detection is the high number of false positives due to the presence of reflections and other bright lesions. One way to solve this problem as proposed by [11, 14] is to provide several other masks (generated in separate pipelines) for these regions and remove falsely classified non-exudate pixels from the final exudate mask. However, we propose to solve this problem directly during training the network and without introducing any additional post-processing steps. The non-exudate pixels are mainly present in set^- and they result in a higher loss when comparing the output of the network to the ground truth. We propose to use the importance sampling approach so that the samples in set^- having a higher loss in previous epochs, have a higher chance to be seen in the next training iteration. In this way, the misleading negative samples are presented to the network more often and the network learns to adapt its weights and parameters accordingly.

If $\text{set}^+ = \{x_1, \ldots, x_M\}$ and $\text{set}^- = \{x_1, \ldots, x_N\}$ are the sets of training samples, then we assign a weight to each sample in these two sets as $W^+ = \{w_1, \ldots, w_M\}$ and $W^- = \{w_1, \ldots, w_N\}$, so that the sum of weights for both W^+ and W^- equals 1. These weights determine the probability of being selected for the next training iteration. Since the patches with exudates need to be treated equally, we define $w_i = 1/M$ for all $w_i \in W^+$. However, for the normal patches $w_i = \frac{l_i}{\sum_{j=1}^{N} l_j}$ for $w_i \in W^-$, where l_j is the loss of the network achieved for the sample $x_j \in \text{set}^-$ in the previous training epoch. For very few epochs at the beginning (e.g. 5) the network is trained using $2K$ samples selected only from set^+. This results in a good initialization of the network. Then in the rest of the iterations, K samples are selected from set^+ and K samples from set^- according to their defined weights. After feeding the minibatch to the network and computing the derivatives of the loss with respect to the output units, one more step is needed in order to avoid the network to get too biased toward examples with high losses. We rescale the gradient of the loss on individual samples of set^- with the scale of $1/\mu_j$, where $\mu_j = N\frac{w_j}{\sum_{j=1}^{N} w_j}$ (see [2] for more details). Finally, the weights of the network are updated accordingly.

3 Experiments

To get the segmentation for the entire image, we just pass the image through the network. The spatial resolution does not change in the network, and the patches are selected large enough to present enough contextual information; therefore, the learned kernels are sufficient for segmenting the full image. Figure 2 from left to right represents the full segmentation (probability maps) of two sample images from the e-ophtha-EX and two images from the DiaRetDB1 dataset, overlaid on top of the original images in the first row. Their corresponding ground truth images are presented in the second row.

In order to evaluate the performance of our method, we use the same evaluation approach as proposed by [14]. In this approach, the evaluation is done

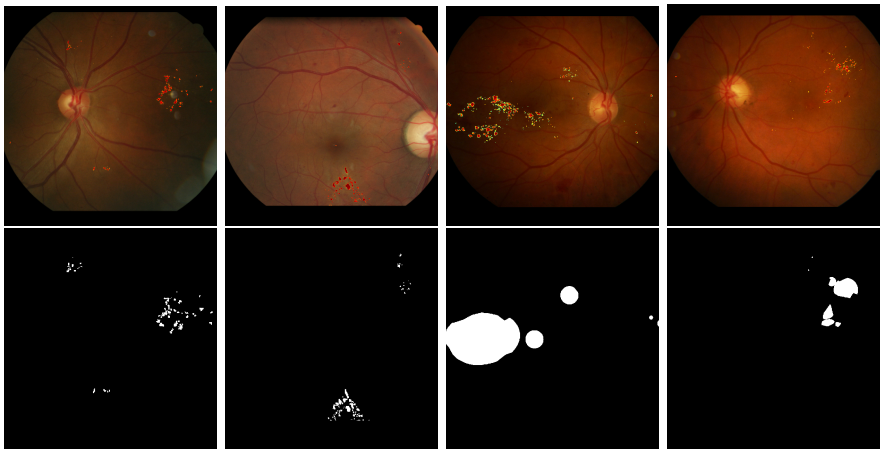

Fig. 2: Four sample segmentation results overlaid on the original images (top row) and their corresponding ground truth images (bottom row).

either at the exudate level or at the image level. At exudate level, one exudate component (defined as a set of connected pixels) is considered to be completely segmented, if there is a certain overlap between that exudate and the ground truth. For the definition of true/false positives and true/false negatives, the reader is referred to [14]. However, the image level validation only evaluates the correct prediction of the presence of exudates in the images, which is more important from the clinical point of view. To this aim, the image level prediction probability is defined as the maximum of all individual probabilities at the pixel level.

The results are reported in Table 1. The second column represents the total number of images in each set and the number of healthy vs. images with exudates (h/ex). The third column represents the F1-score at the exudate level. The next two columns present the F1-score and the area under the ROC curve (AUC) at the image level. Finally, the last column includes the AUC only at the image level reported by the state-of-the-art techniques. The DR2 set was only used for evaluation at the image level because the exudate masks are not available for this dataset. Our proposed method achieved the F1-score of 0.832 at the exudate level on the e-ophtha-EX dataset, which is higher than the score (F1-score = 0.732) reported by [14]. Using the same evaluation approach, the authors in [11] reported the F1-score of 0.78 for the DRiDB dataset [12], which was not available to us for comparison. Based on these results, our method outperforms the state-of-the-art techniques in most of the cases. The performance on the DR2 dataset is also as high as the results reported by [10]. The F1-score values at the exudate level are very good indications showing that the number of false positives is very low, i.e., the network is differentiating the exudate pixels from misleading non-exudate ones very well. The ROC curves for different datasets are also depicted in Fig. 3.

Table 1: The evaluation results on different public datasets

Dataset	size	Exudate	Image		Literature
	total (h/ex)	F1-score	F1-score	AUC	AUC
e-ophtha-EX [14]	82 (47/35)	0.832	0.967	0.994	0.95 [14]
DiaRetDB1 [7]	89 (59/30)	0.819	0.880	0.965	0.95 [14]
DR2 [10]	379 (300/79)	-	0.871	0.972	0.978 [10]

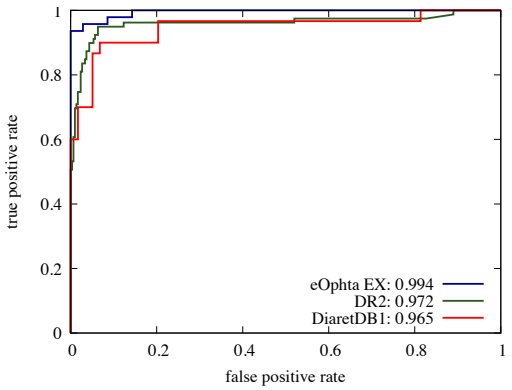

Fig. 3: The ROC curves (at the image level) obtained for different datasets.

4 Conclusion

We presented a novel method for exudate segmentation in retinal images. Our proposed network consists of 20 convolutional layers (9 ResNet blocks). The proposed importance sampling step, which prioritizes the sampling towards highly misleading non-exudate samples, helps to decrease the number of falsely detected non-exudate components to a great extent. Therefore, no additional post-processing steps are needed. Moreover, it helps to achieve a high performance, outperforming the state-of-the-art techniques, in less number of iterations. Even though the main goal of this work was to segment the exudate components, the results show that our network performs very well in deciding about the presence of exudates, which is typically enough for clinicians to take actions. This method might easily be used for detection of other types of lesions if their manual segmentations are available.

Acknowledgement

This project has received funding from the European Union's Seventh Framework Programme, Marie Curie Actions-Initial Training Network, under Grant Agreement No. 607643, "Metric Analysis For Emergent Technologies (MAnET)". It was also supported by the Hé Programme of Innovation, which is partly fi-

nanced by the Netherlands Organization for Scientific Research (NWO) under Grant No. 629.001.003.

References

1. Alghamdi, H.S., Tang, H.L., Waheeb, S.A., Peto, T.: Automatic optic disc abnormality detection in fundus images: A deep learning approach. In: Proceedings of the Ophthalmic Medical Image Analysis Third International Workshop, Held in Conjunction with MICCAI. pp. 17–24 (2016)
2. Canévet, O., Jose, C., Fleuret, F.: Importance sampling tree for large-scale empirical expectation. In: International Conference on Machine Learning. pp. 1454–1462 (2016)
3. Harangi, B., Lazar, I., Hajdu, A.: Automatic exudate detection using active contour model and regionwise classification. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 5951–5954 (2012)
4. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1026–1034 (2015)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
6. The International Council of Ophthalmology (ICO): ICO Guidelines for Diabetic Eye Care (January 2017)
7. Kälviäinen, R.V.J.P.H., Uusitalo, H.: DIARETDB1 diabetic retinopathy database and evaluation protocol. Medical Image Understanding and Analysis p. 61 (2007)
8. Nesterov, Y.: A method for unconstrained convex minimization problem with the rate of convergence O(1/k2). Soviet Mathematics Doklady 27(2), 372–376 (1983)
9. Perdomo, O., Otalora, S., Rodríguez, F., Arevalo, J., González, F.A.: A novel machine learning model based on exudate localization to detect diabetic macular edema. In: Proceedings of the Ophthalmic Medical Image Analysis Third International Workshop, Held in Conjunction with MICCAI. pp. 137–144 (2016)
10. Pires, R., Jelinek, H.F., Wainer, J., Valle, E., Rocha, A.: Advancing bag-of-visual-words representations for lesion classification in retinal images. PloS One 9(6), e96814 (2014)
11. Prentašić, P., Lončarić, S.: Detection of exudates in fundus photographs using deep neural networks and anatomical landmark detection fusion. Computer Methods and Programs in Biomedicine 137, 281–292 (2016)
12. Prentašić, P., Lončarić, S., Vatavuk, Z., Bencic, G., Subasic, M., Petkovic, T., Dujmovic, L., Malenica-Ravlic, M., Budimlija, N., Tadic, R.: Diabetic retinopathy image database (DRiDB): a new database for diabetic retinopathy screening programs research. In: 8th International Symposium on Image and Signal Processing and Analysis. pp. 711–716. IEEE (2013)
13. Raman, R., Nittala, M.G., Gella, L., Pal, S.S., Sharma, T.: Retinal sensitivity over hard exudates in diabetic retinopathy. Journal of Ophthalmic and Vision Research 10(2), 160 (2015)
14. Zhang, X., Thibault, G., Decencière, E., Marcotegui, B., Laÿ, B., Danno, R., Cazuguel, G., Quellec, G., Lamard, M., Massin, P., et al.: Exudate detection in color retinal images for mass screening of diabetic retinopathy. Medical Image Analysis 18(7), 1026–1043 (2014)