Deep Retinal Image Segmentation: A FCN-Based Architecture with Short and Long Skip Connections for Retinal Image Segmentation

Conference Paper · October 2017

DOI: 10.1007/978-3-319-70093-9_76

CITATIONS
29

READS
861

6 authors, including:

Jie Yang
Hefei University,China,Hefei
909 PUBLICATIONS **18,055** CITATIONS
SEE PROFILE

Yu QIAO
Shanghai Jiao Tong University
67 PUBLICATIONS **589** CITATIONS
SEE PROFILE

Deep Retinal Image Segmentation: A FCN-Based Architecture with Short and Long Skip Connections for Retinal Image Segmentation

Zhongwei Feng[1], Jie Yang[1(✉)], Lixiu Yao[1], Yu Qiao[1], Qi Yu[2], and Xun Xu[2]

[1] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China
{fengzhongwei,jieyang,lxyao,qiaoyu}@sjtu.edu.cn
[2] Shanghai General Hospital, Shanghai Jiao Tong University, Shanghai, China
1683289010@qq.com, 820478428@qq.com

Abstract. This paper presents Deep Retinal Image Segmentation, a unified framework of retinal image analysis that provides both optic disc and exudates segmentation. The paper presents a new formulation of fully Convolutional Neural Networks (FCNs) that allows accurate segmentation of the retinal images. A major modification in these retinal image segmentation tasks are to improve and speed-up the FCNs training by adding short and long skip connections in standard FCNs architecture with class-balancing loss. The proposed method is experimented on the DRIONS-DB dataset for optic disc segmentation and the privately dataset for exudates segmentation, which achieves strong performance and significantly outperforms the-state-of-the-art. It achieves 93.12% sensitivity (Sen), 99.56% specificity (Spe), 89.90% Positive predictive value (PPV) and 90.93% F-score for optic disc segmentation while 81.35% Sen, 98.76% Spe, 81.64% PPV and 81.50% F-score for exudates segmentation respectively.

Keywords: Optic disc segmentation · Exudates segmentation · Convolutional neural network · Skip connections · Class-balancing loss

1 Introduction

Retinal image segmentation is key for the diagnosis of ophthalmological diseases such as diabetes and hypertension. However, manual segmentation of retinal images by ophthalmologist is both time consumed and lack accuracy and. Accordingly, automated segmentation systems can play an important role in improving the segmentation accuracy as well as diagnosis accuracy.

Optic Disc Segmentation: Prior attempts of optic disc segmentation includes handcrafted features [1] and morphology [2]. Morales et al. [3] use morphological

D. Liu et al. (Eds.): ICONIP 2017, Part IV, LNCS 10637, pp. 713–722, 2017.
https://doi.org/10.1007/978-3-319-70093-9_76

operators with Principal Component Analysis to obtain the optic disc in retinal images. In the last ten years, machine learning based methods have gained attentions, which is a powerful tool for feature classification. In [4], the authors propose an automatic method of segmenting OD and optic cup based on a statistical model technique. A superpixel classification method is proposed in [5]. However, handcrafted features has limit to different datasets. Recently, Convolutional Neural Networks (CNNs) is applied in the optic disc segmentation task. Lim et al. [6] describe a comprehensive solution based on applying convolutional neural networks to feature exaggerated inputs emphasizing disc pallor without blood vessel obstruction. Deep Retinal Image Understanding (DRIU) [7] uses FCNs [8] based on VGG-16 [9] with two set of specialized layers to solve both retinal vessels and optic disc segmentation. However, the models have large number of parameters, which are difficult to fit extremely small datasets.

Exudates Segmentation: Previous work on exudates segmentation can be broadly divided into unsupervised methods and supervised methods. Harangi et al. [10] use an active contour based method to extract accurate borders of the candidates and let a boosted Naive Bayes classifier eliminate the false candidates. In [11], the authors propose a svm-based method using the Gabor features and GLCM features to classify retinal images into exudates or nonexudates. Ardiyanto et al. [12] use maximum entropy principle to determines a reasonable threshold value for separating exudates areas. CNN-based methods are also arised in the exudates segmentation task. Prentašić et al. [13] uses CNNs combined with high level knowledge about landmark points, which contains vessels and optic disc, in order to increase the accuracy of exudate detection. However, its performance depends upon landmark points detection, which is a difficult problem factually. Perdomo et al. [14] work on LeNet [15] convolutional neural network architecture to improve the classification of nonexudates and exudates in retinal images to correctly diagnosis the disease. Both existed CNN-based methods on exudates segmentation have been trained by patch-based approach, which consumes a lot of computations and time.

Contribution: In this paper, we propose a modified FCN architecture by adding short and long connections for both optic disc and exudates segmentation, which is much more lightweight in number of parameters. Taking advantage of FCN, our proposed model can achieve end-to-end training, which consumes less time compared with patch-based training. To the best of our knowledge, our work is the first of its kind to bring the combined advantage of FCNs approach for exudates segmentation in retinal images. The rest of the paper is organized as follows: Sect. 2 describes the proposed methodology in detail. In Sect. 3 shows the experimental results. Finally, in Sect. 4 we conclude our paper with a summary.

2 Materials and Methods

Figure 1 presents a pipeline of our method for both optic disc and exudates segmentation. Only Contrast Limited Adaptive Histogram Equalization (CLAHE) is used as the preprocessing for the proposed method.

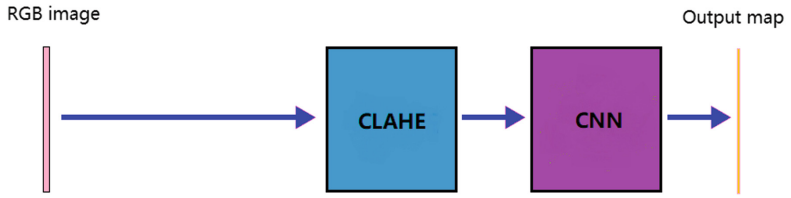
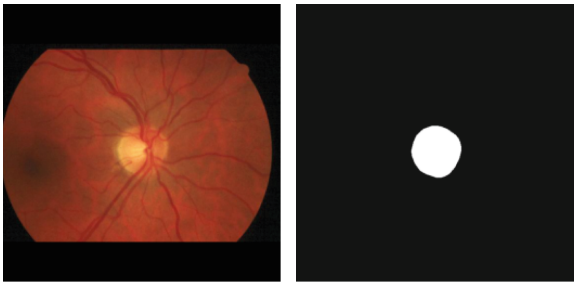


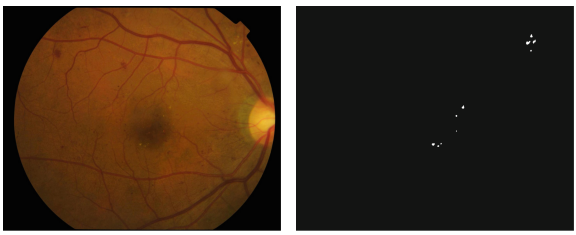
Fig. 1. Pipeline of the proposed method for the task of both optic disc and exudates segmentation.

2.1 Dataset Description

In this paper, results of optic disc segmentation are reported for the publicly available dataset DRIONS-DB [16], which contains groundtruth for optic disc. DRIONS-DB contains 110 full eye fundus images in total, which is divided into two parts for training and testing and each part contains 55 images equally. Figure 2(a) and (b) shows an example from the DRIONS-DB dataset. The results of exudates are tested on the privately dataset with precise groundtruth segmentation, which manually labeled by experienced ophthalmologists in Shanghai



(a) DRIONS-DB images (b) Ground Truth



(c) Exudates images (d) Ground Truth

Fig. 2. Examples from datasets. (a), (b) are retinal images and its corresponding groundtruth from DRIONS-DB. (c), (d) are retinal images and its corresponding groundtruth from the EX dataset.

No. 1 Hospital. The exudates dataset has 62 full eye fundus images totally, which is divided into the trainset with 42 images and the rest for testing. Figure 2(c) and (d) shows an example from the exudates dataset. Limited to GPU memory, all the images are resize to $256 * 256$ for training.

2.2 Data Augmentation

As for retinal image segmentation tasks there is very little training images available. On the other hand, CNN models easily have the chance of overfitting the training data. However, data augmentation is used to help prevent this behaviour. Data augmentation is the process to generate new samples by applying transformations to the original images. In this paper, we extend keras [17] ImageDataGenerator to real-time SegImageGenerator for both optic disc and exudates segmentation tasks. In particular, during training phase, we apply same transformations on image and corresponding groundtruth, such as flip, crop & resize, rotation, shift and etc.

2.3 Network Architecture

In this section, a unified method is proposed for segmentation of optic disc and exudates. Our approach is primarily based on deep learning techniques, which have outperformed the-state-of-the-art in many visual tasks. As for the classification task, residual networks [18] with short skip connections prominently performs the-state-of-the-art. Furthermore, CNNs have been applied successfully to a large variety of general recognition tasks such as object detection [19], semantic segmentation [8], and contour detection [20].

The U-net [21] is a FCNs architecture for image segmentation that accepts image as an input and returns softmap as an output, which shows good performance for biomedical image segmentation. The standard U-net has only long skip connections by recovering spatial information lost during downsampling, which has capable of training on little datasets and achieving segmentation competitive with patch-based methods. Inspired by the residual network [18] with short skip connections, which outperforms the-state-of-the-art in many classification tasks. We extend standard U-net by replacing regular convolutional layers with modified residual blocks, which are similar to the ones introduced in residual networks, in order to speed-up training and improve precise segmentation. The proposed architecture is illustrated in Fig. 3(a). Its input to the simple block in the proposed architecture is a $3 \times 256 \times 256$ image from the training set. The simple block, consisting of a 1×1 convolution, followed by a batch normalization [22] layer for speed-up training and a rectified linear unit, is illustrated in Fig. 3(b). The encoding path follows the typical architecture of U-net, but replacing regular convolutional layers with modified residual blocks. Each modified residual block, which is showed in Fig. 3(c), consists of the repeated application 3 times of the basic block. In order to equal the channels of its input and output for element-wise addition, a extra regular convolutional layer on shortcut will be added if

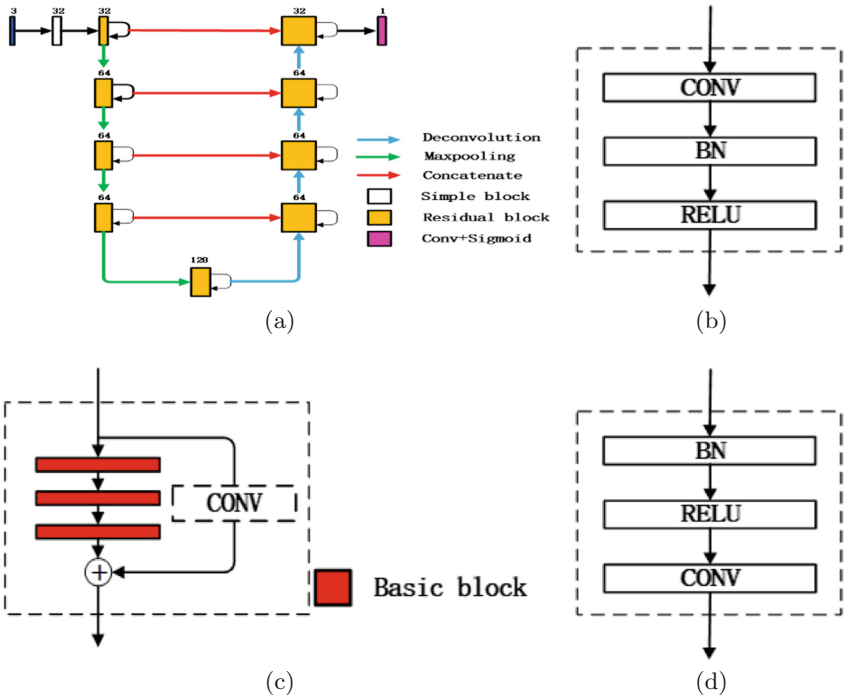


Fig. 3. Proposed CNN architecture for retinal image segmentation.(a) U-net with short skip connections built from residual blocks (b) simple block (c) residual block and (d) basic block

necessary (Fig. 3(c) dotted rectangle on shortcut). The basic block consists of a batch normalization layer, a rectified linear unit and a 3×3 convolution with padding for same size, which is illustrated in Fig. 3(d). Each modified residual block followed by a 2×2 max pooling operation with 2 pixels stride for downsampling to reduce the amount of parameters and computation. In order to concate different-size feature maps through long skip connections, an upsampling of the feature map followed by a 2×2 deconvolution [8] is used. This important architecture is that allow the network to propagate context information to higher resolution layers for more precise segmentation. After a concatenation, each followed by a modified residual block in decoding path. At the final convolutional layer a 1×1 convolution is used to map multi-channel maps to the desired number of classes. Finally, a sigmoid operation is applied on the last convolutional layer to scale the softmap. The output is of dimension $1 \times 256 \times 256$.

For training the network, the segmentation tasks are learnt by class-balancing loss function originally proposed in [23] for contour detection in natural images. The class-balancing loss function is then defined as:

$$
\begin{aligned}
L(y_i, \hat{y}_i) &= -\beta \sum_{i \in Y_+} (y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)) \\
&\quad - (1 - \beta) \sum_{i \in Y_-} (y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)) \\
&= -\beta \sum_{i \in Y_+} \log \hat{y}_i \\
&\quad - (1 - \beta) \sum_{i \in Y_-} \log(1 - \hat{y}_i)
\end{aligned} \tag{1}
$$

where $\hat{y}_i$ is the predicted segmentation map obtained by the last sigmoid layer while $y_i \in \{0, 1\}$ is the ground truth. The multiplier β is used to achieve the balance of the large number of background compared to foreground pixels. Y_- and Y_+ denote the background and foreground sets of the ground truth Y, respectively. In this case, we use $1 - \beta = |Y_+|/|Y|$, $\beta = |Y_-|/|Y|$.

2.4 Training Parameters

Training consists in an iterative propagation through the network and modification of its weights, which are initialized by xavier method. The cycle of presenting all training examples, called an epoch, is split into smaller units called batches. In this approach, we use mini-batch stochastic gradient descent with momentum at 0.95 and fixed the learning rate to be 10^{-3}. The model, which includes new 400 batches that generated by real-time SegImageGenerator in each epoch, is trained 500 epochs with a batch size of 4. The implementation is based on keras framework [17], which performs all computation on GPUs in single precision arithmetic. The experiments are conducted on Intel Core i7-6700 CPU with a NVIDIA TitanX card. In practical, these initial parameters is same for optic disc and exudates segmentation.

3 Experiment and Discussions

In test phase, we present the performance of networks tested on the test set in terms of Sen, Spe, PPV and F-score, which is obtained on pixel-wise. In this case, positive decision is made when the output of the unit associated with the positive class in the sigmoid output layer is greater than the 0.5; otherwise, negative decision is made. In fact, this threshold can be arbitrarily selected from [0, 1], which leads to different result.

Optic Disc Segmentation: Figure 4 shows exemplary segmentations produced by our proposed model for optic disc segmentation. For the task of optic disc segmentation, we compare our solution with the method from DRIU, which is the best method that we have found for investigated datasets. In Table 1, we present the performance of networks tested on the test set. The results show that our method performs a little better than DRIU comparison on DRIONS-DB dataset, which has strong performance and significantly outperforms the-state-of-the-art for optic disc segmentation.

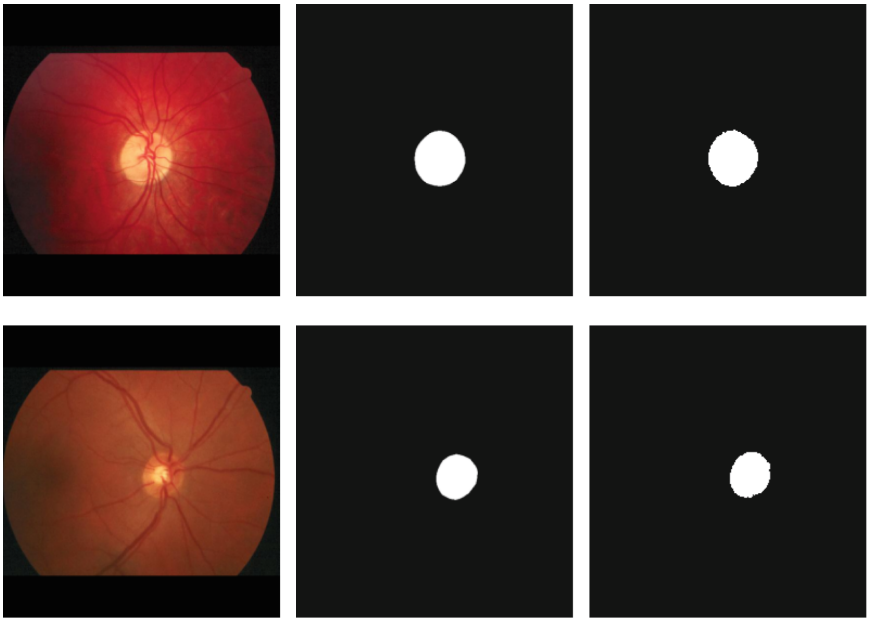


Fig. 4. Output map produced by our proposed method on the DRIONS-DB: first column, original retinal images; middle column, expert annotations; last column, results obtained by our method

Table 1. Performance on optic disc segmentation.

Method	Sen	Spe	PPV	F-score
Yin et al. [4]	.8542	.9826	-	-
Lim et al. [6]	.8742	-	.8535	.8636
Maninis et al. [7]	.9252	**.9972**	.8758	.9082
proposed method	**.9312**	.9956	**.8990**	**.9093**

Exudates Segmentation: Figure 5 shows accuracy segmentations produced by our method for exudates segmentation. In Table 2, we present the performance tested on the test set. The proposed method has strong performance and significantly outperforms the-state-of-the-art. Interestingly, taking a closer look at some true negative of our technique (see Fig. 5 red rectangles on the middle), we observe that almost true negative is very small. Although many convolution operators are applied on, it is hard to learn useful features to classify very small exudates in retinal images. Recently, we extend our proposed architecture to multi-scale inputs, which is expected to solve this problem.

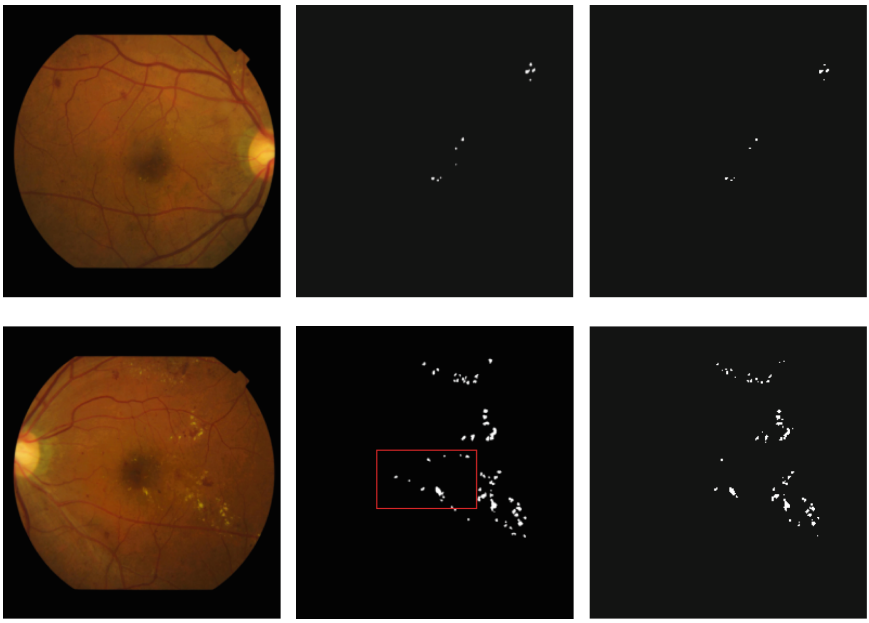


Fig. 5. Output map produced by our proposed method on the EX dataset:first column, original retinal images; middle column, expert annotations; last column, results obtained by our method (Color figure online)

Table 2. Performance on exudates segmentation.

Method	Sen	Spe	PPV	F-score
Ardiyanto et al. [12]	.1690	**.9960**	.6940	-
Harangi et al. [10]	.7123	-	.6613	.6854
Prentaic et al. [13]	.7854	.9954	.7834	.7812
proposed method	**.8135**	.9876	**.8164**	**.8150**

4 Conclusion

Combining with class-balancing loss, we presented a fully convolutional neural network with short and long skip connections for both optic disc and exudates segmentation. The proposed architecture has high-capability to learn hierarchical features and context information from raw pixel data without handcrafted features and any prior domain knowledge.

In this paper, the proposed method has strong performance and significantly outperforms the-state-of-the-art for both retinal optic disc and exudates segmentation. Taking advantage of FCN with short and long skip connections, our method has full potential of carrying out more robust and precise segmentation than traditional methods. So, we currently work on verifying this claim for hemorrhages and microaneurysms segmentation on retinal images.

Acknowledgments. This research is partly supported by NSFC, China (No: 81600776), Committee of Science and Technology, Shanghai, China (No: 16411962100) and (No. 17JC1403000)

References

1. Youssif, A.A.H.A.R., Ghalwash, A.Z., Ghoneim, A.A.S.A.R.: Optic disc detection from normalized digital fundus images by means of a vessels' direction matched filter. IEEE Trans. Med. Imaging **27**(1), 11–18 (2008)
2. Walter, T., Klein, J.-C.: Segmentation of color fundus images of the human retina: detection of the optic disc and the vascular tree using morphological techniques. In: Crespo, J., Maojo, V., Martin, F. (eds.) ISMDA 2001. LNCS, vol. 2199, pp. 282–287. Springer, Heidelberg (2001). doi:10.1007/3-540-45497-7_43
3. Morales, S., Naranjo, V., Angulo, J., Alcaniz, M.: Automatic detection of optic disc based on PCA and mathematical morphology. IEEE Trans. Med. Imaging **32**(4), 786–796 (2013)
4. Yin, F., Liu, J., Wong, D.W.K., Tan, N.M., Cheung, C., Baskaran, M., Aung, T., Wong, T.Y.: Automated segmentation of optic disc and optic cup in fundus images for glaucoma diagnosis. In: 2012 25th International Symposium on Computer-Based Medical Systems (CBMS), pp. 1–6. IEEE (2012)
5. Cheng, J., Liu, J., Xu, Y., Yin, F., Wong, D.W.K., Tan, N.M., Tao, D., Cheng, C.Y., Aung, T., Wong, T.Y.: Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. IEEE Trans. Med. Imaging **32**(6), 1019–1032 (2013)
6. Lim, G., Cheng, Y., Hsu, W., Lee, M.L.: Integrated optic disc and cup segmentation with deep learning. In: 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 162–169. IEEE (2015)
7. Maninis, K.-K., Pont-Tuset, J., Arbeláez, P., Van Gool, L.: Deep retinal image understanding. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 140–148. Springer, Cham (2016). doi:10.1007/978-3-319-46723-8_17
8. 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 (2015)
9. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint (2014). arXiv:1409.1556
10. Harangi, B., Lazar, I., Hajdu, A.: Automatic exudate detection using active contour model and regionwise classification. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 5951–5954. IEEE (2012)
11. Ruba, T., Ramalakshmi, K.: Identification and segmentation of exudates using SVM classifier. In: 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1–6. IEEE (2015)
12. Ardiyanto, I., Nugroho, H.A., Buana, R.L.B.: Maximum entropy principle for exudates segmentation in retinal fundus images. In: 2016 International Conference on Information & Communication Technology and Systems (ICTS), pp. 119–123. IEEE (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. Perdomo, O., Arevalo, J., Gonzalez, F.A.: Convolutional network to detect exudates in eye fundus images of diabetic subjects. In: 12th International Symposium on Medical Information Processing and Analysis, p. 101600T. International Society for Optics and Photonics (2017)
15. Lenet, B., Komorowski, R., Wu, X.Y., Huang, J., Grad, H., Lawrence, H., Friedman, S.: Antimicrobial substantivity of bovine root dentin exposed to different chlorhexidine delivery vehicles. J. Endod. **26**(11), 652–655 (2000)
16. Carmona, E.J., Rincón, M., Garcia-Feijoó, J., Martínez-de-la Casa, J.M.: Identification of the optic nerve head with genetic algorithms. Artifi. Intell. Med. **43**(3), 243–259 (2008)
17. Chollet, F., et al.: Keras (2015). https://github.com/fchollet/keras
18. 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)
19. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **38**(1), 142–158 (2016)
20. Yang, J., Price, B., Cohen, S., Lee, H., Yang, M.H.: Object contour detection with a fully convolutional encoder-decoder network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 193–202 (2016)
21. 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). doi:10.1007/978-3-319-24574-4_28
22. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint (2015). arXiv:1502.03167
23. Xie, S., Tu, Z.: Holistically-nested edge detection. In: Proceedings of IEEE International Conference on Computer Vision (2015)