

Optic Disc and Cup Segmentation Methods for Glaucoma Detection with Modification of U-Net Convolutional Neural Network¹

A. Sevastopolsky

*Department of Mathematical Methods of Forecasting, Faculty of Computational Mathematics and Cybernetics,
Moscow State University, Moscow, Russia
e-mail: artem.sevastopolsky@gmail.com*

Abstract—Glaucoma is the second leading cause of blindness all over the world, with approximately 60 million cases reported worldwide in 2010. If undiagnosed in time, glaucoma causes irreversible damage to the optic nerve leading to blindness. The optic nerve head examination, which involves measurement of cup-to-disc ratio, is considered one of the most valuable methods of structural diagnosis of the disease. Estimation of cup-to-disc ratio requires segmentation of optic disc and optic cup on eye fundus images and can be performed by modern computer vision algorithms. This work presents universal approach for automatic optic disc and cup segmentation, which is based on deep learning, namely, modification of U-Net convolutional neural network. Our experiments include comparison with the best known methods on publicly available databases DRIONS-DB, RIM-ONE v.3, DRISHTI-GS. For both optic disc and cup segmentation, our method achieves quality comparable to current state-of-the-art methods, outperforming them in terms of the prediction time.

Keywords: glaucoma detection, eye fundus, image segmentation, computer vision, optic disc segmentation, optic cup segmentation, convolutional neural network, deep learning, U-Net

DOI: 10.1134/S1054661817030269

1. INTRODUCTION

Glaucoma is the second leading cause of blindness all over the world, with approximately 60 million cases reported worldwide in 2010, and an increase by 20 million is expected in 2020 [1, 2]. If left unnoticed, glaucoma can cause irreversible damage to the optic nerve leading to blindness. Therefore, diagnosing glaucoma at early stages is very important [1].

Optic nerve examination includes eye fundus test, which requires a doctor localizing areas of optic disc and optic cup (central part of optic disc) and finding their borders. Presence of glaucoma can be identified by noticing optic nerve cupping, i.e., increase of optic cup in size. One of the main indicators of the disease is cup-to-disc ratio (CDR)—a ratio between heights of cup and disc [1]. It is considered one of the most representative features of optic disc and cup areas for glaucoma detection, and, according to [3], eye with CDR of at least 0.65 is usually considered as glaucomatous in clinical practice. Figure 1 shows an example of healthy and glaucoma-suspicious eye.

Segmentation of the optic disc and cup and determination of the CDR are very time-consuming tasks currently performed only by professionals. As stated in [5], according to a research, full segmentation of optic

disc and cup requires about eight minutes per eye for a skilled grader. Solutions for automated analysis and assessment of glaucoma can be very valuable in various situations, such as mass screening and medical care in countries with significant lack of qualified specialists [6, 7].

There are several approaches to development of computer vision algorithms for glaucoma detection based on eye fundus images. First approach is to determine the presence of the disease directly from fundus images, which involves either manual or automatic extraction of image features, derived from color, position and pairwise relation of pixels. Another approach is to build algorithms for optic disc and cup segmentation, then, based on that, read out disc and cup dimension and from that judge on presence of the disease. In this work we investigate the latter pipeline, since it can provide more transparent and reliable solution for a medical doctor.

Recognition quality and prediction time are the major requirements to the solution for automatic segmentation of eye parts. In order for a computer to be a decision-making system or at least an automatic eye scanner, it must make segmentation errors very seldom. Prediction time is also very important, especially when it is required to analyze large number of pictures in a small amount of time. Training time may be a concern in case retraining of an algorithm on larger database is needed frequently. However, exact require-

¹The article is published in the original.

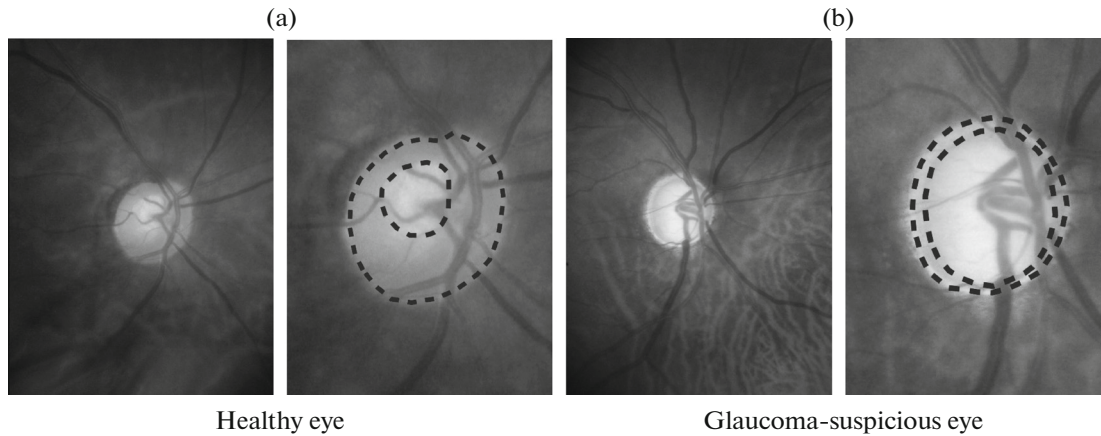


Fig. 1. An example of healthy and glaucoma-suspicious eye from RIM-ONE v.3 [4] database. Right-hand picture of each example contains enlarged optic disc area, where optic disc border is indicated by outer dashed line, optic cup border—by inner dashed line. Note that CDR is larger for glaucoma-suspicious eye.

ments to the method depend on a specific setting of an automatic assessment system.

2. RELATED WORK

In this section we give an overview of several methods for optic disc and cup segmentation that have been evaluated by their authors on publicly available datasets with both images and groundtruth provided.

For optic disc segmentation task, authors of [8] use Fully-convolutional neural network [9] based on VGG-16 net [10] and transfer learning technique. They achieve superhuman quality of recognition in terms of Dice score (see Section 3 of this paper) and boundary error (mean distance between the boundary of the result and that of the ground truth), since obtained results are more consistent with a gold standard than a second human annotator used as control.

For optic cup segmentation task, authors of [11] use 2-layer multi-scale convolutional neural network trained with boosting. Training process pipeline is multi-stage and includes patches preparation and neural network training. For pre-processing, entropy filtering [12] in $L^*a^*b^*$ color space is performed for extracting the most important points of an image, followed by contrast normalization and standardization of patches. Gentle AdaBoost [13] algorithm is then used to train convolutional filters, which are represented as linear regressors for small patches. At the test time, image propagation through the network is followed by unsupervised graph cut [14]. The method was evaluated on DRISHTI-GS [15, 16] database, and it outperformed all other existing methods in terms of Intersection-over-Union score and Dice score (see Section 3 of this paper). However, it is necessary to note that this method crops images by area of their optic disc (cup) before performing segmentation of the optic disc (cup). It makes the method not appli-

able to new, unseen images of full eye fundus, since it requires a bounding box of optic disc and cup to be available in advance.

The paper [17] suggests an improvement to the aforementioned method in the training procedure for convolutional filters. Evaluation on DRISHTI-GS and RIM-ONE v.3 [4] databases for optic disc and cup is provided. Compared to the previous method, it does not require the images to be cropped by the area of optic cup for its segmentation, which makes the solution applicable to previously unseen images.

Method from the paper [8] has several drawbacks. It uses a deep neural network which takes a long time to train, model is large in terms of size of the file with network parameters and amount of required GPU memory. Authors of the paper were not pursuing a goal of the optic cup segmentation, which is a more challenging task than the optic disc segmentation. Besides, we were unable to reproduce the reported results. Methods from [11, 17] are very complicated, hard to program and to reproduce the results. Being prepared for execution on CPU, they also have large prediction time. As written before, [11] method required images to be cropped by the area of optic cup in advance, which is another drawback of a method. Some methods that are not mentioned in this section, such as [5, 18, 19], have mostly been evaluated either on datasets that are not currently publicly available, or on very small datasets, or used metrics dependent on proportion between classes, thus making it harder to compare with them.

3. THE PRESENTED APPROACH

In this section, the universal method is proposed for segmentation of optic disc and cup. Our approach is primarily based on deep learning techniques, which have made a revolution in all tasks of computer vision

Table 1. Comparison of methods for optic disc segmentation. “–” indicates that the result is not reported. Training time is computed as a product of one epoch time and average number of epochs

	DRIONS-DB		RIM-ONE v.3		Training time on RIM-ONE v.3	Prediction time	Parameters
	IOU	<i>Dice</i>	IOU	<i>Dice</i>			
Our approach	0.89	0.94	0.89	0.95	26 s × 382 = 9932 s	0.1 s	6.6×10^5
DRIU [8]	0.88	0.97	0.89	0.96	56 s × 200 = 11200 s	0.13 s	1.5×10^7
Zilly et al. [17]	–	–	0.89	0.94	3296 s	5.3 s	1890

Table 2. Comparison of methods for optic cup segmentation. “–” indicates that the result is not reported

	DRISHTI-GS		RIM-ONE v.3		Prediction time, s
	IOU	<i>Dice</i>	IOU	<i>Dice</i>	
Our approach	0.75	0.85	0.69	0.82	0.06
Zilly et al. [17]	0.85	0.87	0.80	0.82	5.3
BCF [11]	0.86	0.83	–	–	–

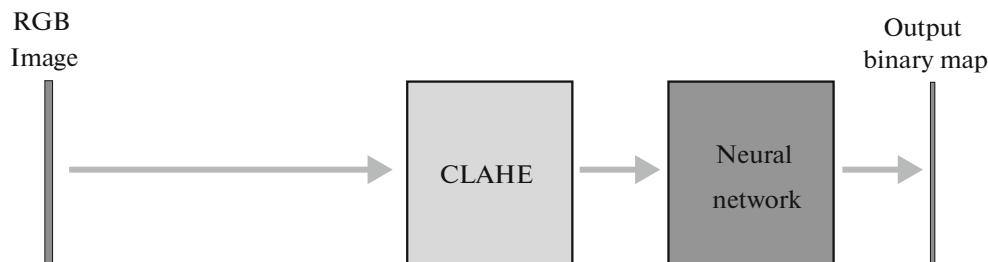
in the last years and currently provide state-of-the-art solutions in image classification, segmentation and many other image recognition tasks. Another advantage of convolutional neural networks as main tools of deep learning is their universality, as the same network can usually recognize various patterns in different images and for different objects.

Figure 2 presents a pipeline of our method for optic disc segmentation, Fig. 3—for optic cup segmentation. Contrast Limited Adaptive Histogram Equalization (CLAHE) [20] is used as a pre-processing for both methods. It equalizes contrast by changing color of image regions and interpolating the result across them. For optic cup, we firstly crop the images by bounding box of optic disc (with margin from each side), which can be acquired from trained algorithm for optic disc.

Core component of the method is a convolutional neural network built upon U-Net [21]. It is a neural network for image segmentation that accepts image as an input and returns probability map as an output. U-Net was introduced as a Fully-convolutional neural network capable of training on extremely small datasets and achieving results competitive with sliding-

window based models. Trained with specific data augmentation and enhancement techniques, it outperforms existing methods on several biomedical image segmentation challenges [21].

The architecture presented in the paper is depicted in Fig. 4. Like the original U-Net, it consists of contracting path (left side) and an expansive path (right side). Contracting path structurally repeats a typical architecture of convolutional part of the classification network, e.g., VGG-16 [10]. On the expansive path, information is merged from layers of contracting path of appropriate resolution and layers of expansive path of lower resolution, so that a whole network recognizes patterns at several scales. Input image is firstly passed through a convolutional layer with filters of 3×3 pixels spatial resolution; number of filters in a layer is shown in the figure above a blue rectangle representing layer’s output. Afterwards, Dropout regularization [22] and ReLu activation function ($f(x) = \max(0, x)$) are applied. The same is repeated again, and Max Pooling operation is applied, reducing image width and height by two. Image is then passed through aforementioned sequence of layers multiple times, until

**Fig. 2.** Pipeline of the proposed method for the task of optic disc segmentation.

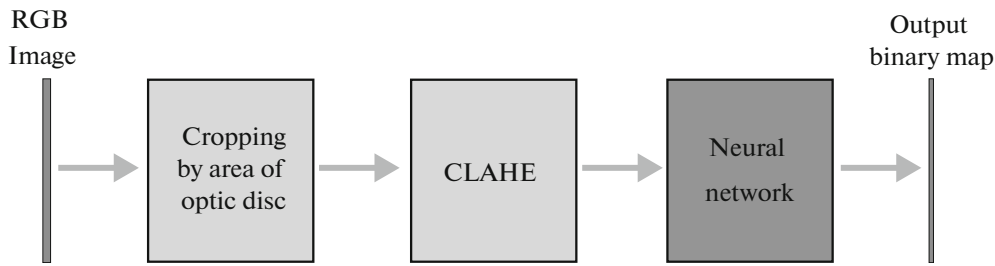


Fig. 3. Pipeline of the proposed method for the task of optic cup segmentation.

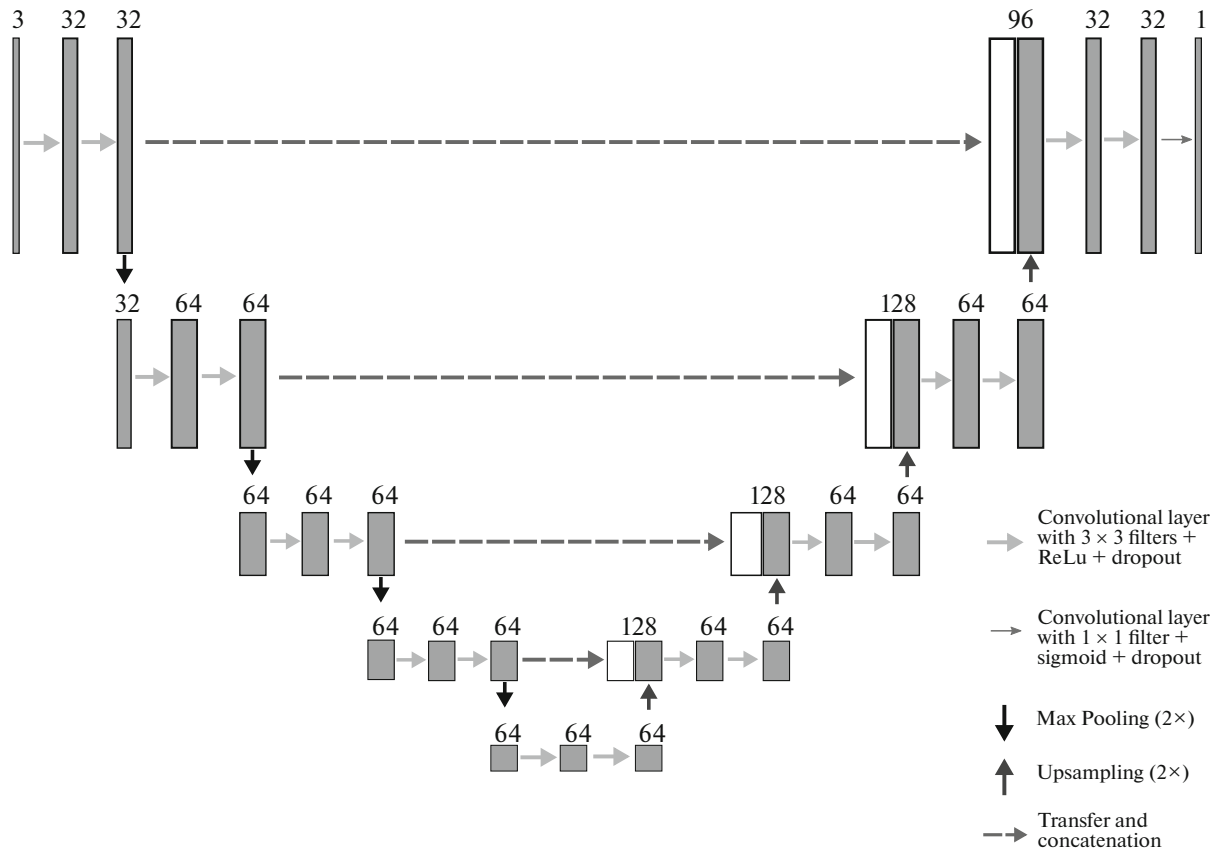


Fig. 4. Architecture of neural network employed in our method.

resolution is low enough. On the expansive path, the same convolutional layers are applied, interleaved with Upsampling layers, which raise image width and height by two in a trivial way.

Compared to original U-Net, the presented modification has less filters in all convolutional layers and does not possess an increasing number of filters for decreasing resolution. Our experiments have shown that these changes do not lower quality of recognition for our tasks, but make the architecture much more lightweight in terms of number of parameters and training time. As a loss function, we use $l(A, B)$:

$$l(A, B) = -\log d(A, B),$$

where

$$d(A, B) = \frac{2 \sum_{i,j} a_{ij} b_{ij}}{\sum_{i,j} a_{ij}^2 + \sum_{i,j} b_{ij}^2},$$

where $A = (a_{ij})_{i=1,j=1}^{H,W}$ is a predicted output map, containing probabilities that each pizel belongs to the foreground, and $B = (b_{ij})_{i=1,j=1}^{H,W}$ is a correct binary output map.

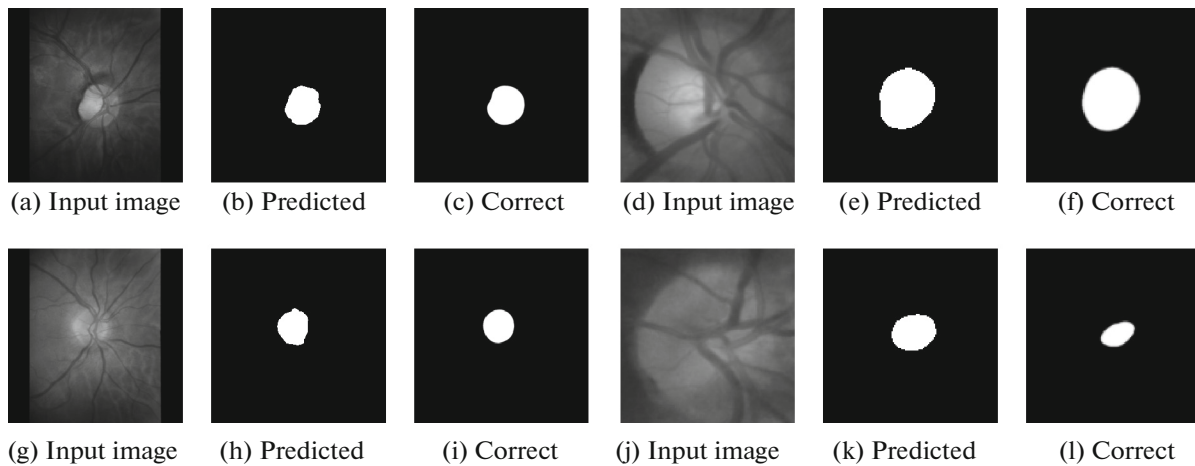


Fig. 5. Visual comparison of the predicted results and correct segmentation on RIM-ONE v.3 for the optic disc (a–c, g–i) and cup (d–f, j–l). On (d–f, j–l) region of the optic disc is shown as an input image. For optic disc: (a–c) best case (IOU = 0.93, Dice = 0.97), (g–i) worst case (IOU = 0.80, Dice = 0.90); for optic cup: (d–f) best case (IOU = 0.93, Dice = 0.97), (j–l) worst case (IOU = 0.46, Dice = 0.64).

$d(A, B)$ is an extension of Dice score for binary images $Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|}$; if A and B contain only

binary values, $d(A, B)$ and $Dice(A, B)$ are equal, but $d(A, B)$ also supports values that lie in $(0, 1)$. This extension allows us to compute gradient of the loss function. Stochastic Gradient Descent (SGD) with momentum [23] was used as an optimization method. During the training, data augmentation was used to enlarge the training set by artificial examples. Images were subject to random rotations, zooms, shifts and flips.

It is necessary to note that the proposed method does not require any preliminary cropping of input images to area of the optic disc, as it can segment the optic disc and the optic cup on a full eye fundus image. Detailed comparison of the presented method with the existing ones is given in the Section 4.

4. EXPERIMENTS

This section of the paper contains comparison between our solution and existing methods for both considered tasks. Results are reported for publicly available datasets DRIONS-DB [24], RIM-ONE v.3 [4], DRISHTI-GS [15, 16], which contain groundtruth segmentation for optic disc (and some for optic cup as well). DRIONS-DB contains 110 full eye fundus images with optic disc segmentation; RIM-ONE v.3—159 images cropped by optic disc area, such that its diameter occupies about a fifth part of an image side length, with optic disc and cup segmentation; DRISHTI-GS—50 full eye fundus images with optic disc and cup segmentation. We evaluate the quality of trained algorithms by Intersection-over-

Union (IOU) score: $\frac{|A \cap B|}{|A \cup B|}$ and Dice score: $\frac{2|A \cap B|}{|A| + |B|}$,

where $A = (a_{ij})_{i=1}^H_{j=1}^W$ is a predicted output map, containing probabilities that each pixel belongs to the foreground, and $B = (b_{ij})_{i=1}^H_{j=1}^W$ is a correct binary output map. These quality measures do not depend on image scale, object scale and class imbalance. Dice score is also equal to F_1 score—harmonic mean of precision and recall.

We used a learning rate of 10^{-3} for optic disc and a learning rate of 3×10^{-4} for optic cup segmentation. Momentum was set to 0.95, mini-batch of size 1 was used in order to minimize required amount of GPU memory. Resolution of input images was set to 256×256 for optic disc and to 512×512 for optic cup segmentation before their cropping. Region of interest was then resized to 128×128 by bilinear interpolation.

For the task of optic disc segmentation, we compare our solution with the method from [8] paper (further referred as DRIU, as the name of the paper suggests), which is the best method that we have found in terms of IOU and Dice score functions for investigated datasets. For the task of optic cup segmentation, we compare with the method from [11] (further referred as BCF, as the name of the paper suggests) and from [17]. Score estimates are computed by cross-validation with 5 folds.

The presented algorithms were implemented on GPU with Python 2.7 programming language and Keras framework for training of neural networks (with Theano backend [25]). CLAHE implementation from Scikit-Image library was also used. All estimates of computational time are given for Amazon Web Services [26] g2.2xlarge instance with one NVIDIA GRID (Kepler GK104) GPU and Intel Xeon E5-2670

CPU for 256×256 images; estimate of Zilly et al. [17] method's prediction time is given for a 2.66 GHz quad-core CPU, as reported. Prediction time of BCF [11] is expected to be close to Zilly et al. [17] prediction time, since these methods are very similar.

The results of the experiments indicate that the proposed method not only demonstrates quality competitive with quality of the existing methods in a majority of score metrics, but also possesses lowest prediction time, lowest training time among deep learning solutions, has small number of parameters (whole model can be saved in a file of only 5 MB; DRIU model requires about 120 MB) and is very easy to program with the use of modern frameworks. Despite that we gave estimates of prediction time for a machine equipped with modern (though not top level) GPU, for GPU with lower performance a prediction time can be only a few times larger. These advantages make the proposed method being a good solution for automatic glaucoma assessment on mobile devices.

CONCLUSIONS

In this paper we show that our method based on modified U-Net neural network can provide results similar or better than existing methods for the tasks of optic disc and cup segmentation on eye fundus images. The same method, applied to both tasks, achieves high quality of segmentation, which proves its applicability to various problems of image recognition. Advantages of the proposed solution also include its simplicity, simple programming with the use of modern frameworks and lowest possible prediction time. Experiments results and visual comparison show that automatic optic disc segmentation can be done at the quality competitive with human. However, optic cup is more challenging to recognize, which is supported by the fact that its border is much more subtle. We believe that there is a room for improvement for optic cup segmentation, and further research is needed.

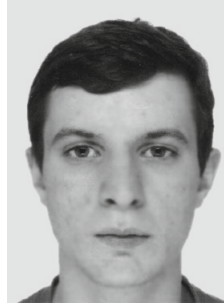
ACKNOWLEDGMENTS

We are especially grateful to A.G. D'yakonov, Professor, Dr. Sci. (Moscow State University), for supporting and supervising this work. We would like to thank L.M. Mestetskii, Professor, Dr. Tech. (Moscow State University), for initiating and supporting ophthalmological research at the department. We are also grateful to Youth Laboratories company and especially to K. Kiselev for provided computational resources.

REFERENCES

1. A. Almazroa, R. Burman, K. Raahemifar, and V. Lakshminarayanan, "Optic disc and optic cup segmentation methodologies for glaucoma image detection: a survey," *J. Ophthalmol.* **2015**, (2015).
2. H. A. Quigley and A. T. Broman, "The number of people with glaucoma worldwide in 2010 and 2020," *British J. Ophthalmol.* **90** (3), 262–267 (2006).
3. M. U. Akram, A. Tariq, S. Khalid, M. Y. Javed, S. Abbas, and U. U. Yasin, "Glaucoma detection using novel optic disc localization, hybrid feature set and classification techniques," *Australasian Phys. Eng. Sci. Med.* **38** (4), 643–655 (2015).
4. F. Fumero, S. Alayón, J. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, "Rim-one: An open retinal image database for optic nerve evaluation," in *Proc. 24th IEEE Int. Symp. on Computer-Based Medical Systems (CBMS)* (Bristol, 2011), pp. 1–6.
5. G. Lim, Y. Cheng, W. Hsu, and M. L. Lee, "Integrated optic disc and cup segmentation with deep learning," in *Proc. 27th IEEE Int. Conf. on Tools with Artificial Intelligence (ICTAI)* (Vietri sur Mare, 2015), pp. 162–169.
6. A. Bastawrous, H. K. Rono, I. A. Livingstone, H. A. Weiss, S. Jordan, H. Kuper, and M. J. Burton, "Development and validation of a smartphone-based visual acuity test (peek acuity) for clinical practice and community-based fieldwork," *JAMA Ophthalmol.* **133** (8), 930–937 (2015).
7. V. Lodhia, S. Karanja, S. Lees, and A. Bastawrous, "Acceptability, usability, and views on deployment of peek, a mobile phone mhealth intervention for eye care in Kenya: qualitative study," *JMIR mHealth uHealth* **4** (2) (2016).
8. K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, "Deep retinal image understanding," in *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2016), pp. 140–148.
9. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (Boston, MA, 2015), pp. 3431–3440.
10. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556 (2014).
11. J. G. Zilly, J. M. Buhmann, and D. Mahapatra, "Boosting convolutional filters with entropy sampling for optic cup and disc image segmentation from fundus images," in *Proc. Int. Workshop on Machine Learning in Medical Imaging* (Springer, 2015), pp. 136–143.
12. R. Gonzalez, R. Woods, and S. Eddins, *Digital Image Processing Using MATLAB* (Prentice-Hall, Upper Saddle River, NJ, 2003).
13. H. Dögan and O. Akay, "Using adaboost classifiers in a hierarchical framework for classifying surface images of marble slabs," *Expert Syst. Appl.* **37** (12), 8814–8821 (2010).
14. M. B. Salah, A. Mitiche, and I. B. Ayed, "Multiregion image segmentation by parametric kernel graph cuts," *IEEE Trans. Image Processing* **20** (2), 545–557 (2011).
15. J. Sivaswamy, S. Krishnadas, A. Chakravarty, G. Joshi, A. S. Tabish, et al., "A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis," *JSM Biomed. Imaging Data Pap.* **2** (1) (2015).
16. J. Sivaswamy, S. Krishnadas, G. D. Joshi, M. Jain, and A. U. S. Tabish, "Drishti-gs: retinal image dataset for optic nerve head (onh) segmentation," in *Proc. 11th*

- IEEE Int. Symp. on Biomedical Imaging (ISBI)* (Beijing, 2014), pp. 53–56.
17. J. Zilly, J. M. Buhmann, and D. Mahapatra, “Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation,” *Comput. Med. Imag. Graph.* **55**, 28–41 (2017).
 18. H. Li and O. Chutatape, “Automated feature extraction in color retinal images by a model based approach,” *IEEE Trans. Biomed. Eng.* **51** (2), 246–254 (2004).
 19. J. Jose and J. Kuruvilla, “Detection of red lesions and hard exudates in color fundus images,” *Int. J. Eng. Comput. Sci.* **3** (10), 8583–8588 (2014).
 20. R. Szeliski, *Computer Vision: Algorithms and Applications* (Springer Science and Business Media, 2010).
 21. O. Ronneberger, P. Fischer, and T. Brox, “U-net: convolutional networks for biomedical image segmentation,” in *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2015), pp. 234–241.
 22. G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” arXiv:1207.0580 (2012).
 23. I. Sutskever, J. Martens, G. E. Dahl, and G. E. Hinton, “On the importance of initialization and momentum in deep learning,” in *Proc. Int. Conf. on Machine Learning* (Atlanta, 2013).
 24. E. J. Carmona, M. Rincón, J. García-Feijoó, and J. M. Martínez-de-la Casa, “Identification of the optic nerve head with genetic algorithms,” *Artificial Intellig. Med.* **43** (3), 243–259 (2008).
 25. Theano Development Team, “Theano: A Python framework for fast computation of mathematical expressions,” arXiv 1605.02688 (May 2016).
 26. Amazon web services. <https://aws.amazon.com>



Artem Sevastopolsky (born in 1996) received his Bachelor degree from Lomonosov Moscow State University, faculty of Computational Mathematics and Cybernetics, department of Mathematical Methods of Forecasting. His research interests include machine learning, computer vision, deep learning, image and video processing.