Deep Retinal Image Understanding

Kevis-Kokitsi Maninis[1], Jordi Pont-Tuset[1], Pablo Arbeláez[2], and Luc Van Gool[1,3]

[1]ETH Zürich [2]Universidad de los Andes [3]KU Leuven

Abstract. This paper presents Deep Retinal Image Understanding (DRIU), a unified framework of retinal image analysis that provides both retinal vessel and optic disc segmentation. We make use of deep Convolutional Neural Networks (CNNs), which have proven revolutionary in other fields of computer vision such as object detection and image classification, and we bring their power to the study of eye fundus images. DRIU uses a base network architecture on which two set of specialized layers are trained to solve both the retinal vessel and optic disc segmentation. We present experimental validation, both qualitative and quantitative, in four public datasets for these tasks. In all of them, DRIU presents super-human performance, that is, it shows results more consistent with a gold standard than a second human annotator used as control.

Keywords: Retinal vessel segmentation, optic disc segmentation, deep learning, convolutional neural networks, retinal image understanding

1 Introduction

Retinal image understanding is key for ophthalmologists while assessing widely spread eye diseases such as glaucoma, diabetic retinopathy, macular degeneration, and hypertension, among others. Although these diseases can lead to severe visual impairment and blindness if left untreated, early diagnosis and appropriate treatment, coupled with periodic examination by specialists, have proven to be determinant factors for controlling their evolution, which translates into better prognosis and an improved quality of life for patients. Given that several risk factors associated to these diseases, such as sedentarism, obesity and aging, are related to lifestyle, their frequency among the general population is growing. This situation has stressed the need for automated methods to assist ophthalmologists in retinal image understanding, and has sparked the interest for this field among the medical image analysis community.

Two anatomical structures are of particular interest for specialists when performing diagnostic measurements on eye fundus images: the blood vessel network and the optic disc. Consequently, most prior work on automated analysis of retinal images has focused on the segmentation of these two structures. Classic methods for addressing the task of blood vessel segmentation involve hand crafted filters like line detectors [17,14] and vessel enhancement techniques [20,26,5]. Approaches that rely on powerful machine learning techniques have emerged over the last years. In [15] the authors combine different kinds of vessel features and segment them with fully connected conditional random fields. In [1] a gradient boosting framework is proposed for learning filters in a supervised way. Algorithms that are able to enhance fine structures given a regression of retinal vessels have also been developed recently [19,9]. Prior work on optic

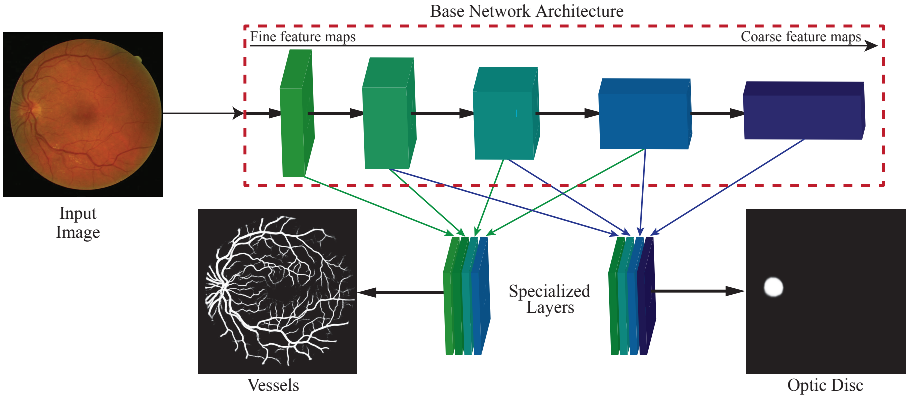

Fig. 1. Overview of DRIU: Given a base CNN, we extract side feature maps and design specialized layers to perform blood vessel segmentation (left) and optic disc segmentation (right).

disc segmentation includes morphological operators [23] and hand crafted features [25] Morales *et al.* [13] use morphology along with Principal Component Analysis (PCA) to obtain the structure of the optic disk inside a retinal image. A superpixel classification method is proposed in [3].

In the last five years, deep learning techniques have revolutionized the field of computer vision. Deep Convolutional Neural Network (CNN) architectures were initially designed for the task of natural image classification [12], and recent research has led to impressive progress in solving that problem. At the core of these approaches, lies a *base network* architecture, starting from the seminal AlexNet [12], to the more complex and more accurate VGGNet [18] and the inception architecture of GoogLeNet [22]. Furthermore, CNNs have been applied successfully to a large variety of general recognition tasks such as object detection [8], semantic segmentation [10], and contour detection [24]. In the domain of retinal image understanding, CNNs have been used for retinal vessel segmentation in [7] to classify patch features into different vessel classes. For optic disc segmentation, the authors of [27] use a CNN to extract features, which they post-process to obtain binary segmentations. Instead, our work is based on applying a CNN end-to-end, both for retinal vessel segmentation and optic disc detection, efficiently, since we avoid the redundant computations from a patch-based approach.

In this paper, we leverage the latest progress on deep learning to advance the field of automated retinal image interpretation. We design a CNN architecture that specializes a base network for the tasks of segmenting blood vessels and optic discs in fundus images. An overview of our approach, which we call Deep Retinal Image Understanding (DRIU), is presented in Figure 1. DRIU is both highly efficient and highly accurate: at inference time, it requires a single forward pass of the CNN to segment both the vessel network and the optic disc and, as the experimental results will show, DRIU reaches or surpasses the performance of trained human specialists for both tasks on four publicly available annotated datasets.

2 CNNs for Retinal Image Understanding

We approach retinal vessel segmentation and optic disc detection as an image-to-image regression task, by designing a novel CNN architecture. We start from the VGG [18] network, originally designed for large-scale natural image classification. For our purposes, the fully connected layers at the end of the network are removed, so it mainly consists of convolutional layers coupled with Rectified Linear Unit (ReLU) activations. The use of four max pooling layers in the architecture separates the network into five stages (as in Figure 1), each stage consisting of several convolutional layers. Between the pooling layers, feature maps of the same stage that are generated by convolutions with different filters have the same size. As we proceed deeper in the network, the information becomes coarser due to the decrease in size, which is a key factor for generalization. We call this part of our architecture the "base network". The layers of the base network are already pre-trained on millions of images, which is necessary for training deep architectures. To effectively use the information from feature maps with different sizes, we draw inspiration from the "inception" architecture of GoogLeNet [22], which adds supervision at multiple internal layers of the network, and we connect task-specific "specialized" convolutional layers to the final layer of each stage. Each specialized layer produces feature maps in K different channels, which are resized to the original image size and concatenated, creating a volume of fine-to-coarse feature maps. We append one last convolutional layer which linearly combines the feature maps from the volume created by the specialized layers into a regressed result. In our experiments, we used $K = 16$. The majority of convolutional layers employ 3×3 convolutional filters for efficiency, except the ones used for linearly combining the outputs (1×1 filters).

For training the network, we adopt the class-balancing cross entropy loss function originally proposed in [24] for the task of contour detection in natural images. We denote the training dataset by $S = \{(X_n, Y_n), n = 1, ..., N\}$, with X_n being the input image and $Y_n = \{y_j^{(n)}, j = 1, ..., |X_n|\}, y_j^{(n)} \in \{0, 1\}$ the predicted pixel-wise labels. For simplicity, we drop the subscript n. The loss function is then defined as:

$$\mathcal{L}(\mathbf{W}) = -\beta \sum_{j \in Y_+} \log P(y_j = 1|X; \mathbf{W}) - (1 - \beta) \sum_{j \in Y_-} \log P(y_j = 0|X; \mathbf{W}) \qquad (1)$$

where $\mathbf{W}$ denotes the standard set of parameters of the CNN, which are trained with backpropagation. The multiplier β is used to handle the imbalance of the substantially greater number of background compared to foreground pixels, which in our case are the vessels or the optic disc. Class balancing is necessary when we have severely biased ground truths (e.g.: approximately 10% of the pixels are vessels, while the others are background). Y_+ and Y_- denote the foreground and background sets of the ground truth Y, respectively. In this case, we use $\beta = |Y_-|/|Y|$. The probability $P(.)$ is obtained by applying a sigmoid $\sigma(.)$ to the activation of the final convolutional layer.

We use the same network architecture for both retinal vessel segmentation and optic disc segmentation. We found that the coarse feature maps of the final stage do not help with vessel detection since they contain coarse information which erases the thin vessels, whereas the finest ones of the first stage do not help with detecting the coarse structure of the optic disc. Thus, we construct two separate feature map volumes, one for each task. For retinal vessel segmentation, the volume contains features from the 4 finer stages, while for optic disc detection we use the 4 coarser stages (see Figure 1). Our final result for both tasks is a probability map, in which a pixel detected as vessel/disc is assigned a higher score.

At training time, we fine-tune the entire architecture (base network and specialized layers) for 20000 iterations. We use stochastic gradient descent with momentum, operating on one image per iteration. Due to the lack of data, the learning rate is set to a very small number ($lr = 10^{-8}$), which is gradually decreased as the training process proceeds. We augment the datasets using standard techniques, by rotating and scaling the images, as a pre-processing step. We also substract the mean value of the training images for each colour channel.

At testing time, there is no need for any pre-processing of the data. The entire architecture runs on a GPU, and operates on the original RGB channels of a retinal image. The average execution time for retinal vessel and optic disc segmentation on an NVIDIA TITAN-X GPU is 85 milliseconds (ms) for DRIVE and 104 ms for STARE, which is orders of magnitude faster than the current state-of-the-art [15,1,7]. The same applies for optic disc segmentation, where our algorithm processes an image of DRIONS-DB in 65 ms and 110 ms for the larger images of RIM-ONE dataset.

3 Experimental Validation

We experiment on eye fundus images to segment both the blood vessels and the optic disc . In both cases, and for each database, we split the data into separate *training* and *test* sets, we learn all the parameters of DRIU on the training set, and then evaluate the final model on the previously unseen test set. Since CNNs are high-capacity models, it is a standard practice to augment the training set by rotating and scaling the images, in order to avoid learning regularities caused by the limited size of the dataset, such as the location of the optic disc in the training images, which would lead to overfitting and poor generalization. Additionally, we keep the learning rate of the network very low (fine-tuning with $lr = 10^{-8}$).

Blood Vessel Segmentation: We experiment on the DRIVE [21] and STARE [11] datasets (40 and 20 images, respectively). Both contain manual segmentations of the blood vessels by two expert annotators. We use the segmentations of the first annotator as the gold standard to train/test our algorithm. The ones from the second are evaluated against the gold standard to measure human performance. For DRIVE we use the standard train/test split and for STARE we use the split defined in [9], according to which the first 10 images consist the training set and the last 10 the test set. We compare DRIU with the current state-of-the-art [7,1,15] as well as some traditional approaches for retinal vessel segmentation [17,20]. We also retrain HED [24] (state of the art in generic contour detection with CNNs) on DRIVE and STARE using their public code.

We compare the techniques by binarizing the soft-map result images at multiple confidence values and computing the pixel-wise precision-recall between the obtained mask and the gold standard, resulting in one curve per technique. Figure 2 shows both qualitative and quantitative results. As a summary measure (in brackets in the legend) we compute the Dice coefficient or F1-measure (equivalent [16] to the Jaccard index) of the optimal point (marked in the lines).

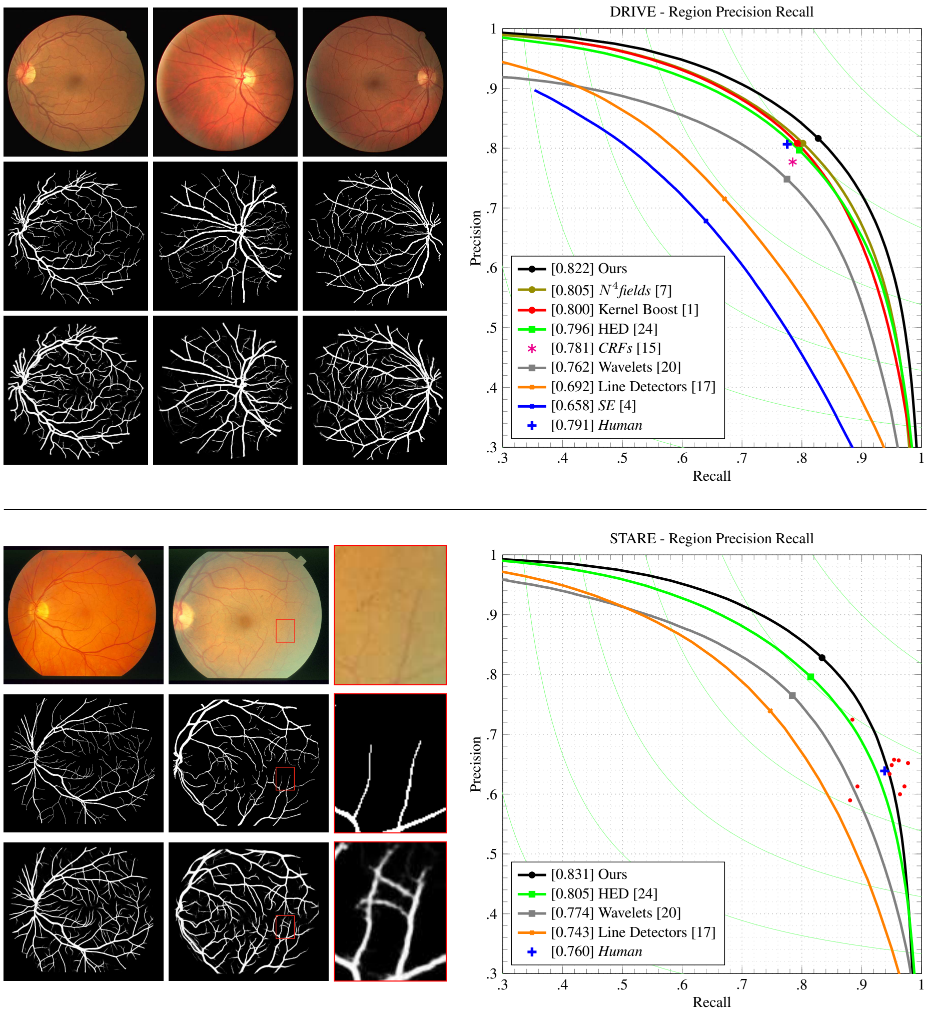

Fig. 2. Vessel Segmentation on DRIVE (top) and STARE (bottom). **Left**: top row, eye fundus images; middle row, human expert annotations; bottom row, results obtained by our method. **Right**: Region precision recall curves, methods in italics refer to pre-evaluated results. The red dots indicate the performance of the second human annotator on each image of the test set.

The results show that DRIU performs better than all methods of comparison on both datasets, in all operating regimes. Also in both datasets, DRIU provides more consistent detections with respect to the gold standard than the second human expert (better F measure). Interestingly, taking a closer look at some false positives of our technique (see Figure 2 red rectangles on the bottom half), we observe that, although very weak, there are actually two vessels where DRIU signals vessels, although the human annotator did not notice them.

Optic Disc Segmentation: We experiment on the DRIONS-DB [2] and RIM-ONE (r3) [6] datasets (110 and 159 images, respectively). Both contain manual segmentations of the optic disc by two expert annotators. As for the vessels, we use the segmentations of the first annotator to train/test our algorithm. We split into training and testing sets (60/50 and 99/60, respectively). Given the nature of the results, apart from the region precision-recall curves, we also measure the boundary error as the mean distance between the boundary of the result and that of the ground truth.

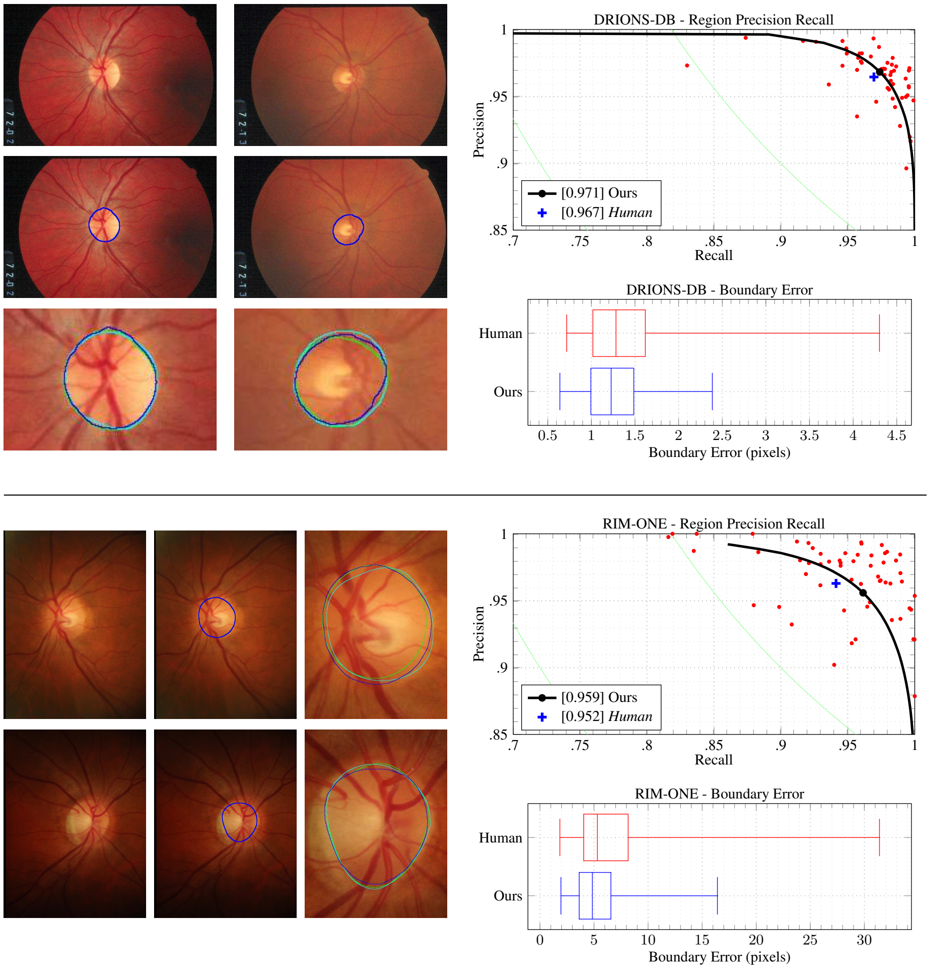

Fig. 3. Optic Disc Segmentation on DRIONS-DB (top) and RIM-ONE (bottom). **Left**: top row, eye fundus images; middle row, our segmentation; bottom row, detail of our segmentation (blue) against the two human annotations (green and cyan). **Right**: Region precision-recall curves (top) and boundary error (bottom). The red dots on the curve indicate the performance of the second human annotator on each image of the test set.

Figure 3 shows the results of the qualitative and quantitative evaluation performed on these two datasets. Focusing first on the region precision-recall curves, DRIU shows super-human performance, meaning that it presents results more coherent to the gold standard than the second human annotator.

In terms of boundary accuracy, the box-plots show the distribution of errors (limits of the 4 quartiles). In both datasets DRIU presents results that not only have a lower median error, but also show less dispersion, so more consistency. The qualitative results corroborate that DRIU is robust and consistent, even in the more challenging and diverse scenarios of RIM-ONE database.

4 Conclusions and Discussion

We presented DRIU, a method for retinal vessel and optic disc segmentation that is both fast and accurate. DRIU brings the power of CNNs, which have proven groundbreaking in other fields of computer vision, to retinal image analysis by the use of a base shared CNN network and per-task specialized layers. The experimental validation in four public datasets, both qualitative and quantitative, shows that DRIU has super-human performance in these tasks[1].

The impact of an automated solution to the problems of vessel and optic disc segmentation goes beyond assisting specialists in the initial diagnosis of eye diseases. Accurate and repeatable measurements can be an invaluable tool for monitoring their evolution. Looking forward, our technology also has the potential of changing medical practice, as it offers the possibility of carrying out robust comparative statistical analyses on large populations.

Acknowledgements: Research funded by the EU Framework Programme for Research and Innovation - Horizon 2020 - Grant Agreement No. 645331 - EurEyeCase. We thank NVIDIA Corporation for donating the GPUs used in this project.

References

1. Becker, C., Rigamonti, R., Lepetit, V., Fua, P.: Supervised feature learning for curvilinear structure segmentation. In: MICCAI (2013)
2. Carmona, E.J., Rincón, M., García-Feijoó, J., Martínez-de-la Casa, J.M.: Identification of the optic nerve head with genetic algorithms. AIIM 43(3), 243–259 (2008)
3. 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 T-MI 32(6), 1019–1032 (2013)
4. Dollár, P., Zitnick, C.L.: Structured forests for fast edge detection. In: ICCV (2013)
5. Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A.: Blood vessel segmentation methodologies in retinal images–a survey. Computer methods and programs in biomedicine 108(1), 407–433 (2012)

[1] All the resources of this paper, including code and pre-trained models to reproduce the results, are available at: `http://www.vision.ee.ethz.ch/~cvlsegmentation/`

6. Fumero, F., Alayón, S., Sanchez, J., Sigut, J., Gonzalez-Hernandez, M.: Rim-one: An open retinal image database for optic nerve evaluation. In: CBMS. pp. 1–6 (2011)
7. Ganin, Y., Lempitsky, V.: N^4-fields: Neural network nearest neighbor fields for image transforms. In: ACCV (2014)
8. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Region-based convolutional networks for accurate object detection and segmentation. IEEE T-PAMI 38(1), 142–158 (2016)
9. Gu, L., Cheng, L.: Learning to boost filamentary structure segmentation. In: ICCV (2015)
10. Hariharan, B., Arbeláez, P., Girshick, R., Malik, J.: Hypercolumns for object segmentation and fine-grained localization. In: CVPR (2015)
11. Hoover, A., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE T-MI 19(3), 203–210 (2000)
12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS (2012)
13. Morales, S., Naranjo, V., Angulo, J., Alcañiz, M.: Automatic detection of optic disc based on pca and mathematical morphology. IEEE T-MI 32(4), 786–796 (2013)
14. Nguyen, U.T., Bhuiyan, A., Park, L.A., Ramamohanarao, K.: An effective retinal blood vessel segmentation method using multi-scale line detection. PR 46(3), 703–715 (2013)
15. Orlando, J.I., Blaschko, M.: Learning fully-connected crfs for blood vessel segmentation in retinal images. In: MICCAI (2014)
16. Pont-Tuset, J., Marques, F.: Supervised evaluation of image segmentation and object proposal techniques. IEEE T-PAMI (2015)
17. Ricci, E., Perfetti, R.: Retinal blood vessel segmentation using line operators and support vector classification. IEEE T-MI 26(10), 1357–1365 (2007)
18. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. ICLR (2015)
19. Sironi, A., Lepetit, V., Fua, P.: Projection onto the manifold of elongated structures for accurate extraction. In: ICCV (2015)
20. Soares, J.V., Leandro, J.J., Cesar Jr, R.M., Jelinek, H.F., Cree, M.J.: Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. IEEE T-MI 25(9), 1214–1222 (2006)
21. Staal, J., Abràmoff, M.D., Niemeijer, M., Viergever, M., Van Ginneken, B., et al.: Ridge-based vessel segmentation in color images of the retina. IEEE T-MI 23(4), 501–509 (2004)
22. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: CVPR (2015)
23. 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: Medical Data Analysis, pp. 282–287. Springer (2001)
24. Xie, S., Tu, Z.: Holistically-nested edge detection. In: ICCV (2015)
25. 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 T-MI 27(1), 11–18 (2008)
26. Zhang, B., Zhang, L., Zhang, L., Karray, F.: Retinal vessel extraction by matched filter with first-order derivative of gaussian. Computers in Biology and Medicine 40(4), 438–445 (2010)
27. Zilly, J.G., Buhmann, J.M., Mahapatra, D.: Boosting convolutional filters with entropy sampling for optic cup and disc image segmentation from fundus images. In: Machine Learning in Medical Imaging, pp. 136–143. Springer (2015)