






# A Single-Resolution Fully Convolutional Network for Retinal Vessel Segmentation in Raw Fundus Images

Ricardo J. Araújo<sup>1,2</sup>(✉) , Jaime S. Cardoso<sup>1,3</sup> , and Hélder P. Oliveira<sup>1,2</sup> 

<sup>1</sup> INESC TEC, Porto, Portugal

`ricardo.j.araujo@inesctec.pt`

<sup>2</sup> Faculdade de Ciências da Universidade do Porto, Porto, Portugal

<sup>3</sup> Faculdade de Engenharia da Universidade do Porto, Porto, Portugal

**Abstract.** The segmentation of retinal vessels in fundus images has been heavily focused in the past years, given their relevance in the diagnosis of several health conditions. Even though the recent advent of deep learning allowed to foster the performance of computer-based algorithms in this task, further improvement concerning the detection of vessels while suppressing background noise has clinical significance. Moreover, the best performing state-of-the-art methodologies conduct patch-based predictions. This, put together with the preprocessing techniques used in those methodologies, may hinder their use in screening scenarios. Thus, in this paper, we explore a fully convolutional setting that takes raw fundus images and allows to combine patch-based training with global image prediction. Our experiments on the DRIVE, STARE and CHASEDB1 databases show that the proposed methodology achieves state-of-the-art performance in the first and the last, allowing at the same time much faster segmentation of new images.

**Keywords:** Retina · Vessel segmentation · Deep learning

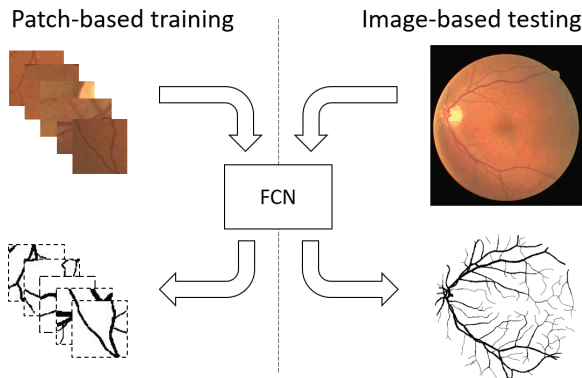
## 1 Introduction

The retina is a tissue layer in the eye of vertebrates that participates in the production of nerve impulses that go to the visual cortex of the brain. Its vascularization is easily assessed in a non-intrusive manner by photography-based mechanisms, such that fundus imaging is often used as a diagnostic means of medical conditions affecting the morphology of vessels, such as hypertension, diabetes, arteriosclerosis, and cardiovascular disease [6]. It has been reported that 10% of all the diabetic patients have diabetic retinopathy, the main cause of blindness among people in the Western civilizations. Therefore, an early treatment is essential, and given that manual analysis by experts is very time consuming, automated vessel analysis is crucial for inclusion in screening programs [8].

This clinical relevance lead to the emergence of a large number of both unsupervised and supervised methodologies. Unsupervised approaches started

to appear before the advent of public databases and use theory from, one or a combination of, matched filters, vessel tracing, mathematical morphology, and scale-space representation [1, 15, 16]. Works resorting to supervised learning use manual annotations and different learning algorithms to find proper mapping functions between hand-crafted features and target segmentation [3, 12]. The advent of deep learning further improved the performance of retinal vessel segmentation. Even though this approach is heavily dependent on labeled data and available databases contain at most dozens of images, researchers resort to dividing retinal images into small patches and transform the problem into a patch classification one [11]. However, this has implications at prediction time, as a patch has to be extracted for each pixel, leading to increased computational costs. This, associated with image preprocessing, which is also commonly conducted, may hinder the use of such systems in scenarios where a large number of images needs to be analyzed on the spot, as is the case of screening programs.

In this paper, we propose a Fully Convolutional Network (FCN) design that is able to segment an unseen image at a single step, even if it is trained in a patch-wise fashion (see Fig. 1).



**Fig. 1.** Fully convolutional networks take images of arbitrary size, allowing to combine patch-based training and image-based prediction.

In practice, an adequate preprocessing facilitates the learning process, even though theory supports that a high number of non-linearities is able to adapt to the structure of data. Thus, in our experiments, we use raw color fundus images, to understand if this network is able to improve the state-of-the-art concerning vessel detection and background noise suppression, and simultaneously keep the prediction process as simple as possible. A FCN was proposed in the past [2], however its performance is significantly inferior to the best performing methods, indicating that other specific network design options may not have been ideal for retinal vessel segmentation.

## 1.1 Main Contributions

The main contributions of this work are:

- A neural network design allowing fast predictions on new data, which is crucial in all applications with high throughput of data, as is the case of screening programs;
- A methodology achieving high performance even being applied to raw fundus images, thus avoiding the need of using expensive preprocessing methods for image normalization.

## 1.2 Document Structure

This Section summarized the relevance and previous work regarding the topic of vessel segmentation in retinal fundus images, and the main contributions of our work; in Sect. 2, we discuss in detail the different options we took for designing the proposed model; in Sect. 3 we briefly describe the datasets used to assess the performance of our methodology, we introduce the conducted experiments and discuss the results; finally, Sect. 4 concludes the work and discusses possible directions for future research.

# 2 Methodology

Here, we discuss the motivations and preliminary empiric findings that led us into designing a fully convolutional network adapted to the specific task of vessel segmentation in raw color fundus images.

## 2.1 Fully Convolutional Network for Vessel Segmentation

Convolutional neural networks (CNNs) have revolutionized the field of computer vision, given their combination of deep hierarchical feature extraction (sequence of convolutional layers) and classification (fully connected layers) blocks. This was the type of deep neural network used in [11], where very small patches of the retina were fed into the model and it outputted the probability of the center pixel being a vessel. This highlights one of the problems of using typical CNNs for segmenting vessels, which is the need to divide a given image into a very large number of small patches and classify each of them, yielding a tremendous computational cost. A second problem is that fully connected layers force all the input images to have the same size.

A FCN design is a more adequate choice for segmentation problems, since it does not use fully connected layers. Thus, it is not mandatory to divide an image in order to obtain a complete segmentation map, which is crucial whenever we require fast predictions, as is the case of retinal screening programs, where a high volume of data is quickly generated. The inputs may also have varying size, making this design much more adaptable to different imaging conditions.

It allows us to train on smaller patches of the images and later still be able to obtain single-pass predictions of the entire images, as is represented in Fig. 1. Note that performing patch-wise training is an engineering option which facilitates avoiding wasting computational effort with portions of the images that do not contain information of the retina fundus.

## 2.2 Specific Design Considerations

After motivating the use of a FCN design for the segmentation of retinal vessels, now we delve into more specific aspects of the proposed network architecture, discussing some options we took based in previous works and empirical findings.

**Spatial Resolution.** Pooling or strided convolutions are commonly used to induce higher-level features to encode more neighborhood information. Recent results [11] suggest that pooling operations seem to not improve the performance of networks that are trained in small images. In preliminary experiments, we found that indeed a single-resolution deep network was more capable than a Unet-like model when extracting small capillaries. Even though the latter is able to combine low- and high-scale features, it seems that a deeper network at a fine scale is able to obtain better representations of small structures of interest, as is the case of very small vessels. Thus, in this work, the image resolution was kept across the entire network, contrarily to the previously proposed FCN [2].

**Activation Units.** All intermediate non-linearities were given by a Leaky Rectified Linear Unit (Leaky ReLU):

$$f(x) = \begin{cases} x & \text{if } x > 0, \\ ax & \text{otherwise} \end{cases} \quad (1)$$

where  $x$  represents the outcome of the previous convolution and  $a$  was set to 0.2. It was used over a ReLU just to allow the network to learn even for negative inputs. In the last layer, we used a Sigmoid activation unit, since we are dealing with a pixel-wise binary problem.

**Batch Normalization.** Whenever the statistics at test time differ from the ones found during training, batch normalization becomes problematic. In fact, this is the case when a model is trained in small retinal patches and at test time is applied to entire retinal images, whose statistics will be inevitably different. In preliminary experiments, we found that using batch normalization was indeed hurting the performance of the models, thus it was not considered in the final design.

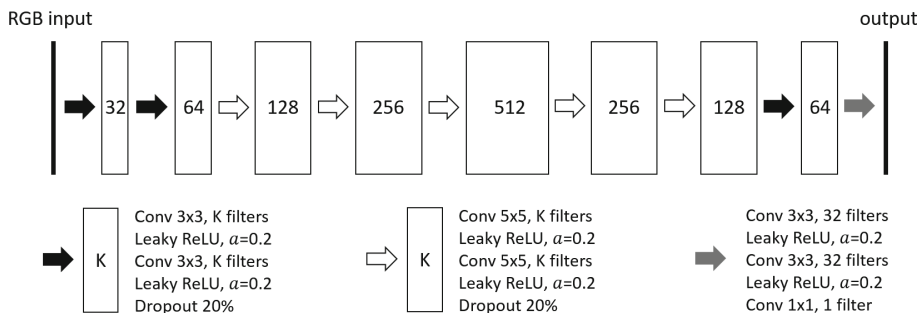
**Dropout.** Turning off some computational connections along the network was useful to create more redundancies and thus obtain more robust models. We found it was also useful to apply dropout at the initial levels of the model, in order to add some noise to the initial representations.

**Loss Function.** Neural networks targeting binary segmentation problems usually minimize the Binary Cross Entropy (BCE) loss, a pixel-wise criterium that exponentially increases as the network becomes more confident when committing a mistake. Note however, that this loss is agnostic to class imbalance, thus it naturally biases models to be more confident identifying the most common class, which in our case, is the background. We are interested in alleviating this effect, in order to obtain models with good sensitivity and that do not simply ignore narrow vessels. Weighting differently each class is an option we consider for reaching fairer models. Furthermore, we used the recently proposed focal loss [10], an extension to the BCE loss that puts more focus in the misclassified examples:

$$FL(p) = -\left(y \cdot \alpha(1-p)^\gamma \cdot \log(p) + (1-y) \cdot (1-\alpha) \cdot p^\gamma \cdot \log(1-p)\right) \quad (2)$$

where  $p \in [0, 1]$  is the probability of class 1 (vessel) outputted by the network,  $y \in \{0, 1\}$  is the binary target variable,  $\gamma \geq 0$  is a focusing parameter, and  $\alpha \in [0, 1]$  allows to give more weight to samples of a certain class.  $\gamma$  was set to 2 in this work. Even though the focal loss by itself is also agnostic to class imbalance, by performing hard training, it helps inducing the model to not ignore the potential hardest cases, such as small capillaries.

After all these considerations, architecture and hyper-parameter tuning was conducted (see Sect. 3.3). The final design we considered for segmenting vessels from raw color fundus images is represented in Fig. 2.



**Fig. 2.** Single-resolution fully convolutional network used in this work for segmenting vessels in raw fundus images.

### 3 Experiments and Results

The datasets and metrics used to assess the performance of our model will be briefly described here, then we provide details regarding how hyper-parameter tuning was conducted to obtain the final neural network design, and finally, we present and discuss the achieved results.

### 3.1 Datasets

Several public benchmarks of retinal vessel segmentation are available. In this paper, we conducted experiments in three of the most commonly used datasets among the literature works, which are the DRIVE [14], STARE [5], and CHASEDB1 [13] datasets.

The DRIVE database results from a diabetic retinopathy screening program in The Netherlands. Among the collected images, 40 photographs have been randomly selected, 7 of which showing signs of early diabetic retinopathy. The images were acquired using a Canon CR5 non-mydratiac 3CCD camera with a 45 degree field of view and later digitized to  $584 \times 565$  pixels.

The STARE database comprises 20 retinal images captured by a TopCon TRV-50 fundus camera and digitized to  $605 \times 700$  pixels. Half of the images are pathological.

Finally, the CHASEDB1 dataset includes retinal images of children from the Child Heart and Health Study in England. 28 fundus images of size  $960 \times 999$  are available, with the particularity that central vessel reflex is abundant.

### 3.2 Model Evaluation

To evaluate how well a map of vessel probabilities fits the ground truth, we calculated the metrics that are commonly used in this task, which are accuracy, sensitivity, and specificity:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

where TP, FP, FN, and TN are the true positive, false positive, false negative, and true negative detections. A limitation of these metrics is that they are evaluated at a threshold of 0.5. Thus, we also considered the commonly used area under the receiver-operator curve (AUC), which seems more ideal for this task, as it better depicts how well a method separates both classes.

### 3.3 Implementation Details

The architecture and hyper-parameters were tuned by randomly picking three images from DRIVE’s training set for validation purposes and using the remaining ones to train varying model configurations, according to the considerations detailed in Sect. 2. Color images were solely normalized to the range  $[0, 1]$ . At each training epoch, 500 batches of  $N$  patches of size  $M \times M$  were fed to the network. Patches were randomly extracted from images at valid positions, where valid means the center pixel belongs to the retinal fundus. Data augmentation

was conducted via random transformations including vertical or horizontal flipping, and rotations in the range  $[-\pi/2, \pi/2]$ . We used the Adam optimizer with the parameters as provided in the original work [7], with the exception of the learning rate, which was initialized to  $1e-4$  and decreased to half every time the validation loss did not decrease for 10 epochs. A loss decrease was only considered if it surpassed the threshold of  $1e-4$ . Early stopping occurred if there were 30 epochs without improvement. Our preliminary experiments achieved best performance in the validation set using the network design present in Fig. 2, and for  $N = 16$  and  $M = 64$ , even though these hyper-parameters did not have a significant impact in the performance of the model.

We trained our final FCN design for 30 epochs. Starting from epoch 10, we performed linear learning rate decay by multiplying it by a constant of 0.75, and after epoch 20 the constant was changed to 0.5. Concerning DRIVE, we trained the network in the 20 images of the training set and evaluated it in the 20 images comprising the test set. Regarding STARE and CHASEDB1, datasets with few images and where a prior division does not exist, we followed the same approach of other researchers [11], which resorted to the *leave-one-out* validation.

### 3.4 Results and Discussion

The results obtained by conducting the described methodology in the referred databases are present in Table 1, along with the performance of state-of-the-art approaches. It is important to notice that the method of Azzopardi et al. [1], where a Combination of Shifted Filter Responses is used to enhance bar-like structures, belongs to the unsupervised category. Additionally, the work of Fraz et al. [3] uses traditional machine learning, where decision trees are ensembled to predict vessel probability from hand-designed features related with orientation and contrast. The rest of the methods included use deep learning techniques. Dasgupta and Singh [2] introduce a FCN design that takes preprocessed images, Fu et al. [4] couple a CNN with a Conditional Random Field to better model long-range interactions, Li et al. [9] perform patch-based segmentation using 3 fully connected layers having 400 neurons each and conduct pre-training by means of an autoencoder, and, finally, Liskowski and Krawiec [11] propose different variants of CNNs for conducting patch-based classification.

The analysis of the results shows that our FCN design is able to combine efficiency and strong predictive capabilities, even when using raw fundus images. By comparing the AUC of the methodologies, it is possible to conclude that the proposed methodology achieved superior performance in the DRIVE and CHASEDB1 databases. We believe that the performance in the STARE database was hindered due to the high variability of the raw color information among the images. This may indicate that preprocessing techniques leading to more uniform images are relevant in this dataset. Regarding DRIVE, we also tested  $\alpha = 0.6$  (give more weight to the vessel class) to better show the compromises we can get between sensitivity and specificity. The results show that we were capable of reaching better compromises in terms of vessel detection and noise suppression in this dataset, as for similar specificity we achieved higher sensitivity than the

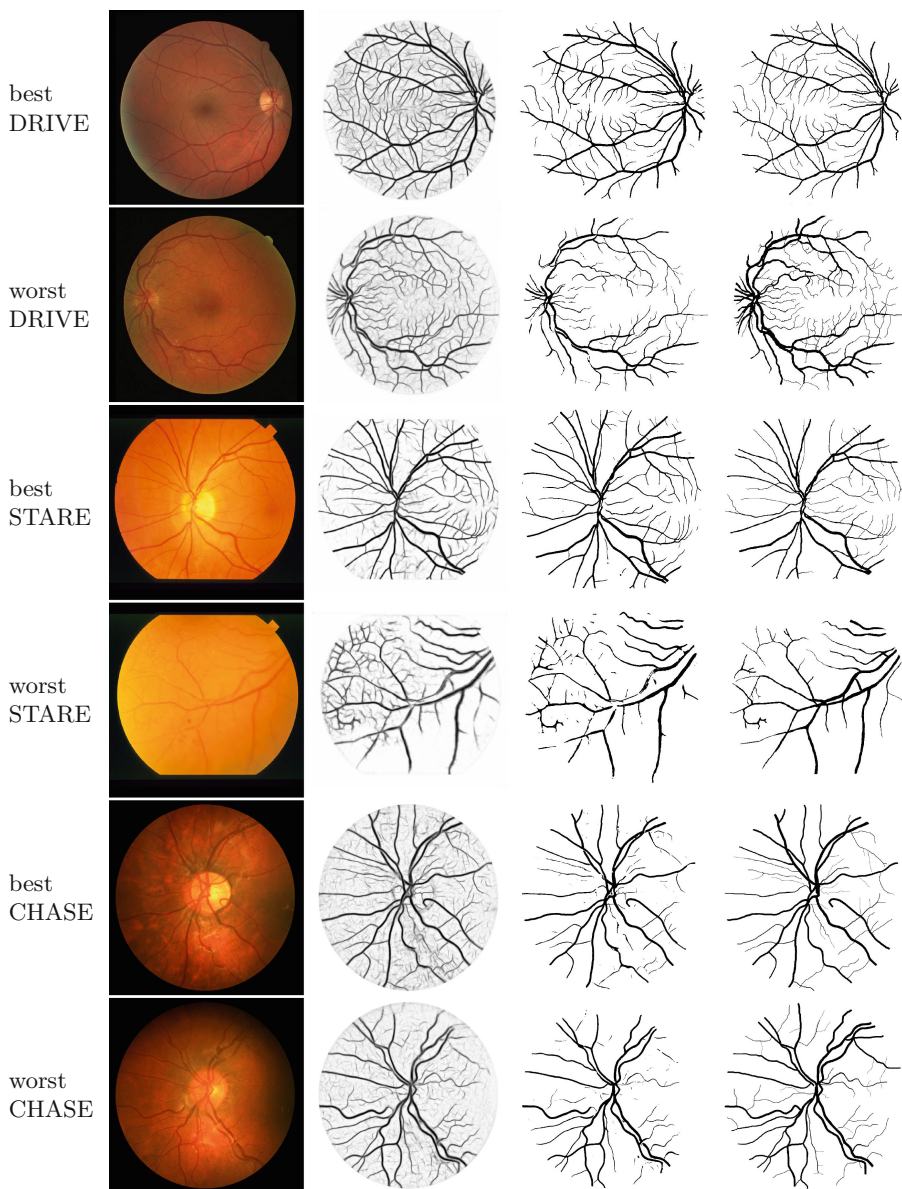
**Table 1.** Performance of the proposed methodology and state-of-the-art approaches in the DRIVE, STARE, and CHASEDB1 databases. Accuracy, sensitivity and specificity are abbreviated as acc, sen, and spe, respectively.

Method	DRIVE				STARE				CHASEDB1			
	AUC	acc	sen	spe	AUC	acc	sen	spe	AUC	acc	sen	spe
Azzopardi et al. [1]	96.1	94.4	76.6	98.1	95.6	95.0	77.2	97.0	94.9	93.9	75.8	95.9
Dasgupta and Singh [2]	97.4	95.3	76.9	98.0	–	–	–	–	–	–	–	–
Fraz et al. [3]	97.5	94.8	74.1	98.1	97.7	95.3	75.5	97.6	97.1	94.7	72.2	97.1
Fu et al. [4]	–	95.2	76.0	–	–	95.8	74.1	–	–	94.9	71.3	–
Li et al. [9]	97.4	95.3	75.7	98.2	98.8	96.3	77.3	98.4	97.2	95.8	75.1	97.9
Liskowski and Krawiec [11]												
balanc.-SP, $s = 3$	97.9	95.1	84.6	96.7	<b>99.3</b>	96.7	92.9	97.1	98.2	94.4	91.6	94.7
balanc.-SP, $s = 5$	97.9	95.3	81.5	97.5	<b>99.3</b>	97.0	90.8	97.7	98.4	95.8	87.9	96.7
no-pool-SP, $s = 5$	97.9	95.4	78.1	98.1	<b>99.3</b>	<b>97.3</b>	85.5	98.6	98.2	96.3	78.2	98.4
Proposed												
$\alpha = 0.5$	<b>98.2</b>	<b>95.6</b>	80.3	97.9	98.7	96.5	82.9	98.0	<b>98.6</b>	<b>96.5</b>	82.1	98.1
$\alpha = 0.6$	<b>98.2</b>	95.4	85.0	96.9	–	–	–	–	–	–	–	–

other methods. Note that by varying  $\alpha$ , we could easily achieve models with very high sensitivity or specificity, thus we stress that it is the compromise that is relevant. Besides, this shows that the AUC metric is the most adequate to inspect the true model’s capacity to distinguish both classes. We did not conduct this experiment in the other databases, since the number of models that are trained in a *leave-one-out* validation setting is very high. The use of focal loss over cross entropy lead to an improvement of 0.2% points regarding the AUC metric, when evaluating the system in the DRIVE database for  $\alpha = 0.5$ . The other metrics did not significantly change with this loss, meaning that it mostly induced the system to become slightly more confident on its predictions. Then, this seems to support that the single-resolution deep architecture was the main reason for our system to significantly outperform the FCN proposed in [2]. Figure 3 shows the best and worst predictions outputted by the proposed methodology for the considered databases, regarding AUC. It is possible to visualize that the model is able to cope with challenging imaging conditions, and even with the presence of severe pathology (4th row of Fig. 3).

Using a Nvidia GeForce GTX 1080 Ti GPU, it took us 2.1, 2.7, and 6.5 s to make a prediction for an image in DRIVE, STARE, and CHASEDB1 databases, respectively. The method of Liskowski and Krawiec [11] takes on average 92 s using the Nvidia GTX Titan GPU. Even though the GPUs are not identical, this strongly suggests that our method is significantly faster, thus being more adequate for real-time applications.





**Fig. 3.** Best and worst results for each database, concerning the AUC metric. From left to right: raw color fundus image, probability map outputted by the proposed methodology, segmentation obtained by thresholding probabilities at 0.5, and ground truth. (Color figure online)

## 4 Conclusion

In this paper we proposed a fully convolutional network to perform vessel segmentation in raw retinal fundus images. This design is more convenient and efficient than state-of-the-art best performing approaches, as it allows to make predictions for unseen images of different sizes at a single step, a trait that becomes relevant in screening scenarios. Our results demonstrate that the proposed method does not necessarily compromise the performance in this task, as it was able to reach state-of-the-art performance in two out of the three tested databases (DRIVE and CHASEDB1). In STARE, the raw images are significantly different from each other, such that preprocessing may be necessary to achieve better results. Thus, for future work, cost efficient preprocessing techniques will be tested for analyzing whether the performance can be improved. Semi-supervised learning will be also targeted, as a means of incorporating unlabeled data in the training process and obtain models that generalize better.

**Acknowledgments.** This work was financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within PhD grant number SFRH/BD/126224/2016 and within project UID/EEA/50014/2019.

## References

1. Azzopardi, G., Strisciuglio, N., Vento, M., Petkov, N.: Trainable cosfire filters for vessel delineation with application to retinal images. *Med. Image Anal.* **19**(1), 46–57 (2015)
2. Dasgupta, A., Singh, S.: A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation. In: *IEEE 14th International Symposium on Biomedical Imaging*, pp. 248–251. IEEE (2017)
3. Fraz, M., et al.: An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Trans. Biomed. Eng.* **59**(9), 2538–2548 (2012)
4. Fu, H., Xu, Y., Lin, S., Kee Wong, D.W., Liu, J.: DeepVessel: retinal vessel segmentation via deep learning and conditional random field. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) *MICCAI 2016*. LNCS, vol. 9901, pp. 132–139. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46723-8\\_16](https://doi.org/10.1007/978-3-319-46723-8_16)
5. Hoover, A., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Med. Imaging* **19**(3), 203–210 (2000)
6. Kanski, J., Bowling, B.: *Clinical Ophthalmology: A Systematic Approach*. Elsevier Health Sciences (2011)
7. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
8. Klonoff, D., Schwartz, D.: An economic analysis of interventions for diabetes. *Diabetes Care* **23**(3), 390–404 (2000)
9. Li, Q., Feng, B., Xie, L., Liang, P., Zhang, H., Wang, T.: A cross-modality learning approach for vessel segmentation in retinal images. *IEEE Trans. Med. Imaging* **35**(1), 109–118 (2016)
10. Lin, T., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* (2018)

11. Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels with deep neural networks. *IEEE Trans. Med. Imaging* **35**(11), 2369–2380 (2016)
12. Marin, D., Aquino, A., Gegúndez-Arias, M., Bravo, J.: A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Trans. Med. Imaging* **30**(1), 146–158 (2011)
13. Owen, C., et al.: Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program. *Invest. Ophthalmol. Vis. Sci.* **50**(5), 2004–2010 (2009)
14. Staal, J., Abràmoff, M., Niemeijer, M., Viergever, M., van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **23**(4), 501–509 (2004)
15. Yin, Y., Adel, M., Bourennane, S.: Retinal vessel segmentation using a probabilistic tracking method. *Pattern Recogn.* **45**(4), 1235–1244 (2012)
16. Zhang, B., Zhang, L., Zhang, L., Karray, F.: Retinal vessel extraction by matched filter with first-order derivative of Gaussian. *Comput. Biol. Med.* **40**(4), 438–445 (2010)