A Deep Neural Network for Vessel Segmentation of Scanning Laser Ophthalmoscopy Images

Maria Ines Meyer[1(✉)], Pedro Costa[1], Adrian Galdran[1], Ana Maria Mendonça[1,2], and Aurélio Campilho[1,2]

[1] INESC-TEC - Institute for Systems and Computer Engineering, Technology and Science, Porto, Portugal
{maria.i.meyer,pedro.costa,adrian.galdran}@inesctec.pt
[2] Faculdade de Engenharia da Universidade do Porto, Porto, Portugal
{amendon,campilho}@fe.up.pt

Abstract. Retinal vessel segmentation is a fundamental and well-studied problem in the retinal image analysis field. The standard images in this context are color photographs acquired with standard fundus cameras. Several vessel segmentation techniques have been proposed in the literature that perform successfully on this class of images. However, for other retinal imaging modalities, blood vessel extraction has not been thoroughly explored. In this paper, we propose a vessel segmentation technique for Scanning Laser Opthalmoscopy (SLO) retinal images. Our method adapts a Deep Neural Network (DNN) architecture initially devised for segmentation of biological images (U-Net), to perform the task of vessel segmentation. The model was trained on a recent public dataset of SLO images. Results show that our approach efficiently segments the vessel network, achieving a performance that outperforms the current state-of-the-art on this particular class of images.

Keywords: Scanning Laser Ophthalmoscopy · Retinal vessel segmentation

1 Introduction

The retina offers a unique and noninvasive insight into the vascular system. A number of diseases cause visible alterations to the retina. Cardiovascular disease, for example, causes changes in the ratio between the diameter of arteries and veins, and can be related to anatomical changes, such as increased vessel tortuosity and alterations in the branching angles at bifurcation points [10]. Complications deriving from diabetes, namely Diabetic Retinopathy (DR), can also be detected through retinal alterations. DR is the main cause of blindness in most developed countries [9]. When detected at an early stage it can be controlled and blindness prevented. DR manifests itself in a number of different ways, depending on the progression stage of the disease. Some of the changes

F. Karray et al. (Eds.): ICIAR 2017, LNCS 10317, pp. 507–515, 2017.
DOI: 10.1007/978-3-319-59876-5_56

that can be associated with DR are retinal hemorrhages and the formation of new vessels, among others.

As a consequence, the retina is routinely examined in medical settings as a means of diagnosis, through the acquisition of retinal images. The standard image modality in this context is color fundus photography. Other types of techniques applied to retinal imaging include Optical Coherence Tomography (OCT), which produces high-resolution cross sectional images of tissues, and Scanning Laser Ophthalmoscopy (SLO), which also produces high-quality images of the eye fundus by using laser beams to acquire the image [15]. In particular, retinal SLO images present some important advantages over standard fundus photography, namely a stronger contrast and definition of the structures. Nevertheless, this class of images has been less studied in the past, due to the lack of publicly available databases.

Among the existing computational retinal image analysis tasks, the segmentation of the retinal vasculature is of particular relevance. Manual retinal vessel segmentation is a costly process that can greatly benefit from an automated approach. For this reason, several methods have been proposed to approach this problem [4,5,8,11]. However, most of the available techniques are designed for standard color retinal images, and may fail to generalize correctly to SLO images.

In this work, we build on an existing Deep Neural Network (DNN) architecture designed for segmentation of cells and neuronal structures in microscopy images [13], and suitably modify the architecture to adapt it to the segmentation of retinal vessel trees. The method is tested on the newly available IOSTAR dataset [16], which contains SLO images and manually segmented vessel networks. We also validate the approach on the RC-SLO dataset, which comprises a selection of region samples displaying hard to segment cases, such as bifurcations or crossings. In both cases, our approach achieves state-of-the-art results, reaching high values for accuracy and for AUC (area under the ROC curve).

In the following sections, we present a brief overview on the particularities of SLO retinal imaging. Next, we give a description of the employed DNN architecture and the training procedure. Finally, we provide an analysis of the obtained results, and we conclude with some remarks on future research.

2 Scanning Laser Ophthalmoscopy

Scanning Laser Ophthalmoscopy (SLO) was first proposed in [15]. In short, the eye fundus is scanned by a laser beam and light is collected only from one retinal point at a time. Different wavelengths enable the imaging of deeper structures in the retina. As a result, this type of imaging has a number of advantages over color fundus photography, namely an improved contrast, lower light exposure, finer detail, and direct digital acquisition. Nevertheless, there are also associated problems, *e.g.* a large sensitivity to movement artifacts, or an expensive acquisition that cannot be performed in true color. A first approach to obtain color SLO images was proposed in [6], by imaging the eye fundus with different laser wavelengths and combining them to form a single image.

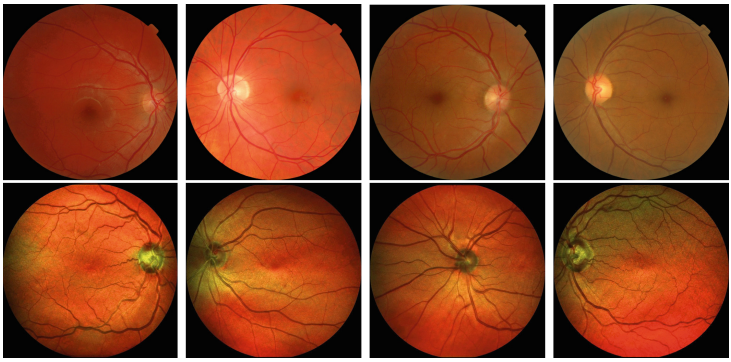


Fig. 1. Comparison between fundus photographs (top) and SLO images (bottom). (Color figure online)

The aforementioned drawbacks contributed to a more widespread use of color fundus photography, which remains the standard in retinal imaging, both in clinical settings and in research. For comparison, some examples of color fundus photographs and SLO images are shown in Fig. 1. An immediate observation is that in SLO images the anatomical structures are represented with higher contrast and detail, and the outer rim of the optic disk has a strong dark green coloring, which corresponds to the neuroretinal rim. These and other differences between both image types may lead to an incorrect generalization of current vessel segmentation methods, that were developed for color fundus photographs.

Few works propose retinal vessel segmentation techniques specifically designed for SLO images [1,12,16]. In [12] a supervised method is presented where the SLO image is pre-processed to enhance the contrast and illumination of the vascular network. The maximum intensity value at each pixel location is extracted at different scales to form an intensity map. Several other parametric maps are extracted, reflecting the width range, standard deviation and local orientation of the vessels, from the combination of the different scales in the previous step. These alternative representations are used as input to a neural network classifier. Similarly to [12], the technique presented in [1] consists of a combination of retinal image enhancement and feature extraction procedures. In the first stage, the SLO image is denoised, and its blood vessels are enhanced by lifting the image to an associated space of positions and orientations. Following, several visual features are computed, including intensities, wavelet responses and an extension of Gaussian derivatives. These features are then supplied to a neural network to produce vessel likelihoods for each pixel. The approach in [16] is related to [1], since the SLO images undergo a similar transformation to a joint positions and orientations space. However, after enhancing the vessel tree, the image is backprojected to the 2D space, and the vasculature is segmented by means of a global thresholding, turning the method into an unsupervised technique.

3 Deep Neural Network for Vessel Segmentation

To achieve vessel segmentation, we propose to consider the problem as a semantic segmentation task. This approach differs from a standard segmentation task in the sense that the aim is to classify each pixel in an image as belonging to one class. In our particular case, the classes to consider are vessel and background.

Deep learning architectures have been widely applied to the semantic segmentation problem. A relevant technique was proposed in [14], based on Fully Convolutional Networks (FCNs). Building on the ideas of FCNs, Ronneberger *et al.* proposed the U-Net architecture [13], which was designed for the segmentation of neuronal structures in electron microscopic recordings and cells in light microscopy images. The U-Net architecture is a powerful deep classifier, that can be trained with a low quantity of images to produce precise segmentations.

The architecture of the U-Net is represented in Fig. 2. The network has two main sections, or paths - a first contracting path, followed by a dimensionally symmetric expanding path. The contracting path consists of consecutive convolutional layers, with stride two and followed by Rectified Linear Unit (ReLU) activations. Assuming that the dimension of the input is a power of two, the stride is selected so that the dimension of the output feature map of the contracting path layers decreases until a dimension of $1 \times 1 \times nf$ (nf: number of filters) is reached. This point in the network marks the beginning of the expanding path. The output of the layer is upsampled to have the same dimension as the previous layer in the contracting path. To compensate for the loss in spacial resolution that results from the multiple downsampling operations, the upsampled feature map is concatenated with the feature map of the corresponding layer in the contracting path. The new feature map serves as input for a new convolutional layer, followed by ReLU activation. This procedure is repeated for the output of each convolutional layer until the output of the expanding path layers reaches the same dimension as the first layer of the network. Finally, the feature vectors are mapped into vessel/background classes by a sigmoid activation function.

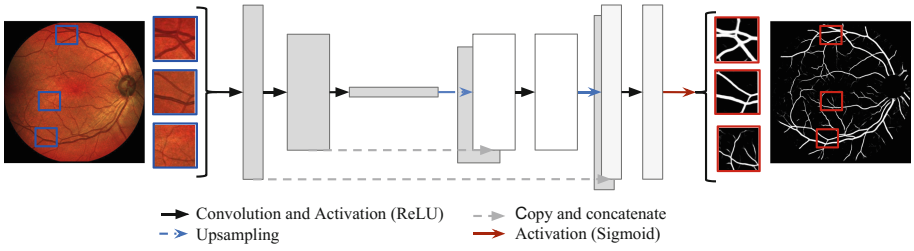


Fig. 2. Overview of the U-net model. Each box represents a multi-channel feature map.

4 Datasets and Implementation

The method was trained using the IOSTAR dataset [16], which includes 30 SLO images taken with an EasyScan camera (provided by i-Optics B.V., the Netherlands). The images have a resolution of 1024×1024 pixels with a 45° field of view (FOV), and were annotated and corrected by two different experts. Annotations in the optic disk region are not available, for which reason an optic disk mask is provided, along with the FOV mask. Further information about the dataset is available in [1] and [16].

From the initial dataset, the first 20 images were employed for training, and the remaining 10 were used to build an independent test set. Images in the training set were divided into overlapping patches of 128×128 pixels. The resulting set of patches was randomized and 10% of this set was used as a validation set.

The method trained on IOSTAR was tested with no further modifications on the RC-SLO vessel patch dataset [16], which is composed of 40 images with dimension 360×320 pixels. This dataset provides a good test because it comprises cases that are difficult to segment, such as central vessel reflex, crossings, bifurcations, background artifacts and high curvature changes. These images were also divided into smaller overlapping patches of 128×128 pixels.

In order to build a vessel tree segmentation of the entire retinal image, we first need to combine the predictions generated locally for each image patch. To achieve this, at the time each patch is produced, the coordinates translating it to its corresponding location in the input image are stored. This allows for a simple reconstruction step to be performed once the predictions for every patch are available. When there is overlap between patches, the resulting local probability maps are simply averaged pixel-wise.

The described DNN was optimized by minimizing the pixel mis-classification error, according to the cross-entropy loss function. The minimization followed a stochastic mini-batch gradient descent, with the gradients computed by standard backpropagation. The applied optimization technique was the Adam optimizer [3] with a learning rate of 1×10^{-4}. The model was trained on the training set, while the loss was monitored on the validation set. Training took approximately 12 h on a NVIDIA GeForce GTX 1080 GPU. Once trained, the model can generate a vessel segmentation for one 128×128 patch in approximately 0.05 s. The implementation was done in Python using Keras [2].

5 Experimental Results and Evaluation

The model was evaluated on the IOSTAR test set. The test images were normalized following the same process as at training time, *i*.e., they were linearly transformed so that they had zero mean and unit standard deviation, and divided into overlapping patches of 128×128 pixels. The vessel segmentation was predicted for each patch, and the image was then reconstructed. The pixel probabilities for overlapping patches were averaged, in order to get the final probability map of the retinal segmentation, as described in Sect. 4.

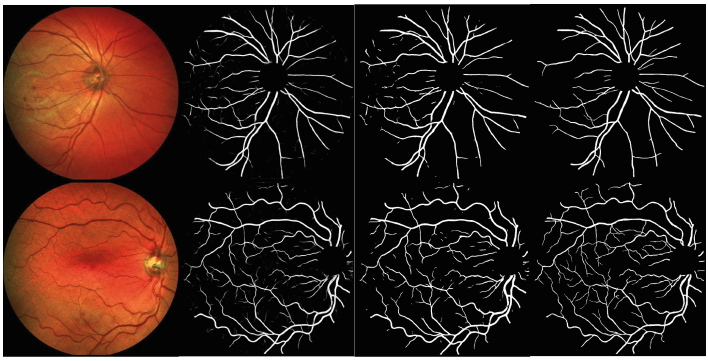
The method was then also evaluated on the RC-SLO dataset. Each 360×320 pixel patch was normalized to have unit standard deviation and mean zero, and divided into smaller patches of 128×128 pixel, to be fed into our pre-trained network. The same reconstruction protocol as described above was used.

The resulting soft segmentations were subjected to ROC analysis. The Mathews Correlation Coefficient (MCC) [7] was computed at different thresholds. MCC is a commonly used score to assess the performance of binary classifiers in the presence of skewed classes, as is our case since there is a more predominant presence of background pixels than vessel pixels. The threshold maximizing the MCC was employed to generate binary segmentations, from which Accuracy (Acc), Sensistivity (Se) and Specificity (Sp) were calculated.

For both datasets the values achieved outperformed the previously existing approaches, with AUC scores of 0.9771 on IOSTAR and 0.9807 on RC-SLO. The derived metrics also reflect a state-of-the-art performance. Table 1 summarizes the results, and compares our method with previous approaches that reported results on the same datasets. Figures 3 and 4 present example results of applying the proposed technique to IOSTAR and RC-SLO images, respectively.

Table 1. Summary of our results and comparison with other previously proposed techniques that also reported performance on the IOSTAR and RC-SLO datasets. (Best values are displayed in bold).

Methods	Approach	Dataset	Se	Sp	Acc	AUC	MCC
BIMSO [1]	Supervised	IOSTAR	0.7863	0.9747	0.9501	0.9615	0.7752
LAD-OS [16]	Unsupervised	IOSTAR	0.7545	0.9740	0.9514	0.9626	0.7318
LAD-OS [16]	Unsupervised	RC-SLO	0.7787	0.9710	0.9512	0.9626	0.7327
OURS	Supervised	IOSTAR	**0.8038**	**0.9801**	**0.9695**	**0.9771**	**0.7920**
OURS	Supervised	RC-SLO	**0.8090**	**0.9794**	**0.9623**	**0.9807**	**0.7905**



(a) (b) (c) (d)

Fig. 3. Segmentation results for IOSTAR dataset. (a) Real image. (b) Output probability map. (c) Segmentation of the probability map at threshold that optimizes MCC. (d) Ground truth annotation. (Color figure online)

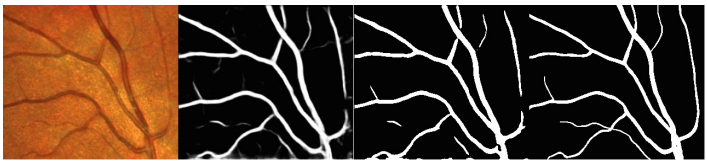


Fig. 4. Example of segmentation result for RC-SLO dataset. (a) Real image; (b) Output probability map. Note the artifacts at the bottom of the image; (c) Segmentation of the probability map at the threshold that optimizes MCC; (d) Ground truth annotation.

Visual inspection of the results in Fig. 3 shows that our technique sometimes under-segments difficult cases, such as thin vessels. We also observe some vessel discontinuities, indicating that the method can still be improved. For the RC-SLO dataset some results present artifacts at the bottom part, as exemplified in Fig. 4(b) and (c). This could be due to the model being trained on patches extracted from entire images, which contained the round FOV borders.

6 Conclusions and Future Work

In this work we proposed a supervised method to segment the retinal vessel tree of Scanning Laser Opthalmoscopy images. Our approach is based on the U-Net DNN architecture, designed for image segmentation. The method was trained on the IOSTAR dataset and achieved above state-of-the-art results on both IOSTAR and RC-SLO datasets.

An important limitation of the present work is its inability to reliably segment vessels within the optic disk, which is a direct result of missing ground truth segmentations in this region, where the background is very different from the remaining image. We are optimistic that if the annotated ground truth contained information regarding these vessels our method could have resolved well. An issue to address is the appearance of artifacts at the bottom part of some of the RC-SLO segmentations. It is noteworthy that even with the presence of such artifacts the performance of the method was very good.

Another limitation of the proposed technique is that it was trained for SLO images, and we can expect it to generalize badly to standard retinal images. A future line of research will be to perform transfer learning from one problem to the other. Visually, vessel segmentation from SLO and from standard retinal images are similar problems. For this reason, after training the method on SLO images, we can take advantage of the representations learned by the neural network, fine-tuning its weights so that it performs well also on standard retinal images. This would avoid the necessity of carrying out a complete re-training. Training the same model simultaneously on both image types is also appealing, since this could confer the neural network more generalization capability.

Finally, the proposed technique could be improved to segment patches of any size and resolution. For this, the method should be modified to account

for multiple scales. This is a well known problem in semantic segmentation, and could be overcome by altering the architecture so that it could support images of different resolutions and patch sizes. This could be relevant to avoid re-training the network each time it is presented with images of different resolutions.

Acknowledgments. This work is financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and the European Regional Development Fund (ERDF), within the project "NanoSTIMA: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics/NORTE-01-0145-FEDER-000016".

References

1. Abbasi-Sureshjani, S., Smit-Ockeloen, I., Zhang, J., Ter Haar Romeny, B.: Biologically-Inspired supervised vasculature segmentation in SLO retinal fundus images. In: Kamel, M., Campilho, A. (eds.) ICIAR 2015. LNCS, vol. 9164, pp. 325–334. Springer, Cham (2015). doi:10.1007/978-3-319-20801-5_35
2. Chollet, F.: Keras (2015). https://github.com/fchollet/keras
3. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. In: International Conference on Learning Representations, pp. 1–13 (2014)
4. Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels with deep neural networks. IEEE Trans. Med. Imaging **35**(11), 2369–2380 (2016)
5. Maninis, K.-K., Pont-Tuset, J., Arbeláez, P., 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
6. Manivannan, A., Kirkpatrick, J.N., Sharp, P.F., Forrester, J.V.: Novel approach towards colour imaging using a scanning laser ophthalmoscope. Br. J. Ophthalmol. **82**(4), 342–345 (1998)
7. Matthews, B.: Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta (BBA) - Protein Structure **405**(2), 442–451 (1975)
8. Mendonça, A.M., Campilho, A.: Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. IEEE Trans. Med. Imaging **25**(9), 1200–1213 (2006)
9. Mohamed, Q., Gillies, M., Wong, T.: Management of diabetic retinopathy: a systematic review. JAMA **298**(8), 902–916 (2007)
10. Nguyen, T.T., Wong, T.Y.: Retinal vascular changes and diabetic retinopathy. Curr. Diab. Rep. **9**(4), 277–283 (2009)
11. Orlando, J.I., Blaschko, M.: Learning fully-connected CRFs for blood vessel segmentation in retinal images. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8673, pp. 634–641. Springer, Cham (2014). doi:10.1007/978-3-319-10404-1_79
12. Pellegrini, E., Robertson, G., Trucco, E., MacGillivray, T.J., Lupascu, C., van Hemert, J., Williams, M.C., Newby, D.E., van Beek, E., Houston, G.: Blood vessel segmentation and width estimation in ultra-wide field scanning laser ophthalmoscopy. Biomed. Opt. Express **5**(12), 4329–4337 (2014)

13. 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
14. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)
15. Webb, R.H., Hughes, G.W., Pomerantzeff, O.: Flying spot TV ophthalmoscope. Appl. Opt. **19**(17), 2991–2997 (1980)
16. Zhang, J., Dashtbozorg, B., Bekkers, E., Pluim, J.P.W., Duits, R., ter Haar Romeny, B.M.: Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. IEEE Trans. Med. Imaging **35**(12), 2631–2644 (2016)