Vessel Segmentation on Low-Resolution Retinal Imaging

Paulo Coelho[1,2]
paulo.coelho@ipleiria.pt
José Camara[3]
jcrcamara@hotmail.com
Hasan Zengin[4]
hasalp38@gmail.com
João M. F. Rodrigues[5]
jrodrig@ualg.pt
António Cunha[2,6]
acunha@utad.pt

[1] Escola Superior de Tecnologia e Gestão, Politécnico de Leiria
[2] INESC TEC - Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência Porto, Portugal
[3] Universidade Aberta, Porto, Portugal
[4] Mehmet Akif Ersoy University, Turkey
[5] LARSyS e ISE, Universidade do Algarve, Portugal
[6] Universidade de Trás-os-Montes e Alto Douro Vila Real, Portugal

Abstract

Retinal vessel segmentation process highlights a set of signals that will serve to aid for diagnosing of various retinal pathologies and lead to a more accurate diagnosis. This paper presents a framework for automatic vessel segmentation applied to lower-resolution retinal images taken with a smartphone equipped with D-EYE lens. A private dataset was assembled and annotated, and two CNN based models were trained for automatic localisation retinal areas and vessel segmentation. A Faster R-CNN that achieved a 96% correct detection of all regions with a Mean Absolute Error (MAE) of 39 pixels, and a U-Net that reached a Dice Coefficient (DICE) of 0.7547.

1 Introduction

Retinal imaging is a technique that allows recording digitally the rear of the eye. These are typically taken by expensive machines like fundus cameras, that produce high-quality and high-resolution retinal images for analysis. Then, with vessel segmentation, interference from other anatomical structures are filtered, helping to obtain the focus of interest on posterior segment structures of the eye. Manually segmenting retinal veins requires minutia, is a burdening task, time- and cost-consuming. Therefore, investigations of automatic or semi-automatic methods for vessel segmentation have been evolving to assist specialists [1]. However, the use of low-cost lenses, such as D-EYE [2], can bring several advantages such as greater portability, ease of use, greater patient comfort, lower costs and so can be an assessment for unprivileged or remote populations. The drawback is the lower quality of the photos obtained when compared to fundus cameras and as consequence not having the necessary sharpness when used in eyes with small pupils, in eyes with opacity of media (keratitis, cataract), or in very bright environments.

The latest trends in research show the extensive use of convolution neural networks (CNN) for the segmentation of retinal vessels and detection of the disease, beyond many other methods [1]. Nevertheless, these methods are all focused on segmenting vessels with high-resolution retinal images. There is still a lack of studies to evaluate the effectiveness of automatic methods to segment retinal vessels in this type of image. These low-resolution and low-quality retinal images create extra difficulties in the use of traditional vessel segmentation methods.

This paper presents a framework focused on the vessels segmenting on lower resolution retinal images taken with a smartphone equipped with D-EYE lens. The framework has two main steps: (A) The detection of the optic disc region using a Faster R-CNN and (B) Visible vessel segmentation made by U-Net, both trained with a customised dataset. The dataset was created with 26 retina videos around the optic disc, with lower-resolution images, and two annotated subsets, one with the localisation of the visible retinal area and other with vessel segmentation.

2 Methodology

This work is divided into two experiences (see the pipeline in Figure 1), applied to several datasets as follows.

Dataset

A dataset of 26 low-resolution videos of the optic papilla under myosis (undilated pupil) was captured from the left and right eyes of 19 volunteers. The videos were split into single images and organised in two different datasets: dataset1 (DS1), with a total of 6060 frames with 1920 x 1080 pixels to be used in the detection of the visible retinal area (from 18 videos were gathered 3881 frames to train, from 3 videos were gathered 776 frames to validation, and from 5 videos were used 1375 frames to test); and the dataset2 (DS2), with a total of 347 frames with 320x320 pixels pixels to be used in the segmentation of the retinal veins (from 2 videos were gathered 252 frames to train, from 1 video were gathered 40 frames to validation, and from 1 video were used 55 frames to test). Additionally, as dataset3 (DS3), a training set of retina public dataset [3] was used for pre-training a segmentation CNN. It is composed of 20 colour images with 565x584 pixels, resulting in 2810 patches with 80x80 pixels (for training, from 14 frames resulted in 1967 patches, for training from 3 frames resulted 421 patches and for testing, from 3 images resulted 422 patches).

Setup

In the first part of the experience, the detection of the retinal visible area (A) consists of computing the location of a rectangle defined by P1 and P2 (see Figure 1), that encloses the visible area in the image (the area of interest). In this case, input images have 1920x1080 pixels (from DS1), due to the D-EYE low lens aperture the area of interest has up to 320x320 pixels. For this purpose, it was used a Faster R-CNN model [4]. To evaluate the model, the Mean Absolute Error (MAE), which is a commonly used metric since it permits to measure the accuracy for continuous variables. In this particular case, four variables were used, two for the coordinates of the upper left corner (P1) and the other two for the lower right corner (P2).

In the second part of the experience (B), also depicted in Figure 1, the vessels segmentation was done within the detected retinal areas, with a U-Net [5] model pre-trained with the DS3 and tuned and evaluated with DS2. To measure the success of the model, it was used the Dice Coefficient (DICE), which is a relative metric that provides a similarity measure between predicted and ground truth segmentations.

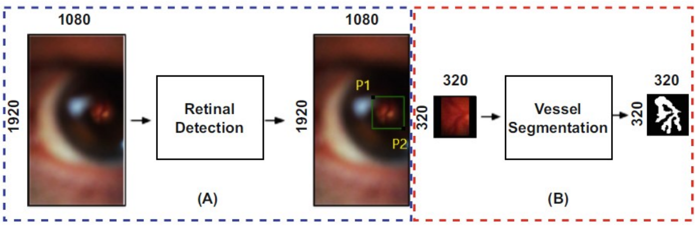

Figure 1: Pipeline diagram for the proposed low-resolution vessel segmentation framework.

For retinal detection (A), the Faster R-CNN model was used to detect retinal visible area detection. It was implemented in TensorFlow, with features pre-trained with Inception Resnet V2 and finetuned with the private dataset1. For augmenting dataset1, rotations were applied with 90-degree steps. The model was trained with default parametrisation: l2 regulariser of 0.01, truncated normal initialiser of 0.01, maxpool kernel size of 2, maxpool stride of 2, localisation loss weight of 2, objectness loss weight of 1, score converter Softmax, momentum optimiser with learning rate 0.0002, momentum optimiser value of 0.9.

In terms of retinal vessel segmentation (B), The U-Net model was implemented using Keras, with a TensorFlow backend. For training the U-Net model, it was used the binary cross-entropy as loss function, and Adam's optimiser with 10^{-3} learning rate based on Ange Tato and Roger Nkambou's work [6] used to achieve faster a stable convergence.

3 Results and discussion

The framework was evaluated for the retinal visible area detection and for vessels segmentation test sets.

The Faster R-CNN obtained results for retinal visible area detection are organised in 10 classification scores with intervals of 0.1 (represents the level of confidence of the detection), and can be seen in Table 1.

Table 1: Testset evaluation of the Faster R-CNN model for retinal visible area detection.

Classification score	Frequency	P1 MAE (pixels)*	P2 MAE (pixels)*	P1 and P2 MAE (pixels)*
0.0	30 (2%)	311 (414)	252 (321)	281 (371)
0.1	4 (0%)	46 (28)	41 (20)	43 (25)
0.2	12 (1%)	63 (41)	40 (31)	51 (38)
0.3	6 (0%)	47 (27)	30 (8)	38 (22)
0.4	7 (1%)	37 (23)	41 (26)	39 (24)
0.5	11 (1%)	76 (69)	50 (37)	63 (57)
0.6	11 (1%)	91 (60)	37 (20)	64 (52)
0.7	14 (1%)	61 (62)	35 (26)	48 (49)
0.8	29 (2%)	67 (50)	36 (20)	52 (41)
0.9	1,251 (91%)	47 (49)	28 (13)	37 (37)
Total images	1,375			

* MAE: mean (standard deviation)

The detection was very successful as 91% of the test images were detected with the classification score equal to or greater than 0.9 (see example in Figure 2, left). As the confidence score decreases, the MAE errors keep approximately constant until it reaches the score interval 0.0, where it increases for the mean of 281 and a standard deviation of 371 (see example in Figure 2, left). It was considered reasonable to use a threshold above 0.5 to accept the areas as valid-regions, achieving 96% of correct detected for all regions, with MAE of 39 pixels.

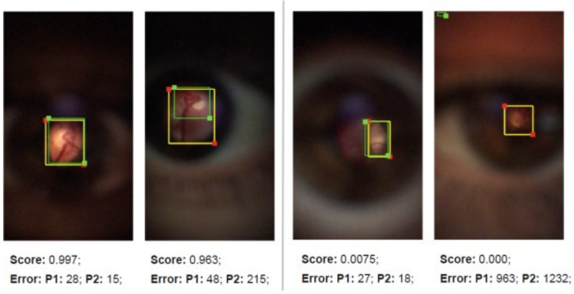

Figure 2: Example of the best and the worst sum of errors P1 and P2 in intervals 0.9 (left) and 0.0 (right). Yellow rectangles are the ground truth and the green are the model prediction.

The U-Net model was trained first with DS3 (Model 1), then trained at the junction of DS2 and DS3 testsets (Model 2) and later retrained Model 1 with DS2 (Model 3) to tune the network with D-EYE retinal data. The attained results are summarised in Table2.

Table 2: Results of the Model 1, Model 2 and Model 3

	Model 1	Model 2	Model 3
(DS3) testset	0.7824	–	0.5784
(DS2 & DS3) testset	0.7474	0.7312	–
(DS2) testset	0.4797	0.7547	0.5580

Model 1 has a reasonable Dice coefficient that seem adequate for the task (0.7824). Observing Figure 3 A), it can be seen that the best result (first column) has achieved a DICE of 0.935 in a patch where vessels are wide and well visible. The model predictions (row 3) have the same structure but seem wider than the ground truth (row 2). At the second column, it can be seen the worst prediction (DICE of 0.0512). At the original image, vessels are thin, almost imperceptible and quite different from the vessels expected to find in low-resolution images. It was selected another image patch with thin veins that seem to us more similar to the ones expected (third column). In this case, the predicted image preserves the structure; it also seem wider than the ground-truth and achieved a DICE of 0.8571. To observe how the Model 1 performs with the low-resolution images, it was evaluated in the DS2 testset, obtaining a low DICE value (0.4797). In the fourth and fifth columns, it can be seen the best and worst predictions. Both patch images are very dark, and veins are poorly visible - the image-patch of fifth is the poorest. The best-predicted segmentation (DICE of 0.8009) is very good, considering the visibility of the veins and though the difficulty of manually creating the ground-truth.

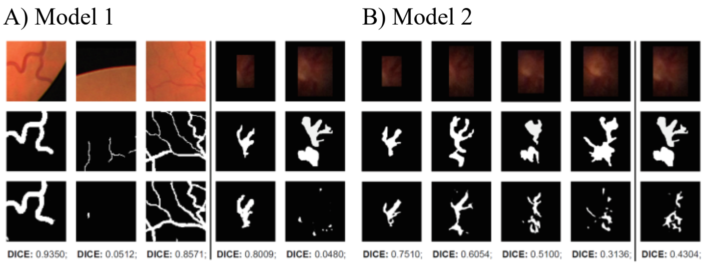

Figure 3: A) 1st row: original patches: the first three are from DS3 testset, respectively best, worst and reasonable predictions; and the last two from DS2 testset, respectively best and worst predictions. 2nd row, ground truth patches, and the 3rd row, Model 1 predictions. B) 1st row: original patches: the first four are from DS2 testset, respectively best, two in-between, and worst predictions; and for comparison, the worst case of Model 1 low-resolution prediction, is presented in the last column. 2nd row, ground truth patches. 3rd row, Model 2 predictions.

The Model 2 achieved a DICE of 0.7312 at the junction of DS2 and DS3 testsets that is lower than the obtained for Model 1 (DICE of 0.7474), but it achieved better on the DS2 testset (DICE of 0.7547). Examples of predicted images of Model 2 can be seen in Figure 3 B). To illustrate the Model 2 predictions of dataset DS2 testset, were chosen four images: the best prediction (DICE: 0.7510), two in-between predictions (DICE: 0.6054, DICE: 0.5100) and the worst prediction (DICE: 0.4304). For comparison with Model 1, column 5 has the prediction results of the worst-patch image predicted by Model 1 (see Figure 3 A), column 5). It can be seen that segmentations are much better: in the first two cases, the structure is all connected as in the ground truth, the other two (where veins are less visible in patch images) have several discontinuities in the structure. In column 5, one can see that Model 2 produces a much better segmentation (DICE: 0.4304) than produced by Model 1 (see Figure 3 A), lower right image).

The last tests made were for Model 3, by doing a posterior train of the Model 1 with DS2, but the results were worse than with Model 1.

4 Conclusions

In this paper, a framework for vessels segmenting on lower-resolution retinal images was proposed, evaluated, and the attained results were presented. A dataset of train models was assembled and annotated for automatic localisation of retinal areas and for vessel segmentation. For the framework, two CNN-based models were successfully trained, a Faster R-CNN that achieved a 96% correct detection of all regions with a MAE of 39 pixels, and a U-Net that achieved a DICE of 0.7547. This study is a precursor to future works to the determination of eye diseases, such as glaucoma and diabetes, applied to low-resolution images.

References

[1] Singh, N., Kaur, L.: A survey on blood vessel segmentation methods in retinal images. In: 2015 International Conference on Electronic Design, Computer Networks & Automated Verification (EDCAV), pp. 23–28. IEEE (January 2015).

[2] The Portable Ophthalmoscope for Your iPhone | Hand held fundus camera price| D-EYE for Humans | D-EYE, https://www.d-eyecare.com/en%20PT/product, Accessed: 2020-09-18

[3] Staal, J., Abramoff, 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).

[4] Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39(6), 1137–1149 (2017).

[5] 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).

[6] Tato, A., Nkambou, R.: Workshop track -ICLR 2018 Improving Adam Optimizer, pp. 1–4 (2018)