UNCERTAINTY-AWARE ARTERY/VEIN CLASSIFICATION ON RETINAL IMAGES

Adrian Galdran[⋆,†]*, M. Meyer*[⋆]*, P. Costa*[⋆]*, Mendonça*[∗,‡]*, A. Campilho*[∗,‡]

[⋆] Institute for Systems and Computer Engineering, Technology and Science, Porto, Portugal
[†] École de Technologie Supérieure, University of Quebec, Canada
‡ Faculdade de Engenharia, Universidade de Porto, Portugal

ABSTRACT

The automatic differentiation of retinal vessels into arteries and veins (A/V) is a highly relevant task within the field of retinal image analysis. However, due to limitations of retinal image acquisition devices, specialists can find it impossible to label certain vessels in eye fundus images. In this paper, we introduce a method that takes into account such uncertainty by design. For this, we formulate the A/V classification task as a four-class segmentation problem, and a Convolutional Neural Network is trained to classify pixels into background, A/V, or uncertain classes. The resulting technique can directly provide pixelwise uncertainty estimates. In addition, instead of depending on a previously available vessel segmentation, the method automatically segments the vessel tree. Experimental results show a performance comparable or superior to several recent A/V classification approaches. In addition, the proposed technique also attains state-of-the-art performance when evaluated for the task of vessel segmentation, generalizing to data that was not used during training, even with considerable differences in terms of appearance and resolution.

Index Terms— Artery/Vein Classification, Retinal Vessel Segmentation, Uncertainty

1. INTRODUCTION

Within the field of retinal image analysis, the study of the vasculature of the retina from eye fundus images plays a key role in early disease diagnosis. In particular, discrimination of retinal vessels into arteries and veins is a key step for the extraction of retinal biomarkers and the analysis of several vision-threatening diseases. For instance, an abnormally low ratio between arteriolar and venular widths (AVR) is known to be predictive of several cardiovascular diseases, as well as Diabetic Retinopathy (DR) [12]. Also, glaucoma has been associated with the hemodynamics of retinal arteries [13].

Previous approaches to identify arteries and veins in a retinal image mostly consist of classifying pixels on a vasculature segmentation, available beforehand, into suitable classes. Hence, the task is typically formulated as a pixel-wise binary classification problem, and different techniques have been developed to solve it. Early attempts tried to derive discriminative visual features, used afterwards to optimize a classifier [6, 12]. Another family of techniques is based on treating the retinal vasculature as a directed graph, and reasoning about its topological properties [1, 2, 20]. Recently, deep learning techniques [11, 18] have been proposed, by-passing manual feature engineering through the usage of Convolutional Neural Networks to automatically learn the optimal representation of vessel pixels for its classification as arteries or veins.

One of the key contributions of this work is its focus on vessel pixels for which the correct classification is uncertain even for human experts, see Fig. 1. Uncertainty in medical imaging has recently attracted much interest [7]. Several strategies exist for modeling ambiguous expert opinion, the simplest of them being training independent models on different labels and fusing their outcome. More advanced approaches can be found in [7], applied to lung nodule segmentation. In the retinal image analysis context, multiple expert criteria was exploited in [9] to improve reliability on DR grading. However, to the best of our knowledge no previous work has focused on pixel-level uncertainty in vessel labeling for the task of vasculature segmentation or classification.

In summary, the contribution of this paper is threefold:

1. Uncertainty in A/V prediction is modeled in a straightforward manner from data annotations. This enables the direct prediction of uncertain pixels avoiding complex mechanisms to capture this phenomenon.
2. In contrast with most previous methods, the resulting model does not require a previously segmented vessel tree. Furthermore, even if it has not been explicitly optimized for segmenting vessels from the retinal background, it achieves state-of-the-art results in this task.
3. As opposed to previous approaches, which require retraining when faced with new data, our solution produces results comparable to current A/V classification techniques on data that was not used for training. This holds true even when such data consists of retinal images with markedly different visual characteristics and resolutions.

Corresponding author: Adrian Galdran (adrian.galdran@inesctec.pt).
Research funded by ERDF European Regional Development Fund, Operational Programme for Competitiveness and Internationalisation - COMPETE 2020, and the FCT - project CMUP-ERI/TIC/0028/2014.

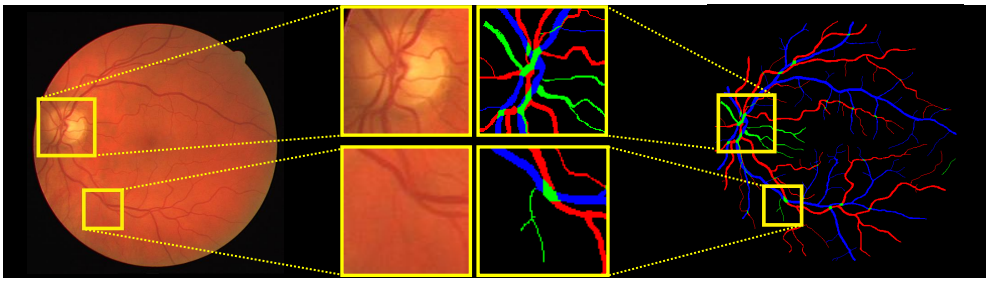

Fig. 1. Region containing uncertain pixels (left), labeled as green (right). Center: zoomed-in details.

2. METHODOLOGY

In this section we describe the details of the proposed method for classifying each pixel on a retinal image into one of the four classes of interest, *i.e.* background ($\mathbf{B}$), artery ($\mathbf{A}$), vein ($\mathbf{V}$), and uncertain ($\mathbf{U}$), shown in Fig. 1. An implementation for reproducing the results in this paper is open-sourced[1].

2.1. Data Preparation

The available data contained manual expert annotations at each pixel specifying if it belonged to $\mathbf{B}$, $\mathbf{A}$, $\mathbf{V}$, $\mathbf{U}$, or if the pixel was on a crossing of an artery and a vein. In the latter case, it is not possible to know to which of both classes a pixel corresponds, so we added crossing pixels into the $\mathbf{U}$ class.

Before training our model, each retinal image was pre-processed by applying the technique introduced in [17], which compensates illumination and exposure based on fog removal techniques [3]. Finally, all images were resized to a common resolution of 512×512.

2.2. Architecture

A Fully-Convolutional Neural Network was built for the task at hand. The employed encoder-decoder architecture follows the same principles as the popular U-NET [16]: a downsampling path with four convolutional layers, each of them with a max-pooling operation, and a symmetrical decoding path with bilinear upsampling operations followed by convolutional layers, with skip-connections to preserve information at different scales. Learnable filters were of 3×3 size and a stride of 1 in the encoding path, and 1×1 in the decoding path. Adequate symmetric padding was applied where necessary, and Parametric Rectified Linear Units were adopted as non-linearities after convolutional layers. Batch-Normalization was employed as well in the encoding path.

2.3. Training Process

The available training data was distributed into a training and validation set (75%, 25%), and the standard four-class cross-entropy error was minimized with ADAM optimizer and a constant learning rate of $2\mathrm{e}{-4}$ and a batch-size of 2. Typical data augmentation strategies were also applied (horizontal/vertical flipping, color perturbations, slight blurring).

After 300 epochs of training, signs of overfitting were observed in the validation set's error. Consequently, training was early-stopped, the training and the validation data were merged, and the model was re-trained for 300 epochs.

2.4. Building Predictions

Out-of-training images underwent the same process as training data, *i.e.* the illumination was compensated with the same method, and resizing to a 512×512 resolution was applied. Furthermore, in order to favor the generalization ability of the model, we applied color transfer of new images with respect to a source image in the training set. This was done by mapping chromatic statistics from the reference image to the target images following the method of [15]. The optimal training image to which color statistics should be mapped was selected based on performance analysis in the training set.

In order to achieve more consistent predictions, test-time augmentation was applied. This consisted of flipping and mirroring the image, generating predictions for each transformed image, applying the corresponding inverse transformations, and averaging the results.

3. EXPERIMENTAL EVALUATION

3.1. Data

For performance analysis, we selected five retinal image datasets containing images of remarkably different appearance and resolutions, both optic-disc and macula-centered:

1. **RITE [5]:** the **R**etinal **I**mages vessel **T**ree **E**xtraction dataset contains 40 macula-centered images with a relatively low resolution of 565×584 pixels. Data comes already separated into 20 training and 20 test images. The training set from RITE was used to optimize our model.
2. **INSPIRE [1]:** INSPIRE contains 40 optic-disc centered retinal images with a resolution of 2392×2048 pixels. It provides A/V annotations only at vessel centerlines.

[1] https://github.com/agaldran/a_v_uncertain

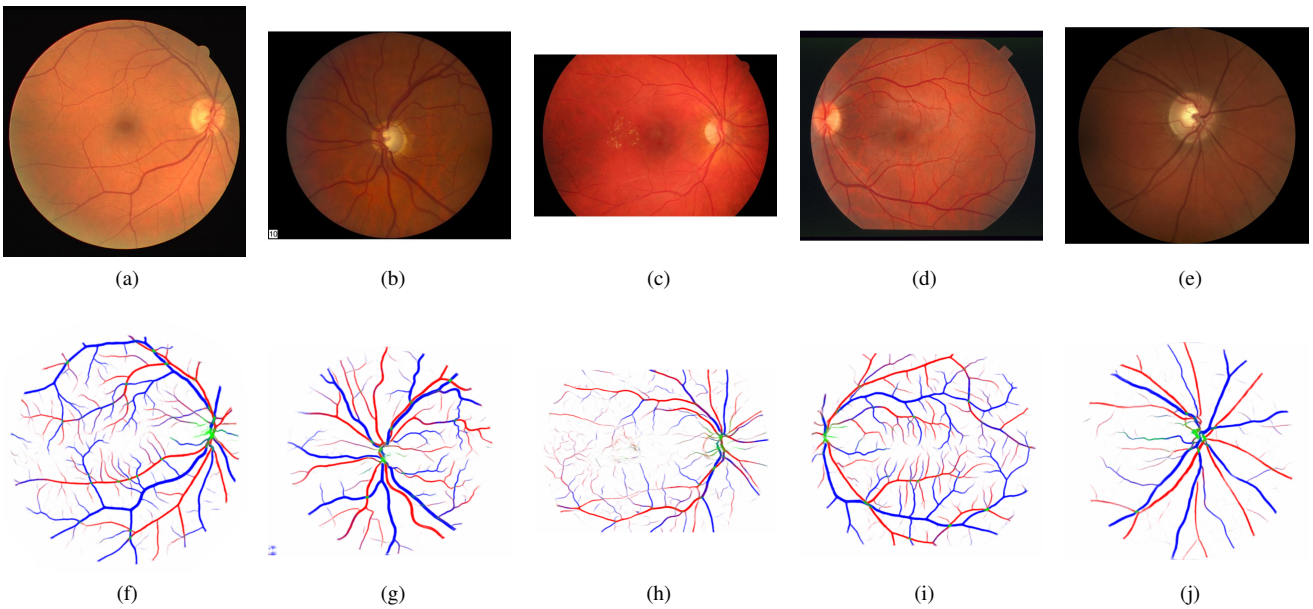

Fig. 2. Visual results on images extracted from (a) RITE [5] (b) INSPIRE [1] (c) HRF [8] (d) STARE [4] (e) LES-AV [13].

3. **HRF [8]:** The **H**igh-**R**esolution **F**undus Image Database contains 30 high-resolution macula-centered images (3504×2336), with expert ground-truth for vessel segmentation.
4. **STARE [4]:** The **ST**ructured **A**nalysis of the **RE**tina dataset contains 20 605×700 macula-centered retinal images. STARE only provides vessel labels.
5. **LES-AV [13]:** The LES-AV dataset contains 22 optic-disc centered images with a resolution of 1620×1444 pixels. No results have been reported on A/V classification for this dataset at the time of writing.

3.2. Qualitative Evaluation

The proposed method processes a retinal fundus image, returning pixel-wise predictions containing the likelihood of belonging to one of four classes of interest, namely background, artery, vein, or uncertain. Several examples of images coming from the different datasets outlined above are provided in Fig. 2. We can see that, in all cases, the model produces meaningful uncertain predictions in the optic disc area, as well as some other vessels from the periphery of the retina, where it becomes harder to assess the class of each pixel. Likewise, most vessel crossings are correctly predicted as uncertain.

3.3. Quantitative Evaluation

The method introduced in this paper does not require a pre-segmented vessel tree to produce predictions. Hence, the ideal evaluation would consist of considering the problem as a multi-class segmentation task. Unfortunately, the vast majority of previous approaches are binary classification techniques that sort vessel pixels into artery or vein. Hence, to enable objective comparisons, we redistribute background and uncertain probabilities among both classes of interest (A/V), and compute performance also as a binary classification problem, disregarding pixels that do not belong to vessels.

In order to also analyze the performance of our method for the task of vessel segmentation, we merged probabilities of artery, vein, or uncertain, into a single class, and again computed performance metrics in a binary classification setting.

3.3.1. Artery/Vein Classification

After training the proposed model in part of the RITE dataset, we build predictions for each image for the remaining data, as well as for the INSPIRE and LES-AV images. We followed the process outlined above to generate two-class predictions on each vessel pixel, and computed Area Under the ROC Curve (AUC), Accuracy (Acc), Sensitivity ($Sens$), Specificity ($Spec$), and F1-score for each dataset. The resulting performance measures are displayed in Table 1. The threshold maximizing the Youden index ($Sens + Spec - 1$) over the training set of RITE was selected to compute Accuracy, Sensitivity, Specificity, and F1-score for every other dataset.

3.3.2. Vessel Segmentation

The technique introduced in this paper operates without the need of a vessel tree in order to produce A/V predictions. However, this is of little practical value if the method cannot generate accurate vessel/background predictions. For this reason, we also report performance on this task. Vessel segmentations were generated as described in the beginning of

	RITE [5]					INSPIRE [1]					LES-AV [13]				
	AUC	*Acc*	*Sens*	*Spec*	F1	AUC	*Acc*	*Sens*	*Spec*	F1	AUC	*Acc*	*Sens*	*Spec*	F1
GrBs. [1]	-	0.85	0.90	0.84	-	-	0.85	0.90	0.84	-	-	-	-	-	-
TpEs. [2]	-	0.92	0.92	0.92	-	-	0.92	0.90	0.91	-	-	-	-	-	-
GenS. [6]	0.78	0.72	0.71	0.74	-	0.95	0.92	0.90	0.91	-	-	-	-	-	-
AVR. [12]	0.88	-	0.80	0.80	-	0.84	-	-	-	-	-	-	-	-	-
UA-AV	0.95	0.89	0.89	0.90	0.88	0.89	0.80	0.82	0.78	0.80	0.94	0.86	0.88	0.85	0.86

Table 1. Performance Comparison of AV classification methods tested on RITE, LES-AV, INSPIRE.

	RITE [5]					HRF [8]					STARE-A [4]				
	AUC	*Acc*	*Sens*	*Spec*	F1	AUC	*Acc*	*Sens*	*Spec*	F1	AUC	*Acc*	*Sens*	*Spec*	F1
DRIU [10]	0.98	0.93	0.95	0.93	0.78	-	-	-	-	-	0.96	0.93	0.90	0.94	0.75
CRF [14]	0.95	0.95	0.79	0.96	0.78	0.94	0.94	0.79	0.96	0.72	0.97	0.96	0.78	0.98	0.76
JLoss [19]	0.98	0.95	0.77	0.98	-	-	0.94	0.79	0.96	-	0.98	0.96	0.76	0.98	-
UA-AV	0.98	0.93	0.94	0.93	0.77	0.94	0.91	0.85	0.91	0.62	0.96	0.95	0.80	0.97	0.79

Table 2. Performance Comparison of vessel segmentation methods tested on RITE, HRF, STARE-A.

this section. Results are displayed in Table 2.

For a fair comparison with previous work, in the STARE dataset we used the same train/test partition (STARE-A, STARE-B) as in [14][2]. We report performance on the STARE-A sub-dataset, which was also used in [10]. It is important to notice that these three state-of-the-art vessel segmentation methods [14, 10, 19], compared in Table 2, re-train their models over a subset of each dataset before reporting results. The optimal cut-off threshold for each dataset is also computed in each different training set, which affects performance measures for each technique. In contrast, our proposed technique produces binary predictions with the same threshold in all datasets. This threshold was initially derived from ROC analysis on the training set of RITE and never altered during the subsequent experiments.

4. DISCUSSION AND FUTURE WORK

As shown in Table 1, the A/V classification of vessel pixels is solved by the proposed method in the RITE dataset with an accuracy similar to current techniques. Its performance is only surpassed by the graph-based approach of [2], which also achieves top results in the INSPIRE dataset. However, it is of great importance to remark that this and the other compared methods [1, 6, 12] were re-trained on INSPIRE, whereas our technique was tested without further adjustment. This is of much relevance in clinical practice, since it is not reasonable to expect that for a new set of images to be processed there will always be available ground-truth to re-train a given model. In addition, experiments on the LES-AV dataset (again without re-training) resulted in a similarly high performance. It is worth noting the wide visual variability among these three datasets, illustrated in Figs. (2a), (2b), and (2e).

In addition, the proposed model exhibits two other important features. First, when faced with hard-to-classify pixels, it can return a prediction of uncertainty, as shown in Figs. (2f)–(2j). This may be useful in order to disregard such vessels when extracting retinal biomarkers. Second, the method operates without the need of a previously available retinal vasculature segmentation. Interestingly, even if it was not trained for the purpose of retinal vessel segmentation, our technique provides state-of-the-art performance in this task, as shown in Table 2. Again, it is important to stress that vessel segmentation results were obtained without any further adjustment of the method to different datasets, which was not the case for the compared methods [10, 14]. The vessel segmentation capability and the generalization power of the presented method strongly favor its introduction in existing automatic vasculature analysis pipelines.

Future research directions will include the analysis of more powerful deep architectures, together with domain-adapted regularization techniques to prevent overfitting and preserve the generalization capability of the approach while improving its performance.

[2]See supplementary results of [14] at `https://ignaciorlando.github.io/` for details about the partition into train and test set of the STARE dataset.

5. REFERENCES

[1] B. Dashtbozorg, A. M. Mendonca, and A. Campilho. An automatic graph-based approach for artery/vein classification in retinal images. *IEEE Transactions on Image Processing*, 23(3):1073–1083, 2014.

[2] R. Estrada, M. J. Allingham, P. S. Mettu, S. W. Cousins, C. Tomasi, and S. Farsiu. Retinal Artery-Vein Classification via Topology Estimation. *IEEE Transactions on Medical Imaging*, 34(12):2518–2534, Dec. 2015.

[3] A. Galdran, A. Alvarez-Gila, A. Bria, J. Vazquez-Corral, and M. Bertalmio. On the Duality Between Retinex and Image Dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[4] A. Hoover, V. Kouznetsova, and M. Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, 19(3):203–210, Mar. 2000.

[5] Q. Hu, M. D. Abràmoff, and M. K. Garvin. Automated Separation of Binary Overlapping Trees in Low-Contrast Color Retinal Images. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, LNCS, pages 436–443, 2013.

[6] F. Huang, B. Dashtbozorg, T. Tan, and B. M. ter Haar Romeny. Retinal artery/vein classification using genetic-search feature selection. *Computer Methods and Programs in Biomedicine*, 161:197–207, July 2018.

[7] S. A. A. Kohl, B. Romera-Paredes, C. Meyer, J. De Fauw, J. R. Ledsam, K. H. Maier-Hein, S. M. A. Eslami, D. J. Rezende, and O. Ronneberger. A Probabilistic U-Net for Segmentation of Ambiguous Images. In *Neural Information Processing Systems*, 2018.

[8] R. Kolar, T. Kubena, P. Cernosek, A. Budai, J. Hornegger, J. Gazarek, O. Svoboda, J. Jan, E. Angelopoulou, and J. Odstrcilik. Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. *IET Image Processing*, 7(4):373–383, June 2013.

[9] J. Krause, V. Gulshan, E. Rahimy, P. Karth, K. Widner, G. S. Corrado, L. Peng, and D. R. Webster. Grader Variability and the Importance of Reference Standards for Evaluating Machine Learning Models for Diabetic Retinopathy. *Ophthalmology*, 125(8):1264–1272, 2018.

[10] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool. Deep Retinal Image Understanding. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, LNCS, pages 140–148, 2016.

[11] M. I. Meyer, A. Galdran, P. Costa, A. M. Mendonça, and A. Campilho. Deep Convolutional Artery/Vein Classification of Retinal Vessels. In *Image Analysis and Recognition*, LNCS, pages 622–630, 2018.

[12] M. Niemeijer, Xiayu Xu, P. Gupta, B. van Ginneken, J. C. Folk, and M. D. Abramoff. Automated Measurement of the Arteriolar-to-Venular Width Ratio in Digital Color Fundus Photographs. *IEEE Transactions on Medical Imaging*, 30(11):1941–1950, Nov. 2011.

[13] J. I. Orlando, J. Barbosa Breda, K. van Keer, M. B. Blaschko, P. J. Blanco, and C. A. Bulant. Towards a Glaucoma Risk Index Based on Simulated Hemodynamics from Fundus Images. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, LNCS, pages 65–73, 2018.

[14] J. I. Orlando, E. Prokofyeva, and M. B. Blaschko. A Discriminatively Trained Fully Connected Conditional Random Field Model for Blood Vessel Segmentation in Fundus Images. *IEEE Transactions on Biomedical Engineering*, 64(1):16–27, Jan. 2017.

[15] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21(4):34–41, Aug. 2001.

[16] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, LNCS, pages 234–241, Oct. 2015.

[17] B. Savelli, A. Bria, C. Marrocco, M. Molinara, F. Tortorella, A. Galdran, and A. Campilho. Illumination Correction by Dehazing for Retinal Vessel Segmentation. In *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, 2017.

[18] R. A. Welikala, P. J. Foster, P. H. Whincup, A. R. Rudnicka, C. G. Owen, D. P. Strachan, and S. A. Barman. Automated arteriole and venule classification using deep learning for retinal images from the UK Biobank cohort. *Computers in Biology and Medicine*, 90:23 – 32, 2017.

[19] Z. Yan, X. Yang, and K.-T. Cheng. Joint Segment-Level and Pixel-Wise Losses for Deep Learning Based Retinal Vessel Segmentation. *IEEE Transactions on Biomedical Engineering*, 65(9):1912–1923, Sept. 2018.

[20] Y. Zhao, J. Xie, P. Su, Y. Zheng, Y. Liu, J. Cheng, and J. Liu. Retinal Artery and Vein Classification via Dominant Sets Clustering-Based Vascular Topology Estimation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, LNCS, pages 56–64, 2018.