

Vascular Topology Rectification Network for Automated Retinal Artery/Vein Segmentation

Author:

Heng, Z; Pagnucco, M; Meijering, E; Song, Y

Publication details:

Proceedings International Symposium on Biomedical Imaging
v. 00

pp. 1 - 4

979-8-3503-1334-5 (ISBN)

1945-7928 (ISSN); 1945-8452 (ISSN)

Event details:

2024 IEEE International Symposium on Biomedical Imaging (ISBI)
GREECE, Athens
2024-05-27 - 2024-05-30

Publication Date:

2024-01-01

Publisher DOI:

<https://doi.org/10.1109/ISBI56570.2024.10635229>

License:

<https://creativecommons.org/licenses/by/4.0/>

Link to license to see what you are allowed to do with this resource.

Downloaded from <http://hdl.handle.net/1959.4/106742> in <https://unsworks.unsw.edu.au> on 2026-02-27

VASCULAR TOPOLOGY RECTIFICATION NETWORK FOR AUTOMATED RETINAL ARTERY/VEIN SEGMENTATION

Zhan Heng, Maurice Pagnucco, Erik Meijering, Yang Song

School of Computer Science and Engineering, University of New South Wales, Australia

ABSTRACT

Retinal artery/vein (A/V) segmentation is an important step in screening various diseases. However, accurate retinal A/V segmentation is still a challenging task, and even state-of-the-art methods often fail to respect the correct topology of the vascular structure, which is crucial for clinical application. For better preservation of topology and connectivity in fully automated A/V segmentation, we propose a two-stage deep learning framework that consists of a multi-task U-Net with a customized vessel aggregation module for initial A/V segmentation, and a vascular topology rectification network based on persistent homology. Extensive experiments on public DRIVE-AV and HRF-AV datasets show that the proposed method delivers competitive pixel-level accuracy and the best topology-level accuracy against state-of-the-art methods.

Index Terms— retinal artery/vein segmentation, fundus image, persistent homology

1. INTRODUCTION

Retinal imaging enables non-invasive observation of human fundi. Particularly, morphological changes in retinal artery/vein (A/V) could be indicators of various pathological conditions. For example, retinal A/V fractal dimension (FD) is linked to lower visual acuity [1]; diabetic retinopathy often leads to narrowing of retinal veins, quantifiable as the arteriolar-to-venous ratio (AVR) [2]. Since manual A/V annotations can be prohibitively time-consuming while requiring expertise of ophthalmologists, accurate automated segmentation could be highly beneficial for computer-aided diagnosis and large-scale screening of relevant diseases.

With recent advances in deep learning, methods based on fully convolutional networks (FCN) have attained decent pixel-level performance in A/V segmentation [3, 4]. However, these earlier works typically restrict a pixel to have a single label [3], or treat pixels at A/V crossings as “uncertain” class [4], prohibiting the identification of A/V crossings, in which case the vascular topology could still be misrepresented despite achieving good pixel-wise accuracies. More recent studies on A/V segmentation tend to focus more on the preservation of vascular topology. [5] integrates domain knowledge of ophthalmology by adopting a multi-input mod-

ule and a spatial activation mechanism. [6, 7, 8, 9] use adversarial learning to guide the networks to focus on topological information of A/V. Among others, TW-GAN [9], the state-of-the-art method on A/V segmentation, includes a topology ranking discriminator to rank the topology correctness of the ground truth mask, the generated A/V segmentation mask and an intentionally shuffled mask. Besides, methods have also been proposed for the problem of curvilinear structure segmentation, though their applications have not been explored extensively in retinal A/V segmentation. For instance, [10] leverages both spatial attention and channel attention to improve curvilinear structure segmentation. In [11, 12], persistent homology is incorporated into deep neural networks to promote correct topology in segmentation results. [13] proposes cIDice loss which demonstrates that a loss function applied along an object’s centerline improve the topology correctness. However, capturing the curvilinear nature of blood vessels remains a challenge for neural networks, and intra-segment errors and broken segmentation (Fig. 2b), can still be found in segmentation results.

To obtain topologically correct retinal A/V segmentation, we introduce a two-stage deep learning framework (Fig. 1) that aims to make pixel-wise-accurate A/V segmentation in Stage 1, and rectify the topological errors in Stage 2. More specifically, Stage 1 of the framework consists of a multi-task U-Net with a vessel aggregation module that focuses on processing the visual information of fundus images. Stage 2 of the framework is a U-Net with dual decoders that incorporates persistent homology [14], a topological feature extraction method in computational topology, to improve the connectivity of the initial A/V segmentation result. Extensive experiments are conducted to evaluate the proposed method on the public DRIVE-AV[15] and HRF-AV[3] datasets. The proposed method achieves state-of-the-art performance in both pixel-level and topology-level metrics.

2. METHODS

2.1. Multi-Task U-Net with Vessel Aggregation Module

The objective of Stage 1 is to perform an initial A/V segmentation on the input fundus image. While the U-Net architecture [16] has been widely adopted for image segmenta-

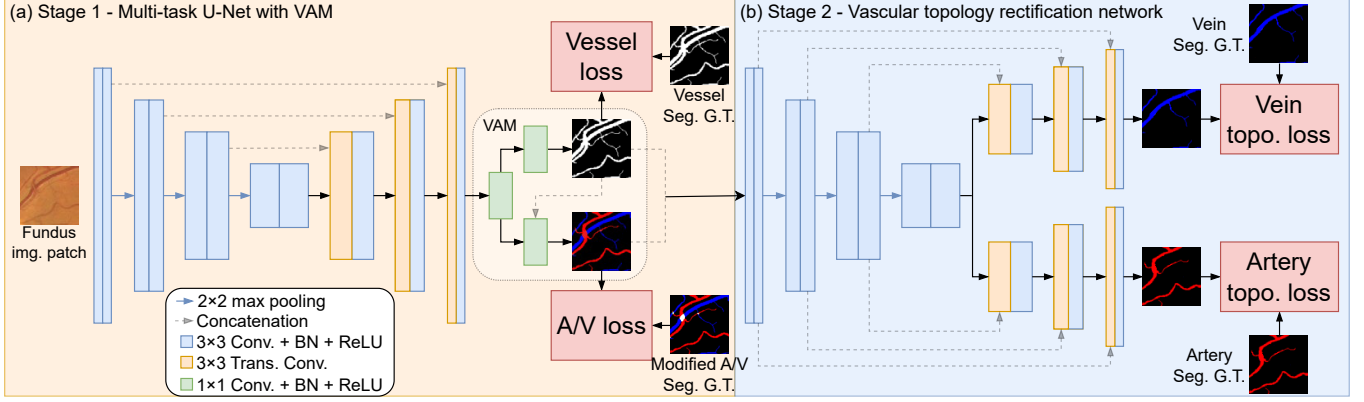


Fig. 1: An overview of the proposed method. (a) Stage 1 - multi-task U-Net with vessel aggregation module for initial A/V segmentation; (b) Stage 2 - Vascular topology rectification network that rectifies the topological errors in results from Stage 1.

tion, including retinal vessel segmentation and A/V segmentation, we propose to adopt a multi-task U-Net facilitated by the proposed vessel aggregation module (VAM) for better results. Besides the main task of A/V segmentation, an auxiliary task of “vessel segmentation” is set up, where pixels are simply classified into vessel/non-vessel without distinguishing between artery and vein. In this way, the easier auxiliary task will act as additional supervision by joint training, and also assist the main task of A/V segmentation directly by providing an a priori vessel segmentation map.

In VAM, the feature map from the final layer of the U-Net goes through different branches of 1×1 convolution layers, batch normalization and ReLU to obtain the vessel segmentation map and A/V segmentation maps. The vessel segmentation map is concatenated with the intermediate feature maps to generate the A/V segmentation maps. Since the segmentation task is highly imbalanced, a binary cross entropy (BCE)-Dice loss is calculated to represent the vessel segmentation loss. Similarly, the A/V segmentation loss is the BCE-Dice loss averaged over artery and vein probability maps. Moreover, we apply a weight map to the A/V segmentation loss function, so foreground (A/V) pixels, background (non-vessel) pixels and uncertain pixels (see Section 2.3) would have weights of 1, 0.1 and 0 respectively. Therefore, uncertain pixels will not intervene with training, and more focus is given to foreground pixels.

2.2. Vascular Topology Rectification Network with Expanded Topological Focal Loss

The initial A/V segmentation maps generated in Stage 1 still contain broken segments and intra-segment errors due to the sole focus on pixel-level accuracy. The vascular topology rectification network, i.e., Stage 2 of the proposed framework, takes the initial vessel and A/V segmentation probability maps from Stage 1 as input, and attempts to further refine them with a focus on rectifying topological errors. For this,

we adopt a U-Net with dual decoders that output arteries and veins probability maps separately. It is important to carefully design a loss function that can pinpoint the pixels that would affect the vascular topology. Therefore, we refer to the concept of persistent homology in computational topology.

Persistent Homology. We use S_p to denote the binary segmentation mask obtained from a segmentation probability map $\mathbf{X}$ given a real-valued threshold p . Sweeping p from 1 to 0 ($1 \geq p_1 \geq p_2 \geq \dots \geq 0$) will result in a sequence of segmentation masks that have monotonically growing foreground pixels:

$$\emptyset \subseteq S_1 \subseteq S_{p_1} \subseteq S_{p_2} \subseteq \dots \subseteq S_0 \quad (1)$$

As p varies, foreground pixels in S_p will collectively form different structures, giving rise to new topological structures and causing existing ones to disappear. Based on the idea of persistent homology [14], we record the 1-dimensional Betti number β_1 , i.e. the number of holes, in the binary segmentation masks after padding the borders with foreground pixels. When β_1 changes, we record the pixels that caused the change in the topological structures as the “critical pixels”. To account for noise, critical pixels with probabilities that are too close to adjacent pixels (where the maximal difference is no more than 0.03) are filtered out. A loss function based on these critical pixels essentially guides the network to perform better at topologically critical locations [11]. However, these critical pixels are usually isolated pixels, which can make training inefficient. To improve on this, we propose an expanded topological focal loss for A/V segmentation.

Expanded Topological Focal Loss for A/V Segmentation. It is observed that in the initial A/V prediction map generated in Stage 1, the broken areas (e.g. A/V crossings) are usually relatively small in size, but larger than a single pixel. Therefore, it could be argued that the small neighborhoods around critical pixels are important for correct connectivity as well. In order to speed up training and make better use of the relatively computationally expensive persistent homology,

Table 1: Left: A/V segmentation performance on DRIVE-AV and HRF-AV datasets compared with state-of-the-art methods. Right: Ablation study results of the proposed method on DRIVE-AV dataset. Best performance marked in bold.

Datasets	DRIVE-AV				HRF-AV				Dataset	DRIVE-AV			
Methods	Sen	F1	PR	Betti	Sen	F1	PR	Betti	Modules	Sen	F1	PR	Betti
U-Net [16]	53.6	65.6	65.9	16.9	63.4	67.1	72.2	11.7	U-Net	53.6	65.6	65.9	16.9
CS-Net [10]	52.0	63.6	63.0	-	62.5	66.1	69.2	-	+VAM	59.7	69.8	85.2	16.2
MNNSA [5]	59.8	61.8	65.4	-	62.8	64.3	69.1	-	Trained with complete G.T.	53.2	62.0	76.2	18.8
TR-GAN [6]	67.8	65.6	67.4	-	63.1	67.5	72.5	-	+U-NetDD	77.3	72.8	86.2	12.8
BMFN [7]	69.9	70.0	71.2	14.5	67.7	71.7	74.0	8.8	+TopoLoss [11]	79.4	70.9	82.1	12.3
TW-GAN [9]	78.2	66.3	65.6	13.3	76.4	75.7	82.0	4.7	+ETFL	80.3	76.0	84.7	11.9
Proposed	80.3	76.0	84.7	11.9	80.1	75.0	81.1	3.1					

we expand the critical pixel set to the union of k -hop neighborhoods of all critical pixels. Here k is empirically set to 2, resulting in 5×5 pixels critical patches. A focal loss $\mathcal{L}_{et}$ is computed over the expanded set of critical pixels $\mathcal{C}$ to focus even more on the hard-to-classify critical pixels.

$$\mathcal{L}_{et} = \frac{1}{|\mathcal{C}|} \sum_{(i,j) \in \mathcal{C}} \alpha(1 - \exp(-\mathcal{H}^{(i,j)}))^{\gamma} \mathcal{H}^{(i,j)} \quad (2)$$

$\mathcal{H}^{(i,j)}$ denotes the BCE loss between ground truth segmentation labels and output segmentation probability map, at position (i, j) . In practice, the expanded topological loss is calculated on non-overlapping patches, separately for arteries and veins and their average is used for training, with α and γ set at 1 and 0.5 empirically. The A/V segmentation BCE-Dice loss in Stage 1 is also used in combination to ensure the overall pixel-level correctness of the result.

2.3. Modified A/V Ground Truth for Training

It is observed that a standard U-Net, or even the improved U-Net proposed in Section 2.1, tends to perform poorly when trained with the A/V ground truth with all the A/V crossings properly labelled. This is expected, as pixels at vessel crossings present visual features of the vessel type on top, but annotations would suggest they should be labelled as both artery and vein, which could confuse the model. To solve this problem, when training Stage 1 we modify the A/V crossing labels in ground truth to the uncertain class, which are omitted in loss calculation. This encourages Stage 1 to focus more on visual information and classify pixels at vessel crossings into the type of vessel on top (Fig. 2). For Stage 2 we train with the complete ground truth to improve vascular connectivity. In this way, visual information and topological information are disentangled to an even greater extent.

3. EXPERIMENTS

Datasets and Implementation. The proposed method is tested on two publicly available datasets. DRIVE-AV[15] contains 40 fundus images with a resolution of 565×584

pixels with a 45° field of view (FOV). The 40 images are equally divided into a training set and a testing set. The original split is used to ensure comparability. HRF-AV[3] contains 45 fundus images with a resolution of 3504×2336 pixels and a 60° FOV, including 15 images of healthy subjects, 15 images of diabetic retinopathy and 15 images of glaucoma. As encouraged by [3], the first five images from each sub-category are used for evaluation, and the rest are used for training. We use the A/V ground truth provided by [9] for correct annotations around vessel crossings.

For training, 384×384 pixels patches are randomly extracted from images. Random rotations, flippings, and Contrast Limited Adaptive Histogram Equalization (CLAHE) are applied. Our implementation uses PyTorch 1.9 and GUDHI. Both stages are optimized with the Adam optimizer and a batch size of 4, with a learning rate of 0.005. When training Stage 2, the proposed expanded topological focal loss takes effect from the 30th epoch. Training typically converges after 1,500 epochs for Stage 1 and 100 epochs for Stage 2. At the inference phase, 384×384 pixels patches are extracted with a stride of 192 pixels, which are stitched together and the averages are taken for overlaps.

Evaluation Metrics. For evaluation, we follow the evaluation scheme in [7] and calculate the sensitivity, F1-score and Area Under the Curve of the Precision-Recall Curve (PR) separately for arteries and veins, and report the weighted scores based on pixel counts. Furthermore, to assess the topological correctness of the results, we also report the Betti number error [11] of the proposed method, which measures the topological consistency between the ground truth and prediction.

Experimental Results. Table 1 (left) summarizes the quantitative evaluation results of our proposed method, compared with existing methods. The competing methods includes recent methods achieving state-of-the-art pixel-level performance [5, 6, 7, 9]. We report the results in [7] where available, and perform evaluation on the other methods if pretrained models are publicly released. On the DRIVE-AV dataset, our proposed method outperforms competing methods in all metrics, some by a large margin. On the HRF-AV dataset, our proposed method achieves the highest sensitivity in our comparison, improving the second place by nearly 4%.

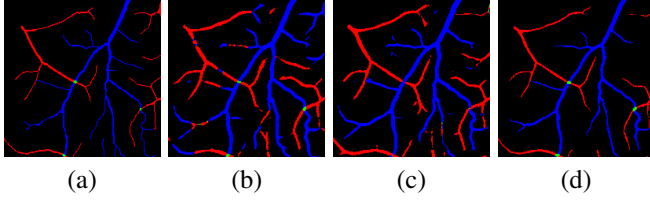


Fig. 2: Illustration of different training strategies. (a) Ground truth A/V annotations; (b) Initial A/V segmentation trained with complete labels (including vessel crossings); (c) Initial A/V segmentation if trained with modified labels (excluding vessel crossings); (d) Rectified result from Stage 2.

TW-GAN [9] still obtained the best F1 and AUC-PR, but the advantage over our method is limited. The proposed method achieves the lowest Betti number error on both datasets, which corresponds to the best ability to preserve the correct vascular topology and connectivity.

Fig. 2 shows some visualization of results produced by the proposed method. As can be seen in (b), when trained with complete A/V ground truth, the initial A/V segmentation identifies some vessel crossings, but struggles to generate continuous vessel segments. If vessel crossing pixels are excluded from training, the output in (c) is significantly improved except for the false negatives around vessel crossings. Fig. 2(d) demonstrates that the proposed topology rectification network in Stage 2 improves vascular topology and eliminates noisy artifacts.

Ablation Study. We investigate the individual influence and effectiveness of the proposed modules. As listed in Table 1 (right), the inclusion of VAM brings a consistent improvement in all metrics compared to a baseline U-Net. When trained with the complete ground truth labels, almost all metrics drop considerably. While attaching a U-Net with dual decoders (U-NetDD) as described in Section 2.1 improves the performance, the proposed expanded topological focal loss (ETFL) for A/V segmentation further enhances both pixel-wise and topological correctness, outperforming the Topo loss [11] which is also based on persistent homology.

4. CONCLUSION

This paper proposes a novel method for retinal A/V segmentation that focuses on visual information and vascular topological information in two cascaded stages. Experiments on two public datasets have verified the proposed method’s state-of-the-art performance, showing accurate pixel-level segmentation and more topologically correct retinal vascular topology. In future work, we will work on enforcing explicit topological constraints for A/V classification, and explore the possibility of extending the application to other segmentation tasks.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research used public data, ethical approval was not required as confirmed by the license attached with the data.

6. REFERENCES

- [1] T. E. Tan, et al., “Global Assessment of Retinal Arteriolar, Venular and Capillary Microcirculations Using Fundus Photographs and Optical Coherence Tomography Angiography in Diabetic Retinopathy,” *Sci. Rep.*, vol. 9, no. 1, pp. 11751, 2019.
- [2] R. Klein, et al., “The Relation of Retinal Vessel Caliber to the Incidence and Progression of Diabetic Retinopathy: XIX: The Wisconsin Epidemiologic Study of Diabetic Retinopathy,” *Arch. Ophthalmol.*, vol. 122, no. 1, pp. 76–83, 2004.
- [3] R. Hemelings, et al., “Artery–vein segmentation in fundus images using a fully convolutional network,” *Computerized Medical Imaging and Graphics*, vol. 76, pp. 101636, 2019.
- [4] A. Galdran, et al., “Uncertainty-Aware Artery/Vein Classification on Retinal Images,” *ISBI*, pp. 556–560, 2019.
- [5] W. Ma, et al., “Multi-task Neural Networks with Spatial Activation for Retinal Vessel Segmentation and Artery/Vein Classification,” *MICCAI*, pp. 769–778, 2019.
- [6] W. Chen, et al., “TR-GAN: Topology Ranking GAN with Triplet Loss for Retinal Artery/Vein Classification,” *MICCAI*, pp. 616–625, 2020.
- [7] Y. Zhou, et al., “Learning to Address Intra-segment Misclassification in Retinal Imaging,” *MICCAI*, pp. 482–492, 2021.
- [8] S. Luo, Z. Heng, M. Pagnucco, and Y. Song, “Two-Stage Topological Refinement Network for Retinal Artery/Vein Classification,” *ISBI*, pp. 1–4, 2022.
- [9] W. Chen, et al., “TW-GAN: Topology and width aware GAN for retinal artery/vein classification,” *Med. Image Anal.*, vol. 77, pp. 102340, 2022.
- [10] L. Mou, et al., “CS-Net: Channel and Spatial Attention Network for Curvilinear Structure Segmentation,” *MICCAI*, pp. 721–730, 2019.
- [11] X. Hu, F. Li, D. Samaras, and C. Chen, “Topology-Preserving Deep Image Segmentation,” *NeurIPS*, 2019.
- [12] J. R. Clough, et al., “Explicit Topological Priors for Deep-Learning Based Image Segmentation Using Persistent Homology,” *IPMI*, pp. 16–28, 2019.
- [13] S. Shit et al., “clDice - a Novel Topology-Preserving Loss Function for Tubular Structure Segmentation,” *CVPR*, pp. 16560–16569, 2021.
- [14] N. Otter, et al., “A roadmap for the computation of persistent homology,” *EPJ Data Science*, vol. 6, no. 1, pp. 1–38, 2017.
- [15] Q. Hu, M. D. Abràmoff, and M. K. Garvin, “Automated Separation of Binary Overlapping Trees in Low-Contrast Color Retinal Images,” *MICCAI*, pp. 436–443, 2013.
- [16] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *MICCAI*, pp. 234–241, 2015.