

Retinal Vascular Network Reconstruction and Classification via Dominant Sets Clustering

Journal:	<i>Transactions on Medical Imaging</i>
Manuscript ID	TMI-2019-0468
Manuscript Type:	Full Paper
Date Submitted by the Author:	29-Apr-2019
Complete List of Authors:	<p>Zhao, Yitian; Chinese Academy of Sciences, Ningbo Institute of Industrial Technology</p> <p>Xie, Jianyang; Chinese Academy of Sciences, Ningbo Institute of Industrial Technology</p> <p>Zhang, Huaizhong; Edge Hill University, Department of Computer Science</p> <p>Zheng, Yalin; University of Liverpool, Department of Eye and Vision Science</p> <p>Zhao, Yifan; Cranfield University, EPSRC Centre for Innovative Manufacturing in Through-life Engineering Services</p> <p>Qi, Hong; Peking University Third Hospital, Department of Ophthalmology</p> <p>Zhao, Yangchun; Second Affiliated Hospital of Zhejiang Chinese Medical University, Department of Ophthalmology</p> <p>Su, Pan; Chinese Academy of Sciences, Ningbo Institute of Industrial Technology</p> <p>Liu, Jiang; Ningbo Institute of Material Technology and Engineering, Chinese Academy of Sciences, iMED</p> <p>Liu, Yonghuai; Edge Hill University, Department of Computer Science</p>
Keywords:	Optical Imaging/OCT/DOT < Imaging modalities, Eye < Object of interest, Vessels < Object of interest, Pattern recognition and classification < General methodology, Shape analysis < General methodology
Specialty/Area of Expertise:	

Retinal Vascular Network Reconstruction and Classification via Dominant Sets Clustering

Yitian Zhao, Jianyang Xie, Huaizhong Zhang, Yalin Zheng, Yifan Zhao, Hong Qi, Yangchun Zhao, Pan Su*, Jiang Liu and Yonghuai Liu*

Abstract—The estimation of vascular network topology in complex networks is important in understanding the relationship between vascular changes and a wide spectrum of diseases. Automatic classification of the retinal vascular trees into arteries and veins is of direct assistance to the ophthalmologist in terms of diagnosis and treatment of eye disease. However, it is challenging due to their projective ambiguity and subtle changes in appearance, contrast and geometry in the imaging process. In this paper, we propose a novel method that is capable of making the artery/vein (A/V) distinction in retinal color fundus images based on vascular network topological properties. To this end, we adapt the concept of *dominant sets clustering* and formalize the retinal blood vessel topology estimation and the A/V classification as a pairwise clustering problem. The graph is constructed through image segmentation, skeletonization and identification of significant nodes. The edge weight is defined as the inverse Euclidean distance between its two end points in the feature space of intensity, orientation, curvature, diameter, and entropy. The reconstructed vascular network is classified into arteries and veins based on their intensity and morphology. The proposed approach has been applied to five public databases (INSPIRE, IOSTAR, VICAVER, DRIVE and WIDE) and achieved high accuracies of 95.1%, 94.2%, 93.8%, 91.1%, and 91.0%, respectively. Furthermore, we have made manual annotations of the blood vessel topologies for INSPIRE, IOSTAR, VICAVER, and DRIVE databases, and these annotations are released for public access so as to facilitate researchers in the community.

Index Terms—Retinal images, dominant sets clustering, blood vessel, vascular topology, Artery/vein classification

I. INTRODUCTION

The retinal blood vascular network is the only vascular network of the human body that is visible to a non-invasive imaging approach. In consequence, automated analysis of retinal vascular structure is the most common way to support examination, diagnosis and treatment of many diseases [1], [2],

This work was supported by National Science Foundation Program of China (61601029), Zhejiang Provincial Natural Science Foundation (LZ19F010001), and Ningbo Natural Science Foundation (2018A610055).

Y. Zhao, J. Xie, P. Su, and J. Liu are with Ningbo Institute of Industrial Technology, Chinese Academy of Sciences, China; H. Zhang and Y. Liu are with Department of Computer Science, Edge Hill University, UK; Y. Zheng is with Department of Eye and Vision Science, University of Liverpool, UK; H. Qi is with Department of Ophthalmology, Peking University Third Hospital, China; YF. Zhao is with School of Aerospace, Transport and Manufacturing, Cranfield University, UK; YC. Zhao is with the Second Affiliated Hospital of Zhejiang Chinese Medical University.

J. Liu is also with Department of Computer Science and Engineering, Southern University of Science and Technology, China

P. Su is also with School of Control and Computer Engineering, North China Electric Power University, China

* denotes corresponding author: supan@nimte.ac.cn, and yonghuai.liu@edgehill.ac.uk

such as diabetic retinopathy, hypertension and other cardiovascular diseases [3], [4]. Retinal arteriolar constriction, or arteriovenous nicking, significant dilatation and elongation of main arteries and veins, or vascular caliber and tortuosity change are frequently associated with the aforementioned diseases [5]. It is crucial to identify and distinguish the structures of individual blood vessels from the entire blood vessel network in a given fundus image. In particular, the Arteriolar-to-Venular Ratio (AVR) is considered to be an important characteristic sign that quantifies the severity of a wide spectrum of diseases [1], [6], [7], for example, low AVR - i.e., narrowing of arteries and widening of veins - is a direct biomarker for diabetic retinopathy. By contrast, a high AVR has been associated with higher cholesterol levels and inflammatory markers [8]. For these clinical observations, it would be of direct benefit if the retinal vascular network could be distinguished anatomically into different blood vessel branches, or separated into arteries and veins.

In practice, ophthalmologists use color and morphological information to discriminate between arteries and veins, since the arteries contain more oxygen and appear brighter than the veins, and thinner than neighboring veins [9]. These features of the retinal vasculature are usually captured by fundus photography, due to its lower cost and ease of use. Manual classification of retinal blood vessels is time consuming and subject to human errors. Therefore, an automatic vascular tracing method for topology reconstruction, or classification of blood vessels as arteries and veins, is highly desirable in seeking to overcome time constraints and avoid human errors. This calls for precise description of the vascular structure in terms of its color, topological, geometrical and morphological properties as derived from retinal images. However, it is highly challenging to discriminate arteries from veins, or perform any other measurement of interest (e.g., topology estimation), if the blood vessel widths are small, even after accurately segmenting the vasculature from the given fundus images. For example, existing methods often fail to trace correctly when faced with the occurrence of bifurcation and crossover at junction points: (i) the bifurcation – different blood vessel segments are from one blood vessel tree, and (ii) crossovers – two blood vessels overlap due to the projection of a 3D human eye to a 2D fundus image. These intersections usually lead to difficulty in predicting whether given blood vessels contacting a junction belong to the same blood vessel tree, or a different tree, due to their projective ambiguity and subtle changes in appearance, contrast and geometry in the imaging process.

To address these problems, we propose a novel topology

estimation and arteries/veins (A/V) classification method by adapting the concept of dominant sets clustering (DOS), in which a dominant set is used to represent a vascular tree. The novel method includes three main steps: (i) we re-conceptualize our previous work [10] and extend it for both the retinal blood vessel topology estimation and classification, (ii) formalize them as a pairwise clustering problem, and (iii) validate the proposed method over five public retinal datasets. DOS is a graph-theoretic approach, and is a well-known generalization of the notion of maximal cliques to an edge-weighted graph. It has been proven to work well in data clustering and image segmentation [11]–[13]. We aim to classify the entire retinal blood vessel network, not just the most prominent blood vessels, with a view to clarifying the underlying topology - how different blood vessels are anatomically connected to each other - and to identify and distinguish the structure of individual blood vessels from the entire blood vessel network. To classify the reconstructed vascular network into arteries and veins, we consider domain-specific knowledge about how they perfuse the retina, including angular, morphological, and textural properties of all blood vessel segments of the junction. We also take into account that arteries and veins usually alternate near the optic disc.

The main contributions of this paper are as follows:

- The concept of dominant sets clustering has been introduced to tackle the challenging problem of vasculature analysis, and proved to be an effective way of addressing the problem of tracing crossovers.
- The proposed method can split the entire blood vessel graph into several individual branches as subtrees, and is capable of demonstrating how different blood vessels are anatomically connected to each other. In addition, the A/V classification is undertaken on the topology-assigned blood vessel network, rather than the entire blood vessel segments.
- The proposed method has been validated quantitatively using five publicly accessible datasets, with promising results. In addition, the manual annotations of blood vessel topologies of four datasets were established as the ground truth, and have been released for public access¹.

The remainder of this paper is organized as follows. In Section II, the existing methods for retinal A/V classification and topology reconstruction are briefly reviewed. Section III details the proposed blood vessel topology estimation and A/V classification method. Section IV introduces the datasets, ground truth and evaluation metrics we will use. In Sections V and VI, we present the experimental results and discuss the effectiveness, robustness, and efficiency respectively of the proposed method. Finally, in Section VII we draw conclusions, and indicate directions for future work.

II. RELATED WORKS

In the past decade, extensive work has been carried out on automatic retinal blood vessel segmentation [14]–[18]. However, automated A/V classification and vascular topology

reconstruction are still understudied, despite their significance to understanding the structure and distribution of the blood vessels. The majority of existing blood vessel classification methods make use of machine learning techniques, using local features and adding structural information from the vascular tree.

Martinez-Perez et al. [19] proposed a semi-automatic retinal blood vessel analysis method that is capable of measuring and quantifying the geometrical and topological properties. It requires a human expert first to classify a branch as either vein or artery. Kondermann et al. [20] classified blood vessels by utilizing the Support Vector Machines (SVM) and neural networks. Relevant features were extracted from the blood vessel profile and the regions of interest centred at the pixels on the blood vessel centerlines. Niemeijer et al. [1], [21] found the k-Nearest Neighbour (kNN) classifier performing the best, by testing four different classifiers in the task of distinguishing between arteries and veins from a feature vector containing 27 elements. These elements were generated from blood vessel centerline pixels: a soft label was assigned to each centerline, implying the likelihood of the blood vessel being a vein. Mirsharif et al. [22] classified the blood vessels into arteries and veins by using multiple classifiers, and found that the Linear Discriminant Analysis (LDA) classifier had the best performance. Relan et al. [23] automatically classified retinal blood vessels as arteries or veins using color features and a Gaussian Mixture Model (GMM). Vazquez et al. [3] combined color-based clustering and blood vessel tracking to differentiate arteries from veins, and the tracking strategy based on the minimal path approach was employed to support the resulting classification by voting. Girard and Chriet [24] trained a convolutional neural network (CNN) for the task of assigning blood vessel pixels into arteries or veins. This approach propagated the blood vessel graph by using the minimum spanning tree. Huang et al. [25] introduced four new features to avoid distortions resulting from lightness inhomogeneity, and the accuracy of the A/V classification was improved by using the LDA classifier.

Due to demanding precise segmentation results in most of the existing methods, the ambiguity of small and mid-sized blood vessels makes the A/V classification a very difficult computational task. Pathological conditions and intensity inhomogeneities further complicate the task of A/V classification. More recently, the analysis of graphs extracted from the retinal blood vessel structure has been utilized to assist in blood vessel classification [8], [9], [26], [27]. This approach classifies the entire vascular tree by determining the type of all intersection pixels (graph nodes) and then assigning a label to each blood vessel segment, so as to reconstruct the underlying blood vessel topology with more accurate classification of small and mid-sized blood vessels. Rothaus et al. [26] presented an automated graph separation algorithm to distinguish between arteries and veins. Dashtbozorg et al. [8] proposed a method which first split the vascular graph into subgraphs by embedding the graph nodes and applying intensity features, and then, using LDA, assigned a label to each subgraph stating whether it was artery or vein. Joshi et al. [9], [27] separated their vascular graph into different subgraphs by using Dijkstra's shortest-

¹<http://imed.nimte.ac.cn/vetovessel-topology-groundtruth.html>

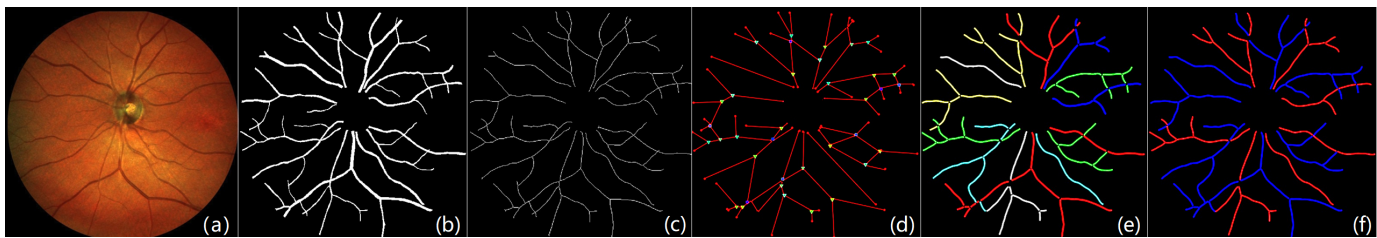


Fig. 1: Overview of the proposed method. (a) Original image. (b) Extracted blood vessels. (c) Skeletonized blood vessels. (d) Graph generated with significant nodes overlaid. (e) Estimated vascular network topology. (f) Classified arteries and veins: arteries shown in red, and veins in blue.

path algorithm, then labeled each subgraph as either artery or vein using a fuzzy C-means clustering algorithm. Estrada et al. [5], [28] utilized a global likelihood model to capture the structural plausibility of each blood vessel, and employed a graph-theoretic method to estimate the overall blood vessel topology with domain-specific knowledge and applied three features, local growth, overlap and color, to accurately classify the A/V types. Lyu et al. [29] used a curvature orientation histogram to extract blood vessel landmarks from the blood vessel centerline, and separated the different blood vessel trees by curvature orientation clustering. De et al. [30], [31] proposed a graph-theoretical approach to reconstruct the blood vessel network from topological information. They adapted the label propagation over directed graphs, and by this method the graph was partitioned into disjoint subgraphs. However, these topology reconstruction or subgraph estimation methods often fail to achieve accurate results due to the difficulty in identifying the presence of crossovers at blood vessel junctions, as the caliber, angular or other measurements are unreliable at points of abrupt change from one blood vessel to another.

III. METHOD

The proposed method adapts the dominant set clustering for retinal topology reconstruction and A/V classification, and the overview of the method is illustrated in Fig. 1. Fig. 1(a) shows an example retinal color fundus image from the public IOSTAR dataset, and Fig. 1(b) illustrates its extracted vasculature. Then a skeletonization step is applied to the extracted blood vessels, as shown in Fig. 1(c). A vessel graph is generated by removing the junction pixels from Fig. 1(c), and Fig. 1(d) depicts the generated graph with significant nodes overlaid. The blood vessel network is separated into several individual tree branches in different colors, as shown in Fig. 1(e). Finally, these tree branches are classified into two categories: arteries (red) and veins (blue). The generation of the graph, the separation and classification of the blood vessel network into arteries and veins are detailed below in Section III-A, Sections III-B and III-C, and Section III-D respectively.

A. Graph generation

Firstly, the optic disc is masked in order to avoid morphological complications due to the irregular and highly tortuous

blood vessels at the optic disc: the superpixel-based optic disc segmentation [32] was utilized for its fast speed and effectiveness. Then the infinite perimeter active contour with hybrid region (IPACHR) method proposed in [33] is employed to automatically segment the retinal blood vessels for its effectiveness in detecting vessels with irregular and oscillatory boundaries, as shown in Fig. 1(b). An iterative morphology thinning operation [34] is finally performed on the extracted blood vessels to obtain a single-pixel-wide skeleton map, as shown in Fig. 1(c).

The vascular bifurcations, crossovers, and blood vessel ends (terminal points) may then be extracted from the skeleton map by locating intersection points (pixels with more than two neighbors) and terminal points (pixels with one neighbor). All the intersection points and their neighbors are then removed from the skeleton map, producing an image with clearly separated blood vessel segments. A blood vessel graph can be generated from this skeleton map by linking the first and last nodes in the same blood vessel segment. The generated graph will usually include some misrepresentations of the blood vessels: typical errors include *node splitting*, *missing links* and *false links*. Correction of these errors was achieved by employing the strategy proposed in [8], which considers the local vessel calibers and angles to reconnect the missing links and correct the false links. Fig. 1(d) shows the blood vessel graph, in which red dots indicate terminal points, green triangles bifurcations, and blue squares crossover points.

The intersections extracted from the skeleton map are critical to the final topology estimation. In the following, the details of partitioning the points in each intersection into different vascular segments are described. The analysis of the intersections is broken down into three categories, according to the number of points involved in each intersection: connecting points (2), bifurcation points (3, 4), and crossovers/meeting points (3, 4, 5), where the number in the brackets after each category indicates the number of vascular segments connected to that intersection. Again, the method proposed by Dashtbozorg et al. [8] is used to handle the cases of nodes of degree 2. For the more complicated categories, nodes of degree 3 to 5, a clustering method based on the dominant sets is proposed as follows.

The concept of dominant sets arises from the study of graph theory, by which a continuous formulation of the maximum clique problem is defined in [11]. An undirected graph G

with weighted edges is represented as $G = (V, E, \omega)$, where V is a set of nodes, edge set $E \subseteq V \times V$ indicates all the connections of the relevant nodes, and $\omega : E \rightarrow \mathbb{R}^+$ is the positive weight function. Fig. 2 shows an example of the edge-weighted graph extracted from a representative junction of a retinal blood vessel network.

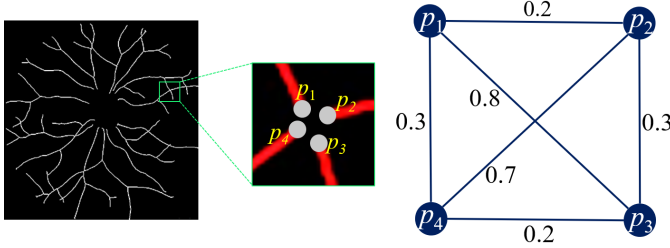


Fig. 2: An example of the edge-weighted graph.

In the context of vascular topology estimation and arteries/veins identification, V is a set of extracted pixels from a retinal color fundus image and ω represents the similarity among the pixels in V . A $|V| \times |V|$ symmetric matrix $A = \{a_{ij}\}$ is used to represent the weighted graph G , which is named an adjacency matrix. The value of a_{ij} is derived by a similarity measure defined in the feature space of the pixels. Here, we define $a_{ij} = 0$ for $i = j$, which indicates that the generated graph G does not include self-loop.

B. Dominant sets

A dominant set can be formally defined based on the values of similarity among the nodes in V . Let $S \subseteq V$ be a nonempty subset of nodes, $p_i \in S$ and $p_j \in S \setminus \{p_i\}$, where $S \setminus \{p_i\}$ indicates the nodes in the set S excluding the node p_i . The relative similarity between p_i and p_j with respect to the average similarity between p_j and its neighbours in $S \setminus \{p_i\}$ can be defined as:

$$\phi_{S \setminus \{p_i\}}(p_i, p_j) = a_{ij} - \frac{1}{|S| - 1} \sum_{p_k \in S \setminus \{p_i\}} a_{jk}. \quad (1)$$

where $|\cdot|$ denotes the cardinality of a set. For example, Fig. 2 depicts an edge-weighted graph generated for an intersection which is extracted from Fig. 1, where the weights of the edges indicate the similarity amongst four pixels $\{p_1, p_2, p_3, p_4\}$. Given $S = \{p_1, p_2, p_3\}$ we have: $\phi_{\{p_1, p_3\}}(p_2, p_1) = a_{21} - (a_{11} + a_{13})/2 = -0.2$ and $\phi_{\{p_1, p_3\}}(p_2, p_3) = a_{23} - (a_{31} + a_{33})/2 = -0.1$. It can be observed that $\phi_S(p_i, p_j)$ may be either positive or negative.

The weight $W_S(p_i)$ of p_i with regard to S is defined recursively as:

$$W_S(p_i) = \begin{cases} 1, & \text{if } |S| = 1 \\ \sum_{p_j \in S \setminus \{p_i\}} \phi_{S \setminus \{p_i\}}(p_i, p_j) W_{S \setminus \{p_i\}}(p_j), & \text{otherwise.} \end{cases} \quad (2)$$

where $W_S(p_i)$ expresses the similarity between node p_i and the nodes of $S \setminus \{p_i\}$ with respect to the mutual similarity amongst the nodes in $S \setminus \{p_i\}$. Finally, the total weight of S is calculated as $W(S) = \sum_{p_i \in S} W_S(p_i)$. Take the calculation of $W_{\{p_1, p_3, p_4\}}(p_4)$ as an example (shown in Fig. 3), we have:

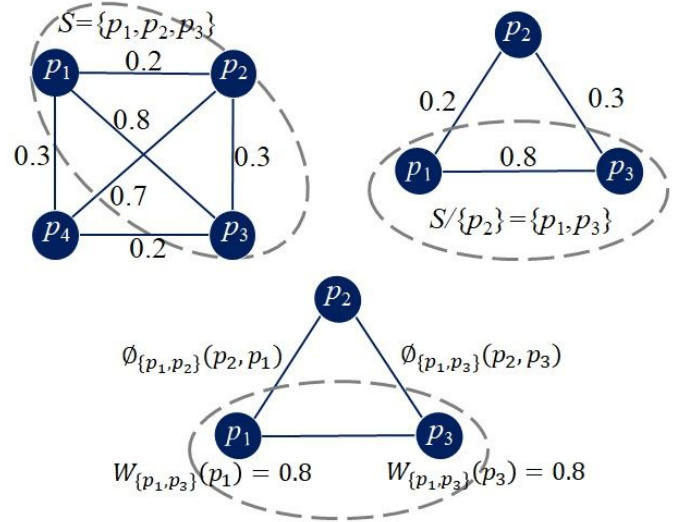


Fig. 3: An example of decomposed recursive calculation of similarity over a graph for dominant sets.

$$W_{\{p_1, p_3\}}(p_1) = W_{\{p_1, p_3\}}(p_3) = 0.8, W_{\{p_1, p_2, p_3\}}(p_2) = \phi_{\{p_1, p_3\}}(p_2, p_1)W_{\{p_1, p_3\}}(p_1) + \phi_{\{p_1, p_3\}}(p_2, p_3)W_{\{p_1, p_3\}}(p_3) = -0.24, \text{ and } W_{\{p_1, p_3, p_4\}}(p_4) = -0.24.$$

Formally, a non-empty subset S of nodes V , $S \subseteq V$ such that $W(S') > 0$ for any non-empty subset $S' \subseteq S$ is said to be a dominant set if:

$$W_{S'}(p_i) > 0, \text{ for all } p_i \in S' \quad (3)$$

and

$$W_{S' \cup \{p_j\}}(p_j) < 0, \text{ for any } p_j \notin S'. \quad (4)$$

Take the graph shown in Fig. 3 as an example: $W(\{p_1, p_3\}) = W_{\{p_1, p_3\}}(p_1) + W_{\{p_1, p_3\}}(p_3) = 1.6$, $W(\{p_1\}) = W(\{p_3\}) = 1$, and $W_{\{p_1, p_3\}}(p_1) > 0$, $W_{\{p_1, p_3\}}(p_3) > 0$, $W_{\{p_1, p_2, p_3\}}(p_2) < 0$, $W_{\{p_1, p_3, p_4\}}(p_4) < 0$. Therefore, p_1 and p_3 form a dominant set. Similarly, p_2 and p_4 form another dominant set.

In general, the weights of edges within the dominant set of an edge-weighted graph should be large, representing high internal homogeneity or similarity [11]. By contrast, the weights of edges which link to the dominant sets externally will be small. Therefore, the dominant set is a proper solution to identify branches of a vascular tree, because the similarity of two points from the same branch should be large within an intersection, while that of two points belonging to different branches should be small.

Dominant sets can be identified by local solutions of a standard quadratic program:

$$\begin{aligned} & \text{maximize}_{\mathbf{x}} && f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x} \\ & \text{subject to} && \mathbf{x} \in \Delta, \end{aligned} \quad (5)$$

where

$$\Delta = \left\{ \mathbf{x} \in \mathbb{R}^{|V|} : \sum_{i=1}^{|V|} x_i = 1 \text{ and } x_i \geq 0 \text{ for all } i = 1, \dots, |V| \right\}.$$

A strict local solution \mathbf{x}^* of Eqn. (5) is named the weighted characteristic vector, where $x_i > 0$ means that the node p_i in

Algorithm 1: RDDOS(A , MaxIteration)

Inputs: A : the symmetric matrix $A = \{a_{ij}\}$ of a weighted graph G ;
Outputs: D : a set of nodes which forms a dominant set of G ;

- 1: initialize a vector $\mathbf{x}^{(0)} \in \mathbb{R}^{|V|}$ which satisfies: $x_i^{(0)} \geq 0$ for all $i = 1, \dots, |V|$ and $\sum_{i=1}^{|V|} x_i^{(0)} = 1$
- 2: $D = \emptyset$
- 3: **for** $t = 0 : \text{MaxIteration} - 1$ **do**
- 4: **for** $i = 1 : |V|$ **do**
- 5: $x_i^{(t+1)} = x_i^{(t)} \frac{(A\mathbf{x}^{(t)})_i}{\mathbf{x}^{(t)\top} A\mathbf{x}^{(t)}}$
- 6: **end for**
- 7: **end for**
- 8: **for each** $x_i^{(\text{MaxIteration})} > 0$ **do**
- 9: $D = D \cup p_i$
- 10: **end for**

question is in a dominant set of G . An effective optimization approach for solving Eqn. (5) is given by the *replicator dynamics* [12], [35]:

$$x_i^{(t+1)} = x_i^{(t)} \frac{(A\mathbf{x}^{(t)})_i}{\mathbf{x}^{(t)\top} A\mathbf{x}^{(t)}}, \quad (6)$$

where $i = 1, 2, \dots, |V|$. It has been proven that as t in Eqn. (6) increases, for any initialization of $\mathbf{x}^{(0)} \in \Delta$, the trajectory remains in Δ and the objective function $f(\mathbf{x})$ in Eqn. (5) is either strictly increasing or constant. Therefore, $\mathbf{x}^{(0)}$ can be initialized by generating $|V|$ random numbers and then normalizing them. In the following algorithms, we use $S = \text{RDDOS}(A, \text{MaxIteration})$ in Algorithm 1 to represent the procedure of extracting a dominant set S from a weighted graph given its adjacency matrix A by using replicator dynamics. The stopping criterion of the dynamic system can be set with a maximal iteration number MaxIteration . In this paper, MaxIteration is empirically set to 30 in the experiments. Take the graph shown in Fig. 3 as an example, the matrix A is:

$$\begin{pmatrix} 0.0 & 0.2 & 0.8 & 0.3 \\ 0.2 & 0.0 & 0.3 & 0.7 \\ 0.8 & 0.3 & 0.0 & 0.2 \\ 0.3 & 0.7 & 0.2 & 0.0 \end{pmatrix},$$

given that $\mathbf{x}^{(0)}$ is initialized as $(0.4, 0.3, 0.2, 0.1)$, after the first iteration of the replicator dynamics defined in Eqn. (6), $\mathbf{x}^{(1)} = (0.35, 0.22, 0.30, 0.13)$. \mathbf{x} is finally converged to $(0.5, 0.0, 0.5, 0.0)$ after 15 iterations, and we have $x_1 > 0$ and $x_3 > 0$ which indicates that p_1 and p_3 form a dominant set, as expected.

C. Topology estimation via dominant sets clustering

A peeling-off strategy has been proposed in Algorithm 2, which shows the complete procedure of partitioning the points involved in the intersections into branches. It iteratively extracts a subset of points belonging to the same branch (a dominant set S) each time by using Eqn. (6) and repeating the process with the remaining points in the set $V = V \setminus S$. The identification of different vascular branches at an intersection is carried out by identifying one vascular branch first, and then

Algorithm 2: TopologyEstimation(I)

Inputs: V : a set of points that is associated with intersections;
Outputs: S^* : a partition of V , each element of which includes points of the same branch;

- 1: initialize the $|V| \times |V|$ symmetric matrix $A = \{a_{ij}\}$ by calculating $a_{ij} = \omega(i, j)$ with respect to given features;
- 2: $S^* = \emptyset$
- 3: **while** $V \neq \emptyset$ **do**
- 4: $S = \text{RDDOS}(A, 30)$
- 5: **for each** $p_i \in S$ **do**
- 6: remove the column and row with respect to p_i from A
- 7: **end for**
- 8: $V = V \setminus S$
- 9: $S^* = S^* \cup \{S\}$
- 10: **end while**

identifying another from the remaining nodes and segments, and so on. Therefore, the peeling-off strategy is a direct, intuitive implementation of this procedure [13]. As shown in the aforementioned example of Fig. 3, given that $\mathbf{x}^{(0)}$ is initialized as $(0.4, 0.3, 0.2, 0.1)$, the first iteration of Algorithm 2 identifies the dominant set including p_1 and p_3 , subsequently the second iteration identifies another including p_2 and p_4 , as expected.

It is also worth noting that the number of points processed in the process of identifying a dominant set is greatly reduced over different iterations, from the number of intersections in a vascular network after the pre-processing step proposed in Section III to those excluding the vascular branches already identified. This leads to an increase of computational efficiency in the topology estimation.

Each point p_i in V is represented as a feature vector \mathbf{F}_i : its intensities in R , G , and B channels, orientations, curvatures, blood vessel diameters, and entropy. The values of each feature are then linearly normalized individually into the unit interval $[0, 1]$. The weight $\omega(i, j)$ of an edge connecting points p_i and p_j is finally estimated as: $\omega(i, j) = 1/(\|\mathbf{F}_i - \mathbf{F}_j\| + \epsilon)$, where $\|\cdot\|$ denotes the Euclidean distance between \mathbf{F}_i and \mathbf{F}_j . Note that a tiny value $\epsilon = 0.000001$ is used in case of $\mathbf{F}_i = \mathbf{F}_j$.

For each pixel in a vessel segment, 23 features are measured which are listed in TABLE I and are mainly related to the orientation, diameter, and curvature. Most of these features were previously used for topological reconstruction [8], [9] and vessel classification [1], [20] tasks. While other features can be calculated straightforward, the main ones are defined and explained as follows: the orientation of a pixel is defined as the including angle (in radians) between its lying segment and the positive direction of the x -axis, a measurement between $[0, \pi]$ [9]; the diameter of a pixel is measured as a median value of 15 measurements of diametric length between the vessel edges, and passing through the skeleton pixels of the end region [9]; the curvature C at pixel p is computed as $\frac{\Delta x_p \Delta^2 y_p - \Delta^2 x_p \Delta y_p}{[(\Delta x_p)^2 + (\Delta y_p)^2]^{3/2}}$, where x and y are two different coordinates, and Δ and Δ^2 are the first and second order derivatives of the image at the pixel p .

We employed the sequential forward floating selection for feature selection [1], which starts with an empty feature set and

TABLE I: List of features for the representation of points for the proposed DOS classifier.

No.	Features
1-6	avg. and std. of the intensities within the segment in RGB channels
7-10	avg. and std. of the orientations and curvatures of each centerline pixels
11-12	avg. and std. of the blood vessel diameters of each blood vessel segment
13-18	avg. and std. of the intensities of all centerline pixels under a Gaussian blurring ($\sigma = 4$) in RGB channels
19-23	entropy of intensity in RGB channels, orientation and curvature of each centerline pixels

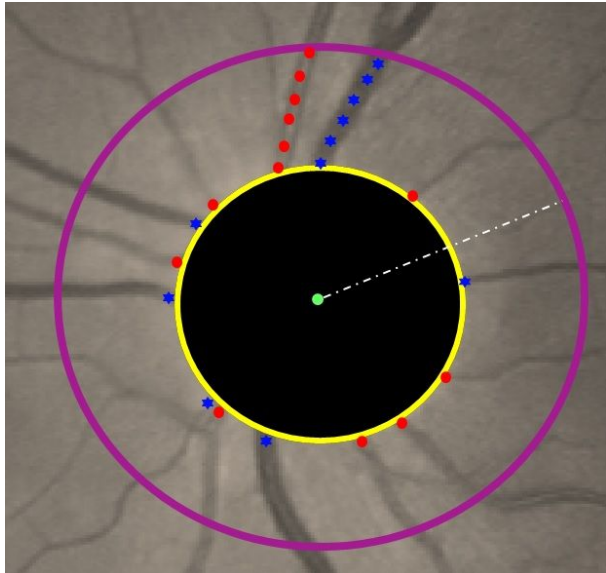


Fig. 4: Definition of a ring for the separation of arteries (red circles) from veins (blue stars).

adds or removes features when this improves the performance of the classifier. (We refer readers to TABLE I for more details about the selected features.) Fig. 1(e) illustrates the estimated vascular network with the added topological information.

D. Arteries/Veins classification via dominant sets clustering

As aforementioned, several supervised learning methods have been exploited for the task of A/V identification. Their performance is limited due to unavailability of sufficient training data, and complicated structure and training process. In this paper, we take into account the reconstructed vascular topology information, utilize the difference of structural information between arteries and veins reflected in retinal images, and propose the ratio of ‘dominant pixels’ as the criterion for an unsupervised A/V identification.

As each branch of retinal blood vessels is composed of many pixels in a retinal image, significant information for discriminating arteries from veins can be derived from their characteristics, and the configuration of the pixels. Dominant pixels are defined as those which form an extracted dominant set amongst all the pixels in a branch of a retinal blood vessel. The dominant pixel ratio is defined as the ratio of dominant

Algorithm 3: DOS-AVIdentification(I)

Inputs: I : a set of pixels that is associated with vascular branches in the retinal color fundus image;

Outputs: L : a label of I which indicates its group;

- 1: initialize the $|I| \times |I|$ symmetric matrix $A = \{a_{ij}\}$ by calculating $a_{ij} = \omega(i, j)$ with respect to given features;
- 2: $S = \text{RDDOS}(A, 10000)$
- 3: **if** $\frac{|S|}{|I|} \geq 0.5$ **then**
- 4: $L = \text{Group-A}$
- 5: **else**
- 6: $L = \text{Group-B}$
- 7: **end if**

pixels to all the pixels in a branch of a retinal blood vessel. Based on these definitions, an effective and concise rule is proposed to separate all the branches of blood vessels into two groups: a branch with a high dominant pixel ratio (given 0.5 or greater) is assigned to Group-A, while Group-B includes those branches with a low dominant pixel ratio. Group-A and Group-B are then assigned as either artery or vein based on their intensity and morphology: arteries are brighter in color than veins, and are thinner than neighboring veins [20]. In this work, we define a vessel segment as ‘artery’ if the average intensity value of the pixels inside the ring centred at any pixel on its centerline with the same radius as the optic disc is larger than 0.48 (with the intensity values lying in the unit interval [0, 1]) (Fig. 4). Otherwise the vessel segment is assigned as ‘vein’. The threshold value was chosen empirically. Algorithm 3 shows the details for the identification of Group-A/Group-B pixels via DOS.

The power of the dominant pixel ratio definition for A/V identification is derived from the effectiveness of DOS in disclosing the hidden structure in arteries and veins. As with the dominant sets-based vascular topology estimation, the weights assigned to the pixels of a vascular branch is responsible for the quality of A/V identification. Once again the features in TABLE I are used to calculate the weights $\omega(i, j)$. Fig. 1(f) illustrates the A/V classification result.

IV. MATERIALS

A. Datasets

Topology reconstruction: Five public retinal datasets were used - the Iowa Normative Set for Processing Images of the RETina (INSPIRE) by the University of Iowa Hospital [1]; the VARPA Images for the Computation of the Arterio/Venular Ratio (VICAVR) captured at the hospital of Conxo, Santiago de Compostela, Spain [3]; the images acquired with an ultra-wide-field device (WIDE) by Duke University [5]; the Scanning Laser Ophthalmoscopy technique-based dataset (IOSTAR) from Eindhoven University of Technology [36]; and the well-known Digital Retinal Images for Blood vessel Extraction (DRIVE) dataset [37]. Except the WIDE dataset, the others used for topology estimation do not contain manually-labeled blood vessel topological information. In consequence, we have annotated them manually, and released these annota-

tions² online.

A/V classification: Four of the above mentioned datasets were used: INSPIRE, VICAVER, WIDE, and DRIVE. An image analysis expert manually classified and then an ophthalmologist checked and corrected the blood vessel segments of the INSPIRE, WIDE, and DRIVE datasets into arteries and veins. For the VICAVER dataset, three experts manually labeled the blood vessel segments as arteries or veins, and agreements between them were then used as groundtruth. However, the experts only classified the blood vessels found within a radius three times that of the optic disc: in other words, the A/V classification labelling for this dataset is incomplete. For a comprehensive visual illustration of A/V labels and fair comparison with other datasets, we asked again the image analysis expert to classify and then the ophthalmologist to check and correct all the unlabeled blood vessels as arteries or veins for this dataset.

It is worth noting that the manual labeling for the INSPIRE and WIDE datasets is at blood vessel centerline pixel level, whereas that for the VICAVER, DRIVE, and IOSTAR datasets is at blood vessel segment pixel level instead. TABLE II summarizes the important details of all these datasets.

B. Topology ground truth estimation

As noted above, the DRIVE, INSPIRE, VICAVER, and IOSTAR datasets have no ground truth for blood vessel topology, so two experts were invited to manually label the topological information of the retinal vascular structure by using a graph editing software we developed for this task. Two experts independently labeled each blood vessel segment or centerline for all the datasets, based on the types of available manual annotations or automatic segmentation of the blood vessel structure. The consensus between them was then used as ground truth. Note, we obtained the topology ground truth of DRIVE, INSPIRE, and IOSTAR based on manual segmentations, and for VICAVER, we generated the topology ground truth on automatic segmented vessels by using the method proposed in [33]. For each image, every blood vessel was labeled with a distinct color (or individual label) to clearly distinguish between different blood vessel trees, as shown in Fig. 1(e).

Thus, the blood vessel segments, or blood vessel centerlines used for topology estimation were extracted either by human graders manually or applying an automatic blood vessel segmentation method [33]: i.e., the DRIVE and IOSTAR datasets include the manual annotations of the retinal blood vessels for each image, so the topology reconstruction was made on manual annotated blood vessel patterns; for the VICAVER dataset, the topology estimation was made at the automatically segmented blood vessels by using the automated segmentation method [33]: for the INSPIRE dataset, the image analysis expert graded and then the ophthalmologist checked and corrected the topology at the blood vessel centerlines, which were generated by the method given in [5].

C. Evaluation metrics

To the best of our knowledge, there is no single metric that can measure the performance of a topology estimation procedure, because of the differing properties of multiple scales of the vasculars, i.e., blood vessels of different sizes exhibit properties that cannot be captured by a single metric. Therefore, in this work, we measure the performance of the topology estimation method by calculating the overlap rate of the correctly identified nodes/intersections. Let A be the total number of the nodes of a blood vessel tree, and B be the number of the nodes that have been correctly identified (the vertices of the estimated topology tree have been assigned labels identical to the ground truth). The overlap rate is then defined as $p = \frac{B}{A} \times 100\%$. Intuitively, p is the percentage of nodes that are correctly identified by the proposed method [30].

To compare the A/V classification performance of the proposed method with the corresponding ground truth as annotated by our human graders, the following metrics were employed: *sensitivity* (Se) = $TP/(TP + FN) \times 100\%$, *specificity* (Sp) = $TN/(TN + FP) \times 100\%$, and *balanced accuracy* ($B-Acc$) = $\frac{Se+Sp}{2}$, where TP , TN , FP and FN denote true positive, true negative, false positive, and false negative, respectively. For the sake of fair comparison and easy implementation, all the evaluations were undertaken at blood vessel centerline pixel level. In this work, we interpret arteries as positives and veins as negatives. Se shows the ability of a given method to detect arteries, while Sp reveals how well it detects veins. $B-Acc$ indicates the overall classification performance, and thus reflects the trade-off between sensitivity and specificity [28].

V. EXPERIMENTAL RESULTS

In order to validate the effectiveness of the proposed DOS algorithm for retinal vascular structure analysis, we have thoroughly evaluated it over five publicly accessible datasets as described in the last section.

A. Topology estimation

Fig. 5 illustrates five vascular topology reconstruction results, each derived from one of these retinal datasets: INSPIRE, IOSTAR, VICAVER, DRIVE, and WIDE, respectively. Compared with the manual annotations, as shown in the second column of Fig. 5, it is clear from visual inspection that our method is able to identify most nodes correctly: only a few cases were incorrectly traced, and these were located at crossovers, as shown in the last column of Fig. 5 - the red squares indicate incorrectly-traced significant nodes.

To better facilitate observation and objective evaluation of the performance of the proposed method in the reconstruction of blood vessel topology, the overlap rates p of the relevant significant nodes were also calculated. TABLE III presents the performance measurements of the proposed method in identifying connectivity of these nodes at bifurcations (BIF) only, crossovers (CRO) only, and all nodes, by counting the number of true positives (TP: correctly identified nodes). As expected, the p scores for bifurcation points for all the five

²<http://imed.nimte.ac.cn/vetovessel-topology-groundtruth.html>

TABLE II: Details of the retinal image datasets used, including the availability of the ground truth for each dataset, and the type of manual annotations.

Datasets	No. Img.	Size	FOV	Camera	Ground truth	Label Type
INSPIRE [†]	40	2392 × 2048	30°	Carl Zeiss Meditec	Topology estimation & AV classification	Centerline
VICAVR [†]	100	768 × 576	45°	Cannon CR6-45NM	Topology estimation & AV classification	Segment
DRIVE [†]	40	565 × 584	45°	Cannon CR5-NM3-CCD	Topology estimation & AV classification	Segment
WIDE	30	1440 × 900	45°	Optos 200Tx	Topology estimation & AV classification	Centerline
IOSTAR [†]	30	1024 × 1024	45°	EasyScan	Topology estimation	Segment

[†] These datasets originally had no blood vessel topological groundtruth. We manually annotated them.

TABLE III: Performance of the proposed method on topology reconstruction at node and segment level, respectively, over different datasets.

	INSPIRE	IOSTAR	VICAVR	DRIVE	WIDE
# BIF	1998	1213	4955	2478	3678
# TP	1945	1167	4799	2288	3551
<i>p</i> score	97.3%	96.2%	96.8%	92.3%	96.5%
# CRO	778	482	1421	832	1230
# TP	697	431	1182	728	1107
<i>p</i> score	89.7%	89.4%	83.2%	87.5%	90.0%
<i>ave. p</i>	95.1%	94.2%	93.8%	91.1%	94.9%
Accuracy	97.5%	95.7%	94.6%	93.5%	96.4%

datasets are much higher than those at crossovers, since it is a relatively simpler node analysis task. The average overlap score over both bifurcations and crossovers was also calculated for all the datasets, and it reveals that the proposed method has the best performance on the INSPIRE dataset with an average $p = 95.1\%$, while yielding the lowest score on the DRIVE dataset, with an average $p = 91.1\%$. This is due to the relatively low resolution and image contrast of this dataset.

In addition, the accuracy of the topology reconstruction at blood vessel centerline level was obtained, and it can be observed that the accuracy scores are higher than the corresponding average overlap rate: for example, we see an accuracy score of 97.5%, but an average overlap rate of only 95.1% for the INSPIRE dataset. This is because the number of blood vessel segments is much larger than the number of significant nodes (bifurcation and crossovers).

As we mentioned above, there is no single metric that is able to directly and objectively measure the performance of a topology reconstruction method. The most common method as described in [5], [30] was topological structure-guided A/V classification. Therefore, in the following subsection, we continue to make use of the proposed DOS-based topology estimation method to guide the task of A/V classification.

B. A/V classification

After the reconstruction of the blood vessel topology, the complete blood vessel network has been separated into several individual branches as sub-trees with an individual label (distinct color). The final goal is to assign these labels to one of two classes: artery or vein.

The features listed in TABLE I and the DOS classifier were utilized again to classify these individual labels into two clusters, A and B . For each sub-tree v , the probability of its being A is estimated by the number of blood vessel pixels

classified by DOS as A : $P_A^v = n_A^v / (n_A^v + n_B^v)$, where n_A^v is the number of pixels classified as A , and n_B^v is the number of pixels classified as B . For each sub-tree, the higher probability is used to define whether the sub-tree is assignable to category A or B . Clusters A and B are then assigned as artery and vein, respectively, based on their intensity and morphology: arteries are brighter in color than veins, and are thinner than neighboring veins [20].

Fig. 6 shows the A/V classification results of the proposed method over four sample images, one from each dataset, based on their topological information. Overall, the proposed method can correctly distinguish most of the A/V labels on all the four datasets, in comparison with the corresponding manual annotations. The arteries are shown in red and veins in blue. However, our A/V classification relies heavily on the prior results of topology reconstruction: if a blood vessel segment was incorrectly identified at the stage of topology reconstruction, it might be very likely to be falsely labeled during subsequent A/V classification. This is because a blood vessel segment must share its label with all its downstream segments, as evidenced by the figures in the second and third columns of Fig. 6 - the incorrectly traced veins are labeled in green, while the falsely identified arteries are labeled in yellow.

In order to better demonstrate the superiority of the proposed method, TABLE IV reports the comparison of our method with the state-of-the-art A/V ones over four datasets, in terms of centerline pixel-level sensitivity, specificity, and balanced accuracy. It is shown that our method correctly identifies over 93.5% of the A/V labels of all the datasets, outperforming all the compared methods, with a single exception that its Sp score on the DRIVE dataset is 1.4% lower than that of [28]. Nevertheless, the proposed method is able to correctly identify the majority of arteries and veins by relying on the topological constraint.

The intra-observer scores are also provided for the INSPIRE and WIDE datasets in TABLE IV. It can be seen that the sensitivities of the proposed method are very close to those of the human observers: our method obtains competing sensitivities, with 96.8% and 96.2%, compared to 97.2% and 97.0% by the intra-observers for the INSPIRE and WIDE datasets, respectively.

Overall, the proposed method has the best performance on the INSPIRE dataset, with a balanced accuracy of 96.4%. This dataset contains images with a higher resolution, and considerably fewer and simpler blood vessel bifurcations and crossovers (e.g., an average number 40 vs. 102 of circuits

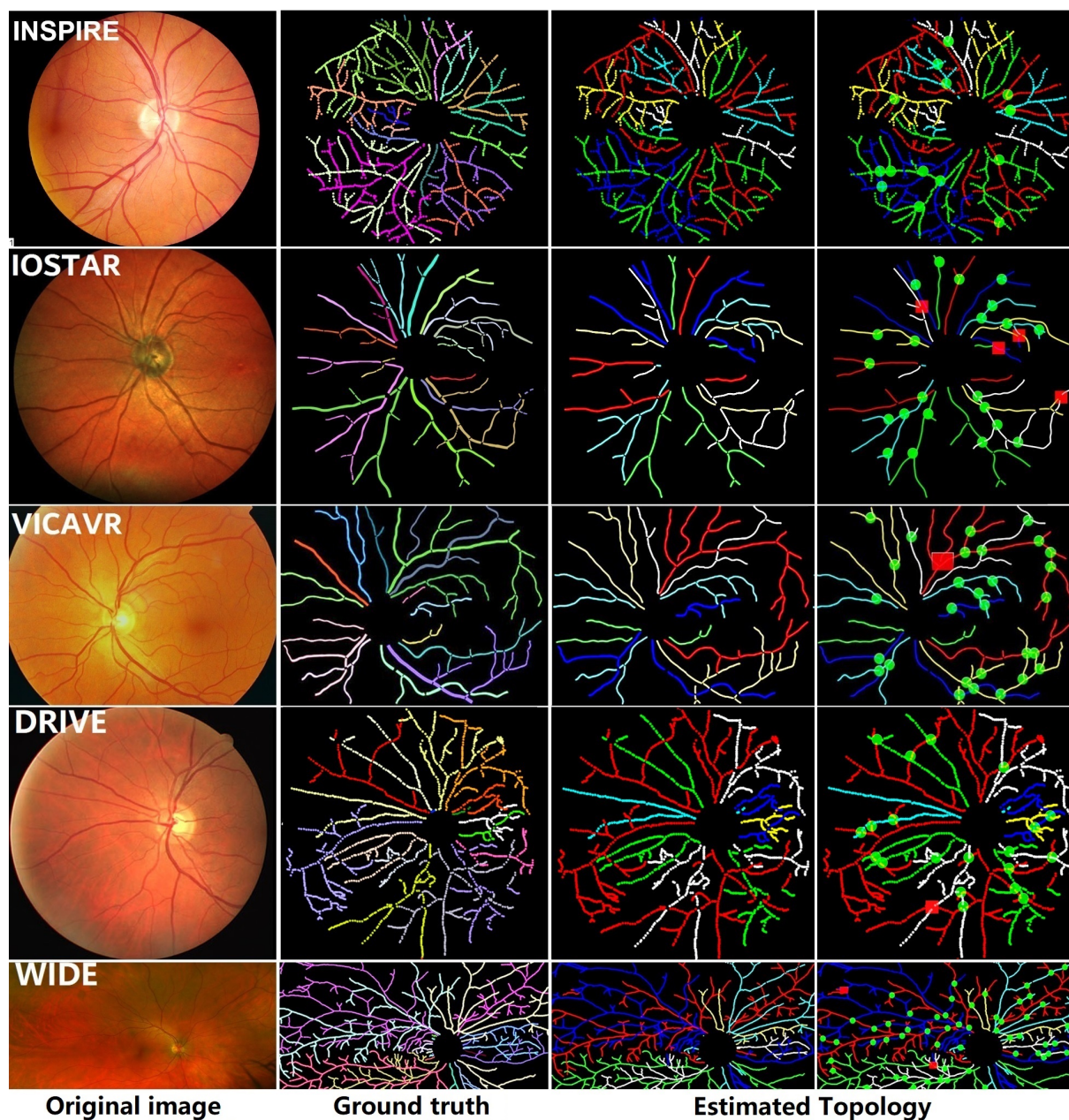


Fig. 5: Examples of vascular topology estimation results. From left to right column: original image, manual annotation, result of the proposed method, and the highlighted correctly (Green discs) and incorrectly (Red squares) identified connections.

when compared to the WIDE dataset). This finding is also evidenced by the performance on the topology estimation, with the highest accuracy of 97.5%, when compared with other datasets in TABLE III.

All the above-reported performances of different methods over the DRIVE dataset were obtained when using the manual annotation produced by Qureshi et al. [40] as ground truth. It is noted that the original authors of the DRIVE dataset also released another manual A/V annotation [37]. We refer them as GT-1 and GT-2, respectively. Fig. 7 shows two examples of these two manual A/V annotations.

In order to demonstrate our A/V classification method over two different manual annotations, and the agreement of those manual annotations, we computed the *false discovery rate*

(FDR) [41] in three different scenarios: Auto vs GT-1, Auto vs GT-2, and GT-1 vs GT-2. FDR is defined as the fraction of the total number of pixels incorrectly identified over the total number of pixels in ground truth. It can be observed from TABLE V that the proposed automatic A/V classification method has a larger FDR than the human annotations, as expected, and the two human annotations also have an FDR of as large as 0.071 ± 0.014 . Interestingly, Estrada et al. [28] also reported a very close agreement score of 96% between GT-1 and GT-2. These differences between the annotations given by different human graders imply the difficulty of the A/V classification problem.

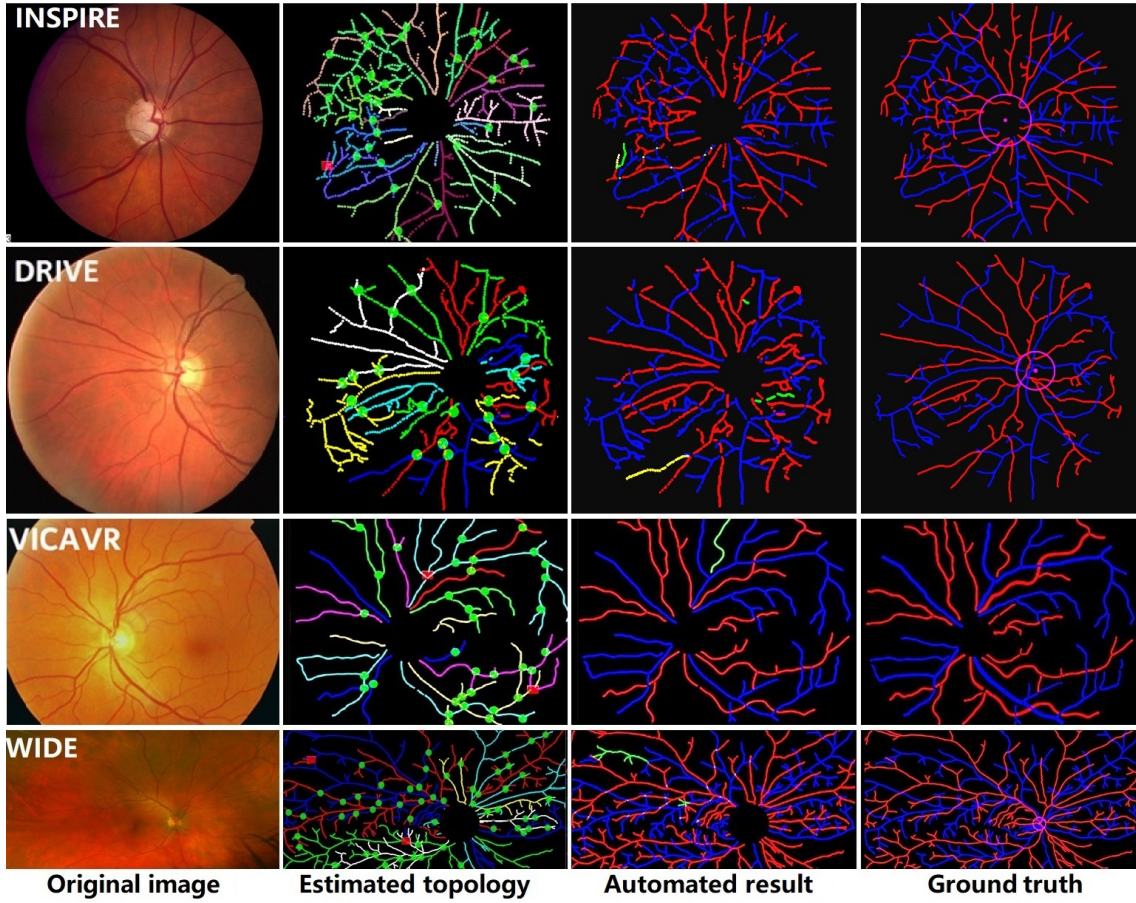


Fig. 6: A/V classification results of the proposed method over four different datasets. From left to right column: original image, blood vessel topology, A/V classification result, and corresponding manual annotations.

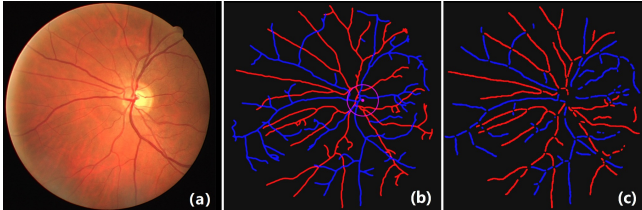


Fig. 7: Different manual A/V annotation results of the DRIVE dataset. (a) original image, (b) manual annotation by [40] (GT-1), (c) manual annotation by [37] (GT-2).

TABLE V: False discovery rate of different A/V classification methods over two different manual annotations: GT-1 and GT-2. Auto refers to the proposed automated A/V classification method.

	Auto vs GT-1	Auto vs GT-2	GT-1 vs GT-2
FDR	0.093±0.009	0.088±0.013	0.071±0.014

C. Computational complexity

The proposed method has a computational complexity of $O(N^2)$ in topology estimation, where N is the total number of nodes in a blood vessel graph, a computational complexity of $O(M \times P^2)$ in label assignment of each vessel branch,

where M is the total number of individual vessel branches in a vessel graph, and P is the average number of pixels of an individual vessel branch and a computational complexity of (M) in identifying the labels of the vessel branches as either artery or vein. Therefore, the entire computational complexity of the proposed method is $O(N^2) + O(M \times P^2)$ and thus $O(N^2)$ where $M \leq N$, $P \leq N$, and $MP \leq N$.

The average running times of each image with standard deviation in seconds for overall A/V classification and topology estimation of our method are: $72.61 \pm 16.32s$ and $2.83 \pm 0.41s$ over INSPIRE; $24.43 \pm 2.31s$ and $2.10 \pm 0.33s$ over DRIVE; $28.98 \pm 0.48s$ and $2.57 \pm 1.01s$ over VICA VR; and $33.77 \pm 2.01s$ and $5.84 \pm 1.37s$ over WIDE respectively. Note, we did not perform the A/V classification over IOSTAR due to lack of manual annotation, so we only report the running time of its topology estimation as $1.93 \pm 0.80s$. All the experiments were carried out in MATLAB2015a on a PC with an Intel Xeon E5-2695 v4 CPU, 2.10GHz, and 128GB RAM.

In the literature, only the method proposed by Estrada et al. [28] reported the computation time of their A/V classification on a Toshiba Satellite X870 laptop with a 2.4Ghz Intel I7 quad-core processor and 32GB of RAM: $117.68 \pm 34.10s$ over INSPIRE; $131.32 \pm 33.40s$ over DRIVE; and $777.35 \pm 330.52s$ over WIDE.

TABLE IV: The sensitivity Se , specificity Sp and balanced accuracy $B-Acc$ of different A/V classification methods over different datasets.

	Classifier	INSPIRE			DRIVE			VICA VR			WIDE		
		Se	Sp	B-Acc	Se	Sp	B-Acc	Se	Sp	B-Acc	Se	Sp	B-Acc
Second human grader	-	97.2%	97.0%	97.1%	-	-	-	-	-	-	97.0%	97.9%	97.4%
Niemeijer et al. [1]	kNN	78.0%	78.0%	78.0%	80.0%	80.0%	80.0%	-	-	-	-	-	-
Muramatsu et al. [38]	LDA	-	-	-	-	-	93.0%	-	-	-	-	-	-
Vazquez et al. [3]	k-means	-	-	-	-	-	-	-	-	88.8%	-	-	-
Mirsharif et al. [22]	LDA	-	-	-	-	-	84.1%	-	-	-	-	-	-
Relan et al. [23]	GMM	92.7%	48.5%	70.8%	-	-	-	-	-	-	-	-	-
Dashbozorg et al. [8]	LDA	91.0%	86.0%	88.5%	90.0%	84.0%	87.0%	-	-	89.8%	-	-	-
Lyu et al. [29]	k-means	90.2%	79.4%	84.8%	-	-	-	-	-	-	-	-	-
Pellegrini et al. [39]	graph	-	-	-	-	-	-	-	-	-	-	-	92.5%
Girard et al. [24]	CNN	-	-	-	92.3%	93.1%	92.7%	-	-	-	-	-	-
Huang et al. [25]	LDA	-	-	85.1%	-	-	-	-	-	90.6%	-	-	-
Estrada et al. [28]	graph	91.5% $\pm 0.1\%$	90.2% $\pm 0.1\%$	90.9% $\pm 0.1\%$	93.0% $\pm 0.06\%$	94.1% $\pm 0.07\%$	93.5% $\pm 0.05\%$	-	-	-	91.0% $\pm 0.06\%$	90.9% $\pm 0.06\%$	91.0% $\pm 0.06\%$
Proposed	DOS	96.8% $\pm 0.08\%$	95.7% $\pm 0.08\%$	96.4% $\pm 0.06\%$	94.2% $\pm 0.07\%$	92.7% $\pm 0.07\%$	93.5% $\pm 0.07\%$	95.4% $\pm 0.1\%$	93.8% $\pm 0.09\%$	94.6% $\pm 0.09\%$	96.2% $\pm 0.1\%$	94.2% $\pm 0.06\%$	95.2% $\pm 0.08\%$

VI. DISCUSSIONS AND CONCLUSIONS

In this work, we proposed a DOS-based topology estimation method, and further utilized the obtained topological information to separate arteries from veins in retinal fundus images. In other words, the blood vessel classification problem has been formalized as a pairwise clustering problem. In this section, we carry out a comprehensive comparison between the proposed method and the state-of-the-art label propagation (or topology estimation) and A/V classification methods.

A. Comparison to the state-of-the-art label propagation / topology estimation methods

As suggested by De et al. [31], the topology estimation task can be reformulated as topological label propagation over the directed graph. Therefore, in this section, the p values at blood vessel centerline pixel level obtained by the proposed DOS method is compared with those of five state-of-the-art label propagation and topology estimation methods: Loopy Belief Propagation (LBP) [42], Symmetrized Graph Laplacian (SGL) [43], Zero-mode Free Laplacian (ZFL) [44], Matrix-Forest Theorem of Directed graphs (MFTD) [30], [45], and Heuristic Search Algorithm (HSA) [5]. Of these methods, LBP, SGL, and ZFL are essentially label propagation approaches, for which the source codes have already been developed by the machine learning community; MFTD and HSA are state-of-the-art topology reconstruction methods for tree-like structures: their codes are available from the authors. However, the HSA was designed for the WIDE dataset, and is thus applicable only to this dataset.

In our experiments, the recommended parameters from the original source code or literature were used. These methods all share the same blood vessel extraction and optic disc removal step, as well as the preparation process of converting the skeleton maps into undirected graphs or digraphs, as described in Section III.A. The experimental results are presented in TABLE VI.

Overall, our DOS method consistently outperforms the other label propagation methods [42]–[44] and existing topology estimation methods [5], [30] by a rather large margin. Again,

TABLE VI: The overlapping rate (p score) of different label propagation and topology estimation methods over different datasets.

	INSPIRE	IOSTAR	VICA VR	DRIVE	WIDE
LBP [42]	76.6%	75.9%	74.6%	71.9%	72.3%
ZFL [44]	84.3%	84.3%	83.1%	82.7%	81.7%
SGL [43]	81.0%	79.8%	82.0%	80.9%	78.8%
MFTD [30]	91.1%	90.9%	87.6%	84.9%	87.5%
HSA [5]	-	-	-	-	89.6%
DOS	95.1%	94.2%	93.8%	91.1%	94.9%

all the methods yielded their best results over the INSPIRE dataset due to its simpler branches, and their second best over the WIDE dataset due to the wider field-of-view of this dataset.

Fig. 8 also presents for visual comparison the topology reconstruction results of the competing methods on an example image from the DRIVE dataset. Grey and red colored disks are used here to represent correct and incorrect predictions in various bifurcation and crossover scenarios. It can be seen that the proposed method produces better topology reconstruction results than its competitors, by observing the bifurcations and crossovers where these have been highlighted. This is because MFTD takes into consideration only the angular or intensity properties of neighbouring blood vessel segments forming a junction for the construction of their current weight matrix [30]. Generally speaking, the proposed method achieves the best performance, which suggests the advantages of our DOS approach for vascular topology reconstruction. However, the proposed method also makes errors in some cases, since it may suffer from failures at the segmentation and skeletonization stages that may misrepresent the topological structures of the retinal blood vessels for the classification phase.

B. Comparison to the state-of-the-art A/V classifiers

To facilitate better comparison with the results of our classifier, we also applied the most commonly-used A/V classifiers to the topology-assigned structures derived by our method. The comparative A/V classifiers chosen were LDA, GMM, k-means clustering, and SVM. For the supervised classifiers (LDA and SVM), we randomly assigned half of the images

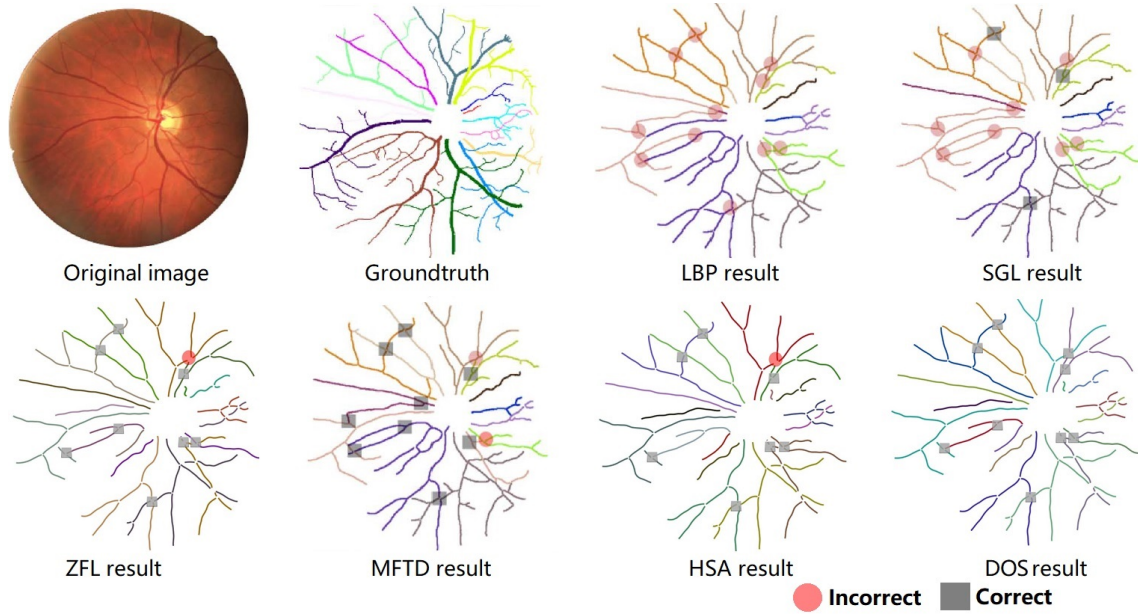


Fig. 8: The results of different methods in determining the labels at bifurcations and crossovers. Note: the results of LBP, SGL, and MFTD were quoted from [30] for convenience.

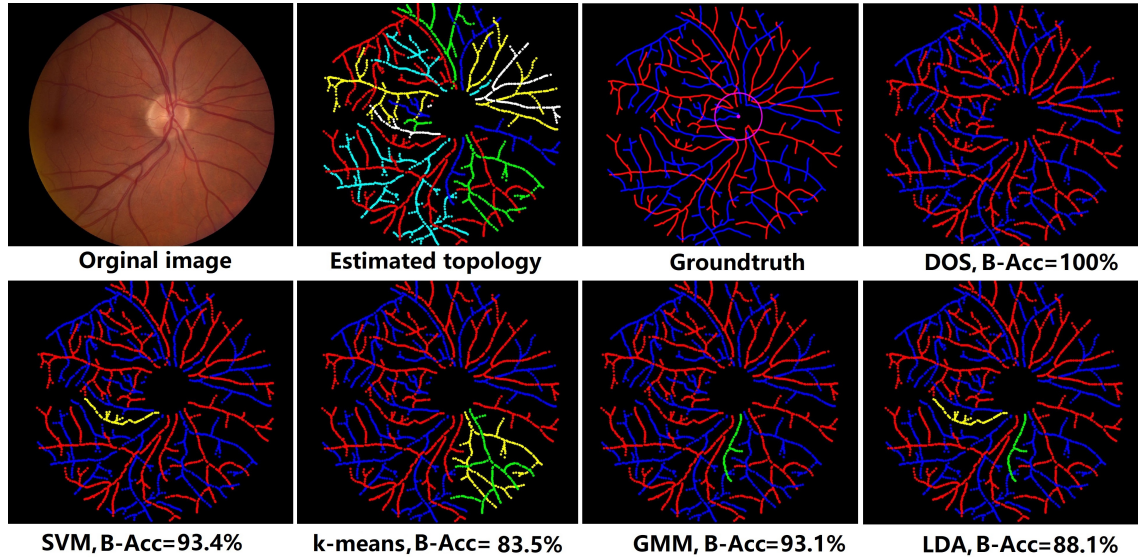


Fig. 9: Blood vessel classification results of a randomly selected image from the INSPIRE dataset by using different A/V classifiers over the topology-assigned blood vessel structures produced by the proposed method.

as a training set, and the remainder of the images as the test set. The same features listed in TABLE I were used to train the LDA and SVM classifiers. TABLE VII demonstrates how well the competing classifiers succeeded in classifying the topology-assigned network into arteries or veins. It can be seen that our DOS method clearly outperformed the compared classification methods, with a balanced accuracy ($B\text{-Acc}$) of 96.4%, 93.5%, 94.6%, and 95.2% on the INSPIRE, DRIVE, VICAVER, and WIDE datasets, respectively.

Fig. 9 illustrates the A/V classification performance of different classifiers on our DOS-guided blood vessel topology (Green: incorrectly traced veins; Yellow: incorrectly traced). It shows that the proposed DOS method achieved the highest bal-

anced accuracy in this case, with 100% on a randomly-selected image from the INSPIRE dataset. The k-means clustering classifier produced the worst results with a $B\text{-Acc}$ score of 83.5%: this is because one of its major drawbacks is the naive use of the mean value for the representation of each cluster, as a result of which the clusters obtained may not be repeatable and lack consistency. The GMM method obtained relatively higher scores than the other three compared methods, with a $B\text{-Acc}$ score of 93.1%: it is more flexible in terms of cluster shape representation than k-means and SVM, and the clusters can be shaped as ellipses, rather than being restricted to hyperspheres in k-means clustering or hyperplanes in SVM.

From a comparison of TABLE IV and TABLE VII, it can be

TABLE VII: The balanced accuracy and standard deviation of different classifiers over the topology-assigned structure derived by the proposed method from different datasets.

	INSPIRE	DRIVE	VICAVR	WIDE
LDA	91.6% \pm 0.10%	89.6% \pm 0.11%	91.1% \pm 0.10%	90.6% \pm 0.12%
GMM	92.1% \pm 0.09%	87.4% \pm 0.06%	91.0% \pm 0.07%	89.3% \pm 0.08%
k-means	88.3% \pm 0.12%	85.8% \pm 0.11%	90.5% \pm 0.10%	90.1% \pm 0.12%
SVM	93.9% \pm 0.07%	88.3% \pm 0.09%	90.6% \pm 0.06%	89.2% \pm 0.08%
DOS	96.4% \pm 0.06%	93.5% \pm 0.07%	94.6% \pm 0.09%	95.2% \pm 0.08%

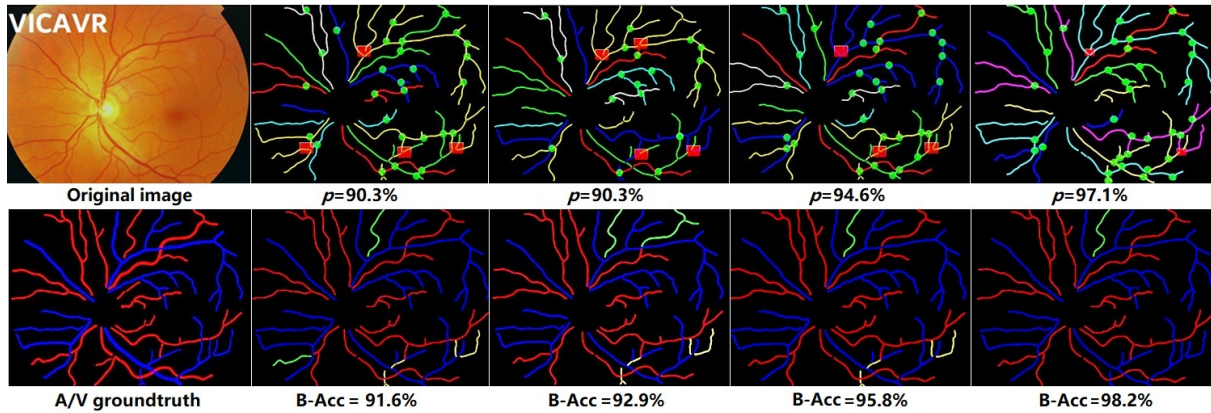


Fig. 10: The topology estimation (top row) and A/V classification (bottom row) results of the proposed method with and without removing different types of features for the representation of the pixels of interest. From second to last column: intensity-based features removed, orientation-based features removed, diameter-based features removed, and all features retained.

TABLE VIII: Topology estimation and A/V classification results ($p/B-Acc$) of the proposed method with and without removing different types of features for the representation of the pixels of interest.

	INSPIRE	DRIVE	VICAVR	WIDE	IOSTAR
<i>All features retained</i>	95.1% / 96.4%	91.1% / 93.5%	93.8% / 94.6%	94.9% / 95.2%	95.1% / -
<i>Diameter-based features removed</i>	94.2% / 94.4%	89.1% / 91.5%	92.0% / 93.3%	93.4% / 94.5%	94.1% / -
<i>Orientation-based features removed</i>	92.9% / 93.2%	86.8% / 88.9%	89.9% / 91.8%	92.5% / 93.7%	93.0% / -
<i>Intensity-based features removed</i>	91.0% / 92.1%	85.1% / 88.5%	88.3% / 90.6%	91.9% / 92.1%	92.2% / -

observed that the method proposed by Dashtbozorg et al. [8] yielded a $B-Acc$ score of 88.5% over the INSPIRE dataset: they used similar features to train the LDA classifier. After the LDA classifier was applied to our DOS generated blood vessel topology, the $B-Acc$ score has been dramatically increased to 91.6%. In a similar manner to the GMM classifier, the $B-Acc$ score has been increased from 70.8% [23] to 92.1%. These results imply that the accurate identification of the underlying network topology is key to improving the accuracy of A/V classification.

C. Feature importance analysis

The feature weights for nodes and blood vessel segments are critical to the accuracy of subsequent topology estimation and A/V classification. In real applications, the feature weights are usually determined by a similarity measure, where only certain features are taken into consideration. Many strategies have been explored for the selection of these critical features, in an effort to identify more compact and better quality feature subsets for various tasks. Such techniques typically involve the use of an individual feature significance evaluation, or a measurement of feature subset consistency, that work together with a search algorithm to determine a quality

subset of features. However, most of the feature selection techniques are supervised, i.e., the labels of instances are required for evaluating or measuring the significance of the feature subsets. For the purposes of this topology estimation and A/V classification work, it is prohibitively time-consuming to have the vascular branches annotated by human experts, and hence benchmarking public data sets are rare in the literature. Therefore, the features used here to determine the weights ($\omega(i, j)$ in Algorithm 3) were empirically selected, as listed in TABLE I.

We reported the performances of topology estimation and A/V classification results of the proposed method in TABLE VIII, in terms of removing the intensity-based, orientation-based, and diameter-based features, respectively from the feature candidates. It can be seen that intensity-based features affect most of the proposed classifier, as the p scores of the topology estimation are significantly decreased by 4.1%, 6.0%, 5.5%, 3.0%, and 2.9% over INSPIRE, DRIVE, VICAVR, WIDE, and IOSTAR, respectively after their removal. Similarly, the $B-Acc$ scores of A/V classification are also decreased by a significant margin: 4.3%, 5.0%, 4.0%, 3.1%, and 2.9% over INSPIRE, DRIVE, VICAVR, and WIDE datasets.

Fig. 10 illustrates the resulting topology estimation and

A/V classification labels of a randomly selected image from the VICAVR database (Green: incorrectly traced veins; Yellow: incorrectly traced), after removing the intensity-based, orientation-based, and diameter-based features individually from the feature representation of the pixels of interest. It demonstrates that intensity-based features are the most important: more incorrectly identified connections occurred, and more incorrectly classified arteries and veins took place after their removal.

D. Comparison between automatic and manual segmentations

All the above-reported results are based on ground truth segmentation, with a single exception for the VICAVR dataset. In this subsection, in order to characterize how the errors of automatic and manual segmentations affect their topological reconstructions and final A/V classifications, we have repeated our topology estimation and AV classification methods over the automated vessel segmentation results of the INSPIRE dataset.

To compare the vessel segmentation performance of the automated method with the corresponding ground truth, we computed the *sensitivity* and *false discovery rate* (FDR) [41] between the predicted centerlines and ground truth centerlines. (the INSPIRE dataset provides the centerline of the retinal blood vessels only.) Sensitivity is the fraction of the number of centerline pixels in the correctly detected segments (true positives) over the total number of centerline pixels in the ground truth. FDR is defined as the fraction of the total number of centerline pixels incorrectly detected as vessel segments (false positives) over the total number of centerline pixels of the traced vessels in the ground truth. The use of *specificity*, defined as the number of pixels correctly rejected as non-vessel structures (true negatives), is not adequate for the evaluation of this segmentation task, since the vast majority of the pixels do not belong to blood vessels. It is worth noting that, as is customary in the evaluation of the methods extracting one pixel-wide curves [41], a three-pixel tolerance region around the manually traced nerves is considered to be a true positive. In other words, a predicted centerline point is considered as true positive if there are no more than three pixels from the nearest ground truth centerline point. The experimental results are presented in Figure 11 and Table IX.

Figure 11 shows that the proposed method has produced perfect topology estimation and A/V classification results without any error on manual vessel annotations of an image randomly selected from the INSPIRE dataset. While only one node was incorrectly identified, indicated by the red rectangle in Fig. 11(c), and one vessel segment was incorrectly classified on the automatic segmented vessels, indicated by the yellow line in Fig. 11(e), since a tiny vessel was mis-detected and leads to the generation of a partly incorrect vessel graph that causes further errors in the vessel topology reconstructions and A/V classifications.

In addition, the percentage of nodes that were correctly identified (p score) and accuracy of the topology reconstruction at blood vessel centerline pixel-level were calculated. As can be observed in TABLE IX, the proposed topology

estimation method achieved a similar p score on automatic and manual segmentations: 93.5% and 95.1%, respectively. However, it yielded significantly low scores of accuracy and B - Acc , by 8.6% and 8.7%, respectively. As aforementioned, the accuracy and B - Acc were calculated at centerline-level. The automated segmentation performance achieved only 83.8% in sensitivity, in comparison with that on manual segmentation. This is because a large portion of tiny vessels were mis-segmented.

However, the obtained p score implies that the proposed method is able to correctly identify most of the nodes in automatically detected vessels. These results show that our topology estimation and its subsequent AV classification, are relatively robust in terms of applying to either automatic or manual vessel segmentation results at node-level, while the overall performance (pixel-level) is still dependent on the completeness of the vessel segmentation.

E. Conclusions

Automated identification of the anatomical connectivity of different blood vessels, and classification of those blood vessels into arteries and veins, respectively, are essential for the automated assessment of vascular changes.

In this paper, we have proposed a novel artery/vein classification method based on vascular topological characteristics. Our framework combines graph-theoretic methods with DOS to accurately analyse the retinal vasculature. The concept of DOS in this work was successfully adapted to formalize the topology estimation and A/V classification as a pairwise clustering problem. The core issues then go to work on graph generation and edge weight definition. They have been achieved through image segmentation, skeletonization and identification of significant nodes. The latter is defined as the inverse Euclidean distance between the two end points of an edge in the feature space, where each node is represented as a 23 dimensional feature vector about intensity, orientation, curvature, diameter and entropy. The significance of our method is that it is capable of classifying the whole vascular network, and does not restrict itself to specific regions of interest.

The proposed method has accurately reconstructed the vascular topology and classified the blood vessel types into arteries and veins on five publicly accessible retinal datasets. The results show that our method produces better performance when compared with the state-of-the-art topology estimation and A/V classification ones. It can be expected that the proposed method could be a powerful tool for analyzing vasculature for better management of a wide spectrum of vascular-related diseases.

In addition, we have manually labeled the blood vessel topology for four publicly-available retinal datasets (INSPIRE, VICAVR, DRIVE, and IOSTAR), and these annotations have been released for public access to help other researchers in the community in performing research and development on the same and related topics. For future work, we will test our method on other retinal datasets (e.g., RITE [46]) and neuronal datasets (e.g., DIADEM [47]), and refine the initial graph generation to improve the accuracy of the retinal topology reconstruction.

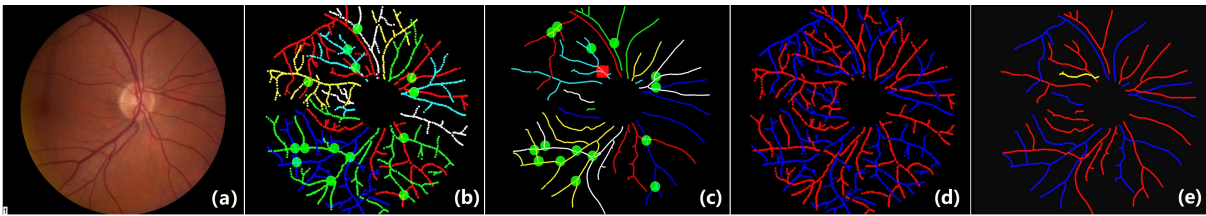


Fig. 11: Illustrative results of the proposed method for topology estimation and A/V classification applied on automatic and manual segmentations. (a) A randomly selected image from INSPIRE; (b)-(c) the topology estimation results applied to manual and automatic segmentations; (d)-(e) A/V classification results on manual and automatic segmentations.

TABLE IX: The performance of the proposed method for vessel topology estimation and A/V classification to automatic and manual segmentations at the node-level and vessel centerline pixel-level over the INSPIRE dataset.

	vessel segmentation		topology estimation				A/V classification
	sensitivity	FDR	# nodes	# TP	p	accuracy	B-Acc
manual segmentation	-	-	2776	2642	95.1%	97.5%	96.4%
automatic segmentation	83.8%	0.071	2480	2321	93.5%	88.9%	87.7%

REFERENCES

- [1] M. Niemeijer, X. Xu, A. Dumitrescu, B. van Ginneken, J. Folk, and M. Abramoff, "Automated measurement of the arteriolar-to-venular width ratio in digital color fundus photographs," *IEEE Trans. Med. Imaging*, vol. 30, no. 11, pp. 1941–1950, 2011.
- [2] J. Xie, Y. Zhao, Y. Zheng, P. Su, J. Liu, and Y. Wang, "Retinal vascular topology estimation via dominant sets clustering," in *Proc. ISBI*, 2018, pp. 1458–1462.
- [3] S. Vázquez, B. Cancela, N. Barreira, and M. Saez, "Improving retinal artery and vein classification by means of a minimal path approach," *Mach. Vis. Appl.*, vol. 24, no. 5, pp. 919–930, 2013.
- [4] T. Na, J. Xie, Y. Zhao, Y. Zhao, Y. Liu, Y. Wang, and J. Liu, "Retinal vascular segmentation using superpixel-based line operator and its application to vascular topology estimation," *Medical physics*, vol. 45, no. 7, pp. 3132–3146, 2018.
- [5] R. Estrada, C. Tomasi, S. Schmidler, and S. Farsiu, "Tree topology estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1688–1701, 2015.
- [6] Y. Zhao, J. Zhao, J. Yang, Y. Liu, Y. Zhao, Y. Zheng, L. Xia, and Y. Wang, "Saliency driven vasculature segmentation with infinite perimeter active contour model," *Neurocomputing*, vol. 259, pp. 201–209, 2017.
- [7] Y. Zhao, Y. Zheng, Y. Liu, Na, and Y. W. J. Liu, "Automatic 2d/3d vessel enhancement in multiple modality images using a weighted symmetry filter," *IEEE Trans. Med. Imaging*, pp. 1–1, 2017.
- [8] B. Dashtbozorg, A. M. Mendonça, and A. Campilho, "An automatic graph-based approach for artery/vein classification in retinal images," *IEEE Trans. Image Processing*, vol. 23, no. 3, pp. 1073–1083, 2014.
- [9] V. Joshi, J. Reinhardt, M. Garvin, and M. Abramoff, "Automated method for identification and artery-venous classification of vessel trees in retinal vessel networks," *PLoS ONE*, vol. 9, no. 2, pp. 1–12, 2014.
- [10] 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 *Proc. MICCAI*, 2018, pp. 109–118.
- [11] M. Pavan and M. Pelillo, "Dominant sets and hierarchical clustering," in *Proc. ICCV*, 2003, pp. 362–369.
- [12] E. Zemene and M. Pelillo, "Interactive image segmentation using constrained dominant sets," in *Proc. ECCV*, 2016, pp. 278–294.
- [13] E. Zemene, L. Alemu, and M. Pelillo, "Dominant sets for "constrained" image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2018.
- [14] Y. Zhao, Y. Zheng, Y. Liu, J. Yang, Y. Zhao, D. Chen, and Y. Wang, "Intensity and compactness enabled saliency estimation for leakage detection in diabetic and malarial retinopathy," *IEEE Trans. Med. Imaging*, vol. 36, no. 1, pp. 51–63, 2017.
- [15] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable COS-FIRE filters for vessel delineation with application to retinal images," *Med. Image Anal.*, vol. 19, pp. 46–57, 2015.
- [16] Y. Zhao, Y. Liu, X. Wu, S. Harding, and Y. Zheng, "Retinal vessel segmentation: An efficient graph cut approach with retinex and local phase," *PLoS ONE*, vol. 10, p. e0122332, 2015.
- [17] M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "Blood vessel segmentation methodologies in retinal images - a survey," *Comput. Meth. Prog. Bio.*, vol. 108, pp. 407–433, 2012.
- [18] J. Soares and M. Cree, "Retinal vessel segmentation using the 2D Gabor wavelet and supervised classification," *IEEE Trans. Med. Imaging*, vol. 25, pp. 1214–1222, 2006.
- [19] M. Martínez-Pérez, A. Hughes, A. Stanton, S. Thom, N. Chapman, A. Bharath, and K. Parker, "Retinal vascular tree morphology: a semi-automatic quantification," *IEEE Trans. Biomed. Engineering*, vol. 49, no. 8, pp. 912–917, 2002.
- [20] C. Kondermann, D. Kondermann, and M. Yan, "Blood vessel classification into arteries and veins in retinal images," in *Proc. SPIE Medical Imaging*, 2007, p. 651247.
- [21] M. Niemeijer, B. Ginneken, and M. Abramoff, "Automatic classification of retinal vessels into arteries and veins," in *Proc. SPIE Medical Imaging*, vol. 7260, 2009, pp. 72 601F–1–72 601F–8.
- [22] Q. Mirsharif, F. Tajeripour, and H. Pourreza, "Automated characterization of blood vessels as arteries and veins in retinal images," *Comp. Med. Imag. and Graph.*, vol. 37, no. 7-8, pp. 607–617, 2013.
- [23] D. Relan, T. MacGillivray, L. Ballerini, and E. Trucco, "Retinal vessel classification: Sorting arteries and veins," in *Proc. EMBC*, 2013, pp. 7396–7399.
- [24] F. Girard and F. Cheriet, "Artery/vein classification in fundus images using CNN and likelihood score propagation," in *Proc. IEEE Global Conf. on Sig. and Info. Process.*, 2017, pp. 720–724.
- [25] F. Huang, B. Dashtbozorg, and B. H. Romeny, "Artery/vein classification using reflection features in retina fundus images," *Mach. Vis. Appl.*, vol. 29, no. 1, pp. 23–34, 2018.
- [26] K. Rothaus, X. Jiang, and P. Rhiem, "Separation of the retinal vascular graph in arteries and veins based upon structural knowledge," *Image Vision Comput.*, vol. 27, no. 7, pp. 864–875, 2009.
- [27] V. Joshi, M. Garvin, J. Reinhardt, and M. Abramoff, "Automated method for the identification and analysis of vascular tree structures in retinal vessel network," in *Proc. SPIE Medical Imaging*, 2011, p. 79630I.
- [28] R. Estrada, M. J. Allingham, P. S. Mettu, S. W. Cousins, C. Tomasi, and S. Farsiu, "Retinal artery-vein classification via topology estimation," *IEEE Trans. Med. Imaging*, vol. 34, no. 12, pp. 2518–2534, 2015.
- [29] X. Lyu, Q. Yang, S. Xia, and S. Zhang, "Construction of retinal vascular trees via curvature orientation prior," in *Proc. BIMB*, 2016, pp. 375–382.
- [30] J. De, L. Cheng, X. Zhang, F. Lin, H. Li, K. Ong, W. Yu, Y. Yu, and S. Ahmed, "A graph-theoretical approach for tracing filamentary structures in neuronal and retinal images," *IEEE Trans. Med. Imaging*, vol. 35, no. 1, pp. 257–72, 2016.
- [31] J. De, X. Zhang, F. Lin, and L. Cheng, "Transduction on directed graphs via absorbing random walks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 7, pp. 1770–1784, 2018.
- [32] J. Cheng, J. Liu, Y. Xu, F. Yin, D. Wong, N. Tan, D. Tao, C. Cheng, T. Aung, and T. Wong, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE Trans. Med. Imaging*, vol. 32, no. 6, pp. 1019–1032, 2013.

- [33] Y. Zhao, L. Rada, K. Chen, S. Harding, and Y. Zheng, "Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images," *IEEE Trans. Med. Imaging*, vol. 34, no. 9, pp. 1797–1807, 2015.
- [34] P. Bankhead, J. McGeown, and T. Curtis, "Fast retinal vessel detection and measurement using wavelets and edge location refinement," *PLoS ONE*, vol. 7, p. e32435, 2009.
- [35] M. Pavan and M. Pelillo, "Dominant sets and pairwise clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 167–172, 2007.
- [36] J. Zhang, B. Dashtbozorg, E. J. Bekkers, J. P. W. Pluim, R. Duits, and B. M. ter Haar Romeny, "Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores," *IEEE Trans. Med. Imaging*, vol. 35, pp. 2631–2644, 2016.
- [37] J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 501–509, 2004.
- [38] C. Muramatsu, Y. Hatanaka, T. Iwase, T. Hara, and H. Fujita, "Automated selection of major arteries and veins for measurement of arteriolar-to-venular diameter ratio on retinal fundus images," *Comput. Med. Imag. Grap.*, vol. 35 6, pp. 472–80, 2011.
- [39] E. Pellegrini, G. Robertson, T. J. MacGillivray, J. van Hemert, G. Houston, and E. Trucco, "A graph cut approach to artery/vein classification in ultra-widefield scanning laser ophthalmoscopy," *IEEE Trans. Med. Imaging*, vol. 37, pp. 516–526, 2018.
- [40] T. Qureshi, M. Habib, A. Hunter, and B. Al-Diri, "A manually-labeled, artery/vein classified benchmark for the drive dataset," in *Proc. CBMS*, 2013, pp. 485–488.
- [41] P. Guimarães, J. Wigdahl, and A. Ruggeri, "A fast and efficient technique for the automatic tracing of corneal nerves in confocal microscopy," *Transl. Vis. Sci. Techn.*, vol. 5, p. Article 7, 2016.
- [42] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proc. ICUAL*, 1999, pp. 467–475.
- [43] D. Zhou and B. Schölkopf, "Learning from labeled and unlabeled data on a directed graph," in *Proc. ICML*, 2005, pp. 1036–1043.
- [44] H. Cai, X. Wang, H. Huang, and C. Ding, "Simultaneous image classification and annotation via biased random walk on tri-relational graph," in *Proc. ECCV*, 2012, pp. 823–836.
- [45] L. Cheng, J. De, X. Zhang, F. Lin, and H. Li, "Tracing retinal blood vessels by matrix-forest theorem of directed graphs," in *Proc. MICCAI*, 2014, pp. 626–633.
- [46] Q. Hu, M. D. Abramoff, and M. K. Garvin, "Automated construction of arterial and venous trees in retinal images," *Journal of Med. Imaging*, vol. 2, p. 044001, 2015.
- [47] T. A. Gillette, K. M. Brown, and G. A. Ascoli, "The DIADEM metric: Comparing multiple reconstructions of the same neuron," *Neuroinfo*, vol. 9, no. 2-3, pp. 233–245, 2011.

Retinal Vascular Network Reconstruction and Classification via Dominant Sets Clustering

Yitian Zhao, Jianyang Xie, Huaizhong Zhang, Yalin Zheng, Yifan Zhao, Hong Qi, Yangchun Zhao, Pan Su*,
Jiang Liu and Yonghuai Liu*

Abstract—The estimation of vascular network topology in complex networks is important in understanding the relationship between vascular changes and a wide spectrum of diseases. Automatic classification of the retinal vascular trees into arteries and veins is of direct assistance to the ophthalmologist in terms of diagnosis and treatment of eye disease. However, it is challenging due to their projective ambiguity and subtle changes in appearance, contrast and geometry in the imaging process. In this paper, we propose a novel method that is capable of making the artery/vein (A/V) distinction in retinal color fundus images based on vascular network topological properties. To this end, we adapt the concept of *dominant sets clustering* and formalize the retinal blood vessel topology estimation and the A/V classification as a pairwise clustering problem. The graph is constructed through image segmentation, skeletonization and identification of significant nodes. The edge weight is defined as the inverse Euclidean distance between its two end points in the feature space of intensity, orientation, curvature, diameter, and entropy. The reconstructed vascular network is classified into arteries and veins based on their intensity and morphology. The proposed approach has been applied to five public databases (INSPIRE, IOSTAR, VICAVER, DRIVE and WIDE) and achieved high accuracies of 95.1%, 94.2%, 93.8%, 91.1%, and 91.0%, respectively. Furthermore, we have made manual annotations of the blood vessel topologies for INSPIRE, IOSTAR, VICAVER, and DRIVE databases, and these annotations are released for public access so as to facilitate researchers in the community.

Index Terms—Retinal images, dominant sets clustering, blood vessel, vascular topology, Artery/vein classification

I. INTRODUCTION

The retinal blood vascular network is the only vascular network of the human body that is visible to a non-invasive imaging approach. In consequence, automated analysis of retinal vascular structure is the most common way to support examination, diagnosis and treatment of many diseases [1], [2],

This work was supported by National Science Foundation Program of China (61601029), Zhejiang Provincial Natural Science Foundation (LZ19F010001), and Ningbo Natural Science Foundation (2018A610055).

Y. Zhao, J. Xie, P. Su, and J. Liu are with Ningbo Institute of Industrial Technology, Chinese Academy of Sciences, China; H. Zhang and Y. Liu are with Department of Computer Science, Edge Hill University, UK; Y. Zheng is with Department of Eye and Vision Science, University of Liverpool, UK; H. Qi is with Department of Ophthalmology, Peking University Third Hospital, China; YF. Zhao is with School of Aerospace, Transport and Manufacturing, Cranfield University, UK; YC. Zhao is with the Second Affiliated Hospital of Zhejiang Chinese Medical University.

J. Liu is also with Department of Computer Science and Engineering, Southern University of Science and Technology, China

P. Su is also with School of Control and Computer Engineering, North China Electric Power University, China

* denotes corresponding author: supan@nimte.ac.cn, and yonghuai.liu@edgehill.ac.uk

such as diabetic retinopathy, hypertension and other cardiovascular diseases [3], [4]. Retinal arteriolar constriction, or arteriovenous nicking, significant dilatation and elongation of main arteries and veins, or vascular caliber and tortuosity change are frequently associated with the aforementioned diseases [5]. It is crucial to identify and distinguish the structures of individual blood vessels from the entire blood vessel network in a given fundus image. In particular, the Arteriolar-to-Venular Ratio (AVR) is considered to be an important characteristic sign that quantifies the severity of a wide spectrum of diseases [1], [6], [7], for example, low AVR - i.e., narrowing of arteries and widening of veins - is a direct biomarker for diabetic retinopathy. By contrast, a high AVR has been associated with higher cholesterol levels and inflammatory markers [8]. For these clinical observations, it would be of direct benefit if the retinal vascular network could be distinguished anatomically into different blood vessel branches, or separated into arteries and veins.

In practice, ophthalmologists use color and morphological information to discriminate between arteries and veins, since the arteries contain more oxygen and appear brighter than the veins, and thinner than neighboring veins [9]. These features of the retinal vasculature are usually captured by fundus photography, due to its lower cost and ease of use. Manual classification of retinal blood vessels is time consuming and subject to human errors. Therefore, an automatic vascular tracing method for topology reconstruction, or classification of blood vessels as arteries and veins, is highly desirable in seeking to overcome time constraints and avoid human errors. This calls for precise description of the vascular structure in terms of its color, topological, geometrical and morphological properties as derived from retinal images. However, it is highly challenging to discriminate arteries from veins, or perform any other measurement of interest (e.g., topology estimation), if the blood vessel widths are small, even after accurately segmenting the vasculature from the given fundus images. For example, existing methods often fail to trace correctly when faced with the occurrence of bifurcation and crossover at junction points: (i) the bifurcation – different blood vessel segments are from one blood vessel tree, and (ii) crossovers – two blood vessels overlap due to the projection of a 3D human eye to a 2D fundus image. These intersections usually lead to difficulty in predicting whether given blood vessels contacting a junction belong to the same blood vessel tree, or a different tree, due to their projective ambiguity and subtle changes in appearance, contrast and geometry in the imaging process.

To address these problems, we propose a novel topology

estimation and arteries/veins (A/V) classification method by adapting the concept of dominant sets clustering (DOS), in which a dominant set is used to represent a vascular tree. The novel method includes three main steps: (i) we re-conceptualize our previous work [10] and extend it for both the retinal blood vessel topology estimation and classification, (ii) formalize them as a pairwise clustering problem, and (iii) validate the proposed method over five public retinal datasets. DOS is a graph-theoretic approach, and is a well-known generalization of the notion of maximal cliques to an edge-weighted graph. It has been proven to work well in data clustering and image segmentation [11]–[13]. We aim to classify the entire retinal blood vessel network, not just the most prominent blood vessels, with a view to clarifying the underlying topology - how different blood vessels are anatomically connected to each other - and to identify and distinguish the structure of individual blood vessels from the entire blood vessel network. To classify the reconstructed vascular network into arteries and veins, we consider domain-specific knowledge about how they perfuse the retina, including angular, morphological, and textural properties of all blood vessel segments of the junction. We also take into account that arteries and veins usually alternate near the optic disc.

The main contributions of this paper are as follows:

- The concept of dominant sets clustering has been introduced to tackle the challenging problem of vasculature analysis, and proved to be an effective way of addressing the problem of tracing crossovers.
- The proposed method can split the entire blood vessel graph into several individual branches as subtrees, and is capable of demonstrating how different blood vessels are anatomically connected to each other. In addition, the A/V classification is undertaken on the topology-assigned blood vessel network, rather than the entire blood vessel segments.
- The proposed method has been validated quantitatively using five publicly accessible datasets, with promising results. In addition, the manual annotations of blood vessel topologies of four datasets were established as the ground truth, and have been released for public access¹.

The remainder of this paper is organized as follows. In Section II, the existing methods for retinal A/V classification and topology reconstruction are briefly reviewed. Section III details the proposed blood vessel topology estimation and A/V classification method. Section IV introduces the datasets, ground truth and evaluation metrics we will use. In Sections V and VI, we present the experimental results and discuss the effectiveness, robustness, and efficiency respectively of the proposed method. Finally, in Section VII we draw conclusions, and indicate directions for future work.

II. RELATED WORKS

In the past decade, extensive work has been carried out on automatic retinal blood vessel segmentation [14]–[18]. However, automated A/V classification and vascular topology

reconstruction are still understudied, despite their significance to understanding the structure and distribution of the blood vessels. The majority of existing blood vessel classification methods make use of machine learning techniques, using local features and adding structural information from the vascular tree.

Martinez-Perez et al. [19] proposed a semi-automatic retinal blood vessel analysis method that is capable of measuring and quantifying the geometrical and topological properties. It requires a human expert first to classify a branch as either vein or artery. Kondermann et al. [20] classified blood vessels by utilizing the Support Vector Machines (SVM) and neural networks. Relevant features were extracted from the blood vessel profile and the regions of interest centred at the pixels on the blood vessel centerlines. Niemeijer et al. [1], [21] found the k-Nearest Neighbour (kNN) classifier performing the best, by testing four different classifiers in the task of distinguishing between arteries and veins from a feature vector containing 27 elements. These elements were generated from blood vessel centerline pixels: a soft label was assigned to each centerline, implying the likelihood of the blood vessel being a vein. Mirsharif et al. [22] classified the blood vessels into arteries and veins by using multiple classifiers, and found that the Linear Discriminant Analysis (LDA) classifier had the best performance. Relan et al. [23] automatically classified retinal blood vessels as arteries or veins using color features and a Gaussian Mixture Model (GMM). Vazquez et al. [3] combined color-based clustering and blood vessel tracking to differentiate arteries from veins, and the tracking strategy based on the minimal path approach was employed to support the resulting classification by voting. Girard and Chriet [24] trained a convolutional neural network (CNN) for the task of assigning blood vessel pixels into arteries or veins. This approach propagated the blood vessel graph by using the minimum spanning tree. Huang et al. [25] introduced four new features to avoid distortions resulting from lightness inhomogeneity, and the accuracy of the A/V classification was improved by using the LDA classifier.

Due to demanding precise segmentation results in most of the existing methods, the ambiguity of small and mid-sized blood vessels makes the A/V classification a very difficult computational task. Pathological conditions and intensity inhomogeneities further complicate the task of A/V classification. More recently, the analysis of graphs extracted from the retinal blood vessel structure has been utilized to assist in blood vessel classification [8], [9], [26], [27]. This approach classifies the entire vascular tree by determining the type of all intersection pixels (graph nodes) and then assigning a label to each blood vessel segment, so as to reconstruct the underlying blood vessel topology with more accurate classification of small and mid-sized blood vessels. Rothaus et al. [26] presented an automated graph separation algorithm to distinguish between arteries and veins. Dashtbozorg et al. [8] proposed a method which first split the vascular graph into subgraphs by embedding the graph nodes and applying intensity features, and then, using LDA, assigned a label to each subgraph stating whether it was artery or vein. Joshi et al. [9], [27] separated their vascular graph into different subgraphs by using Dijkstra's shortest-

¹<http://imed.nimte.ac.cn/vetovessel-topology-groundtruth.html>

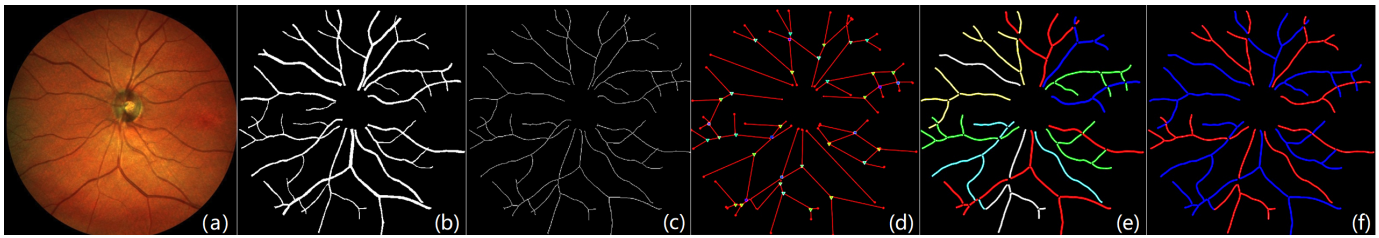


Fig. 1: Overview of the proposed method. (a) Original image. (b) Extracted blood vessels. (c) Skeletonized blood vessels. (d) Graph generated with significant nodes overlaid. (e) Estimated vascular network topology. (f) Classified arteries and veins: arteries shown in red, and veins in blue.

path algorithm, then labeled each subgraph as either artery or vein using a fuzzy C-means clustering algorithm. Estrada et al. [5], [28] utilized a global likelihood model to capture the structural plausibility of each blood vessel, and employed a graph-theoretic method to estimate the overall blood vessel topology with domain-specific knowledge and applied three features, local growth, overlap and color, to accurately classify the A/V types. Lyu et al. [29] used a curvature orientation histogram to extract blood vessel landmarks from the blood vessel centerline, and separated the different blood vessel trees by curvature orientation clustering. De et al. [30], [31] proposed a graph-theoretical approach to reconstruct the blood vessel network from topological information. They adapted the label propagation over directed graphs, and by this method the graph was partitioned into disjoint subgraphs. However, these topology reconstruction or subgraph estimation methods often fail to achieve accurate results due to the difficulty in identifying the presence of crossovers at blood vessel junctions, as the caliber, angular or other measurements are unreliable at points of abrupt change from one blood vessel to another.

III. METHOD

The proposed method adapts the dominant set clustering for retinal topology reconstruction and A/V classification, and the overview of the method is illustrated in Fig. 1. Fig. 1(a) shows an example retinal color fundus image from the public IOSTAR dataset, and Fig. 1(b) illustrates its extracted vasculature. Then a skeletonization step is applied to the extracted blood vessels, as shown in Fig. 1(c). A vessel graph is generated by removing the junction pixels from Fig. 1(c), and Fig. 1(d) depicts the generated graph with significant nodes overlaid. The blood vessel network is separated into several individual tree branches in different colors, as shown in Fig. 1(e). Finally, these tree branches are classified into two categories: arteries (red) and veins (blue). The generation of the graph, the separation and classification of the blood vessel network into arteries and veins are detailed below in Section III-A, Sections III-B and III-C, and Section III-D respectively.

A. Graph generation

Firstly, the optic disc is masked in order to avoid morphological complications due to the irregular and highly tortuous

blood vessels at the optic disc: the superpixel-based optic disc segmentation [32] was utilized for its fast speed and effectiveness. Then the infinite perimeter active contour with hybrid region (IPACHR) method proposed in [33] is employed to automatically segment the retinal blood vessels for its effectiveness in detecting vessels with irregular and oscillatory boundaries, as shown in Fig. 1(b). An iterative morphology thinning operation [34] is finally performed on the extracted blood vessels to obtain a single-pixel-wide skeleton map, as shown in Fig. 1(c).

The vascular bifurcations, crossovers, and blood vessel ends (terminal points) may then be extracted from the skeleton map by locating intersection points (pixels with more than two neighbors) and terminal points (pixels with one neighbor). All the intersection points and their neighbors are then removed from the skeleton map, producing an image with clearly separated blood vessel segments. A blood vessel graph can be generated from this skeleton map by linking the first and last nodes in the same blood vessel segment. The generated graph will usually include some misrepresentations of the blood vessels: typical errors include *node splitting*, *missing links* and *false links*. Correction of these errors was achieved by employing the strategy proposed in [8], which considers the local vessel calibers and angles to reconnect the missing links and correct the false links. Fig. 1(d) shows the blood vessel graph, in which red dots indicate terminal points, green triangles bifurcations, and blue squares crossover points.

The intersections extracted from the skeleton map are critical to the final topology estimation. In the following, the details of partitioning the points in each intersection into different vascular segments are described. The analysis of the intersections is broken down into three categories, according to the number of points involved in each intersection: connecting points (2), bifurcation points (3, 4), and crossovers/meeting points (3, 4, 5), where the number in the brackets after each category indicates the number of vascular segments connected to that intersection. Again, the method proposed by Dashtbozorg et al. [8] is used to handle the cases of nodes of degree 2. For the more complicated categories, nodes of degree 3 to 5, a clustering method based on the dominant sets is proposed as follows.

The concept of dominant sets arises from the study of graph theory, by which a continuous formulation of the maximum clique problem is defined in [11]. An undirected graph G

with weighted edges is represented as $G = (V, E, \omega)$, where V is a set of nodes, edge set $E \subseteq V \times V$ indicates all the connections of the relevant nodes, and $\omega : E \rightarrow \mathbb{R}^+$ is the positive weight function. Fig. 2 shows an example of the edge-weighted graph extracted from a representative junction of a retinal blood vessel network.

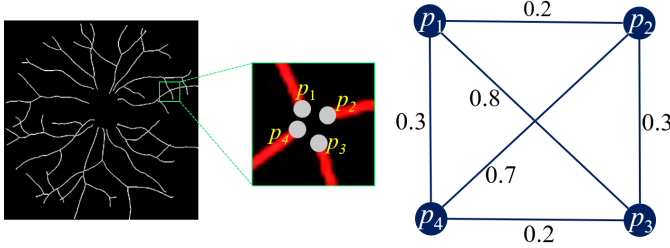


Fig. 2: An example of the edge-weighted graph.

In the context of vascular topology estimation and arteries/veins identification, V is a set of extracted pixels from a retinal color fundus image and ω represents the similarity among the pixels in V . A $|V| \times |V|$ symmetric matrix $A = \{a_{ij}\}$ is used to represent the weighted graph G , which is named an adjacency matrix. The value of a_{ij} is derived by a similarity measure defined in the feature space of the pixels. Here, we define $a_{ij} = 0$ for $i = j$, which indicates that the generated graph G does not include self-loop.

B. Dominant sets

A dominant set can be formally defined based on the values of similarity among the nodes in V . Let $S \subseteq V$ be a nonempty subset of nodes, $p_i \in S$ and $p_j \in S \setminus \{p_i\}$, where $S \setminus \{p_i\}$ indicates the nodes in the set S excluding the node p_i . The relative similarity between p_i and p_j with respect to the average similarity between p_j and its neighbours in $S \setminus \{p_i\}$ can be defined as:

$$\phi_{S \setminus \{p_i\}}(p_i, p_j) = a_{ij} - \frac{1}{|S| - 1} \sum_{p_k \in S \setminus \{p_i\}} a_{jk}. \quad (1)$$

where $|\cdot|$ denotes the cardinality of a set. For example, Fig. 2 depicts an edge-weighted graph generated for an intersection which is extracted from Fig. 1, where the weights of the edges indicate the similarity amongst four pixels $\{p_1, p_2, p_3, p_4\}$. Given $S = \{p_1, p_2, p_3\}$ we have: $\phi_{\{p_1, p_3\}}(p_2, p_1) = a_{21} - (a_{11} + a_{13})/2 = -0.2$ and $\phi_{\{p_1, p_3\}}(p_2, p_3) = a_{23} - (a_{31} + a_{33})/2 = -0.1$. It can be observed that $\phi_S(p_i, p_j)$ may be either positive or negative.

The weight $W_S(p_i)$ of p_i with regard to S is defined recursively as:

$$W_S(p_i) = \begin{cases} 1, & \text{if } |S| = 1 \\ \sum_{p_j \in S \setminus \{p_i\}} \phi_{S \setminus \{p_i\}}(p_i, p_j) W_{S \setminus \{p_i\}}(p_j), & \text{otherwise.} \end{cases} \quad (2)$$

where $W_S(p_i)$ expresses the similarity between node p_i and the nodes of $S \setminus \{p_i\}$ with respect to the mutual similarity amongst the nodes in $S \setminus \{p_i\}$. Finally, the total weight of S is calculated as $W(S) = \sum_{p_i \in S} W_S(p_i)$. Take the calculation of $W_{\{p_1, p_3, p_4\}}(p_4)$ as an example (shown in Fig. 3), we have:

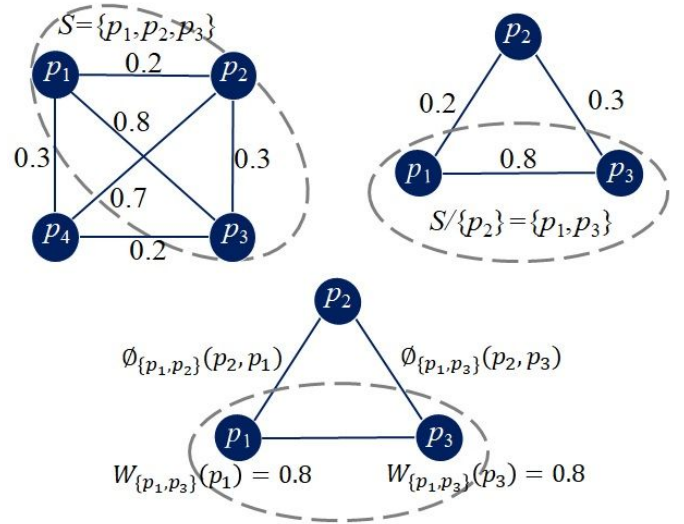


Fig. 3: An example of decomposed recursive calculation of similarity over a graph for dominant sets.

$$W_{\{p_1, p_3\}}(p_1) = W_{\{p_1, p_3\}}(p_3) = 0.8, W_{\{p_1, p_2, p_3\}}(p_2) = \phi_{\{p_1, p_3\}}(p_2, p_1)W_{\{p_1, p_3\}}(p_1) + \phi_{\{p_1, p_3\}}(p_2, p_3)W_{\{p_1, p_3\}}(p_3) = -0.24, \text{ and } W_{\{p_1, p_3, p_4\}}(p_4) = -0.24.$$

Formally, a non-empty subset S of nodes V , $S \subseteq V$ such that $W(S') > 0$ for any non-empty subset $S' \subseteq S$ is said to be a dominant set if:

$$W_{S'}(p_i) > 0, \text{ for all } p_i \in S' \quad (3)$$

and

$$W_{S' \cup \{p_j\}}(p_j) < 0, \text{ for any } p_j \notin S'. \quad (4)$$

Take the graph shown in Fig. 3 as an example: $W(\{p_1, p_3\}) = W_{\{p_1, p_3\}}(p_1) + W_{\{p_1, p_3\}}(p_3) = 1.6$, $W(\{p_1\}) = W(\{p_3\}) = 1$, and $W_{\{p_1, p_3\}}(p_1) > 0$, $W_{\{p_1, p_3\}}(p_3) > 0$, $W_{\{p_1, p_2, p_3\}}(p_2) < 0$, $W_{\{p_1, p_3, p_4\}}(p_4) < 0$. Therefore, p_1 and p_3 form a dominant set. Similarly, p_2 and p_4 form another dominant set.

In general, the weights of edges within the dominant set of an edge-weighted graph should be large, representing high internal homogeneity or similarity [11]. By contrast, the weights of edges which link to the dominant sets externally will be small. Therefore, the dominant set is a proper solution to identify branches of a vascular tree, because the similarity of two points from the same branch should be large within an intersection, while that of two points belonging to different branches should be small.

Dominant sets can be identified by local solutions of a standard quadratic program:

$$\begin{aligned} & \text{maximize}_{\mathbf{x}} && f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x} \\ & \text{subject to} && \mathbf{x} \in \Delta, \end{aligned} \quad (5)$$

where

$$\Delta = \left\{ \mathbf{x} \in \mathbb{R}^{|V|} : \sum_{i=1}^{|V|} x_i = 1 \text{ and } x_i \geq 0 \text{ for all } i = 1, \dots, |V| \right\}.$$

A strict local solution \mathbf{x}^* of Eqn. (5) is named the weighted characteristic vector, where $x_i > 0$ means that the node p_i in

Algorithm 1: RDDOS(A , $MaxIteration$)

Inputs: A : the symmetric matrix $A = \{a_{ij}\}$ of a weighted graph G ;
Outputs: D : a set of nodes which forms a dominant set of G ;

- 1: initialize a vector $\mathbf{x}^{(0)} \in \mathbb{R}^{|V|}$ which satisfies: $x_i^{(0)} \geq 0$ for all $i = 1, \dots, |V|$ and $\sum_{i=1}^{|V|} x_i^{(0)} = 1$
- 2: $D = \emptyset$
- 3: **for** $t = 0 : MaxIteration - 1$ **do**
- 4: **for** $i = 1 : |V|$ **do**
- 5: $x_i^{(t+1)} = x_i^{(t)} \frac{(A\mathbf{x}^{(t)})_i}{\mathbf{x}^{(t)\top} A\mathbf{x}^{(t)}}$
- 6: **end for**
- 7: **end for**
- 8: **for each** $x_i^{(MaxIteration)} > 0$ **do**
- 9: $D = D \cup p_i$
- 10: **end for**

question is in a dominant set of G . An effective optimization approach for solving Eqn. (5) is given by the *replicator dynamics* [12], [35]:

$$x_i^{(t+1)} = x_i^{(t)} \frac{(A\mathbf{x}^{(t)})_i}{\mathbf{x}^{(t)\top} A\mathbf{x}^{(t)}}, \quad (6)$$

where $i = 1, 2, \dots, |V|$. It has been proven that as t in Eqn. (6) increases, for any initialization of $\mathbf{x}^{(0)} \in \Delta$, the trajectory remains in Δ and the objective function $f(\mathbf{x})$ in Eqn. (5) is either strictly increasing or constant. Therefore, $\mathbf{x}^{(0)}$ can be initialized by generating $|V|$ random numbers and then normalizing them. In the following algorithms, we use $S=RDDOS(A, MaxIteration)$ in Algorithm 1 to represent the procedure of extracting a dominant set S from a weighted graph given its adjacency matrix A by using replicator dynamics. The stopping criterion of the dynamic system can be set with a maximal iteration number $MaxIteration$. In this paper, $MaxIteration$ is empirically set to 30 in the experiments. **Take the graph shown in Fig. 3 as an example, the matrix A is:**

$$\begin{pmatrix} 0.0 & 0.2 & 0.8 & 0.3 \\ 0.2 & 0.0 & 0.3 & 0.7 \\ 0.8 & 0.3 & 0.0 & 0.2 \\ 0.3 & 0.7 & 0.2 & 0.0 \end{pmatrix},$$

given that $\mathbf{x}^{(0)}$ is initialized as $(0.4, 0.3, 0.2, 0.1)$, after the first iteration of the replicator dynamics defined in Eqn. (6), $\mathbf{x}^{(1)} = (0.35, 0.22, 0.30, 0.13)$. \mathbf{x} is finally converged to $(0.5, 0.0, 0.5, 0.0)$ after 15 iterations, and we have $x_1 > 0$ and $x_3 > 0$ which indicates that p_1 and p_3 form a dominant set, as expected.

C. Topology estimation via dominant sets clustering

A peeling-off strategy has been proposed in Algorithm 2, which shows the complete procedure of partitioning the points involved in the intersections into branches. It iteratively extracts a subset of points belonging to the same branch (a dominant set S) each time by using Eqn. (6) and repeating the process with the remaining points in the set $V = V \setminus S$. The identification of different vascular branches at an intersection is carried out by identifying one vascular branch first, and then

Algorithm 2: TopologyEstimation(I)

Inputs: V : a set of points that is associated with intersections;
Outputs: S^* : a partition of V , each element of which includes points of the same branch;

- 1: initialize the $|V| \times |V|$ symmetric matrix $A = \{a_{ij}\}$ by calculating $a_{ij} = \omega(i, j)$ with respect to given features;
- 2: $S^* = \emptyset$
- 3: **while** $V \neq \emptyset$ **do**
- 4: $S = RDDOS(A, 30)$
- 5: **for each** $p_i \in S$ **do**
- 6: remove the column and row with respect to p_i from A
- 7: **end for**
- 8: $V = V \setminus S$
- 9: $S^* = S^* \cup \{S\}$
- 10: **end while**

identifying another from the remaining nodes and segments, and so on. Therefore, the peeling-off strategy is a direct, intuitive implementation of this procedure [13]. As shown in the aforementioned example of Fig. 3, given that $\mathbf{x}^{(0)}$ is initialized as $(0.4, 0.3, 0.2, 0.1)$, the first iteration of Algorithm 2 identifies the dominant set including p_1 and p_3 , subsequently the second iteration identifies another including p_2 and p_4 , as expected.

It is also worth noting that the number of points processed in the process of identifying a dominant set is greatly reduced over different iterations, from the number of intersections in a vascular network after the pre-processing step proposed in Section III to those excluding the vascular branches already identified. This leads to an increase of computational efficiency in the topology estimation.

Each point p_i in V is represented as a feature vector \mathbf{F}_i : its intensities in R , G , and B channels, orientations, curvatures, blood vessel diameters, and entropy. The values of each feature are then linearly normalized individually into the unit interval $[0, 1]$. The weight $\omega(i, j)$ of an edge connecting points p_i and p_j is finally estimated as: $\omega(i, j) = 1/(\|\mathbf{F}_i - \mathbf{F}_j\| + \epsilon)$, where $\|\cdot\|$ denotes the Euclidean distance between \mathbf{F}_i and \mathbf{F}_j . Note that a tiny value $\epsilon = 0.000001$ is used in case of $\mathbf{F}_i = \mathbf{F}_j$.

For each pixel in a vessel segment, 23 features are measured which are listed in TABLE I and are mainly related to the orientation, diameter, and curvature. Most of these features were previously used for topological reconstruction [8], [9] and vessel classification [1], [20] tasks. While other features can be calculated straightforward, the main ones are defined and explained as follows: the orientation of a pixel is defined as the including angle (in radians) between its lying segment and the positive direction of the x -axis, a measurement between $[0, \pi]$ [9]; the diameter of a pixel is measured as a median value of 15 measurements of diametric length between the vessel edges, and passing through the skeleton pixels of the end region [9]; the curvature C at pixel p is computed as $\frac{\Delta x_p \Delta^2 y_p - \Delta^2 x_p \Delta y_p}{[(\Delta x_p)^2 + (\Delta y_p)^2]^{3/2}}$, where x and y are two different coordinates, and Δ and Δ^2 are the first and second order derivatives of the image at the pixel p .

We employed the sequential forward floating selection for feature selection [1], which starts with an empty feature set and

TABLE I: List of features for the representation of points for the proposed DOS classifier.

No.	Features
1-6	avg. and std. of the intensities within the segment in RGB channels
7-10	avg. and std. of the orientations and curvatures of each centerline pixels
11-12	avg. and std. of the blood vessel diameters of each blood vessel segment
13-18	avg. and std. of the intensities of all centerline pixels under a Gaussian blurring ($\sigma = 4$) in RGB channels
19-23	entropy of intensity in RGB channels, orientation and curvature of each centerline pixels

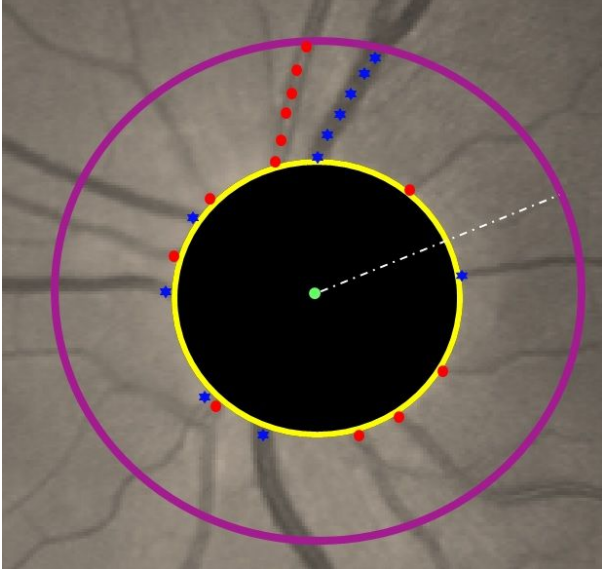


Fig. 4: Definition of a ring for the separation of arteries (red circles) from veins (blue stars).

adds or removes features when this improves the performance of the classifier. (We refer readers to TABLE I for more details about the selected features.) Fig. 1(e) illustrates the estimated vascular network with the added topological information.

D. Arteries/Veins classification via dominant sets clustering

As aforementioned, several supervised learning methods have been exploited for the task of A/V identification. Their performance is limited due to unavailability of sufficient training data, and complicated structure and training process. In this paper, we take into account the reconstructed vascular topology information, utilize the difference of structural information between arteries and veins reflected in retinal images, and propose the ratio of ‘dominant pixels’ as the criterion for an unsupervised A/V identification.

As each branch of retinal blood vessels is composed of many pixels in a retinal image, significant information for discriminating arteries from veins can be derived from their characteristics, and the configuration of the pixels. Dominant pixels are defined as those which form an extracted dominant set amongst all the pixels in a branch of a retinal blood vessel. The dominant pixel ratio is defined as the ratio of dominant

Algorithm 3: DOS-AVIdentification(I)

Inputs: I : a set of pixels that is associated with vascular branches in the retinal color fundus image;

Outputs: L : a label of I which indicates its group;

- 1: initialize the $|I| \times |I|$ symmetric matrix $A = \{a_{ij}\}$ by calculating $a_{ij} = \omega(i, j)$ with respect to given features;
- 2: $S = \text{RDDOS}(A, 10000)$
- 3: **if** $\frac{|S|}{|I|} \geq 0.5$ **then**
- 4: $L = \text{Group-A}$
- 5: **else**
- 6: $L = \text{Group-B}$
- 7: **end if**

pixels to all the pixels in a branch of a retinal blood vessel. Based on these definitions, an effective and concise rule is proposed to separate all the branches of blood vessels into two groups: a branch with a high dominant pixel ratio (given 0.5 or greater) is assigned to Group-A, while Group-B includes those branches with a low dominant pixel ratio. Group-A and Group-B are then assigned as either artery or vein based on their intensity and morphology: arteries are brighter in color than veins, and are thinner than neighboring veins [20]. **In this work, we define a vessel segment as ‘artery’ if the average intensity value of the pixels inside the ring centred at any pixel on its centerline with the same radius as the optic disc is larger than 0.48 (with the intensity values lying in the unit interval [0, 1]) (Fig. 4). Otherwise the vessel segment is assigned as ‘vein’. The threshold value was chosen empirically.** Algorithm 3 shows the details for the identification of Group-A/Group-B pixels via DOS.

The power of the dominant pixel ratio definition for A/V identification is derived from the effectiveness of DOS in disclosing the hidden structure in arteries and veins. As with the dominant sets-based vascular topology estimation, the weights assigned to the pixels of a vascular branch is responsible for the quality of A/V identification. Once again the features in TABLE I are used to calculate the weights $\omega(i, j)$. Fig. 1(f) illustrates the A/V classification result.

IV. MATERIALS

A. Datasets

Topology reconstruction: Five public retinal datasets were used - the Iowa Normative Set for Processing Images of the RETina (INSPIRE) by the University of Iowa Hospital [1]; the VARPA Images for the Computation of the Arterio/Venular Ratio (VICAVR) captured at the hospital of Conxo, Santiago de Compostela, Spain [3]; the images acquired with an ultra-wide-field device (WIDE) by Duke University [5]; the Scanning Laser Ophthalmoscopy technique-based dataset (IOSTAR) from Eindhoven University of Technology [36]; and the well-known Digital Retinal Images for Blood vessel Extraction (DRIVE) dataset [37]. Except the WIDE dataset, the others used for topology estimation do not contain manually-labeled blood vessel topological information. In consequence, we have annotated them manually, and released these annota-

tions² online.

A/V classification: Four of the above mentioned datasets were used: INSPIRE, VICAVER, WIDE, and DRIVE. **An image analysis expert manually classified and then an ophthalmologist checked and corrected** the blood vessel segments of the INSPIRE, WIDE, and DRIVE datasets into arteries and veins. For the VICAVER dataset, three experts manually labeled the blood vessel segments as arteries or veins, and agreements between them were then used as groundtruth. However, the experts only classified the blood vessels found within a radius three times that of the optic disc: in other words, the A/V classification labelling for this dataset is incomplete. For a comprehensive visual illustration of A/V labels and fair comparison with other datasets, we asked again **the image analysis expert to classify and then the ophthalmologist to check and correct** all the unlabeled blood vessels as arteries or veins for this dataset.

It is worth noting that the manual labeling for the INSPIRE and WIDE datasets is at blood vessel centerline pixel level, whereas that for the VICAVER, DRIVE, and IOSTAR datasets is at blood vessel segment pixel level instead. TABLE II summarizes the important details of all these datasets.

B. Topology ground truth estimation

As noted above, the DRIVE, INSPIRE, VICAVER, and IOSTAR datasets have no ground truth for blood vessel topology, so two experts were invited to manually label the topological information of the retinal vascular structure by using a graph editing software we developed for this task. Two experts independently labeled each blood vessel segment or centerline for all the datasets, based on the types of available manual annotations or automatic segmentation of the blood vessel structure. The consensus between them was then used as ground truth. **Note, we obtained the topology ground truth of DRIVE, INSPIRE, and IOSTAR based on manual segmentations, and for VICAVER, we generated the topology ground truth on automatic segmented vessels by using the method proposed in [33].** For each image, every blood vessel was labeled with a distinct color (or individual label) to clearly distinguish between different blood vessel trees, as shown in Fig. 1(e).

Thus, the blood vessel segments, or blood vessel centerlines used for topology estimation were extracted either by **human graders manually or applying** an automatic blood vessel segmentation method [33]: i.e., the DRIVE and IOSTAR datasets include the manual annotations of the retinal blood vessels for each image, so the topology reconstruction was made on manual annotated blood vessel patterns; for the VICAVER dataset, the topology estimation was made at the automatically segmented blood vessels by using the automated segmentation method [33]: for the INSPIRE dataset, **the image analysis expert graded and then the ophthalmologist checked and corrected** the topology at the blood vessel centerlines, which were generated by the method given in [5].

C. Evaluation metrics

To the best of our knowledge, there is no single metric that can measure the performance of a topology estimation procedure, because of the differing properties of multiple scales of the vasculars, i.e., blood vessels of different sizes exhibit properties that cannot be captured by a single metric. Therefore, in this work, we measure the performance of the topology estimation method by calculating the overlap rate of the correctly identified nodes/intersections. Let A be the total number of the nodes of a blood vessel tree, and B be the number of the nodes that have been correctly identified (the vertices of the estimated topology tree have been assigned labels identical to the ground truth). The overlap rate is then defined as $p = \frac{B}{A} \times 100\%$. Intuitively, p is the percentage of nodes that are correctly identified by the proposed method [30].

To compare the A/V classification performance of the proposed method with the corresponding ground truth as annotated by our human graders, the following metrics were employed: *sensitivity* (Se) = $TP/(TP + FN) \times 100\%$, *specificity* (Sp) = $TN/(TN + FP) \times 100\%$, and *balanced accuracy* ($B-Acc$) = $\frac{Se+Sp}{2}$, where TP , TN , FP and FN denote true positive, true negative, false positive, and false negative, respectively. For the sake of fair comparison and easy implementation, all the evaluations were undertaken at blood vessel centerline pixel level. In this work, we interpret arteries as positives and veins as negatives. Se shows the ability of a given method to detect arteries, while Sp reveals how well it detects veins. $B-Acc$ indicates the overall classification performance, and thus reflects the trade-off between sensitivity and specificity [28].

V. EXPERIMENTAL RESULTS

In order to validate the effectiveness of the proposed DOS algorithm for retinal vascular structure analysis, we have thoroughly evaluated it over five publicly accessible datasets as described in the last section.

A. Topology estimation

Fig. 5 illustrates five vascular topology reconstruction results, each derived from one of these retinal datasets: INSPIRE, IOSTAR, VICAVER, DRIVE, and WIDE, respectively. Compared with the manual annotations, as shown in the second column of Fig. 5, it is clear from visual inspection that our method is able to identify most nodes correctly: only a few cases were incorrectly traced, and these were located at crossovers, as shown in the last column of Fig. 5 - the red squares indicate incorrectly-traced significant nodes.

To better facilitate observation and objective evaluation of the performance of the proposed method in the reconstruction of blood vessel topology, the overlap rates p of the relevant significant nodes were also calculated. TABLE III presents the performance measurements of the proposed method in identifying connectivity of these nodes at bifurcations (BIF) only, crossovers (CRO) only, and all nodes, by counting the number of true positives (TP: correctly identified nodes). As expected, the p scores for bifurcation points for all the five

²<http://imed.nimte.ac.cn/vetovessel-topology-groundtruth.html>

TABLE II: Details of the retinal image datasets used, including the availability of the ground truth for each dataset, and the type of manual annotations.

Datasets	No. Img.	Size	FOV	Camera	Ground truth	Label Type
INSPIRE [†]	40	2392 × 2048	30°	Carl Zeiss Meditec	Topology estimation & AV classification	Centerline
VICAVR [†]	100	768 × 576	45°	Cannon CR6-45NM	Topology estimation & AV classification	Segment
DRIVE [†]	40	565 × 584	45°	Cannon CR5-NM3-CCD	Topology estimation & AV classification	Segment
WIDE	30	1440 × 900	45°	Optos 200Tx	Topology estimation & AV classification	Centerline
IOSTAR [†]	30	1024 × 1024	45°	EasyScan	Topology estimation	Segment

[†] These datasets originally had no blood vessel topological groundtruth. We manually annotated them.

TABLE III: Performance of the proposed method on topology reconstruction at node and segment level, respectively, over different datasets.

	INSPIRE	IOSTAR	VICAVR	DRIVE	WIDE
# BIF	1998	1213	4955	2478	3678
# TP	1945	1167	4799	2288	3551
<i>p</i> score	97.3%	96.2%	96.8%	92.3%	96.5%
# CRO	778	482	1421	832	1230
# TP	697	431	1182	728	1107
<i>p</i> score	89.7%	89.4%	83.2%	87.5%	90.0%
<i>ave. p</i>	95.1%	94.2%	93.8%	91.1%	94.9%
Accuracy	97.5%	95.7%	94.6%	93.5%	96.4%

datasets are much higher than those at crossovers, since it is a relatively simpler node analysis task. The average overlap score over both bifurcations and crossovers was also calculated for all the datasets, and it reveals that the proposed method has the best performance on the INSPIRE dataset with an average $p = 95.1\%$, while yielding the lowest score on the DRIVE dataset, with an average $p = 91.1\%$. This is due to the relatively low resolution and image contrast of this dataset.

In addition, the accuracy of the topology reconstruction at blood vessel centerline level was obtained, and it can be observed that the accuracy scores are higher than the corresponding average overlap rate: for example, we see an accuracy score of 97.5%, but an average overlap rate of only 95.1% for the INSPIRE dataset. This is because the number of blood vessel segments is much larger than the number of significant nodes (bifurcation and crossovers).

As we mentioned above, there is no single metric that is able to directly and objectively measure the performance of a topology reconstruction method. The most common method as described in [5], [30] was topological structure-guided A/V classification. Therefore, in the following subsection, we continue to make use of the proposed DOS-based topology estimation method to guide the task of A/V classification.

B. A/V classification

After the reconstruction of the blood vessel topology, the complete blood vessel network has been separated into several individual branches as sub-trees with an individual label (distinct color). The final goal is to assign these labels to one of two classes: artery or vein.

The features listed in TABLE I and the DOS classifier were utilized again to classify these individual labels into two clusters, A and B . For each sub-tree v , the probability of its being A is estimated by the number of blood vessel pixels

classified by DOS as A : $P_A^v = n_A^v / (n_A^v + n_B^v)$, where n_A^v is the number of pixels classified as A , and n_B^v is the number of pixels classified as B . For each sub-tree, the higher probability is used to define whether the sub-tree is assignable to category A or B . Clusters A and B are then assigned as artery and vein, respectively, based on their intensity and morphology: arteries are brighter in color than veins, and are thinner than neighboring veins [20].

Fig. 6 shows the A/V classification results of the proposed method over four sample images, one from each dataset, based on their topological information. Overall, the proposed method can correctly distinguish most of the A/V labels on all the four datasets, in comparison with the corresponding manual annotations. The arteries are shown in red and veins in blue. However, our A/V classification relies heavily on the prior results of topology reconstruction: if a blood vessel segment was incorrectly identified at the stage of topology reconstruction, it might be very likely to be falsely labeled during subsequent A/V classification. This is because a blood vessel segment must share its label with all its downstream segments, as evidenced by the figures in the second and third columns of Fig. 6 - the incorrectly traced veins are labeled in green, while the falsely identified arteries are labeled in yellow.

In order to better demonstrate the superiority of the proposed method, TABLE IV reports the comparison of our method with the state-of-the-art A/V ones over four datasets, in terms of centerline pixel-level sensitivity, specificity, and balanced accuracy. It is shown that our method correctly identifies over 93.5% of the A/V labels of all the datasets, outperforming all the compared methods, with a single exception that its Sp score on the DRIVE dataset is 1.4% lower than that of [28]. Nevertheless, the proposed method is able to correctly identify the majority of arteries and veins by relying on the topological constraint.

The intra-observer scores are also provided for the INSPIRE and WIDE datasets in TABLE IV. It can be seen that the sensitivities of the proposed method are very close to those of the human observers: our method obtains competing sensitivities, with 96.8% and 96.2%, compared to 97.2% and 97.0% by the intra-observers for the INSPIRE and WIDE datasets, respectively.

Overall, the proposed method has the best performance on the INSPIRE dataset, with a balanced accuracy of 96.4%. This dataset contains images with a higher resolution, and considerably fewer and simpler blood vessel bifurcations and crossovers (e.g., an average number 40 vs. 102 of circuits

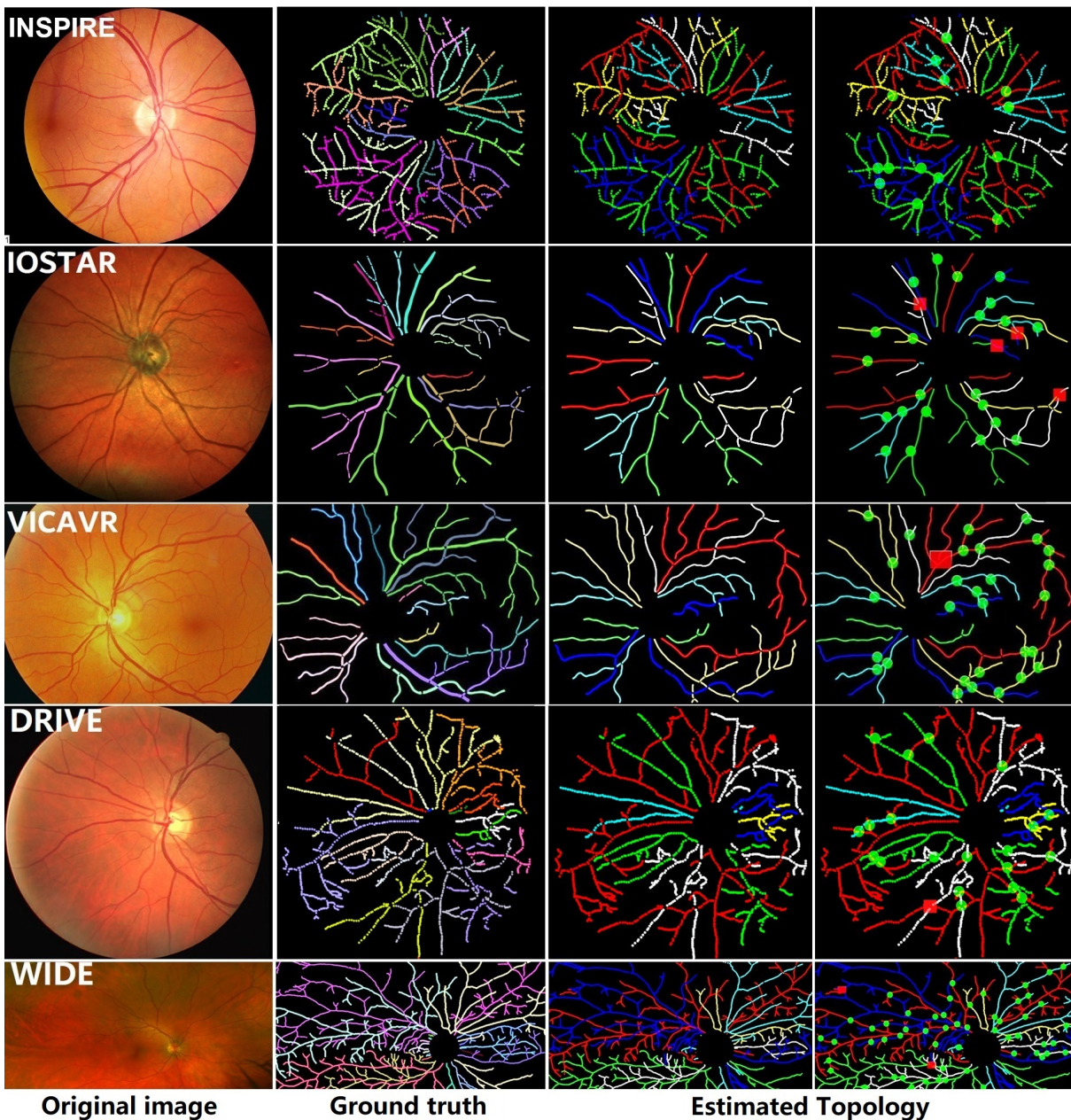


Fig. 5: Examples of vascular topology estimation results. From left to right column: original image, manual annotation, result of the proposed method, and the highlighted correctly (Green discs) and incorrectly (Red squares) identified connections.

when compared to the WIDE dataset). This finding is also evidenced by the performance on the topology estimation, with the highest accuracy of 97.5%, when compared with other datasets in TABLE III.

All the above-reported performances of different methods over the DRIVE dataset were obtained when using the manual annotation produced by Qureshi et al. [40] as ground truth. It is noted that the original authors of the DRIVE dataset also released another manual A/V annotation [37]. We refer them as GT-1 and GT-2, respectively. Fig. 7 shows two examples of these two manual A/V annotations.

In order to demonstrate our A/V classification method over two different manual annotations, and the agreement of those manual annotations, we computed the *false discovery rate*

(FDR) [41] in three different scenarios: Auto vs GT-1, Auto vs GT-2, and GT-1 vs GT-2. FDR is defined as the fraction of the total number of pixels incorrectly identified over the total number of pixels in ground truth. It can be observed from TABLE V that the proposed automatic A/V classification method has a larger FDR than the human annotations, as expected, and the two human annotations also have an FDR of as large as 0.071 ± 0.014 . Interestingly, Estrada et al. [28] also reported a very close agreement score of 96% between GT-1 and GT-2. These differences between the annotations given by different human graders imply the difficulty of the A/V classification problem.

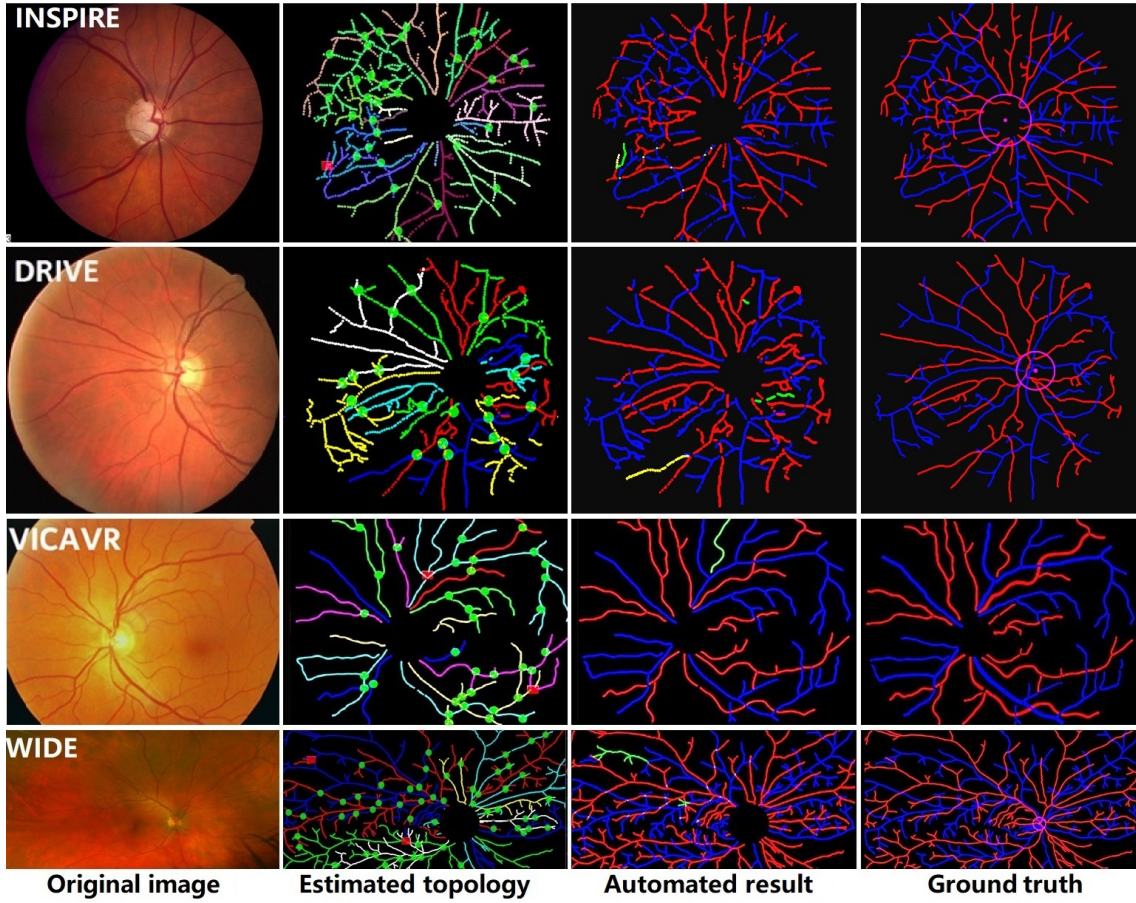


Fig. 6: A/V classification results of the proposed method over four different datasets. From left to right column: original image, blood vessel topology, A/V classification result, and corresponding manual annotations.

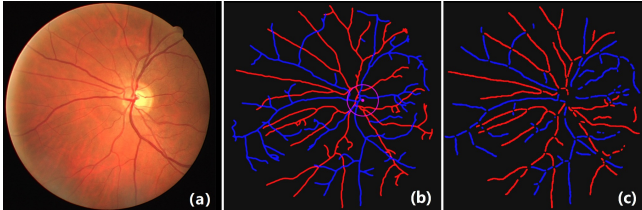


Fig. 7: Different manual A/V annotation results of the DRIVE dataset. (a) original image, (b) manual annotation by [40] (GT-1), (c) manual annotation by [37] (GT-2).

TABLE V: False discovery rate of different A/V classification methods over two different manual annotations: GT-1 and GT-2. Auto refers to the proposed automated A/V classification method.

	Auto vs GT-1	Auto vs GT-2	GT-1 vs GT-2
FDR	0.093±0.009	0.088±0.013	0.071±0.014

C. Computational complexity

The proposed method has a computational complexity of $O(N^2)$ in topology estimation, where N is the total number of nodes in a blood vessel graph, a computational complexity of $O(M \times P^2)$ in label assignment of each vessel branch,

where M is the total number of individual vessel branches in a vessel graph, and P is the average number of pixels of an individual vessel branch and a computational complexity of (M) in identifying the labels of the vessel branches as either artery or vein. Therefore, the entire computational complexity of the proposed method is $O(N^2) + O(M \times P^2)$ and thus $O(N^2)$ where $M \leq N$, $P \leq N$, and $MP \leq N$.

The average running times of each image with standard deviation in seconds for overall A/V classification and topology estimation of our method are: $72.61 \pm 16.32s$ and $2.83 \pm 0.41s$ over INSPIRE; $24.43 \pm 2.31s$ and $2.10 \pm 0.33s$ over DRIVE; $28.98 \pm 0.48s$ and $2.57 \pm 1.01s$ over VICA VR; and $33.77 \pm 2.01s$ and $5.84 \pm 1.37s$ over WIDE respectively. Note, we did not perform the A/V classification over IOSTAR due to lack of manual annotation, so we only report the running time of its topology estimation as $1.93 \pm 0.80s$. All the experiments were carried out in MATLAB2015a on a PC with an Intel Xeon E5-2695 v4 CPU, 2.10GHz, and 128GB RAM.

In the literature, only the method proposed by Estrada et al. [28] reported the computation time of their A/V classification on a Toshiba Satellite X870 laptop with a 2.4Ghz Intel I7 quad-core processor and 32GB of RAM: $117.68 \pm 34.10s$ over INSPIRE; $131.32 \pm 33.40s$ over DRIVE; and $777.35 \pm 330.52s$ over WIDE.

TABLE IV: The sensitivity Se , specificity Sp and balanced accuracy $B-Acc$ of different A/V classification methods over different datasets.

	Classifier	INSPIRE			DRIVE			VICA VR			WIDE		
		Se	Sp	B-Acc	Se	Sp	B-Acc	Se	Sp	B-Acc	Se	Sp	B-Acc
Second human grader	-	97.2%	97.0%	97.1%	-	-	-	-	-	-	97.0%	97.9%	97.4%
Niemeijer et al. [1]	kNN	78.0%	78.0%	78.0%	80.0%	80.0%	80.0%	-	-	-	-	-	-
Muramatsu et al. [38]	LDA	-	-	-	-	-	93.0%	-	-	-	-	-	-
Vazquez et al. [3]	k-means	-	-	-	-	-	-	-	-	88.8%	-	-	-
Mirsharif et al. [22]	LDA	-	-	-	-	-	84.1%	-	-	-	-	-	-
Relan et al. [23]	GMM	92.7%	48.5%	70.8%	-	-	-	-	-	-	-	-	-
Dashbozorg et al. [8]	LDA	91.0%	86.0%	88.5%	90.0%	84.0%	87.0%	-	-	89.8%	-	-	-
Lyu et al. [29]	k-means	90.2%	79.4%	84.8%	-	-	-	-	-	-	-	-	-
Pellegrini et al. [39]	graph	-	-	-	-	-	-	-	-	-	-	-	92.5%
Girard et al. [24]	CNN	-	-	-	92.3%	93.1%	92.7%	-	-	-	-	-	-
Huang et al. [25]	LDA	-	-	85.1%	-	-	-	-	-	90.6%	-	-	-
Estrada et al. [28]	graph	91.5% $\pm 0.1\%$	90.2% $\pm 0.1\%$	90.9% $\pm 0.1\%$	93.0% $\pm 0.06\%$	94.1% $\pm 0.07\%$	93.5% $\pm 0.05\%$	-	-	-	91.0% $\pm 0.06\%$	90.9% $\pm 0.06\%$	91.0% $\pm 0.06\%$
Proposed	DOS	96.8% $\pm 0.08\%$	95.7% $\pm 0.08\%$	96.4% $\pm 0.06\%$	94.2% $\pm 0.07\%$	92.7% $\pm 0.07\%$	93.5% $\pm 0.07\%$	95.4% $\pm 0.1\%$	93.8% $\pm 0.09\%$	94.6% $\pm 0.09\%$	96.2% $\pm 0.1\%$	94.2% $\pm 0.06\%$	95.2% $\pm 0.08\%$

VI. DISCUSSIONS AND CONCLUSIONS

In this work, we proposed a DOS-based topology estimation method, and further utilized the obtained topological information to separate arteries from veins in retinal fundus images. In other words, the blood vessel classification problem has been formalized as a pairwise clustering problem. In this section, we carry out a comprehensive comparison between the proposed method and the state-of-the-art label propagation (or topology estimation) and A/V classification methods.

A. Comparison to the state-of-the-art label propagation / topology estimation methods

As suggested by De et al. [31], the topology estimation task can be reformulated as topological label propagation over the directed graph. Therefore, in this section, the p values at blood vessel centerline pixel level obtained by the proposed DOS method is compared with those of five state-of-the-art label propagation and topology estimation methods: Loopy Belief Propagation (LBP) [42], Symmetrized Graph Laplacian (SGL) [43], Zero-mode Free Laplacian (ZFL) [44], Matrix-Forest Theorem of Directed graphs (MFTD) [30], [45], and Heuristic Search Algorithm (HSA) [5]. Of these methods, LBP, SGL, and ZFL are essentially label propagation approaches, for which the source codes have already been developed by the machine learning community; MFTD and HSA are state-of-the-art topology reconstruction methods for tree-like structures: their codes are available from the authors. However, the HSA was designed for the WIDE dataset, and is thus applicable only to this dataset.

In our experiments, the recommended parameters from the original source code or literature were used. These methods all share the same blood vessel extraction and optic disc removal step, as well as the preparation process of converting the skeleton maps into undirected graphs or digraphs, as described in Section III.A. The experimental results are presented in TABLE VI.

Overall, our DOS method consistently outperforms the other label propagation methods [42]–[44] and existing topology estimation methods [5], [30] by a rather large margin. Again,

TABLE VI: The overlapping rate (p score) of different label propagation and topology estimation methods over different datasets.

	INSPIRE	IOSTAR	VICA VR	DRIVE	WIDE
LBP [42]	76.6%	75.9%	74.6%	71.9%	72.3%
ZFL [44]	84.3%	84.3%	83.1%	82.7%	81.7%
SGL [43]	81.0%	79.8%	82.0%	80.9%	78.8%
MFTD [30]	91.1%	90.9%	87.6%	84.9%	87.5%
HSA [5]	-	-	-	-	89.6%
DOS	95.1%	94.2%	93.8%	91.1%	94.9%

all the methods yielded their best results over the INSPIRE dataset due to its simpler branches, and their second best over the WIDE dataset due to the wider field-of-view of this dataset.

Fig. 8 also presents for visual comparison the topology reconstruction results of the competing methods on an example image from the DRIVE dataset. Grey and red colored disks are used here to represent correct and incorrect predictions in various bifurcation and crossover scenarios. It can be seen that the proposed method produces better topology reconstruction results than its competitors, by observing the bifurcations and crossovers where these have been highlighted. This is because MFTD takes into consideration only the angular or intensity properties of neighbouring blood vessel segments forming a junction for the construction of their current weight matrix [30]. Generally speaking, the proposed method achieves the best performance, which suggests the advantages of our DOS approach for vascular topology reconstruction. However, the proposed method also makes errors in some cases, since it may suffer from failures at the segmentation and skeletonization stages that may misrepresent the topological structures of the retinal blood vessels for the classification phase.

B. Comparison to the state-of-the-art A/V classifiers

To facilitate better comparison with the results of our classifier, we also applied the most commonly-used A/V classifiers to the topology-assigned structures derived by our method. The comparative A/V classifiers chosen were LDA, GMM, k-means clustering, and SVM. For the supervised classifiers (LDA and SVM), we randomly assigned half of the images

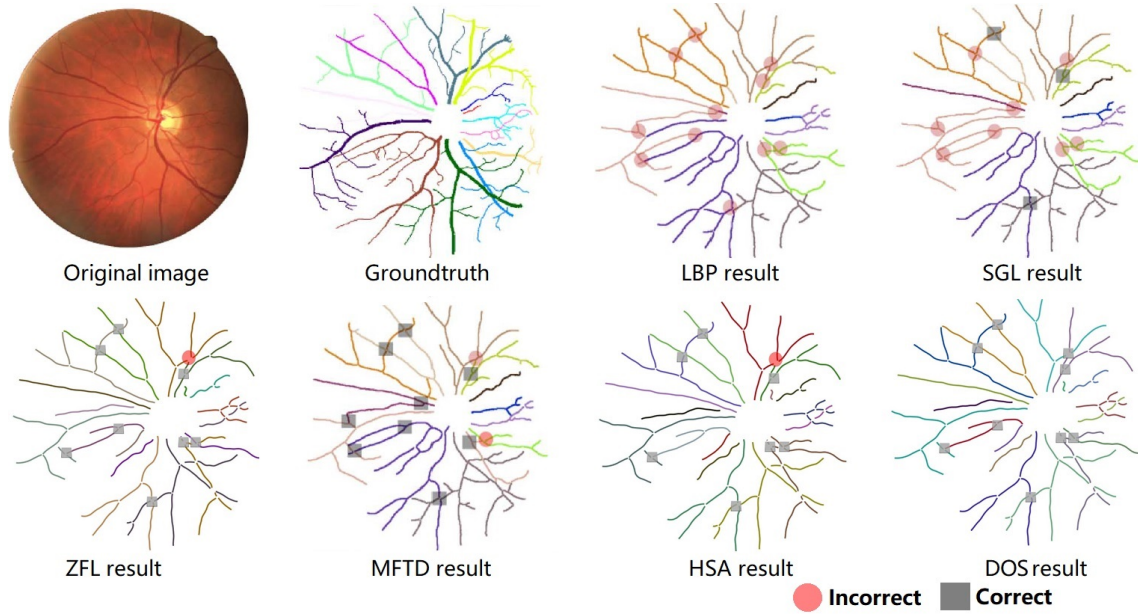


Fig. 8: The results of different methods in determining the labels at bifurcations and crossovers. Note: the results of LBP, SGL, and MFTD were quoted from [30] for convenience.

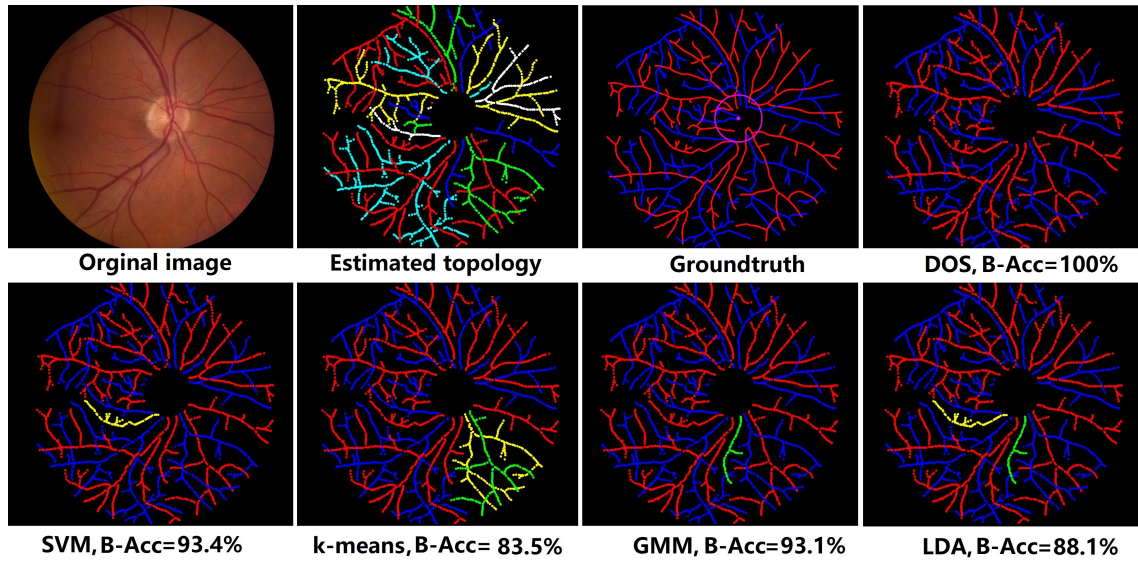


Fig. 9: Blood vessel classification results of a randomly selected image from the INSPIRE dataset by using different A/V classifiers over the topology-assigned blood vessel structures produced by the proposed method.

as a training set, and the remainder of the images as the test set. The same features listed in TABLE I were used to train the LDA and SVM classifiers. TABLE VII demonstrates how well the competing classifiers succeeded in classifying the topology-assigned network into arteries or veins. It can be seen that our DOS method clearly outperformed the compared classification methods, with a balanced accuracy ($B-Acc$) of 96.4%, 93.5%, 94.6%, and 95.2% on the INSPIRE, DRIVE, VICAVER, and WIDE datasets, respectively.

Fig. 9 illustrates the A/V classification performance of different classifiers on our DOS-guided blood vessel topology (Green: incorrectly traced veins; Yellow: incorrectly traced). It shows that the proposed DOS method achieved the highest bal-

anced accuracy in this case, with 100% on a randomly-selected image from the INSPIRE dataset. The k-means clustering classifier produced the worst results with a $B-Acc$ score of 83.5%: this is because one of its major drawbacks is the naive use of the mean value for the representation of each cluster, as a result of which the clusters obtained may not be repeatable and lack consistency. The GMM method obtained relatively higher scores than the other three compared methods, with a $B-Acc$ score of 93.1%: it is more flexible in terms of cluster shape representation than k-means and SVM, and the clusters can be shaped as ellipses, rather than being restricted to hyperspheres in k-means clustering or hyperplanes in SVM.

From a comparison of TABLE IV and TABLE VII, it can be

TABLE VII: The balanced accuracy and standard deviation of different classifiers over the topology-assigned structure derived by the proposed method from different datasets.

	INSPIRE	DRIVE	VICAVR	WIDE
LDA	91.6% \pm 0.10%	89.6% \pm 0.11%	91.1% \pm 0.10%	90.6% \pm 0.12%
GMM	92.1% \pm 0.09%	87.4% \pm 0.06%	91.0% \pm 0.07%	89.3% \pm 0.08%
k-means	88.3% \pm 0.12%	85.8% \pm 0.11%	90.5% \pm 0.10%	90.1% \pm 0.12%
SVM	93.9% \pm 0.07%	88.3% \pm 0.09%	90.6% \pm 0.06%	89.2% \pm 0.08%
DOS	96.4% \pm 0.06%	93.5% \pm 0.07%	94.6% \pm 0.09%	95.2% \pm 0.08%

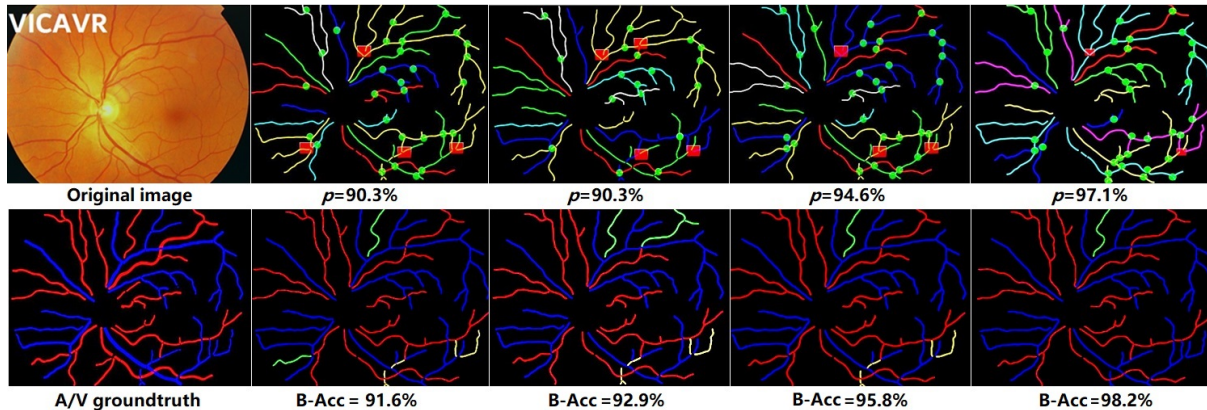


Fig. 10: The topology estimation (top row) and A/V classification (bottom row) results of the proposed method with and without removing different types of features for the representation of the pixels of interest. From second to last column: intensity-based features removed, orientation-based features removed, diameter-based features removed, and all features retained.

TABLE VIII: Topology estimation and A/V classification results ($p/B-Acc$) of the proposed method with and without removing different types of features for the representation of the pixels of interest.

	INSPIRE	DRIVE	VICAVR	WIDE	IOSTAR
All features retained	95.1% / 96.4%	91.1% / 93.5%	93.8% / 94.6%	94.9% / 95.2%	95.1% / -
Diameter-based features removed	94.2% / 94.4%	89.1% / 91.5%	92.0% / 93.3%	93.4% / 94.5%	94.1% / -
Orientation-based features removed	92.9% / 93.2%	86.8% / 88.9%	89.9% / 91.8%	92.5% / 93.7%	93.0% / -
Intensity-based features removed	91.0% / 92.1%	85.1% / 88.5%	88.3% / 90.6%	91.9% / 92.1%	92.2% / -

observed that the method proposed by Dashtbozorg et al. [8] yielded a $B-Acc$ score of 88.5% over the INSPIRE dataset: they used similar features to train the LDA classifier. After the LDA classifier was applied to our DOS generated blood vessel topology, the $B-Acc$ score has been dramatically increased to 91.6%. In a similar manner to the GMM classifier, the $B-Acc$ score has been increased from 70.8% [23] to 92.1%. These results imply that the accurate identification of the underlying network topology is key to improving the accuracy of A/V classification.

C. Feature importance analysis

The feature weights for nodes and blood vessel segments are critical to the accuracy of subsequent topology estimation and A/V classification. In real applications, the feature weights are usually determined by a similarity measure, where only certain features are taken into consideration. Many strategies have been explored for the selection of these critical features, in an effort to identify more compact and better quality feature subsets for various tasks. Such techniques typically involve the use of an individual feature significance evaluation, or a measurement of feature subset consistency, that work together with a search algorithm to determine a quality

subset of features. However, most of the feature selection techniques are supervised, i.e., the labels of instances are required for evaluating or measuring the significance of the feature subsets. For the purposes of this topology estimation and A/V classification work, it is prohibitively time-consuming to have the vascular branches annotated by human experts, and hence benchmarking public data sets are rare in the literature. Therefore, the features used here to determine the weights ($\omega(i, j)$ in Algorithm 3) were empirically selected, as listed in TABLE I.

We reported the performances of topology estimation and A/V classification results of the proposed method in TABLE VIII, in terms of removing the intensity-based, orientation-based, and diameter-based features, respectively from the feature candidates. It can be seen that intensity-based features affect most of the proposed classifier, as the p scores of the topology estimation are significantly decreased by 4.1%, 6.0%, 5.5%, 3.0%, and 2.9% over INSPIRE, DRIVE, VICAVR, WIDE, and IOSTAR, respectively after their removal. Similarly, the $B-Acc$ scores of A/V classification are also decreased by a significant margin: 4.3%, 5.0%, 4.0%, 3.1%, and 2.9% over INSPIRE, DRIVE, VICAVR, and WIDE datasets.

Fig. 10 illustrates the resulting topology estimation and A/V classification labels of a randomly selected image from the VICAVR database (Green: incorrectly traced veins; Yellow: incorrectly traced), after removing the intensity-based, orientation-based, and diameter-based features individually from the feature representation of the pixels of interest. It demonstrates that intensity-based features are the most important: more incorrectly identified connections occurred, and more incorrectly classified arteries and veins took place after their removal.

D. Comparison between automatic and manual segmentations

All the above-reported results are based on ground truth segmentation, with a single exception for the VICAVR dataset. In this subsection, in order to characterize how the errors of automatic and manual segmentations affect their topological reconstructions and final A/V classifications, we have repeated our topology estimation and AV classification methods over the automated vessel segmentation results of the INSPIRE dataset.

To compare the vessel segmentation performance of the automated method with the corresponding ground truth, we computed the *sensitivity* and *false discovery rate* (FDR) [41] between the predicted centerlines and ground truth centerlines. (the INSPIRE dataset provides the centerline of the retinal blood vessels only.) Sensitivity is the fraction of the number of centerline pixels in the correctly detected segments (true positives) over the total number of centerline pixels in the ground truth. FDR is defined as the fraction of the total number of centerline pixels incorrectly detected as vessel segments (false positives) over the total number of centerline pixels of the traced vessels in the ground truth. The use of *specificity*, defined as the number of pixels correctly rejected as non-vessel structures (true negatives), is not adequate for the evaluation of this segmentation task, since the vast majority of the pixels do not belong to blood vessels. It is worth noting that, as is customary in the evaluation of the methods extracting one pixel-wide curves [41], a three-pixel tolerance region around the manually traced nerves is considered to be a true positive. In other words, a predicted centerline point is considered as true positive if there are no more than three pixels from the nearest ground truth centerline point. The experimental results are presented in Figure 11 and Table IX.

Figure 11 shows that the proposed method has produced perfect topology estimation and A/V classification results without any error on manual vessel annotations of an image randomly selected from the INSPIRE dataset. While only one node was incorrectly identified, indicated by the red rectangle in Fig. 11(c), and one vessel segment was incorrectly classified on the automatic segmented vessels, indicated by the yellow line in Fig. 11(e), since a tiny vessel was mis-detected and leads to the generation of a partly incorrect vessel graph that causes further errors in the vessel topology reconstructions and A/V classifications.

In addition, the percentage of nodes that were correctly identified (p score) and accuracy of the topology reconstruction at blood vessel centerline pixel-level were calculated.

As can be observed in TABLE IX, the proposed topology estimation method achieved a similar p score on automatic and manual segmentations: 93.5% and 95.1%, respectively. However, it yielded significantly low scores of accuracy and B - Acc , by 8.6% and 8.7%, respectively. As aforementioned, the accuracy and B - Acc were calculated at centerline-level. The automated segmentation performance achieved only 83.8% in sensitivity, in comparison with that on manual segmentation. This is because a large portion of tiny vessels were mis-segmented.

However, the obtained p score implies that the proposed method is able to correctly identify most of the nodes in automatically detected vessels. These results show that our topology estimation and its subsequent AV classification, are relatively robust in terms of applying to either automatic or manual vessel segmentation results at node-level, while the overall performance (pixel-level) is still dependent on the completeness of the vessel segmentation.

E. Conclusions

Automated identification of the anatomical connectivity of different blood vessels, and classification of those blood vessels into arteries and veins, respectively, are essential for the automated assessment of vascular changes.

In this paper, we have proposed a novel artery/vein classification method based on vascular topological characteristics. Our framework combines graph-theoretic methods with DOS to accurately analyse the retinal vasculature. The concept of DOS in this work was successfully adapted to formalize the topology estimation and A/V classification as a pairwise clustering problem. The core issues then go to work on graph generation and edge weight definition. They have been achieved through image segmentation, skeletonization and identification of significant nodes. The latter is defined as the inverse Euclidean distance between the two end points of an edge in the feature space, where each node is represented as a 23 dimensional feature vector about intensity, orientation, curvature, diameter and entropy. The significance of our method is that it is capable of classifying the whole vascular network, and does not restrict itself to specific regions of interest.

The proposed method has accurately reconstructed the vascular topology and classified the blood vessel types into arteries and veins on five publicly accessible retinal datasets. The results show that our method produces better performance when compared with the state-of-the-art topology estimation and A/V classification ones. It can be expected that the proposed method could be a powerful tool for analyzing vasculature for better management of a wide spectrum of vascular-related diseases.

In addition, we have manually labeled the blood vessel topology for four publicly-available retinal datasets (INSPIRE, VICAVR, DRIVE, and IOSTAR), and these annotations have been released for public access to help other researchers in the community in performing research and development on the same and related topics. For future work, we will test our method on other retinal datasets (e.g., RITE [46]) and neuronal datasets (e.g., DIADEM [47]), and refine the initial graph

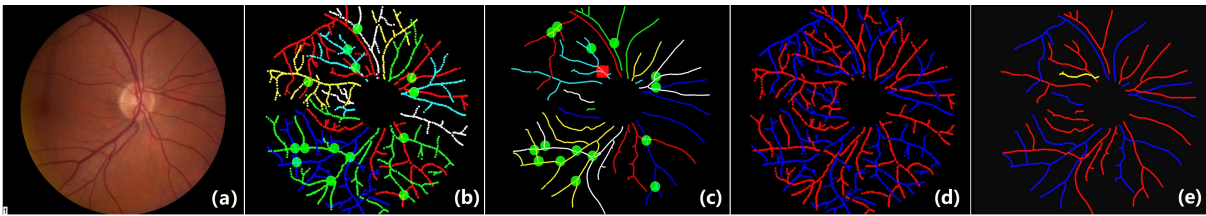


Fig. 11: Illustrative results of the proposed method for topology estimation and A/V classification applied on automatic and manual segmentations. (a) A randomly selected image from INSPIRE; (b)-(c) the topology estimation results applied to manual and automatic segmentations; (d)-(e) A/V classification results on manual and automatic segmentations.

TABLE IX: The performance of the proposed method for vessel topology estimation and A/V classification to automatic and manual segmentations at the node-level and vessel centerline pixel-level over the INSPIRE dataset.

	vessel segmentation		topology estimation				A/V classification
	sensitivity	FDR	# nodes	# TP	p	accuracy	B-Acc
manual segmentation	-	-	2776	2642	95.1%	97.5%	96.4%
automatic segmentation	83.8%	0.071	2480	2321	93.5%	88.9%	87.7%

generation to improve the accuracy of the retinal topology reconstruction.

REFERENCES

- [1] M. Niemeijer, X. Xu, A. Dumitrescu, B. van Ginneken, J. Folk, and M. Abramoff, "Automated measurement of the arteriolar-to-venular width ratio in digital color fundus photographs," *IEEE Trans. Med. Imaging*, vol. 30, no. 11, pp. 1941–1950, 2011.
- [2] J. Xie, Y. Zhao, Y. Zheng, P. Su, J. Liu, and Y. Wang, "Retinal vascular topology estimation via dominant sets clustering," in *Proc. ISBI*, 2018, pp. 1458–1462.
- [3] S. Vázquez, B. Cancela, N. Barreira, and M. Saez, "Improving retinal artery and vein classification by means of a minimal path approach," *Mach. Vis. Appl.*, vol. 24, no. 5, pp. 919–930, 2013.
- [4] T. Na, J. Xie, Y. Zhao, Y. Liu, Y. Wang, and J. Liu, "Retinal vascular segmentation using superpixel-based line operator and its application to vascular topology estimation," *Medical physics*, vol. 45, no. 7, pp. 3132–3146, 2018.
- [5] R. Estrada, C. Tomasi, S. Schmidler, and S. Farsiu, "Tree topology estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1688–1701, 2015.
- [6] Y. Zhao, J. Zhao, J. Yang, Y. Liu, Y. Zhao, Y. Zheng, L. Xia, and Y. Wang, "Saliency driven vasculature segmentation with infinite perimeter active contour model," *Neurocomputing*, vol. 259, pp. 201–209, 2017.
- [7] Y. Zhao, Y. Zheng, Y. Liu, Na, and Y. W. J. Liu, "Automatic 2d/3d vessel enhancement in multiple modality images using a weighted symmetry filter," *IEEE Trans. Med. Imaging*, pp. 1–1, 2017.
- [8] B. Dashtbozorg, A. M. Mendonça, and A. Campilho, "An automatic graph-based approach for artery/vein classification in retinal images," *IEEE Trans. Image Processing*, vol. 23, no. 3, pp. 1073–1083, 2014.
- [9] V. Joshi, J. Reinhardt, M. Garvin, and M. Abramoff, "Automated method for identification and artery-venous classification of vessel trees in retinal vessel networks," *PLoS ONE*, vol. 9, no. 2, pp. 1–12, 2014.
- [10] 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 *Proc. MICCAI*, 2018, pp. 109–118.
- [11] M. Pavan and M. Pelillo, "Dominant sets and hierarchical clustering," in *Proc. ICCV*, 2003, pp. 362–369.
- [12] E. Zemene and M. Pelillo, "Interactive image segmentation using constrained dominant sets," in *Proc. ECCV*, 2016, pp. 278–294.
- [13] E. Zemene, L. Alemu, and M. Pelillo, "Dominant sets for "constrained" image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2018.
- [14] Y. Zhao, Y. Zheng, Y. Liu, J. Yang, Y. Zhao, D. Chen, and Y. Wang, "Intensity and compactness enabled saliency estimation for leakage detection in diabetic and malarial retinopathy," *IEEE Trans. Med. Imaging*, vol. 36, no. 1, pp. 51–63, 2017.
- [15] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable COS-FIRE filters for vessel delineation with application to retinal images," *Med. Image Anal.*, vol. 19, pp. 46–57, 2015.
- [16] Y. Zhao, Y. Liu, X. Wu, S. Harding, and Y. Zheng, "Retinal vessel segmentation: An efficient graph cut approach with retinex and local phase," *PLoS ONE*, vol. 10, p. e0122332, 2015.
- [17] M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "Blood vessel segmentation methodologies in retinal images - a survey," *Comput. Meth. Prog. Bio.*, vol. 108, pp. 407–433, 2012.
- [18] J. Soares and M. Cree, "Retinal vessel segmentation using the 2D Gabor wavelet and supervised classification," *IEEE Trans. Med. Imaging*, vol. 25, pp. 1214–1222, 2006.
- [19] M. Martínez-Pérez, A. Hughes, A. Stanton, S. Thom, N. Chapman, A. Bharath, and K. Parker, "Retinal vascular tree morphology: a semi-automatic quantification," *IEEE Trans. Biomed. Engineering*, vol. 49, no. 8, pp. 912–917, 2002.
- [20] C. Kondermann, D. Kondermann, and M. Yan, "Blood vessel classification into arteries and veins in retinal images," in *Proc. SPIE Medical Imaging*, 2007, p. 651247.
- [21] M. Niemeijer, B. Ginneken, and M. Abramoff, "Automatic classification of retinal vessels into arteries and veins," in *Proc. SPIE Medical Imaging*, vol. 7260, 2009, pp. 72 601F–1–72 601F–8.
- [22] Q. Mirsharif, F. Tajeripour, and H. Pourreza, "Automated characterization of blood vessels as arteries and veins in retinal images," *Comp. Med. Imag. and Graph.*, vol. 37, no. 7–8, pp. 607–617, 2013.
- [23] D. Relan, T. MacGillivray, L. Ballerini, and E. Trucco, "Retinal vessel classification: Sorting arteries and veins," in *Proc. EMBC*, 2013, pp. 7396–7399.
- [24] F. Girard and F. Cheriet, "Artery/vein classification in fundus images using CNN and likelihood score propagation," in *Proc. IEEE Global Conf. on Sig. and Info. Process.*, 2017, pp. 720–724.
- [25] F. Huang, B. Dashtbozorg, and B. H. Romeny, "Artery/vein classification using reflection features in retina fundus images," *Mach. Vis. Appl.*, vol. 29, no. 1, pp. 23–34, 2018.
- [26] K. Rothaus, X. Jiang, and P. Rhiem, "Separation of the retinal vascular graph in arteries and veins based upon structural knowledge," *Image Vision Comput.*, vol. 27, no. 7, pp. 864–875, 2009.
- [27] V. Joshi, M. Garvin, J. Reinhardt, and M. Abramoff, "Automated method for the identification and analysis of vascular tree structures in retinal vessel network," in *Proc. SPIE Medical Imaging*, 2011, p. 796301.
- [28] R. Estrada, M. J. Allingham, P. S. Mettu, S. W. Cousins, C. Tomasi, and S. Farsiu, "Retinal artery-vein classification via topology estimation," *IEEE Trans. Med. Imaging*, vol. 34, no. 12, pp. 2518–2534, 2015.
- [29] X. Lyu, Q. Yang, S. Xia, and S. Zhang, "Construction of retinal vascular trees via curvature orientation prior," in *Proc. BIMB*, 2016, pp. 375–382.
- [30] J. De, L. Cheng, X. Zhang, F. Lin, H. Li, K. Ong, W. Yu, Y. Yu, and S. Ahmed, "A graph-theoretical approach for tracing filamentary structures in neuronal and retinal images," *IEEE Trans. Med. Imaging*, vol. 35, no. 1, pp. 257–72, 2016.
- [31] J. De, X. Zhang, F. Lin, and L. Cheng, "Transduction on directed graphs via absorbing random walks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 7, pp. 1770–1784, 2018.
- [32] J. Cheng, J. Liu, Y. Xu, F. Yin, D. Wong, N. Tan, D. Tao, C. Cheng,

- T. Aung, and T. Wong, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE Trans. Med. Imaging*, vol. 32, no. 6, pp. 1019–1032, 2013.
- [33] Y. Zhao, L. Rada, K. Chen, S. Harding, and Y. Zheng, "Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images," *IEEE Trans. Med. Imaging*, vol. 34, no. 9, pp. 1797–1807, 2015.
- [34] P. Bankhead, J. McGeown, and T. Curtis, "Fast retinal vessel detection and measurement using wavelets and edge location refinement," *PLoS ONE*, vol. 7, p. e32435, 2009.
- [35] M. Pavan and M. Pelillo, "Dominant sets and pairwise clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 167–172, 2007.
- [36] J. Zhang, B. Dashtbozorg, E. J. Bekkers, J. P. W. Pluim, R. Duits, and B. M. ter Haar Romeny, "Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores," *IEEE Trans. Med. Imaging*, vol. 35, pp. 2631–2644, 2016.
- [37] J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 501–509, 2004.
- [38] C. Muramatsu, Y. Hatanaka, T. Iwase, T. Hara, and H. Fujita, "Automated selection of major arteries and veins for measurement of arteriolar-to-venular diameter ratio on retinal fundus images," *Comput. Med. Imag. Grap.*, vol. 35 6, pp. 472–80, 2011.
- [39] E. Pellegrini, G. Robertson, T. J. MacGillivray, J. van Hemert, G. Houston, and E. Trucco, "A graph cut approach to artery/vein classification in ultra-widefield scanning laser ophthalmoscopy," *IEEE Trans. Med. Imaging*, vol. 37, pp. 516–526, 2018.
- [40] T. Qureshi, M. Habib, A. Hunter, and B. Al-Diri, "A manually-labeled, artery/vein classified benchmark for the drive dataset," in *Proc. CBMS*, 2013, pp. 485–488.
- [41] P. Guimarães, J. Wigdahl, and A. Ruggeri, "A fast and efficient technique for the automatic tracing of corneal nerves in confocal microscopy," *Transl. Vis. Sci. Techn.*, vol. 5, p. Article 7, 2016.
- [42] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proc. ICUAL*, 1999, pp. 467–475.
- [43] D. Zhou and B. Schölkopf, "Learning from labeled and unlabeled data on a directed graph," in *Proc. ICML*, 2005, pp. 1036–1043.
- [44] H. Cai, X. Wang, H. Huang, and C. Ding, "Simultaneous image classification and annotation via biased random walk on tri-relational graph," in *Proc. ECCV*, 2012, pp. 823–836.
- [45] L. Cheng, J. De, X. Zhang, F. Lin, and H. Li, "Tracing retinal blood vessels by matrix-forest theorem of directed graphs," in *Proc. MICCAI*, 2014, pp. 626–633.
- [46] Q. Hu, M. D. Abramoff, and M. K. Garvin, "Automated construction of arterial and venous trees in retinal images," *Journal of Med. Imaging*, vol. 2, p. 044001, 2015.
- [47] T. A. Gillette, K. M. Brown, and G. A. Ascoli, "The DIADEM metric: Comparing multiple reconstructions of the same neuron," *Neuroinfo.*, vol. 9, no. 2-3, pp. 233–245, 2011.

Associate Editor's Comments:

Q1: All reviewers see scientific merit in this work. In addition, it is seen very positive that the annotation of several databases is made publicly available. On the other hand, they also pointed out a number of technical issues to be addressed. In particular, Reviewer B raised the issue of inconsistency (use of ground truth segmentations for performance evaluation). As remarked by that reviewer, it is indeed not straightforward to compare AV classifications based on two different segmentations (automatic vs. manual segmentations). But currently, it remains not clear if the proposed approach is robust to errors in the automated segmentation. It is thus essential to “present some alternative way to characterize how errors in the initial segmentation affect their topological reconstructions and final classification” (Reviewer B).

A: Many thanks for your suggestions and comments. We have done the experiments on both automatic and manual segmentations, and the comparison of the proposed method for topology estimation and A/V classification based on automatic and manual segmentation results at the node-level and vessel centerline pixel-level as demonstrated at TABLE IX and Fig.11. The experimental settings and discussions are provided in Section VI.D. The results show that our method is relatively robust to the segmentation.

Q2: One minor question from my side: “Experts one and two independently labeled each vessel segment or centerline for all the datasets, based on the types of available manual annotations of the vessel structure. The consensus between them was then used as ground truth”. How is this done in the case of two diverging labelings only? By discussion between the two labeling experts?

A: Yes, a discussion is required between two experts to make a consensus decision.

Reviewer: 1

General comments

The authors propose a dominant sets clustering method for graph generation of retinal vasculature, followed by artery/vein classification. The paper is well organized, well written, showing interesting results. In general, the paper is good but there are several issues that need to be addressed. My comments below are organized by section.

Concerns to be addressed:

I - INTRODUCTION

Q1: being the vessels biased by the authors, a question arises if there is any bias; is he/she a ophthalmology or image analysis expert?

A: The ground truth of topological reconstruction was manually annotated by an image analysis expert; and all the annotations made were then checked and corrected by an ophthalmologist to reduce bias. Such details have been added into the revised paper in Sections IV.A and IV.B.

III - METHOD

II. B - Graph generation:

The description in II.B, follows closely [7, 31], which is OK as the references are given, and may help the reader in understanding the clustering method. However, it has some issues and details that are not complete, which makes the understanding of this more difficult, for instance:

Q2: When stating that $w: E \rightarrow R$ is the positive weight function, the application in to R is not correct; it must be to R^+ .

A: Thanks, and it has been corrected. (page 3)

Q3: You need to define the values of a_{ij} for $i=j$. For graphs without self-loops is it $a_{ij} = 0$?

A: Thanks for pointing it out. We amended the definition on Page 4.

Q4: Please change metric to measure, as in similarity metric. A metric is a distance function, that follows a set of properties, that are not shared by a similarity measure (in the case of this paper is the inverse of a distance).

A: Many thanks for this constructive suggestion. We have replaced the 'metric' with 'measure'. (pages 3-4)

Q5: The numeric example in Fig 2., is equal (with different values) with an example in [31]. Better would to use a more realistic and illustrative examples from the graphs extracted from the retinal blood vessels.

A: Many thanks for the suggestion. We have replaced the numeric example by a more illustrative diagram which shows the extraction of a graph at an intersection. (See Figs. 2 and 3 in page 4)

Q6: This illustrative example could be further used for illustration of a couple of iterations of the algorithms used for Topology estimation and A/V identification.

A: Many thanks for the suggestion. We have illustrated the application of replicator dynamics on the graph shown in Fig 3. (pages 4 and 5)

Q7: As for the similarity measure, the authors use the inverse of a distance. What happens if the distance is equal to zero?

A: To avoid the special case where the denominator is zero, a small number $\epsilon=0.000001$ was added. (page 5)

TABLE I:

Q8: This tables defines qualitatively the features, but the reader would appreciate the formal definition of all of them, eg for orientation, curvature, and diameters.

A: All the features were previously used for topological reconstruction [8], [9] and vessel classification [1], [20] tasks. While other features can be calculated straightforward, the main ones, orientation, diameter and curvature, are defined and explained in Section C. (page 5)

II-C - Arteries/Veins Classification via Dominant Sets Clustering

Q9: In this statement "arteries are brighter in color than veins and are thinner than neighboring veins" the authors need to clarify what neighborhood is considered and how is computed; furthermore, the intensity difference is valid only locally and is very difficult to establish for thinner vessels; this needs to be addressed in depth, and how is this applied.

A: After all the vessel branches have been classified into two groups, we need to assign them into artery and vein. We may compute the average diameters or intensity values of vessel segments located at the ring range, see Fig. 4 (the radius of the ring is the same as the radius of the optic

disc), and we defined this ring range as 'neighborhood', or 'Region of Interest'. In practice, either the diameters or intensity values may be used to assign the artery or vein. In this work, we define a vessel segment as 'artery' if the average intensity value of the pixels inside the ring centred at any pixel on its centerline with the same radius as the optic disc is larger than 0.48 (with the intensity values lying in the unit interval [0, 1]). Otherwise the vessel segment is assigned as 'vein'. The threshold value was chosen empirically.

IV - MATERIALS

Q10: The results in Fig 3 needs a qualitative analysis: there are several particularities on the results they need an explanation. To illustrate this please look at the results for INSPIRE: in the ground truth (GT), the top yellow vessel, has a final segment that is not present but there is a segmentation result; on the other hand, for the same vessel; a final bottom segments present in the GT are not in the result; there are segments without any correct/incorrect labels.

A: Many thanks for your suggestion. We have replaced a new figure with correct segments and labels. (See Fig. 5 in page 9)

B – A/V Classification

Q11: Please clarify what is a sub-graph.

A: To avoid confusion, we replaced it with 'sub-tree'. A sub-tree represents an individual vessel branch, i.e., a vessel was labeled with a distinct color (or individual label). This explanation has been added at the very beginning of Section V.B. (page 8)

Q12: The estimates on the evaluation measures need to be accompanied by the corresponding standard deviations.

A: We have added the standard deviations in TABLE IV and VII. Note, only the method proposed by Estrada et al. [28] reported their performances with standard deviations, and those of other methods were not available. (pages 11 and 13)

Q13: Indications on the computation time per image will help, particularly for obtaining the topology.

A: Thanks for the suggestion. We have provided the analyses of computational complexity and computation times for topology estimation and A/V classification of our method in Section V.C. (page 10)

VI. DISCUSSIONS AND CONCLUSIONS

Q14: Organize tables V and VI in the same manner, with datasets in the columns (or rows) and the methods to compare in row (or columns). Also in both, use bold for the best.

A: We have re-formatted Table VII in the same manner as Table VI, and the best scores are highlighted in bold. (pages 11 and 13)

Q15: The readers would appreciate to have further discussions on: the relevance of the used features, (part of this discussion is made for VICAVR, but it may also for all the other datasets; the results can be shown in a simple table); the reasons for the more frequent errors and hints how to improve.

A: We have added the TABLE VIII to further demonstrate the performances of feature importance analysis, and comprehensive discussions are provided in the last two paragraphs of Section VI.C (pages 12-13)

Q16: Table VI uses KNN and in the text we have K-means and KNN; please correct this.

A: Corrected. (pages 11-12)

REFERENCES

Q17: Update reference [6] as MICCAI proceedings are already available. One author is missing in reference [28]

A: Corrected. (updated as references [10] and [33])

Minor comments:

Index terms:

Q18: The term “vessels” does not seem adequate in the context of the paper. My suggestion: vessel --> blood vessel.

A: Corrected accordingly.

Notation:

Q19: The notation for representing vectors and scalars are the same, which creates confusion, eg for representing the feature vector f_i and a feature component of this vector.

A: Corrected. (page 5)

Q20: The meaning of the symbols $| \cdot |$ and $|| \cdot ||$ must be defined.

A: They have been defined accordingly in the revised version. $| \cdot |$ indicates the cardinality of a set. $|| \cdot ||$ indicates the Euclidean distance. (pages 4 and 5)

Q21: Pg 1; 2nd column: Topological and geometrical properties: Why geometrical; what type of geometrical properties; why colour and morphology is not mentioned here.

A: The sentence has been rephrased as: “This calls for precise description of the vascular structure in terms of its color, topological, geometrical and morphological properties as derived from retinal images” in Section I.

Q22: Fig: 1: Fig 1 needs a brief description. Furthermore, and to be precise, it not describes an overview of the method but the input/output of different functions that need to be presented and briefly described.

A: More informative description is provided at the beginning of Section.III (page 3)

Q23: Typos/formatting: Pg 2: 2nd column: veinsk-means --> veins k-means; Pg 2: line before in III. METHOD; Pg 9: line before in B. Comparison to the state-of-the-art A/V classifiers

A: Corrected.

Reviewer: 2

Comments to the Author

The authors present a graph-based approach to automatically segment and classify retinal vessels into either arteries or veins from a single fundus image. Overall, their proposed approach is promising, but they need to address the following issues in the manuscript.

Major comments:

Q1: Although the authors present their approach as a fully automated pipeline---from image to AV classification---as far as I can tell, all their results are based on ground-truth segmentations (except for the VICAVR dataset). While I understand that it is not straightforward to compare AV classifications based on two different segmentations (i.e., automatic vs. manual segmentations), at this point it is not clear if their approach is robust to errors in the automated segmentation. Ideally, the authors should repeat their topology and AV classification results for one of the datasets, probably DRIVE or INSPIRE, using their automatic segmentations. I realize that manually annotating these segmentations is time-consuming, so it might not be feasible in the available time-frame. In lieu of that, the authors should present some alternative way to characterize how errors in the initial segmentation affect their topological reconstructions and final classification.

A: Many thanks for these constructive suggestions. We have done the experiments on both automatic and manual segmentations, and the comparison of the proposed method for topology estimation and A/V classification based on automatic and manual segmentation results at the node-level and vessel centerline pixel-level as demonstrated at TABLE IX and Fig.11. The experimental settings and discussions are provided in Section VI.D. The results show that our method is relatively robust to the segmentation. In conclusion, our topology estimation and its subsequent A/V classification are relatively robust in terms of applying on either automatic or manual vessel segmentation results at node-level, while the overall performance (pixel-level) is still dependent on the completeness of the vessel segmentation. For more details, please refer to Section VI.D. (pages 13-14)

Q2: The authors state that there are no manual AV annotations for either the DRIVE or INSPIRE datasets, but the authors of Estrada et al. 2015-b (citation [24] in the manuscript) made their graph-based annotations for these two datasets publically available. Granted, their website, http://people.duke.edu/~sf59/Estrada_TMI_2015_dataset.htm, is not very clear about which annotations for which datasets are available, but it would be useful to know how similar these different graph-based AV annotations are, e.g., using a rater agreement metric. Also, the original authors of the DRIVE dataset recently issued an AV-labeled version of their pixel-level vessel segmentations, so it would be interesting to compare the agreement to those labels, as well.

A: In fact, we used the manual annotation provided by Estrada as the ground truth for evaluation. However, we also noticed that the original authors of the DRIVE dataset also released their AV-labels. Therefore, in the revision, we refer these two annotations as GT1 and GT2, respectively, and provided the AV classification results between GT1 vs GT2, auto vs GT1, auto vs GT2, so as to further compare the agreement to those labels. For more details, please refer to the last paragraph of Section V.B, Fig 7, and TABLE V. (pages 10-11)

Q3: The technical description of Dominant Set Clustering could be made clearer. For example, the authors switch the indices i and j from Equations 1 and 2, which makes it harder to follow. In this case, i is not in S in the first equation, but it is in S in the second equation. The example is Fig.2 is also not very clear. Perhaps they could show the recursive subsets in a visual manner, to guide the reader as to what is being computed and how.

A: Thanks for the suggestion. In the revised version, we have reformed Eqn. (1) as p_i is in S , and p_j is in $S/\{p_i\}$. In so doing, Equations (1) and (2) are consistent in the domains of p_i , p_j , and S . Furthermore, we also provided illustrative calculations based on Fig. 2 as successive examples, which may help the reader to better understand the recursive definition of dominant set. (page 4)

Q4: The list of features in Table 1 is intuitive, but still needs justification. What process did the authors use to pick those 23 features? Was it empirical or based on domain knowledge? In either case, did they use some form of feature selection to trim this list down from a larger set of candidates?

A: All the features were previously used for topological reconstruction [8], [9] and vessel classification [1], [20] tasks. We employed the sequential forward floating selection for feature selection, which starts with an empty feature set and adds or removes features when this improves the performance of the classifier. (page 5)

Q5: Fig. 7 does show that trimming this list further affects the results, but perhaps adding some features (e.g., distance from optic nerve) might improve the results. On a similar note, Fig. 7 is only illustrative, but it would be useful to see this analysis over an entire dataset.

A: Thanks for your suggestion. Fig.7 now is updated as Fig. 10. We have added the TABLE VIII to further demonstrate the performances of feature importance analysis, and a comprehensive discussion is provided at the last two paragraphs of Section VI.C (pages 12-13)

Minor comments:

Q6: The reference to Fig. 1 at the end of the first paragraph of Section IV.B is incorrect.

A: Corrected. (page 6)

Q7: There is no space between heading III, Method, and the previous paragraph.

A: Corrected. (page 3)

Q8: The authors should an overview of their pipeline to the beginning of Section III. They have a summary in Fig. 1, but it would be better to also have this run-down in the text, itself.

A: A brief description of the outline of the proposed framework has been added at the beginning of Section III. (page 3)

Q9: The authors should name and provide a brief description of the vessel segmentation, thinning, optic nerve detection methods, etc. that they use in their pipeline. Right now, they only list the reference, which is very inconvenient for the reader.

A: The name of corresponding approaches and their description were given in the revised version. (page 3)

Reviewer: 3

Comments to the Author

Thanks for doing such a great job on this paper. It was a pleasure to read, and is a strong contribution to the field, particularly the effort that has been spent on additional labelling of these widely used datasets.

Thanks a lot!

Q1: I have spotted a small number of typos. These are detailed below:

Abstract: "feature space about intensity, orientation,... "-> "feature space of intensity, orientation ..."

Page 1, Right Hand Column (RHC), first paragraph: remove comma after "benefit"

Page 2: RHC, 3rd para, 2nd line: some missing words or punctuation after "and veins".

Page 3: LHC, section B, 3rd line: "each intersection to different" -> ... ""each intersection into different"

Page 4: Eqn 2: two small issues with Eqn 2, and Eqn 1.

A: All corrected.

Q2: For Eqn 2, The summation is over nodes p_j , whereas the expansion summation is actually over the node *indices* j . I would suggest modifying the summation range term $p_j \in S\{p_i\}$ to read $j : p_j \in S\{p_i\}$

A: Thanks for pointing it out. While S is a set of nodes, the summation over the nodes does make sense. Essentially, the summation is implemented over all the relevant nodes. For coding, it may be implemented over the subscripts of the relevant nodes instead. (page 4)

Q3: Now, W_s depends on the definition given in Eqn 1; I am not sure about this issue, so I have addressed it second: I am *sure* about this point, but what does not appear to be explicit is whether S should exclude the weighting a_{ji} in the summation over k .

A: Thanks for the suggestion. In the revised version, we have reformed Eqn. (1) as p_i is in S , and p_j is in $S/\{p_i\}$. Therefore, a_{ji} (or a_{ij}) is excluded in the summation over k . (page 4)

Q4: Page 2/3), since you have clear algorithms (Algorithms 1 and Algorithms 2) it seems odd that there is not an algorithm provided for RDDOS (it is in the form of Eqns 1 and 2, but given the ambiguity in the expressions, why not include it as an algorithm as well?)

A: Thanks for pointing it out. We have added the algorithm RDDOS in Algorithm 1. (page 5)

Q5: Page 8: Figure 4 should be enhanced a bit, particularly the 3rd column. 4th column is fine!

A: Updated as Figure 5. (page 9)

Otherwise, I am very happy with this paper, and I think it represents useful and establishes some new state-of-the-art approaches to vasculature labelling and, therefore, analysis.

29 April 2019

IEEE-TMI
Editorial office

Dear Editor

On behalf myself and my co-authors, I am submitting our revised manuscript entitled “*Retinal Vascular Network Reconstruction and Classification via Dominant Sets Clustering*” for consideration of publication as a research article.

The previous version (ID: TMI-2018-1136) was handled by AE: Xiaoyi Jiang, and a major revision was suggested.

We have addressed all comments one by one either by improving the manuscript following the suggestions or replying to those questions. We have also made a number of minor changes to further improve the quality of the paper. We believe our paper will benefit readers from a wide range of research interests.

Your faithfully,

Dr. Yitian Zhao

Cixi Institute of Biomedical Engineering,
Ningbo Institute of Industrial Technology,
Chinese Academy of Sciences,
Ningbo, China