Multi-Task Siamese Network for Retinal Artery/Vein Separation via Deep Convolution Along Vessel

Zhiwei Wang, Xixi Jiang, Jingen Liu, *Member, IEEE*, Kwang-Ting Cheng, *Fellow, IEEE*, and Xin Yang, *Member, IEEE*

***Abstract*— Vascular tree disentanglement and vessel type classification are two crucial steps of the graph-based method for retinal artery-vein (A/V) separation. Existing approaches treat them as two independent tasks and mostly rely on ad hoc rules (e.g. change of vessel directions) and hand-crafted features (e.g. color, thickness) to handle them respectively. However, we argue that the two tasks are highly correlated and should be handled jointly since knowing the A/V type can unravel those highly entangled vascular trees, which in turn helps to infer the types of connected vessels that are hard to classify based on only appearance. Therefore, designing features and models isolatedly for the two tasks often leads to a suboptimal solution of A/V separation. In view of this, this paper proposes a multi-task siamese network which aims to learn the two tasks jointly and thus yields more robust deep features for accurate A/V separation. Specifically, we first introduce Convolution Along Vessel (CAV) to extract the visual features by convolving a fundus image along vessel segments, and the geometric features by tracking the directions of blood flow in vessels. The siamese network is then trained to learn multiple tasks: i) classifying A/V types of vessel segments using visual features only, and ii) estimating the similarity of every two connected segments by comparing their visual and geometric features in order to disentangle the vasculature into individual vessel trees. Finally, the results of two tasks mutually correct each other to accomplish final A/V separation. Experimental results demonstrate that our method can achieve accuracy values of 94.7%, 96.9%, and 94.5% on three major databases (DRIVE, INSPIRE, WIDE) respectively, which outperforms recent state-of-the-arts.**

***Index Terms*— Siamese network, artery-vein separation, deep learning, multi-task learning.**

Manuscript received February 26, 2020; accepted March 6, 2020. Date of publication March 11, 2020; date of current version August 31, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61872417, in part by the JD AI Research (the Grapevine Scholar Plan), and in part by the Fundamental Research Funds for the Central Universities under Grant 22019kfyRCPY118 and Grant 2020kfyXGYJ026. *(Corresponding author: Xin Yang.)*

Zhiwei Wang, Xixi Jiang, and Xin Yang are with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: zhiweiwang@hust.edu.cn; xixijiang@hust.edu.cn; xinyang2014@hust.edu.cn).

Jingen Liu is with JD AI Research, Mountain View, CA 94039 USA (e-mail: jingen.liu@jd.com).

Kwang-Ting Cheng is with the School of Engineering, The Hong Kong University of Science and Technology, Hong Kong (e-mail: timcheng@ust.hk).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2020.2980117

I. Introduction

AUTOMATED identification of arteries and veins in digital fundus images is one of the fundamental techniques for computer-aided diagnosis of systemic diseases including diabetes, hypertension, arteriosclerosis, vascular disorders, etc. [1]–[4]. These diseases are known to asymmetrically affect arteriolar and venular geometric structures (e.g. tortuosity, crossing/branching angle, etc.) [5]. In particular, changes in arteriolar-to-venular ratio (AVR), which measures the ratio in retinal arteriolar vs. venular diameter, are considered to be correlated with a wide variety of cardiovascular diseases [6]. Many efforts [7]–[18] have been made to deal with the task of separating the retinal vasculature into arteries and veins after vessels being segmented. Given a fundus image and its corresponding binary map of vessel segmentation, existing solutions of A/V separation can be summed up into two categories, graph-based and dense methods [9].

The graph-based method typically begins with building a graph to represent vasculature based on a skeletonized segmentation map, where graph nodes indicate vessel conjunctions and graph links indicate vessel segments. The vasculature is then disentangled into multiple trees by graph analyzing. Finally, a binary label is assigned to each tree via performing vessel type classification. Zhao *et al.* [14] empirically defined 23 handcrafted features for estimating the similarity of every two connected vessel segments, and then proposed a dominant-sets clustering approach, which disconnects links with low similarity and re-connects ones with high homogeneity, to disentangle the graph into multiple vessel trees. The disentangled vessel trees are further assigned A/V types according to their green-channel intensities. In [19] the authors further examined the validity of their proposed model on extra datasets, and achieved the state-of-the-art performance of A/V separation. However, the dominant-sets clustering approach mainly focuses on accurately disentangling vascular trees, while vessel type classification is relatively neglected. Dashtbozorg *et al.* [17] developed four rule-based algorithms for specific graph nodes with degrees 2 to 5 to perform accurate vascular tree disentanglement. An LDA classifier is then utilized to predict A/V type of each vessel pixel based on 30 handcrafted features. The outcomes of tree disentanglement and type classification mutually rectify each other to obtain the result of A/V separation. Srinidhi *et al.* [18] adopted a similar

framework as [17] but employed more sophisticated rules and features. They proposed a vessel keypoint descriptor (VKD) to identify vessel keypoints including bifurcation, crossover and end nodes in a vascular graph, and utilized a depth-first search (DFS) approach to split the graph into individual trees according to several VKD-derived rules. A 66-d feature vector encoding both visual and geometric information is designed to identify A/V types of vessel pixels, and the classification result is then utilized to label vessel trees via mutual correction.

Comparing to the graph-based method which considers both vascular tree disentanglement and vessel type classification, the dense method mainly focuses on employing distinguishable features and machine learning models for accurate pixel-wise A/V classification. Niemeijer *et al.* [16] proposed to extract a set of features including intensity and derivative information from each vessel pixel and employed a k-Nearest Neighbor (kNN) classifier to predict vessels' A/V types. Zamperini *et al.* [8] demonstrated that the mixture of color and contrast information inside and outside vessels, and positional information can provide effective features for accurate vessel type classification. Huang *et al.* [15] proposed four new features associated with the lightness reflection of vessels to robustly model differences between arteries and veins. Recently, deep learning techniques show a great potential in dense methods. Meyer *et al.* [9] adopted a Fully-connected Convolutional Network (FCN) [20] for A/V classification and achieved an extremely high accuracy on the DRIVE dataset. Girard *et al.* [13] utilized a convolutional neural network (CNN) to jointly segment and classify vessel pixels into arteries and veins, and achieved a promising A/V separation performance on the CT-DRIVE database.

Despite some progress, both graph-based and dense methods share a common inherent flaw that the two tasks, i.e., vascular tree disentanglement and vessel type classification, are treated independently. However, non-trivial improvements brought by a mutual correction [17], [18] between results of the two tasks intuitively imply that some common features should be shared among them, which could positively guide each other. For instance, knowledge for A/V type classification can facilitate us to uncouple vessels which are intertwined or cling to each other in parallel. On the other hand, the types spread from vessel trunks could help to label connected branches which could be very difficult to classify due to low resolution and/or imaging artifacts. As a result, distilling and transferring knowledge across the two tasks can be a potential solution to further improve the performance of A/V separation [21].

Besides, extracting effective CNN features that are beneficial to both tasks remains understudied. Existing dense methods typically use a regular convolution operation to extract features in a whole image vertically and horizontally, resulting in noises from background and/or other neighboring vessel segments. Although such noises can be filtered out by vessel masks, the features that mix visual and geometric information could degrade the A/V separation performance. This is because visual and geometric features play different roles in vascular tree disentanglement and vessel type classification. For tree disentanglement, visual and geometric information are complementary in different situations, while visual information is often decisive for vessel type classification. Without decoupling visual information from geometric information in CNN features, they can hardly excel in the two tasks.

In this paper, we aim at exploring the potential of deep learning in the graph-based A/V separation method. To this end, we present a multi-task siamese network which can learn effective deep learning features to jointly handle both tasks, i.e., vascular tree disentanglement and vessel type classification. Specifically, we design a new fashion of convolution operation, named *Convolution Along Vessel* (CAV), which is customized for vessel-like objects. Instead of convolving a whole image pixel by pixel orderly, we constrain the convolutional kernels 'walk' along individual vessel segments. Such operation is implemented by sequentially conducting vessel horizontalization and convolution operation. Since CAV normalizes growing directions of all vessels via straightening and aligning them in the same direction, i.e. horizontal direction, the CAV-derived features are geometric-invariant visual features. The counterpart geometric features are thus obtained by recording the walking tracks of CAV kernels (i.e. directions of vessel blood flow). By doing so, the CAV decouples visual and geometric information, each of which can maximize its own function in different tasks. A siamese network is then introduced to extract features from a pair of vessel segments, followed by two linear mappings for training the network in a manner of multi-task learning as shown in Fig. 1. One of the mappings classifies A/V types based on CAV-derived visual features, and the other uses both visual and geometric features to examine whether the segment pair has to be disconnected or not. We expect that the knowledge (i.e. parameters) distilled from one task could be used by the network for guiding the other, which in turn leads to a better convergence. In addition, the final result of A/V separation is obtained by mutual correction between the outputs of two tasks, where for each vessel segment the unanimous result is retained, and the contradictory one is determined by majority voting.

In summary, our key contributions are listed as follows:

- We design a CAV for decoupling visual and geometric information in CNN features, which excelling deep features for different tasks, i.e., vascular tree disentanglement and vessel type classification.
- We introduce a multi-task siamese network to learn deep features beneficial to both tasks via jointly solving problems of disentangling vascular trees and classifying A/V types.
- Extensive and comprehensive experimental results on three challenging public datasets, i.e, DRIVE, INSPIRE, and WIDE, demonstrate the validity of CAV operation and multi-task siamese network for accurate A/V separation, and a superior performance over existing state-of-the-art methods.

II. Method

Fig. 1 illustrates the proposed multi-task siamese network which consists of two main techniques, i.e., CAV to decouple visual and geometric vessel features and multi-task learning to train the siamese network. Fig. 2 illustrates the framework

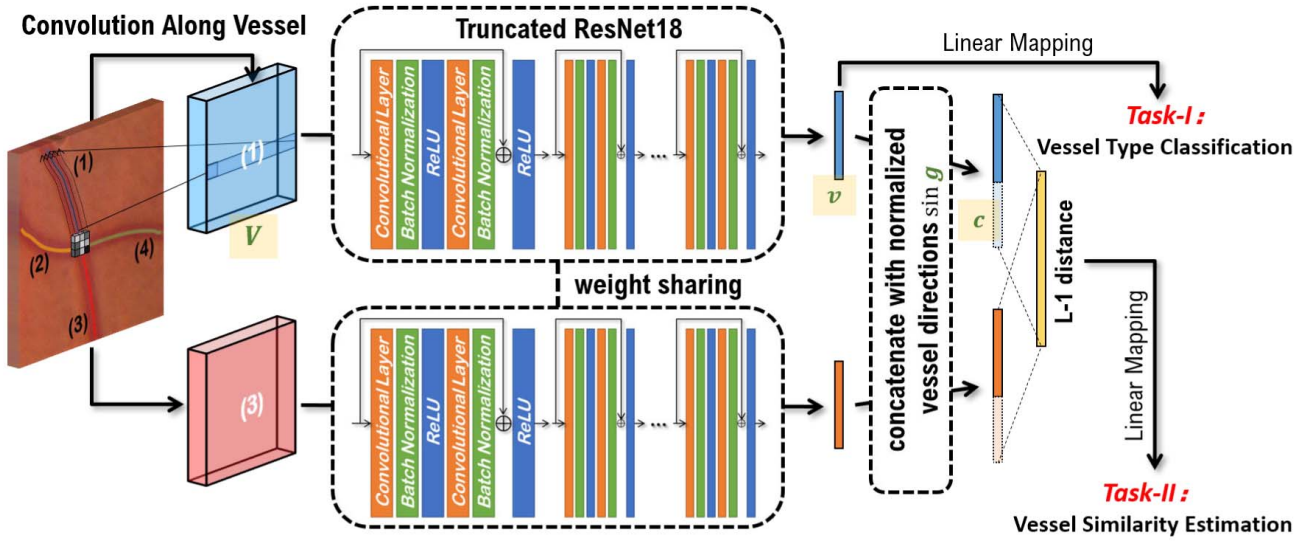

Fig. 1. The framework of our proposed multi-task siamese network for A/V separation. Convolution along vessel (CAV) decouples visual and geometric features of each vessel segments.

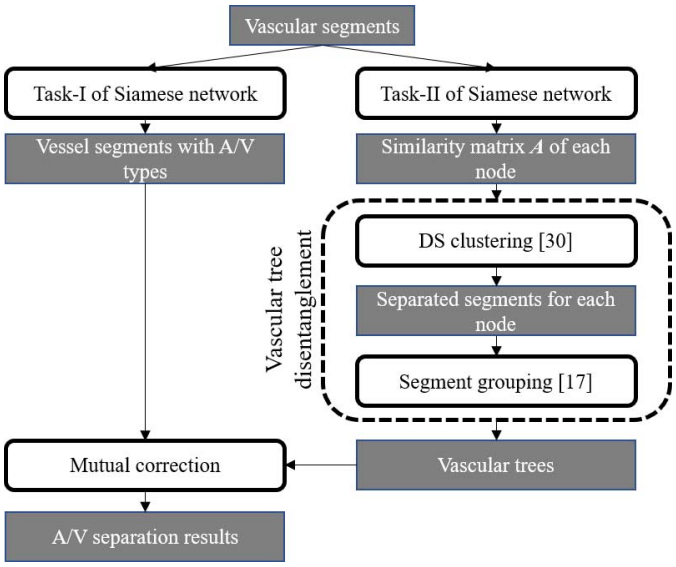

Fig. 2. Framework of our method in the inference phase.

of our method in the inference phase, including vessel type classification (Task-I), vascular network separation (Task-II + vessel tree disentanglement) and mutual correction. In the following, we will detail the two techniques first, followed by methods of vascular tree disentanglement and mutual correction. Before that, we provide a notation table which is presented in Table I to make the paper easier to follow.

TABLE I
NOTATIONS AND CORRESPONDING DESCRIPTIONS TO MAKE THIS WORK EASY TO READ

Notion	Description
$\boldsymbol{C}$	Target control points in the original image
$\boldsymbol{S}$,$\tilde{\boldsymbol{S}}$	Source control points in the transferred image and its homogeneous form
l_{in}, l, l_{ou}	Points at inner parallel, centerline and outer parallel curves of a vessel segment
$\boldsymbol{T}$	Transform matrix
$\boldsymbol{R}$	Radial basis function (RBF) matrix
z,$\tilde{z}$	A pixel's location in the transferred image and its homogeneous form
p	The matching pixel's location in the original image
$\boldsymbol{V}$,$\boldsymbol{v}$	Geometric-invariant visual feature map and visual feature vector
$\boldsymbol{g}$	Geometric feature vector
$\boldsymbol{A}$	Symmetric matrix calculated by the learned Siamese network
$\boldsymbol{c}$	The complete feature vector of both visual and geometric features
G	Group label

A. Feature Extraction via Convolution Along Vessel

1) Three Steps to Locate Individual Vessel Segments: As shown in Fig. 3, we utilize three steps to construct the vascular graph from a given digital fundus image, where individual vessel segments can be located as links of the graph.

Specifically, we first obtain a binary segmentation map of vasculature from a fundus image using the segmentation method proposed in [22]. Most segmentation approaches typically apply Conditional Random Fields (CRFs) inference as an independent post-processing step to improve the pixel-wise segmentation results since CRF is able to refine weak and coarse pixel-level predictions to produce sharp boundaries and fine-grained segmentations. The segmentation method we used in this work goes further and jointly learns the parameters of CNN and CRF by combining the strengths of both CNNs and CRF-based graphical models in a single end-to-end trainable deep network. For the DRIVE dataset, the segmentation

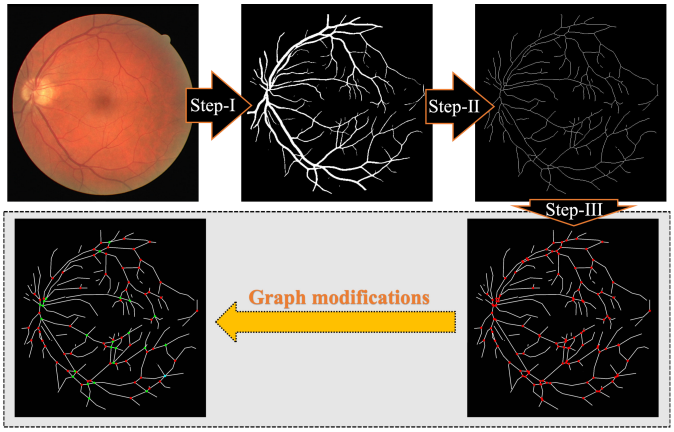

Fig. 3. Three steps to locate individual vessel segments, where three types of graph modifications are used in Step-III. For the graph nodes, nodes with degree 3, 4, and 5 are denoted by red, green, and cyan color respectively.

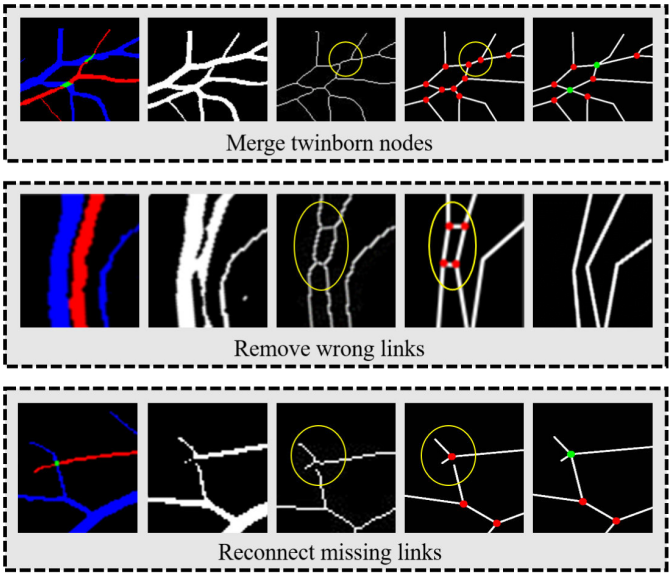

Fig. 4. From left to right image in each row: groundtruth, segmentation map, skeleton map, initial graph, refined graph. The graph error is marked by yellow ellipse. For the graph nodes, nodes with degree 3 and 4 are denoted by red and green color respectively.

method achieves 97.92% and 95.67% of AUC and accuracy respectively, which are very close to the recent state-of-the art method [23] where 97.8% and 95.5% of AUC and accuracy are reported.

Second, we compute the skeleton (i.e. vessel centerline image) by applying the iterative thinning algorithm presented in [24]. This algorithm removes border pixels until the object shrinks to a minimally connected stroke [17].

Third, we build a graph representing the vascular network by finding the branching/crossing points and endpoints on a skeleton map, and connecting those points with graph links as shown in Fig. 3. The initially built graph is rough and contains three types of graph errors, i.e. twinborn nodes (denoted by the red dots in the yellow circle of the first row in Fig. 4), wrong links (denoted by the horizontal edges between the red dots in the second row of Fig. 4) and missing links (denoted

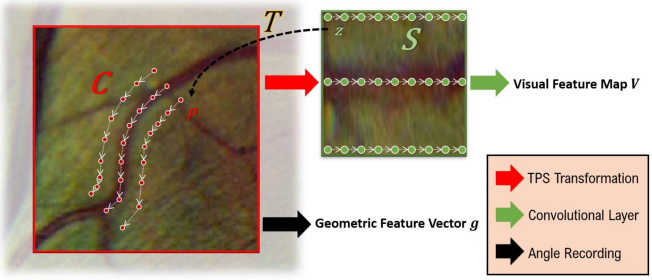

Fig. 5. The mechanism of CAV for decoupling the feature of a vessel segment into visual and geometric feature vectors. TPS transformation is first utilized to straighten a squiggly vessel segment, followed by a regular convolutional kernel to extract a purely visual feature map. The corresponding geometric feature vector is then derived by recording the angles of vessel growth.

by the yellow circle of the third row in Fig. 4). To tackle the graph errors, we refine the graph based on [17]. Specifically, to merge twinborn nodes, we follow [17] to check each two adjacent nodes with degree 3. If the distance between two nodes is sufficiently small and a link in one node (common link excluded) has the same orientation as another link in the other node (the same for the remaining links), then the two corresponding nodes are merged.

To remove wrong links, we follow [17] to check the distance between every two adjacent nodes with degree 3. If the distance is sufficiently small, we then check the angle between the links connected to each node. If two of its links have an identical orientation and are almost perpendicular to the third link (the common link between two nodes), then the common link is a wrongly-detected link and should be removed.

To reconnect missing links, we follow [17] to check the distance from a degree 1 node (i.e. an endpoint) to other nodes with a degree 3. If the distance is sufficiently small, then the nodes will be connected with a new link. According to [17], all the distance thresholds in graph refinement are adaptively determined depending on the vessel calibers and angles. Fig. 4 shows the results before and after graph refinement.

After obtaining the refined vascular graph, every vessel segment can be located by finding its corresponding link in the graph.

2) Details of CAV for Feature Extraction: Given a vessel segment, the dual solution of programming a convolutional kernel to 'walk' along a vessel is to first estimate deformation of transforming the squiggly vessel to a straight vessel, and then straighten the vessel before convolution. To this end, CAV is implemented by approximating a Thin Plate Spline (TPS) transformation to rectify a vessel into a straight one, followed by a regular convolutional layer as shown in Fig. 5. Specifically, we calculate two parallel curve segments which are respectively toward and away from the center of curvature of a vessel segment. The vertical distance between the two calculated parallel curves is set to three times of the average caliber of the vessel segment to include some context information. To improve computing efficiency, we utilize the Douglas-Peucker (DP) algorithm [25] to sparsify these curves into multiple points without losing too much geometric information.

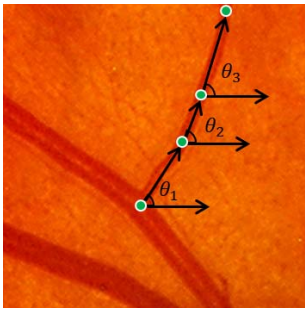

Fig. 6. Geometric feature vector g is obtained by recording the walking tracks of CAV kernels (i.e. directions of vessel blood flow).

We denote the sparse points as *target control points* $\boldsymbol{C}$ in the original image, and *source control points* $\boldsymbol{S}$ in the transferred image respectively. Specifically, $\boldsymbol{C} = [\boldsymbol{l}_{in}, \boldsymbol{l}, \boldsymbol{l}_{ou}] \in \mathbb{R}^{2\times 3m}$, where $\boldsymbol{l}_{in}, \boldsymbol{l}, \boldsymbol{l}_{ou}$ refer to points at inner parallel, centerline and outer parallel curves respectively, as indicated by red dots in the left image of Fig. 5, and m is the number of sparse points at each curve. $\boldsymbol{S} = [\boldsymbol{s}_1, \cdots, \boldsymbol{s}_{3m}] \in \mathbb{R}^{2\times 3m}$ are those at the intersections of regular, equal-spaced grids, as indicated by green dots in the right image of Fig. 5. Based on pairs of source control points $\boldsymbol{S}$ and target control points $\boldsymbol{C}$, the transform matrix T is computed as:

$$\boldsymbol{T} = \left(\boldsymbol{L}^{-1} \cdot \begin{bmatrix} \boldsymbol{C}^T \\ \boldsymbol{0}^{3\times 2} \end{bmatrix}\right)^T \in \mathbb{R}^{2\times(3m+3)} \tag{1}$$

where $\boldsymbol{L}$ is derived from source control points as:

$$\boldsymbol{L} = \begin{bmatrix} \tilde{\boldsymbol{S}}^T & \boldsymbol{R} \\ \boldsymbol{0}^{3\times 3} & \tilde{\boldsymbol{S}} \end{bmatrix} \in \mathbb{R}^{(3m+3)\times(3m+3)} \tag{2}$$

where $\tilde{\boldsymbol{S}} = \begin{bmatrix} \boldsymbol{1}^{3m\times 1} & \boldsymbol{S}^T \end{bmatrix}^T \in \mathbb{R}^{3\times 3m}$ is the homogeneous coordinates of $\boldsymbol{S}$ and $\boldsymbol{R}$ is a symmetric matrix, each entry is $r_{i,j} = \phi(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|_2)$, where $\phi(d) = d^2\, ln(d^2)$ is the radial basis function (RBF).

Based on $\boldsymbol{T}$, the matching relation (shown as back dash line in Fig. 5) between a pixel z in the transferred image and the matching pixel $\boldsymbol{p}$ in the original image is defined as:

$$\boldsymbol{p} = \boldsymbol{T} \cdot \begin{bmatrix} \tilde{\boldsymbol{z}}^T & \phi(\|\boldsymbol{z} - \boldsymbol{s}_1\|_2) & \cdots & \phi(\|\boldsymbol{z} - \boldsymbol{s}_{3m}\|_2) \end{bmatrix}^T \in \mathbb{R}^2 \tag{3}$$

where $\tilde{\boldsymbol{z}} = \begin{bmatrix} 1, & \boldsymbol{z}^T \end{bmatrix}^T$, and $\boldsymbol{z} = [x, y]^T$ is the x, y-coordinates of the pixel.

Accordingly, we can use Eq. 3 and the bilinear interpolation to straighten any given vessel segment as shown in Fig. 5. In theory, the TPS transformed images can have an arbitrary size. For convenience of the following convolution, we resize them into a fixed size of 64×64, and then apply a regular convolution with the kernel size of 3×3 and one pixel zero padding, yielding a geometric-invariant visual feature map denoted by $\boldsymbol{V}$ shown as the blue or pink transparent block in Fig. 1.

The CAV operation can also obtain the corresponding geometric feature vector $\boldsymbol{g}$ where each entry is an angle from a sparse control point to the next along the vessel. Specifically, Fig. 6 visualizes the sparse points on the centerline of the vessel segment $\boldsymbol{l} = [x_1, y_1; x_2, y_2; \ldots; x_m, y_m]^T \in \mathbb{R}^{2\times m}$, where each column $[x_n, y_n], n = 1, \ldots, m$ is the x-y coordinates of a point, denoted by the green dots. Therefore, the geometric feature vector $\boldsymbol{g}$ can be calculated as Eq. 4.

$$\boldsymbol{g} = \left[\theta_1, \theta_2, \ldots, \theta_{m-1}\right] \in \mathbb{R}^{m-1} \tag{4}$$

where $\theta_n = \cos^{-1} \frac{y_{n+1}-y_n}{\sqrt{(x_{n+1}-x_n)^2+(y_{n+1}-y_n)^2}} \in [0, \pi]$ and $n = 1, \ldots, m-1$.

Note that the dimension of $\boldsymbol{g}$ could also be arbitrary, we just resize it to be 16-dimensional for the computational convenience of the following process.

B. Multi-Task Learning via Siamese Network

1) Task-I: Vessel Type Classification: In this task, we have the siamese network focus on differentiating arteries from veins. Specifically, we utilize a ResNet18 [26] with the last prediction layer truncated to further process the feature map $\boldsymbol{V}$ and add a new fully-connected (FC) layer to yield a 16-d deep visual feature vector $\boldsymbol{v}$. Each entry of $\boldsymbol{v}$ is activated by the sigmoid function $\sigma(x) = 1/[1 + exp(-x)]$, and thus has a float value ranging from 0 to 1.

We introduce a FC layer followed by a softmax function as a linear mapping, which takes $\boldsymbol{v}$ as the input and predict two probabilities, i.e. pa_I and pv_I, indicating whether the vessel is arterial or venous. Accordingly, the loss of the Task-I is a binary cross-entropy loss which is formulated as:

$$\mathcal{L}_{\text{Task-I}} = -y_I \log(pa_I) - (1 - y_I) \log(pv_I) \tag{5}$$

where $pa_I + pv_I = 1$, and $y_I \in \{0, 1\}$ is the ground-truth label of vessel type indicating that the vessel segment belongs to an artery ($y = 1$) or vein ($y = 0$).

2) Task-II: Vessel Similarity Estimation: Retinal arteries and veins do not intersect in the 3D space except at the capillaries. However, the 3D-to-2D projection via imaging techniques will lose 3D structural information, yielding a vascular network with arteries and veins incorrectly attached and interlaced. In order to recover those 3D relationships, we need to split those incorrectly connected vessel segments at each graph node into two groups. To this end, we feed the siamese network pairs of locally connected vessel segments, and train the network to estimate the similarity between them. Specifically, the input two segments first go through the CAV layer and truncated ResNet18 sequentially, resulting in two 16-d visual vectors, i.e., $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. Meanwhile, the CAV tracks the angle of vessel directions for two segments respectively, yielding two geometric vectors $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$. The value range of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ is from 0 to 1 due to the activation of sigmoid. To make both visual and geometric features have the same scale, we normalize $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ into $\sin \boldsymbol{g}_1$ and $\sin \boldsymbol{g}_2$. The complete feature of a vessel segment is formed by concatenating both visual and geometric feature vectors formulated as:

$$\boldsymbol{c} = \begin{bmatrix} \boldsymbol{v}^T & \sin \boldsymbol{g}^T \end{bmatrix}^T \in \mathbb{R}^{32\times 1}, \quad \boldsymbol{v}, \boldsymbol{g} \in \mathbb{R}^{16\times 1} \tag{6}$$

After obtaining features of two samples, i.e., $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$, the typical training routine of the siamese network is to minimize the contrastive loss [27] or the triplet loss [28]. Both losses were designed to make the network learn to

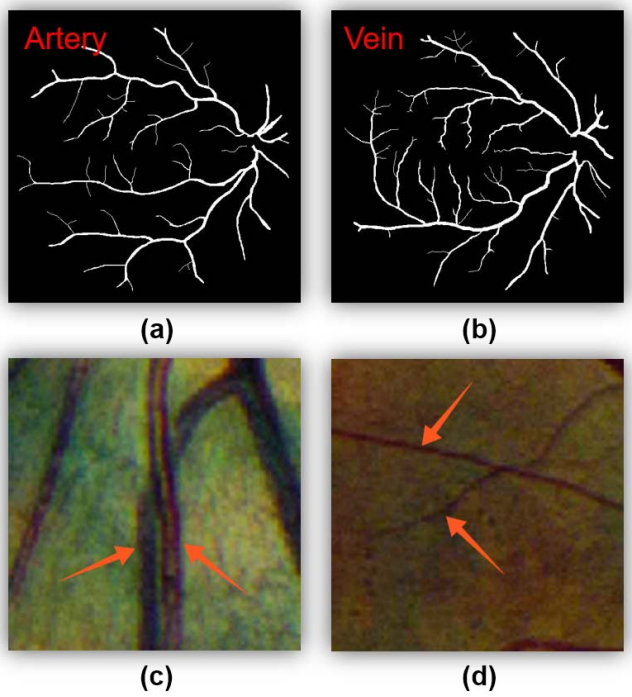

Fig. 7. (a) and (b) Are binary vessel maps of artery and vein vascular network. (c) and (d) Are two pairs of crossover vessels.

minimize the feature distance between two samples if they come from the same class and maximize the distance if they are from different classes. However, in this work, we utilize a linear mapping (i.e. a FC layer) to directly estimate the similarity between two segments of each pair. In detail, we first calculate the L1 distance between the two obtained feature vectors, i.e., $|\boldsymbol{c}_1 - \boldsymbol{c}_2|$. The linear mapping followed by a sigmoid function is then utilized to predict a probability p_{II}, a likelihood that the two segments are from the them class. At last, we train the siamese network to minimize the loss defined as:

$$\mathcal{L}_{\text{Task-II}} = -\left[y_{II}\log(p_{II}) + (1 - y_{II})\log(1 - p_{II})\right] \quad (7)$$

where y_{II} is dynamically calculated based on Eq. 8.

$$y_{II} = \begin{cases} 1 & \boldsymbol{c}_1 \text{ and } \boldsymbol{c}_2 \text{ are with the same vessel type} \\ 0 & \boldsymbol{c}_1 \text{ and } \boldsymbol{c}_2 \text{ are with different vessel types} \end{cases} \quad (8)$$

3) Validity Analysis and Training Details: The most advantage of our proposed CAV is allowing us to decouple visual and geometric features under the framework of deep learning advances, which in turn enables the siamese network to efficiently learn distinctive representations with distracting information suppressed. For example, when comparing the arteriosus network with the venous network which are shown in Fig. 7(a) and Fig. 7(b), the difference in the geometric aspect can be hardly observed. For vessel segments from a global perspective, growing directions of vessels appear to have no distinguishable patterns among arteries and veins. The network could be confused if we force it to discriminate arteries from veins based on vessel orientations, and therefore we only input the visual features to the siamese network in the Task-I.

On the other hand, visual and geometric information benefits the Task-II in different situations. For most graph nodes where an artery and a vein intersect or touch, both visual and geometric features (e.g. homogeneity of color, intensities, directional consistency, etc.) can be used for revealing the 3D structure of vessels. For those interlaced arteries and veins with similar orientations as shown in Fig. 7(c), visual features are obviously very decisive for separating vessel segments with different types. However, for those interlaced vessels whose visual information is indistinguishable or broken by imaging artifacts, vessel directions still can be utilized for separation by verifying their continuities as illustrated in Fig. 7(d). Based on above analysis, we concatenate both visual and geometric feature vectors, and expect the siamese network to intelligently use them in the Task-II according to different situations.

For the siamese network, we propose three key rules for training and inference. First, we only input thick vessel segments in Task-I whose average caliber is greater than a threshold since thin vessel segments usually yield indistinctive visual features. The threshold is set according to the vessel thickness distribution of a dataset. Second, we randomly sample vessel segments as inputs without violation of the first rule in Task-I, and only sample segment pairs around graph nodes with degree greater than 2 in the Task-II. For those graph nodes with 2, we handle them using the rule-based algorithms described in [17] in the inference phase. At last, we alternately minimize $\mathcal{L}_{\text{Task-II}}$ and $\mathcal{L}_{\text{Task-I}}$ to avoid neutralization of gradients derived by the two losses.

C. Vascular Tree Disentanglement

The Task-II of the learned siamese network can generate a similarity matrix $\boldsymbol{A}$ for each node, each entry of $\boldsymbol{A}$ indicates the similarity of every pair of vessel segments connecting to the node. To disentangle a vascular graph into several separated trees, each of which shares the same vessel types for all segments, we first apply the Dominant-Set (DS) clustering algorithm [29] to each matrix $\boldsymbol{A}$ to grouping similar vessel segments together for each node.

Specifically, for each node DS clustering estimates a vector $\boldsymbol{x}^* \in \mathbb{R}^{n\times 1}$ by solving Eq. 9. The nonzero components in $\boldsymbol{x}^*$ denote corresponding vessel segments sharing the same type which are then grouped together. The rest of the zero components are then grouped into another cluster.

$$\begin{aligned} &\text{maximize} f\ (\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} \\ &\text{subject to } \boldsymbol{x} \geq \boldsymbol{0} \text{ and } \boldsymbol{e}^T\boldsymbol{x} = 1 \end{aligned} \quad (9)$$

We further group sets of vessel segments based on [17]. That is, we first identify the location of the optic disc center (ODC) using the automatic method based on the entropy of vascular directions [30]. The farthest segment from the ODC is located, and a group label is assigned to it, e.g. G^1. The other segments connected to it through the same graph node are assigned either the same group label G^1 or a new label, e.g. G^2, based on DS clustering. After repeating this procedure until every segment is retrieved, we can obtain a set of groups, i.e. $\{G^i, i = 1, 2, \ldots, N\}$, where each group represents an individual vessel tree, and N is the number of extracted trees.

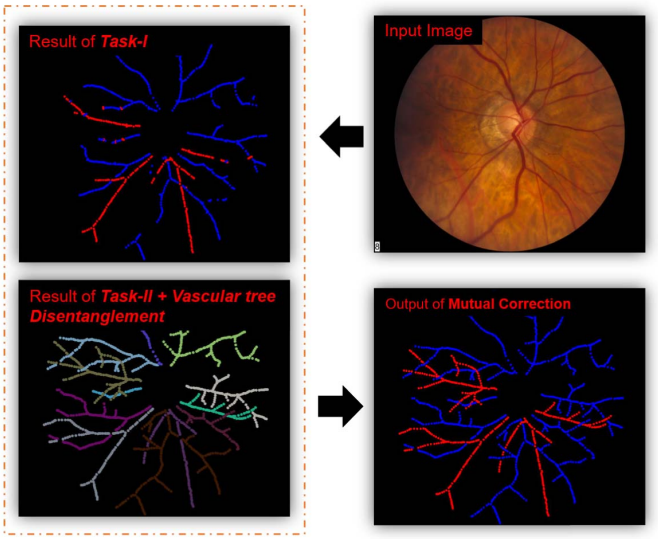

Fig. 8. The example illustrates the effectiveness of mutual correction between outputs of vessel tree disentanglement and A/V type classification of thick vessel segments.

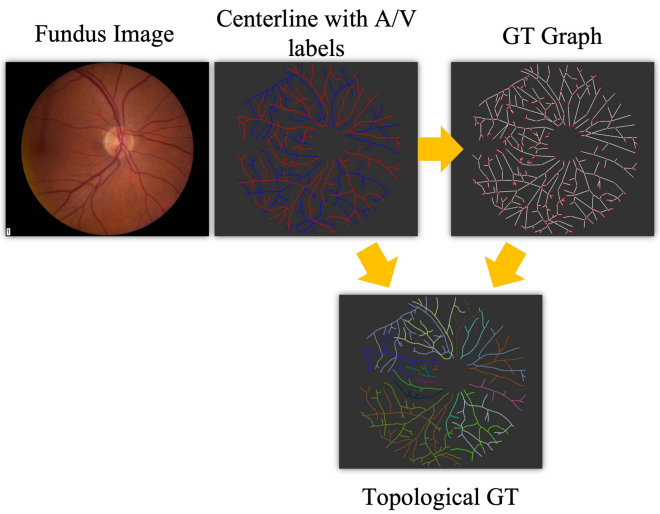

Fig. 9. An exemplar generated topological GT.

D. Mutual Correction for A/V Separation

The result of Task-I is vessel type predictions of all segmented thick vessels as shown in the left-up image in Fig. 8. We denote this result as $\{l^i : \langle pa_I^i, pv_I^i \rangle\}$, where l^i is a vessel segment indexed by i, and $\langle pa_I^i, pv_I^i \rangle$ is its probabilities of being arterial and venous respectively. Note that $\langle pa_I, pv_I \rangle$ could be non-existent if the l is a thin vessel. The result of Task-II is several vessel trees as shown in the left-down image in Fig. 8. We denote this result as $\{G^i : \{l\}\}$, where G^i is one of extract vessel trees indexed by i which is composed of a set of vessel segments $\{l\}$. We sequentially perform two steps for mutual correction of these two results, yielding the final A/V separation:

1. Assign each vessel tree G a vessel type based on the $\langle pa_I, pv_I \rangle$ of all thick vessel segments in G. Suppose there are N thick vessel segments in G, the vessel type P is:
$$P = \begin{cases} \text{artery} & \sum_{i=1}^{N} pa_I^i > \sum_{i=1}^{N} pv_I^i \\ \text{vein} & \text{otherwise} \end{cases} \quad (10)$$
which means that if the majority of segments have a higher possibility of being arterial, the entire vessel tree G will be assigned an arterial label and vice versa. By doing so, all segments including both thin and thick ones should share the same vessel type with the tree they belong to.
2. For segments classified with a strong confidence, keep their initial vessel type instead of enforcing them to share the same type between them and their corresponding vessel tree(s). Specifically, for a vessel segment l^i with $pa_I^i > 0.9$, we keep its vessel type as artery no matter what type is assigned to it in the first step. Similarly, for a vessel segment l^i with $pv_I^i > 0.9$, we keep its vessel type as vein.

The first step corrects the vessel types of those misclassified thick vessel segments and unclassified thin ones by the strategy of majority voting and sharing type within an extracted vessel tree. The second step guarantees that the classified vessel segments with high confidence values will not be affected in the first step in case of accidental errors in vessel tree extraction. Fig. 8 illustrates the result of mutual correction between the two results produced by the siamese network.

III. Materials

A. Datasets and Annotations

We tested our proposed method on three publicly available datasets, namely the Digital Retinal Images for Blood vessel Extraction (DRIVE) [31], the Iowa Normative Set for Processing Images of the REtina (INSPIRE) [6], and the images acquired with an ultra-wide-field device (WIDE) [32].

The DRIVE dataset consists of 40 images with the size of 565×584 and its ground-truth (GT) A/V labels have two versions, i.e. AV-DRIVE [33] and CT-DRIVE [17]. For AV-DRIVE, three different human graders manually labeled all the vessel pixels and the agreements between them were adopted as the final A/V labels. CT-DRIVE only provided A/V labels for the DRIVE test set (i.e. 20 images), where just vessel centerline pixels were marked. To obtain more precise labels, we first skeletonized the label images of AV-DRIVE to get A/V labels of those centerline pixels, and then took a consensus between AV-DRIVE and CT-DRIVE to obtain final A/V labels for images in the test set of the DRIVE. The INSPIRE dataset consists of 40 images with the size of 2392×2048, and the WIDE dataset contains 30 SLO-images with the size of 1440×900. For both INSPIRE and WIDE databases, an image analysis expert manually classified and then an ophthalmologist checked and corrected the blood vessel segments into arteries and veins for centerline pixels.

To evaluate the performance of vascular tree disentanglement, we also obtained topological GT of the retinal vascular structure as shown in Fig. 9. Specifically, we first build a graph representing the vascular network by finding the branching/crossing points and endpoints on the GT vessel

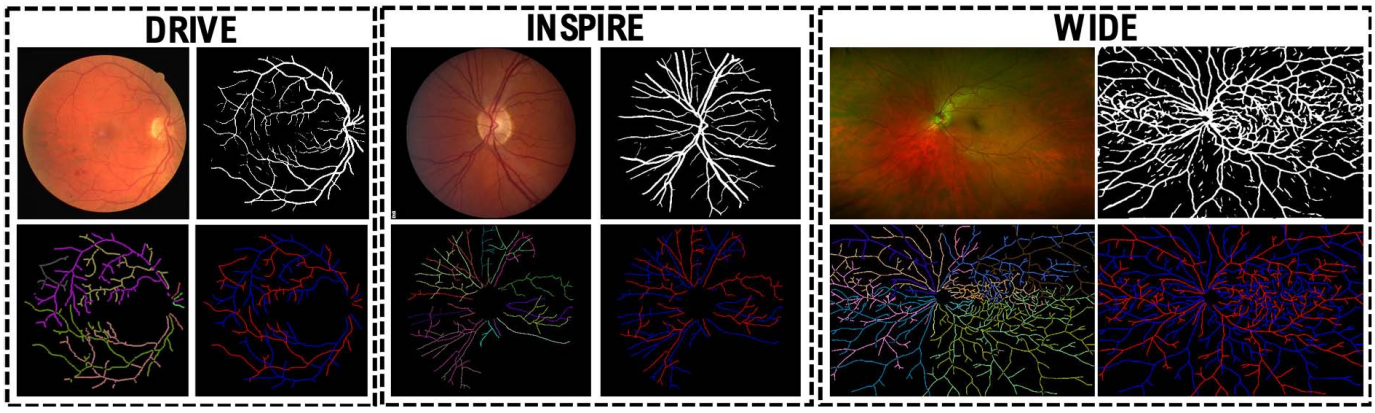

Fig. 10. Samples from the three datasets, i.e. DRIVE, INSPIRE and WIDE visualize characteristics of image quality, complexity of vascular network and A/V vessel spatial distribution among these datasets.

centerline. Then we identify every node and determine whether to disconnect and/or reconnect its links (i.e. segments) or not. This process is errorless since we only need to disconnect links with different A/V labels and reconnect links with the same A/V label. After obtaining the GT graph with identified nodes, we can easily generate the topological GT using the disentanglement approach described in Sec. II-B. Similarly, this process is errorless and thus the topological GT obtained based on the above process is the true GT that can be used to evaluate our proposed method.

Fig. 10 displays samples of raw fundus images, topology GT and A/V GT from which we could observe characteristics of image quality, complexity of vessel structure and A/V spatial distribution among different datasets. Automated vessel segmentation maps are also presented. Note that only DRIVE provides annotated segmentation masks for training a vessel extractor. Therefore, we utilize the vessel segmentation method trained on DRIVE to extract vessels for the three datasets, and the extracted vessels are only used for picking thick vessel segments. For constructing vascular graphs, we utilized automated vessel segmentation maps for DRIVE, and the ground-truth vessel centerline maps for both INSPIRE and WIDE.

B. Evaluation Metrics

To evaluate the A/V classification performance, we used four metrics widely-used in the field of binary classification: Sensitivity (Se), Specificity (Sp), Accuracy (Acc) and Youden's index (YI) [34] defined as follows:

$$Se = \frac{TP}{TP + FN}, \quad Sp = \frac{TN}{TN + FP}$$
$$Acc = \frac{TP + TN}{TP + FP + FN + TN}, \quad YI = Se + Sp - 1 \quad (11)$$

where TP, TN, FP and FN denote true positives, true negatives, false positives, and false negatives respectively.

As artery is treated as the positive in our A/V classification, Se and Sp are used to measure the capability of our algorithm in correctly classifying artery and vein segments, respectively. Youden's index (YI) [34] or Bookmaker Informedness (BM)

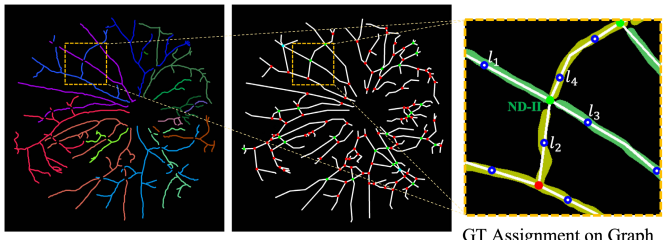

Fig. 11. An example about how we exam the correctness of node identification. From left to right: the topological GT where different colors indicate different group labels; the graph which is built based on the automatic vessel segmentation, where nodes with degree 3, 4, and 5 are denoted by red, green, and cyan color respectively; an image about how to assign group labels on graph links connected to the same node.

metric is one of the well-known diagnostic tests. It is a single statistic that captures the performance of a dichotomous diagnostic test and considers both the Se and Sp equally to compute the trade-off between them. YI reaches its best value at 1 (perfect sensitivity and specificity) and worst at 0.

For the task of vascular tree disentanglement, there lacks an effective evaluation metric to precisely measure the performance of a topology estimation. This is because there are large variations in properties of vessels of different sizes, and the properties cannot be captured by a single metric [19]. Inspired by [19], we calculated the number of the correctly identified nodes (i.e. the vertices of the estimated topology tree that has been assigned labels identical to the ground truth) divided by the total number of nodes of a blood vessel tree. Note that those nodes with a degree equal to 2 are excluded in the calculation since our proposed method did not handle them in the task of vascular disentanglement. Such overlap rate Q is formulated as:

$$Q = \frac{\#\text{ of correctly identified nodes}}{\#\text{ of total nodes}} \quad (12)$$

which reveals the percentage of nodes that are correctly identified by the proposed method [35].

Since the self-built graph and the topological GT might not be perfectly aligned with each other as shown in Fig. 11,

matching between them should be performed for evaluating the performance of node identification based on the metric Q. Instead of matching the nodes, we actually performed the matching process on the vessel segments, a.k.a. graph links.

Specifically, for validating a node, we first locate all links connected to the node, and use the midpoint of the link to represent the position of it. As shown in the rightmost image of Fig. 11, we take a node as an example, which has four connected links $\{l_1, l_2, l_3, l_4\}$ and the position of each link is indicated by the blue ring. For each link l_i, the group label $g|_{l_i} \in \{1, \ldots, N\}$ is defined as the same group label with the nearest vessel pixel, where N is the total number of vascular trees in the topological GT. For instance, the group labels of l_1 and l_3 are the same with the green vessel segments and the group labels of l_2 and l_4 are the same with the yellow vessel segments as shown in Fig. 11.

After the separation based on our proposed Siamese network, the four links will be grouped into at most two groups, i.e. L_1 and L_2, where $L_1 \cap L_2 = \phi$ and $L_1 \cup L_2 = \{l_1, l_2, l_3, l_4\}$. Therefore, if the group labels are identical for all links in both L_1 and L_2, the node can be considered as being correctly identified. Specifically, we calculate two values $\sigma_1 = \Sigma_{l_i \in L_1} (\overline{g_1} - g|_{l_i})^2$ and $\sigma_2 = \Sigma_{l_i \in L_2} (\overline{g_2} - g|_{l_i})^2$, where $\overline{g}$ is the averaged value of group labels. σ_1, σ_2 reflect the numerical fluctuation of group labels in each separated group. Therefore, the node is correctly identified if and only if $\sigma_1 = \sigma_2 = 0$.

For evaluation on INSPIRE and WIDE, we adopted a 2-fold validation strategy, which divides the dataset into two even partitions (i.e., 20 images and 15 images in each partition for INSPIRE and WIDE). One partition is used for training and the rest for testing, and vice-versa. The average results of the two partitions are reported. The DRIVE dataset was pre-partitioned into two sets, each of which contains 20 images. Therefore, we used the partition strategy provided by DRIVE and also reported averaged results by 2-fold cross validation.

IV. Ablation Study

We examine the three key contributions: 1) the new feature extraction operation, i.e. convolution along vessel (CAV) v.s. regular convolution (RC), 2) the vascular tree disentanglement approach, i.e. deep learning-based v.s. rule-based, and 3) the training strategy, i.e. multi-task learning v.s. separate learning. Moreover, we also investigate the impact of segmentation accuracy on the performance of the proposed A/V separation method.

A. Convolution Along Vessel v.s. Regular Convolution

For comparison, both CAV and RC are used to extract the first feature map $\boldsymbol{V}$, which is further processed by the truncated ResNet18 as shown in the Fig. 1. In order to embed RC in the siamese network for feature extraction, we crop an image patch based on a tightly-enclosed box for each vessel segment, and resize the patch into 64×64, and extract the feature map $\boldsymbol{V}$ via a convolutional kernel with the size of 3×3. To filter out the background information, we multiply the feature map extracted by RC with the vessel segmentation mask. We trained a model for each task. For training the model for Task-I, we only minimize the loss defined in Eq. 5. For training the model for Task-II, we only minimize the loss defined in Eq. 7.

1) Vascular Tree Disentanglement: We present the results of two different subsets and both, i.e. nodes with node degree (ND) equal to 3 (denoted as ND-I), nodes with ND greater than 3 (denoted as ND-II), and nodes with ND not less than 3 (denoted as ND-I + II). Nodes in ND-I are the most commonly observed cases in a vascular network, and thus the performance of identifying those nodes significantly affects the overall performance. Nodes in ND-II are mostly crossovers and correct identification of them is crucial for the final A/V separation. The result on ND-I + II is the overall performance of node identification.

Table. II shows the results of CAV and RC for vascular tree disentanglement. We compare them under two different configurations, i.e. w/ and w/o sin $\boldsymbol{g}$, referring to taking the geometric features of vessel segments into consideration or not. As we can see, without geometric information (i.e. w/o sin $\boldsymbol{g}$), features extracted by CAV are comparable with, or even better than those extracted by RC. Such results imply that the features with visual and geometric information twisted together in an unknown manner could be worse than pure visual features for node identification. With usage of geometric features, features extracted by both RC and CAV can improve the performance by $5\% - 10\%$ as expected.

It is noteworthy that the CAV operation outperforms RC by a large margin for nodes in ND-II, which are difficult but important nodes to identify since most of them are crossovers. Incorrectly identification of those nodes could significantly degrade the performance of A/V separation after mutual correction. From this perspective, our proposed CAV achieves superior performances to RC for all three datasets.

2) Vessel Type Classification: We only pick vessel segments whose average caliber is greater than a pre-defined threshold value for both training and testing. According to different resolutions of the three datasets, we set the threshold to 3, 10, 5 pixels for DRIVE, INSPIRE, and WIDE respectively. We calculate the four metrics before and after mutual correction (MC), and MC is performed between classification results and *the groundtruth vessel trees*. Table. III shows comparison results. As can be seen, for those thick vessels (i.e. before MC), the features extracted by CAV are more robust and effective to distinguish arteries from veins than those extracted by RC. This comparison results demonstrate that the visual features play more decisive role in the task of vessel type classification.

In summary, the comparison results in both Tables. II and III well validate the effectiveness of our proposed CAV for extracting features of vessel-like objects. The extracted features with visual and geometric information decoupled are more suitable and effective in the tasks of vascular tree disentanglement and vessel type classification.

B. Deep Learning-Based Approach v.s. Rule-Based Approach

We use CAV for feature extraction and train the siamese model by minimizing the loss only designed for Task-II.

TABLE II
COMPARISON RESULTS OF OVERLAP Q BETWEEN A BASELINE A/V SEPARATION MODEL USING A REGULAR CONVOLUTION (RC) LAYER AND OUR PROPOSED MODEL USING CONVOLUTION ALONG VESSEL (CAV)

		DRIVE			INSPIRE			WIDE		
		ND-I	**ND-II**	**ND-I+II**	**ND-I**	**ND-II**	**ND-I+II**	**ND-I**	**ND-II**	**ND-II**
w/o $\sin g$	RC	79.9%	73.9%	78.3%	81.8%	74.9%	80.4%	81.8%	70.3%	79.1%
	CAV	80.6%	74.7%	79.0%	86.3%	80.0%	85.0%	85.2%	74.5%	82.6%
w/ $\sin g$	RC	88.3%	76.1%	85.1%	90.5%	77.0%	87.7%	92.8%	76.6%	88.9%
	CAV	90.2%	80.2%	87.6%	94.0%	87.9%	92.8%	93.0%	82.7%	90.6%

TABLE III
COMPARISON RESULTS OF THE FINAL A/V SEPARATION BETWEEN A BASELINE (RC) AND OUR PROPOSED MODEL (CAV)

		DRIVE				INSPIRE				WIDE			
		Se	**Sp**	**Acc**	**YI**	**Se**	**Sp**	**Acc**	**YI**	**Se**	**Sp**	**Acc**	**YI**
before MC	RC	71.1%	69.5%	70.0%	40.6%	74.2%	74.8%	74.5%	49.0%	68.6%	63.1%	65.5%	31.7%
	CAV	82.8%	79.2%	80.8%	62.0%	82.3%	80.0%	81.2%	62.3%	82.1%	77.9%	80.0%	60.0%
after MC	RC	91.9%	89.2%	91.3%	81.1%	91.2%	89.6%	90.4%	80.8%	87.1%	83.4%	84.4%	70.5%
	CAV	96.7%	93.4%	95.9%	90.1%	96.7%	96.1%	96.3%	92.8%	96.5%	96.0%	95.2%	92.5%

TABLE IV
COMPARISON RESULTS OF NODE IDENTIFICATION BETWEEN THE RULE-BASED APPROACH AND OUR PROPOSED DL-BASED APPROACH

	DRIVE			INSPIRE			WIDE		
	ND-I	**ND-II**	**ND-I+II**	**ND-I**	**ND-II**	**ND-I+II**	**ND-I**	**ND-II**	**ND-II**
Rule-based approach	93.0%	79.9%	89.5%	93.8%	78.4%	90.6%	93.3%	79.4%	90.0%
DL-based approach	90.2%	80.2%	87.6%	94.0%	87.9%	92.8%	93.0%	82.7%	90.6%

We denote this approach as *DL-based approach*. For comparison, we implement four rules proposed in [17] for node identification, and use these rules to separate the vascular graph into vessel trees. We denote this method as *Rule-based approach*.

Table. IV shows comparison results of DL-based and Rule-based approaches in the task of vascular disentanglement. For ND-I, the performances of both approaches are comparable. For DRIVE, the Rule-based approach is even slightly better than DL-based approach. We believe the reason is that the relatively low resolution of this dataset derives noisy visual features for the DL-based approach, which negatively affects the prediction. However, for ND-II, the DL-based approach significantly outperforms the Rule-based approach. Rule-based approach typically utilizes geometric information like angles, curvatures, etc, to separate those wrongly touched vessel segments while nodes in ND-II could be so complicated that beyond the ability of the rule-based approach to analyze. DL-based approach employs visual information like color, brightness, etc, and are powered by training data for accurately identifying those nodes which cannot be handled by those handcrafted rules.

We also investigate the quantitative impact of the performance of vascular tree disentanglement on final A/V separation. For both rule-based and DL-based approaches, we obtain the final result of A/V separation by performing mutual correction between disentangled vessel trees and *groundtruth A/V types* of thick vessel segments. Comparison results shown in Table. V demonstrate that our DL-based approach can lead to more accurate A/V separation that the rule-based approach.

It is also interesting that when comparing the results in the last row of Table. III and those in Table. V, we found that using perfectly disentangled vessel trees can lead to more accurate A/V separation comparing with using perfect A/V types of thick vessel segments. Such finding raises the question: *Is it necessary to put too many efforts on improving the dense-based method to reach perfect A/V type classification?* From this comparison, we believe that the development of an accurate topology estimator for disentangling vascular network can bring more performance improvement of A/V separation considering that there exists no perfect model for the two tasks.

C. Multi-Task Learning v.s. Separate Learning

For comparison, we train two DL-based models by minimizing $\mathcal{L}_{\text{Task-II}}$ and $\mathcal{L}_{\text{Task-I}}$ independently as the baseline which is denoted as separate learning (SL)-based approach. We also train a DL-based model by alternately minimizing $\mathcal{L}_{\text{Task-II}}$ and $\mathcal{L}_{\text{Task-I}}$, and denote this model as multi-task learning (MTL)-based approach.

TABLE V
COMPARISON RESULTS OF THE FINAL A/V SEPARATION BETWEEN THE RULE-BASED APPROACH AND OUR PROPOSED DL-BASED APPROACH

	DRIVE				INSPIRE				WIDE			
	Se	**Sp**	**Acc**	**YI**	**Se**	**Sp**	**Acc**	**YI**	**Se**	**Sp**	**Acc**	**YI**
Rule-based approach	96.3%	93.7%	95.8%	90.0%	92.8%	92.0%	92.4%	84.8%	95.2%	94.7%	94.0%	89.9%
DL-based approach	96.0%	91.9%	94.8%	87.9%	95.4%	95.8%	95.6%	91.2%	95.1%	95.0%	94.1%	90.1%

TABLE VI
COMPARISON RESULTS OF NODE IDENTIFICATION BETWEEN THE BASELINE SL-BASED APPROACH AND OUR PROPOSED MTL-BASED APPROACH

	DRIVE			INSPIRE			WIDE		
	ND-I	**ND-II**	**ND-I+II**	**ND-I**	**ND-II**	**ND-I+II**	**ND-I**	**ND-II**	**ND-II**
SL-based approach	92.3%	85.6%	90.5%	94.0%	87.9%	92.8%	93.0%	83.3%	90.7%
MTL-based approach	94.4%	89.3%	93.0%	97.6%	90.6%	96.1%	96.9%	90.9%	95.5%

TABLE VII
COMPARISON RESULTS OF THE FINAL A/V SEPARATION BETWEEN THE BASELINE SL-BASED APPROACH AND OUR PROPOSED MTL-BASED APPROACH

	DRIVE				INSPIRE				WIDE			
	Se	**Sp**	**Acc**	**YI**	**Se**	**Sp**	**Acc**	**YI**	**Se**	**Sp**	**Acc**	**YI**
SL-based approach	92.0%	86.6%	90.1%	78.6%	94.8%	94.1%	94.5%	88.9%	91.1%	90.4%	89.8%	81.5%
MTL-based approach	96.9%	92.7%	94.7%	89.6%	97.3%	96.6%	96.9%	93.9%	96.0%	95.0%	94.5%	91.0%

Comparison results of node identification in Table. VI shows that, even if the SL-based approach trained on $\mathcal{L}_{\text{Task-II}}$ only has already achieved decent performance, the MTL-based approach still gains $3\%-4\%$ improvement, demonstrating that the information learned in the task of A/V type classification is indeed helpful for separating wrongly connected segments. Beside the node identification, we also provide the comparison results for the task of A/V separation in Table. VII. For all three datasets, multi-task learning can always boost the performance of the counterpart models trained based on separate learning.

In summary, the results in both Tables. V and VII demonstrate introducing MTL can make the model distill the knowledge from one task to boost the performance of the other task, yielding mutually beneficial results of the two tasks and thus resulting in improvement in final A/V separation.

D. Impact of Segmentation Accuracy

Several retinal vessel segmentation methods [20], [23], [36]–[38] have been proposed and recently push the performance of vessel segmentation to the ultimate limit. In this subsection, we aim to exam the impact of segmentation errors on the final A/V separation results. Specifically, for each testing image we obtained six segmentation maps using five state-of-the-art segmentation methods [20], [23], [36]–[38] as well as our own implementation. For [20], [23], [36] listed in Table VIII, we utilized the source code provided by the authors to obtain the segmentation maps and for [37], [38] we directly utilized the segmentation results provided by the authors. We rank all the methods in a decreasing order of the segmentation accuracy. The performance of node identification and A/V separation based on the respective vessel segmentation maps are presented in Table VIII. From the results we observe that the performance of node identification and A/V separation improves accordingly with the increasing of segmentation accuracy. In general, when vessel segmentation accuracy is above 95% (e.g. [23], [36]–[38]), the accuracy of node identification (i.e. 90.63%∼93.03%) and A/V separation (i.e. 90.81%∼94.67%) is satisfactory. When vessel segmentation accuracy drops below 95% (e.g., 94.69% for FCN [20]), the performance of A/V separation drops greatly and becomes unacceptable.

We further visualize the segmentation errors of our implementation which generally occur at three typical locations: 1) At the boundaries of vessels, yielding thicker or thinner segmented vessels than the groundtruth. 2) At the end of thin vessels, yielding miss segmentation of thin vessel segments. This type of error is the most common in our predicted segmentation maps. 3) At the bifurcation points, breaking the connection of two vessel segments. This type of errors is quite rare and often occurs at thick vessels which are close to the optic disk.

Each type of errors has different impacts on A/V separation. For the first and the second type of errors, the connection relationship of vessel segmentations can be well-preserved, and the influence is thus trivial. For the third type of errors, broken connection of segmentations could impede the effectiveness of mutual correction between the results

TABLE VIII
THE PERFORMANCE (%) OF THE PROPOSED METHOD BASED ON SEVERAL VESSEL SEGMENTATION METHODS WITH DIFFERENT ACCURACY VALUES OVER THE DRIVE DATASET

	Vessel Segmentation			Node Identification	A/V Separation
	Se	**Sp**	**Acc**	**ND-I+II**	**Acc**
Vessel-Net [38]	80.38%	98.02%	95.78%	93.03%	94.67%
Our implementation	79.52%	98.02%	95.67%	93.03%	94.65%
MS-NFN [39]	78.44%	98.19%	95.67%	92.97%	94.57%
CE-Net [23]	-	-	95.50%	91.82%	93.02%
DeepVessel [37]	76.03%	-	95.23%	90.63%	90.81%
FCN [20]	72.03%	97.90%	94.69%	89.72%	84.95%

TABLE IX
COMPARISON RESULTS BETWEEN OUR PROPOSED MODEL AND THE RECENT STATE-OF-THE-ART METHODS OF THE SENSITIVITY Se, SPECIFICITY Sp, ACCURACY Acc AND YOUDEN'S INDEX (YI). BEST PERFORMANCE IS PRESENTED IN BOLD

	DRIVE				INSPIRE				WIDE			
	Se	**Sp**	**Acc**	**YI**	**Se**	**Sp**	**Acc**	**YI**	**Se**	**Sp**	**Acc**	**YI**
Human	96.4%	96.0%	96.1%	92.4%	97.8%	97.1%	97.4%	94.9%	97.0%	97.9%	-	94.9%
Niemeijer *et al.* [6]	80.0%	80.0%	80.0%	60.0%	78.0%	78.0%	78.0%	56.0%	-	-	-	-
Muramatsu *et al.* [39]	90.5%	**96.0%**	92.8%	86.5%	-	-	-	-	-	-	-	-
Mirsharif *et al.* [40]	82.7%	85.7%	84.1%	68.4%	-	-	-	-	-	-	-	-
Relan *et al.* [41]	75.3%	55.8%	64.1%	31.1%	92.7%	48.5%	71.4%	41.2%	-	-	-	-
Dashbozorg *et al.* [17]	90.0%	84.0%	87.4%	74.0%	91.0%	86.0%	88.3%	77.0%	-	-	-	-
Lyu *et al.* [42]	78.1%	87.4%	83.2%	65.5%	90.2%	79.4%	85.1%	69.6%	-	-	-	-
Pellegrini *et al.* [43]	-	-	-	-	-	-	-	-	-	-	-	85.0%
Girard *et al.* [44]	92.3%	93.1%	93.3%	85.4%	-	-	-	-	-	-	-	-
Huang *et al.* [15]	-	-	-	-	-	-	85.1%	-	-	-	-	-
Estrada *et al.* [11]	93.0%	94.1%	-	87.1%	91.5%	90.2%	-	81.7%	91.0%	90.9%	-	81.9%
Zhao *et al.* [19]	94.2%	92.7%	91.1%	86.9%	96.8%	95.7%	95.1%	92.5%	**96.2%**	94.2%	91.0%	90.4%
Srinidhi *et al.* [18]	95.0%	91.5%	93.2%	86.5%	96.9%	96.6%	96.8%	93.5%	92.3%	88.2%	90.2%	80.5%
Ours	**96.9%**	92.7%	**94.7%**	**89.6%**	**97.3%**	96.6%	**96.9%**	**93.9%**	96.0%	**95.0%**	**94.5%**	**91.0%**

of vascular disentanglement and vessel type classification. If disconnected vessels incurred by segmentation errors are classified as a wrong A/V type, they will not be rectified by the A/V types of other vessels which should have been connected.

V. COMPARISON WITH THE STATE-OF-THE-ARTS

In this section, we compare our proposed A/V separation method with several state-of-the-art methods [6], [11], [15], [17]–[19], [39]–[44]. Table. IX shows the comparison results where values of Se, Sp, and Acc of the state-of-the-arts are collected from the literatures. The intra-observer results for the three datasets are also provided which are denoted as *Human* in Table. IX. For DRIVE, the annotations of AV-DRIVE and CT-DRIVE are used for calculating the four evaluation metrics. For INSPIRE, we found another released set of annotations provided by [17], and report its Se, Sp, Acc, and YI against the original annotations. For WIDE, we directly utilized the metric values given by the recent work [19].

It is shown that our method outperforms all existing methods on all datasets, with a single exception that its Sp score on the DRIVE dataset is 3.3% lower than that of [39]. For the DRIVE and WIDE datasets, our method achieves higher values of YI revealing better overall performance than all compared methods by 2%. Although, for INSPIRE our method is just slightly better than the approach proposed by Srinidhi *et al.* [18], both theirs and ours are getting close to saturated performance which is only 1% short of human performance of the YI metric. Besides INSPIRE, it can also be seen that the values of sensitivity and accuracy of our proposed method are very close to those of *Human* on the other two datasets: our method obtains competing sensitivities, with 96.9% and 96.0%, compared to 96.4% and 97.0% by the intra-observers for DRIVE and WIDE respectively, and values of accuracy, with 94.7%, compare to 96.1% by the human for DRIVE. Among three datasets, the proposed method achieves the best performance on INSPIRE. The INSPIRE dataset contains images with the highest resolution, and relatively simpler vessel structures, i.e. fewer bifurcations and crossovers.

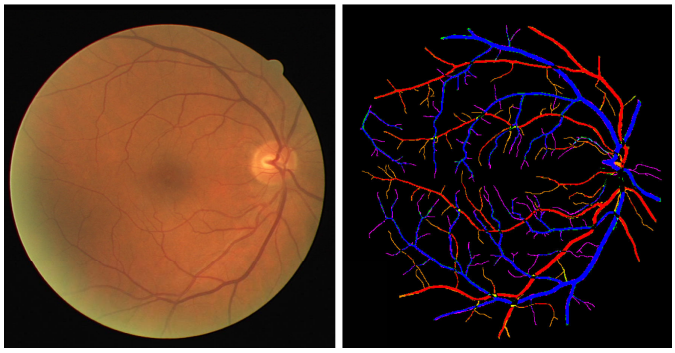

Fig. 12. A sample color-coded result image for DRIVE. The corresponding colors are defined in Table X.

It is noteworthy that we adopt a similar idea as the method proposed by Zhao *et al.* [19] for the task of disentangling vascular network. However, we utilize a siamese network for learning the similarity measure between vessel segments while Zhao *et al.* [19] designed several hand-crafted features including intensity, orientation, diameter, etc, and calculated the similarity based on their distance in the handcrafted feature space. From the comparison results shown in Table. IX, we can observe that the deep learning approach adopted in our method makes the similarity measurement more accurate comparing to those empirically handcrafted features, thus yielding a superior performance of A/V separation. Similar to [19], we also give the computational complexity of our proposed method. Considering the total number of nodes in ND-I + II as scale of the problem or the input size n, the running times of Task-II $T_{II}(n) = 25p + 16q + 9(n - p - q)$, where p and q are the total numbers of nodes with degree equal to 5 and 4 respectively. For the worst case that each node have five links without common links, the running times of Task-I is then $T_I(n) = 5n$. Therefore, the time complexity is $T_I(n) + T_{II}(n) \sim O(n)$.

Furthermore, we for the first time compare our method as a graph-based A/V separation approach with a particular family of A/V separation methods [45], which jointly segments and classifies retinal pixels into three classes, i.e. background, arteriole and venule pixels, and thus the confusion matrix consisting of nine cases should be considered in the validation. To this end, we first transfer our A/V separation results of vessel centerline pixels into the predictions on retinal pixels. Specifically, we match the pixels in the segmentation image with the pixels in the vascular skeleton through the position relationship. Each pixel in the segmentation image shares the same A/V label with its nearest pixel in the vessel centerline processed by our A/V separation method. Since we don't obtain the A/V predictions for the vessel centerline within the optic disc, we simply use the ground truth labels to fill it. A resulting image is visualized in Fig. 12.

Table X compares the 3×3 confusion matrices computed based on our method and [45] on DRIVE. The true background, true vein, and true artery are denoted as $\mathbf{GT}_{\text{bkg}}$, $\mathbf{GT}_{\text{v}}$, and $\mathbf{GT}_{\text{a}}$ and the predicted background, predicted vein, and predicted artery are denoted as $\mathbf{P}_{\text{bkg}}$, $\mathbf{P}_{\text{v}}$, and $\mathbf{P}_{\text{a}}$. For the sake of better comparison, the starred number was normalized to 1. As can be seen in Table X, the segmentation method used in our work misses many vessel pixels. In contrast, Xu *et al.* [45] results in many false positives that wrongly classifies non-vessel pixels as arteries or veins. If only considering the trace of the confusion matrix, our method outperforms Xu *et al.* [45] by a small margin.

TABLE X
COMPARISON RESULTS OF THE TRUE BACKGROUND, VEIN, ARTERY AND PREDICTED BACKGROUND, VEIN, ARTERY ON DRIVE

	Ours			Xu *et al.* [46]		
	$\mathbf{P}_{\text{bkg}}$	$\mathbf{P}_{\text{v}}$	$\mathbf{P}_{\text{a}}$	$\mathbf{P}_{\text{bkg}}$	$\mathbf{P}_{\text{v}}$	$\mathbf{P}_{\text{a}}$
$\mathbf{GT}_{\text{bkg}}$	27.7	0.15	0.13	26.5	0.57	0.69
$\mathbf{GT}_{\text{v}}$	0.25	1.42	0.01	0.07	1.12	0.12
$\mathbf{GT}_{\text{a}}$	0.31	0.03	1.00*	0.07	0.12	1.00*

To better visualize the achieved performances of our method on the three different datasets, we present three examples of results for each dataset. Fig. 13, Fig. 14, and Fig. 15 visualize results of A/V separation for DRIVE, INSPIRE, and WIDE respectively.

VI. DISCUSSION

The experimental results in Section IV and Section V demonstrate the robustness of our proposed method and the effectiveness of individual components, i.e., CAV and multi-task learning, for A/V separation. Despite of obtaining accurate A/V types and vessel similarities, there exist potential factors that could affect the final A/V separation performance of our method. First, we can observe from Section IV-D that improving the accuracy of vessel segmentation could further improve the performance of A/V separation. Therefore, developing more accurate vessel segmentation method and exploring the correlation between segmentation and A/V separation will be explored in our future work. In addition, the errors of graph refinement could also propagate forward and degrade the performance of A/V separation. Those graph errors mainly consist of three types, i.e., missing links, false links and twinborn node. The first and second types of graph errors are very rare and in our experiments they all can be perfectly corrected, and thus have no negative impacts on final A/V separation. The third type of graph errors, i.e., twinborn nodes, is the most common type of errors in the initially built graph. Specifically, for a total of 1613 nodes where vessels cross each other on DRIVE, 1387 nodes (i.e., 86%) were incorrectly extracted as twinborn nodes. Among 1387 nodes, 1347 nodes (i.e., 97.1%) can be perfectly corrected by our graph refinement and will not hurt A/V separation performance. For the remaining 40 nodes which cannot be rectified, as they only redundantly express a graph node but will not break the connection relationship of vessel segments, in our experiment 87.5% of them (i.e. 35 out of 40 nodes)

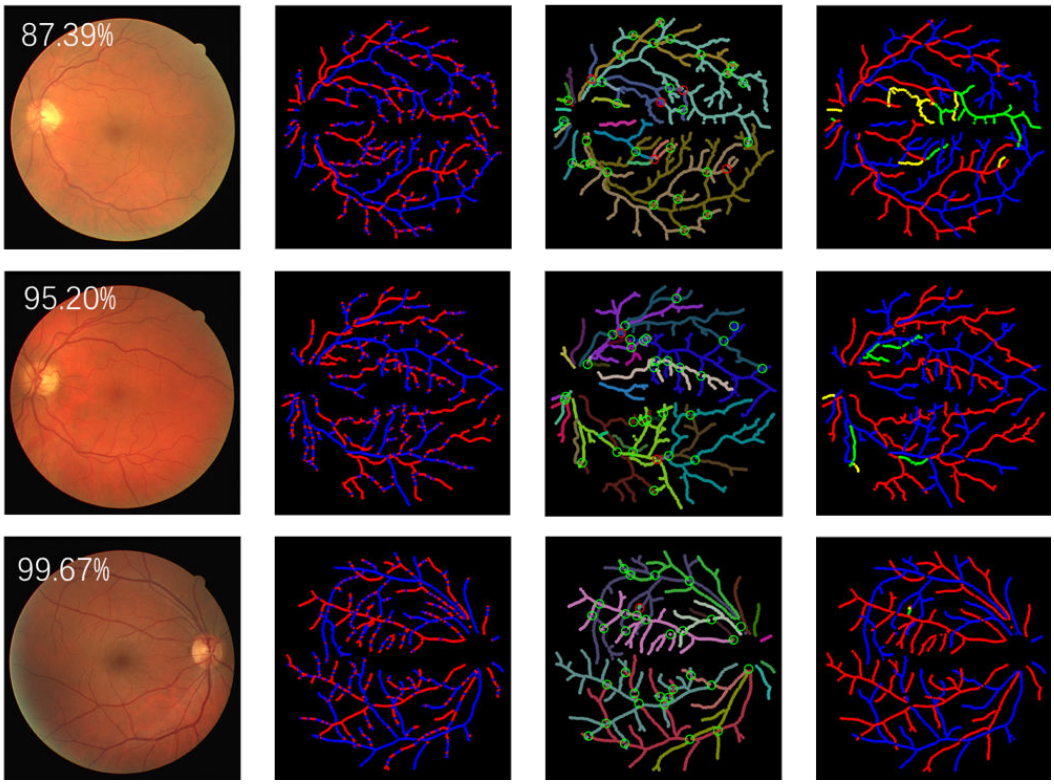

Fig. 13. Three examples which visualize the A/V separation performance of our proposed method on DRIVE. From the left to the right column: the original fundus images overlaid with accuracy values, A/V type classified thick vessels by our method, disentangled vessel trees by our method where red circles mark incorrectly identified nodes and green circles mark correct ones, our predicted A/V separation results where red, blue, green and yellow indicate true positives, true negatives, false positives and false negatives respectively.

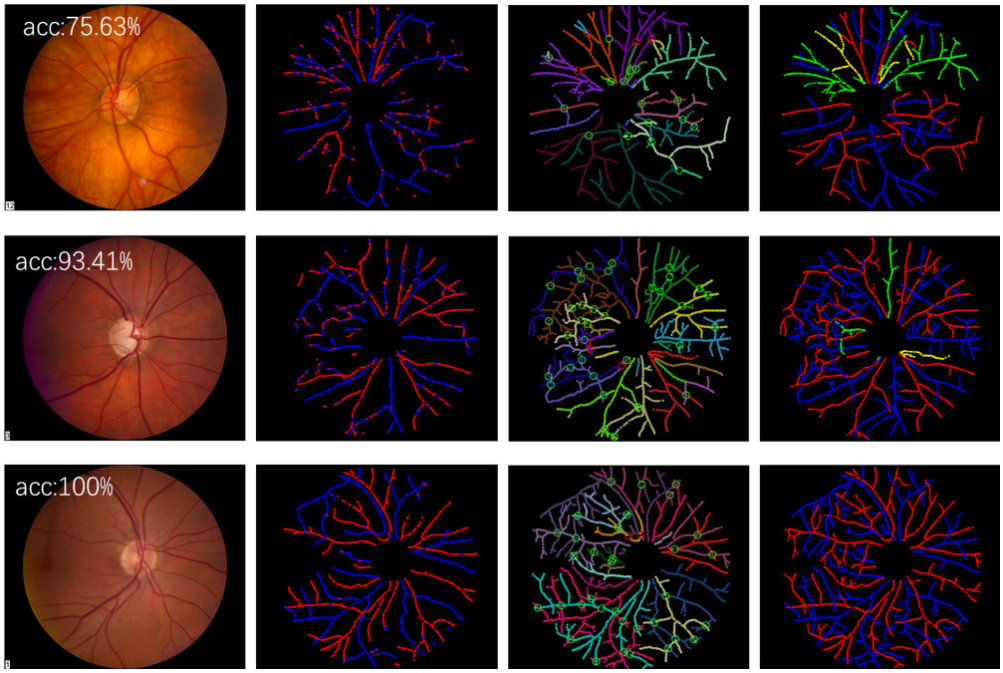

Fig. 14. Three examples which visualize the A/V separation performance of our proposed method on INSPIRE. From the left to the right column: the original fundus images overlaid with accuracy values, A/V type classified thick vessels by our method, disentangled vessel trees by our method where red circles mark incorrectly identified nodes and green circles mark correct ones, our predicted A/V separation results where red circles mark incorrectly identified nodes and green circles mark correct ones, our predicted A/V separation results where red, blue, green and yellow indicate true positives, true negatives, false positives and false negatives respectively.

can be further correctly disentangled and classified by our proposed method. Based on the above discussion, an unified A/V separation framework which can tolerant errors from individual modules and minimize the negative effect of every possible factor will be the exploration in our future work.

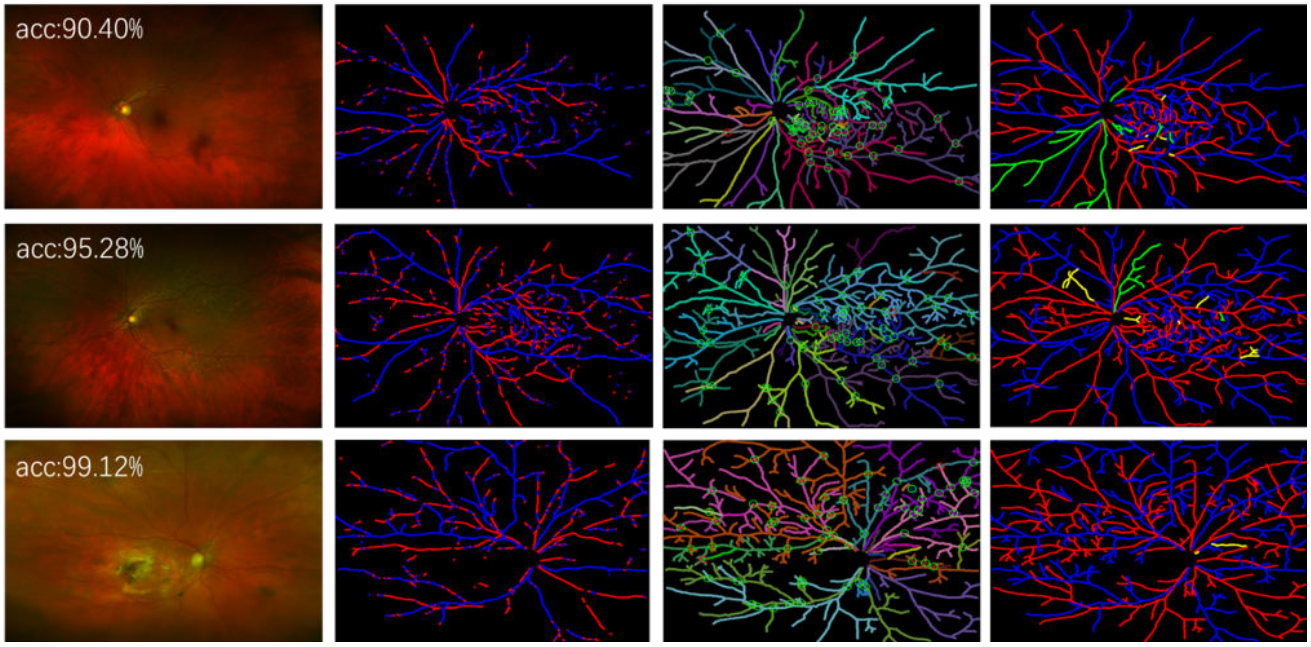

Fig. 15. Three examples which visualize the A/V separation performance of our proposed method on WIDE. From the left to the right column: the original fundus images overlaid with accuracy values, A/V type classified thick vessels by our method, disentangled vessel trees by our method where red circles mark incorrectly identified nodes and green circles mark correct ones, our predicted A/V separation results where red circles mark incorrectly identified nodes and green circle mark correct ones, our predicted A/V separation results where red, blue, green and yellow indicate true positives, true negatives, false positives and false negatives respectively.

VII. Conclusion

In this study, we present an unified multi-task siamese network for accurately separating arteries from veins in retinal fundus image by jointly solving the two tasks, i.e., vascular tree disentanglement and A/V type classification. We design a new fashion of convolution, i.e., CAV, to decouple the CNN features of vessels into visual and geometric information. Based on the decoupled deep features, we train the siamese network to efficiently learn the two crucial tasks by leveraging the visual and geometric features adaptively. For A/V type classification, only visual information is utilized to identify those thick vessels, and for vascular tree disentanglement, both visual and geometric features are used by the siamese network for inferring the true 3D structure of those connected vessel segments. Benefiting from CAV and the multi-task learning, more robust and effective deep features are extracted from vessels, resulting in an accurate A/V separation performance. Comprehensive ablation studies verified the validity of our proposed three key contributions including the new feature extraction operation (i.e., CAV), the deep learning-based approach for vascular tree disentanglement (i.e., the siamese network), and the training strategy of multi-task learning. Extensive comparison results with the recent state-of-the-arts on three widely-used public datasets, i.e., DRIVE, INSPIRE and WIDE, demonstrated our method's superior performance of A/V separation.

References

[1] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, 2010.

[2] C. Y.-L. Cheung, M. K. Ikram, C. Chen, and T. Y. Wong, "Imaging retina to study dementia and stroke," *Progr. Retinal Eye Res.*, vol. 57, pp. 89–107, Mar. 2017.

[3] S. M. Heringa, W. H. Bouvy, E. van den Berg, A. C. Moll, L. J. Kappelle, and G. J. Biessels, "Associations between retinal microvascular changes and dementia, cognitive functioning, and brain imaging abnormalities: A systematic review," *J. Cerebral Blood Flow Metabolism*, vol. 33, no. 7, pp. 983–995, Jul. 2013.

[4] S. McGrory *et al.*, "The application of retinal fundus camera imaging in dementia: A systematic review," *Alzheimer's Dementia, Diagnosis, Assessment Disease Monitor.*, vol. 6, no. 1, pp. 91–107, Jan. 2017.

[5] S. S. Hayreh, B. Zimmerman, M. J. McCarthy, and P. Podhajsky, "Systemic diseases associated with various types of retinal vein occlusion," *Amer. J. Ophthalmol.*, vol. 131, no. 1, pp. 61–77, Jan. 2001.

[6] M. Niemeijer *et al.*, "Automated measurement of the Arteriolar-to-Venular width ratio in digital color fundus photographs," *IEEE Trans. Med. Imag.*, vol. 30, no. 11, pp. 1941–1950, Nov. 2011.

[7] C. Ventura, J. Pont-Tuset, S. Caelles, K.-K. Maninis, and L. Van Gool, "Iterative deep learning for network topology extraction," 2017, *arXiv:1712.01217*. [Online]. Available: http://arxiv.org/abs/1712.01217

[8] A. Zamperini, A. Giachetti, E. Trucco, and K. S. Chin, "Effective features for artery-vein classification in digital fundus images," in *Proc. 25th IEEE Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2012, pp. 1–6.

[9] M. I. Meyer, A. Galdran, P. Costa, A. M. Mendonça, and A. Campilho, "Deep convolutional artery/vein classification of retinal vessels," in *Proc. Int. Conf. Image Anal. Recognit.* Berlin, Germany: Springer, 2018, pp. 622–630.

[10] C. L. Srinidhi, P. Rath, and J. Sivaswamy, "A vessel keypoint detector for junction classification," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 882–885.

[11] 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. Imag.*, vol. 34, no. 12, pp. 2518–2534, Dec. 2015.

[12] S. Abbasi-Sureshjani, I. Smit-Ockeloen, E. Bekkers, B. Dashtbozorg, and B. T. H. Romeny, "Automatic detection of vascular bifurcations and crossings in retinal images using orientation scores," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 189–192.

[13] F. Girard, C. Kavalec, and F. Cheriet, "Joint segmentation and classification of retinal arteries/veins from fundus images," *Artif. Intell. Med.*, vol. 94, pp. 96–109, Mar. 2019.

[14] Y. Zhao *et al.*, "Retinal artery and vein classification via dominant sets clustering-based vascular topology estimation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2018, pp. 56–64.
[15] F. Huang, B. Dashtbozorg, and B. M. T. H. Romeny, "Artery/vein classification using reflection features in retina fundus images," *Mach. Vis. Appl.*, vol. 29, no. 1, pp. 23–34, Jan. 2018.
[16] M. Niemeijer, B. van Ginneken, and M. D. Abràmoff, "Automatic classification of retinal vessels into arteries and veins," in *Proc. Int. Soc. Optics Photon.*, vol. 7260, Feb. 2009, Art. no. 72601F.
[17] B. Dashtbozorg, A. M. Mendonca, and A. Campilho, "An automatic graph-based approach for Artery/Vein classification in retinal images," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1073–1083, Mar. 2014.
[18] C. L. Srinidhi, P. Aparna, and J. Rajan, "Automated method for retinal Artery/Vein separation via graph search Metaheuristic approach," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2705–2718, Jun. 2019.
[19] Y. Zhao *et al.*, "Retinal vascular network topology reconstruction and Artery/Vein classification via dominant set clustering," *IEEE Trans. Med. Imag.*, vol. 39, no. 2, pp. 341–356, Feb. 2020.
[20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
[21] Y. Teh *et al.*, "Distral: Robust multitask reinforcement learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4496–4506.
[22] S. Zheng *et al.*, "Conditional random fields as recurrent neural networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1529–1537.
[23] Z. Gu *et al.*, "CE-net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.
[24] Z. Guo and R. W. Hall, "Parallel thinning with two-subiteration algorithms," *Commun. ACM*, vol. 32, no. 3, pp. 359–373, Mar. 1989.
[25] D. H. Douglas and T. K. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," *Cartographica, Int. J. for Geograph. Inf. Geovisualization*, vol. 10, no. 2, pp. 112–122, 1973.
[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[27] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2006, pp. 1735–1742.
[28] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823.
[29] M. Pavan and M. Pelillo, "Dominant sets and pairwise clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 167–172, Jan. 2007.
[30] A. M. Mendonća, F. Cardoso, A. V. Sousa, and A. Campilho, "Automatic localization of the optic disc in retinal images based on the entropy of vascular directions," in *Proc. Int. Conf. Image Anal. Recognit.* Berlin, Germany: Springer, 2012, pp. 424–431.
[31] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.
[32] R. Estrada, C. Tomasi, S. C. Schmidler, and S. Farsiu, "Tree topology estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1688–1701, Aug. 2015.
[33] T. A. Qureshi, M. Habib, A. Hunter, and B. Al-Diri, "A manually-labeled, artery/vein classified benchmark for the DRIVE dataset," in *Proc. 26th IEEE Int. Symp. Comput.-Based Med. Syst.*, Jun. 2013, pp. 485–488.
[34] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.
[35] J. De *et al.*, "A graph-theoretical approach for tracing filamentary structures in neuronal and retinal images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 257–272, Jan. 2016.
[36] H. Fu, Y. Xu, S. Lin, D. W. K. Wong, and J. Liu, "Deepvessel: Retinal vessel segmentation via deep learning and conditional random field," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2016, pp. 132–139.
[37] Y. Wu *et al.*, "Vessel-net: Retinal vessel segmentation under multi-path supervision," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2019, pp. 264–272.
[38] Y. Wu, Y. Xia, Y. Song, Y. Zhang, and W. Cai, "Multiscale network followed network model for retinal vessel segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Berlin, Germany: Springer, 2018, pp. 119–126.
[39] 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. Graph.*, vol. 35, no. 6, pp. 472–480, Sep. 2011.
[40] S. G. Vázquez *et al.*, "Improving retinal artery and vein classification by means of a minimal path approach," *Mach. Vis. Appl.*, vol. 24, no. 5, pp. 919–930, Jul. 2013.
[41] D. Relan, T. MacGillivray, L. Ballerini, and E. Trucco, "Retinal vessel classification: Sorting arteries and veins," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 7396–7399.
[42] X. Lyu, Q. Yang, S. Xia, and S. Zhang, "Construction of retinal vascular trees via curvature orientation prior," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2016, pp. 375–382.
[43] E. Pellegrini, G. Robertson, T. 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. Imag.*, vol. 37, no. 2, pp. 516–526, Feb. 2018.
[44] F. Girard and F. Cheriet, "Artery/vein classification in fundus images using CNN and likelihood score propagation," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2017, pp. 720–724.
[45] X. Xu *et al.*, "Simultaneous arteriole and venule segmentation with domain-specific loss function on a new public database," *Biomed. Opt. Express*, vol. 9, no. 7, pp. 3153–3166, Jul. 2018.