Masked Vascular Structure Segmentation and Completion in Retinal Images

Yi Zhou, Thiara Sana Ahmed, Meng Wang, Eric A. Newman, Leopold Schmetterer, Huazhu Fu, Jun Cheng, and Bingyao Tan

***Abstract*— Early retinal vascular changes in diseases such as diabetic retinopathy often occur at a microscopic level. Accurate evaluation of retinal vascular networks at a micro-level could significantly improve our understanding of angiopathology and potentially aid ophthalmologists in disease assessment and management. Multiple angiogram-related retinal imaging modalities, including fundus, optical coherence tomography angiography, and fluorescence angiography, project continuous, inter-connected retinal microvascular networks into imaging domains. However, extracting the microvascular network, which includes arterioles, venules, and capillaries, is challenging due to the limited contrast and resolution. As a result, the vascular network often appears as fragmented segments. In this paper, we propose a backbone-agnostic Masked Vascular Structure Segmentation and Completion (MaskVSC) method to reconstruct the retinal vascular network. MaskVSC simulates missing sections of blood vessels and uses this simulation to train the model to predict the missing parts and their connections. This approach simulates highly heterogeneous forms of vessel breaks and mitigates the need for massive data labeling. Accordingly, we introduce a connectivity loss function that penalizes interruptions in the vascular network. Our findings show that masking 40% of the segments yields optimal performance in reconstructing the interconnected vascular network. We test our method on three different types of retinal images across five separate datasets. The results demonstrate that MaskVSC outperforms state-of-the-art methods in maintaining vascular network completeness and segmentation accuracy. Furthermore, MaskVSC has been introduced to different segmentation backbones and has successfully improved performance. The code and 2PFM data are available at: https://github.com/Zhouyi-Zura/MaskVSC.**

***Index Terms*— Topology-aware mask, Vessel structure segmentation and completion, Retinal images.**

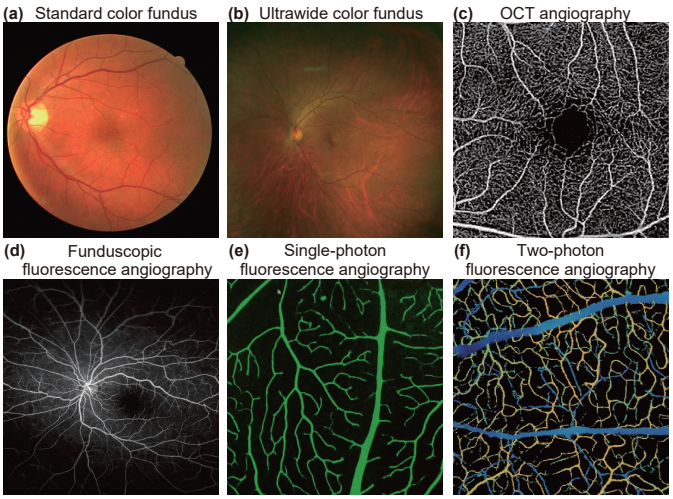

Fig. 1. Retinal vasculature related images. Different in-vivo (a)-(d) and ex-vivo (e)-(f) imaging modalities capture heterogeneous contrasts of the retinal vasculature ((d) is used from [1]). Nevertheless, all the modalities map the continuous, interconnected retinal vascular network onto imaging domains. OCT: optical coherence tomography. Color in f) represents different depths.

This work was funded by grants from the National Medical Research Council (OFLCG/004c/2018-00; MOH-000249-00; MOH-000647-00; MOH-001001-00; MOH-001015-00; MOH-000500-00; MOH-000707-00; MOH-001072-06; MOH-001286-00), National Research Foundation Singapore (NRF2019-THE002-0006 and NRF-CRP24-2020-0001), Agency for Science, Technology and Research (A20H4b0141, C231118001), the Singapore Eye Research Institute & Nanyang Technological University (SERI-NTU Advanced Ocular Engineering (STANCE) Program), and the SERI-IHPC Joint Lab Seeding Funding Program. Corresponding author: Bingyao Tan (Email: tan.bingyao@seri.com.sg).

Yi Zhou and Thiara Sana Ahmed are with Singapore Eye Research Institute, Singapore National Eye Centre, Singapore.

Meng Wang is with Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, United States.

Eric A. Newman is with Department of Neuroscience, University of Minnesota, Minneapolis, Minnesota, United States.

Leopold Schmetterer is with Singapore Eye Research Institute, Singapore National Eye Centre, Singapore; SERI-NTU Advanced Ocular Engineering (STANCE) Program, Singapore; Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore; Institute of Molecular and Clinical Ophthalmology, Basel, Switzerland; School of Chemical and Biomedical Engineering, Nanyang Technological University (NTU), Singapore; Centre for Medical Physics and Biomedical Engineering, Medical University of Vienna, Vienna, Austria; Department of Clinical Pharmacology, Medical University of Vienna, Vienna, Austria; and Rothschild Foundation Hospital, Paris, France.

Huazhu Fu is with Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore.

Jun Cheng is with Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore.

Bingyao Tan is with Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, SERI-NTU Advanced Ocular Engineering Program, Singapore, and Ophthalmology & Visual Sciences Academic Clinical Program, Duke-NUS Medical School, Singapore.

I. Introduction

RETINAL vasculature is responsible for providing oxygen and metabolic supply to retinal neurons. Vessels enter and exit the retina through the optic nerve head, forming a unique closed loop system. The structure of the vasculature is dense, planar, and plexus-based. There are 3 distinct

plexuses: the superficial plexus which has a tree structure, the intermediate and the deep plexus both have a mesh structure. These plexuses are interconnected by vertical anastomoses. Parameters of the retinal angioarchitecture are indicators of retinal health, and are essential for understanding the intricate perfusion of the eye. These parameters include the vessel diameter, localized dilatation or constriction, stenosis, or occlusion, and the presence of neovascularization and/or leakage. Alterations in the angioarchitecture can be associated with various physiological disorders, including glaucoma [2], diabetic retinopathy (DR) [3], retinal vein occlusion [4], and age-related macular degeneration [5]. Additionally, the retinal vasculature is the only vascular structure that is optically visible through the ocular media, making it an optimal organ for investigating the organization of microvasculature.

Multiple imaging modalities map the continuous, interconnected retinal vascular network into the imaging domain, capturing heterogeneous aspects of the retinal vasculature, as illustrated in Fig. 1. For instance, snapshot fundus images primarily capture large vessels, and provide color information about the veins and arteries. Other methods, such as single- or two-photon fluorescence microscopy (2PFM) and optical coherence tomography angiography (OCTA) [6], can image volumetric angiographic images, with capillary detail. Several datasets have been released for fundus vessel segmentation [7]–[10] and artery-vein segmentation [11]–[15], providing essential benchmarks for developing robust segmentation methods.

Accurate segmentation of vascular structures is crucial for the quantitative analysis of vessel-related diseases. Key characteristics of the vascular segmentation task include handling of multi-scale, multi-orientation tubular structures with crossings and branches. Traditional methods for vascular segmentation primarily involve designing filters that match the tubular structure of the vessels [16]–[18]. Morphological operations are also widely used [19]–[21]. However, the structural elements defined by morphological operations are assumed to be linear, which limits their effectiveness when dealing with twisted and blurred vessels. Vessel tracking methods [22]–[24] trace the segments between two selected seed points for a single vessel to discover the best-matched path, but they cannot detect vessels without seed points. Multi-scale methods [25]–[29] detect blood vessels of different thicknesses by altering the image size or changing the filter size. Machine learning algorithms [30]–[35], such as support vector machines and adaptive boosting, can automatically detect retinal vessels, but they are unlearnable.

However, these traditional methods require handcrafted features, which limits their generalization and accuracy. Over the last decade, deep learning (DL)-based methods have emerged as powerful tools for vascular segmentation [36]. Convolutional neural networks (CNNs) have proven to be particularly useful for this task [37]–[41], capturing multi-scale vascular information [42] through a coarse-to-fine segmentation sequence [43] or by emphasizing fine spatial information [44]. Techniques such as boundary detection [45] and edge enhancement [46] have also been effective for fine vessel segmentation. The application of the attention mechanism [47], which captures global associations, has led to the development of various attention modules for retinal vessel segmentation with notable success [48], [49]. Additionally, Graph Neural Networks (GNNs) [50], [51], which construct vessels into a graph structure, are well-suited for processing curvilinear structure data like angiograms [52]–[54].

Despite the good segmentation performance achieved by the methods mentioned above, there are still several challenges remaining:

1) *Manually labeling vessels in retinal images is a cumbersome and time-consuming task, particularly for fine and tiny vessels in images with a low signal-to-noise ratio (SNR). This limits the amount of labeled data available.*
2) *The topological integrity of the angioarchitecure is impacted by disconnected and fragmented vascular structures, especially in small vessels or capillaries.*

To address the challenge of limited data, the use of masking or deletion is becoming increasingly popular. He et al. [55] proposed an efficient self-supervised learning method called Masked Autoencoder (MAE), which enhances the feature learning ability and generalization performance of models by randomly masking input images to generate diverse training samples. Inspired by the above aspects of the work, we propose a novel backbone-agnostic Masked Vascular Structure Segmentation and Completion (MaskVSC) method for predicting interconnected vascular structures with limited manually labeled data. Our main contributions are as following:

- We propose a comprehensive topology-aware mask (TopoMask) strategy by masking random portions of vessel segments to simulate breakpoints or missing branches in angiograms.
- We propose a new connectivity loss function that guides the model to predict the same number of connectivity components as the ground truth, thereby penalizing vessel fragmentation.
- Our MaskVSC method outperforms other state-of-the-art topology-based methods in preserving a more complete vascular structure across five different datasets containing three retinal imaging modalities.

II. Related Works

A. Curvilinear Structures Segmentation

Curvilinear structures, defined as elongated and line-like formations, encompass a variety of natural and man-made structures, including blood vessels. Given the similar challenges these structures face, even in varied contexts, solutions for their analysis often employ analogous methods [56].

Methods for segmenting curvilinear structures can be broadly categorized according to their prediction approaches: 1) Pixel/voxel-based methods [57], [58], which utilize local or global information to compute a feature vector for each pixel/voxel, and then label each one based solely on its features. 2) Edge-based methods [59], which estimate the positions of the curve's borders and subsequently fill the interior region. 3) Shape-based methods [60], [61], which rely on predefined templates and make predictions based on the similarity between the input and these templates. 4) Region-based methods [62], which extract the entire region of

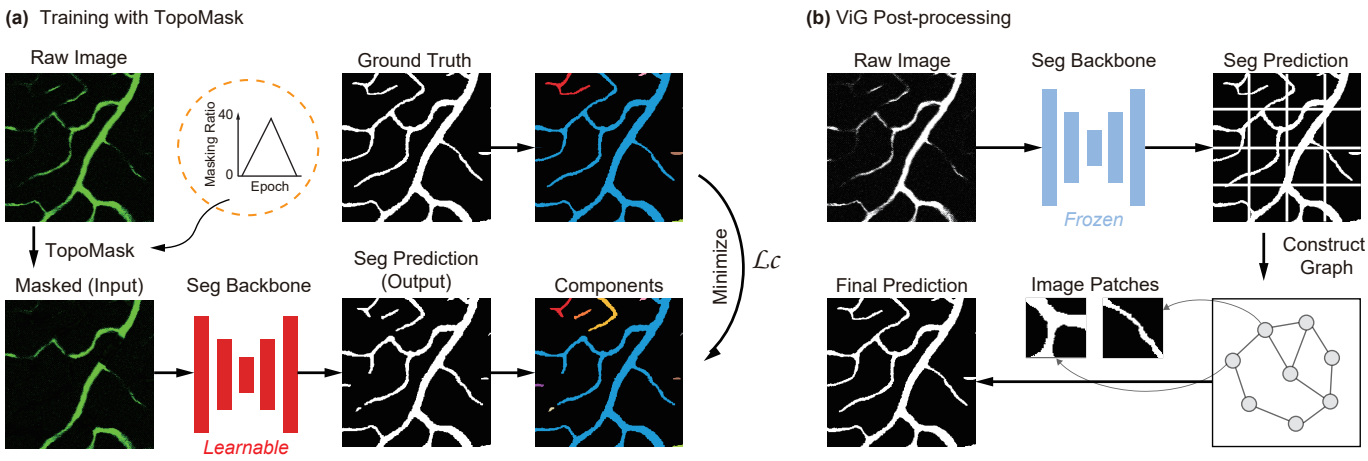

Fig. 2. Schematic overview of our proposed method. (a) Training the segmentation backbone with TopoMask and $\mathcal{L}_C$, where the "Masked (Input)" and "Seg Prediction" are served as the input and output of the segmentation backbone, respectively. (b) Using ViG post-processing the first segmentation prediction. Colors in "components" represent individual isolated components.

interest and may involve post-processing steps to refine the segmentation. 5) User-assisted methods [63], which require varying levels of user involvement, from selecting seed points to providing ongoing feedback throughout the segmentation process. 6) DL-based methods, which leverage labeled data to predict segmentation on unseen samples through learning.

DL-based methods can be further subdivided into: multi-scale [64], multi-task [65], [66], and attention mechanism-based [67], [68]. The performance of these methods is often constrained by the quantity and quality of available annotations. To mitigate this limitation, several advanced learning methods based on weakly supervised [69], semi-supervised [70], self-supervised [71], and unsupervised [72] learning have been proposed. While these methods reduce the dependence on large labeled datasets, they often require complex parameter selection tailored to different modalities. This adds an additional layer of complexity to their implementation.

B. Topology-based Segmentation

The curvilinear or tree-like structure of the vessels in the superficial plexus of the retinal vasculature serves as prior knowledge to design topology-aware loss functions. Improved performances have been achieved using an unsupervised vascular segmentation loss based on active contours [73] and a total variation regularization loss [74]. Shit et al. [75] introduced a loss termed centerline Dice (clDice) that computes the intersection of segmentation and its morphological skeleton. Hu et al. [76] designed a continuous-valued loss (TopoLoss) that enforces the segmentation to have the same Betti number as the ground truth, where the Betti number is a topological invariant for characterizing connectivity and number of disconnected components. Further, they proposed a loss based on discrete Morse Theory (DMT) [77] to identify global structures for better topological accuracy [78]. Later, they used DMT and persistent homology to construct a one-parameter family of structures as the topological representation space [79]. Stucki et al. [80] introduced a Betti Matching error as a topological metric and loss, utilizing induced matchings from persistent homology for spatial matching of topological features. Gupta et al. [81] proposed a topological interaction module to encode this constraint into end-to-end training. They further focused on uncertainty estimation for identifying highly uncertain and error-prone vessel structures [82].

C. Post-processing for Vessel Completion

To better preserve vessel continuity, several post-processing techniques have been developed, including: fully-connected Conditional Random Field (CRF) [83], local gradient consistency extrapolation [84], [85], morphological operation [86], [87], or a probability regularized walk [88]. Additionally, CRF models [89] are often employed in post-processing to enhance segmentation performance. Nevertheless, traditional post-processing modules are unlearnable and can be challenging to optimize. Some studies have attempted to integrate DL-based models [90]–[92], however, semantic segmentation models do not necessarily preserve the vascular topology. Several methods used synthetic disconnected vessel to train a reconnection model for enhancing vascular connectivity after segmentation [93], [94], whereas our MaskVSC integrates connectivity preservation directly into the segmentation model by simulating missing vessel segments and applying a connectivity loss.

III. Method

In this paper, we introduce MaskVSC – a backbone-agnostic method for retinal vascular structure segmentation and completion. It consists of three key components: a topology-aware mask strategy TopoMask, a novel connectivity loss, and a Vision GNN (ViG)-based [95] post-processing. TopoMask randomly masks vessel segments, enabling the model to learn how to reconstruct these masked segments. Segment masking simulates the vessel breakpoints that may appear in real retinal images. The ratio of masked breakpoints is adjusted during training, allowing the model to learn generalized representations of the vessels. The connectivity loss is proposed to enhance vascular continuity by penalizing fragmented segments.

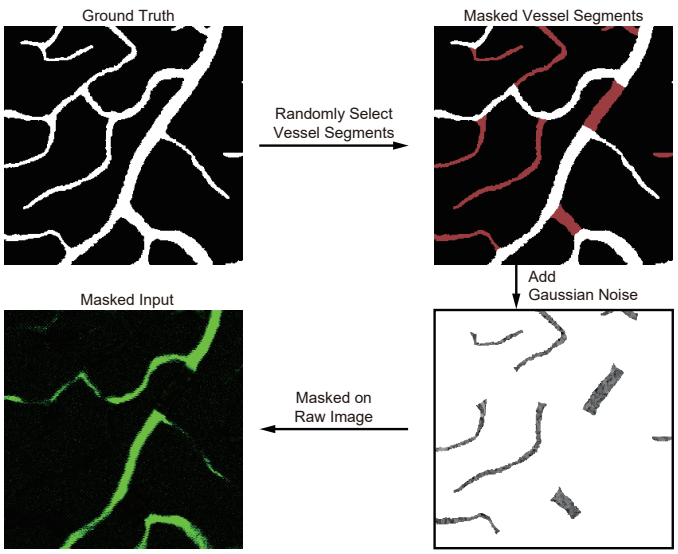

Fig. 3. Flowchart of the TopoMask. We randomly select parts of vessel branches and add Gaussian noise to them, then these noisy branches are masked on the raw image to serve as the input.

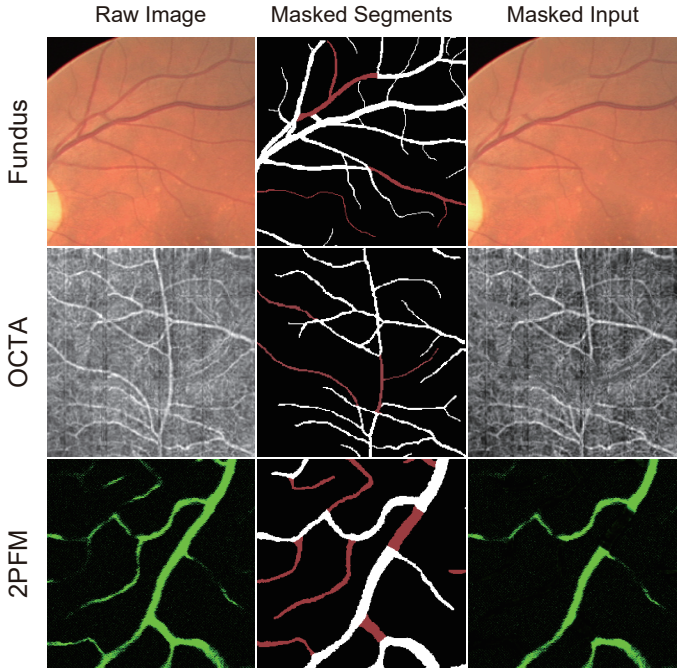

Fig. 4. Raw image, masked vessel segments, and masked input of three different modalities.

Subsequently, a ViG-based model post-processes the initial prediction, leveraging the topological information of retinal vessels by representing the image as a graph structure and processing the nodes and edges using graph convolution.

The training process has two steps. First, we train a vessel segmentation and completion model using TopoMask. Second, we freeze its parameters and proceed to train a ViG-based model to refine the topology of the segmented vessels. During the testing phase, the raw retinal image is fed sequentially into the vessel segmentation and completion model and then the ViG-based model to produce the final predicted vessel map. The flowchart of MaskVSC is shown in Fig. 2, where (a) represents the first step of vessel segmentation and completion, and (b) represents the second step of post-processing.

A. Topology-aware Mask Strategy

Inspired by the masking concept of MAE [55], we initially experimented with using it directly as a pre-training step for our task. However, this approach resulted in unsatisfactory outcomes, particularly in maintaining vascular connectivity, which is crucial for accurate vessel segmentation. This discrepancy may stem from the difference between the target of MAE and our task. Natural images involve complex backgrounds and overlapping objects, emphasizing overall category consistency. In contrast, retinal blood vessels require continuity and topology preservation, necessitating the accurate capture of fine vessels amidst noise and artifacts to maintain network integrity and connectivity. Additionally, MAE is designed for large-scale natural image datasets, typically comprising tens of thousands of images. In the field of medical imaging, data scarcity is a common challenge, leading to potential overfitting and low generalization. To address these challenges and better leverage the characteristics of vascular structures, we propose a topology-aware masking strategy (TopoMask). TopoMask masks random portions of vessel branches to simulate the breakpoints or missing parts in angiograms without disrupting the continuity of other vascular segments, as shown in Fig. 3.

To achieve this, we perform a morphological thinning on the binary ground truth to create a single-pixel-wide skeleton image of the vessels, followed by isolating the skeleton of each vessel segment. Afterward, we randomly select portions of these segment skeletons, dilate the skeletons, with the number of pixels dilated varying according to the dataset, and generate a mask. These binary masks are converted into image backgrounds by filling the masked regions with zero-mean Gaussian noises with different standard deviations (SD), according to the noise characteristics of each imaging modality. This approach can prevent possible gradient explosion or collapse due to model biases. The original and masked images from different modalities are shown in Fig. 4. The masked image serves as the training input and can greatly expand the dataset, thus preventing the model from overfitting on limited data.

It is worth noting that the proportion of masked vessels to all vessels, i.e., the masking ratio, is automatically controlled during training. As the epochs gradually increase, the masking ratio follows a progressive increase from 0, peaking at half of the epochs and then decreasing back to 0, with the first and last 10% of epochs left unmasked for "warm-up" and "cool-down" periods, respectively. This procedure gradually increases the learning difficulty, allowing the model to learn foundational features with complete vessel information early on and gradually adapt to masked conditions. By increasing the mask ratio up to half of the epochs and then decreasing it, the model learns to handle incomplete data effectively, which enhances its robustness and generalization ability. The initial and final unmasked phases stabilize training and consolidate learned features, respectively. This progressive trend of mask-

ing ratio is illustrated as "Masking Ratio" in Fig. 2 under "TopoMask".

B. Connectivity Loss

The vascular network of the retina is a fully interconnected structure, making a fully-interconnected vascular representation essential. While several of the works introduced above have focused on vascular connectivity, these topology- or post-processing-based methods typically emphasize either topology preservation or pixel-level connectivity. They often overlook the prior knowledge that an ideal prediction of vascular structure should contain as few isolated components as possible. Inspired by the connectivity metric in [96], we propose a connectivity loss $\mathcal{L}_C$ to guide the model in generating a more continuous vascular network. $\mathcal{L}_C$ evaluates the connectivity degree of the segmented retinal vessels in the prediction and penalizes vessel fragments by comparing the number of connected components between the prediction and the ground truth, which is denoted as:

$$\mathcal{L}_C = \frac{|\#_C(Y) - \#_C(Y_G)|}{\#(Y_G)}, \tag{1}$$

where Y and Y_G represent the prediction and ground truth, respectively. $\#_C(Y)$ and $\#_C(Y_G)$ denote the number of connected components in Y and Y_G, respectively. $\#(Y_G)$ stands for the number of pixels in Y_G. Specifically, we approximate the number of undifferentiated connected components using a smoothness penalty $S(\cdot)$ that measures pixel differences in spatial dimensions to capture connectivity:

$$\begin{aligned} S(Y) = \frac{1}{N}\Bigg(&\sum_{i,j} |Y(i,j) - Y(i,j+1)| \\ &+ \sum_{i,j} |Y(i,j) - Y(i+1,j)| \Bigg), \end{aligned} \tag{2}$$

where $Y(i,j)$ represents the pixel value at position (i,j) in Y, and N is the total number of pixels in the image, serving as a normalization factor to ensure the penalty is scale-invariant. The first and second terms measure differences between horizontally and vertically adjacent pixels. For Y_G, $S(Y_G)$ is obtained by the same way.

To avoid over-segmenting vessels, such as in an extreme case where all the pixels in the image are labeled to have only one component, the training loss of the vessel segmentation model also includes Binary Cross-entropy (BCE) loss $\mathcal{L}_{BCE}$ - a pixel-wise objective function that directly evaluates the distance between the prediction and the ground truth. Since blood vessels occupy a small portion of the image, BCE loss penalizes the model if over-segmentation occurs and the background pixels are labeled as vessels. This pixel-wise penalty discourages the model from assigning all pixels to a single component and promotes balanced segmentation, effectively reducing over-segmentation tendencies [97]. It is a common loss for binary segmentation tasks so its formula is omitted. Thus, the total objective function $\mathcal{L}_T$ is expressed as:

$$\mathcal{L}_T = \mathcal{L}_{BCE} + \lambda \mathcal{L}_C, \tag{3}$$

where λ is the trade-off parameter between objective functions and is experimentally set to 1 through trials and errors.

C. ViG-based post-processing

After the initial prediction from TopoMask, we utilize the ViG [95] to post-process for better topological feature preservation. ViG treats an image as a graph by dividing it into patches, each treated as a node, and constructs edges by linking nodes according to K-nearest neighbors based on the Euclidean distance. Retinal blood vessels have a sparse structure with an emphasis on connectivity, which is intuitively better modeled by a graph rather than a mesh or normal convolution. ViG considers a vascular image as a graph and processes it based on the connectivity between the nodes. This graph structure enables the model to capture spatial relationships between patches, which is especially useful for maintaining connectivity in vascular structures. We choose ViG due to its ability to encode spatial dependencies and continuity patterns, which are essential for our tasks.

Specifically, ViG divides an image into N patches, which are transformed into feature vectors X. These features become nodes $\mathcal{V} = \{v_1, v_2, \cdots, v_N\}$. For each node v_i, the K nearest neighbors are found, constructing a directed edge e_{ji} from v_j to v_i, forming a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{E}$ denotes all edges. The graph convolution layer then exchanges information between nodes by aggregating neighboring features.

To avoid excessive smoothing due to deep structure, ViG applies a linear layer before and after graph convolution for additional feature diversity and uses a nonlinear activation function to avoid layer collapse. The upgraded module is called the Grapher module. For feature X, the Grapher module can be represented as:

$$Y = \sigma(\text{GraphConv}(XW_{in}))W_{out} + X, \tag{4}$$

where W_{in} and W_{out} are the weights of fully connected layers, and σ is the activation function. Additionally, ViG uses a multi-layer perceptron with two fully-connected layers on each node to further enhance feature transformation capacity and mitigate over-smoothing, forming the feed-forward network (FFN) module:

$$Z = \sigma(YW_1)W_2 + Y, \tag{5}$$

where W_1 and W_2 are the weights of fully connected layers. The Grapher and FFN modules are stacked to form a ViG block, maintaining feature diversity for more thorough learning of discriminative representations. This study chooses ViG-Ti for its lowest computational cost. BCE loss is used as its objective function.

IV. Experiments and Results

A. Datasets

We validate our method on seven datasets containing three different retinal vasculature-related modalities: Fundus (**DRIVE** [30] and **STARE** [16] for vessel segmentation, **RITE** [11] and **LES-AV** [14] for artery-vein segmentation), OCTA (**ROSE** [43] and **OCTA-500** [98]), and **2PFM** from our own collection, with details listed in Table I. The division of training and testing follows the original setup for DRIVE, RITE, LES-AV, ROSE, and OCTA-500. Specifically, DRIVE and RITE are both divided into 20 images for training and 20

TABLE I
DETAILS OF THE DATASETS USED FOR EVALUATION.

Modality	Dataset Name	Number	FOV	Resolution
Fundus	DRIVE [30]	40	45°	584×565
	STARE [16]	20	35°	700×605
	RITE [11]	40	45°	584×565
	LES-AV [14]	21	30°	1444×1620
		1	45°	1958×2196
OCTA	ROSE-1 [43]	117	3×3 mm^2	304×304
	OCTA-500 [98]	300	6×6 mm^2	400×400
		200	3×3 mm^2	304×304
2PFM	2PFM	405	1×1 mm^2	1024×1024

images for testing, and LES-AV is divided into 11 images for training and 11 images for testing. For ROSE, we choose the SVC+DVC pixel-level annotations of ROSE-1 for evaluation and split it into 90 images for training and 27 images for testing. For OCTA-500, we keep its training set (380 images) for training and combine its validation and test sets (120 images) for testing. For STARE, we use a k-fold (k=4) cross-validation method, similar to [68], where 15 images are used for training and the remaining 5 images for testing.

The 2PFM data was acquired from ex-vivo rat retinas, where the rat was perfused with fluorescein-conjugated albumin and gelatin, and the retina was harvested and imaged with a custom-built 2PFM using a 16×, 0.8 numerical aperture objective. The annotation of retinal blood vessels, manually traced by a professional ophthalmologist in the raw images, serves as the ground truth. All 2D images are divided into a training set of 270 images and a testing set of 135 images, with a pixel resolution of 1024×1024. All procedures in this study were conducted in accordance to the ARVO Statement for the Use of Animals in Ophthalmic and Vision Research. Part of the 2PFM dataset is publicly available at our GitHub repository.

B. Implementation Details

The proposed framework is individually trained and tested on each dataset separately. We utilize the Adam optimizer with a learning rate of 2.0×10^{-4} and a momentum of 0.5. The batch size is set to 4, and the maximum number of epochs is set to 200 for each training step. Data augmentation, including random rotation (from -30° to 30°), and random horizontal flipping, is used during training. No augmentation is performed during testing. The resolution of the image is resized to 512×512 for DRIVE and RITE, 1024×1024 for LES-AV, 576×576 for STARE, and 384×384 / 304×304 for OCTA-500, respectively. The image resolution of the remaining ROSE and 2PFM is left unchanged. Zero-mean Gaussian noise added to the masks. The noise's standard deviation is 20 in all three RGB channels for fundus images, 100 for OCTA images, and 20 for 2PFM images, respectively. The proposed method is implemented using Pytorch and is trained on one NVIDIA GeForce RTX 4090 GPU with 24G of memory. Topomask is accompanied by training, hence, the training time is the same with and without it. Specifically, it takes about 1 hour for STARE, 2 hours for DRIVE, RITE, and LES-AV, 4 hours for ROSE-1, and 6 hours for OCTA-500 and 2PFM.

C. Evaluation Metrics

We utilize multiple performance metrics to evaluate different aspects of vascular topology preservation: Dice coefficient ($Dice$), Jaccard index (Jac), centerline Dice ($clDice$) [75]. Also, we evaluate topological correctness using the following metrics: Connectivity (C), Area (A), Length (L) [96], Betti Number error (β^{err}) [76], Betti Matching error (μ^{err}) [80].

$Dice$ and Jac are common metrics for evaluating similarity in medical image segmentation. $clDice$ calculates the intersection of the segmentation masks and their morphological skeleton, maintaining topological invariance up to homotopy equivalence for binary segmentation. The C, A, and L metrics evaluate the degree of fragmentation, overlapping, and coincidence between the prediction and ground truth. They were specifically designed to assess the graphic characteristics of the vascular structure. These three features are combined into a global quality assessment metric, CAL, defined as $CAL = C \times A \times L$. The C, A, and L are defined as:

$$C = 1 - min(1, \frac{|\#_C(Y) - \#_C(Y_G)|}{\#(Y_G)}), \tag{6}$$

$$A = \frac{\#((\delta_\alpha(Y) \cap Y_G) \cup (\delta_\alpha(Y_G) \cap Y))}{\#(Y \cup Y_G)}, \tag{7}$$

$$L = \frac{\#((\delta_\beta(Y) \cap \varphi(Y_G)) \cup (\delta_\beta(Y_G) \cap \varphi(Y)))}{\#(\varphi(Y) \cup \varphi(Y_G))}, \tag{8}$$

where $\#_C(Y)$, $\#_C(Y_G)$, and $\#(Y_G)$ are the same as the corresponding ones in Eq.1. δ_α is a morphological dilation using a disc of α pixels in radius. φ is a skeletonization operation, and δ_β is a morphological dilation with a disc of β pixels in radius to reduce the impact of slight differences in vessel tracing, where β controls sensitivity degree to these differences. For all experiments, α and β are both set to 2 following [96].

The Betti Number is a fundamental metric in topological analysis, with 0- and 1-dimensional Betti Numbers representing the number of connected components and independent holes in a 2D vessel map, respectively. Betti Number error measures the difference between the Betti Numbers of the prediction and the ground truth, computed separately for 0-dimensional (β_0^{err}) and 1-dimensional (β_1^{err}) structures. Betti Matching error is similar to Betti Number error, however, it also considers the spatial location of the 0-dimension (μ_0^{err}) and 1-dimension (μ_1^{err}) components, making it a stricter metric.

D. Results of Comparison Experiments

We compare our proposed MaskVSC method with two categories of state-of-the-art, backbone-agnostic baseline methods: 1) those based on topology-preserving loss, including **clDice** [75], **TopoLoss** [76], and **BettiLoss** [80], and 2) topology-aware methods, including **DMT** [78] and **Gupta et al.** [82]. The U-Net [99] is used as the backbone for all methods. For TopoMask, the masking ratio is set to 40%, as it yields the best performance in Section IV-E. An unpaired t-test is performed to determine whether the improvements over all methods are significant (labeled in **bold**) or not (labeled in *italics*).

TABLE II

RESULTS OF COMPARISON EXPERIMENTS ON SEVEN DIFFERENT RETINAL DATASETS. ALL METHODS USE U-NET [99] AS THE BACKBONE (MEAN ± STANDARD DEVIATION). ↑ MEANS HIGHER VALUES ARE BETTER, AND ↓ VICE VERSA. **BOLD** INDICATES A SIGNIFICANT IMPROVEMENT IN T-TEST, *italics* INDICATES AT LEAST ONE NON-SIGNIFICANT IMPROVEMENT IN T-TEST.

Dataset	Method	Dice (%)↑	Jac (%)↑	clDice (%)↑	CAL (%)↑	β_0^{err} ↓	β_1^{err} ↓	μ_0^{err} ↓	μ_1^{err} ↓
DRIVE	clDice [75]	77.55 ± 1.11	63.54 ± 3.17	81.35 ± 3.32	68.73 ± 5.11	3.08	0.279	3.83	0.434
	TopoLoss [76]	77.63 ± 1.03	63.15 ± 2.72	80.08 ± 3.19	68.68 ± 8.47	2.80	0.285	3.75	0.405
	BettiLoss [80]	76.93 ± 1.07	62.17 ± 2.08	79.06 ± 2.88	67.86 ± 4.11	8.25	0.306	9.83	0.409
	DMT [78]	78.44 ± 0.99	63.32 ± 3.44	82.69 ± 2.92	70.97 ± 6.70	2.74	0.288	3.88	0.375
	Gupta et al. [82]	79.16 ± 1.32	66.27 ± 2.95	83.77 ± 3.64	71.94 ± 7.55	2.30	0.261	2.97	0.365
	Proposed	**80.30 ± 1.07**	**67.19 ± 3.00**	**84.57 ± 3.28**	**73.67 ± 5.16**	**1.99**	**0.254**	**2.63**	**0.348**
STARE	clDice [75]	76.19 ± 1.43	61.72 ± 6.33	82.85 ± 5.83	69.16 ± 9.05	3.99	0.103	3.98	0.102
	TopoLoss [76]	76.77 ± 1.88	62.25 ± 5.95	81.93 ± 5.49	70.37 ± 8.92	3.20	0.105	3.19	0.103
	BettiLoss [80]	75.30 ± 1.34	60.22 ± 5.28	80.98 ± 4.80	67.10 ± 7.69	8.72	0.113	9.02	0.103
	DMT [78]	77.08 ± 2.04	63.94 ± 5.89	82.52 ± 5.04	71.18 ± 9.22	2.91	0.100	2.88	0.109
	Gupta et al. [82]	78.73 ± 2.47	64.82 ± 6.01	84.47 ± 5.55	73.02 ± 11.0	3.00	0.098	2.57	0.102
	Proposed	**79.95 ± 1.74**	**66.15 ± 5.79**	**85.01 ± 5.33**	**73.29 ± 8.63**	**2.04**	**0.094**	**2.33**	**0.100**
ROSE	clDice [75]	72.92 ± 0.95	57.46 ± 4.82	72.34 ± 4.19	56.19 ± 7.76	5.62	0.190	12.4	0.181
	TopoLoss [76]	73.06 ± 1.01	58.52 ± 3.97	71.18 ± 4.11	54.75 ± 6.92	5.60	0.190	12.8	0.193
	BettiLoss [80]	71.24 ± 0.76	55.90 ± 3.54	70.70 ± 3.32	50.43 ± 6.04	9.01	0.212	15.1	0.203
	DMT [78]	74.38 ± 0.98	59.05 ± 4.77	73.26 ± 4.68	55.98 ± 7.50	4.65	0.171	11.0	0.195
	Gupta et al. [82]	74.77 ± 1.52	59.66 ± 5.03	73.45 ± 5.77	56.52 ± 8.11	4.04	0.170	12.0	0.194
	Proposed	**75.25 ± 0.87**	**60.01 ± 4.39**	**73.74 ± 4.36**	**58.41 ± 6.17**	**3.67**	**0.169**	**9.20**	**0.178**
OCTA-500	clDice [75]	74.68 ± 0.84	61.96 ± 5.66	85.93 ± 5.19	72.95 ± 8.98	1.62	0.0788	2.12	0.0890
	TopoLoss [76]	75.01 ± 1.02	62.38 ± 5.18	85.50 ± 5.34	72.01 ± 9.54	1.83	0.0790	2.02	0.0877
	BettiLoss [80]	73.63 ± 0.77	60.81 ± 4.41	82.85 ± 4.76	69.44 ± 8.57	2.18	0.0801	2.63	0.0910
	DMT [78]	75.61 ± 0.85	62.54 ± 6.22	86.17 ± 5.05	73.26 ± 11.5	1.35	0.0787	1.74	0.0875
	Gupta et al. [82]	75.86 ± 0.74	62.76 ± 6.45	86.67 ± 5.13	73.61 ± 12.3	1.23	0.0784	1.62	0.0867
	Proposed	**76.73 ± 0.75**	**63.36 ± 4.96**	**87.94 ± 5.04**	**74.59 ± 9.17**	**1.05**	**0.0781**	**1.38**	**0.0853**
2PFM	clDice [75]	82.06 ± 1.90	67.33 ± 4.91	82.97 ± 4.82	72.99 ± 9.67	2.86	0.130	3.13	0.151
	TopoLoss [76]	82.43 ± 2.07	67.68 ± 3.66	82.54 ± 4.77	73.81 ± 8.95	2.57	0.130	2.92	0.150
	BettiLoss [80]	81.52 ± 1.35	66.20 ± 3.25	80.66 ± 3.26	70.50 ± 6.25	8.20	0.133	6.19	0.159
	DMT [78]	82.90 ± 1.44	68.11 ± 4.21	83.09 ± 4.62	73.93 ± 9.77	2.05	0.129	2.46	0.150
	Gupta et al. [82]	83.77 ± 1.88	68.95 ± 4.73	83.14 ± 5.80	75.17 ± 10.4	1.90	0.128	2.97	0.149
	Proposed	**85.80 ± 1.68**	**69.71 ± 3.95**	**85.21 ± 4.69**	**77.18 ± 8.90**	*1.63*	**0.127**	**2.25**	**0.135**
RITE	clDice [75]	57.19 ± 0.87	38.71 ± 4.32	58.17 ± 2.87	31.14 ± 4.06	5.80	0.184	40.8	1.42
	TopoLoss [76]	57.80 ± 0.79	38.80 ± 4.17	58.91 ± 3.22	31.05 ± 3.82	5.11	0.181	38.2	1.30
	BettiLoss [80]	56.82 ± 0.85	37.49 ± 3.82	57.60 ± 2.92	30.78 ± 3.75	9.44	0.252	60.3	2.17
	DMT [78]	58.04 ± 0.88	39.15 ± 3.96	59.37 ± 2.98	31.80 ± 3.20	4.19	0.162	32.2	1.08
	Gupta et al. [82]	58.38 ± 0.84	39.94 ± 4.08	59.66 ± 3.20	31.95 ± 3.08	4.49	0.144	31.8	0.994
	Proposed	**59.77 ± 0.82**	**41.78 ± 3.91**	**60.18 ± 2.80**	**32.68 ± 3.19**	**3.16**	**0.140**	**30.9**	**0.991**
LES-AV	clDice [75]	59.45 ± 4.19	45.91 ± 7.28	64.08 ± 7.43	33.05 ± 8.28	2.28	0.0304	6.20	0.165
	TopoLoss [76]	60.87 ± 3.22	45.48 ± 10.2	63.88 ± 8.64	32.78 ± 11.4	2.60	0.0328	6.82	0.164
	BettiLoss [80]	59.18 ± 3.85	45.09 ± 9.01	60.50 ± 5.28	30.02 ± 6.32	9.14	0.0523	17.2	0.328
	DMT [78]	61.27 ± 2.47	46.40 ± 10.8	64.44 ± 7.06	34.11 ± 8.63	2.05	0.0309	5.91	0.155
	Gupta et al. [82]	62.08 ± 3.40	46.82 ± 9.25	65.21 ± 6.22	34.18 ± 7.41	1.52	0.0240	5.85	0.143
	Proposed	**63.32 ± 1.84**	**47.19 ± 7.77**	**66.17 ± 6.35**	**35.19 ± 8.07**	**1.05**	**0.0224**	**5.19**	**0.111**

The comparison results, summarized in Table II, show that the MaskVSC significantly outperforms other methods on all five datasets, except for β_0^{err} by Gupta et al. [82] on 2PFM. The superior performance of MaskVSC is demonstrated in its achievement of higher $Dice$ and Jac scores. These scores are achieved because of TopoMask's ability to mimic the blurred vessels by randomly masking segments, which enhances the data topologically during the training to enable the backbone more accurately identify vessel structures in low SNR regions. Additionally, visual inspection of Fig. 5 confirms that the predictions of MaskVSC are more similar to the ground truth. Compared to $Dice$, $clDice$ better captures interconnected vessels and graph similarity, and proposed MaskVSC outperformed $clDice$ due to the connectivity loss $\mathcal{L}_C$, which generates more interconnected predictions. These predictions contain fewer connected components and are closer to the ground truth in terms of graphic features, allowing MaskVSC to also achieve improved results in the CAL metric, which measures the degree of fragmentation between the predictions and the ground truth. The highlighted regions in Fig. 5 visually corroborate that microvasculature is more connected to the major vessels, resulting in a more continuous vascular structure.

For topological correctness, higher β_0^{err} and β_1^{err} indicate more isolated vessel segments and undesirable hole structures, respectively. Betti Matching error μ^{err}, which also considers spatial location, is more stringent than β^{err}. MaskVSC significantly outperforms other methods on β_0^{err} and β_1^{err}, except when compared to β_0^{err} of Gupta et al. [82]. Their method integrates inter-structural uncertainty by modeling topology as a sample of probability distributions and performing joint reasoning. The exception may be explained by the generation of multiple probability maps through perturbation and sampling, which contributed to topology retention but could also potentially result in false positives. While TopoMask also risks generating false positives, the proposed ViG post-processing better captures topological information by flexibly representing irregular retinal vessels as graph structures. Hence, the misclassification of background regions as vessel branches is reduced. In addition, MaskVSC is also significantly optimal for μ_0^{err} and μ_1^{err}, demonstrating its predicted vascular network contains fewer isolated segments and more desirable hole structures at spatial locations. This is also reflected in the artery-vein segmentation task, as can be seen in Fig. 5, where MaskVSC yields more contiguous results for both arteries and veins.

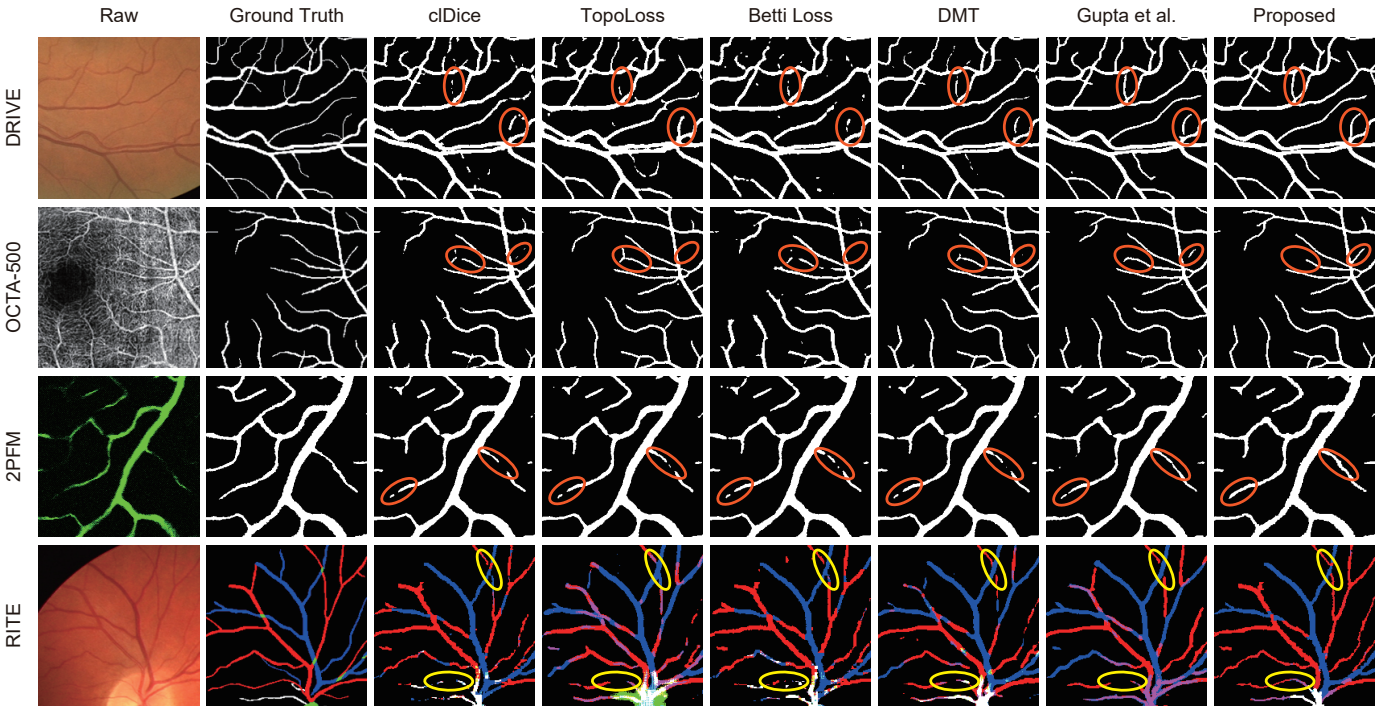

Fig. 5. Visual inspection of comparing different methods on different types of images across DRIVE, OCTA-500, 2PFM, and RITE datasets.

E. Masking Ratio Effect and Parametric Study

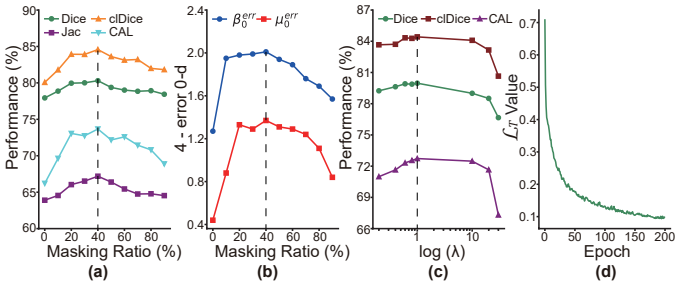

Fig. 6. (a)-(b) Results of masking ratio effect study. (c) Parametric study of $\boldsymbol{\lambda}$ in $\boldsymbol{\mathcal{L}_T}$. (d) Convergence curve for total loss $\boldsymbol{\mathcal{L}_T}$ during training on DRIVE.

We explore which masking ratio yields the best training effect. U-Net [99] with a loss function $\mathcal{L}_T$ is chosen to perform the ablation experiments on the DRIVE dataset without using ViG for post-processing.

The results depicted in Fig. 6(a-b) start with a zero masking ratio, where β_1^{err} and μ_1^{err} are omitted due to minimal changes. The inclusion of TopoMask improves all six metrics regardless of the masking ratio. A 40% masking ratio is identified as optimal, with performance increasing more abruptly from 0-20% and decreasing steadily from 40%-90%. Notably, TopoMask with a masking ratio of 90% still enhances the performance.

The optimal masking ratio of 40% is in line with other computer vision research (10% to 50%) [100]–[102], however, it is lower than MAE's 75% [55]. This discrepancy may stem from differences between natural and biomedical images, as retinal vessels sparsely occupy image space. Additionally, the masking way varies: MAE masks random patches of the original image, whereas our TopoMask masks are based on the retinal vessel topology. Given the presence of breakpoints in the original image, masking a high portion of branches may result in fewer remaining vessel segments, which potentially lower performance.

In addition, we conduct a parametric study on λ in $\mathcal{L}_T$, based on DRIVE dataset, U-Net backbone, and a 40% masking ratio (Fig. 6(c)). The optimal performance metrics are obtained when $\lambda = 1$. When λ increases from 0 to 1, performance metrics progressively increase, indicating the generation of more interconnected vascular networks. However, when $\mathcal{L}_C > \mathcal{L}_{BCE}$, the metric values of segmentation performance decrease sharply. This trend is similar across other datasets. While Jac (not shown) is not optimal with $\lambda = 1$, our primary focus is on accurate segmentation and topology preservation. Therefore, we set $\lambda = 1$ as a trade-off.

F. Performance over Different Backbones

We assess the effectiveness of MaskVSC on different DL segmentation backbones, including the CNN-based U-Net [99] and CS^2-Net [68], and the Transformer-based Swin-Unet [103]. The masking ratio of TopoMask for all three backbones is chosen to be 40%.

The results listed in Table III indicate that integrating MaskVSC with different backbones substantially enhances multiple evaluation metrics across all datasets. MaskVSC enables the primary backbones to perform more prominently in predicting the accuracy and topological integrity of the vascular structure. This notable improvement is not only evident through the segmentation similarity metrics such as $Dice$, Jac, and $clDice$ but also through the metrics assessing topological correctness, including CAL, β^{err}, and μ^{err}. These

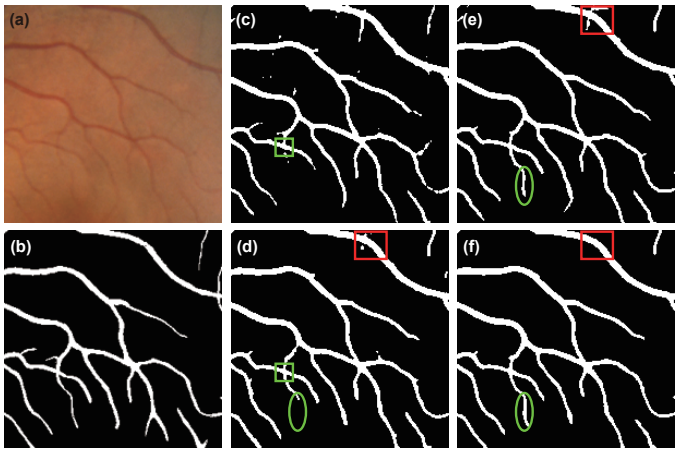

Fig. 7. Visual inspection of ablation study. (a) Raw image, (b) Ground truth, (c) U-Net, (d) U-Net + $\mathcal{L}_C$, (e) U-Net + $\mathcal{L}_C$ + TopoMask, (f) U-Net + $\mathcal{L}_C$ + TopoMask + ViG.

enhancements underscore the strong potential of MaskVSC for retinal vessel segmentation and completion, offering a robust adaptation for various DL backbones.

G. Ablation Study

We conduct an ablation study to evaluate the contribution of the different components of MaskVSC, including the TopoMask, the connectivity loss $\mathcal{L}_C$, and the ViG-based post-processing. The study is implemented on U-Net [99] and the masking ratio of TopoMask is 40%. Table IV shows the results of the ablation study on the DRIVE dataset. Both β_0^{err} and μ_0^{err} are greatly optimized after the addition of $\mathcal{L}_C$, confirming that $\mathcal{L}_C$ effectively reduces the fragmented component in the prediction and enhances the connectivity of the vascular network. This also results in increased $Dice$ and $clDice$, indicating that fewer connectivity components brought by $\mathcal{L}_C$ help in classifying the foreground and background more accurately.

Comparing $\mathcal{L}_C$ with the other topology-preserving losses (clDice [75], TopoLoss [76], BettiLoss [80]) from Table II, although $\mathcal{L}_C$ shows little difference in $Dice$ and $clDice$, it gains a significant margin on both β^{err} and μ^{err}. This demonstrates that $\mathcal{L}_C$ effectively preserves different topologies of the vasculature, bringing the prediction closer to the ground truth in terms of the 0- and 1-dimensional Betti Numbers.

TopoMask further increases $Dice$ and $clDice$, with $clDice$ growing from 80.07% to 82.61%. The 0-d Betti Number also becomes more similar to the ground truth. TopoMask randomly masks the vascular skeleton branches to mimic the missing vessel. The increased metric values suggest these masks help the model adapt to more vessel-absent scenarios on limited data, preventing it from fitting to a sub-optimal saddle point. The β_1^{err} and μ_1^{err} with only TopoMask are slightly lower compared to those with only $\mathcal{L}_C$, probably due to TopoMask causing an increase in false positives, leading to undesirable ring structures. This issue is corrected $\mathcal{L}_C$ and TopoMask are combined.

We utilize a ViG-based post-processing model to further address this issue. The retinal vascular structure's complexity benefits from ViG's non-uniform sampling, which is flexible to adapt to local variations. Additionally, ViG's capability to process nodes and edges in the graph structure allows a better understanding of the vascular network's topology, reducing the possibility of false positives. The visual inspection of Fig. 7 also supports these statements. Comparing Fig. 7(c) and Fig. 7(d), the introduction of $\mathcal{L}_C$ reduces the fragmented segmentation and enhances the continuity (highlighted in green boxes). When TopoMask is introduced (Fig. 7(e)), the model recognizes retinal vessels in low SNR regions (highlighted in green ellipse) but also leads to more false positives (highlighted in red box). This is effectively mitigated with the introduction of ViG (Fig. 7(f)).

We additionally conduct an ablation study on the biphasic setting of the masking ratio, comparing the setting applied in this paper with either only increasing or only decreasing the masking ratio. In the increase-only case, the masking ratio increases from 0 to a peak of 40%, while the decrease-only case does the opposite, with the masking ratio decreasing from 40% to 0 at the 90% of epochs. Both comparison settings leave unmasking for the first and last 10% of epochs as in this paper. The experimental results with U-Net backbone on the DRIVE dataset are listed in Table V. The proposed setting performs the best and the decrease-only case performs the worst. This may be because excessive masking at the start of training leads the model to learn the wrong vessel representation. The increase-only case has a steep drop in the masking ratio in the final "cool-down" phase, which may make the model difficult to adapt from the masked image to the real image.

H. Application in Diabetic Retinopathy Classification

To validate the significance of MaskVSC, we compare the retinal vascular tortuosity, a sensitive, vascular network-related vascular metric in ocular diseases, between healthy subjects and patients with DR. The retinal vasculature is segmented by a U-Net backbone with and without MaskVSC, and tortuosity is calculated as the ratio path length by straight-line distance between endpoints of a vessel segment. Experiments are performed on the OCTA-500 dataset (6×6 mm^2 FOV) with 91 normal and 35 DR OCTA images. The receiver operating characteristic (ROC) curve and area under the curve (AUC) for both methods are displayed in Fig. 8(c), and the examples of healthy and DR OCTA images with overlaid segmentation with/without MaskVSC are in Fig. 8(a-b), respectively. MaskVSC predicts more interconnected vascular patterns that the small vessels are more connected to the main vascular branches. Moreover, MaskVSC helps to improve the discriminative power of vascular tortuosity in classifying DR (AUC: 0.81 vs. 0.71; p = 0.0287, *DeLong test*). It indicates that MaskVSC could provide insights into topological information on the retinal vasculature in DR and other related ocular diseases.

V. Conclusion

In this paper, we propose a backbone-agnostic vascular structure segmentation and completion method called MaskVSC. Our method includes a topology-aware mask strategy (TopoMask) that simulates vessel breakpoints in retinal

TABLE III
RESULTS WITH THREE DIFFERENT BACKBONES ON SEVEN DIFFERENT RETINAL DATASETS.

Dataset	Method	Dice (%)↑	Jac (%)↑	clDice (%)↑	CAL (%)↑	β_0^{err} ↓	β_1^{err} ↓	μ_0^{err} ↓	μ_1^{err} ↓
DRIVE	U-Net [99]	75.93 ± 1.49	60.17 ± 3.80	78.06 ± 3.69	66.95 ± 5.20	10.3	0.306	10.8	0.409
	U-Net [99] + Proposed	**80.30 ± 1.07**	**67.19 ± 3.00**	**84.57 ± 3.28**	**73.67 ± 5.16**	**1.99**	**0.254**	**2.63**	**0.347**
	CS^2-Net [68]	76.83 ± 1.20	62.12 ± 3.09	79.17 ± 3.69	63.73 ± 5.61	2.90	0.281	3.68	0.331
	CS^2-Net [68] + Proposed	**79.94 ± 1.08**	**66.63 ± 2.94**	**85.21 ± 3.33**	**73.28 ± 5.49**	**1.72**	**0.254**	**2.31**	**0.311**
	Swin-Unet [103]	73.98 ± 1.89	58.79 ± 5.19	76.05 ± 4.82	63.39 ± 6.94	6.53	0.297	7.36	0.417
	Swin-Unet [103] + Proposed	**77.08 ± 1.24**	**62.67 ± 3.60**	**80.38 ± 3.85**	**68.00 ± 6.01**	**3.36**	**0.277**	**4.18**	**0.377**
STARE	U-Net [99]	77.83 ± 2.04	63.77 ± 5.47	82.42 ± 5.78	70.48 ± 8.78	3.65	0.111	4.04	0.103
	U-Net [99] + Proposed	**79.95 ± 1.74**	**66.15 ± 5.79**	**85.01 ± 5.33**	**73.29 ± 8.63**	**2.04**	**0.094**	**2.33**	**0.100**
	CS^2-Net [68]	78.19 ± 1.48	62.81 ± 5.71	83.53 ± 4.10	68.91 ± 6.65	1.87	0.0906	2.24	0.0969
	CS^2-Net [68] + Proposed	**79.35 ± 2.17**	**65.27 ± 6.76**	**84.98 ± 5.81**	**71.27 ± 6.23**	*1.73*	**0.0844**	**2.05**	**0.0906**
	Swin-Unet [103]	76.37 ± 2.09	59.39 ± 8.90	79.53 ± 6.39	66.38 ± 10.3	4.73	0.0875	5.09	0.119
	Swin-Unet [103] + Proposed	**78.43 ± 1.71**	**62.67 ± 7.20**	**83.13 ± 4.64**	**70.73 ± 9.13**	**2.85**	**0.0813**	**3.32**	**0.100**
ROSE	U-Net [99]	72.91 ± 0.75	57.35 ± 4.02	71.04 ± 4.00	56.26 ± 5.10	6.36	0.204	13.4	0.204
	U-Net [99] + Proposed	**75.25 ± 0.87**	**60.01 ± 4.39**	**73.74 ± 4.36**	**58.41 ± 6.17**	**3.67**	**0.169**	**9.20**	**0.178**
	CS^2-Net [68]	74.26 ± 1.09	58.91 ± 4.02	74.89 ± 3.23	58.36 ± 4.58	4.49	0.169	10.2	0.178
	CS^2-Net [68] + Proposed	**75.65 ± 0.81**	**60.73 ± 4.00**	**75.69 ± 3.65**	**59.18 ± 5.46**	**3.16**	**0.142**	**8.42**	**0.142**
	Swin-Unet [103]	71.23 ± 1.01	55.10 ± 4.25	74.28 ± 3.90	60.20 ± 4.27	10.2	0.174	13.0	0.194
	Swin-Unet [103] + Proposed	**72.17 ± 0.81**	**56.44 ± 3.88**	**75.62 ± 3.76**	**61.05 ± 4.57**	**6.44**	**0.146**	**9.38**	**0.188**
OCTA-500	U-Net [99]	74.58 ± 0.74	60.67 ± 4.87	85.69 ± 5.14	72.12 ± 9.32	2.17	0.0791	2.52	0.0891
	U-Net [99] + Proposed	**76.73 ± 0.75**	**63.36 ± 4.96**	**87.94 ± 5.04**	**74.59 ± 9.17**	**1.05**	**0.0781**	**1.38**	**0.0853**
	CS^2-Net [68]	76.30 ± 0.77	63.04 ± 6.55	83.87 ± 5.85	66.02 ± 9.91	2.09	0.0992	2.62	0.111
	CS^2-Net [68] + Proposed	**77.09 ± 0.94**	**64.89 ± 6.77**	**84.56 ± 5.78**	**69.15 ± 9.92**	**1.81**	**0.0911**	**2.30**	**0.096**
	Swin-Unet [103]	73.55 ± 1.04	60.90 ± 6.95	83.30 ± 6.77	73.61 ± 10.3	7.18	0.298	9.14	0.374
	Swin-Unet [103] + Proposed	**76.74 ± 0.93**	**64.49 ± 6.90**	**84.12 ± 5.93**	*73.78 ± 9.19*	**4.28**	**0.269**	**6.11**	**0.334**
2PFM	U-Net [99]	84.94 ± 1.65	68.64 ± 4.38	83.21 ± 5.61	75.82 ± 9.67	5.57	0.133	8.65	0.160
	U-Net [99] + Proposed	**85.80 ± 1.68**	**69.71 ± 3.95**	**85.21 ± 4.69**	**77.82 ± 8.90**	**1.63**	**0.127**	**2.25**	**0.135**
	CS^2-Net [68]	85.11 ± 1.74	67.96 ± 7.01	83.39 ± 5.48	74.44 ± 9.70	1.92	0.126	2.64	0.148
	CS^2-Net [68] + Proposed	**86.52 ± 1.76**	**68.71 ± 5.73**	**84.40 ± 3.54**	**75.82 ± 8.28**	**1.71**	**0.120**	**2.37**	**0.138**
	Swin-Unet [103]	81.47 ± 1.93	64.45 ± 6.27	81.82 ± 5.68	78.26 ± 9.42	4.56	0.134	9.88	0.175
	Swin-Unet [103] + Proposed	**84.91 ± 1.73**	**69.53 ± 4.44**	**85.35 ± 3.06**	**83.39 ± 8.69**	**1.48**	**0.127**	**2.26**	**0.145**
RITE	U-Net [99]	56.22 ± 1.08	37.12 ± 3.99	57.07 ± 3.97	28.21 ± 4.02	10.1	0.282	65.5	2.544
	U-Net [99] + Proposed	**59.77 ± 0.82**	**41.78 ± 3.91**	**60.18 ± 2.80**	**32.68 ± 3.19**	**3.16**	**0.140**	**30.9**	**0.991**
	CS^2-Net [68]	56.80 ± 0.68	38.82 ± 2.85	58.06 ± 2.67	29.40 ± 2.98	4.70	0.320	67.6	2.43
	CS^2-Net [68] + Proposed	**58.41 ± 0.66**	**41.10 ± 2.49**	**60.83 ± 3.32**	**33.72 ± 3.88**	**2.19**	**0.151**	**25.5**	**0.97**
	Swin-Unet [103]	58.09 ± 0.87	40.40 ± 2.70	58.22 ± 2.87	31.34 ± 3.25	6.58	0.245	85.3	2.52
	Swin-Unet [103] + Proposed	**60.54 ± 0.92**	**42.31 ± 3.08**	**61.01 ± 2.18**	**35.27 ± 2.99**	**2.09**	**0.111**	**24.2**	**0.94**
LES-AV	U-Net [99]	58.06 ± 1.45	44.37 ± 6.12	64.69 ± 6.07	32.01 ± 7.42	1.40	0.0333	12.9	0.165
	U-Net [99] + Proposed	**63.32 ± 1.84**	**47.19 ± 7.77**	**66.17 ± 6.35**	**35.19 ± 8.07**	**1.05**	**0.0224**	**5.19**	**0.111**
	CS^2-Net [68]	59.15 ± 3.35	44.81 ± 5.43	65.88 ± 4.74	32.00 ± 6.09	3.44	0.162	28.0	1.02
	CS^2-Net [68] + Proposed	**61.47 ± 4.08**	**46.97 ± 6.86**	**68.24 ± 4.06**	**34.58 ± 8.10**	**1.61**	**0.085**	**10.8**	**0.76**
	Swin-Unet [103]	67.23 ± 1.44	49.94 ± 6.12	70.09 ± 6.28	39.91 ± 8.18	1.75	0.0624	16.3	0.412
	Swin-Unet [103] + Proposed	**68.07 ± 1.57**	**52.15 ± 5.56**	**72.11 ± 5.45**	**41.38 ± 9.07**	**1.19**	**0.0333**	**6.44**	**0.247**

TABLE IV
RESULTS OF ABLATION STUDY ON THE DRIVE DATASET. ✓ STANDS FOR "INCLUDED" AND "T-M" MEANS TOPOMASK.

$\mathcal{L}_C$	T-M	ViG	Dice (%)↑	clDice (%)↑	β_0^{err} ↓	β_1^{err} ↓	μ_0^{err} ↓	μ_1^{err} ↓
			75.93	78.06	10.3	0.306	10.8	0.409
✓			77.94	80.07	2.73	0.265	3.56	0.365
	✓		78.59	82.61	2.70	0.273	2.81	0.399
✓	✓		78.97	83.42	2.14	0.263	2.69	0.357
✓	✓	✓	80.30	84.57	1.99	0.254	2.63	0.348

TABLE V
RESULTS OF ABLATION STUDY FOR MASKING RATIO SETTING ON THE DRIVE DATASET.

Setting	Dice (%)↑	clDice (%)↑	β_0^{err} ↓	β_1^{err} ↓	μ_0^{err} ↓	μ_1^{err} ↓
Increase-only	79.12	83.80	2.12	0.266	2.72	0.352
Decrease-only	78.45	82.55	2.60	0.271	2.97	0.378
Proposed	80.30	84.57	1.99	0.254	2.63	0.348

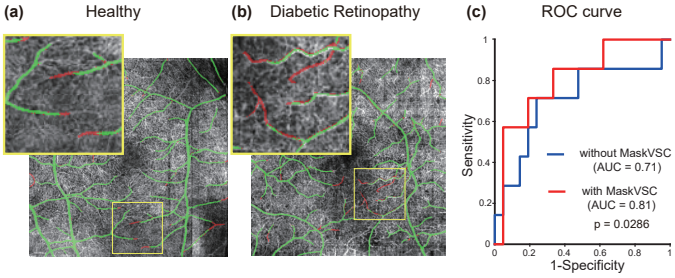

Fig. 8. (a) Healthy OCTA image, (b) DR OCTA image, (c) ROC curve of DR Diagnose. Green represents the intersection of with and without MaskVSC. Red represents the difference set of with MaskVSC with respect to without MaskVSC.

images. Additionally, a connectivity loss $\mathcal{L}_C$ guides the generation of a more connected prediction, and ViG-based post-processing further reduces false positives from TopoMask by performing graphical structural analysis to refine the vascular structure. We conduct comparison experiments on five datasets containing three modalities, and MaskVSC achieves superior segmentation performance and topology preservation compared to other state-of-the-art methods. This method has potential applications in other domains for extracting more continuous and interconnected linear or curvilinear structures, facilitating further studies on microvascular topology.

Lastly, we also notice several aspects that need improvement. First, despite our comprehensive evaluation, the data tested in this study may not fully capture the diversity of diseases. Second, our approach is designed for 2D retinal vessels, but vascular structures are inherently 3D. Developing a framework capable of segmenting and complementing 3D

vascular structures is an important direction for future work.

REFERENCES

[1] L. Ding, M. H. Bawany, A. E. Kuriyan, R. S. Ramchandran, C. C. Wykoff, and G. Sharma, "Recovery-fa19: Ultra-widefield fluorescein angiography vessel detection dataset," 2019. [Online]. Available: https://dx.doi.org/10.21227/m9yw-xs04

[2] A. P. Cherecheanu, G. Garhofer, D. Schmidl, R. Werkmeister, and L. Schmetterer, "Ocular perfusion pressure and ocular blood flow in glaucoma," *Current Opinion in Pharmacology*, vol. 13, no. 1, pp. 36–42, 2013.

[3] K. Y. Tey, K. Teo, A. C. Tan, K. Devarajan, B. Tan, J. Tan, L. Schmetterer, and M. Ang, "Optical coherence tomography angiography in diabetic retinopathy: a review of current applications," *Eye and Vision*, vol. 6, pp. 1–10, 2019.

[4] Y. Muraoka, A. Tsujikawa, T. Murakami, K. Ogino, K. Kumagai, K. Miyamoto, A. Uji, and N. Yoshimura, "Morphologic and functional changes in retinal vessels associated with branch retinal vein occlusion," *Ophthalmology*, vol. 120, no. 1, pp. 91–99, 2013.

[5] B. Pemp and L. Schmetterer, "Ocular blood flow in diabetes and age-related macular degeneration," *Canadian Journal of Ophthalmology*, vol. 43, no. 3, pp. 295–301, 2008.

[6] M. Ang, A. C. Tan, C. M. G. Cheung, P. A. Keane, R. Dolz-Marco, C. C. Sng, and L. Schmetterer, "Optical coherence tomography angiography: a review of current and future clinical applications," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 256, pp. 237–245, 2018.

[7] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2538–2548, 2012.

[8] S. Holm, G. Russell, V. Nourrit, and N. McLoughlin, "Dr hagis—a fundus image database for the automatic extraction of retinal surface vessels from diabetic patients," *Journal of Medical Imaging*, vol. 4, no. 1, pp. 014 503–014 503, 2017.

[9] J. Odstrcilik, R. Kolar, A. Budai, J. Hornegger, J. Jan, J. Gazarek, T. Kubena, P. Cernosek, O. Svoboda, and E. Angelopoulou, "Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database," *IET Image Processing*, vol. 7, no. 4, pp. 373–383, 2013.

[10] K. Jin, X. Huang, J. Zhou, Y. Li, Y. Yan, Y. Sun, Q. Zhang, Y. Wang, and J. Ye, "Fives: A fundus image dataset for artificial intelligence based vessel segmentation," *Scientific Data*, vol. 9, no. 1, p. 475, 2022.

[11] Q. Hu, M. D. Abràmoff, and M. K. Garvin, "Automated separation of binary overlapping trees in low-contrast color retinal images," in *MICCAI*. Springer, 2013, pp. 436–443.

[12] X. Lyu, L. Cheng, and S. Zhang, "The reta benchmark for retinal vascular tree analysis," *Scientific Data*, vol. 9, no. 1, p. 397, 2022.

[13] R. Hemelings, B. Elen, I. Stalmans, K. Van Keer, P. De Boever, and M. B. Blaschko, "Artery–vein segmentation in fundus images using a fully convolutional network," *Computerized Medical Imaging and Graphics*, vol. 76, p. 101636, 2019.

[14] J. I. Orlando, J. Barbosa Breda, K. Van Keer, M. B. Blaschko, P. J. Blanco, and C. A. Bulant, "Towards a glaucoma risk index based on simulated hemodynamics from fundus images," in *MICCAI*. Springer, 2018, pp. 65–73.

[15] J. Van Eijgen, J. Fhima, M.-I. Billen Moulin-Romsée, J. A. Behar, E. Christinaki, and I. Stalmans, "Leuven-haifa high-resolution fundus image dataset for retinal blood vessel segmentation and glaucoma diagnosis," *Scientific Data*, vol. 11, no. 1, p. 257, 2024.

[16] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical Imaging*, vol. 19, no. 3, pp. 203–210, 2000.

[17] Y. Zhao, L. Rada, K. Chen, S. P. Harding, and Y. Zheng, "Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images," *IEEE Transactions on Medical Imaging*, vol. 34, no. 9, pp. 1797–1807, 2015.

[18] J. Zhang, B. Dashtbozorg, E. Bekkers, J. P. Pluim, R. Duits, and B. M. ter Haar Romeny, "Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores," *IEEE Transactions on Medical Imaging*, vol. 35, no. 12, pp. 2631–2644, 2016.

[19] F. Zana and J.-C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1010–1019, 2001.

[20] A. M. Mendonca and A. Campilho, "Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction," *IEEE Transactions on Medical Imaging*, vol. 25, no. 9, pp. 1200–1213, 2006.

[21] M. S. Miri and A. Mahloojifar, "Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 5, pp. 1183–1192, 2010.

[22] F. K. Quek and C. Kirbas, "Vessel extraction in medical images by wave-propagation and traceback," *IEEE Transactions on Medical Imaging*, vol. 20, no. 2, pp. 117–131, 2001.

[23] M. Sofka and C. V. Stewart, "Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures," *IEEE Transactions on Medical Imaging*, vol. 25, no. 12, pp. 1531–1546, 2006.

[24] K. K. Delibasis, A. I. Kechriniotis, C. Tsonos, and N. Assimakis, "Automatic model-based tracing algorithm for vessel segmentation and diameter estimation," *Computer Methods and Programs in Biomedicine*, vol. 100, no. 2, pp. 108–122, 2010.

[25] O. Wink, W. J. Niessen, and M. A. Viergever, "Multiscale vessel tracking," *IEEE Transactions on Medical Imaging*, vol. 23, no. 1, pp. 130–133, 2004.

[26] M. E. Martinez-Perez, A. D. Hughes, S. A. Thom, A. A. Bharath, and K. H. Parker, "Segmentation of blood vessels from red-free and fluorescein retinal images," *Medical Image Analysis*, vol. 11, no. 1, pp. 47–61, 2007.

[27] M. Vlachos and E. Dermatas, "Multi-scale retinal vessel segmentation using line tracking," *Computerized Medical Imaging and Graphics*, vol. 34, no. 3, pp. 213–227, 2010.

[28] K. Hu, Z. Zhang, X. Niu, Y. Zhang, C. Cao, F. Xiao, and X. Gao, "Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function," *Neurocomputing*, vol. 309, pp. 179–191, 2018.

[29] Ç. Sazak, C. J. Nelson, and B. Obara, "The multiscale bowler-hat transform for blood vessel enhancement in retinal images," *Pattern Recognition*, vol. 88, pp. 739–750, 2019.

[30] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[31] J. V. Soares, J. J. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification," *IEEE Transactions on Medical Imaging*, vol. 25, no. 9, pp. 1214–1222, 2006.

[32] E. Ricci and R. Perfetti, "Retinal blood vessel segmentation using line operators and support vector classification," *IEEE Transactions on Medical Imaging*, vol. 26, no. 10, pp. 1357–1365, 2007.

[33] C. A. Lupascu, D. Tegolo, and E. Trucco, "Fabc: retinal vessel segmentation using adaboost," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1267–1274, 2010.

[34] D. Marín, A. Aquino, M. E. Gegúndez-Arias, and J. M. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features," *IEEE Transactions on Medical Imaging*, vol. 30, no. 1, pp. 146–158, 2010.

[35] J. I. Orlando, E. Prokofyeva, and M. B. Blaschko, "A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 16–27, 2016.

[36] M. R. K. Mookiah, S. Hogg, T. J. MacGillivray, V. Prathiba, R. Pradeepa, V. Mohan, R. M. Anjana, A. S. Doney, C. N. Palmer, and E. Trucco, "A review of machine learning methods for retinal blood vessel segmentation and artery/vein classification," *Medical Image Analysis*, vol. 68, p. 101905, 2021.

[37] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, "Deep retinal image understanding," in *MICCAI*, 2016, pp. 140–148.

[38] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 11, pp. 2369–2380, 2016.

[39] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "Dunet: A deformable network for retinal vessel segmentation," *Knowledge-Based Systems*, vol. 178, pp. 149–162, 2019.

[40] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A cross-modality learning approach for vessel segmentation in retinal images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 109–118, 2015.

[41] A. Oliveira, S. Pereira, and C. A. Silva, "Retinal vessel segmentation based on fully convolutional neural networks," *Expert Systems with Applications*, vol. 112, pp. 229–242, 2018.

[42] H. Wu, W. Wang, J. Zhong, B. Lei, Z. Wen, and J. Qin, "Scs-net: A scale and context sensitive network for retinal vessel segmentation," *Medical Image Analysis*, vol. 70, p. 102025, 2021.

[43] Y. Ma, H. Hao, J. Xie, H. Fu, J. Zhang, J. Yang, Z. Wang, J. Liu, Y. Zheng, and Y. Zhao, "Rose: a retinal oct-angiography vessel segmentation dataset and new model," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 928–939, 2020.

[44] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[45] H. Fu, Y. Xu, S. Lin, D. W. Kee Wong, and J. Liu, "Deepvessel: Retinal vessel segmentation via deep learning and conditional random field," in *MICCAI*, 2016, pp. 132–139.

[46] Y. Li, Y. Zhang, W. Cui, B. Lei, X. Kuang, and T. Zhang, "Dual encoder-based dynamic-channel graph convolutional network with edge enhancement for retinal vessel segmentation," *IEEE Transactions on Medical Imaging*, vol. 41, no. 8, pp. 1975–1989, 2022.

[47] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *NeurIPS*, vol. 30, 2017.

[48] X. Li, Y. Jiang, M. Li, and S. Yin, "Lightweight attention convolutional neural network for retinal vessel image segmentation," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1958–1967, 2020.

[49] X. Tan, X. Chen, Q. Meng, F. Shi, D. Xiang, Z. Chen, L. Pan, and W. Zhu, "Oct2former: A retinal oct-angiography vessel segmentation transformer," *Computer Methods and Programs in Biomedicine*, vol. 233, p. 107454, 2023.

[50] M. Gori, G. Monfardini, and F. Scarselli, "A new model for learning in graph domains," in *IJCNN*, vol. 2, 2005, pp. 729–734.

[51] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *ICLR*, 2016.

[52] S. Y. Shin, S. Lee, I. D. Yun, and K. M. Lee, "Deep vessel segmentation by learning graphical connectivity," *Medical Image Analysis*, vol. 58, p. 101556, 2019.

[53] Y. Meng, H. Zhang, Y. Zhao, X. Yang, Y. Qiao, I. J. MacCormick, X. Huang, and Y. Zheng, "Graph-based region and boundary aggregation for biomedical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 41, no. 3, pp. 690–701, 2021.

[54] G. Zhao, K. Liang, C. Pan, F. Zhang, X. Wu, X. Hu, and Y. Yu, "Graph convolution based cross-network multiscale feature fusion for deep vessel segmentation," *IEEE Transactions on Medical Imaging*, vol. 42, no. 1, pp. 183–195, 2022.

[55] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *CVPR*, 2022, pp. 16 000–16 009.

[56] P. Bibiloni, M. González-Hidalgo, and S. Massanet, "A survey on curvilinear object segmentation in multiple applications," *Pattern Recognition*, vol. 60, pp. 949–970, 2016.

[57] D.-S. Huang, W. Jia, and D. Zhang, "Palmprint verification based on principal lines," *Pattern Recognition*, vol. 41, no. 4, pp. 1316–1328, 2008.

[58] X. Qian, M. P. Brennan, D. P. Dione, W. L. Dobrucki, M. P. Jackowski, C. K. Breuer, A. J. Sinusas, and X. Papademetris, "A non-parametric vessel detection method for complex vascular structures," *Medical Image Analysis*, vol. 13, no. 1, pp. 49–61, 2009.

[59] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable cosfire filters for vessel delineation with application to retinal images," *Medical Image Analysis*, vol. 19, no. 1, pp. 46–57, 2015.

[60] J. A. Tyrrell, E. di Tomaso, D. Fuja, R. Tong, K. Kozak, R. K. Jain, and B. Roysam, "Robust 3-d modeling of vasculature imagery using superellipsoids," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 223–237, 2007.

[61] V. Bismuth, R. Vaillant, H. Talbot, and L. Najman, "Curvilinear structure enhancement with the polygonal path image-application to guide-wire segmentation in x-ray fluoroscopy," in *MICCAI*, 2012, pp. 9–16.

[62] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "Iterative vessel segmentation of fundus images," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 7, pp. 1738–1749, 2015.

[63] H. Zhao, J. Kumagai, M. Nakagawa, and R. Shibasaki, "Semi-automatic road extraction from high-resolution satellite image," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives*, vol. 34, 2002.

[64] Y. Wu, Y. Xia, Y. Song, Y. Zhang, and W. Cai, "Multiscale network followed network model for retinal vessel segmentation," in *MICCAI*, 2018, pp. 119–126.

[65] L. Lin, Z. Wang, J. Wu, Y. Huang, J. Lyu, P. Cheng, J. Wu, and X. Tang, "Bsda-net: A boundary shape and distance aware joint learning framework for segmenting and classifying octa images," in *MICCAI*, 2021, pp. 65–75.

[66] J. Hao, T. Shen, X. Zhu, Y. Liu, A. Behera, D. Zhang, B. Chen, J. Liu, J. Zhang, and Y. Zhao, "Retinal structure detection in octa image via voting-based multitask learning," *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3969–3980, 2022.

[67] L. Mou, Y. Zhao, L. Chen, J. Cheng, Z. Gu, H. Hao, H. Qi, Y. Zheng, A. Frangi, and J. Liu, "Cs-net: Channel and spatial attention network for curvilinear structure segmentation," in *MICCAI*, 2019, pp. 721–730.

[68] L. Mou, Y. Zhao, H. Fu, Y. Liu, J. Cheng, Y. Zheng, P. Su, J. Yang, L. Chen, A. F. Frangi *et al.*, "Cs2-net: Deep learning segmentation of curvilinear structures in medical imaging," *Medical Image Analysis*, vol. 67, p. 101874, 2021.

[69] L. Lin, L. Peng, H. He, P. Cheng, J. Wu, K. K. Wong, and X. Tang, "Yolocurvseg: You only label one noisy skeleton for vessel-style curvilinear structure segmentation," *Medical Image Analysis*, vol. 90, p. 102937, 2023.

[70] X. Xu, M. C. Nguyen, Y. Yazici, K. Lu, H. Min, and C.-S. Foo, "Semicurv: Semi-supervised curvilinear structure segmentation," *IEEE Transactions on Image Processing*, vol. 31, pp. 5109–5120, 2022.

[71] T. Shi, X. Ding, L. Zhang, and X. Yang, "Freecos: self-supervised learning from fractals and unlabeled images for curvilinear object segmentation," in *ICCV*, 2023, pp. 876–886.

[72] T. Wang and Q. Dai, "Survs: A swin-unet and game theory-based unsupervised segmentation method for retinal vessel," *Computers in Biology and Medicine*, vol. 166, p. 107542, 2023.

[73] S. Gur, L. Wolf, L. Golgher, and P. Blinder, "Unsupervised microvascular image segmentation using an active contours mimicking neural network," in *ICCV*, 2019, pp. 10 722–10 731.

[74] S. LeeBlaire, A. BoasDavid *et al.*, "Anatomical modeling of brain vasculature in two-photon microscopy by generalizable deep learning," *BME Frontiers*, 2020.

[75] S. Shit, J. C. Paetzold, A. Sekuboyina, I. Ezhov, A. Unger, A. Zhylka, J. P. W. Pluim, U. Bauer, and B. H. Menze, "cldice - a novel topology-preserving loss function for tubular structure segmentation," in *CVPR*, June 2021, pp. 16 560–16 569.

[76] X. Hu, F. Li, D. Samaras, and C. Chen, "Topology-preserving deep image segmentation," *NeurIPS*, vol. 32, 2019.

[77] R. Forman, "Morse theory for cell complexes," *Advances in Mathematics*, vol. 134, no. 1, pp. 90–145, 1998.

[78] X. Hu, Y. Wang, L. Fuxin, D. Samaras, and C. Chen, "Topology-aware segmentation using discrete morse theory," in *ICLR*, 2021.

[79] X. Hu, D. Samaras, and C. Chen, "Learning probabilistic topological representations using discrete morse theory," in *ICLR*, 2023.

[80] N. Stucki, J. C. Paetzold, S. Shit, B. Menze, and U. Bauer, "Topologically faithful image segmentation via induced matching of persistence barcodes," in *ICML*, 2023, pp. 32 698–32 727.

[81] S. Gupta, X. Hu, J. Kaan, M. Jin, M. Mpoy, K. Chung, G. Singh, M. Saltz, T. Kurc, J. Saltz *et al.*, "Learning topological interactions for multi-class medical image segmentation," in *ECCV*, 2022, pp. 701–718.

[82] S. Gupta, Y. Zhang, X. Hu, P. Prasanna, and C. Chen, "Topology-aware uncertainty for image segmentation," *NeurIPS*, vol. 36, 2024.

[83] H. Fu, Y. Xu, D. W. K. Wong, and J. Liu, "Retinal vessel segmentation via deep learning network and fully-connected conditional random fields," in *ISBI*, 2016, pp. 698–701.

[84] J. Zhang, E. Bekkers, D. Chen, T. T. Berendschot, J. Schouten, J. P. Pluim, Y. Shi, B. Dashtbozorg, and B. M. ter Haar Romeny, "Reconnection of interrupted curvilinear structures via cortically inspired completion for ophthalmologic images," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 5, pp. 1151–1165, 2018.

[85] H. Du, X. Zhang, G. Song, F. Bao, Y. Zhang, W. Wu, and P. Liu, "Retinal blood vessel segmentation by using the ms-lsdnet network and geometric skeleton reconnection method," *Computers in Biology and Medicine*, vol. 153, p. 106416, 2023.

[86] R. Damseh, P. Pouliot, L. Gagnon, S. Sakadzic, D. Boas, F. Cheriet, and F. Lesage, "Automatic graph-based modeling of brain microvessels captured with two-photon microscopy," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2551–2562, 2018.

[87] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE*

Transactions on Biomedical Engineering, vol. 59, no. 9, pp. 2538–2548, 2012.

[88] L. Mou, L. Chen, J. Cheng, Z. Gu, Y. Zhao, and J. Liu, "Dense dilated network with probability regularized walk for vessel detection," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1392–1403, 2019.

[89] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[90] Z. Yan, X. Yang, and K.-T. Cheng, "Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1912–1923, 2018.

[91] ——, "A three-stage deep learning model for accurate retinal vessel segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1427–1436, 2018.

[92] Y. Wu, Y. Xia, Y. Song, Y. Zhang, and W. Cai, "Nfn+: A novel network followed network for retinal vessel segmentation," *Neural Networks*, vol. 126, pp. 153–162, 2020.

[93] S. Carneiro-Esteves, A. Vacavant, and O. Merveille, "Restoring connectivity in vascular segmentations using a learned post-processing model," in *International Workshop on Topology-and Graph-Informed Imaging Informatics*. Springer, 2024, pp. 55–65.

[94] ——, "A plug-and-play framework for curvilinear structure segmentation based on a learned reconnecting regularization," *Neurocomputing*, p. 128055, 2024.

[95] K. Han, Y. Wang, J. Guo, Y. Tang, and E. Wu, "Vision gnn: An image is worth graph of nodes," *NeurIPS*, vol. 35, pp. 8291–8303, 2022.

[96] M. E. Gegúndez-Arias, A. Aquino, J. M. Bravo, and D. Marín, "A function for quality evaluation of retinal vessel segmentations," *IEEE Transactions on Medical Imaging*, vol. 31, no. 2, pp. 231–239, 2011.

[97] P. Singh, L. Chen, M. Chen, J. Pan, R. Chukkapalli, S. Chaudhari, and J. Cirrone, "Enhancing medical image segmentation: Optimizing cross-entropy weights and post-processing with autoencoders," in *ICCV*, 2023, pp. 2684–2693.

[98] M. Li, K. Huang, Q. Xu, J. Yang, Y. Zhang, K. Xie, S. Yuan, Q. Liu, and Q. Chen, "Octa-500: a retinal dataset for optical coherence tomography angiography study," *Medical Image Analysis*, p. 103092, 2024.

[99] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.

[100] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *ICLR*, 2021.

[101] H. Bao, L. Dong, S. Piao, and F. Wei, "Beit: Bert pre-training of image transformers," *ICLR*, 2021.

[102] L. Zhou, H. Liu, J. Bae, J. He, D. Samaras, and P. Prasanna, "Self pre-training with masked autoencoders for medical image classification and segmentation," in *ISBI*. IEEE, 2023, pp. 1–6.

[103] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *ECCV*, 2022, pp. 205–218.