This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3312338

Expert-guided Knowledge Distillation for Semi-supervised Vessel Segmentation

Ning Shen, Tingfa Xu, Shiqi Huang, Feng Mu, and Jianan Li

***Abstract*— In medical image analysis, blood vessel segmentation is of considerable clinical value for diagnosis and surgery. The predicaments of complex vascular structures obstruct the development of the field. Despite many algorithms have emerged to get off the tight corners, they rely excessively on careful annotations for tubular vessel extraction. A practical solution is to excavate the feature information distribution from unlabeled data. This work proposes a novel semi-supervised vessel segmentation framework, named EXP-Net, to navigate through finite annotations. Based on the training mechanism of the Mean Teacher model, we innovatively engage an expert network in EXP-Net to enhance knowledge distillation. The expert network comprises knowledge and connectivity enhancement modules, which are respectively in charge of modeling feature relationships from global and detailed perspectives. In particular, the knowledge enhancement module leverages the vision transformer to highlight the long-range dependencies among multi-level token components; the connectivity enhancement module maximizes the properties of topology and geometry by skeletonizing the vessel in a non-parametric manner. The key components are dedicated to the conditions of weak vessel connectivity and poor pixel contrast. Extensive evaluations show that our EXP-Net achieves state-of-the-art performance on subcutaneous vessel, retinal vessel, and coronary artery segmentations. Code is available at `https://github.com/shennbit/EXP-Net`.**

***Index Terms*— Semi-supervised learning, Vessel segmentation, Knowledge distillation.**

I. Introduction

BLOOD vessel segmentation is a critical field in medical image processing that has significantly impacted numerous clinical interventions and medical treatments. Technical advancements in blood vessel segmentation have enabled clinicians to enhance diagnosis, treatment planning, and execution workflows. Among the various blood vessels in the human body, the venous and retinal vessels are of particular importance and require enhanced visualization and extraction. Accurate identification of the location of subcutaneous venous vessels is crucial for clinicians to perform venipuncture procedures precisely [1]. Additionally, observing and analyzing retinal vessels can help physicians identify signs of various cardiovascular problems and systemic diseases to prevent deterioration [2]. However, both subcutaneous and retinal vessel segmentation encounter common challenges, such as complex vascular structure, poor vessel contrast, non-uniform imaging brightness, and background interference, as illustrated in Fig. 1.

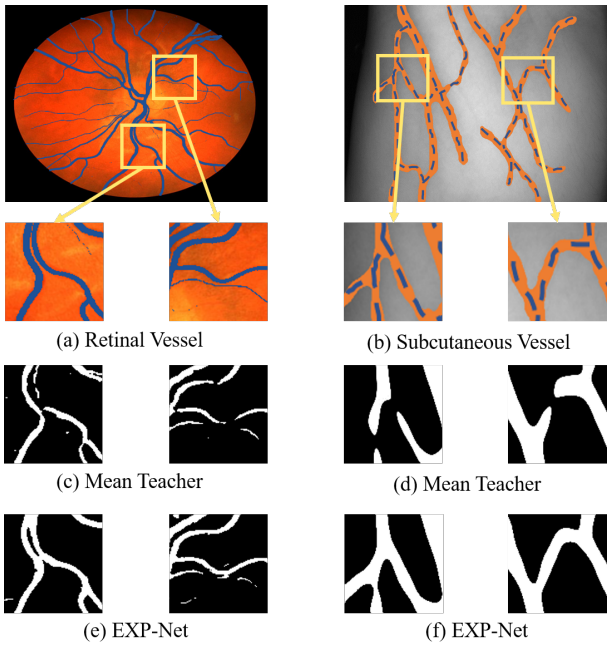

Fig. 1. Typical patterns of the blood vessel. Retinal vessel images often contain numerous tiny blood vessels and complex vascular structures. Additionally, the center of subcutaneous vessels is typically located at local pixel minima with poor contrast, further complicating segmentation. In this context, the proposed EXP-Net plays a critical role in extracting both global and detailed features from limited labeled data, making it particularly valuable for clinical applications.

In the past decade, numerous approaches have emerged for blood vessel segmentation, which can be broadly classified into three categories: multi-scale filtering [3], region growing [4], and active contour models [5]. These techniques draw inspiration from traditional image processing algorithms. More recently, follow-up methods have incorporated innovations from Convolutional Neural Networks (CNNs), which can handle more complex data distributions and extract more powerful features. However, these CNN-based models require extensive labeled data with pixel-level annotations, which are both labor-intensive and time-consuming to obtain.

Ning Shen, Tingfa Xu, Shiqi Huang, Feng Mu, and Jianan Li are with School of Optics and Photonics, Beijing Institute of Technology, Beijing, China (email: shennbit@163.com, ciom_xtf1, lijianan@bit.edu.cn).

Jianan Li is also with the Key Laboratory of Photoelectronic Imaging Technology and System, Ministry of Education of China, Beijing, China.

Tingfa Xu is also with the Key Laboratory of Photoelectronic Imaging Technology and System, Ministry of Education of China, Beijing, China, and with the Chongqing Innovation Center, Beijing Institute of Technology, Chongqing, China.

Corresponding authors: Tingfa Xu and Jianan Li.

Authorized licensed use limited to: Wuhan University. Downloaded on October 12,2023 at 11:03:31 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3312338

Semi-supervised learning is a useful technique for alleviating the stress of limited labeled data by leveraging valuable information from unannotated data to optimize model parameters. In contrast to numerous semi-supervised semantic segmentation algorithms, Mean Teacher [6] stands out by employing knowledge distillation between student and teacher networks. The feedback mechanism in Mean Teacher utilizes a weight-averaged model to enhance target consistency, leading to improved segmentation performance. Another challenge faced by CNN models is their tendency to restrict the relationship between long-range feature information due to the intrinsic locality of convolution operations. The transformer [7] can overcome this limitation by modeling long-range dependencies among token features. Additionally, the complex curve interconnectivity and bifurcation structure of tubular veins severely hinder the development of fine-grained vessel segmentation. One effective solution is skeleton extraction, which addresses the issue of weak vessel connectivity by processing the foreground region into a skeletal remnant that preserves the original area's connectivity.

In light of the aforementioned challenges, we propose a novel semi-supervised framework called EXP-Net for blood vessel segmentation. The proposed framework consists of a student network, a teacher network, and an expert network. The Mean Teacher model forms the foundation of EXP-Net, which uses a small number of labeled data and numerous unlabeled data for semi-supervised segmentation. We further introduce an expert network, which is the core of EXP-Net. This network receives the decoded features from the teacher network and generates improved predictions to correct the student network. The expert network comprises a knowledge enhancement module and a connectivity enhancement module. The knowledge enhancement module uses a transformer-based architecture to model long-range contextual interactions of the multi-scale decoding features of the teacher network. Meanwhile, the connectivity enhancement module improves vessel prediction connectivity from two perspectives, i.e., geometry and topology, which complement each other to extract effective vessel skeleton. By incorporating the skeleton consistency loss, the proposed EXP-Net aims to improve the connection of tubular vessels, especially at curves and bifurcations, as illustrated in Fig. 1.

Concretely, the Mean Teacher model employs the same network structure for both the student and the teacher networks. The teacher network is updated without additional training by averaging the model weights of the student network. The knowledge enhancement module of the expert network employs self-attention based and cross-attention based transformers to collaboratively refine multi-scale decoding features of the teacher network. Additionally, we establish a link between the theoretical properties of topology and geometry to extract vessel skeletons for vessel connectivity enhancement. Specifically, morphological processes are utilized to derive the topological structure skeleton, which expresses the fundamental global topology characteristics while ignoring the position relationship of the vessels. In complement, the geometrical shape skeleton is determined by calculating the distance transformation and locating the central axis of the blood vessel. It is noteworthy that the relevant calculations are carried out in a non-parametric manner.

We conducted extensive experiments on the VESSEL-NIR subcutaneous vessel dataset [1], four retinal vessel datasets, namely DRIVE [8], STARE [9], CHASE_DB1 [10], and HRF [11], and one coronary artery dataset DCA1 [12]. By comparing our proposed EXP-Net with other state-of-the-art fully-supervised and semi-supervised methods, we demonstrate its competitive performance, providing evidence of its superiority and generalizability. In addition, our EXP-Net particularly excels in enhancing vessel connectivity and integrity on various vessel datasets.

To sum up, this work makes the following contributions:

- We propose a novel semi-supervised framework, EXP-Net, for the segmentation of blood vessels. Our approach utilizes an expert network to guide the Mean Teacher in enhancing knowledge distillation.
- Our proposed knowledge enhancement module employs cascaded attention-based transformers to establish long-range dependencies among multi-level token components. This approach effectively extracts relevant vessel features from a global perspective.
- Our connectivity enhancement module improves vessel segmentation by enhancing connectivity at a detailed level. Leveraging the properties of topology and geometry, this module skeletonizes the predictions to reinforce vessel connectivity.
- The proposed EXP-Net framework surpasses the performance of previous state-of-the-art fully-supervised and semi-supervised methods on multiple blood vessel datasets.

II. Related Works

Semi-supervised Segmentation. In the field of medical image segmentation, pixel-level labeling of biomedical data is both labor-intensive and time-consuming. As a result, the research community has placed greater emphasis on semi-supervised segmentation techniques to leverage the valuable information in unannotated data. Existing approaches, such as [13] which aligns the features of the teacher and student network, and [14] which employs a generative model within a Bayesian framework to mitigate overfitting, have shown promising results. Additionally, [15] developed a tripled-uncertainty guided framework to address potential bias in the teacher network caused by annotation scarcity. In contrast, our proposed EXP-Net framework introduces an expert network that guides knowledge distillation to produce more reliable predictions.

Vision Transformer. Motivated by the impressive performance of transformers in natural language processing, the use of vision transformers has emerged for computer vision tasks. The goal of these transformers is to divide images into vision tokens and capture long-range dependencies of sequential data. Prior work, such as [16], has addressed attention collapse issues by regenerating attention maps to increase feature diversity at different layers. Additionally, [17] introduced a dual-branch transformer to combine image patches of different sizes for more robust image features. The effectiveness of

Authorized licensed use limited to: Wuhan University. Downloaded on October 12,2023 at 11:03:31 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3312338

transformers in learning long-range representations has also been demonstrated in medical image processing, as seen in [18], which improves semantic segmentation quality using a hierarchical Swin Transformer. In contrast to prior work, our proposed method leverages cascaded attention-based transformers to optimize the multi-layer decoding features of the Mean Teacher model.

Technologies for Vessel Segmentation. Deep learning has seen significant advancements in blood vessel segmentation algorithms, with neural network-based approaches being particularly impressive. For instance, [19] utilized fully recurrent convolutional networks to capture salient image features and motion signatures at multiple resolution scales. Additionally, [20] integrated a graph neural network into a unified CNN architecture to jointly exploit local appearances and global vessel structures. Meanwhile, [21] proposed a single VSSC Net for segmenting blood vessels in both coronary angiograms and retinal fundus images, incorporating two vessel extraction layers with added supervision. In addition, there are works that specifically address vessel connectivity issues. Araujo et al. [22] proposed a Variational Auto-Encoder that addresses topological inconsistencies in a compressed latent space. Gur et al. [23] incorporated a novel loss term utilizing morphological Active Contours Without Edges. Shit et al. [24] introduced a similarity measure for calculating the intersection of segmentation masks and their corresponding skeleta.

To address the challenges of weak vessel connectivity and poor pixel contrast, we have designed both knowledge and connectivity enhancement modules that respectively focus on vessel feature extraction from global and detailed perspectives. The knowledge enhancement module introduced in this study fortifies the multi-scale decoding capabilities of the teacher network, enabling it to provide global guidance for the student network's learning. Notably, the proposed connectivity enhancement module concurrently addresses both the geometric and topological attributes of vessel trees. We innovately intergrate morphological process (topology) with distance transformation (geometry) to excavate the skeleton feature at a micro level.

III. Method

A. Preliminaries

The Mean Teacher approach is a semi-supervised learning method that utilizes knowledge distillation between the teacher and student networks to improve performance by exploring richer feature information [6]. Our proposed EXP-Net is motivated by this approach and incorporates the Mean Teacher method as its fundamental structure to enhance vessel prediction accuracy. We refer to labeled samples as $\boldsymbol{S}_L = \{(\boldsymbol{X}_i, \boldsymbol{Y}_i)\}_{i=1}^{N}$ and unlabeled samples as $\boldsymbol{S}_U = \{\boldsymbol{X}_j\}_{j=N+1}^{N+M}$. Here, $\boldsymbol{X}_{i/j}$ represents the input vessel image, and $\boldsymbol{Y}_i$ is the corresponding ground-truth label.

To initialize the Mean Teacher method, we compute the relation matrix from the high-level semantic features of each sample. The student network $\boldsymbol{f}_s(\cdot,\cdot)$ is then utilized to form a better teacher network $\boldsymbol{f}_t(\cdot,\cdot)$ via exponential moving average (EMA) of its weights, without any additional training. The teacher network predictions act as corrections to those of the student network, resulting in more accurate predictions. The Mean Teacher objective function is formulated as:

$$L_{mt} = L_s(f_s(X_i,\theta), Y_i) + L_u(f_s(X_j,\theta), f_t(X_j,\theta')), \quad (1)$$

where θ and θ' denote the weights of the student and teacher networks, respectively. $\boldsymbol{L}_s(\cdot,\cdot)$ represents the supervised loss of the labeled $\boldsymbol{X}_i$. $\boldsymbol{L}_u(\cdot,\cdot)$ is the prediction consistency loss of the unlabeled $\boldsymbol{X}_j$.

To update the weight θ'_t of the teacher network at training step t, EMA is applied based on the consecutive weights θ_t:

$$\theta'_t = \alpha\theta'_{t-1} + (1-\alpha)\theta_t, \quad (2)$$

where α is a smoothing coefficient hyperparameter representing the weight decay rate of the student network. By controlling the weight decay rate, the teacher network is more robust in feature condensation than the student network.

B. Overview

Network Composition. Fig. 2 illustrates the comprehensive framework of our EXP-Net, comprising three components: the student, teacher, and expert networks. Adhering to the Mean Teacher paradigm, both the student and teacher networks adopt the U-Net architecture [25] and update weights through the EMA strategy. The expert network, serving as the crux of the EXP-Net, enhances the outputs generated by the teacher network and provides supplementary guidance to the student network. It comprises two modules: a knowledge enhancement module and a connectivity enhancement module.

More specifically, the knowledge enhancement module establishes connections between the multi-level decoding features of the teacher network to further enhance knowledge distillation. To achieve this, we employ one self-attention-based transformer, as well as multiple cross-attention-based transformers to facilitate long-range, dense feature interaction. These transformers are arranged in a sequential manner. Additionally, we introduce the connectivity enhancement module to address the vessel connection problem from the perspective of the developed vessel skeleton characteristics. The module incorporates two complementary skeleton extraction approaches, one from a topological structure perspective and the other from a geometric shape perspective. By analyzing topology and geometry, we can preserve the vessel connection characteristics and simultaneously determine the skeleton positions.

Training Mode. Given that our semi-supervised EXP-Net uses both labeled and unlabeled data, we divide the training process into annotation and pseudo-annotation training. Initially, the student network is trained under the guidance of the labeled data. Throughout the supervised training iterations, we employ the EMA strategy to update the weight of the teacher network. The decoded representations of the teacher network are then passed to the knowledge enhancement module, which uses attention-based transformers to learn long-range dependencies among token features. During the supervised stage, the connectivity enhancement module skeletonizes the predictions of the knowledge enhancement module and the annotations. This

Authorized licensed use limited to: Wuhan University. Downloaded on October 12,2023 at 11:03:31 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3312338

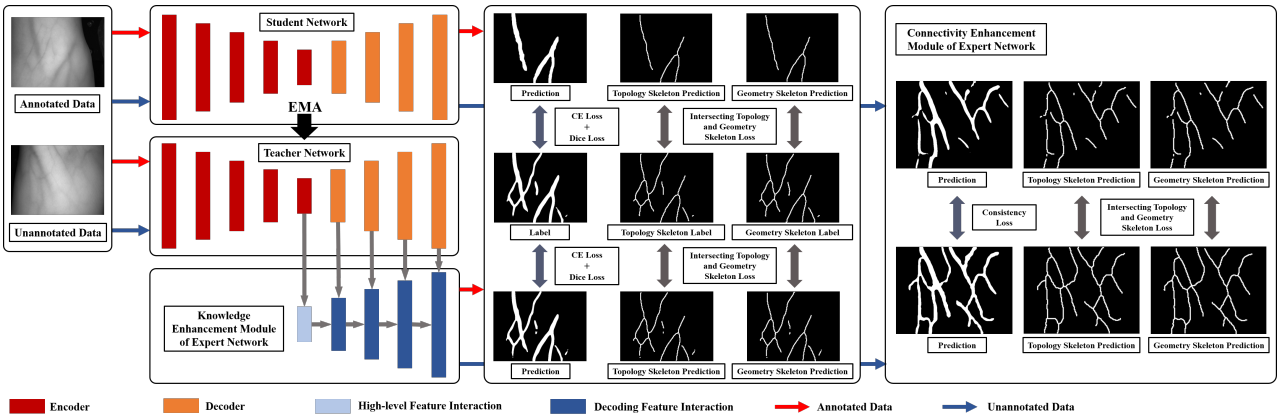

Fig. 2. Overall framework of EXP-Net for blood vessel segmentation. The EXP-Net is built upon the Mean Teacher model as its fundamental structure. In addition, the expert network is responsible for improving the decoded features generated by the teacher network, and it comprises knowledge and connectivity enhancement modules. The grey arrows represent the feature map transmissions.

minimizes the differences between the skeletons and forces the probability maps to achieve detailed vessel connections.

For unlabeled data, the predictions of the expert network serve as pseudo-labels to guide the predictions of the student network in two ways: vessel segmentation optimization and skeleton similarity computation. With the additional unsupervised training, EXP-Net can break through the limitations of annotations and produce promising results.

C. Knowledge Enhancement Module

Module Workflow. The knowledge enhancement module delves into deeper feature information in order to achieve superior rectification of the Mean Teacher. Our approach entails establishing collaboration between high-level and low-level semantic features via long-range dependencies learning through an attention-based transformer. The knowledge enhancement module utilizes both the high-level and low-level decoding features of the teacher network for module training.

Initially, the self-attention based transformer leverages the high-level feature from the encoder-decoder structure, which allows for global reception of the entire image. Subsequently, maps of the decode layers utilize their respective cross-attention based transformers to highlight regions of significant interest in the vessel tokens. As each decoder layer corresponds to a cross-attention based transformer, the knowledge enhancement module achieves dense decoded feature interaction. The details of the interaction are illustrated in Fig. 3.

High-level Feature Interaction. The bottom of the encoder-decoder structure contains the high-level semantic information that covers the entire image. We leverage the self-attention based transformer for high-level feature interaction. The self-attention based transformer consists of multi-head self-attention (MSA) module, multi-layer perceptron (MLP) block, and layer normalizations (LN) [26].

In the self-attention based transformer, the self-attention module takes three matrices as input, namely queries, keys, and values, all of which are obtained from the lower layer of the network [7]. The attention mechanism, denoted as $\boldsymbol{a}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V})$, can be expressed as follows:

$$a(Q, K, V) = SoftMax(\frac{QK^T}{\sqrt{d}})V, \tag{3}$$

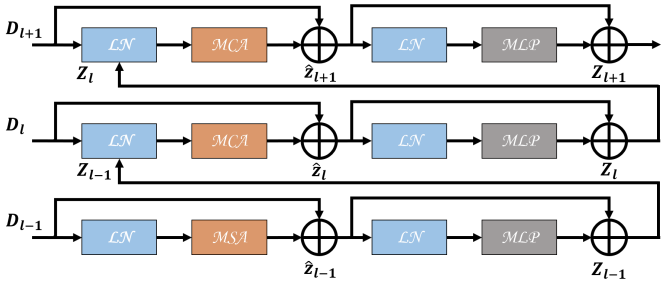

Fig. 3. Illustration of the knowledge enhancement module.

where $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} \in \mathbb{R}^{M^2 \times d}$ denote the query, key, and value matrices, respectively. Here, $\boldsymbol{K}^T$ represents the transpose of $\boldsymbol{K}$, while M^2 and d refer to the number of patches and the dimension of $\boldsymbol{Q}$.

Dense Feature Interaction. To integrate the dense decoded feature of the teacher network, we employ a cross-attention based transformer at each decoder layer. Unlike the self-attention based transformer, the cross-attention based transformer connects in a serial manner, as illustrated in Fig. 3. The l-th layer output $\boldsymbol{Z}_l$ of the continuous cross-attention based transformer can be expressed as follows:

$$\widehat{Z}_l = m(k(Z_{l-1}), k(D_l)) + D_l, \tag{4}$$

$$Z_l = p(k(\widehat{Z}_l)) + \widehat{Z}_l, \tag{5}$$

where $\boldsymbol{D}_l$ and $\boldsymbol{Z}_{l-1}$ denote the feature maps from the decoder and the output of the upper transformer layer, respectively. Here, $\boldsymbol{p}(\cdot)$, $\boldsymbol{m}(\cdot,\cdot)$, and $\boldsymbol{k}(\cdot)$ represent the functions of MLP, multi-head cross-attention (MCA) module, and LN, respectively. The implementation detail of the MCA module is:

$$m(M, N) = SoftMax(\frac{(W_q M)(W_k N)^T}{\sqrt{d}})(W_v N), \tag{6}$$

where $\boldsymbol{M}$ represents $\boldsymbol{k}(\boldsymbol{Z}_{l-1})$, while $\boldsymbol{N}$ represents $\boldsymbol{k}(\boldsymbol{D}_l)$. $\boldsymbol{W_q}$, $\boldsymbol{W_k}$, and $\boldsymbol{W_v}$ are the weight matrices corresponding to the query vector $\boldsymbol{Q}$, key vector $\boldsymbol{K}$, and value vector $\boldsymbol{V}$, respectively. Regarding the cross-attention based transformer, it should be noted that the queries correspond to the feature maps from the decoder of the teacher network, whereas the keys and values correspond to the outputs of the upper transformer layer.

Authorized licensed use limited to: Wuhan University. Downloaded on October 12,2023 at 11:03:31 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3312338

D. Connectivity Enhancement Module

Module Workflow. After the knowledge enhancement module, we introduce a connectivity enhancement module that provides a detailed perspective on highlighting blood vessel connectivity. This module complements the knowledge enhancement module, which offers a global view of vessel feature extraction. Specifically, we analyze the topological structure and geometrical shape of blood vessels to explore their connectivity characteristics. Topology and geometry are two different theoretical concepts. Topology involves abstracting objects into points and lines that are independent of their shape, with the goal of studying the connection between these points and lines. On the other hand, geometry emphasizes the shape and size of structures composed of points and lines.

In order to fully utilize the topological structure and geometrical shape of blood vessels, we employ two distinct skeleton extractions to enhance their connectivity. Skeletonization is a technique that reduces a region of interest to a skeletal remnant by discarding most of the original foreground pixels, while preserving the extent and connectivity of the target region. We will delve into the details of the relevant geometry and topology analyses in the subsequent sections.

Topology Analysis. The topology analysis involves utilizing morphological operations to preserve the overall topology characteristics. Specifically, we use iterative morphological erosion $\boldsymbol{e}(\cdot,\cdot)$ and opening $\boldsymbol{o}(\cdot,\cdot)$ operations to accurately generate the topological vessel skeleton. As the number of iterations increases, the target structure becomes thinner until no further thinning is possible, resulting in the production of the skeleton. The iterative calculation of the topology analysis $\boldsymbol{s}(\boldsymbol{P})$ is expressed as follows:

$$s(P) = \bigcup_{i=0}^{k} (e(P, iE) - o(e(P, iE), E)), \tag{7}$$

$$k = \max\{i | e(P, iE) \neq \emptyset\}, \tag{8}$$

where $\boldsymbol{P}$ represents the binary image to be skeletonized, and $\boldsymbol{E}$ is a structuring element. The structuring element $\boldsymbol{E}$ is a 3×3 square filter filled with 1 for morphological operations.

Here, k denotes the final iteration before the object is eroded to an empty set. By utilizing iterative erosion and opening operations, the resulting skeleton becomes optimally thin, connected, and minimally susceptible to erosion. Moreover, it ensures that the frame maintains the object's topology.

Geometry Analysis. To determine the central axis of the blood vessel, we perform geometry analysis, which involves calculating the distance transformation and locating the medial axis of the target vessels. The medial axis transform is computed as the ridges of the distance transformation, and in mathematics, it is defined as the center line (one-pixel wide) of pixels between the two (or more) edges of a structure.

Specifically, for each point p in the region $\boldsymbol{R}$, we search for the nearest point in the boundary $\boldsymbol{B}$. If there are two or more points in $\boldsymbol{B}$ that are equidistant from p, then p is a skeleton point. The skeleton extraction process in geometry analysis involves gathering all the skeleton points within the region $\boldsymbol{R}$. This collection of skeleton points is then designated as the central axis of the blood vessel.

Each skeleton point is characterized by the property that it maintains a minimum distance from the boundary point. To determine this distance, the required minimum distance $\boldsymbol{d}_s(p, \boldsymbol{B})$ between a point in the target region and its boundary is calculated using the following definition:

$$d_S(p, B) = \inf\{d(p, z) | z \in B\}. \tag{9}$$

Here, the Euclidean distance function $\boldsymbol{d}(p, z)$ is employed. As per the aforementioned theory, the skeleton point is defined as the fitting point p that satisfies the minimum distance. The skeleton within the geometric transform ensures the preservation of location invariance in the medial axis.

Summary. The topology analysis abstracts the size of the target, while the geometry analysis refines the location excursion of the target position. These complementary skeleton analyses enable improved vessel connectivity while preserving the vessel location information. The module is particularly effective for curved and bifurcated vessels and is not limited by the number of parameters required for inference.

E. Objectives

We utilize both a supervised training loss, denoted as $\boldsymbol{L}_s$, and an unsupervised training loss, denoted as $\boldsymbol{L}_u$, as the objective functions of our EXP-Net.

Supervised Training Loss. The student network is trained on labeled data to perform basic segmentation. The teacher network is then updated by averaging the model weights of the student network. The weighted teacher network leverages the labeled data to generate multi-layer decoding features, serving as an intermediate transmission. The expert network's knowledge enhancement module utilizes transformers to recover the spatial and semantic information of multi-layer decoding features. After the training of transformers, the connectivity enhancement module computes the intersecting skeleton consistency between the predictions of the knowledge enhancement module and the annotations.

The supervised training loss $\boldsymbol{L}_s$ for training with annotated data consists of the cross entropy loss $\boldsymbol{L}_{ce}(\cdot,\cdot)$, the Dice loss $\boldsymbol{L}_{dice}(\cdot,\cdot)$, and the intersecting skeleton loss $\boldsymbol{L}_{is}(\cdot,\cdot)$:

$$L_s = L_{ce}(f(X_i), Y_i) + L_{dice}(f(X_i), Y_i) + L_{is}(f(X_i), Y_i), \tag{10}$$

where the intersecting skeleton loss is the sum of the topological skeleton $\boldsymbol{g}_{top}(\cdot,\cdot)$ and the geometrical skeleton $\boldsymbol{g}_{geo}(\cdot,\cdot)$. The intersecting skeleton loss $\boldsymbol{L}_{is}(\cdot,\cdot)$ is defined as:

$$L_{is}(P_i, Y_i) = g_{top}(P_i, Y_i) + g_{geo}(P_i, Y_i), \tag{11}$$

where:

$$g_{top/geo}(P_i, Y_i) = 2 \times \frac{t_{pre}(P_i, Y_i) \times t_{sen}(P_i, Y_i)}{t_{pre}(P_i, Y_i) + t_{sen}(P_i, Y_i)}, \tag{12}$$

$$t_{pre}(P_i, Y_i) = \frac{s(P_i) \cap Y_i}{s(P_i)}, t_{sen}(P_i, Y_i) = \frac{s(Y_i) \cap P_i}{s(Y_i)}. \tag{13}$$

Using the skeleton extraction function $\boldsymbol{s}(\cdot)$, the predictions $\boldsymbol{P}_i$ and annotations $\boldsymbol{Y}_i$ are utilized to calculate the fractions of skeleton precision, denoted as $\boldsymbol{t}_{pre}(\cdot,\cdot)$, and skeleton sensitivity, denoted as $\boldsymbol{t}_{sen}(\cdot,\cdot)$. As in the approach proposed by

Authorized licensed use limited to: Wuhan University. Downloaded on October 12,2023 at 11:03:31 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3312338

[24], the harmonic mean of these measures is employed to maximize both precision and sensitivity metrics.

Unsupervised Training Loss. In the context of unsupervised learning, the unannotated images are utilized by the expert network to enhance vessel feature extraction of the student network. In this regard, the unannotated predictions of the expert network serve as pseudo-annotations to guide the student network. The specialized unsupervised training loss $\boldsymbol{L}_u$ comprises the consistency loss $\boldsymbol{L}_{con}(\cdot,\cdot)$ and the intersecting skeleton loss $\boldsymbol{L}_{is}(\cdot,\cdot)$:

$$L_u = L_{con}(f_s(X_j,\theta), f_e(X_j,\gamma)) + L_{is}(f_s(X_j,\theta), f_e(X_j,\gamma)), \tag{14}$$

$$L_{con}(P_s,P_e) = \frac{1}{n}\sum\nolimits_{i=1}^{n}(p_e(i) - p_s(i))^2, \tag{15}$$

where n is the pixel number of the predicted map, $\boldsymbol{p}_s(i)$ and $\boldsymbol{p}_e(i)$ are the classification results of pixel i of the student and expert networks. θ and γ denote the weights of the student and the expert networks, respectively.

Afterwards, the unannotated predictions of both the student and the expert networks are skeletonized to calculate the intersecting skeleton loss $\boldsymbol{L}_{is}(\cdot,\cdot)$. Unlike the supervised training approach, in the unsupervised training approach, the unannotated skeleton predictions of the expert network are used as labels for optimization. Ultimately, the unsupervised training objectives $\boldsymbol{L}_{is}(\cdot,\cdot)$ and $\boldsymbol{L}_{con}(\cdot,\cdot)$ alleviate the training bottleneck of finite annotations.

IV. Experiments and Results

A. Implementation Details

EXP-Net utilizes U-Net [25] as its base network, with randomly cropped patches of size 480 × 640 as input. The network is trained from scratch using the Adam optimizer [27] with default parameter settings (β_1 = 0.9 and β_2 = 0.99). The initial learning rate is decayed by iterations with a power factor. The batch size consists of 6 images, with 3 annotated and 3 unannotated. The weight decay rate is controlled by the Exponential Moving Average (EMA) hype-parameter α with a value of 0.99. The implementation is based on PyTorch and is performed using two NVIDIA GeForce RTX 3090 GPUs.

The experiments are conducted on six blood vessel image datasets, namely VESSEL-NIR, DRIVE, STARE, CHASE_DB1, HRF and DCA1. Specific parameter settings for each comparison method are described in later subsections. The evaluation on VESSEL-NIR uses four widely-adopted metrics: Pixel-wise Accuracy (PAC), Recall (Rec), Intersection-over-Union (IoU), and Dice Similarity Coefficient (DSC). For retinal vessel and coronary artery segmentation, we adopt five commonly used metrics, namely sensitivity (Sen), specificity (Spe), accuracy (Acc), F1, and Area Under the Curve (AUC).

B. Experiments on Subcutaneous Vessel

Data. To evaluate the effectiveness of our method for subcutaneous vessel segmentation, we conduct experiments on the VESSEL-NIR dataset [1], which contains 1600 annotated and 2000 unannotated samples with a resolution of 480 × 640. The

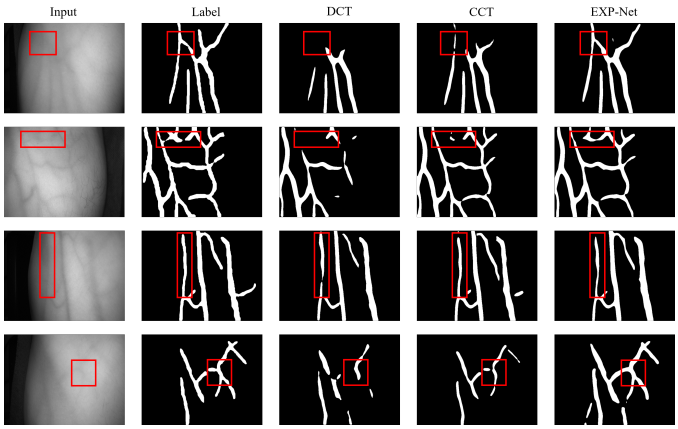

Fig. 4. Qualitative comparison of subcutaneous vessel segmentation. Visualizing vital regions, like thin and highly-curved crossover vessels, stands marked with red boxes.

subcutaneous vessels of the arm, which are located beneath the skin and can be visualized up to 3 mm deep, are imaged using a NIR optical imaging apparatus in the wavelength range of 760 to 1000 nm.

Setups. We partition the 1600 annotated samples into three sets, with 800, 200, and 600 samples assigned for training, validation, and testing, respectively. Furthermore, we utilize the remaining 2000 unannotated samples for unsupervised training. The model is trained in a semi-supervised manner for 10000 iterations, using the Adam optimizer with an initial learning rate of $1e^{-4}$.

Fully-supervised Results. Our EXP-Net is compared against seven state-of-the-art fully-supervised models, and the results with mean value and standard deviation are shown in Table I. Because of the limited amount of labeled data, the comparative methods are trained on 1600 annotated samples. EXP-Net is trained using a combination of 1600 labeled samples and 2000 unlabeled samples. All comparative methods demonstrate feasibility for subcutaneous vessel segmentation and serve as valuable references for our EXP-Net.

Our EXP-Net achieves the best performance in PAC (97.94%), Rec (89.67%), IoU (78.99%), and DSC (88.09%), surpassing all seven fully-supervised models. In comparison to the fully-supervised models, our EXP-Net maximizes the utilization of cost-effective unlabeled data for the extraction of vessel features. The difference in performance among the comparative algorithms is relatively small in terms of PAC. However, the superiority of our EXP-Net is evident in the other three metrics. For instance, Semantic FPN achieves a Rec of 79.01%, which is over 10% lower than our EXP-Net, demonstrating the effectiveness of our method in vessel feature extraction. Our EXP-Net outperforms OCR in PAC, Rec, IoU, and DSC, with performance improvements of 0.75%, 14.35%, 6.23%, and 4.27%, respectively. Although K-Net shows excellent results in most metrics, our EXP-Net still outperforms it with a gain of 3.48% in IoU and 2.49% in DSC. The Genetic U-Net showcases the benefit of achieving good results while upholding high efficiency.

The outstanding performance of EXP-Net is attributed to the knowledge enhancement module, which learns long-range dependencies, and the connectivity enhancement module, which

Authorized licensed use limited to: Wuhan University. Downloaded on October 12,2023 at 11:03:31 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3312338

TABLE I
COMPARISON OF PREDICTIVE PERFORMANCE, PARAMETER NUMBERS, AND INFERENCE TIME PER IMAGE ON SUBCUTANEOUS VESSEL DATASET.

	Method	PAC (%)	Rec (%)	IoU (%)	DSC (%)	Param (M)	Time (ms)
Fully-supervised	Semantic FPN [28]	97.36±1.62	79.01±9.35	75.01±9.09	85.34±7.56	49.35	87
	OCR [29]	97.19±1.54	75.32±11.19	72.76±10.04	83.82±7.00	58.76	99
	CGNet [30]	97.44±1.42	80.59±10.03	75.91±8.59	86.02±5.79	62.63	103
	K-Net [31]	97.37±1.63	82.11±11.38	75.51±9.63	85.60±7.34	51.08	90
	ConvNeXt [32]	97.31±1.69	85.47±10.97	75.97±9.84	85.97±6.75	45.81	80
	Genetic U-Net [33]	97.74±1.52	82.03±9.68	77.37±8.64	86.12±6.97	**17.80**	**38**
	DEDCGCNEE [34]	97.68±1.43	81.05±10.31	77.41±9.13	86.47±5.98	52.47	91
Semi-supervised	DCT [35]	97.54±1.12	82.32±10.90	72.71±8.66	83.91±5.89	31.57	48
	AdvEnt [36]	97.73±1.19	83.93±10.15	76.21±9.04	86.18±6.21	29.89	46
	UA-MT [37]	97.47±1.84	80.06±11.33	72.69±10.43	84.05±7.75	29.64	45
	CCT [38]	97.69±1.11	84.23±9.78	75.92±8.50	86.03±5.81	30.55	46
	CPS [39]	97.37±1.36	80.60±9.12	72.14±9.37	83.68±7.37	29.64	45
	CS-CADA [40]	97.61±1.12	86.45±8.39	76.00±7.65	86.15±5.09	32.64	49
	SSL4DSA [41]	97.53±1.15	85.93±9.14	75.10±8.05	85.53±5.43	29.17	43
	EXP-Net (Ours)	**97.94±0.98**	**89.67±6.87**	**78.99±6.96**	**88.09±4.47**	43.37	66

directs the extraction of fine-grained details.

Semi-supervised Results. To demonstrate the superiority of our proposed method in leveraging both labeled and unlabeled data, we compare it with seven state-of-the-art semi-supervised models. Our EXP-Net proves to be more advantageous in extracting semantic information, particularly in the context of vessel segmentation with limited annotations. With the highest Dice and IoU values, our proposed method outperforms all the other semi-supervised methods. Specifically, DCT achieves results of 82.32% Recall and 97.54% Precision-At-Count, which are even better than several fully-supervised methods. AdvEnt also shows considerable improvement in DSC compared to other semi-supervised segmentation methods. In comparison, EXP-Net surpasses UA-MT by 6.30% and 4.04% in IoU and DSC, respectively. CCT obtains an IoU of 75.92% and a DSC of 86.03% with the assistance of unlabeled data. CPS utilizes both unlabeled data and pseudo labels to enhance the similarity between the predictions of two perturbed networks. The IoU and DSC scores of CS-CADA are inferior to our EXP-Net, especially the gaps are even 2.99% and 1.94%. Unlike other semi-supervised models, EXP-Net's advantage is attributed to the expert network, which leverages valuable feature information from global and detailed views.

Efficiency. Table I presents the number of parameters and inference speed associated with our EXP-Net. Our proposed method exhibits high efficiency, with slightly lower efficiency compared to the semi-supervised models, while significantly outperforming them in terms of vessel segmentation. The integration of the knowledge enhancement module, a crucial component of the EXP-Net, results in an increase in the number of parameters. However, this increase in parameters leads to a considerable improvement in performance. Remarkably, our EXP-Net surpasses Semantic FPN in DSC by 2.75% and saves 5.98M parameters at the same time. Overall, our proposed method achieves both exceptional segmentation performance and fast running speed.

Visualization. Fig. 4 illustrates a visual comparison of the results obtained from our EXP-Net and the currently existing state-of-the-art models. The critical segmentation region is identified with red boxes. Our EXP-Net exhibits superior performance compared to the semi-supervised models DCT and CCT, specifically in the detection of curves and bifurcations. The vessel images of VESSEL-NIR are acquired under challenging conditions with weak image contrast. Our EXP-Net yields more robust segmentation results compared to the semi-supervised models, notably preserving the intricate vessel locations and cross-connections.

C. Experiments on Retinal Vessel

Data. We conducted additional testing on four datasets of retinal vessels, namely DRIVE [8], STARE [9], CHASE_DB1 [10], and HRF [11]. The DRIVE dataset comprises 40 fundus images in color with a resolution of 565 × 584, captured using a Canon CR5 non-mydriatic 3-CCD camera with a field of view of 45°. Among these images, 7 show early signs of mild diabetic retinopathy, while 33 are healthy. The STARE dataset consists of 20 images with a size of 700 × 605 pixels, of which 10 are from pathological indications and the other 10 are from healthy subjects. The images were captured using a Topcon TRV-50 fundus camera with a field of view of 35°. The CHASE_DB1 dataset comprises 28 eye fundus images with a resolution of 990 × 960 pixels, all of which are from the left and right eyes of 14 children. The images were captured using a Nidek NM-200-D camera with a field of view of 30°. Finally, the HRF dataset includes 45 retinal images, categorized into healthy, diabetic retinopathy, and glaucomatous, with 15 images per category. Each image has a resolution of 3504 × 2336 pixels. It is noteworthy that all the manual annotations for the retinal images in the four datasets were performed by experts.

Setups. The DRIVE dataset is comprised of 40 images, with 20 images allocated for training and 20 for testing. Due to the unsupervised training stage of our EXP-Net, we selected 5 images from the testing set for unsupervised learning, while the other 15 images were used for testing. A similar situation arises for the remaining datasets. Since the STARE dataset does not have a uniform division of images into training and test sets, we performed a two-fold cross-validation for evaluation purposes. In the test split of 10 images, 4 images were

Authorized licensed use limited to: Wuhan University. Downloaded on October 12,2023 at 11:03:31 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3312338

TABLE II
COMPARISON WITH STATE-OF-THE-ARTS ON RETINAL VESSEL DATASETS DRIVE, STARE, CHASE_DB1, AND HRF.

	Method	DRIVE					STARE					CHASE_DBI					HRF				
		Sen(%)	Spe(%)	Acc(%)	AUC(%)	F1(%)	Sen(%)	Spe(%)	Acc(%)	AUC(%)	F1(%)	Sen(%)	Spe(%)	Acc(%)	AUC(%)	F1(%)	Sen(%)	Spe(%)	Acc(%)	AUC(%)	F1(%)
Fully-supervised	VSSC Net [21]	78.27	98.21	96.27	97.89	-	**87.38**	98.12	97.37	99.05	-	72.33	98.65	96.33	97.06	-	70.54	98.34	95.57	98.31	-
	ConvNeXt [32]	81.67	98.21	96.25	98.41	81.78	83.73	96.62	97.17	98.98	85.50	83.47	97.75	96.99	98.73	82.53	82.73	98.16	96.37	98.65	80.77
	Genetic U-Net [33]	83.00	97.58	95.77	98.23	**83.14**	86.58	98.46	97.19	**99.21**	**86.30**	**84.63**	98.18	96.67	98.80	82.23	82.20	98.18	96.67	98.72	81.79
	DEDCGCNEE [34]	**83.59**	98.26	**97.05**	**98.66**	82.88	84.05	98.61	**97.51**	98.99	83.63	84.00	98.56	**97.62**	**98.98**	82.61	81.69	98.25	96.95	98.45	80.97
	CS^2-Net [42]	81.54	97.57	95.53	97.84	82.28	83.96	98.13	96.70	98.75	84.20	83.29	97.84	96.51	98.51	81.41	78.90	97.95	96.18	97.58	79.35
	SCS-Net [43]	82.89	98.38	96.97	98.37	-	82.07	98.39	97.36	98.77	-	83.65	98.39	97.44	98.67	-	81.14	98.23	96.87	98.42	-
	Bridge-Net [44]	78.53	98.18	95.65	98.34	82.03	80.02	98.64	96.68	99.01	82.89	81.32	98.40	96.67	98.93	**82.93**	**85.70**	96.90	95.90	98.64	81.53
Semi-supervised	DCT [35]	78.38	98.28	96.53	97.69	79.67	78.17	98.09	96.55	98.24	77.60	80.78	98.57	97.29	97.96	80.88	82.84	98.27	97.11	**98.77**	80.44
	AdvEnt [36]	79.40	98.10	96.51	97.92	79.21	77.81	98.04	96.46	98.36	77.32	76.05	98.67	97.04	97.33	78.50	81.27	98.16	96.92	98.39	79.02
	UA-MT [37]	78.84	98.52	96.87	97.53	81.06	78.91	97.53	96.10	98.21	75.95	76.59	98.59	96.91	97.46	77.08	81.57	98.30	97.05	98.31	79.77
	CCT [38]	78.14	97.88	96.42	97.74	79.63	81.72	97.64	96.40	98.73	78.18	78.14	98.50	97.05	97.72	78.94	79.44	**98.48**	97.05	97.74	74.76
	CPS [39]	78.55	98.49	96.77	97.83	80.68	84.32	96.75	95.79	98.89	76.19	73.94	**98.90**	97.11	97.24	78.46	80.73	98.32	97.00	98.19	79.30
	CS-CADA [40]	80.12	**98.80**	96.58	98.02	80.66	83.64	96.99	95.91	98.27	76.76	79.17	98.21	96.79	97.64	80.41	79.54	97.93	96.55	98.23	79.14
	SSL4DSA [41]	80.57	98.46	96.47	98.11	81.02	80.81	**98.99**	97.36	98.56	78.82	80.42	98.72	96.67	97.83	80.04	80.38	97.94	96.63	97.98	81.32
	EXP-Net (Ours)	81.15	98.37	96.67	98.16	81.79	82.74	98.05	96.77	98.97	83.13	80.86	98.00	96.92	98.54	80.10	81.99	98.42	**97.16**	98.53	**82.17**

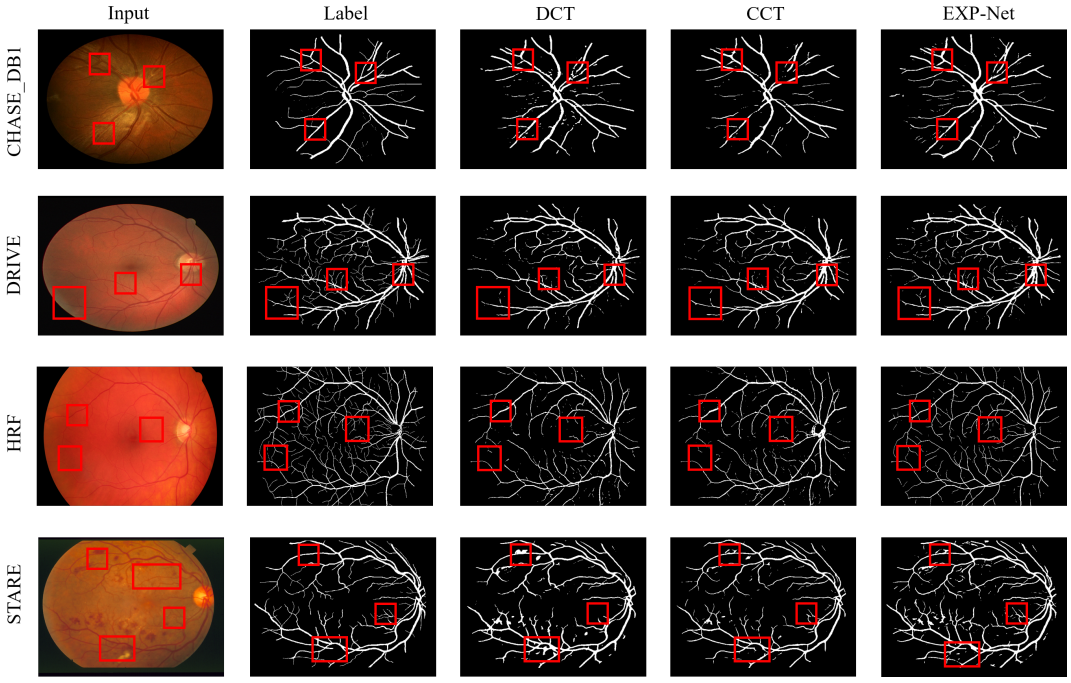

Fig. 5. Qualitative results on retinal vessel datasets. Visualizing vital regions, like complex vascular structures, stands marked with red boxes.

used for unsupervised correction, while the others were used for testing. For CHASE_DB1, the first 20 images were used as the training set, with 3 images allocated for unsupervised correction and the remaining 5 images for testing. Finally, for HRF, the 15 images of each type were evenly distributed into three parts for training, unsupervised correction, and testing. We trained the model for 5000 iterations with an initial learning rate of $5e^{-3}$. For all retinal vessel datasets, our EXP-Net conducts performance evaluations without using FOV masks, which is also the case for all the comparison methods.

Fully-supervised Results. Table II highlights the competitiveness of our EXP-Net when compared to other state-of-the-art methods. Furthermore, the implementation of EXP-Net in a semi-supervised manner necessitates a balance between annotated and unannotated samples, which is often not feasible for retinal vessel datasets due to a lack of annotated samples. It is also noteworthy that our EXP-Net performs well with small amounts of data, and even outperforms state-of-the-art methods in several evaluation metrics.

It can be observed that VSSC Net performs well on all four datasets, particularly on STARE, achieving a Sen of 87.38%. ConvNeXt exhibits outstanding performance compared to other methods, as indicated by the impressive AUC and F1 scores. CS^2-Net integrates local features with global dependencies and normalizations, and the metric scores of Genetic U-Net exceed those of most fully-supervised algorithms on several datasets. The competitive results of SCS-Net suggest that the proposed scale and context-sensitive network effectively captures representative and distinguishing features.

Authorized licensed use limited to: Wuhan University. Downloaded on October 12,2023 at 11:03:31 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3312338

While Bridge-Net is specialized in retinal vessel segmentation, the merit of EXP-Net is demonstrated in its achievements on DRIVE and STARE. Even when compared to fully-supervised models, EXP-Net still achieves competitive results.

It is noteworthy that the evaluation scores of our EXP-Net are marginally lower than those of certain fully-supervised methods. On the DRIVE dataset, our EXP-Net achieves Sen and Acc scores that are 1.74% and 0.30% lower than SCS-Net, respectively. Additionally, DEDCGCNEE gets the scores of 84.00% Sen, 98.56% Spe, and 82.61% F1 on CHASE_DB1, exceeding our results by 3.14%, 0.56%, and 2.51%, respectively. However, the key strength of our EXP-Net lies in its ability to effectively utilize cost-effective unannotated samples and still achieve competitive results. This indicates substantial potential for our EXP-Net to excel in scenarios with scarce annotations. Furthermore, in Section IV-D, we conduct ablation studies to assess the learning capability of our EXP-Net with respect to unlabeled data.

Semi-supervised Results. To assess the guiding capacity of unannotated data in EXP-Net, we benchmarked it against seven cutting-edge semi-supervised models. In general, semi-supervised models exhibit slightly inferior performance compared to fully-supervised models. This can be attributed to the fact that unsupervised training is responsible for a portion of the retinal images. Nevertheless, the semi-supervised algorithms in Table II demonstrated competitive performance in several metrics. Specifically, for CCT, the Sen score surpasses that of Bridge-Net by 1.7% on STARE. Additionally, AdvEnt outperforms multiple fully-supervised models on DRIVE and HRF. UA-MT trails behind our EXP-Net on all four datasets, with a difference in Sen scores of 2.31%, 3.83%, 3.29%, and 0.42%, respectively. The F1 scores of SSL4DSA on all retinal vessel datasets are marginally inferior to our EXP-Net.

Significantly, our EXP-Net exhibits potential for enhancement in partial metric scores. DCT, with 82.84% Sen, surpasses our EXP-Net by 0.85%. Similarly, CPS and CS-CSDA achieve slightly higher Sen scores (84.32% and 83.64%, respectively) compared to our EXP-Net's score of 82.74%. Nevertheless, our EXP-Net demonstrates the most comprehensive performance across all retinal vessel datasets.

Visualization. Figure 5 presents the predictions of comparison algorithms (DCT, CCT) and our EXP-Net on four retinal vessel datasets. Our EXP-Net leverages the knowledge and connectivity enhancement modules to improve the segmentation performance of tiny and inconspicuous blood vessels. The ability of our EXP-Net to mine valid vessel information from unlabeled data for additional learning relevance is particularly noteworthy. Even under the interferences of soft and hard exudates, our EXP-Net demonstrates the ability to perform intelligent identification and segmentation of blood vessels.

D. Ablation Studies

We conducted ablation studies to investigate the effects of the critical components of EXP-Net on five vessel datasets. We performed additional experiments to analyze the impact of labeled and unlabeled data proportion on the results.

Effect of Mean Teacher Model. Table III demonstrates that Mean Teacher performs well on VESSEL-NIR, achieving 72.32%, 83.78%, and 97.45% for IoU, DSC, and PAC, respectively. These results suggest that Mean Teacher specializes in distilling knowledge from the teacher to the student network. To compare the performance of Mean Teacher with that of U-Net, we conducted experiments, which show that Mean Teacher outperforms U-Net by 2.81% and 4.06% for DSC and IoU on VESSEL-NIR, respectively. The generality of EXP-Net is assessed through experiments on four retinal vessel datasets, as presented in Table IV. Mean Teacher forms the foundation of EXP-Net, utilizing knowledge distillation to extract feature information from unannotated data. The linear improvement from U-Net to Mean Teacher suggests that the EMA strategy of Mean Teacher provides inspiration to the teacher network. Our EXP-Net builds on this foundation by further refining the teacher network.

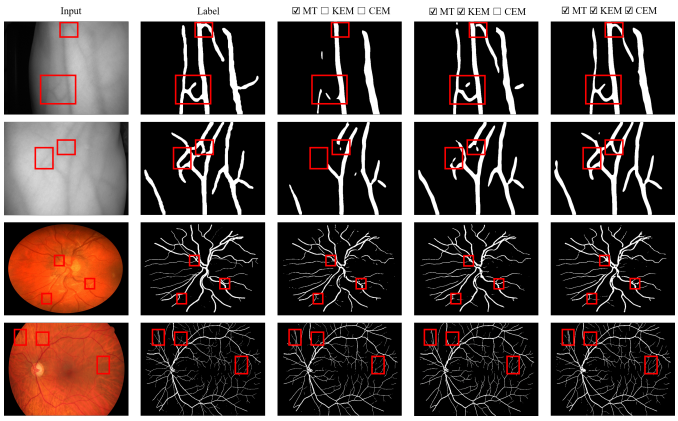

Fig. 6. Visualization of thin and highly-curved crossover vessels in original images, predictions of EXP-Net, and ground-truths. Key segments are marked with red boxes. MT: Mean Teacher, KEM: Knowledge Enhancement Module, CEM: Connectivity Enhancement Module.

TABLE III
ABLATION STUDIES ON VESSEL-NIR. MT: MEAN TEACHER, KEM: KNOWLEDGE ENHANCEMENT MODULE, CEM: CONNECTIVITY ENHANCEMENT MODULE. PA, IT, AND TT REPRESENT PARAMETER NUMBERS, INFERENCE TIME PER IMAGE, AND TRAINING TIME.

U-Net	MT	KEM	CEM	PAC(%)	Rec(%)	IoU(%)	DSC(%)	clDice(%)	Pa(M)	IT(ms)	TT(h)
✓				96.89	83.38	68.26	80.97	82.43	**29.64**	**45**	**0.4**
✓	✓			97.45	88.13	72.32	83.78	85.17	29.64	45	0.9
✓	✓	✓		97.79	88.46	77.23	86.99	88.52	43.37	66	1.6
✓	✓	✓	✓	**97.94**	**89.67**	**78.99**	**88.09**	**91.26**	43.37	66	3.5

Effect of Knowledge Enhancement Module. Table III reveals a notable improvement in the DSC score on VESSEL-NIR, from 83.78% to 86.99%, which can be attributed to the model's ability to establish global connections between token features. Moreover, as demonstrated in Table IV, the Mean Teacher and knowledge enhancement module work in tandem, achieving Sen scores of 80.29%, 81.91%, 79.88%, and 80.76% on DRIVE, STARE, CHASE_DB1, and HRF, respectively. Notably, the knowledge enhancement module delivers notable gains in Sen, Spe, and Acc metrics of 3.18%, 0.49%, and 0.93%, respectively, on HRF, underscoring its corrective role in the model. The consistent improvements across all datasets demonstrate the efficacy of the proposed algorithm, which harnesses vision transformers to accentuate long-range relation

Authorized licensed use limited to: Wuhan University. Downloaded on October 12,2023 at 11:03:31 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3312338

TABLE IV
ABLATION STUDIES ON RETINAL VESSEL DATASETS DRIVE, STARE, CHASE_DB1, AND HRF. MT: MEAN TEACHER, KEM: KNOWLEDGE ENHANCEMENT MODULE, CEM: CONNECTIVITY ENHANCEMENT MODULE.

					DRIVE			STARE			CHASE_DBI			HRF		
Model	U-Net	MT	KEM	CEM	Sen(%)	Spe(%)	Acc(%)	Sen(%)	Spe(%)	Acc(%)	Sen(%)	Spe(%)	Acc(%)	Sen(%)	Spe(%)	Acc(%)
No.1	✓				76.83	98.12	96.21	74.90	**98.20**	96.38	75.09	**98.85**	**97.07**	76.69	97.91	96.11
No.2	✓	✓			78.15	98.02	96.26	77.05	98.07	96.41	76.33	98.46	96.89	77.58	97.89	96.18
No.3	✓	✓	✓		80.29	98.14	96.31	81.91	96.87	95.67	79.88	98.15	97.00	80.76	**98.4**8	97.11
No.4	✓	✓	✓	✓	**81.15**	**98.37**	**96.67**	**82.74**	98.05	**96.77**	**80.86**	98.00	96.92	**81.99**	98.42	**97.16**

TABLE V
RESULTS ON FIVE BLOOD VESSEL DATASETS WITH DIFFERENT PROPORTIONS (**10%**, **25%** AND **50%**) OF LABELED DATA.

	VESSEL-NIR				DRIVE			STARE			CHASE_DB1			HRF		
	PAC (%)	Rec (%)	IoU (%)	DSC (%)	Sen (%)	Spe (%)	Acc (%)	Sen (%)	Spe (%)	Acc (%)	Sen (%)	Spe (%)	Acc (%)	Sen (%)	Spe (%)	Acc (%)
10%	97.48	96.77	58.40	73.74	41.19	**98.91**	93.79	39.75	97.57	93.03	43.86	97.82	94.02	59.31	96.99	93.81
25%	97.19	**97.79**	69.39	81.92	66.68	98.76	95.60	59.67	**98.40**	95.40	70.78	**98.55**	96.56	68.13	96.98	94.49
50%	**97.94**	89.67	**78.99**	**88.09**	**81.15**	98.37	**96.67**	**82.74**	98.05	**96.77**	**80.86**	98.00	**96.92**	**81.99**	**98.42**	**97.16**

features. However, as performance improves, the model's parameter count and inference time also increase.

Effect of Connectivity Enhancement Module. The objective of the connectivity enhancement module is to intensify the connection of tubular vessels. Table III and Table IV present the quantitative results. Specifically, on VESSEL-NIR, the DSC and IoU scores reach 88.09% and 78.99%, respectively. For DRIVE, STARE, and HRF, the Sen scores are 81.15%, 82.74%, and 81.99%, respectively. These results demonstrate that EXP-Net's two critical components provide advantages for blood vessel segmentation from both global and detailed perspectives. In conclusion, the Mean Teacher model and expert network components complement each other to facilitate our EXP-Net's segmentation performance improvement.

Additionally, we employ a novel metric called clDice [24] to quantify vascular connectivity analysis. The inclusion of the connectivity enhancement module (CEM) results in an increase in the clDice score from 86.99% to 88.09%, highlighting the efficacy of the pivotal CEM. As shown in Table III, the training time for CEM is 1.9 hours, utilizing two NVIDIA GeForce RTX 3090 GPUs. During the training phase, the calculation process of topology analysis and geometry analysis in CEM consumes a certain amount of time. But the non-parametric nature of the mathematical operations used in CEM avoids inflating the parameter number, and the lightweight framework of our EXP-Net ensures high inference speed during the testing phase. The visualization results depicted in Fig. 6 effectively demonstrate the effectiveness of each part of our EXP-Net. Our method successfully captures useful vascular features and effectively handles vessel disconnection problems, especially at curves and bifurcations.

Impact of Proportion of Labeled Data. Table V presents the performance of EXP-Net under various proportions of labeled data on five blood vessel datasets. Due to differences in data volume, the impact of varying the proportion of labeled data is evident. In both DRIVE and STARE datasets, the Sen scores at 50% annotations are significantly higher than those at 10%, with differences of 39.96% and 42.99%, respectively. As the proportion decreases from 50% to 10%, the performance of all datasets gradually declines. Notably, even at the 10% proportion, desirable performance is achieved, thanks to the semi-supervised training mode, which fully leverages valuable features from a small amount of labeled and a large amount of unlabeled data. For VESSEL-NIR, we employ four metrics to comprehensively demonstrate the impact of the proportion of labeled data. The IoU and DSC scores at 10% annotations are 58.40% and 73.74%, respectively, which are lower than the 50% annotations by 20.59% and 14.75%, respectively. In conclusion, while labeled data remains essential, our EXP-Net is capable of capturing the overall segmentation performance under varying proportions of labeled data.

TABLE VI
RESULTS ON SUBCUTANEOUS VESSEL DATASET WITH DIFFERENT DATA DISTRIBUTIONS. L: LABELED DATA. U: UNLABELED DATA.

Data Distribution	PAC (%)	Rec (%)	IoU (%)	DSC (%)
1000L+800U	97.52	84.98	73.83	84.78
1000L+1200U	97.56	86.45	74.51	85.23
1000L+1600U	97.68	87.13	76.53	86.90
1000L+2000U	**97.94**	**89.67**	**78.99**	**88.09**

Impact of Proportion of Unlabeled Data. We delve into the significance of the quantity of unlabeled data within our semi-supervised learning mode. Ablation studies are performed on various amounts of unlabeled data while keeping the labeled data constant in Table VI, aiming to assess the learning capability of our EXP-Net with respect to unlabeled data. We maintain a fixed number of labeled data (1000L) and vary the number of unlabeled data (*U) accordingly. The evaluation performance of the configuration 1000L+2000U surpasses that of 1000L+800U in IoU and DSC by 5.16% and 3.31%, respectively. The observed enhancements in the aforementioned metrics can be attributed to the feature extraction from the unlabeled data. The IoU and DSC scores show a gradual increase across four distributions, ranging from 800U to 2000U. The increments in DSC (0.45%, 1.67%, and 1.19%) and IoU (0.68%, 2.02%, and 2.46%) demonstrate the positive

Authorized licensed use limited to: Wuhan University. Downloaded on October 12,2023 at 11:03:31 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3312338

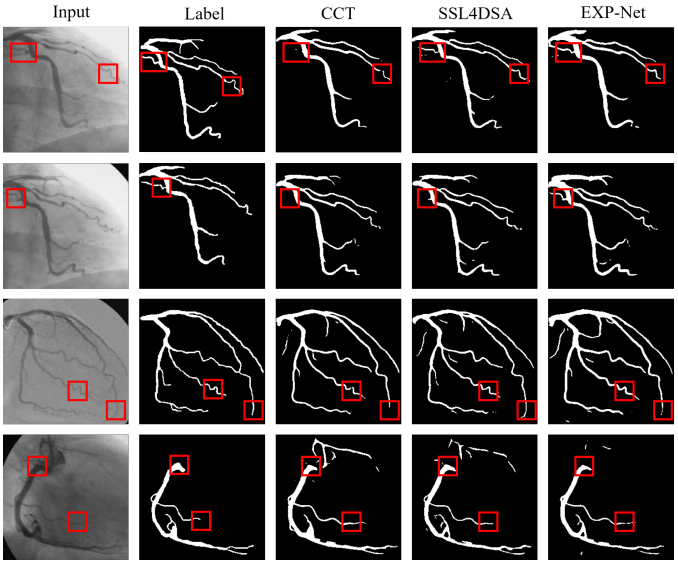

Fig. 7. Qualitative results on coronary artery dataset. Visualizing vital regions, like complex vascular structures, stands marked with red boxes.

impact of unlabeled data on model training. These findings also indicate the development potential of our EXP-Net to achieve better results with more cost-effective unlabeled data. Even in the scenario of 1000L+800U, our EXP-Net achieves commendable IoU and DSC scores of 73.83% and 84.78%, respectively. Overall, our EXP-Net exhibits remarkable learning ability from unlabeled data, which provides substantial assistance in real medical situations.

E. Extension to Coronary Artery Vessel

Data. To comprehensively assess the general applicability and effectiveness of our EXP-Net in semi-supervised vessel segmentation, we utilize two publicly available datasets: Database X-ray Coronary Angiograms (DCA1) [12] and X-ray angiography Coronary Artery Disease (XCAD) [45]. DCA1 consists of 134 X-ray coronary artery images, each with a resolution of 300 × 300 pixels, along with their corresponding ground truth images. All images were curated and annotated by cardiologists at the Cardiology Department of the Mexican Social Security Institute. XCAD encompasses 1621 angiography images acquired during stent placement using a General Electric Innova IGS 520 system. Each image in the XCAD dataset has a resolution of 512 × 512 pixels.

Setups. We partition the 134 images from DCA1 into three sets: 80 for training, 20 for validation, and 34 for testing. Additionally, we utilize all the images from XCAD for unsupervised training. The model is trained for 5000 iterations, starting with an initial learning rate of $1e^{-3}$.

Results. We conduct a comprehensive comparison with four of the latest semi-supervised frameworks to evaluate the effectiveness of EXP-Net in coronary artery segmentation. Each of these semi-supervised frameworks utilizes DCA1 and XCAD for supervised and unsupervised training, respectively. In Table VII, the results of our EXP-Net surpass the performance of other state-of-the-art algorithms in Sen, Acc, and F1. While CS-CADA and SSL4DSA achieve F1 scores of 79.23% and 80.45%, respectively, reflecting competitive results that serve as feasible reference values, the superiority of our EXP-Net is evident in Sen (81.42%) and F1 (80.87%) compared to CPS. Furthermore, while the Sen score of CCT is slightly lower than that of our EXP-Net, the other four metrics are significantly inferior to those of EXP-Net. Ultimately, these exceptional achievements in coronary artery vessel segmentation underscore the generality and effectiveness of EXP-Net.

TABLE VII
COMPARISON WITH STATE-OF-THE-ARTS ON CORONARY ARTERY DATASET DCA1.

Method	Sen(%)	Spe(%)	Acc(%)	AUC(%)	F1(%)
CCT [38]	81.41	98.23	97.45	98.24	78.07
CPS [39]	76.85	98.93	97.73	98.53	78.29
CS-CADA [40]	78.55	98.95	97.86	98.49	79.23
SSL4DSA [41]	80.64	98.91	97.92	98.71	80.45
EXP-Net (Ours)	**81.42**	**99.01**	**98.04**	**98.98**	**80.87**

Visualization. Fig. 7 displays the prediction instances of EXP-Net and the comparison algorithms (CCT and SSL4DSA). Among these comparison algorithms, CCT and SSL4DSA achieve the highest results in Sen (81.41%) and F1 (80.45%), respectively. However, the visualization clearly demonstrates that our EXP-Net exhibits exceptional superiority over CCT and SSL4DSA in detecting tiny blood vessels.

V. CONCLUSION

This paper presents a novel semi-supervised learning framework, EXP-Net, for blood vessel segmentation. The proposed framework is based on the Mean Teacher model and incorporates an expert network to enhance knowledge distillation. The expert network consists of two modules: the knowledge enhancement module and the connectivity enhancement module. The former utilizes vision transformers to extract vessel features with long-range dependencies, while the latter enhances vessel prediction by considering both geometry and topology. These critical components guide fundamental knowledge distillation to generate reliable segmentation, particularly in cases of weak vessel connectivity and poor pixel contrast. We conduct comprehensive experiments to evaluate the effectiveness of our proposed framework. Compared to multiple state-of-the-art models, our EXP-Net achieves competitive performance in several metrics.

REFERENCES

[1] N. Shen, T. Xu, Z. Bian, S. Huang, F. Mu, B. Huang, Y. Xiao, and J. Li, "SCANet: A unified semi-supervised learning framework for vessel segmentation," *IEEE Transactions on Medical Imaging*, 2022.

[2] S. Huang, J. Li, Y. Xiao, N. Shen, and T. Xu, "RTNet: Relation transformer network for diabetic retinopathy multi-lesion segmentation," *IEEE Transactions on Medical Imaging*, vol. 41, no. 6, pp. 1596–1607, 2022.

[3] L. Zhang, M. Fisher, and W. Wang, "Retinal vessel segmentation using multi-scale textons derived from keypoints," *Computerized Medical Imaging and Graphics*, vol. 45, pp. 47–56, 2015.

[4] I. Lázár and A. Hajdu, "Segmentation of retinal vessels by means of directional response vector similarity and region growing," *Computers in Biology and Medicine*, vol. 66, pp. 209–221, 2015.

[5] 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.

Authorized licensed use limited to: Wuhan University. Downloaded on October 12,2023 at 11:03:31 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3312338

[6] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proceedings of the Advances in Neural Information Processing Systems*, vol. 30, 2017.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proceedings of the Advances in Neural Information Processing Systems*, vol. 30, 2017.

[8] 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.

[9] 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.

[10] C. G. Owen, A. R. Rudnicka, R. Mullen, S. A. Barman, D. Monekosso, P. H. Whincup, J. Ng, and C. Paterson, "Measuring retinal vessel tortuosity in 10-year-old children: Validation of the computer-assisted image analysis of the retina (CAIAR) program," *Investigative Ophthalmology & Visual Science*, vol. 50, no. 5, pp. 2004–2010, 2009.

[11] 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.

[12] F. Cervantes-Sanchez, I. Cruz-Aceves, A. Hernandez-Aguirre, M. A. Hernandez-Gonzalez, and S. E. Solorio-Meza, "Automatic segmentation of coronary arteries in X-ray angiograms using multiscale analysis and artificial neural networks," *Applied Sciences*, vol. 9, no. 24, p. 5507, 2019.

[13] H. Wu, Z. Wang, Y. Song, L. Yang, and J. Qin, "Cross-patch dense contrastive learning for semi-supervised segmentation of cellular nuclei in histopathologic images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11 666–11 675.

[14] J. Wang and T. Lukasiewicz, "Rethinking bayesian deep learning methods for semi-supervised volumetric medical image segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 182–190.

[15] K. Wang, B. Zhan, C. Zu, X. Wu, J. Zhou, L. Zhou, and Y. Wang, "Semi-supervised medical image segmentation via a tripled-uncertainty guided mean teacher model with contrastive learning," *Medical Image Analysis*, vol. 79, p. 102447, 2022.

[16] D. Zhou, B. Kang, X. Jin, L. Yang, X. Lian, Z. Jiang, Q. Hou, and J. Feng, "DeepViT: Towards deeper vision transformer," *arXiv preprint arXiv:2103.11886*, 2021.

[17] C.-F. R. Chen, Q. Fan, and R. Panda, "CrossViT: Cross-attention multi-scale vision transformer for image classification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 357–366.

[18] A. Lin, B. Chen, J. Xu, Z. Zhang, G. Lu, and D. Zhang, "DS-TransUNet: Dual swin transformer U-Net for medical image segmentation," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, 2022.

[19] A. I. Chen, M. L. Balter, T. J. Maguire, and M. L. Yarmush, "Deep learning robotic guidance for autonomous vascular access," *Nature Machine Intelligence*, vol. 2, no. 2, pp. 104–115, 2020.

[20] 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.

[21] P. M. Samuel and T. Veeramalai, "VSSC Net: Vessel specific skip chain convolutional network for blood vessel segmentation," *Computer Methods and Programs in Biomedicine*, vol. 198, p. 105769, 2021.

[22] R. J. Araújo, J. S. Cardoso, and H. P. Oliveira, "A deep learning design for improving topology coherence in blood vessel segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 93–101.

[23] S. Gur, L. Wolf, L. Golgher, and P. Blinder, "Unsupervised microvascular image segmentation using an active contours mimicking neural network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 10 722–10 731.

[24] S. Shit, J. C. Paetzold, A. Sekuboyina, I. Ezhov, A. Unger, A. Zhylka, J. P. Pluim, U. Bauer, and B. H. Menze, "clDice-a novel topology-preserving loss function for tubular structure segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 16 560–16 569.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[26] 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 16×16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[28] A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic feature pyramid networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6399–6408.

[29] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," in *Proceedings of the European Conference on Computer Vision*. Springer, 2020, pp. 173–190.

[30] T. Wu, S. Tang, R. Zhang, J. Cao, and Y. Zhang, "CGNet: A light-weight context guided network for semantic segmentation," *IEEE Transactions on Image Processing*, vol. 30, pp. 1169–1179, 2020.

[31] W. Zhang, J. Pang, K. Chen, and C. C. Loy, "K-Net: Towards unified image segmentation," in *Proceedings of the Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 10 326–10 338.

[32] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11 976–11 986, 2022.

[33] J. Wei, G. Zhu, Z. Fan, J. Liu, Y. Rong, J. Mo, W. Li, and X. Chen, "Genetic U-Net: Automatically designed deep networks for retinal vessel segmentation using a genetic algorithm," *IEEE Transactions on Medical Imaging*, vol. 41, no. 2, pp. 292–307, 2021.

[34] 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.

[35] S. Qiao, W. Shen, Z. Zhang, B. Wang, and A. Yuille, "Deep co-training for semi-supervised image recognition," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 135–152.

[36] T.-H. Vu, H. Jain, M. Bucher, M. Cord, and P. Pérez, "ADVENT: Adversarial entropy minimization for domain adaptation in semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2517–2526.

[37] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, 2019, pp. 605–613.

[38] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 674–12 684.

[39] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2613–2622.

[40] R. Gu, J. Zhang, G. Wang, W. Lei, T. Song, X. Zhang, K. Li, and S. Zhang, "Contrastive semi-supervised learning for domain adaptive segmentation across similar anatomical structures," *IEEE Transactions on Medical Imaging*, vol. 42, no. 1, pp. 245–256, 2022.

[41] Y. Pu, Q. Zhang, C. Qian, Q. Zeng, N. Li, L. Zhang, S. Zhou, and G. Zhao, "Semi-supervised segmentation of coronary DSA using mixed networks and multi-strategies," *Computers in Biology and Medicine*, vol. 156, p. 106493, 2023.

[42] L. Mou, Y. Zhao, H. Fu, Y. Liu, J. Cheng, Y. Zheng, P. Su, J. Yang, L. Chen, A. F. Frangi, M. Akiba, and J. Liu, "CS^2-Net: Deep learning segmentation of curvilinear structures in medical imaging," *Medical Image Analysis*, vol. 67, p. 101874, 2021.

[43] 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.

[44] Y. Zhang, M. He, Z. Chen, K. Hu, X. Li, and X. Gao, "Bridge-Net: Context-involved U-net with patch-based loss weight mapping for retinal blood vessel segmentation," *Expert Systems with Applications*, vol. 195, p. 116526, 2022.

[45] Y. Ma, Y. Hua, H. Deng, T. Song, H. Wang, Z. Xue, H. Cao, R. Ma, and H. Guan, "Self-supervised vessel segmentation via adversarial learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 7536–7545.

Authorized licensed use limited to: Wuhan University. Downloaded on October 12,2023 at 11:03:31 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.