SCANet: A Unified Semi-Supervised Learning Framework for Vessel Segmentation

Ning Shen, Tingfa Xu, Ziyang Bian, Shiqi Huang, Feng Mu, Bo Huang, *Member, IEEE*, Yuze Xiao, and Jianan Li

***Abstract*— Automatic subcutaneous vessel imaging with near-infrared (NIR) optical apparatus can promote the accuracy of locating blood vessels, thus significantly contributing to clinical venipuncture research. Though deep learning models have achieved remarkable success in medical image segmentation, they still struggle in the subfield of subcutaneous vessel segmentation due to the scarcity and low-quality of annotated data. To relieve it, this work presents a novel semi-supervised learning framework, SCANet, that achieves accurate vessel segmentation through an alternate training strategy. The SCANet is composed of a multi-scale recurrent neural network that embeds coarse-to-fine features and two auxiliary branches, a consistency decoder and an adversarial learning branch, responsible for strengthening fine-grained details and eliminating differences between ground-truths and predictions, respectively. Equipped with a novel semi-supervised alternate training strategy, the three components work collaboratively, enabling SCANet to accurately segment vessel regions with only a handful of labeled data and abounding unlabeled data. Moreover, to mitigate the shortage of annotated data in this field, we provide a new subcutaneous vessel dataset, VESSEL-NIR. Extensive experiments on a wide variety of tasks, including the segmentation of subcutaneous vessels, retinal vessels, and skin lesions, well demonstrate the superiority and generality of our approach.**

***Index Terms*— NIR vessel imaging, semi-supervised learning, medical image segmentation, recurrent neural network.**

I. Introduction

LOCATING blood vessels is an integral and critical step for various diagnostic and therapeutic interventions, yet still remains a labor-expensive manual procedure that relies on the subjective consciousness of clinicians. Guided by visual inspection and palpation of peripheral veins, the practitioner selects and locates a suitable vein to insert the needle. However, such a seemingly straightforward action faces multiple challenges ranging from lack of clinician experience, inferior patient physiology, to large circumstance variation in practice. In visible peripheral vessel access failure, some optical assistant devices and algorithms are required to instruct the assignments profitably. The risks above have driven the development of the near-infrared (NIR) optical apparatus that enhances image contrast and manifests subcutaneous vessels from skin tissue.

Seeing that the images are produced from progressive visual devices, the conjunction of visual devices and unique segmentation algorithms could liberate the workforce of clinicians and solve many tricky problems. In recent years, deep learning models have made up for the deficiency of visual devices and achieved significant progress in medical image segmentation [1] owning to the powerful capability of automated feature learning to extract abstracted feature representations from data [2]. Deep models usually contain a large number of learnable parameters and thus require massive labeled data with pixel-level annotations for training [3]. However, the pixel-level labeling of biomedical data is extremely labor-expensive and time-consuming, such that in a typical case, only a tiny portion of the data is labeled for model training, leaving a large quantity of unannotated data spare [4]. To alleviate the shortage of training data, there are increasing efforts toward semi-supervised segmentation [5] to exploit the valuable information contained in the unannotated data. Semi-supervised segmentation methods first use a small number of labeled data to optimize model parameters and then adopt abounding unlabeled data to finetune the model further to boost performance. However, issues like confirmation bias [6] drifting model prediction towards noise could easily disturb semi-supervised training. Considering the particular property of subcutaneous vein always in local pixel minimum (Fig. 1), accurate vascular segmentation remains challenging.

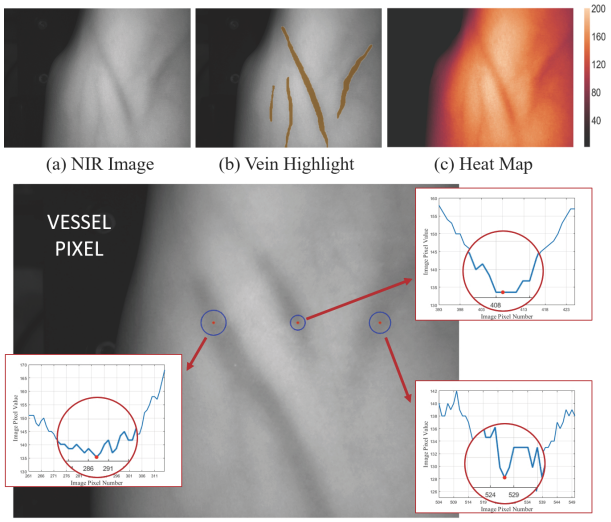
This work presents a novel unified tri-branch framework, SCANet, for semi-supervised peripheral subcutaneous vessel segmentation to tackle the above challenges. Specifically, we first design a multi-scale recurrent neural network equipped with feature fusion as the base segmentation branch (SB). Our motivation stems from the fact that the coarse-to-fine pyramid structure gradually consumes the images of different resolutions and incorporates the useful semantic information into recurrent modules across multiple scales [7]. Combining the predictions from multi-view, feature fusion intensifies the final result outperforming any single scale prediction.

Manuscript received 26 May 2022; revised 14 July 2022; accepted 19 July 2022. Date of publication 21 July 2022; date of current version 31 August 2023. *(Corresponding authors: Tingfa Xu; Jianan Li.)*

Ning Shen, Ziyang Bian, Shiqi Huang, Feng Mu, Bo Huang, Yuze Xiao, and Jianan Li are with the Beijing Institute of Technology, Beijing 100081, China (e-mail: shennbit@163.com; 676286510@qq.com; huangsq@bit.edu.cn; 3120200561@bit.edu.cn; a1039377853@163.com; yuzexiao@bit.edu.cn; lijianan@bit.edu.cn).

Tingfa Xu is with the Chongqing Innovation Center, Beijing Institute of Technology, Beijing 100081, China (e-mail: ciom_xtf1@bit.edu.cn).

Digital Object Identifier 10.1109/TMI.2022.3193150



(a) NIR Image (b) Vein Highlight (c) Heat Map

(d) Vessel Center Pixel

Fig. 1. Examples of (a) NIR arm image, (b) subcutaneous veins highlight, (c) corresponding heat map, and (d) vessel center pixel in local pixel minimum.

SB attempts to individually consume labeled data to train the segmentation model and generate predicted maps.

Second, the consistency decoder branch (CB) uses the midway encoded map and the prediction of SB to synthesize the input sample. Third, the probability map from the segmentation network or label is combined with the original image to train the adversarial learning branch (AB) to narrow the prediction and annotation gap. The motivations of CB and AB are inspired by refining the encoder and decoder of SB respectively. We synthesize the original image and force the encoder of SB to generate representative domain features by successively improving the consistency function. Moreover, the decoder of the segmentation network is usually subject to disturbances like confirmation bias [6], which makes the segmentation result noisy. Therefore, the AB, distinguishing segmentation result from annotation, encourages the decoder to possibly generate the probability map approaching the label. The consistency decoder branch and the adversarial learning branch have complementary advantages to the segmentation branch, thus forming a unified framework SCANet.

The whole training process comprises two steps, *i.e.*, a fully-supervised pretraining stage, and a semi-supervised alternate training stage. The purpose of the former stage is to pretrain the segmentation branch and the consistency decoder branch with a handful of labeled data. The temporary model with earned weights would learn the feature distribution and make a foundation for the follow-up training. The three branches work synergistically based on labeled and unlabeled data in the latter stage. The segmentation branch alternates training with the adversarial learning branch until convergence, while the consistency decoder branch runs through the whole process. It makes the segmentation branch break through the limitation of a small number of annotations and generate more real-like results.

The training of SCANet needs three branches to work together, but the strength of SCANet in running efficiency would boom in the testing phase. The reason is that the consistency decoder branch and the adversarial learning branch in synthesizing the input image and giving the prediction value could be ignored in testing image segmentation. Relying on the segmentation branch alone gives excellent results on various datasets. In this case, the network parameters are 29.64 M, and the processing time for a single image is only 44 ms.

In addition, we build a peripheral subcutaneous vessel dataset VESSEL-NIR to forward vessel imaging processing development, which is impeded by a lack of relevant data. Specifically, we use the guidance of the NIR optical apparatus to generate 3600 samples from different unique subjects. The adopted wavelength range of 760 to 1000 nm allows the device to provide a noticeable increase in vessel image contrast and observe the vessels beneath the skin to 3 mm.

The effectiveness of the proposed method is evaluated on VESSEL-NIR, four retinal vessel datasets (DRIVE [8], STARE [9], CHASE_DB1 [10], and HRF [11]), and one skin lesion dataset (ISIC2018 [12]). The SCANet overtakes most state-of-the-art segmentation models and achieves the competitive results on these medical image datasets, particularly on VESSEL-NIR with the best segmentation performance in DSC of 87.16% and IOU of 77.78%.

To sum up, the main contributions of this work are:

- We design a novel tri-branch semi-supervised semantic segmentation network relying on multi-scale recurrent neural network, consistency decoder, and adversarial learning, achieving outstanding segmentation performance with a small number of labeled data and abounding unlabeled data. The whole training process includes a fully-supervised pretraining and a semi-supervised alternate training stage.
- The multi-scale recurrent neural network accompanied by ConvLSTM units is the primary assignment of extracting semantic features for segmentation. In addition, by integrating predictions from multi-view, the feature fusion algorithm generates better per-pixel segmentation results than any single scale.
- A brand-new subcutaneous vessel dataset VESSEL-NIR[1] is presented in this paper. Extensive experiments on six medical datasets demonstrate that the proposed SCANet outperforms multiple state-of-the-art segmentation models. Some samples of our VESSEL-NIR and source code are available at the link above, and more images will be released soon.

II. Related Works

A. Recurrent Neural Network

Over the past few decades, recurrent neural network (RNN) has come to immense fruition and extensive applications in numerous Natural Language Processing domains. The community development is ascribed to functional characteristics of recursive processing and modeling of historical memory, which are suitable for conducting strong correlations in time and space sequences. RNN is primarily in effect because

[1] https://github.com/shennbit/VESSEL-NIR

of the faculty to overcome the restrictions on traditional approaches [13]. Hopfield [14] introduced a pioneering RNN, which has pattern recognition capability to recover a stored pattern from a corrupted version. Introducing the memory cells to LSTM, ConvLSTM [15] helps the network get around the problems like vanishing gradient and exploding gradient. The popularity of RCNN could be attributable to the incorporation of recurrent connections into each convolutional layer [16]. By integrating the context information, RCNN makes the model deeper and keeps the number of shared parameters static [17].

B. Semi-supervised Medical Image Segmentation

In the condition of labeled data scarcity, semi-supervised segmentation could produce a marked effect [18], [19]. It attempts to utilize part of labeled data and abounding unlabeled data to get more precise results than supervised segmentation [20]. In the sub-fields of semi-supervised segmentation, self-ensembling model and adversarial learning network inspire the area evolution rapidly. The self-ensembling model exploits the student and teacher model to update alternately. Based on the information in the intermediate training steps, the synthetic result could be more splendid than the single latest model. Zhao *et al.* [21] designed a novel self-ensemble architecture and a random patch size training strategy exploiting plenty of unlabeled data. Making use of the organ and lesion images with bounding-box level annotations, Sun *et al.* [22] exploited the teacher-student fashion to form a hierarchical organ-to-lesion attention module. As the core of the adversarial learning network, the discriminator is encouraged to give different scores between ground-truths and segmentation results produced by the generator. The latent space factorization [23], relying on the cycle consistency principle, uses part of labeled images with many unlabeled images together to train a myocardium segmentation neural network. The combination of adversarial training and self-training classification was excavated by Mittal *et al.* [24], whose dual-branch approach reduces both the low-level and the high-level artifacts based on a few labeled data. Regardless of the methods above, many other excellent models come up continuously [25]. UATS [26] combines temporal ensembling with uncertainty-guided self-learning to leverage frequently available unlabeled data. The model of PRS^2 [27] consists of a segmentation network and a pairwise relation network, then the shared encoders transfer the image representations to improve the segmentation performance.

C. Technologies for Vessel Processing

In the community of medical image analysis, vessel processing is indispensable to many clinical diagnoses and manipulations [28]. Significantly, the challenge of difficult subcutaneous vessel access obsesses many clinical workers, and the related sparse technologies make the community develop sluggishly. Li *et al.* [29] proposed a new convex-regional-based gradient model to utilize contextually related regional information to locate and segment vessels in NIR guidance. In the work of [30], the authors built a non-invasive custom optical imaging system to survey the reliability of blood oxygen saturation levels recovered from spectral images. As a new class of mixed-function, the combination of U-Net and reinforcement learning guarantees proper alignment of the projected image regardless of angle, translation, and scale offsets between the NIR measurement and the visible projection [31]. The designed model of [32] is based on a fully recurrent convolutional network, which attempts to capture salient image features and motion signatures at multiple resolution scales. Wei *et al.* [33] devised a condensed but flexible search space to operate retinal vessel segmentation with fewer parameters. In [34], a feature recalibration module and an attention distillation module were presented for accurate airway and artery-vein segmentation.

III. Method

A. Overview

This section provides descriptions from two perspectives, the tri-branch synergy and the training strategy of SCANet, to overview the complete process.

1) Tri-Branch Synergy: SCANet deploys the segmentation branch (SB), the consistency branch (CB), and the adversarial branch (AB) as a unified framework. The SB performs the main semantic segmentation task by equipping a multi-scale recurrent neural network with feature fusion. This branch decomposes the input image into two components: the binary segmentation result representing spatial information and the midway map encoded by the recurrent unit containing high-level concentrated imaging feature. The CB consumes the two representations above and aims to synthesize the original input images. The adversarial branch is responsible for distinguishing between the generated segmentation result and the ground-truth. It concatenates the binary map and corresponding image as the input and is encouraged to determine 1 for ground-truth label maps and 0 for segmentation results derived from the segmentation branch. It is worth noting that the consistency and adversarial branches do not work in the testing phase, as synthesizing the image and giving the prediction value could be omitted for testing image segmentation. The testing consumption of model parameter quantity and running time is parsimonious in this schedule.

2) Training Strategy: The whole training process of SCANet (Fig. 2) contains two steps, *i.e.*, a fully-supervised pretraining stage and a semi-supervised alternate training stage. In the first stage, the segmentation and the consistency branches are pretrained under the guidance of labeled data. The temporary supervised model with earned weights promises the best segmentation performance in the condition of training labeled data. Then, the initialized weight framework of the segmentation branch adopts an alternate training strategy with the adversarial branch, while the consistency branch runs through the whole process. The prediction quality is improved in this manner by using unlabeled data and labeled data. The semi-supervised mode further refined the segmentation branch, including intrinsic supervised learning and alternate training. Finally, the alternate training is repeated for several epochs and generates the maximum weights to obtain high-quality

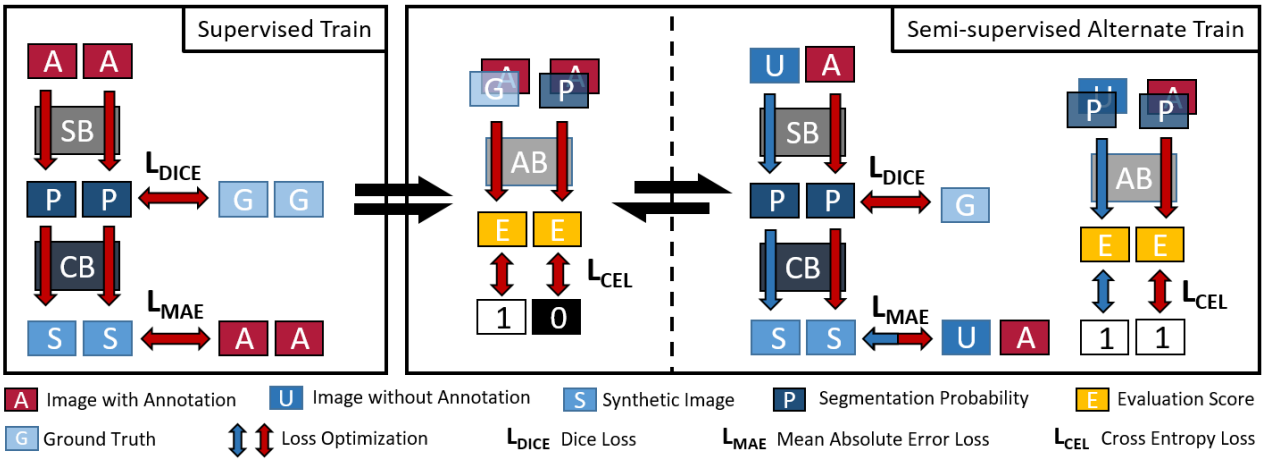


Fig. 2. Illustration of the overall training process. First, the SB and CB in the supervised training consume annotated images to learn the feature distribution and build the follow-up training. Then, the alternate training of AB, CB, and SB work synergistically in a semi-supervised manner. SB: Segmentation Branch, CB: Consistency Branch, AB: Adversarial Branch.

Algorithm 1 The Training Process of SCANet

Input: Dataset $\boldsymbol{S} = \boldsymbol{S_L} \cup \boldsymbol{S_U}$, ending steps k, segmentation branch $\boldsymbol{f_s}(\cdot)$, consistency branch $\boldsymbol{f_c}(\cdot,\cdot)$, adversarial branch $\boldsymbol{f_a}(\cdot,\cdot)$.
Output: Trained model SCANet M.

Fully-supervised pretraining stage:
Sample labeled dataset $\{(\boldsymbol{X_i}, \boldsymbol{Y_i})\}_{i=1}^{n}$ by $\boldsymbol{S_L}$;
Latent map $\boldsymbol{Z_i}$, segmentation map $\boldsymbol{P_i} \leftarrow \boldsymbol{f_s}(\boldsymbol{X_i})$;
Synthesized image $\boldsymbol{S_i} \leftarrow \boldsymbol{f_c}(\boldsymbol{Z_i}, \boldsymbol{P_i})$;
Pre-train $\boldsymbol{f_s}$ and $\boldsymbol{f_c}$ of model M;
Semi-supervised alternate training stage:
Sample unlabeled dataset $\{\boldsymbol{X_j}\}_{j=n+1}^{n+m}$ by $\boldsymbol{S_U}$;
for $1, 2, 3, \cdots, k$ **do**
  Evaluation score $1 \leftarrow \boldsymbol{f_a}(\boldsymbol{X_i}, \boldsymbol{Y_i})$;
  Evaluation score $0 \leftarrow \boldsymbol{f_a}(\boldsymbol{X_i}, \boldsymbol{P_i})$;
  Train $\boldsymbol{f_a}$ of model M;
  $\boldsymbol{Z_j}, \boldsymbol{P_j} \leftarrow \boldsymbol{f_s}(\boldsymbol{X_j})$;
  $\boldsymbol{S_j} \leftarrow \boldsymbol{f_c}(\boldsymbol{Z_j}, \boldsymbol{P_j})$;
  $1 \leftarrow \boldsymbol{f_a}(\boldsymbol{X_j}, \boldsymbol{P_j})$;
  $\boldsymbol{Z_i}, \boldsymbol{P_i} \leftarrow \boldsymbol{f_s}(\boldsymbol{X_i})$;
  $\boldsymbol{S_i} \leftarrow \boldsymbol{f_c}(\boldsymbol{Z_i}, \boldsymbol{P_i})$;
  $1 \leftarrow \boldsymbol{f_a}(\boldsymbol{X_i}, \boldsymbol{P_i})$;
  Train $\boldsymbol{f_s}$ and $\boldsymbol{f_c}$ of model M.
end

segmentation maps. The pseudo-codes in Algorithm 1 clarify the training process of SCANet.

B. Key Network Components

The SCANet is a unified tri-branch framework, the key components of which work synergistically to improve model performance. We show the key components of SCANet in Fig. 3 and recount the architectures as follows.

1) Segmentation Branch: The SB, the principle part of SCANet, undertakes the peripheral subcutaneous vessel segmentation. It combines a multi-scale recurrent neural network with feature fusion to generate per-pixel class predictions. We adopt U-Net [35]/DeepLabV3+ [36] as the base network of SB, whose function is to complete feature encoding and decoding. In this section, the description of SCANet is based on U-Net for brevity. The architecture of SB is a tri-level encoder-decoder segmentation network with recurrent units, as shown in Fig. 3. Each level consists of four convolutional layers with ReLU activation, one ConvLSTM module [15], and three deconvolutional layers. Three Resnet-Blocks follow each layer above. To recover the lost fine-grained spatial information, copy connection links the feature of the convolutional layer to that of the deconvolutional layer by matrix addition. The tail is a convolutional layer with Sigmoid activation, which implements pixel-level classification.

The LSTM-related algorithms are generally applied to temporal data processing. ConvLSTM, added to the multi-level segmentation network for the global feature transfer, is more effective in feature extraction of images. Passing the long short-term memory to the next level is the main function of ConvLSTM. Besides, the motivation for using ConvLSTM is to optimize the encoded feature maps from different levels. Considering the tri-level architecture of SB, ConvLSTM benefits from intensifying coarse-to-fine feature information. Cell state $\boldsymbol{C_l}$ and hidden state $\boldsymbol{H_l}$ undertake the long and short term memory respectively. l is the level index. The structure $\boldsymbol{g}$ of ConvLSTM in Fig. 3 is:

$$C_l, H_l = g(C_{l-1}, H_{l-1}, E_l), \tag{1}$$

where $\boldsymbol{E_l}$ denotes the feature from the encoder of level l. In the first level, the inputs of ConvLSTM are inadequate due to the lack of long or short term memory from previous levels. Compromising in initialization, the cell state $\boldsymbol{C_0}$ and hidden state $\boldsymbol{H_0}$ are superseded by null matrices. As the short-term memory is suited to run at the same level, the hidden state $\boldsymbol{H_l}$ is sent to the decoder of the corresponding level.

Accumulating the segmentation maps from the above levels, the third level of SB would bring out the final prediction. The feature fusion algorithm in the third level makes the result

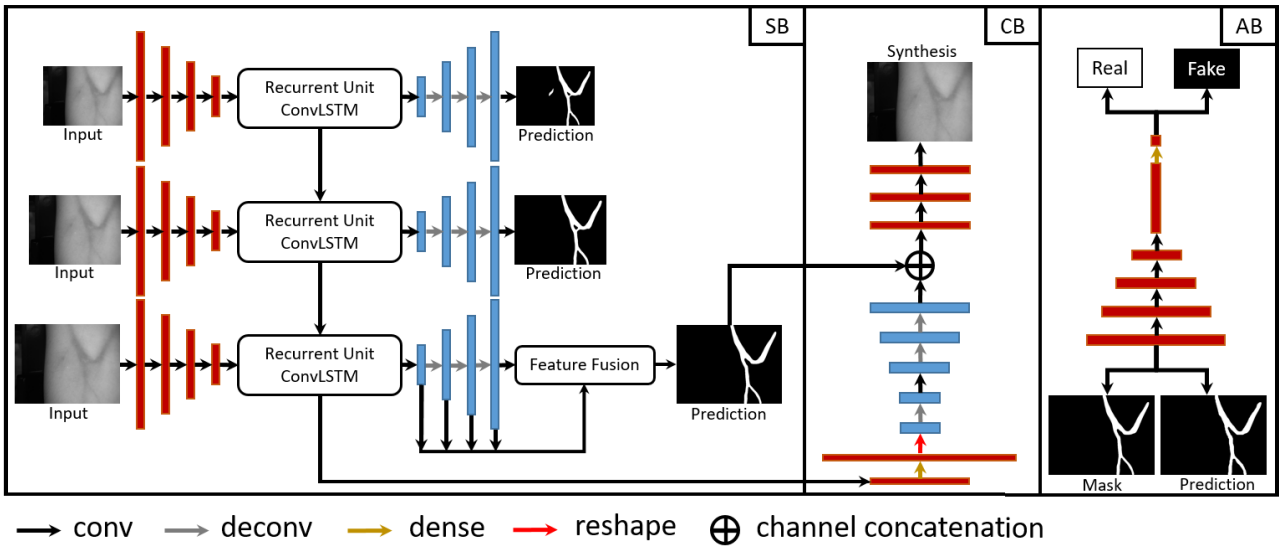


Fig. 3. Overall framework of SCANet. The SB decomposes input image into the binary segmentation result and the midway feature map encoded by recurrent unit. The CB consumes the two mentioned representations and aims to synthesize the original input image. The AB is encouraged to distinguish between segmentation result and ground-truth by giving them different scores.

outperform any single prediction. The output features $\boldsymbol{R_s}$ are resized to the size of the input image and fused recursively:

$$R_s = u(p(F_{s-1})) + p(F_s), \tag{2}$$

where s is the s-th scale in the third level, $\boldsymbol{F_s}$ is the corresponding decoder map, $\boldsymbol{u(*)}$ up-samples $*$ by rate 2, and $\boldsymbol{p(F_s)}$ is the prediction map in the s-th scale.

The loss of SB is the sum of three levels, each of which optimizes respective parameters simultaneously. The sizes of input images are $(w/4, h/4)$, $(w/2, h/2)$ and (w, h). In the first two levels, the input $\boldsymbol{X_l}$ and ground-truth $\boldsymbol{Y_l}$ are down-sampled by the weighted average of pixels in the nearest 2-by-2 neighborhood. l is the level index. Then, the loss $\boldsymbol{L_{SB}}$ adopts the Dice function $\boldsymbol{d}$ to measure the dissimilarity between the produced segmentation output $\boldsymbol{f(\cdot, \theta)}$ and the ground-truth:

$$L_{SB} = \sum_{l=1}^{3} (1 - d(f(X_l, \theta), Y_l)), \tag{3}$$

$$d(f(X_l, \theta), Y_l) = 2 \times [\frac{\sum_{i=1}^{p} (f(X_l, \theta)_i \times Y_{li})}{\sum_{i=1}^{p} (f(X_l, \theta)_i + Y_{li})}], \tag{4}$$

where p is the pixel number of original image, θ is the weight.

2) Consistency Branch: The auxiliary CB guides the two components of SB, the binary segmentation map and the latent representation from the recurrent unit, to synthesize the original input image. The mean absolute error (MAE) between the synthetic and original image is calculated for network optimization. The loss function $\boldsymbol{L_{CB}}$ of CB is:

$$\begin{aligned} L_{CB} &= L_{MAE}(X, r(Z, M, \theta)) \\ &= \frac{1}{p} \sum_{i=1}^{p} (X_i - r(Z, M, \theta)_i), \end{aligned} \tag{5}$$

where θ is the weight, p is the pixel number of original image, $\boldsymbol{M}$ is the segmentation mask of $\boldsymbol{X_i}$, and $\boldsymbol{r(\cdot)}$ represents the consistency network. The latent map $\boldsymbol{Z}$ is passed through the ConvLSTM module and learns more abstracted features from the encoders of SB, as shown in Fig. 3. The high-level semantic feature $\boldsymbol{Z}$ fed into CB not only decides the consistency quality of synthesis but also helps get better segmentation further. The optimization of CB refines $\boldsymbol{Z}$ to contain valid information instead of noise throughout the whole training process. The latent representation filtered by dense blocks and up-sampling layers concatenates with the segmentation map to synthesize the raw image. Our model is encouraged to learn discriminative features for segmentation by benefiting from labeled and unlabeled images. The architecture of CB consists of two dense blocks, two up-sampling layers, three convolutional layers, and three Resnet-Blocks.

3) Adversarial Branch: The task of AB is to match the distribution statistics of predictions and ground-truths. It takes the concatenation of original images and probability maps as input. The probability maps include the binary annotation images and the predicted segmentation maps from SB. In this way, the AB module can learn to distinguish between manual and generated segmentation maps and output the corresponding prediction values. 0 should be assigned to the predicted segmentation maps and 1 to the ground-truths.

The AB consists of four down-sampling blocks and one fully connected dense layer with Sigmoid activation. Each down-sampling block comprises convolutional layer (kernel size 3, stride 2), Leaky-ReLU activation function (parameterized by 0.2), and Batch-Normalization layer. The fully connected layer integrates the features in the front and gives the final verdict (real or fake). We use $\boldsymbol{a(X, Y)} \in [0, 1]$ to denote the scalar probability of the prediction ($\boldsymbol{Y}$ is the ground-truth of $\boldsymbol{X}$). n is the number of annotated samples. Training AB is equivalent to minimizing the following binary

classification loss $\boldsymbol{L_{AB}}$:

$$L_{AB} = L_{CEL}(a(X_n, Y_n), 1) + L_{CEL}(a(X_n, f(X_n, \theta)), 0), \quad (6)$$

$$L_{CEL}(P, Y) = -\frac{1}{k}\sum_{i=1}^{k}[Y_i \cdot log(P_i) + (1 - Y_i) \cdot log(1 - P_i)], \quad (7)$$

where θ is the parameters of SB architecture $\boldsymbol{f(\cdot, \theta)}$, $\boldsymbol{L_{CEL}}$ is the Cross Entropy Loss (CEL). The balance between the numbers of segmentation maps $\boldsymbol{P_i}$ and ground-truths $\boldsymbol{Y_i}$ controls AB to perform prediction stably, k is the number of maps $\boldsymbol{Y_i}$ and $\boldsymbol{P_i}$.

C. Objectives

The training process of SCANet comprises of a fully-supervised pretraining stage and a semi-supervised alternate training stage. The objective functions of these stages are described as follows.

1) Supervised Training Loss: In the fully-supervised pretraining stage, the objective function is the supervised loss $\boldsymbol{L_{sup}}$ which uses the weighted sum of $\boldsymbol{L_{SB}}$ and $\boldsymbol{L_{CB}}$. More specifically, $\boldsymbol{L_{sup}}$ is defined as follows:

$$L_{sup} = \alpha_1 * L_{SB} + \alpha_2 * L_{CB}. \quad (8)$$

Based on the labeled samples, $\boldsymbol{L_{sup}}$ optimizes the segmentation branch and the consistency branch until convergence. It aims to obtain the initial weights and generates segmentation probability maps. α_1 and α_2 adjust the weights between $\boldsymbol{L_{SB}}$ and $\boldsymbol{L_{CB}}$. As the segmentation branch is mainly adopted to learn the distribution of ground-truths and generate the semantic segmentation maps, α_1 is given a larger value to emphasize the relevant loss $\boldsymbol{L_{SB}}$.

2) Alternate Training Loss: In the alternate training process, labeled and unlabeled data are used at the same time. $\boldsymbol{L_{AB}}$, $\boldsymbol{L_{SB}}$, and $\boldsymbol{L_{CB}}$ work synergistically to improve model performance. The three losses are minimized in an alternating fashion until the adversarial branch cannot easily distinguish between ground-truth and the prediction of the segmentation branch. The whole training process is run through by $\boldsymbol{L_{CB}}$. It consumes the input data as a powerful assistant loss function, whether labeled or not. The alternate training loss $\boldsymbol{L_{semi}}$ as the objective function in our semi-supervised framework is:

$$L_{semi} = \lambda_1 * L_{SB} + \lambda_2 * L_{CB} + \lambda_3 * L_{AB}. \quad (9)$$

Since there is no one-to-one ground-truth for the unlabeled data, $\boldsymbol{L_{SB}}$ can not be used to optimize the model in an unsupervised way. The $\boldsymbol{L_{AB}}$ is still applicable because it could work in the guidance of ground-truth and prediction generated in the fully-supervised pretraining stage. Besides, $\boldsymbol{L_{CB}}$ is operated to generate the initial weights in the model initialization stage and optimizes the network parameters in the alternate training stage. The minimizing of loss $\boldsymbol{L_{SB}}$ helps SCANet extract semantic features for segmentation and further give an effective prediction. If the supervised fraction enters the bottleneck period, the training would be optimized by $\boldsymbol{L_{CB}}$ and $\boldsymbol{L_{AB}}$ to increase the reliability of the predictions.

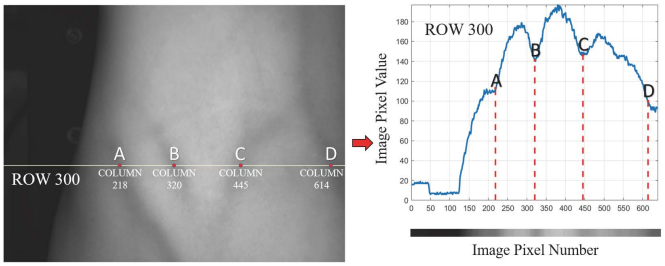


Fig. 4. Cross-sectional gray-level profile of row number 300.

As a weighted sum of $\boldsymbol{L_{SB}}$, $\boldsymbol{L_{CB}}$, and $\boldsymbol{L_{AB}}$, the alternate training loss $\boldsymbol{L_{semi}}$ controls the impact of individual loss to the overall objective function and adjusts the value balance between λ_1, λ_2, and λ_3. A good balance could avoid the issues caused by the distribution diversity of many unlabeled data. The ablation research about the branch weightage is described in Section V-D.

IV. Dataset

A. Data Collection

In this work, we collected a subcutaneous vessel dataset (VESSEL-NIR), including oodles of labeled and unlabeled samples. As a part of the project for autonomous venipuncture research, VESSEL-NIR contains 3600 NIR images from different unique subjects. The study is approved by the Key Laboratory of Photoelectronic Imaging Technology and System, Ministry of Education of China. The NIR imaging technology, providing a noticeable increase in vessel contrast compared to standard visualization, is utilized to visualize invisible subcutaneous vessels. A NIR-sensitive CCD camera takes the stored 8-bit greyscale images of VESSEL-NIR with a cell size of 5.5 μm, flame rate of 25 Hz, power driver of DC 12 $\pm$ 2 V and room temperature of 24 degrees.

B. Data Annotation

Two professional surgeons manually marked the annotated images in two rounds. First, an attending physician marked the ground-truth regions with the technical MITK software tool[2] for image analysis and annotation. Then, the annotations were further checked and corrected by a chief physician in surgery.

C. Data Statistics

The vessel cross-sectional gray-level profile of row number 300 is shown in Fig. 4. The blue curve on the right represents the original gray pixel values. The points A, B, C, and D denote the center positions of vessels, which are at local pixel minimum without pixel continuity and similarity in most cases. The red curve and bars in Fig. 5(a) illustrate the vessel pixel percentage distribution in VESSEL-NIR. The vessel size distribution in Fig. 5(b) is an vital specialty hinging on the effect of deep learning algorithms. The linear descent of the red curve represents that the majority of vessels are not palpable. Based on the data specialty and distribution above, we employ SCANet to segment subcutaneous vessels.

[2]http://www.mitk.org

TABLE I

COMPARISON OF PREDICTIVE PERFORMANCE, PARAMETER NUMBERS, AND INFERENCE TIME PER IMAGE. SCANET_D AND SCANET_U DENOTE THE MODEL BUILT ON DEEPLABV3+ AND U-NET, RESPECTIVELY. TESTS FOR STATISTICAL SIGNIFICANCE (*t*-TEST) OF DSC ARE ALSO PROVIDED. $*/\circ$ INDICATES STATISTICAL DIFFERENCE BETWEEN SCANET_U/SCANET_D AND OTHER METHODS. ($*/\circ : p \leq 0.05, **/\circ\circ : p \leq 0.01, ***/\circ\circ\circ : p \leq 0.001$)

	Fully-supervised				Semi-supervised					
	APCNet [37]	DeepLabV3+ [36]	DANet [38]	PointRend [39]	UAMT [40]	ICT [41]	URPC [42]	CPS [43]	SCANet_U	SCANet_D
DSC (%)	82.99	83.86	84.62	83.83	84.05	84.59	83.24	83.68	**87.16**	**86.74**
IOU (%)	71.69	72.92	74.03	72.88	72.69	73.87	71.54	72.14	**77.78**	**77.06**
PAC (%)	97.08	97.17	97.28	97.16	97.47	97.51	97.33	97.37	**97.85**	**97.85**
Rec (%)	74.20	76.23	77.74	76.22	80.06	82.53	80.32	80.60	**87.22**	**87.11**
Pre (%)	96.09	95.02	94.62	94.89	90.08	88.70	88.13	88.51	**88.44**	**87.41**
Time (ms)	99	95	98	90	45	45	46	45	**44**	**95**
Parameter (M)	59.97	55.91	57.48	49.50	29.64	29.64	29.76	29.64	**29.64**	**55.91**
t-Test	$**_{\circ\circ}$	$*_{\circ}$	$*_{\circ}$	$**_{\circ}$	$***_{\circ\circ\circ}$	$***_{\circ\circ\circ}$	$***_{\circ\circ\circ}$	$***_{\circ\circ\circ}$	-	-

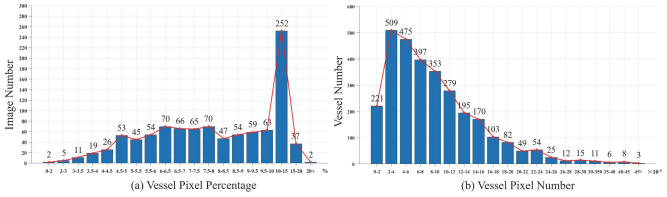


Fig. 5. Illustration of (a) vessel pixel percentage distribution, (b) vessel size distribution.

D. Data Comparison

To our knowledge, VESSEL-NIR is the first open large-scale dataset for subcutaneous vessel segmentation with 3600 vessel images in the pixel resolution of 960 × 1280. The wavelength range of 760 to 1000 nm allows the NIR imaging device to observe the vessels beneath the skin to 3 mm. In this way, the device provides a noticeable increase in vessel image contrast. In addition, an existing subcutaneous vessel dataset is presented by Leli *et al.* [31] and acquired by an IR camera (Raspberry NoIR camera V2). It includes 320 forearm images cropped to the pixel resolution of 512 × 512. The adopted wavelength range of 750 to 900 nm illuminates the area of interest on the forearm. The comparison between VESSEL-NIR and the involved dataset shows that our VESSEL-NIR could provide better flexible applicability in the vessel segmentation-related tasks.

V. EXPERIMENTS AND RESULTS

A. Experimental Settings

1) Implementation Details: Our framework is implemented in Python with the TensorFlow library under the hardware support of two NVIDIA 1080Ti GPUs. We conduct our experiments on six medical image datasets and use U-Net and DeepLabV3+ as the base networks of SCANet. The batch size is 6, containing two annotated and four unannotated images. All training images are resized to the image pixel resolution of 480 × 640. The tri-branch architecture is trained using Adam optimizer with default parameters settings ($\beta_1 = 0.9$ and $\beta_2 = 0.99$). The hype-parameters λ_1, λ_2, and λ_3 of the semi-supervised alternate training are set to 10, 3, and 3. Aligned with semi-supervised alternate training, the α_1 and α_2 of the fully-supervised training are also assigned to 10 and 3. It is important to note that some settings for individual comparison methods are different from others to display the best performance.

2) Evaluation Metrics: Following the literature on vessel segmentation, we use five widely adopted metrics in the evaluation of VESSEL-NIR, including Pixel-wise Accuracy (PAC), Recall (Rec), Precision (Pre), Intersection-Over-Union (IOU), and Dice Similarity Coefficient (DSC). We also exploit three common-used statistical metrics, Sensitivity (Sen), Specificity (Spe), and Accuracy (Acc), to quantitatively evaluate retinal vessel segmentation performance. In the skin lesion dataset experiments, IOU and DSC are employed as the segmentation evaluation metrics.

B. Results of Subcutaneous Vessel Segmentation

1) Data: The available VESSEL-NIR dataset contains 3600 samples. In the semi-supervised manner of subcutaneous vessel segmentation, the 1600 annotated samples are divided into 800, 200, and 600 for training, validation, and testing. The 2000 unannotated samples are used for unsupervised correction of the two assistant branches.

2) Setups: The training epochs of the fully-supervised stage and semi-supervised alternate training stage are fixed to 60 and 300. The learning rate is initialized as $1 \times e^{-5}$ and decays with a momentum of 0.5 with patience of 10 epochs no promotion.

3) Fully-Supervised Results: We implement current state-of-the-art fully-supervised segmentation methods (APCNet [37], DeepLabV3+ [36], DANet [38], and PointRend [39]) on VESSEL-NIR for comparison to verify the effectiveness of SCANet. Quantitative results of the testing set are shown in Table I. All the fully-supervised methods are effective in vessel segmentation, and SCANet achieves competitive performances in most evaluation metrics. Moreover, SCANet_D or SCANet_U denotes the framework using DeepLabV3+ or U-Net as the base network, and the latter tables keep this description.

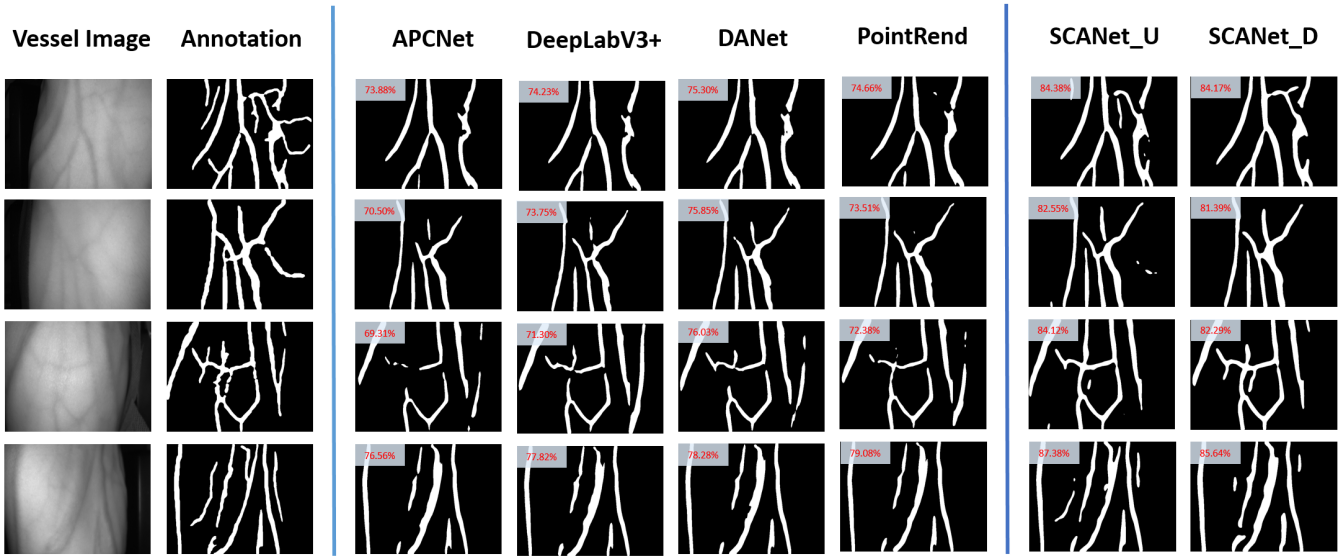


Fig. 6. Qualitative comparison of subcutaneous vessel segmentation with fully-supervised methods. Corresponding DSC score is listed in the top-left corner. SCANet_D and SCANet_U denote the model built on DeepLabV3+ and U-Net, respectively.

Since our task is to perform segmentation on subcutaneous vessel images, the APCNet with multiple well-designed adaptive context modules gets an excellent performance in Pre of 96.09%. In DeepLabV3+, the proposed encoder-decoder structure captures multi-scale features and obtains the vessel segmentation results in Rec of 76.23%, IOU of 72.92%, and DSC of 83.86%. With the integration of two an excellent and flexible attention modules, DANet generates accurate scene segmentation and demonstrates excellent results in DSC of 84.62% and IOU of 74.03%. The PointRend presents image segmentation as a rendering problem, and its results (83.83% DSC, 72.88% IOU) testify to the algorithm universality. Trained in the semi-supervised mode, the proposed SCANet_U gets the best performance in DSC of 87.16%. Besides, the SCANet outperforms all state-of-the-art fully-supervised segmentation methods in terms of DSC, IOU, and Rec by a large margin. We attribute this improvement to the semi-supervised alternate strategy with robust feature representations.

4) Semi-Supervised Results: To further evaluate SCANet, we consider four state-of-the-art semi-supervised models (UAMT [40], ICT [41], URPC [42], and CPS [43]) for quantitative comparisons. The test results of these methods on VESSEL-NIR are presented in Table I. The quantitative results indicate that all semi-supervised methods are competitive in vessel segmentation, and SCANet_U gets the best performance in PAC, Rec, IOU, and DSC.

Encouraging consistent predictions of the same input samples under different disturbances, UAMT even outperforms SCANet slightly in term of Pre. ICT is an efficient consistency regularization technique and achieves great segmentation results in several indices. URPC shows good performances in all evaluation metrics. CPS feeds labeled and unlabeled images into two segmentation networks that share the same structure and adds constraints to make the outputs of the two networks similar to the input samples. It achieves 83.68%, 72.14%, 80.60%, and 97.37% in terms of DSC, IOU, Rec, and PAC, respectively.

5) Analysis: The comparison of total algorithms indicates that SCANet outperforms markedly other fully-supervised and semi-supervised models. We also run two-tail paired t-tests in DSC to analyze the statistical significance between SCANet and other competing methods. p-value less than 0.05 demonstrates a significant difference between comparing algorithms. The asterisk or circle of the bottom row in Table I illustrates the statistically significant improvement in SCANet. The success of SCANet is owed to the tri-branch architecture in semi-supervised alternate training mode. The excellent results of SCANet_U and SCANet_D, overtaking the other fully-supervised and semi-supervised methods in terms of IOU and DSC by a large margin, demonstrate the flexibility of SCANet in base network (U-Net, DeepLabV3+).

6) Efficiency: The model efficiency is researched from two perspectives, *i.e.*, model parameter quantity and time complexity. Though SCANet comprises many moving parts and complicated components, the testing phase is intelligent and flexible. The auxiliary consistency and adversarial branches, synthesizing the input image and giving the prediction value, do not work then. The testing consumption of SCANet_U achieves 29.64 M parameter quantity and 44 ms per image as shown in Table I. The strength of SCANet is reflected in harvesting both superb performance and high efficiency simultaneously. Most of the semi-supervised comparison methods used need a part of the model to perform segmentation for testing, and they are saving and similar in computation consumption.

7) Visualization: Overall, the visualization results between SCANet and the existing state-of-the-art models are shown in Fig. 6 and Fig. 7. As can be seen, SCANet yields segmentation results with accurate boundaries better than other methods. The performances on thin or highly curved crossover vessel structures are presented in Fig. 8 and marked with red boxes.

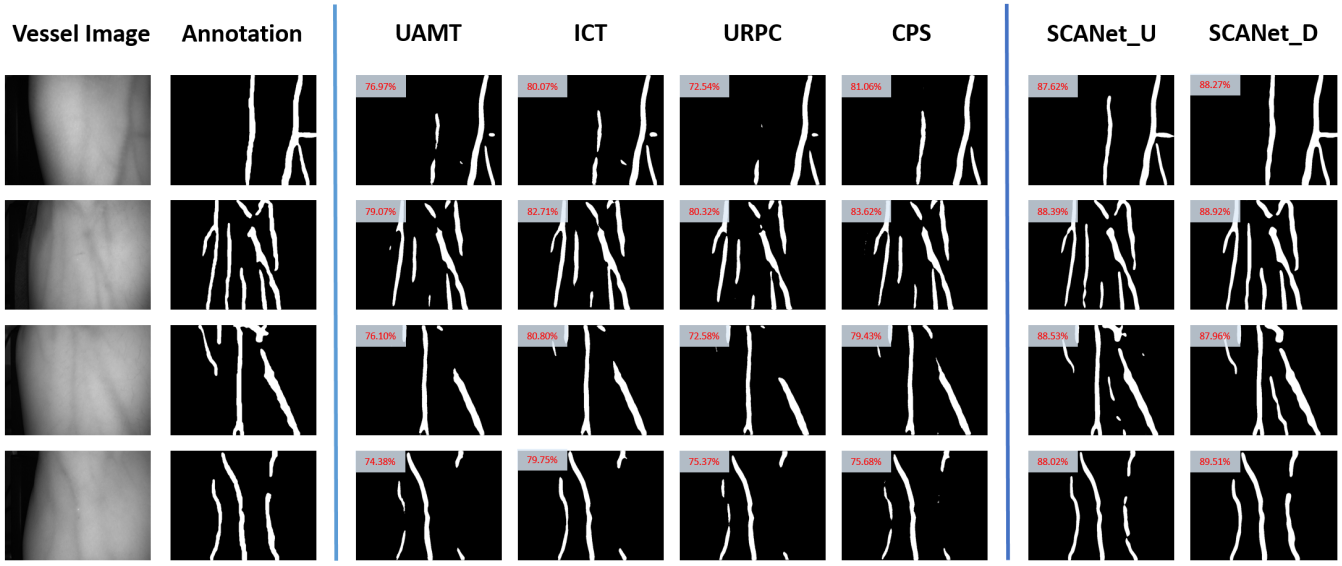


Fig. 7. Qualitative comparison of semi-supervised subcutaneous vessel segmentation. The corresponding DSC score is listed in the top-left corner.

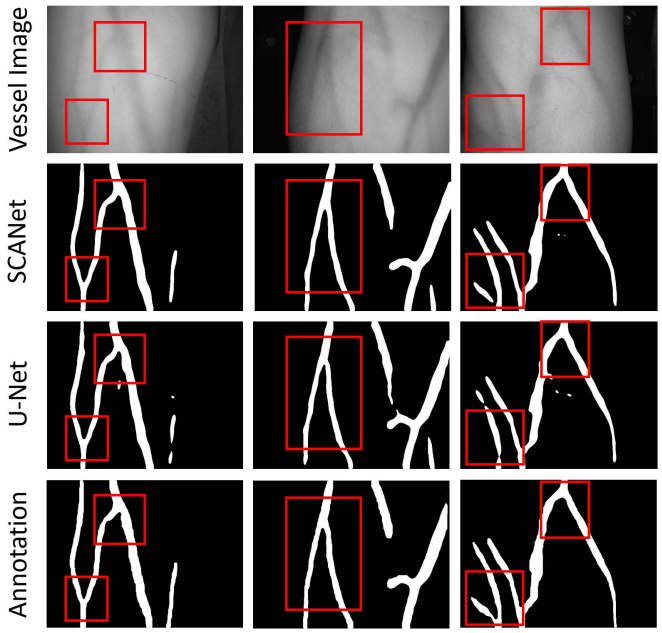


Fig. 8. Visualization of thin and highly-curved crossover vessels in original images, predictions of SCANet and U-Net, and corresponding ground-truths. Key segments marked with red boxes.

The proposed SCANet also maintains excellent predictions in acute problems with low parameter quantity and time consumption.

C. Results of Retinal Vessel Segmentation

1) Data: We implement SCANet on retinal vessel datasets for generalization evaluation, including DRIVE [8], STARE [9], CHASE_DB1 [10], and HRF [11]. The DRIVE dataset consists of 40 fundus images centered on the macula. Each image is acquired using a Canon CR5 non-mydriatic 3-CCD camera and has a dimension of 565 × 584. The STARE dataset is one of the most commonly used fundus icon standard libraries. It contains 20 color fundus images for retinal vascular segmentation. Each image is captured by a Topcon TRV-50 fundus camera and has a dimension of 700 × 605. The CHASE_DB1 dataset includes 28 retinal images taken from the left and right eyes of 14 schoolchildren. Each image has a dimension of 999 × 960. In HRF, there are 45 fundus images from healthy children, diabetic retinopathy, and glaucoma patients. The assignments of training/validation/testing are 25/5/10, 12/4/4, 16/4/8, and 15/5/25 separately.

2) Setups: Noted that the numbers of the images in these datasets are inadequate for semi-supervised learning, which requires abounding data. Compromising in data number, the images trained for the segmentation branch (SB) are also transmitted to the adversarial branch (AB) and consistency branch (CB) for the assistant tasks of adversarial learning and image synthesizing, respectively. There is no interference or negative influence on the synergy between SB, AB, and CB. Sharing image in three branches is an excellent gambit to the dataset with insufficient images and verifies the tri-branch model. The learning rate is initialized as $5 \times e^{-4}$ and decays with a momentum of 0.5 with the patience of 10 epochs with no promotion. The training process performs 50 epochs in the supervised training stage and 100 epochs in the alternate training stage.

3) Results: We conduct comparisons with several latest frameworks to evaluate the effectiveness of SCANet. Table II presents the performances of comparison algorithms and SCANet on three retinal vessel datasets. A comprehensive comparison of index performances between SCANet_D and SCANet_U indicates that DeepLabV3+ is a more suitable base network for retinal vessel segmentation. The Sen of SCANet_D in STARE is significantly higher than other comparison algorithms, especially the most enormous gap is even 7.85%. Through the results of CHASE_DB1 in Table II, SCANet_D has the edge on all indices. The overall results show that the performance of SCANet is highly

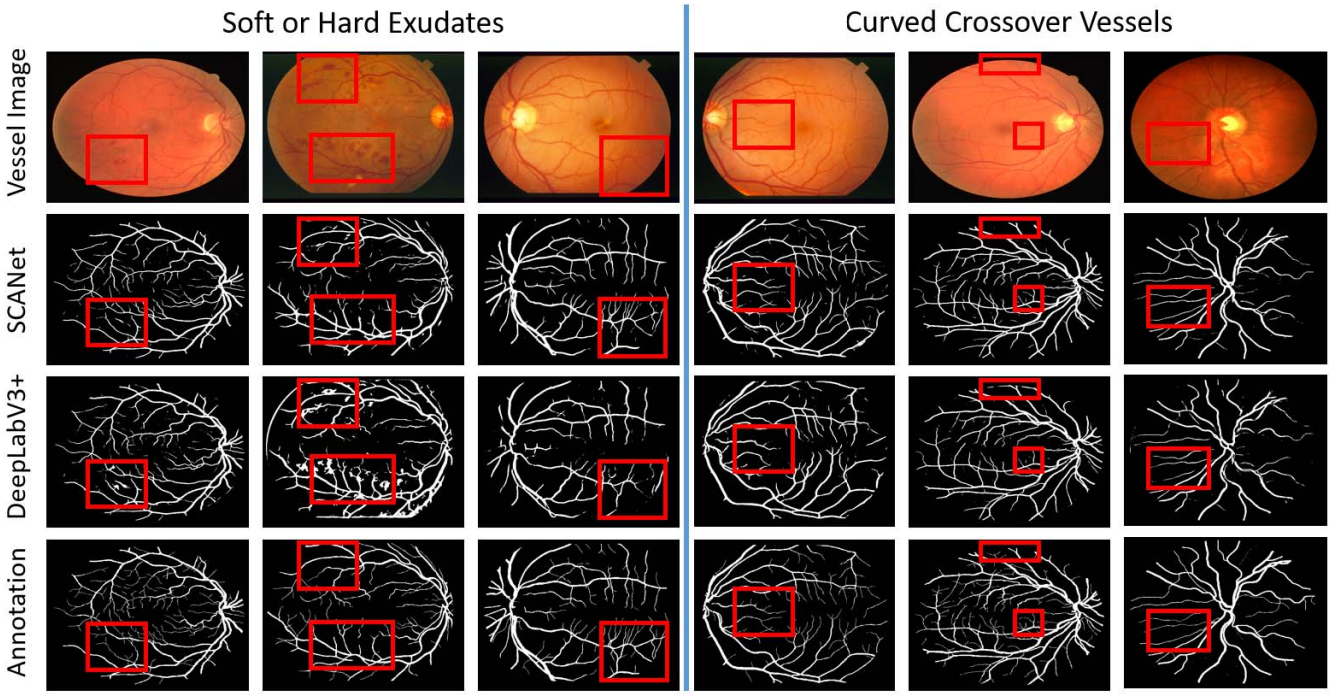


Fig. 9. Qualitative results on retinal vessel datasets. The performances under the interferences of soft or hard exudates are marked with red boxes on the left. The images on the right are the instances of curved crossover vessels.

TABLE II
COMPARISON WITH STATE-OF-THE-ARTS ON DRIVE, STARE, AND CHASE_DB1

Method	DRIVE			STARE			CHASE_DB1		
	Sen(%)	Spe(%)	Acc(%)	Sen(%)	Spe(%)	Acc(%)	Sen(%)	Spe(%)	Acc(%)
DDNET [44]	81.32	97.83	96.07	83.98	97.61	96.98	82.75	97.68	96.48
HAnet [45]	79.91	98.13	95.81	81.86	98.44	96.73	82.39	98.13	96.70
CcNet [46]	76.25	98.09	95.28	77.09	98.48	96.33	-	-	-
Li, et al. [47]	79.21	98.10	95.68	83.52	98.23	96.78	78.18	98.19	96.35
SCS-Net [48]	82.89	98.38	96.97	82.07	98.39	97.36	83.65	98.39	97.44
Yang, et al. [49]	83.53	97.51	95.79	79.46	98.21	96.26	81.76	97.76	96.32
SCANet_U	**80.76**	**98.21**	**96.70**	**81.94**	**97.08**	**95.86**	**80.43**	**98.50**	**97.22**
SCANet_D	**82.08**	**98.00**	**96.62**	**84.94**	**97.52**	**96.51**	**82.53**	**98.21**	**97.37**

TABLE III
COMPARISON WITH STATE-OF-THE-ARTS ON HRF

Method	Sen(%)	Spe(%)	Acc(%)
Attention U-Net [50]	78.55	98.14	96.58
CE-Net [51]	79.56	98.23	96.73
SCS-Net [48]	81.14	98.23	96.87
BSEResU-Net [52]	80.67	97.96	96.37
SkelCon [53]	78.53	98.86	95.90
SCANet_U	**80.89**	**98.44**	**97.09**
SCANet_D	**81.29**	**98.25**	**96.98**

competitive with other state-of-the-art methods. SCANet_D with 82.08%, 84.94%, and 82.53% Sen on DRIVE, STARE, and CHASE_DB1 overtakes multiple state-of-the-art algorithms. Besides, the performances of three retinal vessel datasets in Spe and Acc can be considered reasonable. In Table III, the performances of SCANet_D and SCANet_U on HRF are highly competitive with other state-of-the-art models, even outperforming them in Sen, Spe, and Acc. SCS-Net is an efficient network for the tasks of retinal vessel segmentation, the performances of which on all four datasets are of a great reference value. The results also illustrate that our SCANet could get accurate retinal vessel segmentation in the condition of the different base networks (U-Net and DeepLabV3+).

4) Visualization: The performances of SCANet under the interferences of soft and hard exudates are shown on the left of Fig. 9. The noises and vessel misses caused by interferences are exhibited in the predictions of DeepLabV3+. However, the critical areas marked with red boxes are segmented accurately by SCANet. SCANet also significantly surpasses DeepLabV3+ in comparing curved crossover and fine vessels.

D. Ablation Studies

We provide several ablation analyses about each key component and semi-supervised alternate strategy (SSAS) of SCANet. The related experiments are conducted on VESSEL-NIR. To explore the impact of labeled data proportion, different numbers of labeled images in the training set are operated on VESSEL-NIR, DRIVE, STARE, and CHASE_DB1. In addition, the impact of branch weightage is researched last.

1) Effect of Segmentation Branch: Three baselines, including No.1 (base network U-Net), No.2 (base network + SB without feature fusion), and No.3 (base network + SB) in Table IV

TABLE IV
ABLATION STUDIES. SB: SEGMENTATION BRANCH, FF: FEATURE FUSION, CB: CONSISTENCY BRANCH, AB: ADVERSARIAL BRANCH, SSAS: SEMI-SUPERVISED ALTERNATE STRATEGY

Model	SB	FF	CB	AB	SSAS	PAC (%)	Rec (%)	Pre (%)	IOU (%)	DSC (%)
No.1						96.89	81.93	83.38	68.26	80.97
No.2	✓					97.06	83.80	85.35	71.93	83.45
No.3	✓	✓				97.38	85.01	81.41	72.29	83.70
No.4	✓	✓	✓			97.44	85.39	86.38	73.75	84.69
No.5	✓	✓	✓	✓		97.66	84.48	88.68	74.50	85.23
No.6	✓	✓	✓		✓	97.73	83.43	89.74	74.88	85.44
No.7	✓	✓	✓	✓	✓	**97.85**	**87.22**	**88.44**	**77.78**	**87.16**

are deployed to explore the contribution of SB. The No.2 baseline proclaims that multi-scale recurrent neural network extracts richer semantic information than the base network in DSC of 83.45% to 80.97%. Combining multiple predictions at different scales in the No.3 baseline gets DSC of 83.70% and IOU of 72.79%. As the principal part of SCANet, it shows that the entire SB undertakes the primary assignment of extracting semantic features but consumes massive feature calculations.

2) Effect of Consistency Branch: We evaluate the effect of CB through No.3 and No.4 (base network + SB + CB). In Table IV, all metrics of No.4 steadily raising illustrates the significance of CB. It shows that the assistance of CB promotes the subcutaneous vessel segmentation effectively. Synthesizing the original input image helps SB obtain abstracted feature information and achieve 84.69% DSC, 73.75% IOU, 85.39% Rec. The increases in Rec and Pre denote that CB successfully removes the false positive and false negative regions from the segmentation predictions.

3) Effect of Adversarial Branch: In the evaluation of AB, No.4 and No.5 (base network + SB + CB + AB) are adopted for quantitative analysis. It shows that the involvement of AB makes the performance achieved in DSC of 85.23% and IOU of 74.50%. The improvements in the above metrics are attributed to the capacity of AB to distinguish the segmentation result from annotation. Though partial evaluation metrics are descending slightly, the overall results indicate that the AB module positively affects the whole architecture.

4) Effect of Semi-supervised Alternate Strategy: Moreover, we explore the significance of the semi-supervised learning strategy in No.6 (No.4 + SSAS) and No.7 (No.5 + SSAS). Unlike the fully-supervised mode, SSAS exploits 1000 labeled images and 2000 unlabeled images simultaneously. Both CB and AB in the SSAS mode improve the effectiveness of SCANet. In the condition of scarce labeled data, the SSAS mode may present fantastic enhancement. In Table IV, the performance of No.7 precedes No.6 in DSC and IOU by 1.72% and 2.90%, which is far advance of the base network in all evaluation metrics. Since the unlabeled images work in the No.7 model, they significantly improve to No.5 in DSC by 1.93% and IOU by 3.28%. Overall, the continuous improvements in IOU and DSC from No.1 to No.7 illustrate the abilities of crucial components synergy and training strategy of SCANet.

5) Impact of the Proportion of Labeled Data: Table V lists the results in a semi-supervised model with varying proportions of labeled data on four datasets (VESSEL-NIR, DRIVE, STARE, and CHASE_DB1). Under 50% labeled samples, the performance of SCANet is better than that under other proportions on each dataset. This phenomenon demonstrates the impact of labeled data variance on semi-supervised semantic segmentation. When the number of labeled data varies from 50% to 10%, the minor performance decay on VESSEL-NIR indicates that our method is more robust to the fluctuation in the case of adequate training data. However, SCANet is slightly sensitive to the number change on the dataset with rare labeled data. In DRIVE, STARE, and CHASE_DB1, the DSC scores of SCANet fall by 10.99%, 8.30%, and 10.35% when the number of labeled data varies from 50% to 25%. The main reason causing unstable performance is that the segmentation without sufficient labeled data could not over-rely on assistant branches. Compromising in data number, the images trained for SB are also transmitted to AB and CB in the 50% instances of the three retinal vessel datasets. There is no interference or negative influence occurred in the synergy of the tri-branch.

6) Impact of Branch Weightage: The branch weightage controls the proper contribution of the auxiliary branch to consolidate the individual effectiveness. The analyses about branch weights (hyper-parameters $\lambda_1/\lambda_2/\lambda_3$) are presented in Table VI. Here we choose SCANet_U to implement the experiments about branch balance on VESSEL-NIR. The balance between SB, CB, and AB directly influences the final prediction. Too little weights in CB and AB would weaken the role of unlabeled data and make the semi-supervised SCANet meaningless. If the weights above are assigned too high values, the training focus of SCANet would be far from the main segmentation task. Based on the illustrations, the evaluation performance of 10/3/3 transcending other weightages in IOU and DSC is rational and credible to model optimization.

E. Extension to Skin Lesion Segmentation

1) Data: The experiments above are all about vessel segmentation. To better evaluate the generality and effectiveness of SCANet under massive training data, the publicly available 2017 and 2018 International Skin Imaging Collaboration (ISIC) skin lesion segmentation datasets [12], [54] are used. The datasets are collected from various treatment centers by ISBI, which hosts a challenge to segment lesion boundaries in 2D dermoscopic images. To fully demonstrate the strength of SCANet, we leverage all 2594 images of ISIC2018 to guide the supervised segmentation training, and lay on all 2000 images of ISIC2017 for the assistant unsupervised training of the consistency branch and adversarial branch. Specifically, the 2594 images of ISIC2018 are divided into 1800, 200, and 594 for training, validation, and testing.

2) Setups: The learning rate is initialized as $1 \times e^{-5}$ and decays with a momentum of 0.5 with patience of 10 epochs no promotion. The training process performs 50 epochs in the fully-supervised training stage, and 150 epochs in the semi-supervised alternate training stage.

TABLE V
RESULTS ON FOUR DATASETS WITH DIFFERENT PROPORTIONS (10%, 25% AND 50%) OF LABELED DATA

	VESSEL-NIR					DRIVE			STARE			CHASE_DB1		
	PAC (%)	Rec (%)	Pre (%)	IOU (%)	DSC (%)	Sen (%)	Spe (%)	Acc (%)	Sen (%)	Spe (%)	Acc (%)	Sen (%)	Spe (%)	Acc (%)
10%	95.98	75.50	73.69	58.66	70.71	65.47	97.06	94.23	66.16	96.59	94.24	49.55	99.32	95.73
25%	96.86	80.27	76.88	67.91	80.56	71.99	97.23	95.87	76.64	95.92	95.93	72.18	98.22	96.36
50%	**97.95**	**87.22**	**88.44**	**77.78**	**87.16**	**82.08**	**98.00**	**96.62**	**84.94**	**97.52**	**96.51**	**82.53**	**98.21**	**97.37**

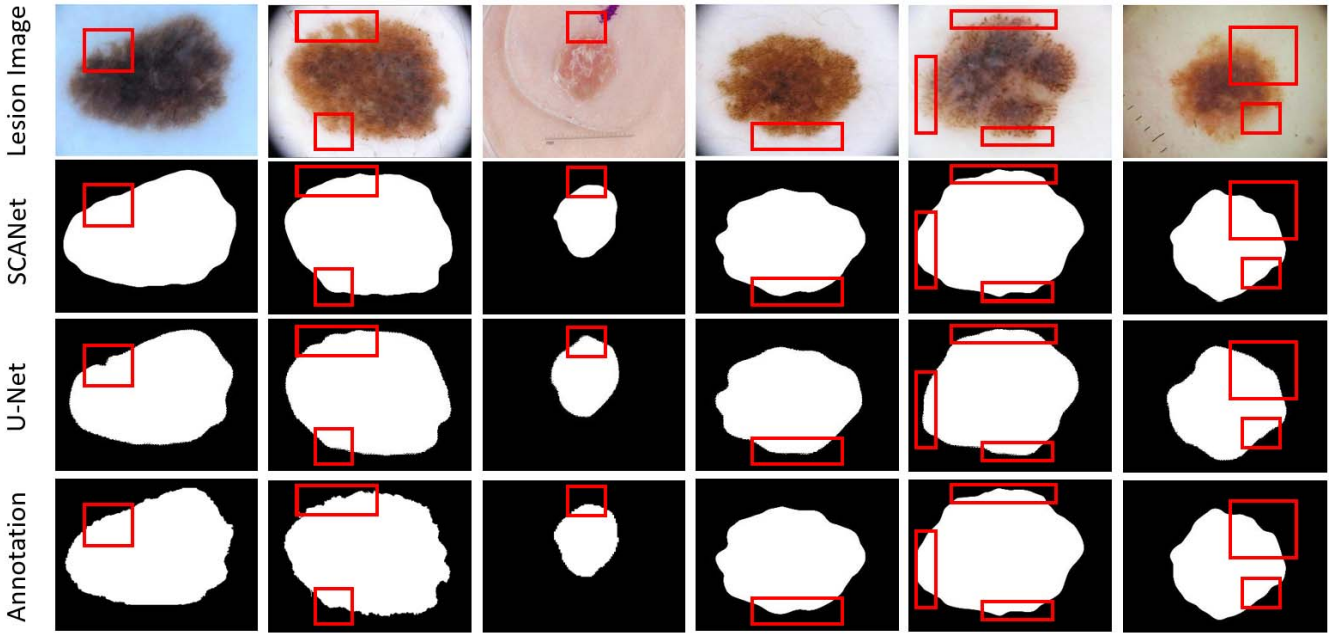


Fig. 10. Qualitative results on ISIC2018. The key segmentation areas are marked with red boxes.

TABLE VI
ABLATIONS ON BALANCING WEIGHTS $\lambda_1/\lambda_2/\lambda_3$

λ_1 / λ_2 / λ_3	PAC (%)	Rec (%)	Pre (%)	IOU (%)	DSC (%)
10 / 1 / 1	97.81	84.37	89.98	75.84	86.16
10 / 3 / 3	**97.85**	**87.22**	**88.44**	**77.78**	**87.16**
10 / 5 / 5	97.79	88.15	86.35	76.83	86.58
10 / 8 / 8	97.89	87.82	87.41	77.48	86.98
10 / 10 / 10	97.70	86.56	86.54	75.87	85.95

TABLE VII
COMPARISON WITH STATE-OF-THE-ARTS ON ISIC2018

	U-Net [35]	DeepLabV3+ [36]	CE-Net [51]	BAT [55]	SCANet_U	SCANet_D
IOU (%)	77.8	79.0	81.6	84.3	**84.6**	**85.4**
DSC (%)	88.5	89.3	89.5	91.2	**91.4**	**91.8**

3) Results: In Table VII, the results of SCANet surpass the other state-of-the-art algorithms in DSC and IOU. The U-Net and DeepLabV3+ get the DSC scores of 88.5% and 89.3%. Nevertheless, SCANet_U and SCANet_D significantly outperform them with little computation memory consumption. As the consistency branch and the adversarial branch do not work in testing phase, the result of SCANet would be better than that of its base network on the same time consumption per image. In the novel methods (CE-Net [51] and BAT [55]), the gaps between them and SCANet are slight. Compared with CE-Net, the superiorities of SCANet_D in DSC of 91.8% and IOU of 85.4% prove the effectiveness of our proposed method on the skin lesion segmentation task. BAT, the performance of which is only second to SCANet, achieves the results of 91.2% and 84.3% in DSC and IOU.

4) Visualization: The prediction instances of SCANet are exhibited in Fig. 10. The boundaries of skin lesions could be segmented accurately in the synergy of three branches without arduous consumption. SCANet is exceptionally superior to U-Net when the image contrast is not strong, and the boundary is not clear. The performances of crucial areas are marked with red boxes.

VI. CONCLUSION

This paper uses a novel semi-supervised learning framework SCANet for vessel segmentation on the peripheral subcutaneous vessel dataset VESSEL-NIR. SCANet deploys multi-scale recurrent neural network, consistency decoder, and adversarial learning as a tri-branch system in the semi-supervised alternate training manner. Extensive

experiments on VESSEL-NIR demonstrate the effectiveness and practical applicability of SCANet. SCANet harvests the best segmentation performance with little computation consumption compared with other state-of-the-art algorithms. The model generality of SCANet is presented by complementary experiments implemented on four public retinal vessel datasets and one skin lesion dataset. We also conduct ablation experiments to evaluate the key components and training strategy of SCANet.

Although our framework has achieved good segmentation results, there are still some limitations. First, the three retinal vessel datasets used are from before 2010. Our future study could extend the experiments to related datasets of much higher quality and resolution. Then, for the task of retinal vessel segmentation, training on a dataset and testing on a different one remains an unsolved challenge. In the future, we may employ the semi-supervised framework to explore and improve cross-dataset performance.

REFERENCES

[1] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12674–12684.

[2] X. Liu *et al.*, "Accurate colorectal tumor segmentation for CT scans based on the label assignment generative adversarial network," *Med. Phys.*, vol. 46, no. 8, pp. 3532–3542, Aug. 2019.

[3] W. Xia *et al.*, "MAGIC: Manifold and graph integrative convolutional network for low-dose CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3459–3472, Dec. 2021.

[4] D. Fan *et al.*, "Inf-Net: Automatic COVID-19 lung infection segmentation from CT images," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2626–2637, Aug. 2020.

[5] Q. Liu, L. Yu, L. Luo, Q. Dou, and P. A. Heng, "Semi-supervised medical image classification with relation-driven self-ensembling model," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3429–3440, Nov. 2020.

[6] E. Arazo, D. Ortego, P. Albert, N. E. O'Connor, and K. McGuinness, "Pseudo-labeling and confirmation bias in deep semi-supervised learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.

[7] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8174–8182.

[8] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.

[9] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, Mar. 2000.

[10] C. G. Owen *et al.*, "Measuring retinal vessel tortuosity in 10-year-old children: Validation of the computer-assisted image analysis of the retina (CAIAR) program," *Investigative Ophthalmol. Vis. Sci.*, vol. 50, no. 5, pp. 2004–2010, 2009.

[11] J. Odstrcilik *et al.*, "Retinal vessel segmentation by improved matched filtering: Evaluation on a new high-resolution fundus image database," *IET Image Process.*, vol. 7, no. 4, pp. 373–383, Jun. 2013.

[12] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 1, pp. 1–9, Dec. 2018.

[13] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent advances in recurrent neural networks," 2017, *arXiv:1801.01078*.

[14] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci. USA*, vol. 79, no. 8, pp. 2554–2558, 1982.

[15] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

[16] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3367–3375.

[17] S. Valipour, M. Siam, M. Jagersand, and N. Ray, "Recurrent fully convolutional networks for video segmentation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 29–36.

[18] M. S. Ibrahim, A. Vahdat, M. Ranjbar, and W. G. Macready, "Semi-supervised semantic image segmentation with self-correcting networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12715–12725.

[19] Y. Wang, P. Tang, Y. Zhou, W. Shen, E. K. Fishman, and A. L. Yuille, "Learning inductive attention guidance for partially supervised pancreatic ductal adenocarcinoma prediction," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2723–2735, Oct. 2021.

[20] Z. Wang *et al.*, "Alleviating semantic-level shift: A semi-supervised domain adaptation method for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 936–937.

[21] Y.-X. Zhao, Y.-M. Zhang, M. Song, and C.-L. Liu, "Multi-view semi-supervised 3D whole brain segmentation with a self-ensemble network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2019, pp. 256–265.

[22] L. Sun, J. Wu, X. Ding, Y. Huang, G. Wang, and Y. Yu, "A teacher-student framework for semi-supervised medical image segmentation from mixed supervision," 2020, *arXiv:2010.12219*.

[23] A. Chartsias *et al.*, "Factorised spatial representation learning: Application in semi-supervised myocardial segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2018, pp. 490–498.

[24] S. Mittal, M. Tatarchenko, and T. Brox, "Semi-supervised semantic segmentation with high-and low-level consistency," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 4, pp. 1369–1379, Apr. 2021.

[25] M. Wang *et al.*, "Semi-supervised capsule cGAN for speckle noise reduction in retinal OCT images," *IEEE Trans. Med. Imag.*, vol. 40, no. 4, pp. 1168–1183, Apr. 2021.

[26] A. Meyer *et al.*, "Uncertainty-aware temporal self-learning (UATS): Semi-supervised learning for segmentation of prostate zones and beyond," *Artif. Intell. Med.*, vol. 116, Jun. 2021, Art. no. 102073.

[27] Y. Xie, J. Zhang, Z. Liao, J. Verjans, C. Shen, and Y. Xia, "Pairwise relation learning for semi-supervised gland segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2020, pp. 417–427.

[28] S. Moccia, E. De Momi, S. El Hadji, and L. S. Mattos, "Blood vessel segmentation algorithms—Review of methods, datasets and evaluation metrics," *Comput. Methods Programs Biomed.*, vol. 158, pp. 71–91, May 2018.

[29] Y. Li *et al.*, "A novel method for low-contrast and high-noise vessel segmentation and location in venipuncture," *IEEE Trans. Med. Imag.*, vol. 36, no. 11, pp. 2216–2227, Nov. 2017.

[30] M. H. Fouad Aref, A. A. R. Sharawi, and Y. H. El-Sharkawy, "Delineation of the arm blood vessels utilizing hyperspectral imaging to assist with phlebotomy for exploiting the cutaneous tissue oxygen concentration," *Photodiagnosis Photodynamic Therapy*, vol. 33, Mar. 2021, Art. no. 102190.

[31] V. M. Leli, A. Rubashevskii, A. Sarachakov, O. Rogov, and D. V. Dylov, "Near-infrared-to-visible vein imaging via convolutional neural networks and reinforcement learning," in *Proc. 16th Int. Conf. Control, Autom., Robot. Vis. (ICARCV)*, Dec. 2020, pp. 434–441.

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

[33] J. Wei *et al.*, "Genetic U-Net: Automatically designed deep networks for retinal vessel segmentation using a genetic algorithm," *IEEE Trans. Med. Imag.*, vol. 41, no. 2, pp. 292–307, Feb. 2022.

[34] Y. Qin *et al.*, "Learning tubule-sensitive CNNs for pulmonary airway and artery-vein segmentation in CT," *IEEE Trans. Med. Imag.*, vol. 40, no. 6, pp. 1603–1617, Jun. 2021.

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2015, pp. 234–241.

[36] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 833–851.

[37] J. He, Z. Deng, L. Zhou, Y. Wang, and Y. Qiao, "Adaptive pyramid context network for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7519–7528.

[38] J. Fu *et al.*, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3146–3154.

[39] A. Kirillov, Y. Wu, K. He, and R. Girshick, "PointRend: Image segmentation as rendering," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9799–9808.

[40] 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 *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2019, pp. 605–613.

[41] V. Verma, K. Kawaguchi, A. Lamb, J. Kannala, Y. Bengio, and D. Lopez-Paz, "Interpolation consistency training for semi-supervised learning," 2019, *arXiv:1903.03825*.

[42] X. Luo *et al.*, "Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2021, pp. 318–329.

[43] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2613–2622.

[44] L. Mou, L. Chen, J. Cheng, Z. Gu, Y. Zhao, and J. Liu, "Dense dilated network with probability regularized walk for vessel detection," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1392–1403, May 2020.

[45] D. Wang, A. Haytham, J. Pottenburgh, O. Saeedi, and Y. Tao, "Hard attention net for automatic retinal vessel segmentation," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 12, pp. 3384–3396, Dec. 2020.

[46] S. Feng, Z. Zhuo, D. Pan, and Q. Tian, "CcNet: A cross-connected convolutional network for segmenting retinal vessels using multi-scale features," *Neurocomputing*, vol. 392, pp. 268–276, Jun. 2020.

[47] X. Li, Y. Jiang, M. Li, and S. Yin, "Lightweight attention convolutional neural network for retinal vessel image segmentation," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1958–1967, Mar. 2021.

[48] 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," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 102025.

[49] L. Yang, H. Wang, Q. Zeng, Y. Liu, and G. Bian, "A hybrid deep segmentation network for fundus vessels via deep-learning framework," *Neurocomputing*, vol. 448, pp. 168–178, 2021.

[50] J. Schlemper *et al.*, "Attention gated networks: Learning to leverage salient regions in medical images," *Med. Image Anal.*, vol. 53, pp. 197–207, Apr. 2019.

[51] Z. Gu *et al.*, "Ce-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[52] D. Li and S. Rahardja, "BSEResU-Net: An attention-based before-activation residual U-Net for retinal vessel segmentation," *Comput. Methods Programs Biomed.*, vol. 205, Jun. 2021, Art. no. 106070.

[53] Y. Tan, K.-F. Yang, S.-X. Zhao, and Y.-J. Li, "Retinal vessel segmentation with skeletal prior and contrastive loss," *IEEE Trans. Med. Imag.*, early access, Mar. 23, 2022, doi: 10.1109/TMI.2022.3161681.

[54] N. C. F. Codella *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 168–172.

[55] J. Wang, L. Wei, L. Wang, Q. Zhou, L. Zhu, and J. Qin, "Boundary-aware transformers for skin lesion segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2021, pp. 206–216.