Semi-Supervised Speckle Noise Reduction in OCT Images With UNet and Swin-Uformer

Yupei Chen, Jiaxiong Li, Zhongzhou Luo, Keyi Fei, Yan Luo, Zhengyu Duan, Jin Yuan, and Peng Xiao

***Abstract*— Speckle noise is the main cause of quality degradation of optical coherence tomography (OCT) images. However, speckle noise reduction is challenging due to the complex cause for statistical modeling and the requirement of a large amount of annotated data for conventional supervised learning strategies. In this article, a novel semi-supervised learning method is proposed for speckle noise reduction in OCT images with limited labeled data. Our method creates pseudo-labels for co-teaching in the training process between a U-shaped convolutional neural network and a U-shaped Transformer with a shifted window to preserve both global information and local details. The proposed scheme encourages the consistency between different streams when the advantages of both are leveraged to compensate each other for better convergence. It shows robustness on both normal and pathological OCT images with different diseases and from different devices. Our method exhibits advantages over several other state-of-the-art methods of speckle noise reduction. To our knowledge, this work is the first attempt to combine convolutional networks and Transformers for semi-supervised speckle noise reduction and achieves promising results on different datasets.**

***Index Terms*— Convolutional neural network (CNN), optical coherence tomography (OCT), semi-supervised learning, speckle noise reduction, Transformer.**

I. Introduction

OPTICAL coherence tomography (OCT) has been widely applied in ocular disease diagnosis for recent decades due to its noninvasive and efficient character. OCT relies on the coherence of optical waves backscattered spatially and temporally from tissue [1], during which process, speckle noise is generated as a main cause of OCT imaging quality degradation. Statistical studies on speckle noise have shown certain characteristics of speckle [2]. Speckle noise arises in all coherent imaging systems due to the effect of environmental conditions on the imaging sensor during imaging acquisition [9]. Speckle noise is common in medical images, such as ultrasound images, OCT images, and synthetic aperture radar (SAR) images. Speckle noise is generated in different ways in different imaging systems. In an ultrasound imaging system, speckle noise occurs when a sound wave beat interferes with little particles or on a scale equivalent to sound wavelength. As for radar images, it occurs due to random variation in return signal [9]. In OCT images, speckle is formed due to the mutual interference of coherent waves with a random set of intensities or random phase shifts. According to Klein et al. [10], there are two types of speckles: signal-carrying speckles and signal-degrading speckles. For speckle noise reduction, the signal-degrading speckle is removed and the signal is kept ideally. Speckle noise reduces the contrast of images and obstacles the observation of fine details in images [11]. When OCT images are corrupted by speckle noise, the image quality degrades, which makes it difficult for feature extraction, recognition, analysis, and quantitative measurement.

Nevertheless, it is difficult to study OCT speckle noise from statistical models [3] since it does not strictly obey any specific statistical distribution [4]. In practice, registration and averaging multiple scans acquired in succession from the same location is considered a standard way for speckle noise reduction in OCT images [5]. Nevertheless, this method is relatively time-consuming and impractical to operate in fast 3-D scanning due to the inability of patients to maintain fixation during examinations and the limitation of imaging speed [6], [7], [8]. In addition, the registration-averaging process might remove subtle but significant details and bring some motion artifacts [9]. Therefore, efficient methods for speckle noise reduction in OCT images are urgently required.

Over the last few decades, lots of efforts have been put into speckle noise reduction of OCT images. On the one hand, hardware-based approaches can reduce the noise from the scanner and detector to some extent, through the improvement of the light source of OCT devices [10]. However, the noise in the imaging system cannot be eliminated. On the other hand, software-based approaches are still the mainstream for speckle noise reduction in OCT images and can be roughly divided into several categories, including filter-based, transform domain-based, sparsity-based, and deep-learning-based

Manuscript received 21 November 2023; revised 6 February 2024; accepted 15 February 2024. Date of publication 28 March 2024; date of current version 5 April 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 82230033 and Grant 82271133, in part by the Department of Science and Technology of Guangdong Province under Grant 2021TX06Y127 and Grant 2021TQ06Y137, in part by the Basic and Applied Basic Research Foundation of Guangdong Province under Grant 2022A1515011486, in part by Guangzhou Science and Technology Plan Project under Grant 202103010001, and in part by the Fundamental Research Funds of the State Key Laboratory of Ophthalmology. The Associate Editor coordinating the review process was Dr. Liuyang Zhang. *(Yupei Chen and Jiaxiong Li contributed equally to this work.) (Corresponding authors: Jin Yuan; Peng Xiao.)*

The authors are with the State Key Laboratory of Ophthalmology and Guangdong Provincial Key Laboratory of Ophthalmology and Visual Science, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou 510060, China (e-mail: yuanjicornea@126.com; xiaop29@mail.sysu.edu.cn).

Digital Object Identifier 10.1109/TIM.2024.3381655

methods [11]. With the rapid development and wide application of deep learning in low-level vision tasks, more and more deep neural network-based methods for OCT image denoising have been proposed in recent years. Tajmirriahi et al. [12] implemented a lightweight mimic convolutional autoencoder for denoising OCT images with high computational efficiency. Devalla et al. [13] utilized a supervised convolutional neural network (CNN)-based U-shaped architecture with residual blocks, dilated convolutions, and multiscale hierarchical feature extractions to denoise OCT images of the optic nerve head. The results outperform the corresponding multiframe B-scans with reduced scanning times and minimal patient discomfort. However, these supervised learning-based methods require a large amount of the so-called clean OCT images which is unlikely to be fulfilled in practice.

To address these problems, in this article, we formulate a novel CNN- and Transformer-based semi-supervised learning for speckle noise reduction in OCT images. By introducing the co-training between the CNN and Transformer, different learning paradigms are implemented, and cross-pseudo-labels are created. In this way, the proposed method combines the advantages of convolutional learning and transformer learning and achieves impressive noise suppression results with a limited number of clean OCT images. The main assumption of this research is that registration and averaging multiple scans acquired in succession from the same location is considered a standard approach to obtaining annotated images for the training process. The novelty of our method lies in two aspects. First, the method uses a limited amount of labeled data during the training process which makes it more practical than traditional supervised CNNs or transformers. As we all know, the acquisition of labeled data is significant in the field of denoising with deep learning. It is even trickier to obtain clean, namely, noise-free or labeled, OCT images for speckle noise reduction. Second, the proposed method co-trains the CNN and Transformer with better convergence and more effective results compared with training with the CNN or Transformers solely. The main contributions of this article are as follows.

1) We present a novel semi-supervised learning scheme for speckle noise reduction in OCT images by co-training between the CNN and Transformer. To the best of our knowledge, this is the first work for semi-supervised co-training between the CNN and Transformer in a noise reduction task, and it is demonstrated that the proposed method outperforms many existing traditional methods and many other semi-supervised methods on both public benchmark and clinical dataset.
2) We formulate the simple but effective weighted joint loss function composed of supervised loss and unsupervised loss as the learning objective in the training process. The bidirectional loss is not forced to be explicitly consistent in the CNN stream and Transformer stream, in which way the advantages of the CNN and Transformer are leveraged to compensate each other for better convergence.
3) We validate the proposed method with objective metrics and retinal layer segmentation performance. Extensive experiments demonstrate the capability of generalization of the proposed method on different datasets from different devices and subjects with various pathologies. The proposed method also achieves comparable results with the supervised baseline.

II. Related Works

A. Traditional Statistical Speckle Noise Reduction Methods

As we mentioned in the previous section, traditional software-based approaches are mainly statistical speckle noise reduction methods and can be roughly divided into several categories, including filter-based, transform domain-based, sparsity-based, and deep-learning-based methods [11]. Filter-based approaches utilize the statistical characteristics of speckle noise and model the statistical noise locally or globally such as alternating sequential filters [14]. Typically, the nonlocal means (NLM) method [15] sets a search window and a similar window. By averaging the windows with different weights, the noisy images are smoothed nonlocally. Block matching and 3D filtering (BM3D) method [16] extracts stacked 3-D similar patches from noisy images for collaborative filtering. These filter-based approaches require laborious efforts of parameter tuning during implementation for different noise levels [17]. Transform domain-based approaches perform noisy image processing in the transform domain [18], typically in frequency, wavelet [19], and complex [20] domains. These methods obtain impressive results on image denoising but bring unexpected artifacts in the transform domain, which might spread to the entire image. Sparsity-based methods reconstruct noise-free images from sparse representation, such as dictionary learning. Multiscale sparsity-based tomographic denoising (MSBTD) [21] and K singular value decomposition (K-SVD) [22] are two typical sparsity-based methods. K-SVD is an iterative method that alternates between sparse coding of the examples based on the current dictionary and a process of updating the dictionary. The convergence is accelerated through the update of the dictionary columns and the sparse representations [23]. These methods are confronted with different problems, such as low efficiency, blurred edges, and oversmoothing, which may result in losing some clinically significant details [11].

B. Semi-Supervised Speckle Noise Reduction in OCT Images

Traditional supervised learning methods require a large amount of annotated data with high quality, which is unlikely to be fulfilled in medical image analysis. This dilemma makes more and more studies focus on semi-supervised methods for OCT image analysis, such as retinal disease classification [24], retinal layer and lesion segmentation [25], [26], medical image registration [27], and so on. Semi-supervised learning is a learning paradigm with regard to the study of how computers and natural systems learn in the presence of both labeled and unlabeled data [28]. In the past few years, many strategies have been proposed for semi-supervised learning, such as pseudo-labeling [29], deep co-training [30], deep adversarial learning [31], mean teacher [32], confidence learning [33], contrastive learning [34], and so on. These methods perform training on CNN-based models with limited labeled data.

For OCT image noise reduction, it is even trickier to obtain clean, namely, noise-free or labeled, OCT images for speckle noise reduction. Therefore, more and more studies focus on methods with limited or even without clean data. An edge-sensitive capsule condition generative adversarial network (GAN) with a small number of parameters was introduced for semi-supervised speckle noise reduction in retinal OCT images by [35] in 2021. It obtains the noise reduction mechanism of the system by learning the common information from paired noisy holographic reconstructed images. The results outperform several traditional smoothing algorithms and are comparable with supervised learning methods. Yin et al. [36] proposed a UNet-based method for speckle noise reduction in coherent imaging without clean images. Guo et al. [37] employed an unsupervised method using nonlocal GAN, which achieved promising results in both quantitative and qualitative aspects. The application of GAN makes it possible to generate noise-free images through unpaired data. However, these methods have some drawbacks due to the application of GAN. On the one hand, the generator is very sensitive to different input noisy images, which may bring artifacts to degrade the image quality. On the other hand, CycleGAN still requires a large amount of unpaired noisy and noise-free data for the generator and discriminator to learn the detailed features. Different from previous work, in this article, we are investigating a CNN and Transformer based semi-supervised learning for speckle noise reduction in OCT images.

C. CNN and Transformer

In the past few decades, convolution has been applied as the main component of deep neural networks for years. CNNs are biologically inspired trainable architectures composed of multiple stages [38]. Each stage is composed of a filter bank layer, a nonlinearity layer, and a feature pooling layer. With multistage architecture, CNNs can learn multilevel hierarchies of features. The capacity is controlled by varying the depth and breadth with strong assumptions about the stationarity of statistics and the locality of pixel dependencies of images [39]. Nevertheless, CNNs do not encode the position and orientation of objects and are incapable of being spatially invariant to the input data [40].

Recently, a Transformer was formulated to take the place of the dominant convolutional architecture in deep neural networks. Transformer is a self-attention-based architecture with high computational efficiency and scalability [41]. Vision Transformer (ViT) attains competitive performance at a large scale in vision tasks [42]. Despite the success in ViT at a large scale, the CNN still outperforms Transformer with similar-sized counterparts when trained on a small amount of data [43]. To employ the benefits of both the CNN and Transformer, convolutional ViT is proposed. The scheme intends to introduce the advantages of the CNN: local receptive fields, shared weights, and spatial subsampling while keeping the merits of the Transformer: dynamic attention, global context fusion, and better generalization.

D. UNet and Swin-Uformer

Among the various CNN architectures, UNet [44] is one of the most commonly adopted in medical image processing, especially in segmentation tasks. Medical images contain both regular patterns from human organs and subtle but significant details. UNet obtains multiscale contextual features with hierarchical feature maps and uses skip concatenation between encoders and decoders to enhance the preservation of subtle details.

Due to the excellent performance of UNet and its variant modifications, U-shaped architecture has been extensively utilized with Transformers for medical image segmentation in the past several years [45]. In addition, a Transformer with shifted window representation, namely, Swin Transformer [46], was brought up with higher efficiency and flexibility to multiscale models using limited self-attention computation to nonoverlapping local windows. Accordingly, the Swin Transformer is capable of efficiently solving pixel-wise vision tasks with content-based interactions between image content and attention weights and long-range dependency modeling [47].

III. Method

A. Overview

The main workflow of the proposed method is depicted in Fig. 1. First, the paired images and unpaired noisy images are split into training batches with a certain ratio for each batch. The training batches are input to train the two streams parallelly, thus pseudo-labeled images are obtained. Then, for paired data, the learning objective is calculated between the pseudo-clean images and the labeled clean images; for unpaired data, the learning objective is calculated between the two streams. Through the design of objective function in parallel training, the scheme intends to introduce the advantages of the CNN: local receptive fields, shared weights, and spatial subsampling while keeping the merits of the Transformer: dynamic attention, global context fusion, and better generalization. The total learning objective is the sum of paired and unpaired loss with the consistency weight. The proposed loss encourages the framework to focus both on local receptive fields and global context fusion. Next, the two streams are built with a UNet and U-shaped Transformer with shifted window, namely, Swin-Uformer, separately. Finally, the framework is evaluated on both public benchmark and clinical dataset. The detailed methods are elaborated below.

B. Semi-Supervised Learning With Cross-Pseudo-Label

The proposed semi-supervised learning method is inspired by several existing works: cross-pseudo-supervision [48], co-teaching [49], and convolutional ViT [43]. Co-teaching teaches two deep neural networks to learn from each other with extremely noisy labels through parallel training in mini-batches for noise-robust results. Cross-pseudo-supervision is a semi-supervised learning method that trains two networks with the same architecture and different weight initialization. The core of these two methods is that they both try to bring disturbance and train the networks for consistency despite the disturbance. Co-teaching introduces noisy labels during supervision and cross-pseudo-supervision introduces different initializations during the training process. However, these strategies are usually applied in binary segmentation tasks but are not suitable for low-level vision tasks. With high

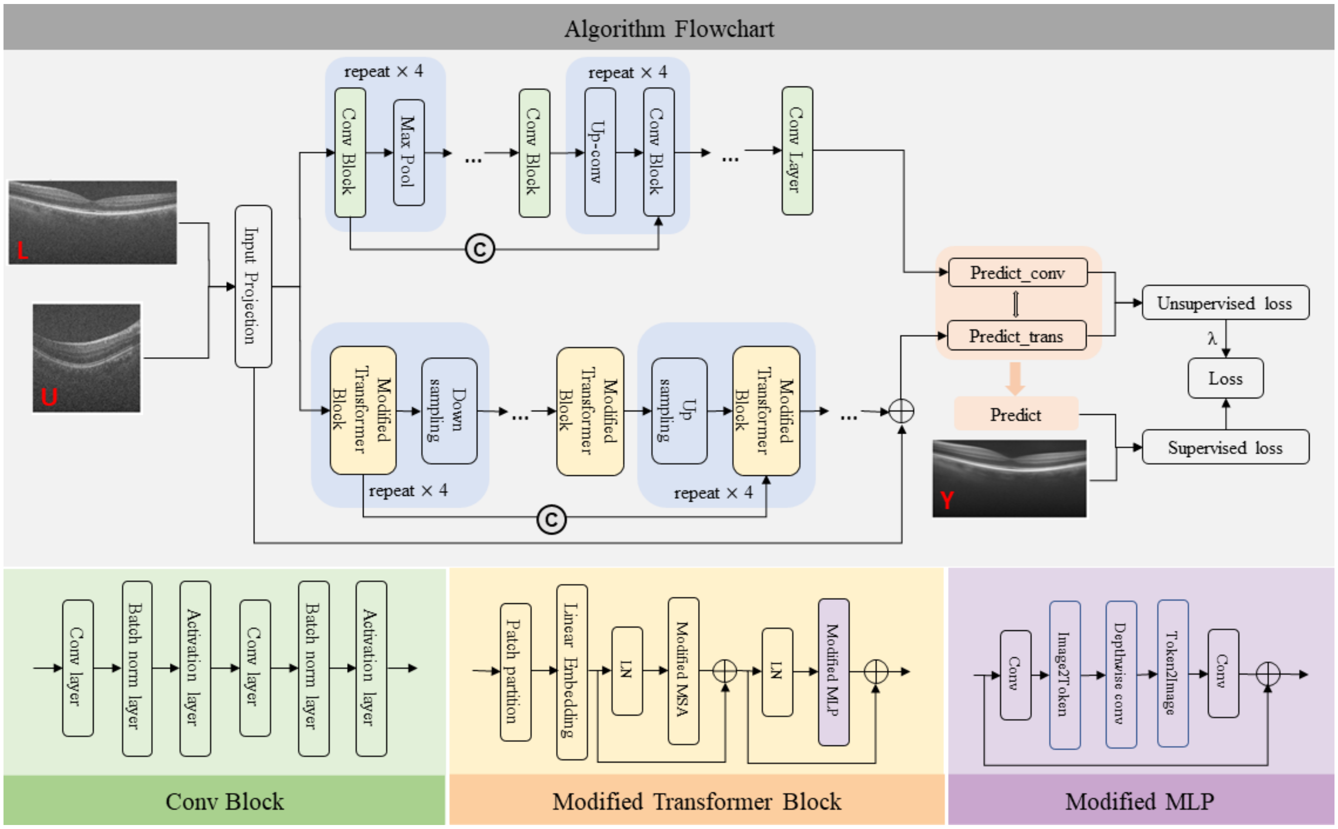

Fig. 1. Algorithmic flowchart of the proposed method.

requirements of different scaling detailed feature extraction, denoising tasks with a limited amount of labeled data face the problem of overfitting and inconsistency of the two streams of learning paradigms. To solve these issues, we introduce a co-training strategy with convolutional and transformer modules. The disturbance is introduced with different network architectures and different learning objectives. Set x_i as the ith input noisy image and y_i as the corresponding labeled clean image. We attain the pseudo-clean images by

$$p_i^{\text{conv}} = f_{\text{conv}}(x_i); \quad p_i^{\text{trans}} = f_{\text{trans}}(x_i) \tag{1}$$

where p_i^{conv} and p_i^{trans} denote the prediction of CNN f_{conv} and Transformer f_{trans} streams, respectively.

C. Learning Objectives

The overall learning objective is composed of supervision loss $\mathcal{L}_{\text{sup}}$ and pseudo-loss $\mathcal{L}_{\text{pseudo}}$, as indicated in the following equation:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{sup}} + \lambda \mathcal{L}_{\text{pseudo}} \tag{2}$$

where λ represents the tradeoff weight of consistency loss subject to sigmoid ramp-ups following the formula $\lambda = 1 + \lambda_0 e^{-5(1-x)^2}$, where λ_0 is a ramp-up ratio and x denotes the ratio between the current epoch number during the training phase and ramp-up length, which is set to be the whole epochs [50]. To guarantee the contribution of the consistency loss at the beginning of the training process, one is added to the formula. The hyperparameter λ_0 is set to be $1e^{-3}$ in our experimental results, which is determined empirically during the parameter fine-tuning phase. The loss function is formulated using the Charbonnier penalty function ℓ_{cb} [51] on the labeled and pseudo-labeled data over the parallel CNN and Transformer streams. Charbonnier penalty function is a differentiable variant of $\mathcal{L}_1$ norm and it is demonstrated that the robust Charbonnier loss function better handles outliers. It was introduced in [52] for training neural networks as $\ell_{\text{cb}}(x) = (x^2 + \epsilon^2)^{1/2}$ and ϵ is empirically set to $1e^{-3}$. The supervision loss ℓ_{sup} is calculated with the noisy input and the labeled clean image, as in the following equation:

$$\ell_{\text{sup}} = \ell_{\text{cb}}(x_i, y_i). \tag{3}$$

The pseudo-loss $\mathcal{L}_{\text{pseudo}}$ is the combination of the CNN and Transformer. For each stream, the pseudo-loss is bidirectional, which means that one is from the CNN to the Transformer, and the other is from the Transformer to CNN, as indicated in the following equations:

$$\ell_{\text{pseudo}}^{\text{conv}} = \ell_{cb}\left(x_i, p_i^{\text{trans}}\right) \tag{4}$$

$$\ell_{\text{pseudo}}^{\text{trans}} = \ell_{cb}\left(x_i, p_i^{\text{conv}}\right). \tag{5}$$

The configuration of bidirectional loss is the core part of the two-stream learning objective. We expect the two different learning paradigms to learn from each other and reach consistency without the enforcement of explicit constraints.

D. Technical Details

1) Network Architecture: The whole framework is based on the U-shaped CNN and U-shaped Transformer backbone, namely, UNet and Swin-Uformer, respectively. Among the various neural network architectures, UNet is one of the most commonly adopted in medical image processing. Medical images contain both regular patterns from human organs and subtle but significant details. UNet obtains multiscale contextual features with hierarchical feature maps and uses skip concatenation between encoders and decoders to enhance the preservation of subtle details. Due to the excellent performance of UNet and its variant modifications, U-shaped architecture is extensively utilized with the Transformer for medical image processing. The architectural intricacies are depicted below.

a) CNN stream: The CNN stream is composed of convolutional blocks with U-shaped concatenation and pooling layers. The network structure is formulated following the architecture settings in [40]. The convolutional block is comprised of a convolutional layer, batch normalization layer, and activation layer. LeakyReLU [53] is utilized in place of rectified linear unit (ReLU) for activation in convolutional blocks to avoid vanishing gradients during the training phase.

b) Transformer stream: We configure the Swin-Uformer similarly to UNet except the convolutional layers are dispensed with attention mechanisms to draw global dependencies between input and output [37]. The standard ViT contains multihead self-attention (MSA), multilayer perceptron (MLP), and positional encoding. Inspired by [54], we add a depth-wise convolutional block to the MLP to improve the capability of leveraging neighboring pixels, which is significant for image denoising [55]. Gaussian error linear unit (GELU) [56] is used as the activation function to avoid the problem of vanishing gradients after each convolution layer. In addition, the standard MSA module is replaced with shifted window partitioning in successive blocks, while other layers remain the same. The shifted window size is set to 8 in our experiments according to [42]. The core part of the Swin-Uformer module is illustrated in Fig. 1.

2) Implementation Details: In our experiments, the Adam weight decay (AdamW) [57] optimizer is adopted with an initial learning rate of 1e-4 and is dropped by a factor of 0.9 every 10 epochs. The batch size was set as 8, in which 2 of them are labeled data. The experiments are trained for 10 000 iterations, equally 150 epochs. The whole scheme was coded in Python based on PyTorch and trained using the NVIDIA Titan Xp GPU with 12G memory.

3) Evaluation Protocol: The method is verified by several objective metrics and retinal layer segmentation performance. To test the robustness of our method, the evaluation is implemented on two datasets—a public benchmark and a clinical dataset collected ourselves with various diseases.

IV. Experiments

A. Dataset

Two datasets are employed to evaluate the proposed method in our experiments. One is a public benchmark with some paired data and some unpaired noisy data [5]. The other is a clinical dataset acquired by our team from patients with various diseases.

1) Public Dataset: The public dataset was first introduced in [5]. In our experiment, 24 spectral domain optical coherence tomography (SDOCT) image pairs from 24 subjects were included. The so-called "noise-free" labeled data was obtained through registration and averaging of repeated B-Scans. In addition, 44 noisy OCT images, including 39 from humans and five from mice, were employed.

2) Clinical Dataset: The second dataset includes 313 1024 × 1536 image pairs extracted from 26 volume scans from 15 subjects, utilizing a commercial ZEISS PlexElite 9000 swept-source OCT angiography (Carl Zeiss Meditec, Inc, Dublin, CA, USA) which is equipped with 200 kHz of scan speed, 1.95 μm of digital axial resolution, 6 mm of A-scan depth, and 1040–1070 nm of optical source center wavelength. For each subject, we scanned a square (∼6× 6 mm) volume centered at the retinal fovea with 1536 A-scans per B-scan and 1024 B-scans per volume. Among the subjects, there are five eyes from three subjects with aged macular disease (AMD), six eyes from three subjects with central serous chorioretinopathy (CSC), four eyes from three subjects with diabetic retinopathy (DR), five eyes from three subjects with macular edema, and six eyes from three subjects with normal eyes. The study was approved by the Institutional Review Board of Sun Yat-sen University, and informed consent was obtained from all subjects involved (No. 2020KYPJ154).

3) Data Preprocessing: To improve the performance of the proposed method and avoid overfitting, data augmentation is implemented during the training process for both streams. In our experiment, augmentation is conducted through random brightness and contrast distortion, random rotation, and flipping, to simulate the practical situations in clinics. Each image is normalized before augmentation. After augmentation, each image is randomly cropped into eight patches with the size of 256 × 256 as training input. The training, validation, and testing set consisted of 256, 27, and 30 images, respectively, which were manually selected from the clinical dataset. Each dataset consists of images from different subjects with different diseases and is distributed roughly even. The public dataset is only used in the testing phase to validate the robustness of the proposed method.

B. Evaluation Metrics

To evaluate the performance of the proposed method for speckle noise reduction in OCT images, several objective indicators were calculated for quantitative evaluation in the testing phase according to [23] and [58]. Some of these indicators were measured over the entire image and some were measured over the manually selected region of interest (ROIs), including one background ROI and three signal ROIs. Fig. 2 shows some example results with normal and different pathological OCT images from the clinical dataset. The background ROIs are marked with green rectangles and the signal ROIs are marked with blue rectangles. The objective indicators are calculated as below.

The signal-to-noise ratio (SNR) is a widely used global performance measure, which is defined as

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\max(I)^2}{\sigma_b^2}\right) \tag{6}$$

where I is the input image and σ_b is the standard deviation of noise in the background region.

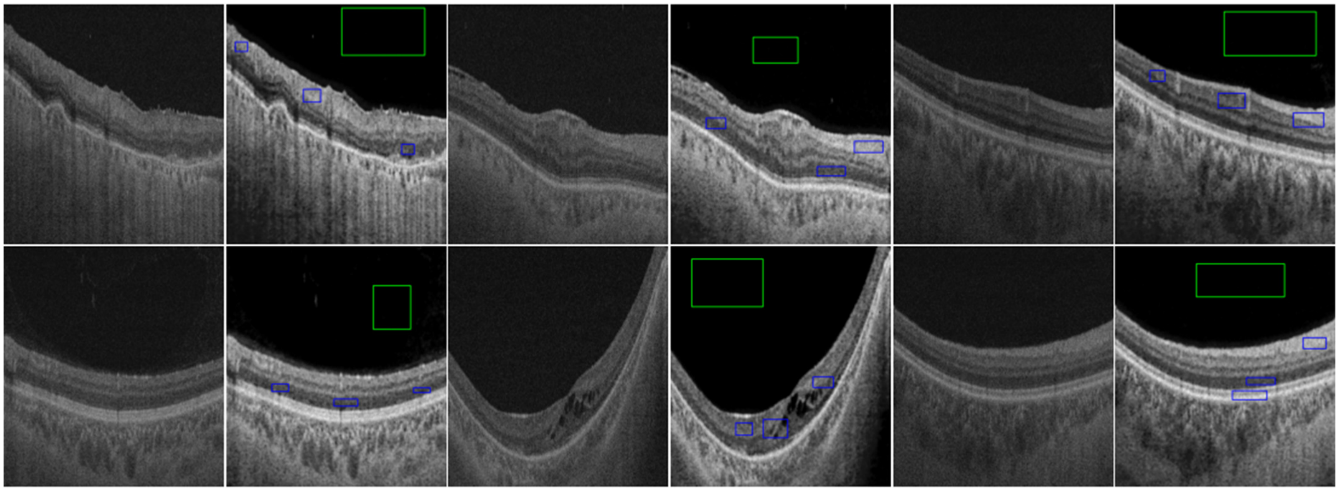

Fig. 2. Examples of denoising results from the test dataset with the proposed method. From left to right in each instance: Original noisy image, denoised image with the proposed method. The green rectangle is selected as the background region and blue rectangles represent the signal ROIs for the calculation of SNR, CNR, EPI, and ENL.

The contrast-to-noise ratio (CNR) is a typical local indicator for image quality measurement. It measures the contrast between the ROI and background noise, the CNR is calculated as

$$\mathrm{CNR} = 10\log_{10}\frac{\mu_m - \mu_b}{\sqrt{\sigma_m^2 + \sigma_b^2}} \tag{7}$$

where μ_m and σ_m^2 denote the mean and variance of the mth ROI, respectively. μ_b and σ_b^2 denote the mean and variance of the background region, respectively.

The edge preservation index (EPI) is defined as

$$\mathrm{TP} = \frac{1}{M}\sum_{m=1}^{M}\frac{\sigma_m^2}{\sigma_m'^{\,2}}\sqrt{\frac{\mu_{\mathrm{den}}}{\mu_{\mathrm{in}}}} \tag{8}$$

where σ_m' is the standard deviation of the mth ROI in the unprocessed input image. μ_{den} and μ_{in} is the mean of the denoised image and noisy image, respectively. Ideally, μ_{den} and μ_{in} are supposed to be equal. The TP measure is averaged over M ROIs. For better texture preservation, TP is bigger.

An equivalent number of looks (ENL) measures the smoothness in the homogenous region. In our experiment, ENL is calculated in the background region which is supposed to be homogenous. ENL is calculated as

$$\mathrm{ENL} = \frac{\mu_b^2}{\sigma_b^2} \tag{9}$$

where μ_b and σ_b^2 denote the mean and variance of the background region, respectively. The homogenous region of the image is smoother when the ENL is larger.

Moreover, to compare the denoised image with the averaged image after registration, the peak signal-to-noise ratio (PSNR) and structure-similarity-index-measure (SSIM) [59] are calculated for evaluation as below

$$\mathrm{PSNR} = 10\log_{10}\frac{\mathrm{MAX}^2}{\frac{1}{\mathrm{MN}}\sum||I_F - I_G||^2} \tag{10}$$

$$\mathrm{SSIM} = \frac{(2\mu_1\mu_2 + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \tag{11}$$

The objective indicators are calculated for the test set as shown in Tables I and II. It shows that the proposed method provides promising results for speckle noise reduction in OCT images. From a subjective perspective, it is observed that the test images are smoothed with enhanced contrast and the detailed retinal structures are well preserved. Moreover, the proposed architecture works for both lesioned and nonlesioned B-scans as shown in Fig. 2.

C. Comparison With Other Methods

To validate the proposed method, several typical traditional denoising methods and some state-of-the-art supervised and semi-supervised denoising methods were implemented for comparison, as is mentioned in Sections I and II, including NLM, wavelet filtering, nonlinear complex diffusion filtering (NCDF) [20], BM3D, K-SVD, supervised UNet [40] and Uformer [41], and another semi-supervised method cGAN [35]. For quantitative evaluation, the objective indicators were calculated and the experimental results are presented in Tables I and II. As shown in Table I, the ENL in NLM is rather low but the CNR is rather high, which might be correlated with background noise. Wavelet had the lowest TP which is likely caused by blurred edges. NCDF has the highest TP and the lowest SNR and CNR, which may be caused by the artifacts near the edge. KSVD provides high ENL and CNR, which may be caused by oversmoothing. BM3D presents fair performance but is far behind the supervised learning methods. With the assistance of unlabeled data, the proposed method outperforms the supervised methods with a limited amount of labeled data. It presents the highest SNR and TP compared with other methods, indicating better smoothing and detail preservation performance. Moreover, the proposed method presents a rather high ENL which shows the ability for background smoothing. Nevertheless, the low CNR indicator implies the inadequacy of contrast preservation. Table II presents higher PSNR and SSIM given the so-called noise-free image acquired from the method described in [60], resembling the ground truth. Furthermore, the proposed method shows

TABLE I
QUANTITATIVE EXPERIMENTAL RESULTS OF DIFFERENT METHODS

Category	Method	SNR	CNR	TP	ENL
	Noisy	29.52±1.96	3.92±0.75	1.17±0.12	32.18±10.40
	Clean	45.68±4.71	8.53±1.09	1.00±0.00	80.27±16.32
Statistical methods	NLM	45.15±2.16	8.18±0.96	0.82±0.10	51.47±24.89
	Wavelet	46.92±3.27	8.08±0.99	0.17±0.03	64.25±34.74
	NCDF	39.72±1.96	7.46±0.86	1.06±0.10	53.98±31.23
	BM3D	45.22±3.32	7.52±0.85	0.66±0.09	76.59±43.85
	K-SVD	42.71±1.97	8.91±1.11	0.60±0.07	91.12±30.72
Supervised learning	UNet	47.74±3.03	8.00±0.85	1.13±0.11	80.72±40.27
	Uformer	46.25±3.93	7.34±1.06	1.05±0.11	86.32±49.16
Semi-supervised learning	**Proposed**	53.57±2.25	7.89±0.78	0.97±0.10	84.49±50.23

TABLE II
QUANTITATIVE EXPERIMENTAL RESULTS OF DIFFERENT METHODS

Category	Method	PSNR	SSIM	NRMSE
	Noisy	19.24±0.60	0.46±0.03	0.38±0.02
Statistical methods	NLM	25.18±1.49	0.79±0.04	0.19±0.03
	Wavelet	23.18±3.16	0.80±0.04	0.25±0.10
	NCDF	23.94±1.44	0.76±0.03	0.22±0.04
	BM3D	23.66±2.60	0.80±0.04	0.24±0.09
	K-SVD	25.77±1.19	0.81±0.05	0.18±0.02
Supervised learning	UNet	24.92±1.31	0.84±0.04	0.20±0.03
	Uformer	23.39±0.96	0.79±0.06	0.23±0.02
Semi-supervised learning	UNet-UNet	22.27±4.02	0.77±0.38	0.33±0.09
	Uformer-Uformer	22.38±3.11	0.79±0.40	0.26±0.07
	cGAN	24.08±1.21	0.88±0.02	0.22±0.03
	Proposed	24.34±1.15	0.81±0.07	0.21±0.03

TABLE III
QUANTITATIVE RESULTS OF DIFFERENT NETWORK STRUCTURES

Methods	SNR	CNR	TP	ENL
UNet-UNet	46.52±3.18	7.13±0.93	0.89±0.17	74.29±23.58
Uformer-Uformer	48.68±4.91	7.64±0.82	0.77±0.09	85.72±48.05
Proposed without MTB	48.39±3.44	7.42±0.85	0.90±0.13	78.53±39.90
Proposed without modified MLP	51.36±2.93	7.16±1.03	0.93±0.08	83.65±32.74
Proposed	53.57±2.25	7.89±0.78	0.97±0.10	84.49±50.23

competitive performance compared with cGAN in our test dataset. However, cGAN requires a large amount of unlabeled training data and creates artifacts in the testing phase of the public dataset.

For qualitative evaluation, Fig. 3 presents the denoising results over a randomly selected OCT B-scan using different denoising approaches. The results show that the proposed method eliminates the speckle noise while preserving the detailed retinal structures well. It is observed that the proposed method shows promising performance in OCT image denoising and outperforms the traditional methods and shows comparable results as the supervised methods.

D. Extended Experiments

To further explore the sophisticated architecture of the proposed method, more experiments were conducted to explore the detailed design of the network structure with quantitative evaluations, as shown in Table III. First, we evaluated the general design of the co-training strategy and compared the proposed method with UNet only and with Uformer only. The results indicate that parallel training with two different streams outperforms training with streams of the same network structure. Moreover, the effectiveness of the proposed network is further verified with experiments with and without modified transformer block or MLP. It is obvious that the modified transformer block and MLP benefit the network and advance the performance of the proposed method. These results demonstrated the rationality and effectiveness of the proposed network structure.

Furthermore, the public dataset introduced in [5] is used to further explore the robustness of the proposed method. Fig. 4 shows some examples of denoising results from the public benchmark, including subjects from both humans and mice, indicating that the proposed method is capable of handling images from different OCT devices and different subjects. We speculate that the proposed method is also applicable in other image modalities to eliminate speckle noise, such as CT or ultrasound images. Future work may involve the possibility of exploring the application in other image modalities if data is available.

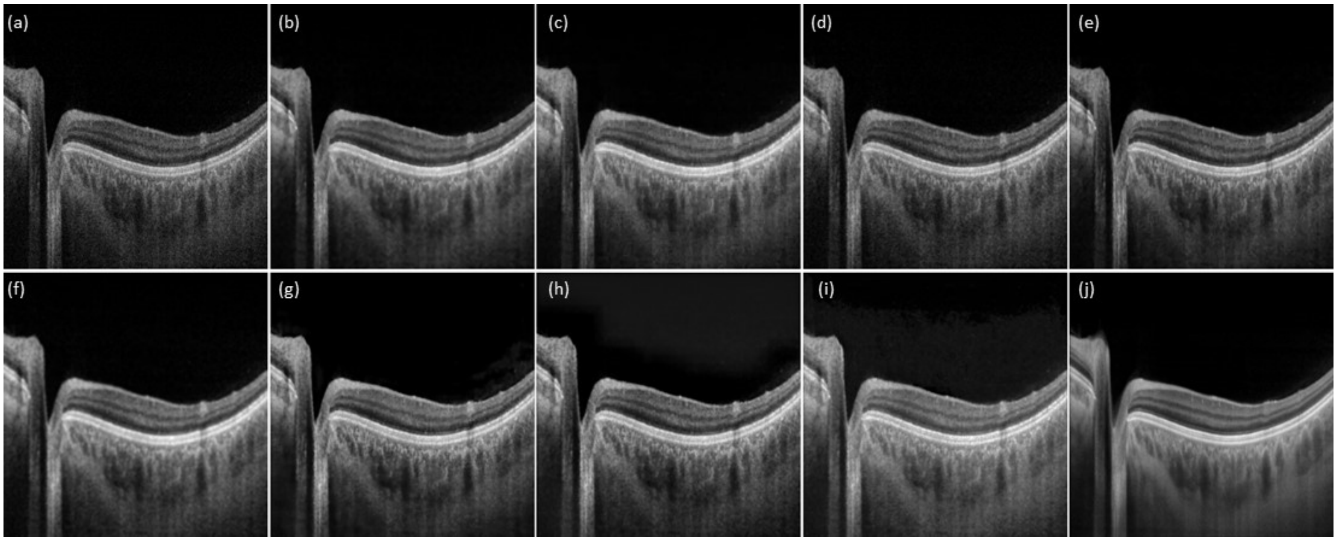

Fig. 3. Example of denoising results of different methods. (a) Original noisy image. (b) NLM. (c) Wavelet. (d) BM3D. (e) NCDF. (f) K-SVD. (g) UNet. (h) Uformer. (i) Proposed. (j) Ground truth.

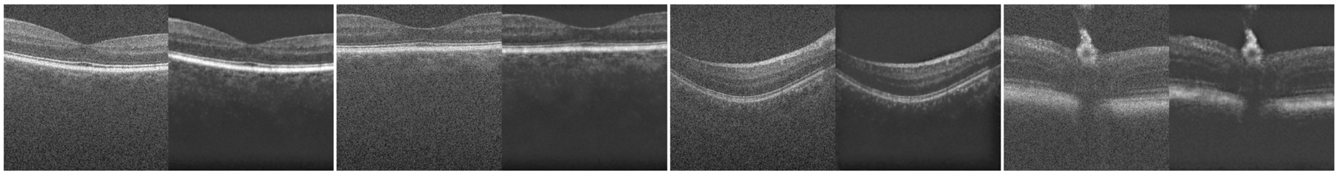

Fig. 4. Examples of denoising results with the proposed method on the public benchmark. In each panel, the original noisy image is on the left side and the denoised image is on the right side.

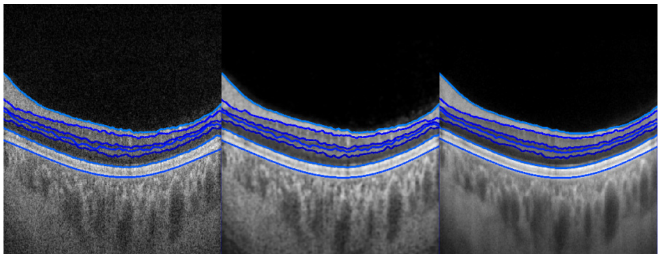

Fig. 5. Retinal segmentation results comparison. From left to right: noisy image, denoised image with the proposed method, ground-truth image.

TABLE IV
QUANTITATIVE RESULTS OF RETINAL LAYER THICKNESS

Method	Noisy	Denoised	Ground Truth
ILM – NFLGCL	26.74±18.81	27.82±18.91	26.48±18.63
NFLGCL – IPLINL	29.33±6.95	28.62±7.71	28.94±5.65
IPLINL - INLOPL	12.78±4.13	12.21±3.74	12.92±2.77
INLOPL - OPLONL	16.02±5.21	16.31±4.23	16.32±3.96
OPLONL - ISOS	34.39±4.83	34.02±4.17	33.79±3.91
ISOS - RPE	37.12±3.21	37.18±1.73	34.97±1.66

E. Application in Retinal Layer Segmentation

To demonstrate how the proposed method facilitates image analysis, more experiments were conducted for retinal OCT image segmentation. Retinal layer segmentation is an important task in OCT image analysis to assist clinical diagnosis. Clinicians can assess retinal diseases quantitatively through retinal layer thickness. We employ the segmentation method in [61] to segment retinal layers after speckle noise reduction with the proposed method in OCT images, to verify how the proposed method promotes OCT image analysis. The corresponding visual results are presented in Fig. 5. The objective indicators were also calculated as shown in Table IV. The results indicate the superior performance after denoising with the proposed method in the application of the retinal layer, showing the potential application in OCT image analysis enhancement.

V. CONCLUSION

In this article, a novel semi-supervised method for speckle noise reduction in OCT images was formulated based on the CNN and Transformer. With the co-training strategy between two streams, different learning paradigms are implemented, and cross-pseudo-labels are created. In this way, the proposed method achieves impressive speckle noise suppression results with a limited number of clean OCT images. The experimental results demonstrated the effectiveness and robustness of the proposed method in OCT image denoising. Overall, it is capable of improving contrast and smoothness, while preserving the detailed retinal features. With comparison, the proposed method outperforms the conventional methods and displays competitive performance as supervised methods. The generality of the proposed method is validated with both normal and pathological data from different OCT devices with different scales and resolutions. Furthermore, it illustrates how it facilitates OCT image analysis with an example of application in retinal layer segmentation.

Besides the promising performance we obtained so far in OCT image denoising, there are a few limitations of our study. First, the computational complexity is rather high during the training stage due to the parallel training design of the proposed method. Although the proposed method was trained with a limited amount of data, the dataset was selected manually according to different diseases and subjects and processed with augmentation. The data processing methods are quite common in this field [12], [62], [63]. Therefore, the images are representative and the variety of the dataset was increased. It shows robustness in both normal and pathological data from different OCT devices using both clinical and public datasets. Nevertheless, we believe that the performance could be further improved with more training data available and other supervised learning methods could be further attempted with more data. To address these problems, future work involves further exploration with more various datasets and more application scenarios, including data from different OCT devices and other medical image modalities such as computed tomography, ultrasound images, and full-field OCT. Moreover, since the proposed method shows the capability of processing multiscale OCT images, future work might involve resolution enhancement with modifications of the proposed method to obtain images of high resolutions with fast scanning. Besides the application with retinal layer segmentation presented in the article, we will also further explore the contribution of the proposed method to other specific applications, such as manual diagnosis of certain ophthalmic pathologies, automatic segmentation of retinal lesions, and so on.

In conclusion, a novel semi-supervised speckle noise reduction method for OCT images was proposed to solve the dilemma of lacking clean OCT images. With co-training between the CNN and Transformer, the proposed scheme encourages consistency between different streams when the advantages of both are leveraged to compensate each other for better convergence. The bidirectional loss is formulated to effectively weigh the supervised and unsupervised loss. The two different learning paradigms learn from each other and reach consistency without the enforcement of explicit constraints. Through thorough experiments and comparisons, it is verified that the proposed method outperforms conventional speckle noise reduction methods and shows competitive results with supervised strategies. In the future, we will extensively explore the generality of different datasets and the possibility of resolution enhancement for OCT images.

References

[1] D. Huang et al., "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[2] J. M. Schmitt, S. H. Xiang, and K. M. Yung, "Speckle in optical coherence tomography," *J. Biomed. Opt.*, vol. 4, no. 1, p. 95, 1999.

[3] M. Bashkansky and J. Reintjes, "Statistics and reduction of speckle in optical coherence tomography," *Opt. Lett.*, vol. 25, pp. 545–547, May 2000.

[4] Y. Huang et al., "Noise-powered disentangled representation for unsupervised speckle reduction of optical coherence tomography images," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2600–2614, Oct. 2021.

[5] L. Fang et al., "Fast acquisition and reconstruction of optical coherence tomography images via sparse representation," *IEEE Trans. Med. Imag.*, vol. 32, no. 11, pp. 2034–2049, Nov. 2013.

[6] R. D. Ferguson, D. X. Hammer, L. A. Paunescu, S. Beaton, and J. S. Schuman, "Tracking optical coherence tomography," *Opt. Lett.*, vol. 29, no. 18, p. 2139, Sep. 2004.

[7] S. Chitchian, M. A. Mayer, A. R. Boretsky, F. J. van Kuijk, and M. Motamedi, "Retinal optical coherence tomography image enhancement via shrinkage denoising using double-density dual-tree complex wavelet transform," *J. Biomed. Opt.*, vol. 17, no. 11, Nov. 2012, Art. no. 116009.

[8] J. Cheng et al., "Speckle reduction in 3D optical coherence tomography of retina by A-scan reconstruction," *IEEE Trans. Med. Imag.*, vol. 35, no. 10, pp. 2270–2279, Oct. 2016.

[9] W. Wu, O. Tan, R. R. Pappuru, H. Duan, and D. Huang, "Assessment of frame-averaging algorithms in OCT image analysis," *Ophthalmic Surg., Lasers Imag. Retina*, vol. 44, no. 2, pp. 168–175, Mar. 2013.

[10] T. Klein, R. André, W. Wieser, T. Pfeiffer, and R. Huber, "Joint aperture detection for speckle reduction and increased collection efficiency in ophthalmic MHz OCT," *Biomed. Opt. Exp.*, vol. 4, no. 4, p. 619, Apr. 2013.

[11] L. Fang, S. Li, Q. Nie, J. A. Izatt, C. A. Toth, and S. Farsiu, "Sparsity based denoising of spectral domain optical coherence tomography images," *Biomed. Opt. Exp.*, vol. 3, no. 5, p. 927, May 2012.

[12] M. Tajmirriahi, R. Kafieh, Z. Amini, and H. Rabbani, "A lightweight mimic convolutional auto-encoder for denoising retinal optical coherence tomography images," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–8, 2021.

[13] S. K. Devalla et al., "A deep learning approach to denoise optical coherence tomography images of the optic nerve head," *Sci. Rep.*, vol. 9, no. 1, p. 14454, Oct. 2019.

[14] W. Zhao and H. Lu, "Medical image fusion and denoising with alternating sequential filter and adaptive fractional order total variation," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 9, pp. 2283–2294, Sep. 2017.

[15] J. Aum, J.-H. Kim, and J. Jeong, "Effective speckle noise suppression in optical coherence tomography images using nonlocal means denoising filter with double Gaussian anisotropic kernels," *Appl. Opt.*, vol. 54, no. 13, p. D43, May 2015.

[16] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[17] M. Li, R. Idoughi, B. Choudhury, and W. Heidrich, "Statistical model for OCT image denoising," *Biomed. Opt. Exp.*, vol. 8, no. 9, p. 3903, Sep. 2017.

[18] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, Jan. 2013.

[19] M. A. Mayer, A. Borsdorf, M. Wagner, J. Hornegger, C. Y. Mardin, and R. P. Tornow, "Wavelet denoising of multiframe optical coherence tomography data," *Biomed. Opt. Exp.*, vol. 3, no. 3, p. 572, Mar. 2012.

[20] R. Bernardes, C. Maduro, P. Serranho, A. Araújo, S. Barbeiro, and J. Cunha-Vaz, "Improved adaptive complex diffusion despeckling filter," *Opt. Exp.*, vol. 18, no. 23, p. 24048, Nov. 2010.

[21] A. Abbasi and A. Monadjemi, "Optical coherence tomography retinal image reconstruction via nonlocal weighted sparse representation," *J. Biomed. Opt.*, vol. 23, no. 3, p. 1, Mar. 2018.

[22] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[23] R. Kafieh, H. Rabbani, and I. Selesnick, "Three dimensional data-driven multi scale atomic representation of optical coherence tomography," *IEEE Trans. Med. Imag.*, vol. 34, no. 5, pp. 1042–1062, May 2015.

[24] Y. Luo, Q. Xu, R. Jin, M. Wu, and L. Liu, "Automatic detection of retinopathy with optical coherence tomography images via a semi-supervised deep learning method," *Biomed. Opt. Exp.*, vol. 12, no. 5, p. 2684, May 2021.

[25] S. Sedai et al., "Uncertainty guided semi-supervised segmentation of retinal layers in OCT images," in *Proc. MICCAI*. Cham, Switzerland: Springer, 2019, pp. 282–290.

[26] X. Liu et al., "Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning," *IEEE Access*, vol. 7, pp. 3046–3061, 2019.

[27] L. Pan, L. Guan, and X. Chen, "Segmentation guided registration for 3D spectral-domain optical coherence tomography images," *IEEE Access*, vol. 7, pp. 138833–138845, 2019.

[28] X. Zhu and A. B. Goldberg, *Introduction to Semi-Supervised Learning*. San Rafael, CA, USA: Morgan & Claypool, 2009.

[29] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Workshop Challenges Represent. Learn.*, 2013, p. 7.

[30] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. 11th Annu. Conf. Comput. Learn. theory*, Jul. 1998, pp. 92–100.

[31] T. Miyato, S.-I. Maeda, M. Koyama, and S. Ishii, "Virtual adversarial training: A regularization method for supervised and semi-supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1979–1993, Aug. 2019.

[32] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," 2017, *arXiv:1703.01780*.

[33] K. Sohn et al., "FixMatch: Simplifying semi-supervised learning with consistency and confidence," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 596–608.

[34] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.

[35] 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.

[36] D. Yin et al., "Speckle noise reduction in coherent imaging based on deep learning without clean data," *Opt. Lasers Eng.*, vol. 133, Oct. 2020, Art. no. 106151.

[37] A. Guo, L. Fang, M. Qi, and S. Li, "Unsupervised denoising of optical coherence tomography images with nonlocal-generative adversarial network," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[38] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proc. IEEE Int. Symp. Circuits Syst.*, Paris, France, May 2010, pp. 253–256.

[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[40] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3859–3869.

[41] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.

[42] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[43] H. Wu et al., "CvT: Introducing convolutions to vision transformers," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, QC, Canada, Oct. 2021, pp. 22–31.

[44] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* (Lecture Notes in Computer Science), vol. 9351, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds. Cham, Switzerland: Springer, Oct. 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[45] O. Petit, N. Thome, C. Rambour, L. Themyr, T. Collins, and L. Soler, "U-Net transformer: Self and cross attention for medical image segmentation," in *Machine Learning in Medical Imaging*, vol. 12966, C. Lian, X. Cao, I. Rekik, X. Xu, and P. Yan, Eds. Cham, Switzerland: Springer, 2021, pp. 267–276.

[46] C.-M. Fan, T.-J. Liu, and K.-H. Liu, "SUNet: Swin transformer UNet for image denoising," 2022, *arXiv:2202.14009*.

[47] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using Swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Montreal, BC, Canada, Oct. 2021, pp. 1833–1844.

[48] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," 2021, *arXiv:2106.01226*.

[49] B. Han et al., "Co-teaching: Robust training of deep neural networks with extremely noisy labels," 2018, *arXiv:1804.06872*.

[50] J. Choi, T. Kim, and C. Kim, "Self-ensembling with GAN-based data augmentation for domain adaptation in semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6829–6839.

[51] A. Bruhn, J. Weickert, and C. Schnörr, "Combining the advantages of local and global optic flow methods," in *Pattern Recognition*, vol. 2449, L. Van Gool, Ed. Berlin, Heidelberg: Springer, 2002, pp. 454–462.

[52] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and accurate image super-resolution with deep Laplacian pyramid networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2599–2613, Nov. 2019.

[53] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, *arXiv:1505.00853*.

[54] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general U-shaped transformer for image restoration," 2021, *arXiv:2106.03106*.

[55] T. Huang, S. Li, X. Jia, H. Lu, and J. Liu, "Neighbor2Neighbor: A self-supervised framework for deep image denoising," *IEEE Trans. Image Process.*, vol. 31, pp. 4023–4038, 2022.

[56] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," 2016, *arXiv:1606.08415*.

[57] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017, *arXiv:1711.05101*.

[58] A. Pizurica et al., "Multiresolution denoising for optical coherence tomography: A review and evaluation," *Current Med. Imag. Rev.*, vol. 4, no. 4, pp. 270–284, Nov. 2008.

[59] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Istanbul, Turkey, Aug. 2010, pp. 2366–2369.

[60] Y. Ma, X. Chen, W. Zhu, X. Cheng, D. Xiang, and F. Shi, "Speckle noise reduction in optical coherence tomography images based on edge-sensitive cGAN," *Biomed. Opt. Exp.*, vol. 9, no. 11, p. 5129, 2018.

[61] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation," *Opt. Exp.*, vol. 18, no. 18, pp. 19413–19428, 2010.

[62] Y. Ma, Q. Yan, Y. Liu, J. Liu, J. Zhang, and Y. Zhao, "StruNet: Perceptual and low-rank regularized transformer for medical image denoising," *Med. Phys.*, vol. 50, no. 12, pp. 7654–7669, Dec. 2023.

[63] S. Li, M. Xiong, B. Yang, X. Zhang, R. Higashita, and J. Liu, "Oct image blind despeckling based on gradient guided filter with speckle statistical prior," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Rhodes Island, Greece, Jun. 2023, pp. 1–5.