Toward Ground-Truth Optical Coherence Tomography via Three-Dimensional Unsupervised Deep Learning Processing and Data

Guangming Ni, *Member, IEEE*, Renxiong Wu, Fei Zheng, Meixuan Li, Shaoyan Huang, Xin Ge, Linbo Liu, *Member, IEEE*, and Yong Liu

***Abstract*— Optical coherence tomography (OCT) can perform non-invasive high-resolution three-dimensional (3D) imaging and has been widely used in biomedical fields, while it is inevitably affected by coherence speckle noise which degrades OCT imaging performance and restricts its applications. Here we present a novel speckle-free OCT imaging strategy, named toward-ground-truth OCT (*t*GT-OCT), that utilizes unsupervised 3D deep-learning processing and leverages OCT 3D imaging features to achieve speckle-free OCT imaging. Specifically, our proposed *t*GT-OCT utilizes an unsupervised 3D-convolution deep-learning network trained using random 3D volumetric data to distinguish and separate speckle from real structures in 3D imaging volumetric space; moreover, *t*GT-OCT effectively further reduces speckle noise and reveals structures that would otherwise be obscured by speckle noise while preserving spatial resolution. Results derived from different samples demonstrated the high-quality speckle-free 3D imaging performance of *t*GT-OCT and its advancement beyond the previous state-of-the-art. The code is available online: https://github.com/Voluntino/tGT-OCT.**

***Index Terms*— Optical coherence tomography, speckle noise, unsupervised learning, volumetric data.**

I. Introduction

OPTICAL coherence tomography (OCT) stands as a pivotal noninvasive biomedical imaging technology and can produce three-dimensional images of various biological tissues with micrometer resolution [1], [2]. Beyond its established utility in ophthalmology, OCT has progressively gained popularity in clinical diagnostic applications across cardiology, dermatology, and other fields [3], [4]. However, an intrinsic challenge of OCT lies in the presence of speckle noise within OCT images, which noticeably compromises their quality and consequently impairs subsequent interpretation and diagnosis.

Many speckle-reduction methods for OCT imaging have been proposed; among these, speckle-modulating OCT (SM-OCT) is a well-regarded method [2]. By using a moving diffuser, SM-OCT can acquire an unlimited number of uncorrelated speckle patterns and effectively remove speckle noise without degrading the spatial resolution of the images, which helps SM-OCT clarify and reveal structures that are otherwise obscured or undetectable. However, SM-OCT uses a moving diffuser in the optical path and has to perform repeated B-scans, which substantially reduces the imaging sensitivity and temporal resolution of the OCT system.

Recently, methods that leverage the robust data-fitting capabilities of deep learning, have been widely used for OCT despeckling. With the rapid development of deep learning-based methods, two categories have emerged: clean-image-required and clean-image-free methods [5]. One of the clean-image-required methods is using noisy-clean pairs to train a supervised network. Specifically, Zhang et al. utilized residual learning and proposed DnCNN to realize noise reduction [6]. Additionally, the generative adversarial network (GAN) has gained immense popularity for supervised OCT despeckling methods, giving rise to approaches such as Caps-cGAN [7], SDSR-OCT [8], DNGAN [9], SiameseGAN [10], and MDR-GAN [11] that have outperformed conventional CNNs in enhancing image quality. Sm-Net OCT [12] involves training a GAN with a customized SM-OCT dataset to despeckle and resolve intricate structures. The other clean-image-required methods needs unpaired clean and noisy images to train a cycleGAN-based network, such as HDcycleGAN [13], SPcycleGAN [14], DRGAN [15], and ADGAN [16]. These unpaired methods disentangle OCT

Manuscript received 14 December 2023; revised 23 January 2024; accepted 2 February 2024. Date of publication 7 February 2024; date of current version 3 June 2024. This work was supported in part by the National Science Foundation of China under Grant 61905036, in part by the China Postdoctoral Science Foundation under Grant 2021T140090 and Grant 2019M663465, in part by the Fundamental Research Funds for the Central Universities (University of Electronic Science and Technology of China) under Grant ZYGX2021J012, and in part by the Medico-Engineering Cooperation Funds from the University of Electronic Science and Technology of China under Grant ZYGX2021YGCX019. *(Guangming Ni and Renxiong Wu are co-first authors.) (Corresponding author: Guangming Ni.)*

Guangming Ni, Renxiong Wu, Fei Zheng, Meixuan Li, Shaoyan Huang, and Yong Liu are with the School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: guangmingni@uestc.edu.cn).

Xin Ge is with the School of Science, Sun Yat-sen University, Shenzhen Campus, Shenzhen 510275, China.

Linbo Liu is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798.

Digital Object Identifier 10.1109/TMI.2024.3363416

images into content and noise domains, approximating the denoising performance of supervised methods. Nonetheless, the dependency of these methods on clean images, which are often challenging to acquire due to repeated lengthy scanning procedures, motivates exploration into methods that don't need clean images.

Consequently, several clean-image-free methods have been explored to address this concern. The Noise2Noise (N2N) strategy [17], using pairs of noisy images of the same scene, has been applied to OCT despeckling [18], [19], [20]. Nevertheless, the reliance of this approach on at least two scans of the same sample location remains a practical barrier. Furthermore, there are single-image denoising techniques that are notable in natural image processing; these include Noise2Void (N2V) [21], Noise2Self (N2S) [22], Neighbor2Neighbor (NBR) [23], etc. In OCT despeckling, the use of NBR, a method with a multiscale pixel patch sampler [24] ensures both despeckling efficacy and structure preservation. The application of Transformer [25], [26] has also been attempted to reduce speckle noise sufficiently and preserve details. These proposed methods represent great efforts to alleviate the problem of detail structure damage in single-image denoising methods but show limited improvement. To resolve finer biological detail structures, Noise2Context [27], Noise2Stack [28], and Noise2Sim [29] have been proposed to augment denoised image details by exploiting shared information from adjacent noisy images within 3D volumetric data. However, reliance solely on short-distance adjacent slices fails to efficiently utilize all of the available 3D OCT data and has limited despeckling performance, especially in non-ophthalmological OCT imaging applications.

Here we present a novel strategy, called to toward-ground-truth OCT (*t*GT-OCT), to achieve speckle-free OCT imaging, that is based on OCT 3D imaging features and unsupervised 3D deep-learning processing. By distinguishing and extracting speckles and structures in random OCT 3D imaging volumetric data using 3D-convolutional neural networks (3D-CNNs), the proposed *t*GT-OCT effectively reduces speckle noise and reveals structures that are otherwise obscured or undetectable while preserving spatial resolution. Qualitative and quantitative results for various 3D OCT images including those of the human retina, other human tissues, meats, and Scotch tape demonstrate the state-of-the-art performance of *t*GT-OCT in despeckling, even achieving microstructure resolution performance comparable to that of SM-OCT, which is regarded as the ideal for OCT despeckling. Meanwhile, our work has also provided a new perspective for studying OCT speckle-free imaging by utilizing the 3D imaging characteristics of OCT alongside unsupervised 3D deep learning processing.

II. Methods

A. 3D Unsupervised Deep Learning Extracts OCT Speckle Patterns in 3D Space

Speckle-modulating OCT uses a moving diffuser and requires repeated scanning of the sample to acquire a large number of uncorrelated speckle patterns and then performs

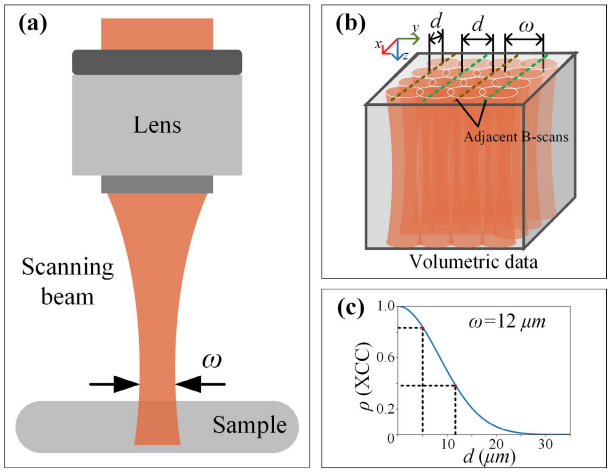

Fig. 1. The correlation of adjacent A-scans depends on the limited beam size and appropriate lateral displacement. (a) The scanning beam size; (b) volumetric data containing uncorrelated and weakly correlated speckle patterns and strongly correlated sample structures when the displacement d is smaller than beam size; (c) the relationship between XCC and lateral displacement.

an averaging operation to effectively remove speckle noise without degrading the spatial resolution of the images. Our previously proposed method, Sm-Net OCT [12], uses deep-learning network to distinguish and extract those large number of uncorrelated speckle patterns in SM-OCT speckle images with SM-OCT speckle-free images (used as ground truth) to generate speckle-free OCT images. OCT speckle patterns depend on scanning voxel sizes and sample structures [2], [30], so OCT 3D volumetric data can contain unlimited uncorrelated speckle patterns; moreover, these speckle patterns can potentially be distinguished and separated in 3D space for OCT speckle-free imaging. Meanwhile, neighboring B-scans also contain mass strongly correlated sample structures in OCT 3D volumetric data, as shown in Fig. 1.

In OCT 3D imaging, the correlation of two A-scans can be expressed using the Pearson cross-correlation coefficient (XCC) [31]. Importantly, the XCC between two adjacent slightly displaced A-scans has an explicit functional dependency on the lateral distance d and can be expressed as (1), where ω, is the Gaussian beam waist of the light beam, which is also the transverse optical resolution of the OCT system.

$$\rho = \exp\left(-\frac{d^2}{\omega^2}\right) \tag{1}$$

As shown in Figs. 1(b) and (c), owing to the limited scanning volume size of OCT, adjacent A-scans can contain both uncorrelated and weakly correlated speckle patterns at suitably minor lateral displacements d, while neighboring B-scans can have strongly correlated structures [32], [33], [34], [35]. Therefore, OCT volumetric data can contain mass uncorrelated speckle patterns and strongly correlated structural information. The strongly correlated structural information can further act as the ground truth for our proposed unsupervised 3D deep learning network described in the following sections; this ground truth helps 3D deep learning network to

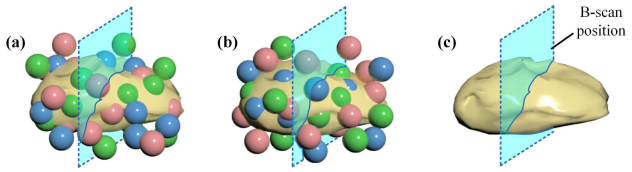

Fig. 2. Schematic of clean volumetric data and paired noisy volumetric data. (a) and (b) are the paired noisy volumes with uncorrelated speckle patterns, (c) is the clean volume. The color balls indicate various speckle noise, the yellow volume indicates target structure and the cyan rectangles mean that the OCT volume consists of multiple B-scans.

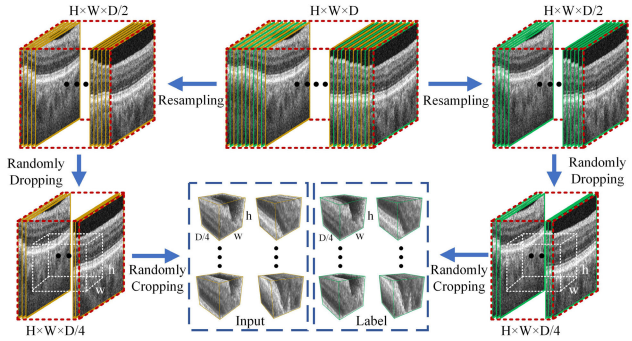

Fig. 3. The pipeline of conducting paired 3D volumetric data. The original volumetric data are resampled adjacently to two similar volumetric data and then divided into paired volumetric blocks by randomly dropping and cropping. one of the paired data is the input and another is the label for training.

further distinguish and extract speckle patterns and generate speckle-free OCT images.

Unsupervised networks [17], [23] analyzing two similar (strongly correlated sample structures) but different noisy (uncorrelated speckle patterns) 2D images have been proposed to reduce noise. Here we extend this principle to the realm of 3D volumetric data. First we consider two noisy volumetric image sets $\mathbf{y} = [\mathbf{y}_0, \mathbf{y}_1, \cdots, \mathbf{y}_{n-1}]$ and $\mathbf{z} = [\mathbf{z}_0, \mathbf{z}_1, \cdots, \mathbf{z}_{n-1}]$, which are sequences of n multiple B-scans. The noisy data $\mathbf{y}$ and $\mathbf{z}$ are independent conditional on the clean data $\mathbf{x} = [\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_{n-1}]$. We aim to train a 3D network $f_\theta(\cdot)$ parametrized by θ by minimizing the function as (2). Fig. 2 shows the schematic of the clean data $\mathbf{x}$ and noisy data $\mathbf{y}$ and $\mathbf{z}$ with substantial uncorrelated speckle patterns.

$$\arg\min_{\theta} \mathbb{E}\| f_\theta(\mathbf{y}) - \mathbf{z}\|^2 \tag{2}$$

Assumed that $\mathbb{E}(\mathbf{y}) = \mathbf{x}$ and $\mathbb{E}(\mathbf{z}) = \mathbf{x} + \varepsilon$, where ε is the gap between the underlying clean data of paired noisy volumetric data $\mathbf{y}$ and $\mathbf{z}$. In general case, $\varepsilon = \mathbb{E}(\mathbf{z}) - \mathbb{E}(\mathbf{y}) \neq 0$. $\mathbb{E}\| f_\theta(\mathbf{y}) - \mathbf{z}\|^2$ can be expressed as

$$\begin{aligned}\mathbb{E}\| f_\theta(\mathbf{y}) - \mathbf{z}\|^2 =& \mathbb{E}\| f_\theta(\mathbf{y}) - \mathbf{x} + \mathbf{x} - \mathbf{z}\|^2 \\ =& \mathbb{E}\| f_\theta(\mathbf{y}) - \mathbf{x}\|^2 + \mathbb{E}\|\mathbf{z} - \mathbf{x}\|^2 \\ & - 2\mathbb{E}\left((f_\theta(\mathbf{y}) - \mathbf{x})^\top (\mathbf{z} - \mathbf{x}) \right) \\ =& \mathbb{E}\| f_\theta(\mathbf{y}) - \mathbf{x}\|^2 + {\sigma_z}^2 \\ & - 2\mathbb{E}\left((f_\theta(\mathbf{y}) - \mathbf{x})^\top (\mathbf{z} - \mathbf{x}) \right). \end{aligned} \tag{3}$$

Due to $\mathbf{y}$ and $\mathbf{z}$ are independent condition on $\mathbf{x}$, we have

$$\begin{aligned}\mathbb{E}\| f_\theta(\mathbf{y}) - \mathbf{z}\|^2 =& \mathbb{E}\| f_\theta(\mathbf{y}) - \mathbf{x}\|^2 + {\sigma_\mathbf{z}}^2 \\ & - 2\mathbb{E}(f_\theta(\mathbf{y}) - \mathbf{x})^\top \mathbb{E}(\mathbf{z} - \mathbf{x}) \\ =& \mathbb{E}\| f_\theta(\mathbf{y}) - \mathbf{x}\|^2 + {\sigma_\mathbf{z}}^2 \\ & - 2\varepsilon \mathbb{E}(f_\theta(\mathbf{y}) - \mathbf{x}), \end{aligned} \tag{4}$$

where $\mathbb{E}\| f_\theta(\mathbf{y}) - \mathbf{x}\|^2$ represents the loss function for training a supervised network using clean images, ${\sigma_\mathbf{z}}^2$ is the variance of $\mathbf{z}$ and it is a constant. Note that if the gap $\varepsilon \to 0$, $\arg\min_\theta \mathbb{E}\| f_\theta(\mathbf{y}) - \mathbf{z}\|^2$ converges to $\arg\min_\theta \mathbb{E}\| f_\theta(\mathbf{y}) - \mathbf{x}\|^2$. In other words, the label for training unsupervised network can be noisy data that are similar to the input noisy data rather than clean data.

B. Conduction of Paired 3D Volumetric Training Data

Here, we prepared paired 3D volumetric data that were similar but not identical for training our unsupervised network; pairs were conducted by processing one OCT 3D volumetric dataset with strongly correlated sample structures and uncorrelated speckle patterns at suitably minor lateral displacements. Fig. 3 shows the processing pipeline for conducting paired data. Specifically, the noisy 3D data with width W, height H and depth D are processed using three steps: (1) The 3D data is divided into two sub-datasets by selecting adjacent B-scans along the lateral direction. One subdata point consists of the (2i-1)-th B-scan while the other subdata point consists of the 2i-th B-scan from the original 3D data, where $i = 1, 2 \cdots, D/2$. (2) B-scans in the subdata are randomly dropped, and in the experiment, either one of two B-scans or two of four B-scans are dropped without repetition. (3) The paired 3D input and label are conducted by random cropping the images with a cropping block of size (h, w, $D/4$). As neighboring B-scans in the original 3D data contain strongly correlated structures, resampling and dropping operations ensure that the paired volumetric data retain similar sample structures but different speckle patterns.

C. Three-Dimensional Datasets

Here we employed two customized volumetric datasets for both training and test, as well as two additional volumetric datasets used exclusively for testing the generalization performance. Detailed characteristics of these datasets are summarized in Table I. The data for the OCT-R1 dataset were collected from 41 human eyes using a OCT scanner (BM-400K BMizar, Topi) at Sichuan Provincial People's Hospital (IRB-2022-258). To enhance the diversity of the data, we conducted scans over two different ranges. For the OCT-N1 dataset, we used a customized OCT setup [1] to collect 46 three-dimensional data from different samples, including Scotch tape, pork, human skin, and placenta. The publicly available dataset OCTA-500 [36] from Optovue OCT devices was also used to verify the adaptability of our proposed method. 200 OCT volumes with a FOV of 3 mm × 3 mm were selected from 500 volumes for training and testing. The OCT-R2 dataset includes five three-dimensional images of the human retina acquired using OCT scanner (Spectralis, Heidelberg). The OCT-N2 dataset contains six 3D data (three noisy data and three clean data) of samples such as Scotch tape, fish, and pork, which were collected by our SM-OCT

TABLE I
THE DETAILS OF TRAINING AND TEST DATASETS

Dataset	OCT Setup	Axial resolution	Lateral displacement	Data size	Subject	Use
OCT-R1	BM-400K BMizar	~3.8 μm	10.0 μm	512×512×512	25	training & test
				1948×1536×1280	16	
OCT-N1	Customized OCT	~1.68 μm	5.10 μm	800×800×800	46	training & test
OCTA-500	RTVue-XR	~3.12 μm	9.87 μm	640×304×304	200	training & test
OCT-R2	Spectralis	~4.1 μm	11.8 μm	512×512×400	5	only test
OCT-N2	Customized SM-OCT	~1.68 μm	5.10 μm	800×420×100	6	only test

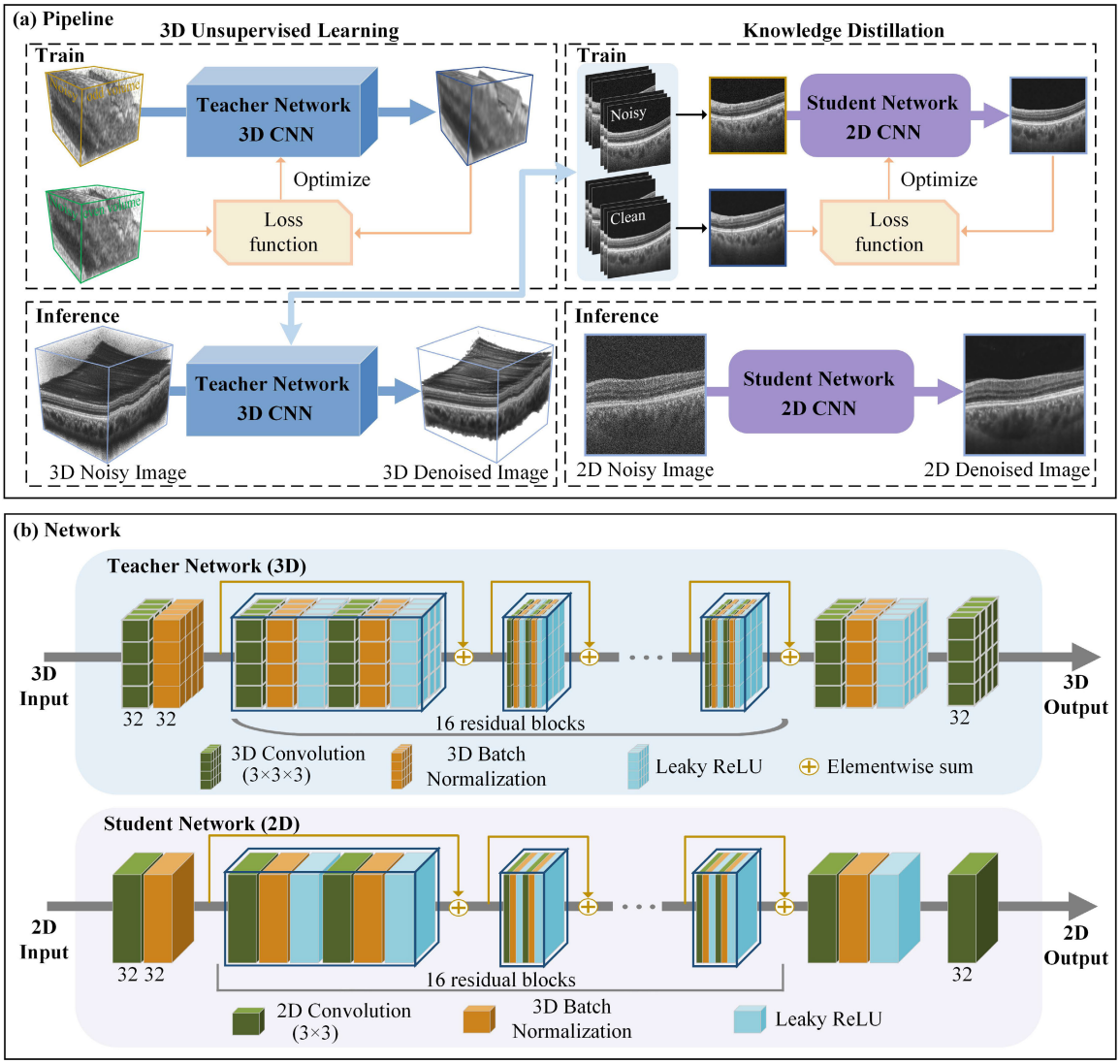

Fig. 4. Pipeline and networks of the proposed *t*GT-OCT. (a) Pipeline, and (b) networks. The proposed *t*GT-OCT contains a 3D CNN with unsupervised learning and a 2D CNN with knowledge distillation. The 3D CNN is trained using paired noisy volumetric data and the noisy-clean pairs for training the 2D CNN are selected from noisy volumes and denoised volumes generated by 3D CNN. During inference and application, the 3D CNN and 2D CNN are employed according to the dimensions of input data and computing resources.

setup [2]. For the generation of clean data references using SM-OCT, we shifted the optical diffuser and simultaneously scanned the same position for 50 times. Data in OCT-R2, OCT-N2 and test sets of OCT-R1, OCT-N1 were only used for the evaluation phase. The datasets OCT-R1[1], OCT-R2[1] and OCTA-500[2] are available for download.

[1]https://tianchi.aliyun.com/dataset/161472
[2]https://ieee-dataport.org/open-access/octa-500

D. Network

Fig. 4(a) illustrates the pipeline of our novel *t*GT-OCT strategy. This comprehensive strategy consists of both a 3D CNN and a 2D CNN to accommodate OCT data with varying dimensions, i.e., for 3D volumetric inputs and single-frame inputs respectively. The 3D CNN is trained with an unsupervised learning strategy that uses paired similar noisy volumetric data: one of which is considered the input while the other is the target. Considering that 3D convolutions

need large computing resources and existing images of public OCT datasets are two-dimensions, we introduce a knowledge distillation mechanism to distill the knowledge of the 3D CNN (as a teacher network) into the 2D CNN (as a student network). For training the 2D CNN, noisy B-scans selected from noisy volume and corresponding clean B-scans selected from denoised volume generated by the 3D CNN are as input and label of the 2D CNN. During inference, the teacher and student networks are employed for denoising single-frame or 3D-volume, respectively.

1) Unsupervised 3D CNN: The presented 3D CNN mentioned above is an evolution of the deep residual network (ResNet), which is commonly used for image super-resolution and denoising. As shown in Fig. 4(b), ResNet consists of a pre-residual layer, sixteen residual blocks and a post-residual layer. The pre-residual layer was a 3D convolutional layer and was followed by a batch normalization layer. The 3D convolutional layer has 32 filters, each with $3 \times 3 \times 3$ kernel size. Each of the residual blocks has two 3D convolutional layers that are identical to the pre-residual layer, followed by batch normalization layers; the network uses the leaky ReLU function as the activation layer. The post-residual layer consists of a $3 \times 3 \times 3$ 3D convolutional layer, batch normalization layer, leaky ReLU function, and a $1 \times 1 \times 1$ 3D convolutional layer. A skip connection is introduced in each residual block to connect the input and output of the block. The size of the feature images passed in the network remains invariant, sustaining a size of $64 \times 64 \times 64$ during training.

2) Knowledge Distillation: In recognition of the resource-intensive nature of 3D convolution and the existence of OCT datasets that only include single noisy images, we introduced a knowledge distillation mechanism to distill the knowledge of the 3D CNN into a 2D CNN. As shown in Fig. 4(b), we denoted the 3D CNN as the teacher network and the 2D CNN as the student network. One noisy B-scan and corresponding denoised B-scan were selected from noisy volumetric data and denoised volumetric data generated by the teacher network, respectively. These frames were used to train the student network with a supervised learning strategy. To retain as many feature maps as possible during knowledge distillation, the structure of the student network was aligned with the teacher network, changing 3D convolution layers in the teacher network to 2D convolution layers. All convolutional layers are 3×3 in size except for the last layer whose size is 1×1.

3) Objective Functions: The unsupervised learning process of the 3D CNN follows the principle of (2), the objective function is defined as (5):

$$\mathcal{L}_{3D} = \frac{1}{HWF} \sum_{i,j,k=1}^{H,W,F} \left\| v_{i,j,k} - p_{i,j,k} \right\|^2, \tag{5}$$

where $v_{i,j,k}$ and $p_{i,j,k}$ are pixels of two similar noisy volumetric data and H, W, and F are the height, width, and frames of the volumetric data, respectively.

For 2D CNN, using mean square error (MSE) and VGG loss is beneficial to improve image quality and maintain high-frequency information [37]. VGG loss is the Euclidean distance between the high-level perceptual features of the generated image and the ground truth extracted by the VGG network. The loss function of 2D CNN can be expressed as:

$$\mathcal{L}_{2D} = \alpha \mathcal{L}_{MSE} + \beta \mathcal{L}_{Vgg}, \tag{6}$$

where α, β are weight coefficients of loss term. $\mathcal{L}_{MSE}$ and $\mathcal{L}_{Vgg}$ are MSE loss and VGG loss, which are defined as:

$$\mathcal{L}_{MSE} = \frac{1}{HW} \sum_{i,j=1}^{H,W} \left\| x_{i,j} - y_{i,j} \right\|^2, \tag{7}$$

$$\mathcal{L}_{Vgg} = \frac{1}{HW} \sum_{i,j=1}^{H,W} \left\| VGG_{16}(x)_{i,j} - VGG_{16}(y)_{i,j} \right\|^2, \tag{8}$$

where VGG_{16} represents the VGG-16 network.

III. Experiment and Results

A. Experimental Setup

We implemented the proposed tGT-OCT 3D deep learning network in the TensorFlow framework. The training was conducted on a platform equipped with an Intel Xeon Silver 4210R CPU, an NVIDIA Quadro RTX A6000 graphics card boasting 48 GB of memory, and a total of 64 GB RAM. Adam was adopted as the optimizer with the momentum 1 and momentum 2 parameters set to 0.5 and 0.999, respectively. The learning rate was set to 5×10^{-5}. The batch size and the number of iterations for training teacher network and student network were set to 4 and 100,000, 8 and 150,000 respectively. The total time for training the teacher network and student network was 30.5 hours on our computing platform. The weights α, β in (6) were set to 0.5 and 1 according to the results of several experiments. We trained two 3D models using the training sets of OCT-R1 and OCT-N1. For testing purposes, we selected 8 volumes in OCT-R1, 8 volumes in OCT-N1, and 10 volumes in OCTA-500, while the remaining volumes were used for training. All volumes in the training sets were denoised by 3D CNN and their B-scans were unpacked to conduct the dataset for training 2D CNN (OCT-R1-KD, OCT-N1-KD and OCTA-500-KD). We excluded B-scans containing large background regions, and the remaining noisy-denoised image pairs were divided into a training set and a validation set at a ratio of 8:2. To expedite training and streamline computational resource utilization, the length, width and height of the volumetric data in training set were randomly cropped to $64 \times 64 \times 64$. The 2D single-frame OCT images were cropped and padded to 480×480. When testing networks, we input the 3D volumes and 2D images in the test sets without cropping the width and height.

When there was no ground truth, signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR) and equivalent number of looks (ENL) [38] were utilized to evaluate the denoising performance. SNR indicates the ratio of signal energy and noise energy, CNR measures the contrast between a feature of interest and background noise and ENL measures smoothness in homogeneous areas. SNR, CNR and ENL are no-reference assessment metrics that give a limited indication of the degree

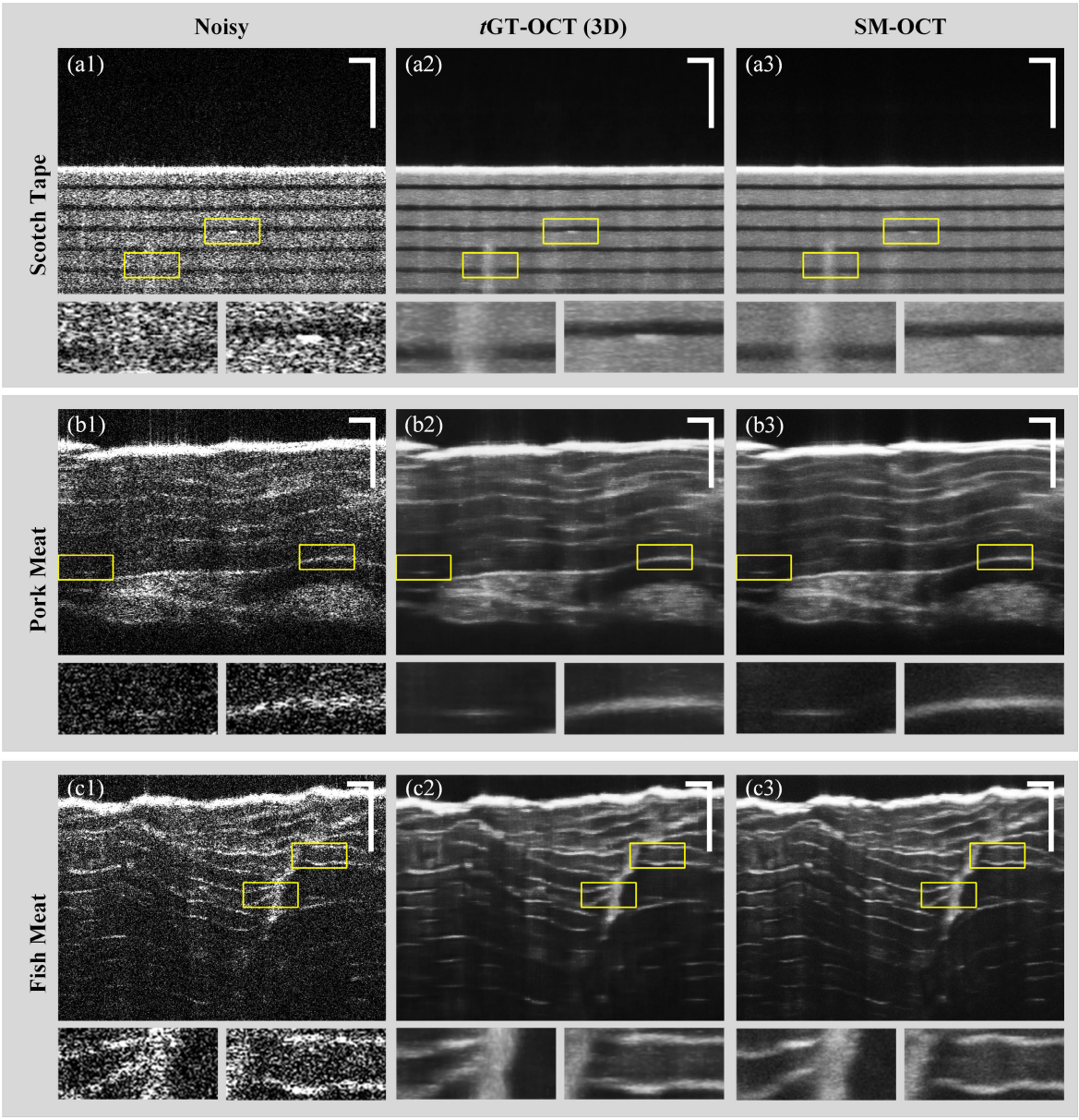

Fig. 5. Image outputs of the proposed *t*GT-OCT 3D network, and comparison with SM-OCT images of Scotch tape, pork meat and fish meat. (a1) - (a3) are images of Scotch tape; (b1) - (b3) are images of pork meat; (c1) - (c3) are images of fish meat. Scale bar: 200 μm.

of denoising. When comparing model outputs with the ground truth, the structure similarity (SSIM) index, peak signal-to-noise ratio (PSNR) and edge preservation index (EPI) [38] were also used. They are calculated respectively to compare structural similarity, signal energy and edge preservation with ground truth.

B. Experimental Results

1) Comparison With Ground Truth: Speckle-modulating OCT can be used to acquire the ideal ground truth image for OCT despeckling and is widely used in supervised deep learning methods [12], [39], [40]. To demonstrate the performance of our proposed *t*GT-OCT, we trained the *t*GT-OCT 3D deep learning network on the OCT-N1 dataset and tested it on the OCT-N2 dataset by comparing the outputs with SM-OCT images. Fig. 5 provides a visual representation, featuring the original noisy images, *t*GT-OCT images, and the corresponding SM-OCT images of Scotch tape, pork meat, and fish meat samples. As shown in Fig. 5(a1) - (a3), the magnified area on the right reveals a subtle structure that is comparable in size to speckle noise, concealed within the noisy image. In contrast, this structural information is clearly visible in the *t*GT-OCT image, which is consistent with the SM-OCT performance. In fish and pork images, the stripes are rendered distinctly in the *t*GT-OCT images, aligning closely with the SM-OCT images.

To perform a more comprehensive comparison, we further compared *t*GT-OCT with other unsupervised and supervised methods on fish meat image. Neighbor2Neighbor (NBR) [23], MAP-SNR [24] and Blind2Unblind (B2U) [41] require only single-noisy images so they were trained with 2D B-scans unpacked from the 3D volume data in train sets of OCT-N1.

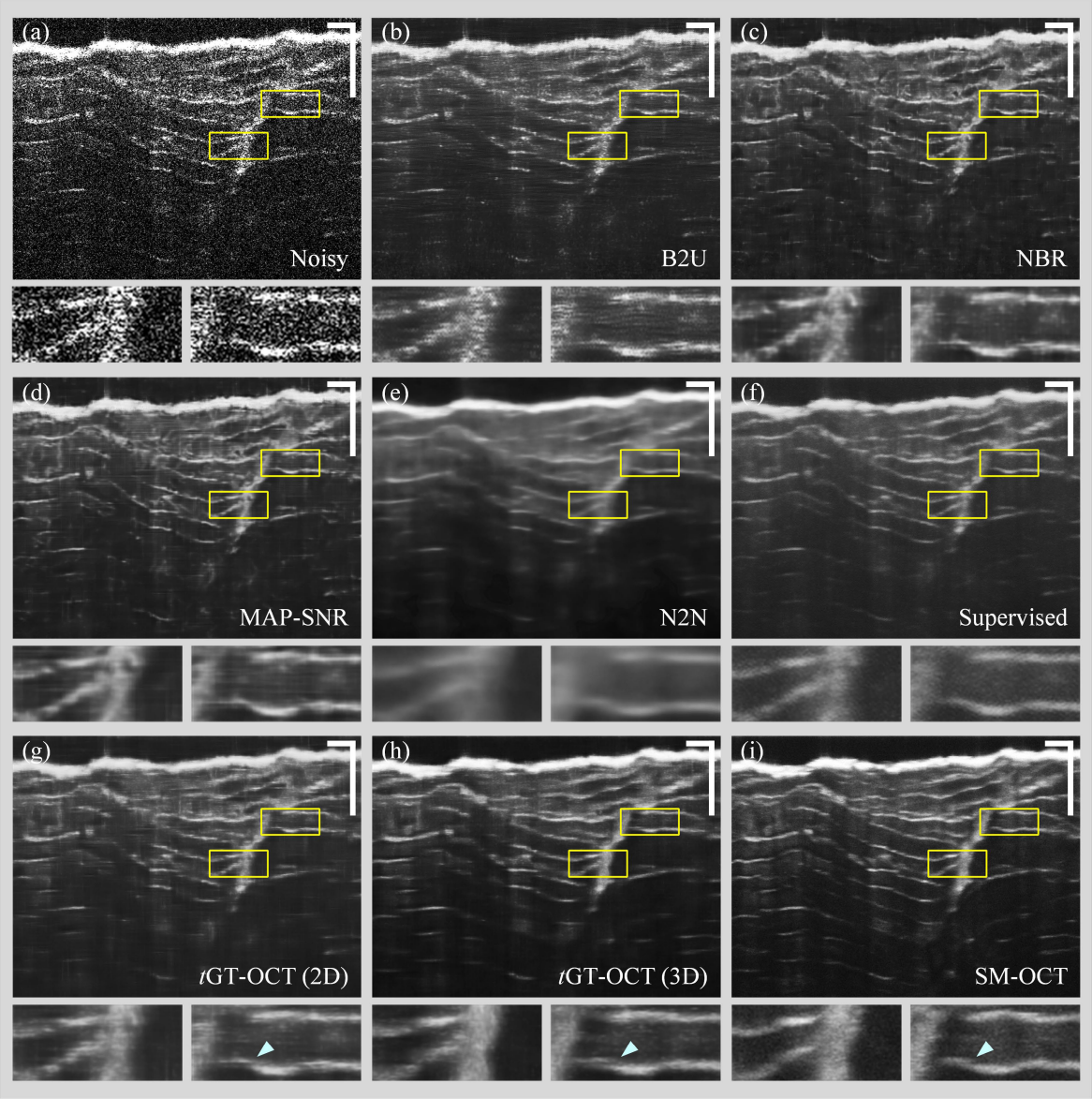

Fig. 6. Visual comparison results on the OCT-N2. (a) is noisy retinal images in volume data; (b) - (i) are results of B2U, NBR, MAP-SNR, N2N, Supervised method, the proposed *t*GT-OCT (2D network), the proposed *t*GT-OCT (3D network) and SM-OCT (ground truth). The cyan arrows indicate that *t*GT-OCT can resolve the detailed structures. Scale bar: 200 μm.

The Noise2Noise(N2N) [17] and supervised method [40] were trained on SM-OCT dataset [12] collected by the same setup as OCT-N1 and OCT-N2. The hyperparameters of all comparative networks were optimized empirically and based on the decreasing trend of the loss function. The representative image in OCT-N2 and its denoising results by different methods are shown in Fig. 6. B2U and NBR remove noise to some extent but has low image quality. N2N is effective in removing noise, but over-smoothing results in unclear detail information. Supervised method reduce noise while preserving details visually due to training with clean images. The proposed *t*GT-OCT (3D) has the best image quality compared to other methods. The detailed structure resolved by *t*GT-OCT is the closest to ground truth, as indicated by the cyan arrows. Table II shows that *t*GT-OCT (3D) has the highest PSNR and SSIM, indicating the closest approximation to ground truth. The N2N and supervised method have PSNR and SSIM scores similar to those of *t*GT-OCT (2D) because they accepted repeated scanned images, obtaining additional effective information. The quantitative results demonstrate that *t*GT-OCT (3D) outperforms other unsupervised and even supervised methods.

2) Comparison Without Ground Truth: Here, we compared the denoising results of different methods on test sets of OCT-R1 and OCT-N1, which don't contain ground truth. NBR, MAP-SNR, B2U, *t*GT-OCT 3D network and 2D network were trained on train sets of the OCT-R1 and OCT-N1 to compare their denoising performance on human retina and other non-retinal samples. Figs. 7(a1) - (a6) present the denoising results of different deep-learning-based methods on representative image of test set of OCT-R1. These results show that the B2U can suppress some noise while retaining some detailed information, although notable residual noise patterns are present.

TABLE II

QUANTITATIVE COMPARISON OF THE PROPOSED *t*GT-OCT AND DIFFERENT DENOISING METHODS ON OCT-N2. (MEAN ± STANDARD DEVIATION)

	Noisy	B2U	NBR	MAP-SNR	N2N	Supervised	*t*GT-OCT(2D)	*t*GT-OCT (3D)
PSNR	13.70±0.48	20.96±1.35	23.36±0.69	23.91±0.49	25.98±1.42	24.54±0.71	25.01±0.28	**28.6±0.43**
SSIM	0.12±0.02	0.44±0.05	0.57±0.03	0.60±0.03	0.69±0.06	0.63±0.01	0.66±0.01	**0.75±0.02**
SNR	10.26±3.98	26.37±2.75	27.37±2.05	29.30±2.24	36.59±6.34	30.79±3.22	36.66±6.04	**37.08±4.36**
CNR	-0.71±1.80	1.68±1.10	2.17±1.42	2.43±1.50	2.65±0.86	2.26±1.54	2.71±1.62	**2.85±1.59**
ENL	0.36±0.03	43.43±23.11	48.60±11.83	68.80±21.32	177.95±96.60	112.96±7.78	292.33±199.92	**458.12±240.60**

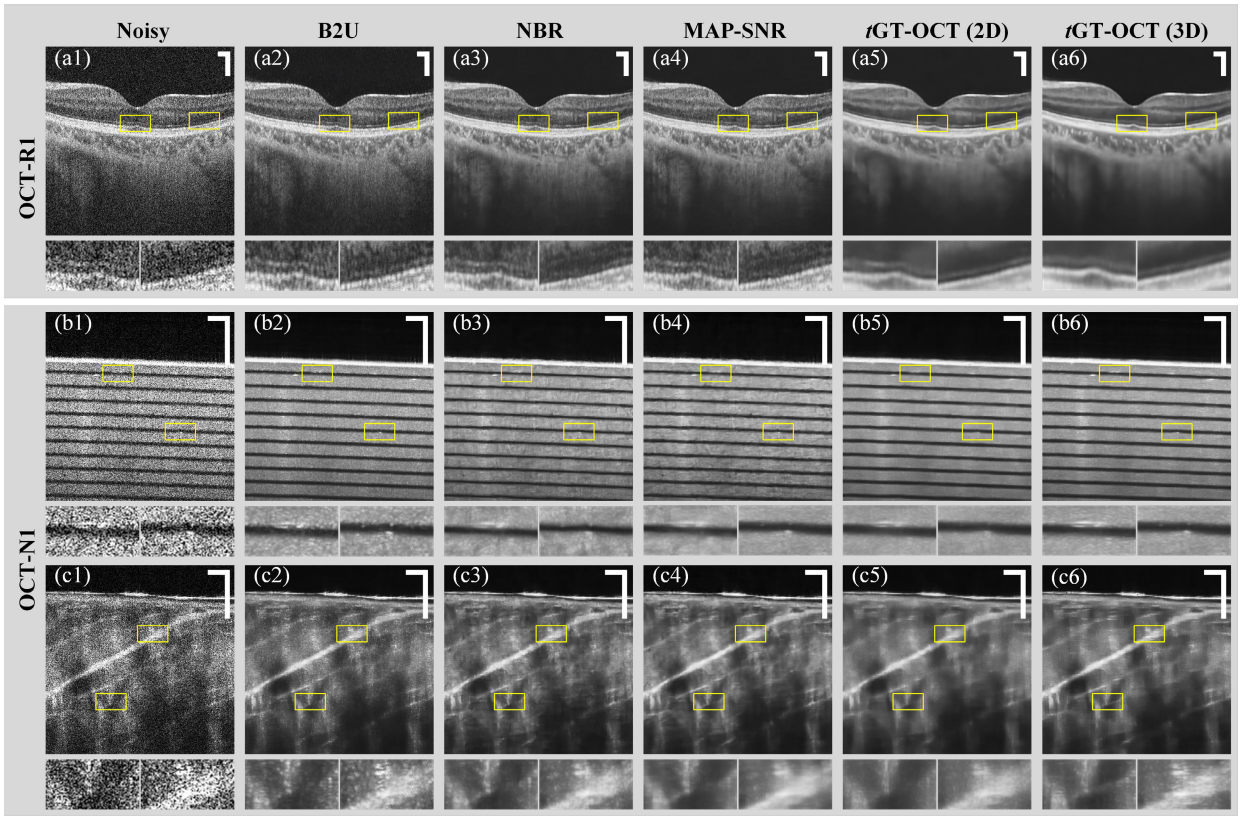

Fig. 7. Visual comparison results on test set of OCT-R1 and OCT-N1. (a1) - (c1) are noisy retinal image, Scotch tape image and pork meat image; (a2) - (c2) are results of B2U; (a3) - (c3) are results of NBR; (a4) - (c4) are results of MAP-SNR; (a5) - (c5) are results of 2D network of the proposed *t*GT-OCT; (a6) - (c6) are results of 3D network of the proposed *t*GT-OCT. Scale bar: 200 μm.

The NBR and MAP-SNR methods achieve a more substantial noise reduction but have compromised image quality due to the introduction of over-smoothing artifacts. In contrast, our proposed *t*GT-OCT (2D) excels in despeckling performance, adeptly suppressing speckle noise while preserving detailed information. *t*GT-OCT (3D) can resolve more microstructures within the choroidal vascular region.

We also compared these approaches using representative noisy Scotch tape image and pork image of test set of OCT-N1, as shown in Figs. 7(b1) - (b6) and (c1) - (c6). It can be seen that B2U's output images still contain remnants of speckle patterns, thereby resulting in poor image quality. The NBR and MAP-SNR methods suppress more noise but cause a smoothing effect that compromises the representation of certain structural details. In the magnified regions on the right side of Figs. 7(c2) - (c4), the pork strips are disconnected, whereas they are continuous in Fig. 7(c6). It is evident that *t*GT-OCT (2D) has better denoising performance than other unsupervised methods when fed into 2D images and *t*GT-OCT (3D) achieves effective microstructure resolution.

Table III shows no-reference evaluation metrics SNR, CNR and ENL values of the proposed *t*GT-OCT and other unsupervised OCT despeckling methodologies on test set of different datasets. *t*GT-OCT (3D) has the highest SNR, CNR and ENL scores in test set of OCT-R1. *t*GT-OCT (2D) has the highest CNR score in test set of OCT-N1.

3) Comparison on Public Dataset: To better verify the advantages and adaptability of the proposed *t*GT-OCT, we engaged in a comprehensive comparison of the public dataset OCTA-500. As shown in Fig. 8, two represent retina B-scans in the test set and their denoised results by different methods were visualized. We can see that our proposed *t*GT-OCT (3D) can reduce noise and make the membrane and layer boundary clearer. The detailed structures, such as blood vessels can be seen after denoising by *t*GT-OCT. The performance of the *t*GT-OCT (2D) is close to 3D network

TABLE III
QUANTITATIVE COMPARISON OF THE PROPOSED *t*GT-OCT AND DIFFERENT DENOISING METHODS ON TEST SET OF OCT-R1, OCT-N1 AND OCTA-500. (MEAN ± STANDARD DEVIATION)

	Metrics	Noisy	B2U	NBR	MAP-SNR	*t*GT-OCT(2D)	*t*GT-OCT(2D)
OCT-R1	SNR	12.18±1.19	30.63±1.19	42.39±1.19	45.09±1.16	48.21±1.39	**48.69±1.23**
	CNR	1.62±0.50	4.08±1.05	4.75±1.60	4.60±1.43	**5.81±1.90**	5.62±2.63
	ENL	4.60±1.46	39.99±9.95	127.98±30.86	116.50±30.70	212.71±63.27	**241.57±63.65**
OCT-N1	SNR	27.74±11.64	39.38±3.25	41.11±1.72	38.41±1.73	41.96±2.50	**43.87±1.56**
	CNR	3.13±1.37	4.81±0.91	5.23±0.85	5.35±0.94	5.34±1.27	**5.73±1.47**
	ENL	32.98±34.01	41.33±10.73	73.97±32.88	99.25±52.12	101.08±50.58	**138.35±95.42**
OCTA-500	SNR	24.95±0.34	41.35±0.35	48.36±0.35	47.58±0.34	47.92±0.40	**51.63±0.39**
	CNR	5.21±0.53	6.98±0.77	7.38±0.92	7.91±1.11	7.16±1.03	**7.49±0.90**
	ENL	140.33±29.80	347.93±175.89	814.75±255.85	710.25±254.01	833.51±183.19	**928.22±287.23**

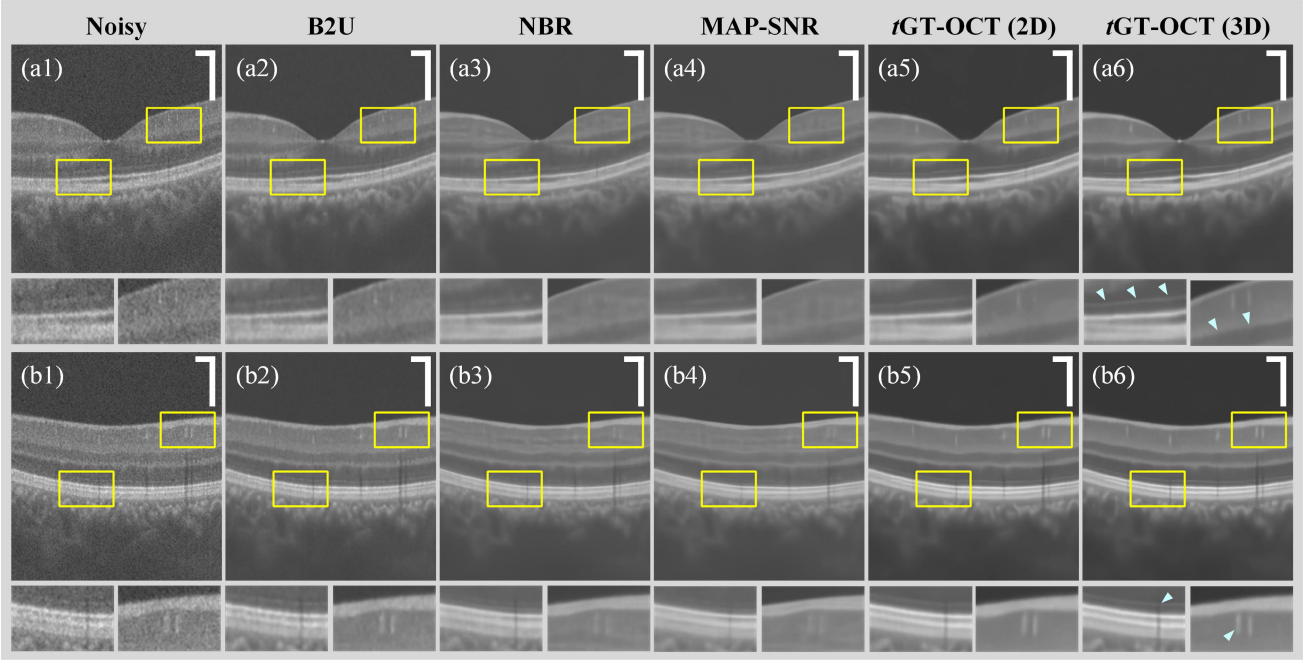

Fig. 8. Visual comparison results on test set of OCTA-500. (a1) - (b1) are noisy retinal images in two volume data; (a2) - (b2) are results of B2U; (b3) - (b3) are results of NBR; (a4) - (b4) are results of MAP-SNR; (a5) - (b5) are results of 2D network of the proposed *t*GT-OCT; (a6) - (b6) are results of 3D network of the proposed *t*GT-OCT. Scale bar: 200 μm.

and superior to other unsupervised denoising methods trained using 2D data. Quantitatively, Table III shows *t*GT-OCT (3D) has the highest SNR, CNR and ENL scores in test set of OCTA-500.

4) Clinical tGT-OCT Performance: Here, we conducted a denoising experiment on clinical OCT images in test sets of OCT-R1 and OCT-N1, including those of human retina, human skin and human placenta samples. The *t*GT-OCT teacher network (3D network) processed the 3D volumetric data, while the *t*GT-OCT student network (2D network) handled the single frames to demonstrate the effectiveness of *t*GT-OCT and knowledge distillation. As shown in Fig. 9, the denoising process was applied to human retinal data with varied lateral scanning ranges (6 mm and 20 mm). Figs. 9(a1) - (c1) show the noisy images, while Figs. 9(a2) - (c2) and Figs. 9(a3) - (c3) depict the *t*GT-OCT results obtained via the *t*GT-OCT 2D and 3D networks. These visual comparisons show the excellent speckle noise suppression and detail preservation capabilities of both networks, as well as their effectiveness across different scanning ranges and pathologies. The magnified regions shown also accentuate the enhanced distinctions between different layers of the retina, affirming the effectiveness of *t*GT-OCT denoising. The *t*GT-OCT 3D network outperforms its *t*GT-OCT 2D counterpart by further resolving micro-structures; this is possible because the 3D network can make full use of 3D spatial information in OCT volumetric data. In Fig. 9(b1), the boundaries of the retinal edema cystic cavities in the magnified orange region are very difficult to distinguish as they are obscured by speckle noise. In Fig. 9(b2), multiple cystic cavities can be distinguished after *t*GT-OCT denoising, while the image shown in Fig. 9(b3), the septum of the cysts is not only clear but also have good continuity. Incomplete posterior vitreous detachment is seen at the point indicated by the cyan arrow in Fig. 9(c3), and the septum produced by detachment can be clearly observed after being denoised by the *t*GT-OCT 3D network, providing a clinical diagnosis.

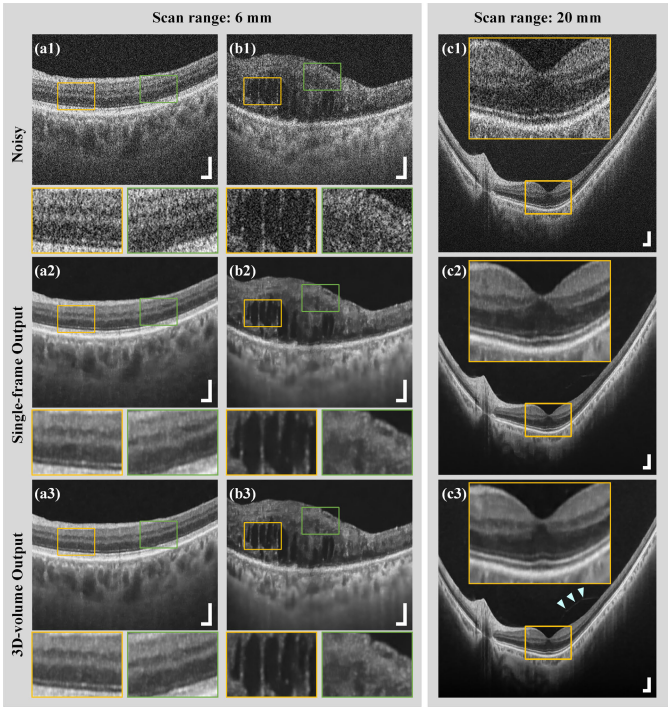

Fig. 9. Denoised images of the proposed *t*GT-OCT on human retina data with different scan ranges. (a1) - (c1) are representative noisy images in test set of OCT-R1; (a2) - (c2) are denosied results of *t*GT-OCT (2D); (a3) - (c3) are denosied results of *t*GT-OCT (3D). Scale bar: 200 μm.

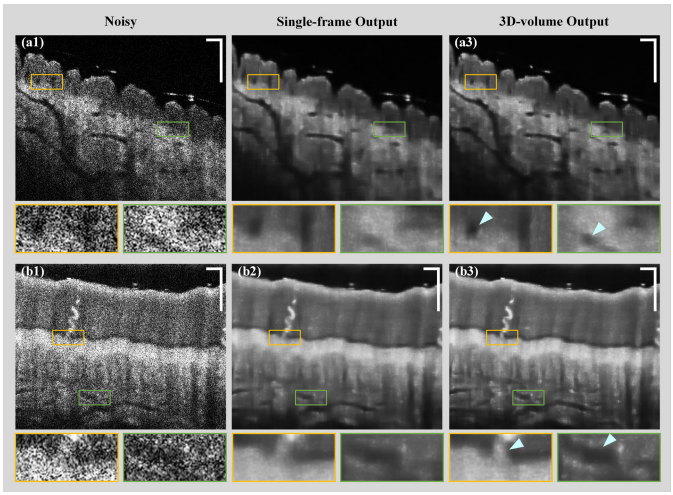

Fig. 10. Denoised images of the human skin output by the proposed *t*GT-OCT. (a1) - (b1) are noisy skin images of the lower arm and fingertip in test set of OCT-N1; (a2) - (b2) are denosied results of *t*GT-OCT (2D); (a3) - (b3) are denosied results of *t*GT-OCT (3D). Scale bar: 200 μm.

Fig. 10 presents cropped noisy images of human skin and the corresponding denoised results of the *t*GT-OCT 2D network and 3D network. Figs. 10(a1) - (a3) shows the skin of the inner side of the lower arm, and Figs. 10(b1) - (b3) show the skin of the fingertip. As indicated by arrows, this figure reveals that the *t*GT-OCT 3D network yields output images with greater image contrast and more discernible blood vessel details compared to those of the *t*GT-OCT 2D network. Detailed information on the sweat duct in the epidermis and epidermal junction can be clearly observed from the image in Fig. 10(b3).

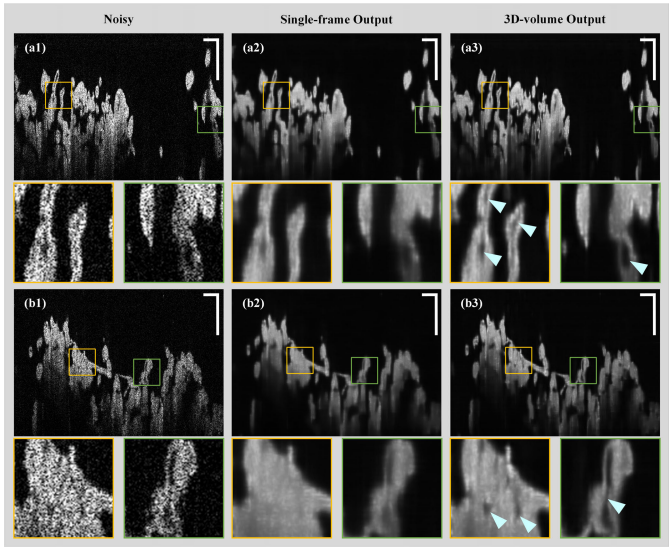

Fig. 11. Denoised images of the human placenta output by the proposed *t*GT-OCT. (a1) - (b1) are noisy placenta images in test set of OCT-R1; (a2) - (b2) are denosied results of *t*GT-OCT (2D); (a3) - (b3) are denosied results of *t*GT-OCT (3D). Scale bar: 200 μm.

Fig. 11 shows the noisy images and *t*GT-OCT output images of the human placenta [1]. Figs. 11(a1) and (b1) show cropped noisy B-scans from the volumetric data, Figs. 11(a2) and (b2) show denoised images output by the *t*GT-OCT 2D network, and Figs. 11(a3) and (c3) show denoised images output by the *t*GT-OCT 3D network. After despeckling with *t*GT-OCT, the shape of the placental villi and the capillaries at the end of the villi can be clearly seen in the magnified area, as indicated by arrows. Compared to the *t*GT-OCT 2D network, the *t*GT-OCT 3D network produces images with notably sharper vessel information, which we attribute to its superior contrast enhancement.

5) Generalization to Different Setups: To verify the generalization capability of our approach, we tested the well-trained network on retinal images acquired from different setups. Fig. 12(a1) is an image from the test set of OCT-R1 obtained with a BM-400K BMizar scanner (Setup A), and Fig. 12(b1) is an image of the OCT-R2 obtained with a Spectralis OCT scanner (Setup B). We ensured that the network had never seen images from the OCT-R2 dataset during training. Both the proposed *t*GT-OCT 2D and 3D networks show excellent generalization capability. As shown in Fig. 12, speckle noise is effectively reduced, preserving fine structures and important biomarkers. The retinal layered structure remains clearly discernible after denoising, with *t*GT-OCT 3D network demonstrating superior detail resolving capabilities. Consistent with the above results, the *t*GT-OCT 3D network has superior detail resolution performance. The posterior vitreous detached membrane and highly reflective foci, indicated by cyan arrows, are notably more observable in the results of the *t*GT-OCT 3D network compared to the *t*GT-OCT 2D network.

IV. Discussion

A. Ablation Study

We used the composite loss function, incorporating both MSE loss and VGG loss, as the objective function for the

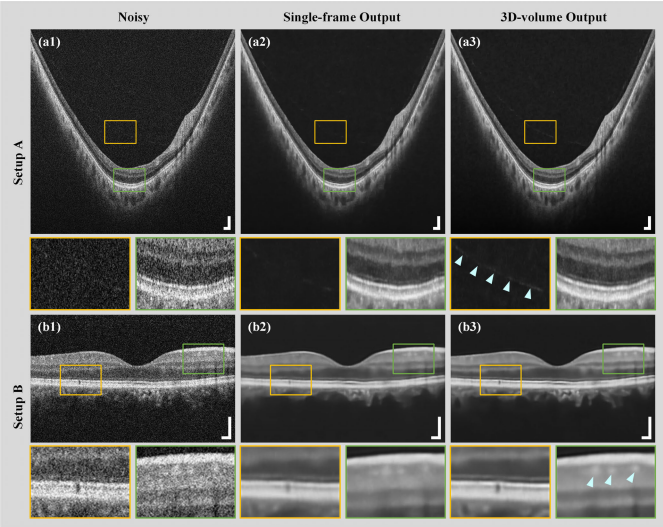

Fig. 12. Denoised images of the proposed *t*GT-OCT obtained from the test data with different retinal OCT setups. (a1) is noisy image acquired by BM-400K BMizar scanner; (a2) and (a3) are denosied results of 2D and 3D networks of *t*GT-OCT. (b1) is noisy image acquired by Spectralis OCT scanner; (b2) and (b3) are denosied results of 2D and 3D networks of *t*GT-OCT. Scale bar: 200 μm.

TABLE IV
ABLATION STUDY RESULTS OF LOSS FUNCTION ON THE VALIDATION SET OF OCT-R1-KD USING PSNR, SSIM AND EPI. (MEAN ± STANDARD DEVIATION)

	PSNR	SSIM	EPI
$\mathcal{L}_{MSE}$	35.136±1.115	0.957±0.009	0.926±0.029
$\mathcal{L}_{VGG}$	24.746±0.357	0.880±0.006	**1.763±0.087**
$0.5\mathcal{L}_{MSE}+1\mathcal{L}_{VGG}$	**35.910±1.200**	**0.962±0.008**	0.965±0.028
$1\mathcal{L}_{MSE}+0.5\mathcal{L}_{VGG}$	35.834±1.202	0.961±0.008	0.950±0.029
$1\mathcal{L}_{MSE}+1\mathcal{L}_{VGG}$	35.815±1.139	0.960±0.008	0.939±0.028

2D CNN. There are two weights in the loss function indicated in (6). The best weight coefficients have been selected through ablation study on the OCT-R1-KD dataset. Multiple models were trained with varying loss function weights and evaluated on the test set of OCT-R1-KD using PSNR, SSIM, and EPI metrics to assess the 2D CNN's ability to mimic the 3D CNN. As shown in Table IV, the $\mathcal{L}_{VGG}$ term is beneficial to maintain edge and high-frequency information. However, solely using $\mathcal{L}VGG$ does not yield satisfactory denoising performance. The composite loss function, incorporating $\mathcal{L}_{VGG}$ and $\mathcal{L}_{MSE}$, can improve image quality while preserving edge information. After several experiments, we determined the optimal weight coefficients to be $\alpha = 0.5$ and $\beta = 1$.

B. Model Efficiency

Table V shows the comparison result of model efficiency. We computed the number of model parameters (#Params), the storage size of the model checkpoint (Size), and mean compute times (mCT) for processing volume data. The comparison experiment was conducted on a platform equipped with GPU NVIDIA RTX A6000 and ten OCT volume data with sizes 512×512×128. Considering that the GPU consumes time to read the data, we fed batches of 2D images with a size of 128 into 2D CNN and fed a 3D volume consisting of 128 images into 3D CNN. We can see that #Params of 3D CNN are more than those of 2D CNN and 3D CNN opetrates slower. In addition, 3D CNN requires significantly more GPU memory during both training and inference. In summary, high-quality denoised images can be achieved using computationally-intensive 3D CNN, but 2D CNN has a higher utility for fast computation and practical engineering deployment.

TABLE V
EFFICIENCY COMPARISON BETWEEN THE 3D CNN AND 2D CNN OF OUR PROPOSED *t*GT-OCT

	*t*GT-OCT (3D)	*t*GT-OCT (2D)
#Params	983,201	330,051
Size	11.2 MB	1.53 MB
mCT (s)	12.02	5.19

C. Application in Retinal Pathologies

Here, we demonstrate the application of our denoising method to remove noise, enhancing the accuracy of retinal segmentation. The data for retinal layer segmentation comes from the OCTA-500 dataset with a FOV of 3 mm × 3 mm. There are 6 labels for retinal layers: internal limiting membrane (ILM), inner plexiform layer (IPL), outer plexiform layer (OPL), inner segment/outer segment (ISOS), retinal pigment epithelium (RPE), and Bruch's membrane (BM). The data for edema segmentation are retinal data that are diagnosed as diabetic macular edema (DME) and age-related macular degeneration (AMD). In the layer segmentation dataset, B-scans comprising 200 volume data are divided into training, validation, and test sets. The training set includes 28,880 B-scans with their labels, and the validation and test set each consists of 15,960 B-scans with corresponding labels. For the edema segmentation dataset, 1,560 B-scans are used for training, 521 for validation, and 433 for testing. For the Drusen segmentation dataset, 2790 B-scans, 931 B-scans, and 394 B-scans are used for training, validation, and testing, respectively. These data were not denoised and were denoised by *t*GT-OCT (3D) to train segmentation networks (U-Net) with different performances.

Fig. 13(a1) - (a2) show noisy and denoised images of a normal retina. Fig. 13(a3) - (a5) are the layer segmentation results and label. It can be seen that the segmentation result of the denoised image is visually closer to the ground truth. Fig. 13(b1) - (b2) show noisy and denoised images of a retina with DME. Fig. 13(b3) - (b5) show the edema segmentation results and label. From the denoised image, the membrane between the edema can be seen more clearly, so the network trained with denoised data can effectively segment the separated edema. Fig. 13(c1) - (c2) show noisy and denoised images of a retina with AMD. Fig. 13(c3) - (c5) show the drusen segmentation results and label. After denoising, the network can more clearly distinguish and segment the edge of Drusen. Table VI demonstrates the evaluation metrics of segmentation on test sets. Quantitatively, these segmentation networks trained with denoised data outperform networks trained with noisy data in all evaluation scores.

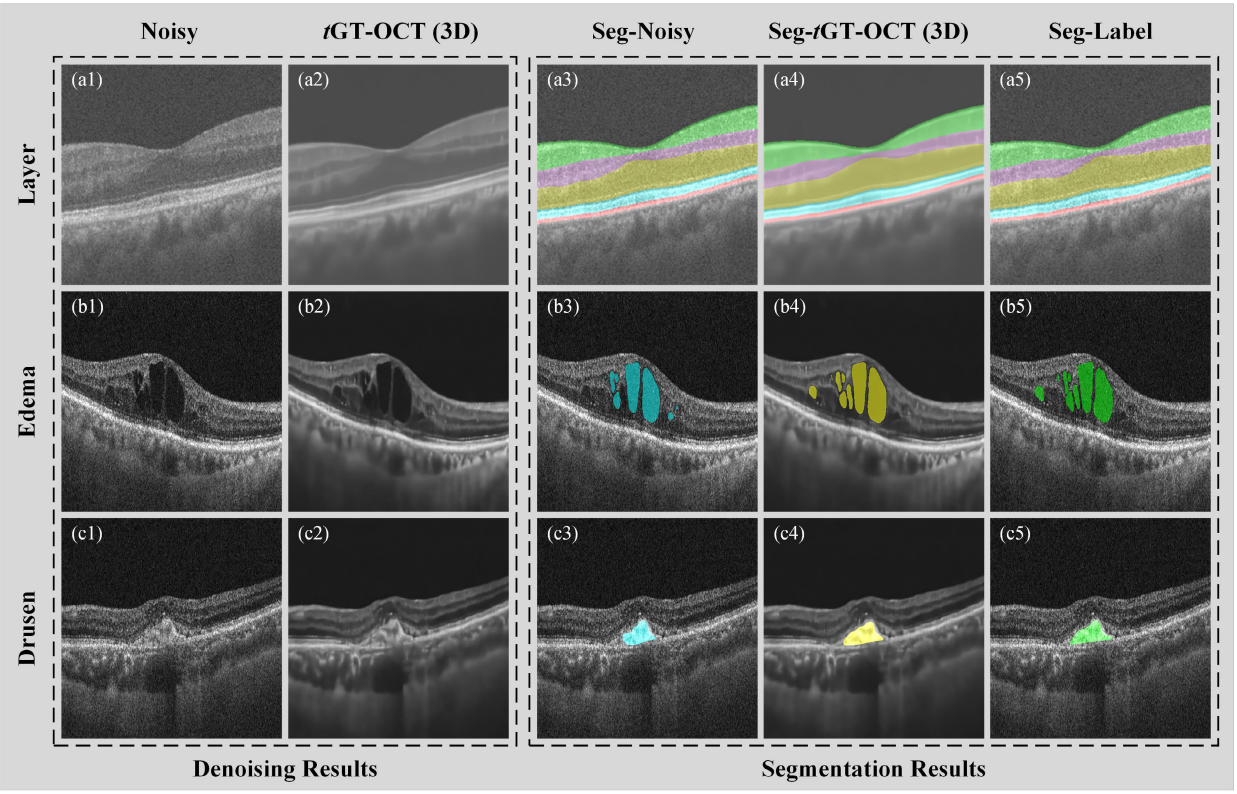

Fig. 13. Comparison of different segmentation performance before and after denoising. (a1) - (a2) are noisy and denoised images of normal retina. (a3) - (a4) are layer segmentation results of noisy image and denoised image by *t*GT-OCT (3D), and (a5) is the segmentation label. (b1) - (b2) are noisy and denoised images of the retina with diabetic macular edema. (a3) - (a4) are edema segmentation results of noisy image and denoised image, and (a5) is the segmentation label. (c1) - (c2) are noisy and denoised images of the retina with age-related macular degeneration. (c3) - (c4) are Drusen segmentation results of noisy image and denoised image, and (c5) is the segmentation label.

TABLE VI
QUANTITATIVE COMPARISON OF LAYER, EDEMA, AND DRUSEN SEGMENTATION RESULTS WITH NOISY AND DENOISED DATA. P. STANDS FOR PRECISION. THE FORMATS OF THE METRICS ARE MEAN (STANDARD DEVIATION)

	Noisy			Denoised		
Layer	P.	Dice	Hd95	P.	Dice	Hd95
ILM	98.07	98.4	187.76	98.32	98.44	169.99
-IPL	(1.38)	(0.96)	(235.82)	(1.38)	(1.02)	(108.82)
IPL	95.66	95.22	258.48	96.10	95.44	249.30
-OPL	(3.13)	(2.89)	(175.02)	(3.11)	(2.82)	(161.31)
OPL	97.91	97.74	230.77	97.40	97.8	225.05
-ISOS	(2.11)	(1.34)	(113.80)	(2.42)	(1.38)	(187.1)
ISOS	96.86	97.21	111.11	97.43	97.33	111.99
-RPE	(2.09)	(1.56)	(60.83)	(1.81)	(1.83)	(53.41)
RPE	94.26	93.97	110.26	95.36	94.64	113.04
-BM	(2.73)	(2.39)	(73.74)	(2.38)	(2.53)	(61.39)
Mean	96.55	96.51	179.68	**96.92**	**96.73**	**173.87**
	(1.43)	(1.65)	(60.67)	**(1.05)**	**(1.45)**	**(56.31)**
Edema	93.45	90.29	934.67	**94.48**	**93.64**	**302.37**
	(2.43)	(1.39)	(640.75)	**(2.01)**	**(0.89)**	**(37.78)**
Drusen	92.54	80.82	869.04	**92.70**	**82.20**	**558.82**
	(2.92)	(1.59)	(558.88)	**(2.44)**	**(1.12)**	**(206.15)**

V. CONCLUSION

Here we have introduced a novel speckle-free OCT imaging strategy that employs an unsupervised 3D deep learning network to distinguish and extract speckle patterns in OCT 3D volumetric data. This approach leveraged the power of the 3D convolutional network and OCT 3D imaging features, negating the necessity for clean images during training. Furthermore, our strategy maximized efficiency by incorporating a knowledge distillation mechanism to train the 2D convolutional network to achieve comparable denoising capabilities, effectively minimizing model complexity and computational demands. Experimental results with different sample images demonstrated that the proposed *t*GT-OCT can clarify and reveal structures that are otherwise obscured or undetectable while preserving spatial resolution; this was achieved by fully using the global information inherent within OCT 3D volumetric data. The presented comparative and generalization studies showed that the proposed *t*GT-OCT can effectively reduce speckle noise in OCT images of different samples and outperforms other deep learning methods, even achieving similar performance SM-OCT. A distilled 2D network, boasting commendable denoising performance and compact file size, stands poised for deployment in actual clinical applications. Our denoising method is applied to retinal layer, edema and drusen segmentation, resulting in improved accuracy of the segmentation network. Meanwhile, this work also leverages a new perspective for studying OCT speckle-free imaging with OCT 3D imaging features and unsupervised 3D deep-learning processing.

REFERENCES

[1] G. Ni et al., "Three-dimensional morphological revealing of human placental villi with common obstetric complications via optical coherence tomography," *Bioeng. Transl. Med.*, vol. 8, no. 1, Jul. 2022, Art. no. e10372.

[2] O. Liba et al., "Speckle-modulating optical coherence tomography in living mice and humans," *Nature Commun.*, vol. 8, no. 1, p. 15845, Jun. 2017.
[3] G. Ni, X. Du, J. Zhang, L. Liu, J. Liu, and Y. Liu, "Single a-line method for fast sample-structure-nondependent dispersion compensation of FD-OCT," *IEEE Photon. Technol. Lett.*, vol. 33, no. 24, pp. 1455–1458, Dec. 2021.
[4] G. Ni, Z. Wang, and C. Zhou, "Optical coherence tomography in biomedicine," in *Biomedical Optical Imaging: From Nanoscopy to Tomography*, J. Xia and R. Choe Eds. AIP Publishing LLC.
[5] M. Geng et al., "Triplet cross-fusion learning for unpaired image denoising in optical coherence tomography," *IEEE Trans. Med. Imag.*, vol. 41, no. 11, pp. 3357–3372, Nov. 2022.
[6] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[7] 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.
[8] Y. Huang et al., "Simultaneous denoising and super-resolution of optical coherence tomography images based on generative adversarial network," *Opt. Exp.*, vol. 27, no. 9, pp. 12289–12307, Apr. 2019.
[9] Z. L. Chen, Z. Y. Zeng, H. L. Shen, X. X. Zheng, P. S. Dai, and P. B. Ouyang, "DN-GAN: Denoising generative adversarial networks for speckle noise reduction in optical coherence tomography images," *Biomed. Signal Process. Control*, vol. 55, Jan. 2020, Art. no. 101632.
[10] N. A. Kande, R. Dakhane, A. Dukkipati, and P. K. Yalavarthy, "SiameseGAN: A generative model for denoising of spectral domain optical coherence tomography images," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 180–192, Jan. 2021.
[11] X. Yu, M. Li, C. Ge, P. P. Shum, J. Chen, and L. Liu, "A generative adversarial network with multi-scale convolution and dilated convolution res-network for OCT retinal image despeckling," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104231.
[12] G. Ni, Y. Chen, R. Wu, X. Wang, M. Zeng, and Y. Liu, "Sm-Net OCT: A deep-learning-based speckle-modulating optical coherence tomography," *Opt. Exp.*, vol. 29, no. 16, pp. 25511–25523, Jul. 2021.
[13] I. Manakov, M. Rohm, C. Kern, B. Schworm, K. Kortuem, and V. Tresp, "Noise as domain shift: Denoising medical images by unpaired image translation," in *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*. Cham, Switzerland: Springer, 2019, pp. 3–10.
[14] M. Wu, W. Chen, Q. Chen, and H. Park, "Noise reduction for SD-OCT using a structure-preserving domain transfer approach," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 9, pp. 3460–3472, Sep. 2021.
[15] 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.
[16] Z. Fu, X. Yu, C. Ge, M. Z. Aziz, and L. Liu, "ADGAN: An asymmetric despeckling generative adversarial network for unpaired OCT image speckle noise reduction," in *Proc. IEEE 6th Optoelectronics Global Conf. (OGC)*, Sep. 2021, pp. 212–216.
[17] D. Wu, K. Gong, K. Kim, X. Li, and Q. Li, "Consensus neural network for medical imaging denoising with only noisy training samples," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2019, pp. 741–749.
[18] Y. Huang, N. Zhang, and Q. Hao, "Real-time noise reduction based on ground truth free deep learning for optical coherence tomography," *Biomed. Opt. Exp.*, vol. 12, no. 4, pp. 2027–2040, Mar. 2021.
[19] B. Qiu et al., "Comparative study of deep neural networks with unsupervised noise2noise strategy for noise reduction of optical coherence tomography images," *J. Biophoton.*, vol. 14, no. 11, Aug. 2021, Art. no. e202100151.
[20] B. Qiu et al., "N2NSR-OCT: Simultaneous denoising and super-resolution in optical coherence tomography images using semisupervised deep learning," *J. Biophoton.*, vol. 14, no. 1, Oct. 2020, Art. no. e202000282.
[21] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2void-learning denoising from single noisy images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 2129–2137.
[22] J. Batson and L. Royer, "Noise2self: Blind denoising by self-supervision," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 524–533.
[23] 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.
[24] Y. Li, Y. Fan, and H. Liao, "Self-supervised speckle noise reduction of optical coherence tomography without clean data," *Biomed. Opt. Exp.*, vol. 13, no. 12, pp. 6357–6372, Nov. 2022.
[25] Q. Zhou, M. Wen, M. Ding, and X. Zhang, "Unsupervised despeckling of optical coherence tomography images by combining cross-scale CNN with an intra-patch and inter-patch based transformer," *Opt. Exp.*, vol. 30, no. 11, pp. 18800–18820, May 2022.
[26] Q. Zhou, M. Wen, B. Yu, C. Lou, M. Ding, and X. Zhang, "Self-supervised transformer based non-local means despeckling of optical coherence tomography images," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104348.
[27] Z. Zhang, X. Liang, W. Zhao, and L. Xing, "Noise2context: Context-assisted learning 3D thin-layer for low-dose CT," *Med. Phys.*, vol. 48, no. 10, pp. 5794–5803, 2021.
[28] M. Papkov et al., "Noise2Stack: Improving image restoration by learning from volumetric data," in *Proc. Int. Workshop Mach. Learn. Med. Image Reconstruct. (MLMIR)*, 2021, pp. 99–108.
[29] C. Niu et al., "Noise suppression with similarity-based self-supervised deep learning," *IEEE Trans. Med. Imag.*, vol. 42, no. 6, pp. 1590–1602, Jun. 2023.
[30] J. M. Schmitt, S. H. Xiang, and K. M. Yung, "Speckle in optical coherence tomography," *J. Biomed. Opt.*, vol. 4, no. 1, pp. 95–105, 1999.
[31] Y. Wang, Y. Wang, A. Akansu, K. D. Belfield, B. Hubbi, and X. Liu, "Robust motion tracking based on adaptive speckle decorrelation analysis of OCT signal," *Biomed. Opt. Exp.*, vol. 6, no. 11, pp. 4302–4316, Oct. 2015.
[32] G. Ni, R. Wu, J. Zhong, and Y. Liu, "Depth-resolved transverse-plane motion tracking with configurable measurement features via optical coherence tomography," *Opt. Exp.*, vol. 30, no. 8, pp. 12215–12227, Mar. 2022.
[33] X. Liu, Y. Huang, and J. U. Kang, "Distortion-free freehand-scanning OCT implemented with real-time scanning speed variance correction," *Opt. Exp.*, vol. 20, no. 15, pp. 16567–16583, Jul. 2012.
[34] A. Ahmad, S. G. Adie, E. J. Chaney, U. Sharma, and S. A. Boppart, "Cross-correlation-based image acquisition technique for manually-scanned optical coherence tomography," *Opt. Exp.*, vol. 17, no. 10, pp. 8125–8136, Apr. 2009.
[35] J. W. Goodman, *Speckle Phenomena in Optics: Theory and Applications*. Greenwood Village, CO, USA: Roberts & Company, 2007.
[36] M. Li et al., "Image projection network: 3D to 2D image segmentation in OCTA images," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3343–3354, Nov. 2020.
[37] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.
[38] 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.
[39] G. M. Ni et al., "Hybrid-structure network and network comparative study for deep-learning-based speckle-modulating optical coherence tomography," *Opt. Exp.*, vol. 30, no. 11, pp. 18919–18938, May 2022.
[40] Z. Dong, G. Liu, G. Ni, J. Jerwick, L. Duan, and C. Zhou, "Optical coherence tomography image denoising using a generative adversarial network with speckle modulation," *J. Biophoton.*, vol. 13, no. 4, Apr. 2020, Art. no. e201960135.
[41] X. Yu et al., "Self-supervised Blind2Unblind deep learning scheme for OCT speckle reductions," *Biomed. Opt. Exp.*, vol. 14, no. 6, pp. 2773–2795, May 2023.