High-Resolution Hierarchical Adversarial Learning for OCT Speckle Noise Reduction

Yi Zhou[1], Jiang Li[2], Meng Wang[1], Weifang Zhu[1], Yuanyuan Peng[1], Zhongyue Chen[1], Lianyu Wang[1], Tingting Wang[1], Chenpu Yao[1], Ting Wang[1], and Xinjian Chen[1,3] (✉)

[1] School of Electronics and Information Engineering, Soochow University,
Suzhou, China
xjchen@suda.edu.cn

[2] Department of Electrical and Computer Engineering, Old Dominion University,
Norfolk, VA, USA

[3] State Key Laboratory of Radiation Medicine and Protection, Soochow University,
Suzhou, China

Abstract. Raw optical coherence tomography (OCT) images typically are of low quality because speckle noise blurs retinal structures, severely compromising visual quality and degrading performances of subsequent image analysis tasks. In this paper, we propose a novel end-to-end cross-domain denoising framework for speckle noise suppression. We utilize high quality ground truth datasets produced by several commercial OCT scanners for training, and apply the trained model to datasets collected by our in-house OCT scanner for denoising. Our model uses the high-resolution network (HRNet) as backbone, which maintains high-resolution representations during the entire learning process to restore high fidelity images. In addition, we develop a hierarchical adversarial learning strategy for domain adaption to align distribution shift among datasets collected by different scanners. Experimental results show that the proposed model outperformed all the competing state-of-the-art methods. As compared to the best of our previous method, the proposed model improved the signal to noise ratio (SNR) metric by a huge margin of $18.13dB$ and only required $25ms$ for denoising one image in testing phase, achieving the real-time processing capability for the in-house OCT scanner.

Keywords: OCT speckle noise reduction · HRNet · GAN.

1 Introduction

OCT is a recent imaging modality for biological tissues [14]. OCT images usually suffer from speckle noises, which degrade quality of OCT images and make automated image analysis challenging. Traditional speckle noise suppression algorithms can be categorized into three groups: 1) Filter-based techniques [1, 4], 2) Sparse transform-based methods [8, 6] and 3) Statistics and low-rank decomposition-based methods [2, 9, 3]. Though these traditional methods are effective for image denoising, there are remaining challenges such as insufficient image feature representation, and some of the methods are time-consuming.

In the past few years, there has been a boom of the application of deep learning to noise suppression. For example, Zhang et al. [16] used the residual learning strategy and batch normalization technique to improve the feedforward denoising convolutional neural network (DnCNN) for target blind Gaussian denoising. Dong et al. [5] proposed a denoising prior driven deep neural network (DPDNN) for image restoration, and both methods were designed based on the additive noise model. We proposed a model, Edge-cGAN [10] based on the conditional generative adversarial network (cGAN) [7] to remove speckle noise, and developed training dataset T by using commercial scanners for data collection. Later, we developed an in-house OCT scanner and improved Edge-cGAN to the second version (Mini-cGAN [17]) to suppress speckle noise for dataset B collected by the in-house scanner. While it achieved good performances on dataset B, Mini-cGAN is not an end-to-end learning model, requiring multiple steps of training. In addition, the testing time complexity is high, making it not suitable for real-time processing.

Different OCT scanners have different characteristics and datasets collected by different scanners may contain distribution shifts. In this paper, our goal is to compensate for the distribution shifts and leverage the high quality ground truth dataset T to achieve effective speckle noise suppression for dataset B collected by our in-house scanner. We propose a novel end-to-end learning framework to achieve our goal and the diagram is shown in Fig. 1. The proposed model 1) utilizes the HRNet backbone to maintain high-resolution representation for image restoration and 2) uses a dynamic hierarchical domain adaptation network to leverage the dataset T for training and apply the trained model to dataset B for noise removal. Our model not only significantly improves image quality as compared with Edge-cGAN and Mini-cGAN but also dramatically reduces testing time, satisfying the real-time requirement of our in-house scanner.

2 Method

2.1 Proposed Model

Overall Architecture: The proposed model (Fig. 1) consists of a generator G, a hierarchical discriminator D_H, and an output alignment discriminator D (omitted in the figure but its learning loss function shown as $\mathcal{L}_D$). Our objective is to leverage high quality dataset T to learn an end-to-end model for speckle noise removal and apply the model to dataset B collected by our in-house scanner. The distribution shift between the two datasets T and B is compensated by the loss function $\mathcal{L}_H$ through hierarchical adversarial learning. In addition, the denoised images $G(T)$ and $G(B)$ generated from the two data domains are aligned by the generative adversarial loss function $\mathcal{L}_D$ during training.

HRNet Backbone for Image Restoration: HRNet [12] is an improved version of the U-Net structure that has been widely applied to image reconstruc-

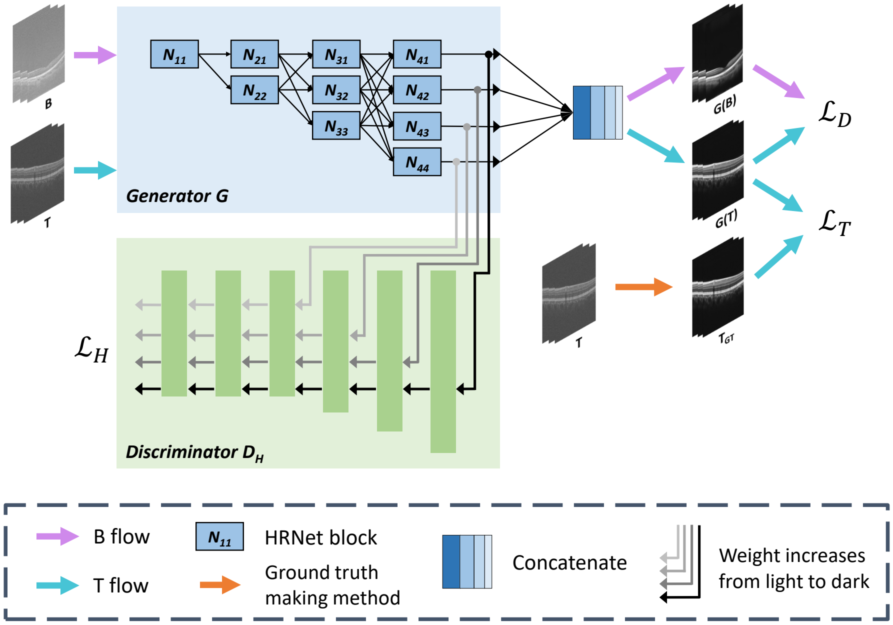


Fig. 1. Diagram of the proposed framework. During training, datasets B and T are alternatively input to the generator G (HRNet). Features of different resolutions are extracted by G and hierarchically combined in the discriminator D_H to perform adversarial learning such that features extracted from dataset B are indistinguishable from those from T. Features from T are then concatenated to generate denoised images as $G(T)$ and are compared to ground truth T_{GT} to minimize $\mathcal{L}_T$. In addition, $G(B)$, images generated from dataset B, are aligned to $G(T)$ by a discriminator D (omitted in this Figure) through the adversarial learning process governed by the loss $\mathcal{L}_D$. During testing, images from B are input to G to generate denoised images as $G(B)$.

tion and segmentation [11]. U-Net uses an encoder to compress input to a low-dimensional latent vector through convolution and pooling and utilizes a decoder to reconstruct input by expanding the compressed vector. There are also skip connections to copy information from encoder directly to decoder located at the same resolution level. Convolutional layers in deep learning models including U-Net extract and magnify useful information from input for classification or regression tasks. However, this convolutional processing only happens in the low resolution levels after pooling in U-Net and other levels copy information directly from encoder to decoder by the skip connections. In contrast, HRNet uses convolutional layers to learn useful information in all resolution levels as shown in Fig. 1. We select $HRNetV2$ layer [12] as output, which concatenates different resolution channels by bilinear interrelation to reconstruct denoised image.

We utilize the combination of mean square error (MSE) and L_1 losses for training: $\mathcal{L}_T = \mathcal{L}_{MSE} + \alpha\mathcal{L}_{L_1}$, as it has been shown in [7] that the L_1 loss led to sparse solutions, which are desired for the OCT speckle noise removal. The MSE loss and L_1 loss can be formulated as follows, respectively,

$$\mathcal{L}_{MSE} = \mathbb{E}_{t,t_{gt}\sim T}[\|t_{gt} - G(t)\|_2], \tag{1}$$

$$\mathcal{L}_{L_1} = \mathbb{E}_{t,t_{gt}\sim T}[\|t_{gt} - G(t)\|_1], \tag{2}$$

where t and t_{gt} represent a raw-clean image pair in training dataset T.

Hierarchical Adversarial Learning: Inspired by [15], we propose a hierarchical adversarial learning structure to compensate domain shift between datasets T and B. We combine outputs from different resolution levels in HRNet (G) using a hierarchial structure as D_H in Fig. 1. Low-level resolution features are domain informative while high-level resolution features are rich in semantic information. Direct stacking these features from different resolution levels for adversarial learning can impair adaptation performance because of the information conflicts among different resolution levels [15]. The hierarchical structure in D_H follows the resolution levels in the encoder part of G so that the conflicts can be mitigated. Each level has a separate objective function and G and D_H play a min-max game to minimize the overall loss, $\mathcal{L}_H = \sum_{k=1}^{K} \gamma_k \mathcal{L}_{h,k}$, where $\mathcal{L}_{h,k}$ is the loss function for level l_k, $k = 1, 2, 3..., K$ and

$$\mathcal{L}_{h,k} = \mathbb{E}_{l_k^t \in G(t)}[\log(D_H(l_k^t))] + \mathbb{E}_{l_k^b \in G(b)}[\log(1 - D_H(l_k^b))], \tag{3}$$

γ_k are mixing coefficients and it increases as k decreases, making the attention focus more on low-level domain information, and l_k^t and l_k^b represent features extracted at the kth layer from datasets T and B, respectively.

Output Alignment: At output of G, we utilize a discriminator D to align the reconstructed images $G(B)$ and $G(T)$ in image space by adversarial learning,

$$\mathcal{L}_D = \mathbb{E}_{t\sim T}[\log(D(G(t)))] + \mathbb{E}_{b\sim B}[\log(1 - D(G(b)))], \tag{4}$$

where b is one image in B. Finally, the total objective function of the framework is,

$$\min_{G} \max_{D_H, D} \mathcal{L}_T + \lambda_0 \mathcal{L}_H + \lambda_1 \mathcal{L}_D, \tag{5}$$

where λ_0, λ_1 are trade-off parameters between the two loss functions. With the above learning processes and loss functions, the generator G can be trained to restore high quality OCT images for our in-house scanner. The training and testing procedure descriptions are provided in the caption of Fig. 1.

2.2 Dataset

High quality dataset T was created in our previous study [10], which was approved by IRB of the University and informed consents were obtained from all subjects. The dataset contains 512 raw OCT images collected by four commercial scanners (Topcon DRI-1 Atlantis, Topcon 3D OCT 2000, Topcon 3D OCT 1000 and Zeiss Cirrus 4000, with their image sizes of 512×992, 512×885, 512×480 and 512×1024 respectively), and the training set is composed of 256 images from Topcon DRI-1 Atlantis and 256 images from Topcon 3D OCT 2000. For each raw image, a clean image was produced by registering and averaging raw images acquired repeatedly at the same location from the same subject. We used flipping along the transverse axis, scaling, rotation, and non-rigid transformations to augment T and increased the data size by 4× for effective training. Dataset B was collected by our recently developed non-commercial in-house OCT scanner and there were totally 1024 raw OCT images, of which 200 images had disease. We also set aside three disease raw images and six normal raw images from the in-house scanner for testing.

2.3 Evaluation Metrics

We utilize four performance metrics to evaluate the proposed model including signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), speckle suppression index (SSI) and edge preservation index (EPI). As Fig.2a) and Fig.3a) show, we manually selected regions of interest (ROIs) in background (green rectangles) and signal (red rectangles) areas for SNR and CNR calculation, and delimitated the blue boundaries for EPI computation. These four metrics are defined as,

$$\begin{aligned} &SNR = 10\lg(\frac{\sigma_s^2}{\sigma_b^2}), CNR = 10\lg(\frac{|\mu_s - \mu_b|}{\sqrt{\sigma_s^2 + \sigma_b^2}}), \\ &SSI = \frac{\sigma_r}{\mu_r} \times \frac{\mu_d}{\sigma_d}, EPI = \frac{\sum_i \sum_j |I_d(i+1,j) - I_d(i,j)|}{\sum_i \sum_j |I_r(i+1,j) - I_r(i,j)|}. \end{aligned} \tag{6}$$

where μ_s, μ_b and σ_s, σ_b denote means and standard deviations of the defined signal and background regions in denoised image, respectively, for SNR and CNR computations. μ_r, σ_r and μ_d, σ_d denote means and standard deviations of raw and denoised images, respectively, for SSI computation. I_r and I_d represent raw and denoised images, and i and j represent longitudinal and lateral coordinates in image. SNR reflects noise level, CNR is the contrast between signal and background, SSI measures ratio between noise and denoised images, and EPI reflects the extent of details of edges preserved in denoised images. A small SSI value and large SNR, CNR and EPI values represent high quality images.

3 Experiments and Results

3.1 Implementation Details

Discriminators D_H and D followed the configurations in [7] consisting of six convolutional layers where the first three used instance normalization. For all the experiments, α was set to 100 in $\mathcal{L}_T$, hyper-parameters λ_0 and λ_1 were

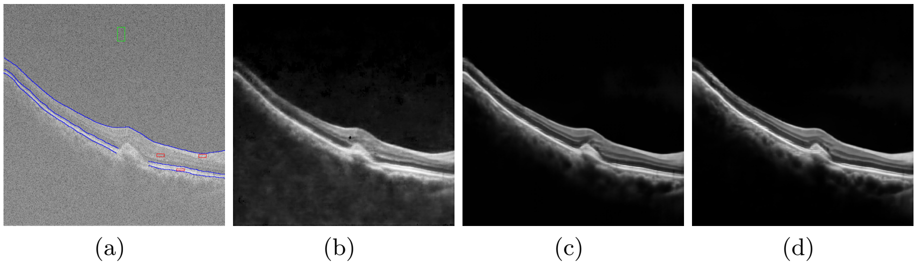


Fig. 2. Ablation study results on a disease OCT image. Red and green rectangles represent signal and background regions manually selected for SNR and CNR calculation, respectively. Blue curves are boundaries manually delimitated for EPI computation. (a) Raw OCT image (b) U-Net (c) HRNet (d) HRNet+Hier (Proposed).

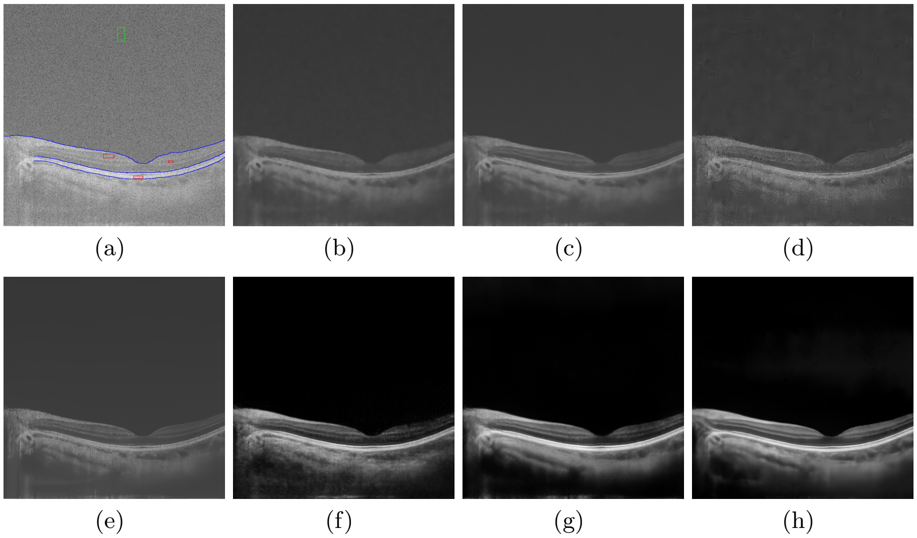
set to 10 in Eq. 5, and $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ in $\mathcal{L}_H$ followed [15] to set to 4, 3, 2 and 1, respectively. We utilized the Adam solver with an initial learning rate of 2.0×10^{-4} and a momentum of 0.5 for optimization. The batch size was set to 2, and the number of training epochs was set to 400. Images were resized to 512×512 during training. The proposed method was implemented using Pytorch and was trained using one NVIDIA RTX 3080 GPU with 10G memory.

3.2 Ablation Study

We conducted ablation study to investigate contribution of each of the three components in the proposed model, including 1) U-Net as backbone for the generator G with output alignment but no hierarchical adversarial learning (U-Net), 2) HRNet as backbone for G with output alignment but no hierarchical adversarial learning (HRNet) and 3) HRNet as backbone for G with both output alignment and hierarchical adversarial learning as the final proposed model (HRNet+Hier). Experimental results in Table 1 show that using HRNet to replace U-Net greatly improved performance of the model, increasing SNR from $22.69dB$ to $35.41dB$. SSI and EPI were also improved with any exception of CNR that slightly degraded. HRNet maintains low- and high-resolution information throughout the entire process, making the denoised images much better quality. When the hierarchical adversarial loss was combined, SNR was improved further to $40.41dB$, EPI also slightly increased and the other two degraded slightly.

Table 1. Ablation study on nine images from the in-house OCT scanner.

Method	SNR (dB)	CNR (dB)	SSI	EPI
Raw image	0.03±0.54	3.80±0.85	1.000±0.00	1.00±0.00
U-net	22.69±7.45	**12.44±1.13**	0.139±0.01	0.83±0.09
HRNet	35.41±12.00	11.35±1.34	**0.087±0.01**	0.94±0.07
HRNet+Hier	**40.41±7.69**	11.15±1.39	0.091±0.01	**0.96±0.07**



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 3. Comparison study results. The red rectangles, green rectangles and blue curves are manually defined as in Fig. 2. (a) Raw image (b) NLM [1] (c) STROLLR [13] (d) DnCNN [16] (e) DPDNN [5] (f) Edge-cGAN [10] (g) Mini-cGAN [17] (h) Proposed (HRNet+Hier).

3.3 Comparison Study

Table 2. Results by competing methods on nine images from the in-house scanner.

Method	SNR (*dB*)	CNR (*dB*)	SSI	EPI	Times(s)
Raw image	0.03±0.54	3.80±0.85	1.000±0.00	1.00±0.00	*None*
NLM [1]	19.50±4.21	9.23±1.98	0.697±0.03	0.55±0.06	0.089±0.006
STROLLR [13]	18.01±3.89	11.03±2.01	0.707±0.03	0.37±0.02	182.203±7.648
DnCNN [16]	14.99±2.27	7.00±1.10	0.670±0.03	0.62±0.03	**0.022±0.003**
DPDNN [5]	34.77±8.40	8.40±1.74	0.684±0.03	0.52±0.03	0.036±0.001
Edge-cGAN [10]	24.35±5.50	11.35±0.97	0.105±0.01	0.87±0.08	0.929±0.014
Mini-cGAN [17]	22.28±4.65	**12.03±1.61**	**0.087±0.01**	0.93±0.11	1.825±0.022
Proposed	**40.41±7.69**	11.15±1.39	0.091±0.01	**0.96±0.07**	0.025±0.008

We compared the proposed method with state-of-the-art methods, including non-local means (NLM) [1], sparsifying transform learning and low-rank method (STROLLR) [13], deep CNN with residual learning (DnCNN) [16], denoising prior driven deep neural network for image restoration (DPDNN) [5], our previous method, Edge-cGAN[10] and its improved version, Mini-cGAN [17]. In these experiments, parameters for traditional methods were set to values so that the models can achieve best results for the application, and deep learning models followed their original configurations. The proposed method and Mini-cGAN compensated distribution shift existed between datasets T and B through adversarial learning. To have a fair comparison, we performed a separate histogram matching process for the testing images for all other competing methods before testing. In addition, we compared efficiency by recording testing time required by each method.

Visual inspection of Fig. 3 reveals that the proposed method achieved the best result (Fig. 3h). Our previous methods Mini-cGAN and Edge-cGAN ranked the second and third (Fig. 3g and Fig. 3f), respectively. Denoised images by all other methods have low visual qualities. Table 2 shows qualitative performance metrics indicating that the proposed method achieved the best SNR and EPI. Mini-cGAN obtained the best CNR and SSI, and DnCNN achieved the best computational efficiency, requiring only $22ms$ to process one image. It is worth noting that the proposed model ranked the second and only needed $25ms$ to denoise one image.

4 Discussion

The proposed method achieved the best visual quality as shown in Fig. 3. NLM (Fig. 3b) and STROLLR (Fig. 3c) suffered from excessive smoothing, leading to blurred regions at the boundaries between adjacent layers. The background regions were not very clean either. DnCNN (Fig. 3d) performed well in retina areas but left artifacts in background. DPDNN (Fig. 3e) obtained a very clean background, however, the interlayer details were not well maintained. Edge-cGAN (Fig. 3f) and Mini-cGAN (Fig. 3g) improved visual quality of the denoised image significantly. However, the signal was still weak in the top right retina area. The proposed model achieved the best image contrast, preserved the most details in the layers under retina, and resulted in a much sharper enhanced image (Fig. 3h).

The metrics of SNR and EPI represent signal to noise ratio and edge preservation performances in denoised images, respectively. The proposed model achieved the best SNR and EPI (Table 2), indicating that it restored the strongest signal by suppressing speckle noise and preserved the desired sharp detail information. In terms of testing time, DnCNN took $3ms$ less than the proposed model. However, image quality by DnCNN was much worse. As compared to Mini-cGAN, the new model improved SNR by a huge margin of $18.13dB$ to $40.41dB$ and the testing time was accelerated by a factor of 73. The computational efficiency satisfied real-time requirement of our in-house scanner. DPDNN achieved the second best SNR of $34.77dB$, its other metrics including CNR, SSI and EPI were much worse than those by the proposed model, which can be confirmed in Fig. 3e.

CNR and SSI represent the contrast and speckle suppression performances in result images, respectively. Mini-cGAN uses U-Net as backbone and generates denoised images by averaging multiple overlapped patches outputted by the trained model during testing. In contrast, the proposed model utilizes the high-resolution HRNet to generate denoised images directly without averaging. The averaging processing in Mini-cGAN reduced variance and led to slightly better CNR and SSI with a cost of much longer testing time. We tested to conduct the same averaging process in the proposed model to generate denoised images and it did improve the CNR metric slightly but degraded SNR and required much longer testing time. We concluded that the HRNet was able to restore the strongest signal because of its unique structure and the repetition step in testing was not necessary.

5 Conclusion

In this paper, we proposed a novel end-to-end cross-domain denoising framework that significantly improved speckle noise suppression performance in OCT images. The proposed model can be trained and tested with OCT images collected by different scanners, achieving automatic domain adaptation. We utilized the HRNet backbone to carry high-resolution information and restored fidelity images. In addition, we developed a hierarchical adversarial learning module to achieve the domain adaptation. The novel model improved SNR by huge margins as compared to our previous models and all competing state of the arts, achieved a testing time of $0.025s$, and satisfied real-time process requirement.

References

1. Aum, J., Kim, J.h., Jeong, J.: Effective speckle noise suppression in optical coherence tomography images using nonlocal means denoising filter with double gaussian anisotropic kernels. Applied Optics **54**(13), D43–D50 (2015)
2. Cameron, A., Lui, D., Boroomand, A., Glaister, J., Wong, A., Bizheva, K.: Stochastic speckle noise compensation in optical coherence tomography using non-stationary spline-based speckle noise modelling. Biomedical optics express **4**(9), 1769–1785 (2013)
3. Cheng, J., Tao, D., Quan, Y., Wong, D.W.K., Cheung, G.C.M., Akiba, M., Liu, J.: Speckle reduction in 3d optical coherence tomography of retina by a-scan reconstruction. IEEE transactions on medical imaging **35**(10), 2270–2279 (2016)
4. Chong, B., Zhu, Y.K.: Speckle reduction in optical coherence tomography images of human finger skin by wavelet modified bm3d filter. Optics Communications **291**, 461–469 (2013)
5. Dong, W., Wang, P., Yin, W., Shi, G., Wu, F., Lu, X.: Denoising prior driven deep neural network for image restoration. IEEE transactions on pattern analysis and machine intelligence **41**(10), 2305–2318 (2018)
6. Fang, L., Li, S., Cunefare, D., Farsiu, S.: Segmentation based sparse reconstruction of optical coherence tomography images. IEEE transactions on medical imaging **36**(2), 407–421 (2016)
7. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)

8. Kafieh, R., Rabbani, H., Selesnick, I.: Three dimensional data-driven multi scale atomic representation of optical coherence tomography. IEEE transactions on medical imaging **34**(5), 1042–1062 (2014)
9. Li, M., Idoughi, R., Choudhury, B., Heidrich, W.: Statistical model for oct image denoising. Biomedical Optics Express **8**(9), 3903–3917 (2017)
10. Ma, Y., Chen, X., Zhu, W., Cheng, X., Xiang, D., Shi, F.: Speckle noise reduction in optical coherence tomography images based on edge-sensitive cgan. Biomedical optics express **9**(11), 5129–5146 (2018)
11. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
12. Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., Xiao, B.: Deep high-resolution representation learning for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 1–1 (2020). https://doi.org/10.1109/TPAMI.2020.2983686
13. Wen, B., Li, Y., Bresler, Y.: When sparsity meets low-rankness: Transform learning with non-local low-rank constraint for image restoration. In: 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP). pp. 2297–2301. IEEE (2017)
14. Wojtkowski, M., Leitgeb, R., Kowalczyk, A., Bajraszewski, T., Fercher, A.F., et al.: In vivo human retinal imaging by fourier domain optical coherence tomography. Journal of biomedical optics **7**(3), 457–463 (2002)
15. Xue, Y., Feng, S., Zhang, Y., Zhang, X., Wang, Y.: Dual-task self-supervision for cross-modality domain adaptation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. pp. 408–417 (2020)
16. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Transactions on Image Processing **26**(7), 3142–3155 (2017)
17. Zhou, Y., Yu, K., Wang, M., Ma, Y., Peng, Y., Chen, Z., Zhu, W., Shi, F., Chen, X.: Speckle noise reduction for oct images based on image style transfer and conditional gan. IEEE Journal of Biomedical and Health Informatics (2021). https://doi.org/10.1109/JBHI.2021.3074852