Simultaneous Noise Reduction and Layer Segmentation for Visible Light Optical Coherence Tomography in Human Retina

Tianyi Ye[1], Jingyu Wang[2], Ji Yi[1,2,†]

1. Department of Biomedical Engineering, Johns Hopkins University, Baltimore USA
2. Department of Ophthalmology, Johns Hopkins University, Baltimore USA

[†]Correspondence: jiyi@jhu.edu

Abstract: Visible light optical coherence tomography (VIS-OCT) is an emerging imaging modality that uses shorter wavelength in visible light range than conventional near infrared (NIR) light. It provides one-micron level axial resolution to improve image contrast to better separate stratified retinal layers, as well as provides microvascular oximetry with spatio-spectral analysis. However, due to the practical limitation of laser safety and comfort, the permissible illumination power is much lower than NIR OCT which can be challenging to obtain high quality VIS-OCT images and subsequent image analysis particularly in pathological eyes. Therefore, improving VIS-OCT image quality by denoising is an essential step in the overall workflow in VIS-OCT clinical applications. In this paper, we provide the first VIS-OCT retinal image dataset from normal eyes, including retinal layer annotation and noisy-clean image pairs. We propose an efficient co-learning deep learning framework for noisy-input segmentation embedded with a self-supervised denoising process. The same neural network performed both denoising and segmentation tasks simultaneously. The task performance is benchmarked qualitatively and quantitatively. The significant improvement of segmentation (2% higher Dice coefficient compared to segmentation-only process) for certain layers is observed when available annotation drops to 25%, indicating a potential angle for annotation-efficient training.

Keywords: VIS-OCT, denoising, segmentation, co-learning

1. Introduction

Optical Coherence Tomography (OCT) is a widely used imaging technique in ophthalmology to evaluate anatomical layers in the retina to diagnose retinal abnormalities and ocular diseases. Current commercialized OCT devices use near-infrared (NIR) light source, either at around 850nm by spectral domain OCT or 1050nm by swept source OCT. The recent emerging visible light optical coherence tomography (VIS-OCT) uses a shorter wavelength visible light centered around 550nm for its illumination. The resulting advantage is higher imaging contrast and one-micron level resolution, which not only enable more accurate analysis for 2D/3D retinal layers but also allows clinicians to perceive and assess the minor morphological changes [1, 2] for potential early detection and treatment of glaucoma, diabetic retinopathy, sickle cell retinopathy, and vein occlusion [3]. Another advantage of VIS-OCT is its spatio-spectral analysis within the microvasculature for label-free oximetry (*i.e.,* measuring hemoglobin oxygen saturation, sO_2) [4]. Song *et al.* demonstrates VIS-OCT reflectivity and spectroscopy of peripapillary retinal nerve fiber layer (pRNFL) are correlated with severity of glaucoma, compared with standard-of-care OCT thickness measurements.[5] Wang *et al.* uses a dual-channel OCT system with wavelengths in the visible and near-infrared light ranges to validates the system's ability to detect structural anomalies in both *en face* and B-scan images with simultaneous retinal OCT angiography (OCTA)

and measurement of sO2 in parafoveal vessels that are around 20–30 μm in diameter. Their work suggests potentially far-reaching clinical implications for the early detection and diagnosis of retinal vascular diseases.[6] Other than measuring hemoglobin oxygen saturation of arterioles, venules and capillaries, the spectral contrast provided by VIS-OCT can also be used for single-scan OCTA [7]. The challenge for VIS-OCT is the more stringent illumination power limit (~0.1-0.25 mW at eye pupil) than NIR OCT (~1-2 mW) and more susceptible to poor optical quality in aging and pathological eyes. Both will lead to degradation of image quality as well as the ability to perform accurate segmentation, a prerequisite for all VIS-OCT quantitative analysis. Therefore, denoising VIS-OCT image and efficient/accurate layer segmentation become critical steps in the general workflow in VIS-OCT.

Recently, several supervised deep learning (DL)-based denoising methods [8, 9] have been proposed, with the need for paired clean images, which are not feasible in some clinical situations. To alleviate such limitations, Huang *et al*. [10] proposed an unsupervised disentanglement method based on CycleGAN [11] with an unpaired noisy-clean dataset. Qiu *et al.* [10] conducted a comparative study to investigate the denoising performance of OCT images over different deep neural networks through an unsupervised Noise2Noise (N2N) [12] strategy with two realizations of the noisy dataset for the same content. At the same time, automatic segmentation of retinal layers has been intensively studied using DL, outperforming traditional graphical and machine learning (ML) methods in several ways [14, 15, 16, 17, 18]. However, most of these methods require a large amount of manual annotation by clinicians for training, which is time-consuming and subjective to variation among clinicians and devices. In addition, the diversity of training data significantly affects the performance of test data. Therefore, annotation-efficient approaches are necessary. Such approaches include the semi-supervised method [19] and the weakly-supervised method [20].

While many DL approaches have been shown to be successful in each task, considering the real-world clinical scenarios, in which both denoising and segmentation are necessary, an efficient way to solve both the problems simultaneously is significant. In this paper, we reported several important progresses on an efficient DL framework for VIS-OCT image processing. First, we collected and published the first VIS-OCT dataset for both denoising and segmentation tasks, and importantly, their combination, with noisy-clean image pairs and ten manually delineated retinal boundaries. Second, along with the dataset, we provide a co-learning framework based on residual-UNet for simultaneous denoising and segmentation (named DenoiSegOCT) with the noisy VIS-OCT image itself. Inspired yet different from DenoiSeg [21], our approach extends the self-supervised strategy (N2V) [22] to inherent noise instead of synthesized Gaussian noise and extends the 3-class segmentation (background, foreground, and edges), which removes the class information in the annotated label and has large gaps between foreground parts, to 10-class segmentation. Such co-learning process significantly reduces the hyperparameters tuning time and simultaneously provides both denoised image and prediction of segmentation. The framework provides generally considerable performance compared to UNet [23] baseline that only segments while do both tasks at the same time. In addition, compared to segmentation-only frameworks (i.e., UNet and ours without denoising), we observe a significant improvement of segmentation for certain layers when available annotation drops to 25%. As for the denoising task, the denoising performance is qualitatively better than using N2V alone.

2. Methods

2.1. VIS-OCT dataset

This paper presents the first VIS-OCT retina image dataset for further machine-learning research. The data include retinal B-scans obtained on our 2nd Gen dual-channel VIS-OCT system [24] for 12 subjects, all displayed normal-appearing VIS-OCT images with varying image quality. The provided dataset includes noisy-clean image pairs, wherein the clean image is obtained by averaging 16 or 32 B-scans at the same position of the retina. It also includes 10 manually delineated retinal boundaries (Fig. 1), each of which is individually reviewed and edited. The dataset consists of 105 B-scans with a size of 512×512 in diverse positions of the retina.

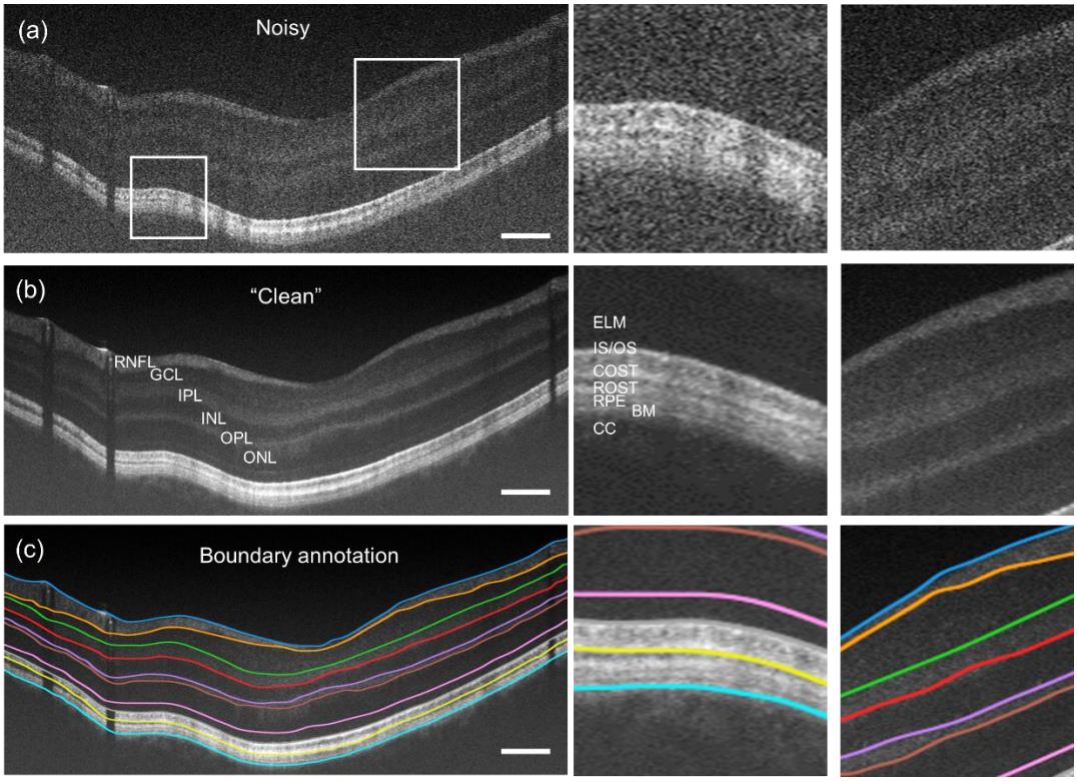

Fig. 1. A noisy-clean (a-b) B-scan pair with 10 manually delineated retinal boundaries and corresponding layers (c): the retinal nerve fiber layer (RNFL); the ganglion cell layer (GCL); the inner plexiform layer (IPL); the inner nuclear layer (INL); the outer plexiform layer (OPL); the outer nuclear layer (ONL); the external limiting membrane (ELM); the inner segment (IS); the outer segment (OS); the cone outer segment tip (COST); the rod outer segment tip (ROST) and retinal pigment epithelium (RPE); and finally the Bruch's Membrane (BM). Zoom-in views from two small regions in inter and outer retina were displayed for comparison.

2.2. Network architecture

In the proposed DenoiSegOCT framework (Fig. 2), we utilize UNet-like encoder-decoder architecture for both denoising and segmentation tasks. The base block of each level of encoders and decoders includes 2 convolutional kernels with the residual operation to take advantage of residual learning [25] that improves the gradient flow during the optimization process. The depth

= 5 is necessary for two reasons to *1)* provide large enough receptive field to capture the global information (i.e., the order of the retinal layers, which is the important prior anatomical knowledge in the task, and *2)* deep enough (i.e., more parameters and non-linear units) for the model to learn high-level features like the certain object of the layers and the order of the layers. The initial number of feature maps = 96 is for the enough model ability to represent the learned features.

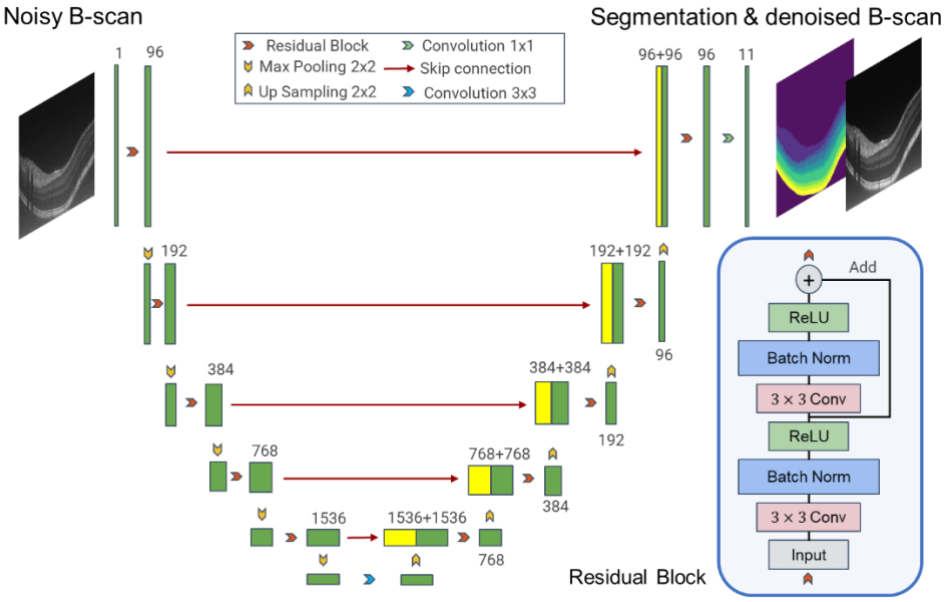

Fig. 2. Architecture of DenoiSegOCT

2.3. Self-supervised denoising

We use the Noise2Void strategy (N2V) to reduce the noise and inherent speckle noise of VIS-OCT by adding an additional channel to the output layer of the network. N2V is a self-supervised denoising method that randomly selects and modifies several spots (as blind spots of the network) in the image and then trains a neural network to restore the modified spots to original pixel values by optimizing mean squared error (MSE) loss of the two. During the training process, the network learns to predict the modified spot by looking at a surrounding area of it, which is the receptive field of the network. Since the random speckle noise in OCT is pixel-independent and the underneath content is pixel-dependent, such process can restore the content information degraded by the noise. In our setup, each selected spot in the image is modified by a randomly selected pixel of its surrounding areas, and the masking strategy in N2V is applied for the modification and optimization.

2.4. Ten-class segmentation

The pixel-wise ground truth mask, including 9 retinal layers and the background for the 10-class segmentation task, is created by filling pixels between the delineated retinal boundaries. We use a weighted cross-entropy loss with empirical weights for each class to alleviate the data-imbalance problem.

2.5. Co-learning strategy

The output layer of the network consists of eleven channels, of which one channel is the output of the denoised image and the remaining ten channels provide the probability that each pixel belongs to the corresponding category. The network is jointly trained by optimizing a combined loss with a task weight factor α and class weight factors wi. For a training batch the combined loss is calculated as:

$$L_c = \alpha L_d(f(x_{ms}), x_{os}) + (1-\alpha) \sum_i w_i L_{si}(f(x_m), y) \quad (1)$$

where L_d and L_{si} are the MSE loss for denoising and cross entropy loss for segmentation of class i, x_{ms} and x_{os} are the modified spots and original spots of the input image, respectively, $f(\mathrm{x})$ denotes the output of the network given an input x, x_m is the modified image and y is the corresponding pixel-wise label.

3. Experiments

3.1. Baselines and comparisons

We select the standard UNet architecture with 64 initial feature maps and a depth of 4 using both noisy and clean images as input to evaluate the baseline segmentation performance for our dataset (named UNet clean/noisy). Also, we conduct experiments of sequential strategy that first denoise the image using a N2V model pretrained on our dataset and then train the UNet baseline (named N2V+UNet). The pretrained N2V also serves as a baseline for self-supervised denoising. The amount of annotated training data is set to 100%, 50%, 25%, and 10% to evaluate the label-efficient property of DenoiSegOCT. The loss of segmentation is set to zero for those images whose labels are unavailable. We then set the weight factor α to zero to make it a pure segmentation task to evaluate the importance of the denoising process.

3.2. Dataset and preprocessing

To the best of our knowledge, currently none of the public datasets are available for simultaneous speckle noise reduction and segmentation. Therefore, we use our VIS-OCT dataset for experiments. Our VIS-OCT dataset contains retinal images of 12 subjects (105 B-scans), of which 3 subjects (51 B-scans) are used for training, 4 subjects (17 Bscans) are used for validation and 5 subjects (37 B-scans) are used for test samples.

The network was implemented using Tensorflow and optimized by Adam optimizer. The initial learning rate is 0.0004 and is halved if the loss on the validation set is not decreasing over ten epochs. The data is augmented with horizontal flipping and then replicated by a factor of 2, followed by random cropped to 256×256 with a batch size of 12 during training. 1.5 percent of the pixels in the input image were modified during the random batch cropping.

The task weight factor α is set to 0.5 and class weight factors wi are set to 0.5 for the background and 1 for the other classes. The network was trained for 200 epochs and the model with the lowest validation loss is selected to test.

4. Results

All the segmentation experiments are repeated 8 times and the mean and standard deviation is presented.

4.1. Self-supervised denoising

We quantitively evaluate our denoising performance using peak signal noise ratio (PSNR) and structure similarity index (SSIM), as shown in (Table 1). Both N2V and DenoiSegOCT can significantly increase the average PSNR/SSIM of noisy images. Note that the annotation amount for segmentation is not significantly affecting the self-supervised denoising and the small fluctuation of PSNR and SSIM does not indicate the visual difference. However, from the visualized denoised images we observe that our framework provides restoration result with sharper edges and higher contrast, which is not reflected in the difference of PSNR and SSIM.

Table 1. PSNR and SSIM off 2V denoised image(N2V), and DenoiSegOCT with 100%, 50%, 25% and 10% annotation available (DenoiSegOCT100 to DenoiSegOCT10).

Method	Noisy	N2V	DenoiSegOCT100	DenoiSegOCT50	DenoiSegOCT25	DenoiSegOCT10
PSNR	22.90	27.26(±0.2470)	**27.29**(±0.1503)	27.17(±0.1552)	26.96(±0.2760)	27.20(±0.2478)
SSIM	0.2639	**0.5296**(±0.0083)	0.5199(±0.0054)	0.5180(±0.0055)	0.5145(±0.0063)	0.5168(±0.0069)

4.2. Segmentation

The Dice coefficient is used to evaluate the segmentation performance (Table 2). With 100% annotation available (Dice100), our DenoiSegOCT using noisy input can achieve a comparable Dice coefficient compared to UNet baseline with clean input and the sequential strategy N2V+UNet, while superior to UNet baseline using noisy input, which indicates the model capability is necessary using our architecture. The DenoiSegOCT becomes more advantageous as the number of available annotations decreases. The ablation study that compares using our architecture with and without denoising process with noisy inputs indicates a significant improvement (2% higher Dice over all test data and all 8 experiments) for GCL, IPL and INL segmentation when available annotation drops to 25% (Table 3). These layers are the blur and ambiguous parts in the images because of the low SNR (weak signal and/or strong noises). At those parts, the edges and shapes are difficult to detect by both human, and machine, who captures the features to distinguish layers then segments them. Normally the supervision of annotation can narrow the gap if there is a fair amount of training data. However, we believe that when the annotation drops a lot, the denoising process takes over and helps capture similar and useful features for segmentation. OPL, however, is the most challenging layer because it is thin and irregular, resulting unstable training with few annotations. This observation aligns with [15]. Also, introducing denoising process can increase the task difficulty, resulting slight performance drop on other layers that are easy to perceive even with noise. Thus, the average improvement of all layers is not significant.

Table 2. Dice coefficient for UNet baselines with clean input and noisy input, N2V+UNet, and our DenoiSegOCT. Dice100, Dice50, Dice25 and Dice10 represent the 100%, 50%, 25%, and 10% of the annotation are available, respectively.

Method	Dice100	Dice50	Dice25	Dice10
UNet(clean)	**0.8852**(±0.0074)	0.8600(±0.0294)	0.7637(±0.0520)	0.6007(±0.0705)
UNet(noisy)	0.8710(±0.0032)	0.8513(±0.0064)	0.8312(±0.0109)	0.7631(±0.0142)
UNet(N2V)	0.8771(±0.0052)	**0.8607**(±0.0047)	0.8369(±0.0050)	0.7472(±0.0202)
Ours(noisy)	0.8740(±0.0042)	0.8602(±0.0103)	**0.8437**(±0.0123)	**0.7652**(±0.0268)

Table 3. Dice coefficient of background (BG) and all 9 retinal layers. w100: 100% segmentation annotation available using DenoiSegOCT with denoising process.

Layers	BG	RNFL	**GCL**	**IPL**	**INL**	OPL	ONL	IS/OS	COST	ROST+RPE	Avg
w100	0.9874	0.8641	0.8339	0.8106	0.8782	0.7524	0.9315	0.9086	0.8951	0.8690	0.8604
w/o100	0.9842	0.8517	0.8410	0.8240	0.8765	0.7517	0.9338	0.9099	0.8940	0.8641	0.8607
w50	0.9859	0.8334	0.8078	0.8064	0.8641	0.7338	0.9141	0.9010	0.8865	0.8576	0.8450
w/o50	0.9814	0.8398	0.8064	0.8148	0.8687	0.7385	0.9257	0.9020	0.8884	0.8542	0.8487
w25	0.9732	0.8128	**0.8111**	**0.7899**	**0.8359**	0.6818	0.9074	0.8919	0.8869	0.8366	0.8283
w/o25	0.9714	0.8279	0.7903	0.7627	0.8103	0.6888	0.9097	0.8926	0.8879	0.8516	0.8247

4.3. Visualization of the overall result

We visualize the denoising and segmentation performance for two test samples (Fig. 3). As for denoising, Comparing N2V (c1, c2) to DenoiSegOCT(d1, d2) with 50% annotation, higher contrast and shaper edges are observed using DenoiSegOCT than that using N2V alone. This is however not reflected in the PSNR and SSIM. As for segmentation with only 25% annotation, comparing DenoiSegOCT without denoising process (f1, f2) to DenoiSegOCT with denoising process (h1, h2) one can see that with the denoising process the segmentation performance of GCL, IPL and INL is significantly improved.

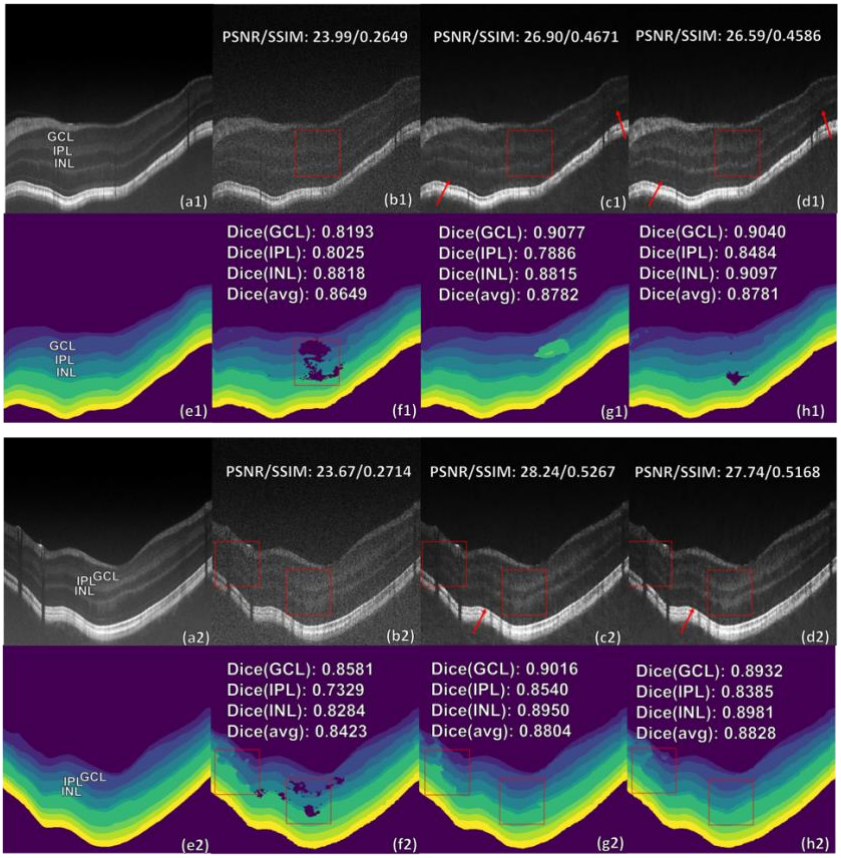

Fig. 3. Overall denoising visualization with 50% annotation and segmentation visualization with 25% annotation. Two test samples are displayed. For a sample x: (ax) is the clean ground truth, (bx) is the noisy input, (cx) is the N2V denoised image, (dx) is the DenoiSegOCT denoised image, (ex) is the manual segmentation mask, (fx),(gx),(hx) are the predicted segmentation masks using ours without denoising, using UNet with image denoised by N2V, and using ours with denoising, respectively.

5. Conclusion

In this paper, we present the first VIS-OCT retinal image dataset for future data-driven method research. We proposed an efficient co-learning framework DenoiSegOCT to simultaneously reduce speckle noise and segment 9 retinal layers and background. Performing denoising and segmentation simultaneously also reduces the complexity and time compared with optimizing the two tasks separately. Given the qualitative and quantitative results, our framework achieves considerable performance with self-denoising and segmentation baselines and provides an angle for few-shot or small dataset learning of segmentation, which is common in many clinical scenarios. We plan to extend the proposed VIS-OCT dataset in the future. With more training data, the 3-D model can be introduced to leverage the intra-frame information to further improve performance of both tasks. In addition, the co-learning effect should be enhanced because the self-denoising and segmentation can use more data to help each other.

Acknowledgements

This study was supported by NIH R01NS108464, and R01EY032163.

Reference

[1] Shau Poh Chong, Marcel Bernucci, Harsha Radhakrishnan, and Vivek J Srinivasan, “Structural and functional human retinal imaging with a fiber-based visible light oct ophthalmoscope,” *Biomedical optics express*, vol. 8, no. 1, pp. 323–337, 2017.

[2] Ji Yi, Siyu Chen, Xiao Shu, Amani A Fawzi, and Hao F Zhang, “Human retinal imaging using visible-light optical coherence tomography guided by scanning laser ophthalmoscopy,” *Biomedical optics express*, vol. 6, no. 10, pp. 3701–3713, 2015.

[3] Xiao Shu, Lisa Jane Beckmann, and Hao F Zhang, “Visible-light optical coherence tomography: a review,” *Journal of biomedical optics*, vol. 22, no. 12, pp. 121707, 2017.

[4] Song W, Shao W, Yi W, Liu R, Desai M, Ness S, Yi J. Visible light optical coherence tomography angiography (VIS-OCTA) facilitates local microvascular oximetry in the human retina. *Biomed Opt Express*. 2020 Jun 30;11(7):4037-4051. doi: 10.1364/BOE.395843. PMID: 33014584; PMCID: PMC7510897.

[5] Weiye Song, Sui Zhang, Yumi Mun Kim, Natalie Saldlak, Marissa G. Fiorello, Manishi Desai, Ji Yi, "Visible light optical coherence tomography of peripapillary retinal nerve fiber layer reflectivity in glaucoma", TVST, 11(28), 2022

[6] Jingyu Wang, Andrew Baker, Manju L. Subramanian, Nicole H. Siegel, Xuejing Chen, Steven Ness, Ji Yi, "Simultaneous visible light optical coherence tomography and near infrared OCT angiography in retinal pathologies: a case study", Experimental Biology and Medicine, December 14, 2021.

[7] Winkelmann, J.A., Eid, A., Spicer, G. et al. Spectral contrast optical coherence tomography angiography enables single-scan vessel imaging. Light Sci Appl 8, 7 (2019).

[8] Yuhui Ma, Xinjian Chen, Weifang Zhu, Xuena Cheng, Dehui Xiang, and Fei Shi, “Speckle noise reduction in optical coherence tomography images based on edgesensitive cgan,” *Biomedical optics express*, vol. 9, no. 11, pp. 5129–5146, 2018.

[9] Devalla, S. K., Subramanian, G., Pham, T. H., Wang, X., Perera, S., Tun, T. A., ... & Girard, M. J. (2019). A deep learning approach to denoise optical coherence tomography images of the optic nerve head. *Scientific reports*, 9(1), 1-13.

[10] Yongqiang Huang, Wenjun Xia, Zexin Lu, Yan Liu, Hu Chen, Jiliu Zhou, Leyuan Fang, and Yi Zhang, “Noise-powered disentangled representation for unsupervised speckle reduction of optical coherence tomography images,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 10, pp. 2600–2614, 2020.

[11] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[12] Bin Qiu, Shuang Zeng, Xiangxi Meng, Zhe Jiang, Yunfei You, Mufeng Geng, Ziyuan Li, Yicheng Hu, Zhiyu Huang, Chuanqing Zhou, et al., "Comparative study of deep neural networks with unsupervised noise2noise strategy for noise reduction of optical coherence tomography images," *Journal of Biophotonics*, vol. 14, no. 11, pp. e202100151, 2021.

[13] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila, "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.

[14] Mike Pekala, Neil Joshi, TY Alvin Liu, Neil M Bressler, D Cabrera DeBuc, and Philippe Burlina, "Deep learning based retinal oct segmentation," Computers in biology and medicine, vol. 114, pp. 103445, 2019.

[15] Abhijit Guha Roy, Sailesh Conjeti, Sri Phani Krishna Karri, Debdoot Sheet, Amin Katouzian, Christian Wachinger, and Nassir Navab, "Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," Biomedical optics express, vol. 8, no. 8, pp. 3627–3642, 2017.

[16] Cecilia S Lee, Ariel J Tyring, Nicolaas P Deruyter, Yue Wu, Ariel Rokem, and Aaron Y Lee, "Deep-learning based, automated segmentation of macular edema in optical coherence tomography," Biomedical optics express, vol. 8, no. 7, pp. 3440–3448, 2017.

[17] Stefanos Apostolopoulos, Sandro De Zanet, Carlos Ciller, Sebastian Wolf, and Raphael Sznitman, "Pathological oct retinal layer segmentation using branch residual u-shape networks," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2017, pp. 294–301.

[18] Yufan He, Aaron Carass, Yihao Liu, Bruno M Jedynak, Sharon D Solomon, Shiv Saidha, Peter A Calabresi, and Jerry L Prince, "Structured layer surface segmentation for retina oct using fully convolutional regression networks," Medical image analysis, vol. 68, pp. 101856, 2021.

[19] Xiaoming Liu, Jun Cao, Tianyu Fu, Zhifang Pan, Wei Hu, Kai Zhang, and Jun Liu, "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, 2018.

[20] Jing Wang, Wanyue Li, Yiwei Chen, Wangyi Fang, Wen Kong, Yi He, and Guohua Shi, "Weakly supervised anomaly segmentation in retinal oct images using an adversarial learning approach," *Biomedical optics express*, vol. 12, no. 8, pp. 4713–4729, 2021.

[21] Tim-Oliver Buchholz, Mangal Prakash, Deborah Schmidt, Alexander Krull, and Florian Jug, "Denoiseg: joint denoising and segmentation," in *European Conference on Computer Vision*. Springer, 2020, pp. 324–337.

[22] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug, "Noise2void-learning denoising from single noisy images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2129–2137.

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[24] Wang, J., Nolen, S., Song, W., Shao, W., Yi, W., & Yi, J. (2022). Second-generation dual-channel visible light optical coherence tomography enables wide-field, full-range, and shot-noise limited retinal imaging. BioRxiv 2022.10.05.511048; doi: https://doi.org/10.1101/2022.10.05.511048

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.