Title

Learning to Correct Axial Motion in Oct for 3D Retinal Imaging

Authors

Wang, Yiqian

Warter, Alexandra

Cavichini-Cordeiro, Melina

et al.

Publication Date

2021-09-01

DOI

10.1109/icip42928.2021.9506620

Peer reviewed

HHS Public Access

Author manuscript

Proc Int Conf Image Proc. Author manuscript; available in PMC 2022 August 09.

Published in final edited form as:

Proc Int Conf Image Proc. 2021 September ; 2021: 126–130. doi:10.1109/icip42928.2021.9506620.

LEARNING TO CORRECT AXIAL MOTION IN OCT FOR 3D RETINAL IMAGING

Yiqian Wang[1], **Alexandra Warter**[2], **Melina Cavichini-Cordeiro**[2], **William R. Freeman**[2], **Dirk-Uwe G. Bartsch**[2], **Truong Q. Nguyen**[1], **Cheolhong An**[1]

[1]Department of Electrical and Computer Engineering, University of California, San Diego

[2]Jacobs Retina Center, Shiley Eye Institute, La Jolla, California, USA

Abstract

Optical Coherence Tomography (OCT) is a powerful technique for non-invasive 3D imaging of biological tissues at high resolution that has revolutionized retinal imaging. A major challenge in OCT imaging is the motion artifacts introduced by involuntary eye movements. In this paper, we propose a convolutional neural network that learns to correct axial motion in OCT based on a single volumetric scan. The proposed method is able to correct large motion, while preserving the overall curvature of the retina. The experimental results show significant improvements in visual quality as well as overall error compared to the conventional methods in both normal and disease cases.

Index Terms—

Motion correction; optical coherence tomography; eye movement; deep learning; retinal imaging

1. INTRODUCTION

Optical Coherence Tomography (OCT) is a powerful technique for non-invasive 3D imaging of biological tissues at μm resolution that has become one of the most important diagnostic modality in retinal imaging [1, 2, 3]. In OCT imaging, the sample is probed with a low-coherent infrared beam and the magnitude of the reflected or backscattered light along the beam axis (A-scan, Z axis of Fig. 1) is measured at different depth by interference. Sequential cross-sectional images (B-scan, XZ plane of Fig. 1) are acquired by raster-scanning the infrared beam transversely through the sample and a 3D volume can be formed by stacking the cross-sectional images as illustrated in Fig. 1 (a) and (b). The direction in which each B-scan is acquired is also called *fast scanning axis*, whereas the orthogonal direction where B-scans are stacked is called the *slow scanning axis*. The plane spanned by the two axes is called the *coronal* or *en-face plane*.

A major challenge in OCT imaging is the axial and horizontal motion artifacts introduced by involuntary eye movements. Even when the patient fixates upon a fixed object, the eye still carries out small and rapid movements including rapid microsaccades, high frequency tremors, and slow drifts with various frequency and magnitude [4]. However, even with high-speed Spectral Domain-OCT (SD-OCT) devices that can acquire more than 300,000

A-scans per second [5], it typically takes several seconds to acquire a 3D volume [6]. Therefore, artifacts due to the fixational eye movements are almost inevitable [4]. These involuntary eye movements would introduce both axial and coronal distortion, which leads to discontinuity in display and visualization of the 3D OCT data as shown in Fig. 1 (a), where the motion artifacts cause discontinuity pointed by red arrows. It has also been shown that eye motion artifacts compromises subsequent analysis including retinal layer segmentation, OCT-Angiography (OCT-A) imaging, and detection of retinal diseases [7].

Existing motion correction approaches include both prospective hardware eye tracking systems that compensate for motion [8, 9], and retrospective software solutions that use image registration techniques to align each 2D B-scan [5, 10, 11]. Even though hardware solutions may yield more accurate correction results, they are unfortunately not available for every OCT device. Existing software solutions either create overly smoothed output that flattens the true curvature of retina based on a single volume [12, 10], or rely on multiple OCT volumes or multimodal reference images [5, 11] which adds extra burden for clinical tests.

In this paper, we propose a convolutional neural network that corrects axial motion in OCT based on a single volume, while recovering the overall curvature of the retina. Since OCT is the most important imaging modality in most retinal diseases [2, 3], the proposed method will significantly improve the accuracy of many down-stream tasks such as retina disease classification, detection, and segmentation [7]. To the best of our knowledge, this is the first deep learning based method applied to the OCT motion correction problem, and experiment results show significant improvement compared with conventional methods in both normal and disease cases.

2. RELATED WORKS

Even though involuntary eye motion is a major problem in OCT imaging, the motion correction problem has not been fully resolved based on current literature. Two review papers published in 2017 [14] and 2019 [4] extensively discussed and summarized existing works on the OCT motion correction problem. The fast B-scans are often considered artifact-free as the frame acquisition rate of modern devices is faster than the expected motion in the retina [4]. Therefore, most works correct axial and coronal movement by treating fast B-scans as rigid bodies. Since axial movement is often more significant compared to coronal movement in magnitude [5], we only focus on axial motion correction in this paper.

Most approaches fall in two major categories: *prospective approaches* and *retrospective approaches* [4]. Prospective approaches are mostly hardware-based solution. By using additional eye tracking devices [8, 9], these hardware solutions often yield more accurate alignment results [14]. Some other approaches are highly dependent on special scanning patterns [15] and signal acquisition techniques [16] to recover a motion-free OCT volume, yet they are difficult to implement and cannot correct eye movements in regular OCT scans. The retrospective approaches are mostly software-based solutions. Potsaid et al. proposed a method that corrects axial motion using orthogonal OCT scans to both horizontal and

vertical directions [5]. Kraus et al. [11, 17] proposed another method using orthogonal scans, which can correct both axial and coronal motion in two stages, and has been widely used as a standard correction in OCT-A imaging [4]. Since these methods require more than one OCT volumes, they increase (double) the time needed for clinical examinations and add extra burden to limited medical resources.

Some other methods aim to correct eye motion based on a single OCT volume. Antony et al. [12] utilized segmentation of the retinal pigment epithelium (RPE) layer based on thin-plate spline fitting. However, the method flattens the natural curvature of the retina, which is undesirable for observing diseases that alters the RPE layer. Xu et al. proposed a method based on particle filter [10], but the method also tended to flatten the retina and was validated only on synthetic motions within 2–10 pixel range. The algorithm of Montuoro et al. [6] can maintain the curvature of the retina to some extend based on a local symmetry assumption of the segmented RPE but the assumption does not always hold for retina with diseases. Fu et al. [18] proposed a saliency-based method to correct the axial and coronal motion, but the authors only tested the coronal correction on synthetic data with motion fewer than 15 pixels.

3. PROPOSED METHOD

It is crucial to develop a motion correction algorithm that does not depend on auxiliary hardware and operates on a single OCT volume. Furthermore, the algorithm should be robust to diseases and large motion artifacts, while preserving the overall curvature of the retina. We propose a convolutional neural network to correct motion artifacts in OCT retinal imaging so that it requires a single OCT volume and retrieves a smooth and anatomically correct 3D representation.

We adopt a U-Net [13] like structure with residual blocks to predict a displacement map based on a single OCT volume. The proposed method operates on the 3D OCT volume $\mathbf{V} \in \mathbb{R}^{H \times W \times N}$ formed by stacking 2D B-scans, where H and W are the width and height of each B-scan and N is the number of B-scans. As shown in Fig. 2, the Z axis of the input OCT volume is treated as channels, while the X, Y axes are considered as spatial dimensions. The output is a displacement map $\mathbf{D} \in \mathbb{R}^{W \times N}$ where each pixel contains a displacement value to Z axis. Negative displacement shifts the A-scan upwards and positive displacement shifts it downwards. The magnitude of displacement denotes the number of pixel to be shifted divided by a normalization factor Z_{norm} for better numerical stability. Finally, the motion corrected OCT volume $\mathbf{V}_{\text{out}}$ can be obtained by

$$\mathbf{V}_{\text{out}}(z, x, y) = \mathbf{V}(z - Z_{\text{norm}}\mathbf{D}(x, y), x, y)\,. \tag{1}$$

As illustrated in Fig. 2, 1 × 1 convolution is applied at the first layer to compress the number of channels and the other convolutions are 3 × 3. Instance normalization (IN) is applied after convolutions in order to normalize over the spatial dimensions without being influenced by other volumes in the same batch. For the three down sampling blocks denoted by red arrows, 2 × 1 convolution with stride 2 × 1 is adopted to downsample the X dimension by 2, while

keeping the resolution on the Y dimension unchanged, because the number of B-scans N in our dataset is significantly smaller than the width of B-scans W. Similarly, 2 × 1 transposed convolution with stride 2 × 1 is used to upsample X dimension by 2 in three upsampling blocks which are indicated by purple arrows.

We also include the segmentation of the inner limiting membrane (ILM) and the retinal pigment epithelium (RPE) layer for better performance. As shown in Fig. 2, we first normalize two segmentation boundaries and concatenate them with the output of the network of Fig. 2, and then apply two additional layers with 1 × 1 convolution to get the final displacement prediction. Denoting the two segmentation boundaries with $\mathbf{B} \in \mathbb{R}^{2 \times W \times N}$ where $\mathbf{B}(0, x, y)$ and $\mathbf{B}(1, x, y)$ entries denote Z coordinates of the ILM and RPE layers at pixel (x, y). The overall retinal curvature $\mathbf{P} \in \mathbb{R}^{2 \times W \times N}$ is first computed by

$$\mathbf{P}(z, x, y) = \mathbf{B}(z, x, 0) + \frac{\mathbf{B}(z, x, N-1) - \mathbf{B}(z, x, 0)}{N-1} y \tag{2}$$

where $z \in \{0, 1\}$ and $(x, y) \in [0, W-1] \times [0, N-1]$. Then, the normalized boundaries $\mathbf{B}'$ can be obtained by

$$\mathbf{B}' = (\mathbf{P} - \mathbf{B}) / Z_{\text{norm}}. \tag{3}$$

The loss function includes a displacement L1 loss and a smoothness loss. Denoting the predicted displacement $\mathbf{D}$ and the ground truth displacement $\mathbf{D}_{\text{GT}} \in \mathbb{R}^{W \times N}$, the displacement L1 loss is given by

$$\mathscr{L}_{\text{disp}}(\mathbf{D}; \mathbf{D}_{\text{GT}}) = \underset{x, y}{\text{mean}}(\mathbf{M}(x, y) |\mathbf{D}(x, y) - \mathbf{D}_{\text{GT}}(x, y)|), \tag{4}$$

where $|\cdot|$ denotes absolute value and $\mathbf{M} \in \mathbb{R}^{W \times N}$ is a predefined mask in [0,1] to assign more weight at the center and less weight to the boundary of the OCT volume, since the region of interest is often placed at the center in clinical practice. The second term is a smoothness loss to enforce smoothness along the fast-scanning axis, defined by

$$\mathscr{L}_{\text{smooth}}(\mathbf{D}) = \sum_{s=1,2} \underset{x, y}{\text{mean}}(|\mathbf{D}^s(x+1, y) - \mathbf{D}^s(x, y)|), \tag{5}$$

where $\mathbf{D}^1$ denotes the displacement at original resolution, and $\mathbf{D}^2$ denotes the displacement downscaled by 2 to the X axis. Let λ_{smooth} be a weighting factor and the total loss is a weighted sum of the two terms

$$\mathscr{L} = \mathscr{L}_{\text{disp}}(\mathbf{D}; \mathbf{D}_{\text{GT}}) + \lambda_{\text{smooth}} \mathscr{L}_{\text{smooth}}(\mathbf{D}) \tag{6}$$

In order to obtain ground truth (artifacts-free) volumes and corresponding displacement maps, pairs of horizontal and vertical 3D OCT volumes with motion artifacts are collected, and each volume is corrected with its orthogonal reference using the motion correction

algorithm in [5], as illustrated in Fig. 3. Note that we only use horizontal and vertical pair volumes for ground truth but the proposed network takes only one single (horizontal or vertical) 3D OCT volume as input.

We augment training data with random flipping on X and Y axis of the input OCT volume and segmentation boundaries to prevent over-fitting. We also add random displacement for augmentation on top of the existing eye motion. An N-dimensional Gaussian random vector following $\mathcal{N}(0, 1)$ is first generated, and its cumulated sum is computed and normalized to get the Y dimensional augmentation $\mathbf{N}_{\mathrm{Y}} \in \mathbb{R}^{W}$ where $\mathbf{N}_{\mathrm{Y}}(0) = \mathbf{N}_{\mathrm{Y}}(1) = 0$, as shown in Fig. 4 (a). The augmentation for X dimension is generated by interpolating between 0 and a random number drawn from $\mathcal{N}(0, 1)$, as shown Fig. 4 (b). Finally, the total augmentation $\mathbf{N} = \mathbf{1}_{W}\mathbf{N}_{\mathrm{Y}}^{T} + \mathbf{N}_{\mathrm{X}}\mathbf{1}_{N}^{T}$ is applied to the input OCT volume and the Y augmentation is subtracted from the ground truth displacement, $\mathbf{D}'_{\mathrm{GT}} = \mathbf{D}_{\mathrm{GT}} - \mathbf{1}_{W}\mathbf{N}_{\mathrm{Y}}^{T}$ where $\mathbf{1}_{n}$ denotes a n-dimensional vector of ones.

At inference time, a linear function is fitted to the X axis of the predicted displacement $\mathbf{D}$ via linear least squares as a post-processing step in order to guarantee that the resulting fast B-scans in $\mathbf{V}_{\mathrm{out}}$ have no distortion except shearing.

4. EXPERIMENTAL RESULTS

In the experiment, we compare the motion correction performance of our proposed network to that of three other methods [6, 12, 18] which operates on a single OCT volume. We evaluate the performance of motion correction algorithms on our dataset collected by Jacobs Retina Center. The dataset of 55 eyes contains paired horizontal and vertical OCT volumes which are obtained by Heidelberg Spectralis in an imaging volume of 1.9 × 5.8 × 5.8 (mm^3) with 20 degree field of view. All the volumes come with instrument's segmentation boundaries of 11 retinal layers. Among 55 horizontal and 55 vertical volumes, the dimensions of 9 volumes are 496 × 512 × 25, while those of the remaining 101 volumes are 496 × 512 × 49. 55×horizontal and 55 vertical OCT volumes in total 110 are divided into 75, 10, 25 for training, validation, and test, respectively. The dataset includes both healthy subjects as well as patients with a variety of diseases such as wet and dry AMD (age-related macular degeneration), ERM (epi-retinal membrane), macular edema, diabetic retinopathy, retinal detachment, macular hole, chorioretinopathy, and posterior vitreous detachment.

The proposed network with 436K parameters is implemented in PyTorch. Reflected padding is used for convolutions and dropouts with $p = 0.2$ is used on every resolution level. We set $Z_{\mathrm{norm}} = 10$ and $\lambda_{\mathrm{smooth}} = 0.5$ for the normalization factor and the smoothness weight, respectively, which are tuned for the best performance. The model is trained with Adam optimizer with weight decay 10^{-3}, batch size 4, an initial learning rate of 10^{-3} and exponential decay with momentum 0.99 for 500 max epochs. The best model is selected based on lowest validation loss and tested on the test set. The approach for ground truth acquisition [5] and two comparison methods [6, 12] are implemented in Python. The axial correction step in [18] is implemented in MATLAB based on the original authors' implementation of saliency detection [19]. We evaluate the algorithms based on two

aspects: the smoothness of OCT volume and the pixel-wise mean absolute error (MAE). After motion correction, the smoothness of OCT volume is measured using the average of normalized mutual information (NMI) [20] between neighboring B-scans (MNMI), and higher MNMI indicates more smoothness.

$$\mathrm{MNMI}(\mathbf{V}_{\mathrm{out}}) = \underset{y}{\mathrm{mean}}(\mathrm{NMI}(\mathbf{V}_{\mathrm{out}}(z, x, y), \mathbf{V}_{\mathrm{out}}(z, x, y+1))).$$

MAE between the predicted and ground truth displacement measures the overall accuracy and preservation of retinal curvature and smaller MAE shows better performance.

$$\mathrm{MAE}(\mathbf{D}; \mathbf{D}_{\mathrm{GT}}) = Z_{\mathrm{norm}} \underset{x,\, y}{\mathrm{mean}}(|\mathbf{D}(x, y) - \mathbf{D}_{\mathrm{GT}}(x, y)|).$$

The qualitative results of different registration methods are shown in Fig. 5. Row (1) and (2) show an example OCT volume with disease, while row (3) and (4) show another example with large motion. The method of [12] in column (b) flattens the RPE and results in errors when the disease alters RPE in (b2). Method [6] of column (c) can smooth axial motion without flattening the retina, but it results in an unnatural curvature at large motion in (c2) and (c4) that does not resemble the ground truth. The results of [18] in column (d) are smooth in most B-scans, but the errors lead to abrupt discontinuities. Our proposed method in column (e) can reduce the motion in the input volume while recovering the overall curvature.

The quantitative results of MNMI and MAE are shown in Table 1, where each entry shows the mean and standard deviation value. We first compute the baseline MNMI and MAE before correction and the MNMI for ground truth correction [5]. As method [12] flattens the retina, it achieves the highest MNMI at 0.5927, but it also yeilds larger MAE than the input. Method [18] also tends to over-smooth the input, reflected by a MNMI larger than the ground truth and an increased MAE. Method [6] improves the MNMI and reduces the MAE compared with the input by a small margin. Overall, our method achieves the lowest MAE at 7.86 pixels, and the MNMI is close to the smoothness of ground truth.

5. CONCLUSION

In conclusion, we propose a deep learning approach that uses the convolutional neural network to predict a displacement map from a single horizontal or vertical OCT volume. The experimental results show that the proposed method is able to correct large motion while recovering the retinal curvature, achieving significant improvements compared to the conventional methods. In future work, we will extend our proposed network to support coronal motion correction besides axial motion. The proposed method will lead to better display and visualization of 3D OCT volumes and benefit subsequent processing including retinal layer segmentation and OCT-A imaging.

REFERENCES

[1]. Huang David, Swanson Eric A, Lin Charles P, Schuman Joel S, Stinson William G, Chang Warren, Hee Michael R, Flotte Thomas, Gregory Kenton, Puliafito Carmen A, et al. , "Optical coherence tomography," science, vol. 254, no. 5035, pp. 1178–1181, 1991. [PubMed: 1957169]

[2]. Abràmoff Michael D, Garvin Mona K, and Sonka Milan, "Retinal imaging and image analysis," IEEE reviews in biomedical engineering, vol. 3, pp. 169–208, 2010. [PubMed: 22275207]

[3]. Ţălu Simona-Delia, "Optical coherence tomography in the diagnosis and monitoring of retinal diseases," International Scholarly Research Notices, vol. 2013, 2013.

[4]. Sánchez Brea Luisa, Andrade De Jesus Danilo, Shirazi Muhammad Faizan, Pircher Michael, van Walsum Theo, and Klein Stefan, "Review on retrospective procedures to correct retinal motion artefacts in oct imaging," Applied Sciences, vol. 9, no. 13, pp. 2700, 2019.

[5]. Potsaid Benjamin, Gorczynska Iwona, Srinivasan Vivek J, Chen Yueli, Jiang James, Cable Alex, and Fujimoto James G, "Ultrahigh speed spectral/fourier domain oct ophthalmic imaging at 70,000 to 312,500 axial scans per second," Optics express, vol. 16, no. 19, pp. 15149–15169, 2008. [PubMed: 18795054]

[6]. Montuoro Alessio, Wu Jing, Waldstein Sebastian, Gerendas Bianca, Langs Georg, Simader Christian, and Schmidt-Erfurth Ursula, "Motion artefact correction in retinal optical coherence tomography using local symmetry," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2014, pp. 130–137.

[7]. Lauermann JL, Woetzel AK, Treder M, Alnawaiseh M, Clemens CR, Eter N, and Alten Florian, "Prevalences of segmentation errors and motion artifacts in oct-angiography differ among retinal diseases," Graefe's Archive for Clinical and Experimental Ophthalmology, vol. 256, no. 10, pp. 1807–1816, 2018.

[8]. Ferguson R Daniel, Hammer Daniel X, Paunescu Lelia Adelina, Beaton Siobahn, and Schuman Joel S, "Tracking optical coherence tomography," Optics letters, vol. 29, no. 18, pp. 2139–2141, 2004. [PubMed: 15460882]

[9]. Tao Yuankai K, Farsiu Sina, and Izatt Joseph A, "Interlaced spectrally encoded confocal scanning laser ophthalmoscopy and spectral domain optical coherence tomography," Biomedical optics express, vol. 1, no. 2, pp. 431–440, 2010. [PubMed: 21258478]

[10]. Xu Juan, Ishikawa Hiroshi, Wollstein Gadi, Kagemann Larry, and Schuman Joel S, "Alignment of 3-d optical coherence tomography scans to correct eye movement using a particle filtering," IEEE transactions on medical imaging, vol. 31, no. 7, pp. 1337–1345, 2012. [PubMed: 22231171]

[11]. Kraus Martin F, Potsaid Benjamin, Mayer Markus A, Bock Ruediger, Baumann Bernhard, Liu Jonathan J, Hornegger Joachim, and Fujimoto James G, "Motion correction in optical coherence tomography volumes on a per a-scan basis using orthogonal scan patterns," Biomedical optics express, vol. 3, no. 6, pp. 1182–1199, 2012. [PubMed: 22741067]

[12]. Antony Bhavna, Abramoff Michael D, Tang Li, Ramdas Wishal D, Vingerling Johannes R, Jansonius Nomdo M, Lee Kyung-moo, Kwon Young H, Sonka Milan, and Garvin Mona K, "Automated 3-d method for the correction of axial artifacts in spectral-domain optical coherence tomography images," Biomedical optics express, vol. 2, no. 8, pp. 2403–2416, 2011. [PubMed: 21833377]

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

[14]. Baghaie Ahmadreza, Yu Zeyun, and D'Souza Roshan M, "Involuntary eye motion correction in retinal optical coherence tomography: Hardware or software solution?," Medical image analysis, vol. 37, pp. 129–145, 2017. [PubMed: 28208100]

[15]. Chen Yiwei, Hong Young-Joo, Makita Shuichi, and Yasuno Yoshiaki, "Three-dimensional eye motion correction by lissajous scan optical coherence tomography," Biomedical optics express, vol. 8, no. 3, pp. 1783–1802, 2017. [PubMed: 28663866]

[16]. Ksenofontov Sergey Yu, Shilyagin Pavel A, Terpelov Dmitry A, Gelikonov Valentin M, and Gelikonov Grigory V, "Numerical method for axial motion artifact correction in retinal spectral-domain optical coherence tomography," Frontiers of Optoelectronics, pp. 1–9, 2020.

[17]. Kraus Martin F, Liu Jonathan J, Schottenhamml Julia, Chen Chieh-Li, Budai Attila, Branchini Lauren, Ko Tony, Ishikawa Hiroshi, Wollstein Gadi, Schuman Joel, et al. , "Quantitative 3d-oct motion correction with tilt and illumination correction, robust similarity measure and regularization," Biomedical optics express, vol. 5, no. 8, pp. 2591–2613, 2014. [PubMed: 25136488]

[18]. Fu Huazhu, Xu Yanwu, Wong Damon Wing Kee, and Liu Jiang, "Eye movement correction for 3d oct volume by using saliency and center bias constraint," in 2016 IEEE Region 10 Conference (TENCON). IEEE, 2016, pp. 1536–1539.

[19]. Fu Huazhu, Cao Xiaochun, and Tu Zhuowen, "Cluster-based co-saliency detection," IEEE Transactions on Image Processing, vol. 22, no. 10, pp. 3766–3778, 2013. [PubMed: 23629857]

[20]. Estévez Pablo A, Tesmer Michel, Perez Claudio A, and Zurada Jacek M, "Normalized mutual information feature selection," IEEE Transactions on neural networks, vol. 20, no. 2, pp. 189–201, 2009. [PubMed: 19150792]

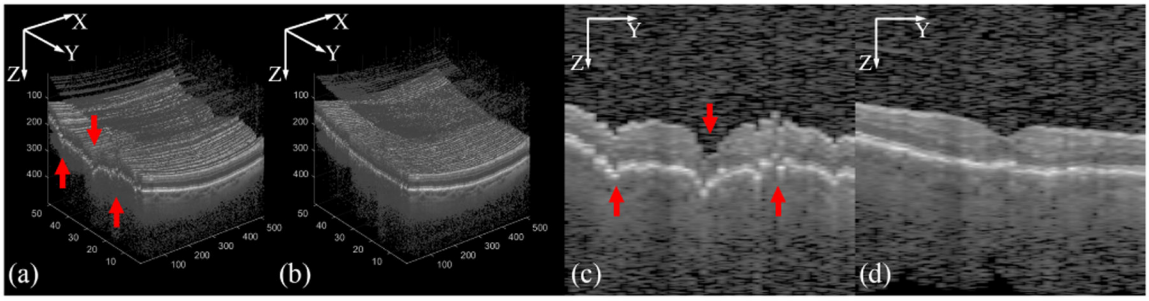


Fig. 1:

(a) 3D OCT volume with axial motion artifacts indicated with red arrows, (b) motion corrected volume by our proposed method, (c) cross-sectional B-scan (YZ plane) with motion artifacts, (d) motion corrected cross-sectional B-scan by our proposed method.

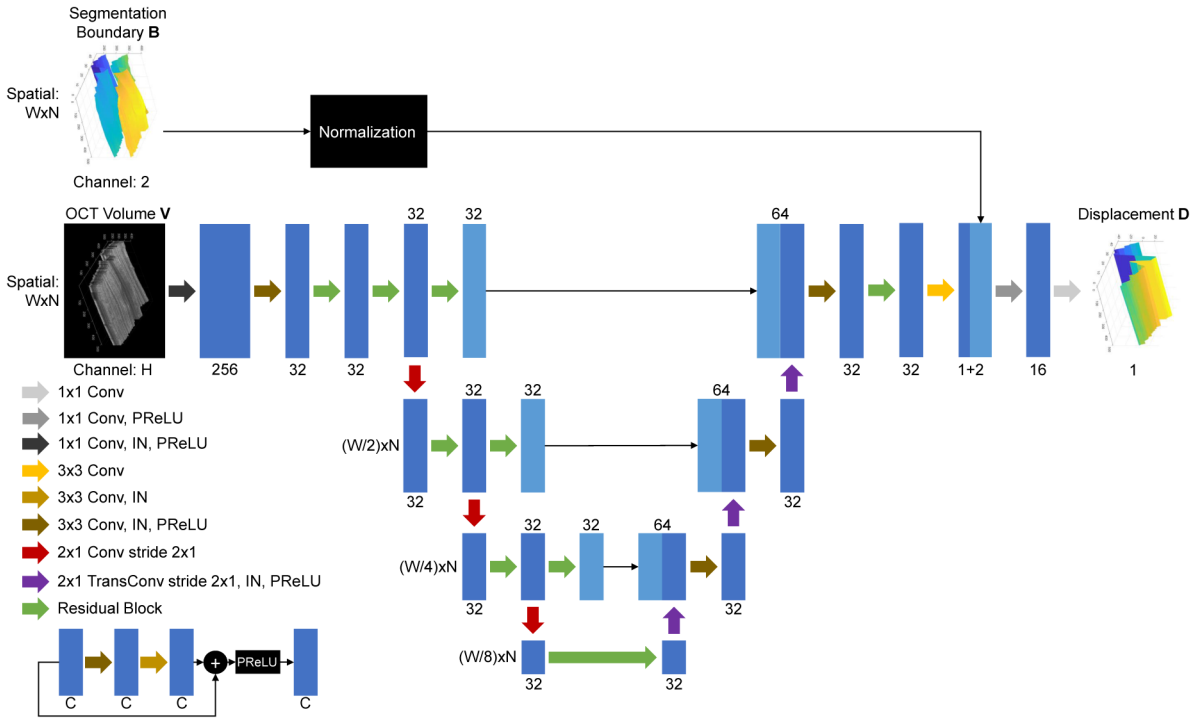


Fig. 2:

Architecture of the proposed OCT motion correction network where a U-Net [13] like structure with residual blocks is adopted. Here IN operation denotes Instance Normalization.

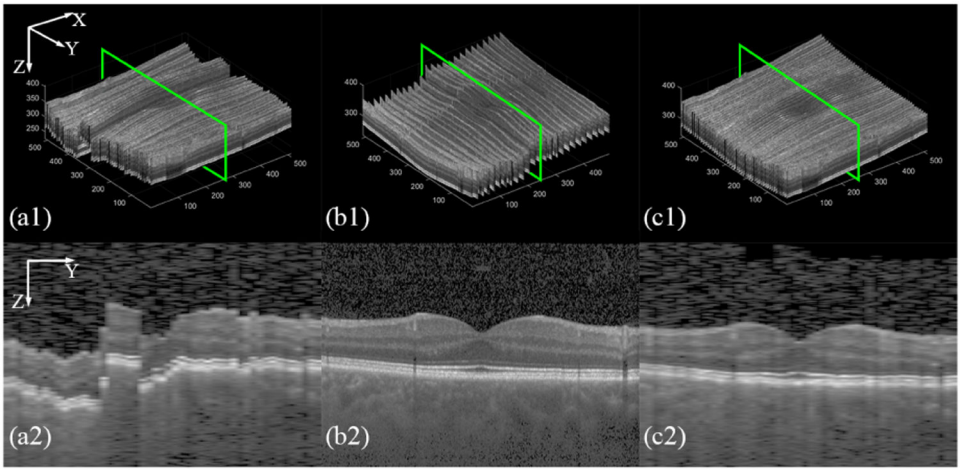
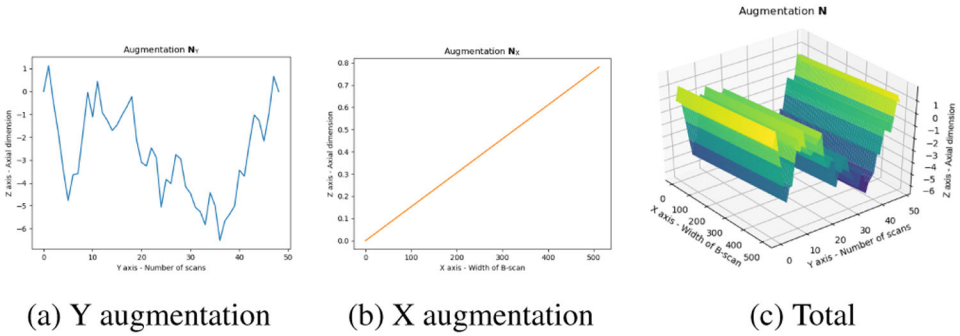


Fig. 3:

Orthogonal method [5] for ground truth acquisition. Column (a) shows the horizontal volume, (b) shows the paired vertical volume, (c) shows the motion-corrected horizontal volume using the motion correction algorithm in [5]. Row (1) shows the 3D volumes, and row (2) shows the cross-sectional B-scan.



(a) Y augmentation (b) X augmentation (c) Total

Fig. 4:

Data augmentation with random displacement. (a) Augmentation for Y dimension $\mathbf{N}_Y$, (b) augmentation for X dimension $\mathbf{N}_X$, (c) Total augmentation $\mathbf{N}$ applied to the input OCT volume.

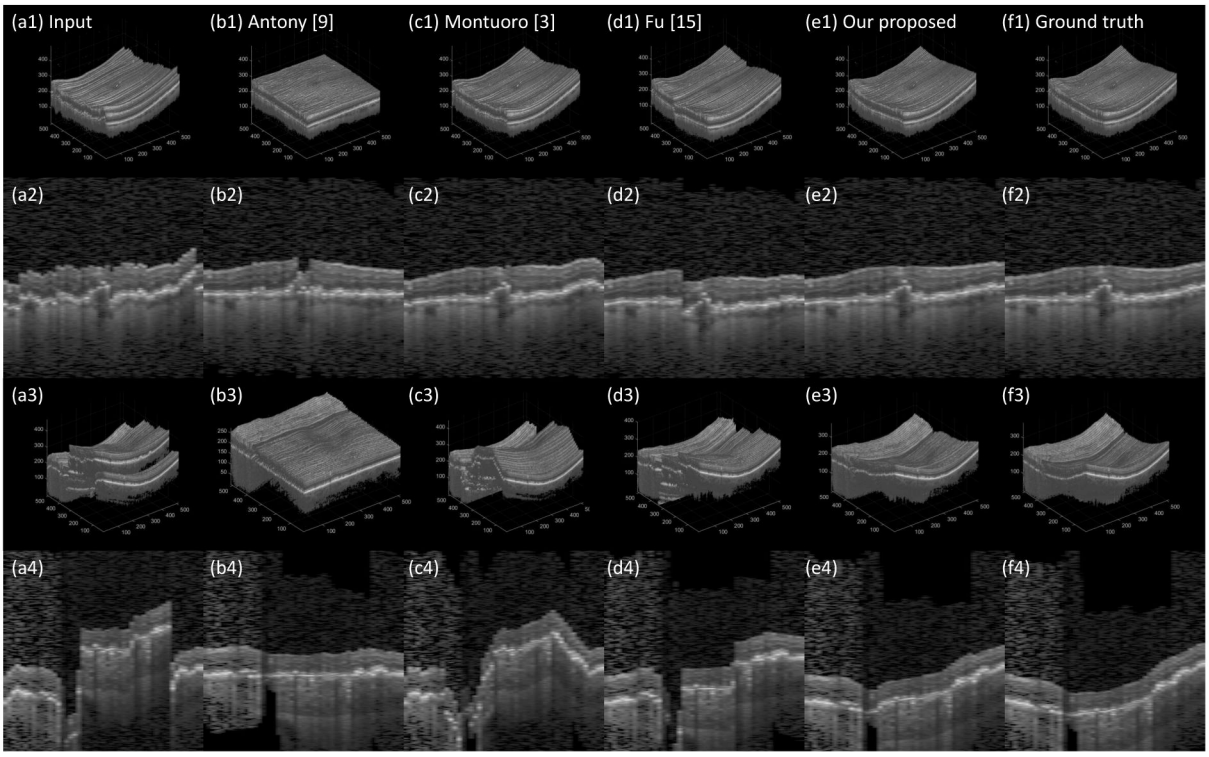


Fig. 5:

Qualitative result of different methods on the test set. Row (1) shows the 3D volume of an example scan with disease, and row (2) shows the cross-sectional B-scans. Row (3) and (4) show another volume with large motion. Gamma correction at 2.2 is applied for display.

Table 1:

Quantitative result of different methods on the test set.

Method	MNMI	MAE
Before correction	0.5811 (±0.0219)	22.39 (±17.52)
Ground truth [5]	0.5901 (±0.0200)	-
Antony et al [12]	**0.5927 (±0.0190)**	28.28 (±10.63)
Montuoro et al [6]	0.5831 (±0.0215)	20.28 (±16.35)
Fu et al [18]	0.5922 (±0.0223)	26.94 (±14.49)
Our proposed	0.5898 (±0.0196)	**7.86 (±5.75)**