IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 July 2023, accepted 11 September 2023, date of publication 18 September 2023,
date of current version 27 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3317011

RESEARCH ARTICLE

Retinal OCT Layer Segmentation via Joint Motion Correction and Graph-Assisted 3D Neural Network

YIQIAN WANG[1], **CARLO GALANG**[2], **WILLIAM R. FREEMAN**[2], **ALEXANDRA WARTER**[2], **ANNA HEINKE**[2], **DIRK-UWE G. BARTSCH**[2], **TRUONG Q. NGUYEN**[1], **(Fellow, IEEE), AND CHEOLHONG AN**[1]

[1]Department of Electrical and Computer Engineering, University of California, San Diego, CA 92093, USA
[2]Jacobs Retina Center, Shiley Eye Institute, University of California, San Diego, CA 92093, USA

Corresponding author: Cheolhong An (chan@eng.ucsd.edu)

This work was supported in part by the National Eye Institute under Grant 1R01EY033847.

ABSTRACT Optical Coherence Tomography (OCT) is a widely used 3D imaging technology in ophthalmology. Segmentation of retinal layers in OCT is important for diagnosis and evaluation of various retinal and systemic diseases. While 2D segmentation algorithms have been developed, they do not fully utilize contextual information and suffer from inconsistency in 3D. We propose neural networks to combine motion correction and segmentation in 3D. The proposed segmentation network utilizes 3D convolution and a novel graph pyramid structure with graph-inspired building blocks. We also collected one of the largest OCT segmentation dataset with manually corrected segmentation covering both normal examples and various diseases. The experimental results on three datasets with multiple instruments and various diseases show the proposed method can achieve improved segmentation accuracy compared with commercial softwares and conventional or deep learning methods in literature. Specifically, the proposed method reduced the average error from 38.47% to 11.43% compared to clinically available commercial software for severe deformations caused by diseases. The diagnosis and evaluation of diseases with large deformation such as DME, wet AMD and CRVO would greatly benefit from the improved accuracy, which impacts tens of millions of patients.

INDEX TERMS Retinal imaging, motion correction, OCT, vessel segmentation, deep learning.

I. INTRODUCTION

Optical Coherence Tomography (OCT) is a 3D imaging technology widely used in ophthalmology. An infrared beam is used to obtain the cross-sections of the retina in vivo at high resolution [1]. The role of OCT imaging is crucial in diagnosing and monitoring both retinal and systemic diseases [2], including age-related macular degeneration (AMD), diabetic macular edema (DME), glaucoma, multiple sclerosis (MS), and so on.

In OCT imaging, the back-scattered intensities of infrared beam represent 1D depth (A-scan, Z axis of Fig. 1). By moving the beam in a raster scanning pattern, a sequence of 2D cross-sectional images (B-scan, XZ plane of Fig. 1) can be acquired. Finally, a 3D OCT volume can be formed by stacking the B-scans (XZ planes) to the Y axis. The *fast scanning axis* refers to the direction where B-scans are acquired (X axis of Fig. 1), and the *slow scanning axis* refers to the direction where B-scans are stacked (Y axis of Fig. 1).

Cross-sectional imaging of OCT is useful for observing the layered structure of the retina, and changes of the retinal layers are critical indicators of both retinal and systemic diseases [3]. For example, thinning of the retinal nerve fiber layer (RNFL) and ganglion cell layer (GCL) is frequently used for assessment of glaucoma. The overall retinal thickness is often used for assessment of DME and choroidal neovascularization (CNV) [2]. It is therefore important to develop an accurate segmentation method for retinal layers to assess these changes automatically. In particular, recent studies reveal that the thickness and vessel density of RNFL

The associate editor coordinating the review of this manuscript and approving it for publication was Essam A. Rashed.

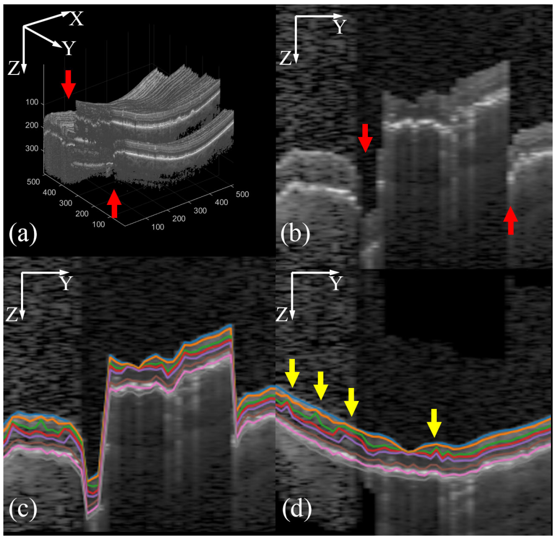


FIGURE 1. **OCT motion artifacts and segmentation. (a) The axial motion artifacts in 3D OCT volume indicated with red arrows, (b) slow B-scan (YZ plane) with motion artifacts, (c) 2D segmented layers with OCT motion artifacts, (d) 2D segmented layers after OCT motion correction, with 3D inconsistency indicated by yellow arrows.**

is related to Alzheimer disease and Parkinson's disease [4], [5], and joint OCT-A vessel density estimation with layer segmentation algorithm could be used to develop a clear and non-invasive tool for early detection of these CNS disorders.

Many OCT layer segmentation approaches have been proposed [6], [7], [8], [9], and commercial OCT systems also provide their own segmentation softwares [10]. However, segmentation error is prevalent with these approaches and compromises the quality of downstream tasks such as OCT-A projections [11]. Recent deep learning segmentation neural networks [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22] lead to significant improvement of accuracy thanks to publicly available annotated datasets [9], [20], [23]. Most networks are modified based on the 2D U-net [24] architecture, which has achieved remarkable performance in numerous image segmentation tasks. However, most deep learning methods are applied on singular 2D B-scan slices. These methods ignore the 3D contextual information within neighboring B-scans, which are especially important for segmenting OCT with diseases. Therefore, the 2D methods are limited in accuracy and their segmentation results lack 3D consistency.

Motion artifact is one of the major reasons that hinder the development of 3D contextual information [25]. Motion artifacts in OCT can be caused by involuntary head motion, respiration, pulsation, or fixational eye movements during the imaging process [25]. These involuntary motions lead to axial and coronal misalignment between neighboring B-scans, shown in sub-figures (b) and (c) in Fig. 1, respectively, where major motion artifacts are indicated by red arrows. The axial motion introduces discontinuities in the slow B-scan as in sub-figure (b), which results in discontinuities in the 2D segmented layers in sub-figure (c). After motion correction, the discontinuities are reduced in sub-figure (d), but the layers lack 3D consistency.

The first segmentation approach utilizing 3D information was proposed by Garvin et al. (OCTExplorer) [26], [27], which applied 3D ‘‘feasibility'' constraints to reduce failures of 2D graph-based approach. Nevertheless, the motion artifacts were removed by flattening the bottom surface of the retina, which also removed the retinal curvature. Besides constraining the 2D segmentation, 3D information can also be used for denoising upon correction of motion artifacts. In the RETOUCH OCT Fluid Detection and Segmentation challenge [28], the winner team [29] performed bounded variation 3D smoothing to reduce speckle noise as pre-processing. However, the 3D information was not fully utilized as their segmentation network was still trained on 2D slices.

The main objective of this paper is to propose a novel method for 3D segmentation of retinal layers in OCT imaging, utilizing neural networks and motion correction. In this paper, we propose a deep learning method that combines motion correction and 3D segmentation. A motion correction neural network first corrects the axial motion artifacts in the input OCT volume, and then a graph-assisted 3D neural network with a novel graph pyramid structure is trained with 3D input and 3D output. We also collected a OCT segmentation dataset with manually corrected segmentation for 1470 B-scan slices covering both normal examples and various diseases. The performance of the proposed method is compared with commercial OCT software solutions, as well as several state-of-the-art methods in literature. Experimental results show that motion correction and 3D contextual information enhance the accuracy of the OCT layer segmentation.

II. RELATED WORK

With the development of OCT imaging systems in the past two decades, many OCT segmentation methods have been proposed. However, segmentation errors with various diseases and motion artifacts are still prevalent with existing segmentation algorithms [11].

Most existing OCT layer segmentation algorithms are 2D image-based, meaning that the segmentation predictions are based on a single B-scan, and applied slice-by-slice for 3D data [6], [7], [8], [9], [30], [31]. Some methods predict the 1D boundaries between each retinal layer, while other methods predict the 2D pixel-wise label. The advantages of predicting 1D boundaries include topology guaranteed layers (i.e. the first boundary will always be above the second boundary) and robustness to outlier regions. However, the 1D boundary method is not able to precisely characterize the layers in retinal disease such as the example illustrated in Fig. 2. Even for human corrected segmentation boundaries, it is not possible to precisely follow the shape of the disease to the one-to-one mapping nature of the 1D boundaries.

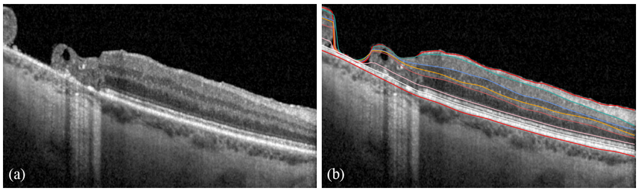


FIGURE 2. Limitations of 1D segmentation boundaries. (a) An example OCT B-scan with retinal disease, (b) human corrected 1D segmentation boundaries. The boundaries could not precisely follow the shape of the disease.

Conventional methods that predict the 1D boundaries include level set methods [6], [7], but these methods take extremely long computational time for up to hours per OCT volume. Graph-based methods are another popular category of algorithm [8], [31], [32], which post-process the pixel-wise prediction from machine learning classifiers [9], [12]. However, these conventional methods are difficult to generalize to various retinal diseases, and require manually established features and extensive parameter tuning.

Thanks to several public datasets with available annotation [9], [20], [23], deep learning-based methods have improved the accuracy of the pixel-wise label prediction via end-to-end training [12], [13], [14], [15], [16], [17], [19], [21], [22]. RelayNet [14] was one of the first deep learning application in retinal layer segmentation. It modified the 2D U-Net architecture, and used the weighted cross-entropy loss to penalize error near each boundary. Other methods combined deep learning classifiers with conventional post-processing to obtain layer boundaries from pixel-wise prediction. Fang et al. [12] combined a patch-based neural network with graph search post-processing to segment 9 layer boundaries. Pekala et al. [19] proposed a dense U-net classifier with post-processing using Gaussian process regression to segment 5 layer boundaries. He et al. [15] proposed to use a second neural network to correct topology based on initial U-Net prediction. Some methods aimed to predict the 1D boundary based on end-to-end regression networks. He et al. [16] proposed cascaded U-Nets to learn the thickness map of each layer achieving topology constraints. The architecture was later improved [17], [18], [22] by multi-task training and including X, Y coordinates as input. One of the state-of-the-art methods was the MGU-Net [20] which combined U-Net with graph-convolution inspired global reasoning blocks [33], achieving the highest Dice coefficient reported on the DME dataset [9].

A major problem that impedes the development of 3D segmentation approaches is that involuntary motion causes misalignment artifacts between neighboring B-scans in 3D OCT imaging. Therefore, motion correction is required to recover the motion-free 3D OCT volume. Some OCT systems integrate eye-tracking hardware to compensate for eye-motion, and there are also post-processing algorithms to correct motion after OCT acquisition [25]. Axial movement is observed to be more significant than coronal movement in magnitude [34], and the higher axial resolution also result in larger axial shift in pixels [34], [35]. Hence, many methods solely focus on correction of axial motion between B-scans [27], [36], [37].

The first OCT segmentation approach to utilize 3D information was proposed by Garvin et al. (OCTExplorer) [26], [27], which could segment 7 boundaries for macular centered OCT scans. The approach first flattened the bottom surface of the retina to remove motion artifacts (along with retinal curvature) and then enhanced 2D graph-based methods by additional 3D "feasibility" constraints. The feasibility constraints took advantage of 3D contextual information and enforced smoothness in neighboring surfaces and surface distance constraints. They demonstrated that their proposed method with 3D information could reduce segmentation failure compared to the 2D graph-based approaches. Besides conventional approaches, DeepMind [13] proposed a 3D segmentation network taking 9 consecutive B-scans that could segment 15 classes of features to aid disease classification. The major limitation was that only two retinal layers could be identified by the segmentation network, namely the neurosensory retina and the RPE. Mukherjee et al. [21] also verified that the 3D neural network architectures outperform their 2D counterparts for OCT segmentation, but their networks were only trained to segment 3 layer boundaries. It is therefore promising to combine motion correction and 3D neural networks in retinal OCT segmentation.

In this paper, we propose a 3D OCT segmentation pipeline that combines two neural networks to correct motion artifacts and perform segmentation based on volumetric data, which enables utilization of 3D contextual information to achieve improved accuracy. Compared to our previous work [38], one of the major innovations is that the N-to-1 sliding window-based 2D neural network is replaced with a N-to-N 3D convolution based graph-assisted neural networks. The proposed approach avoids redundant computation imposed by overlapping windows while improving global reasoning of 3D information. In this paper, we also collected one of the largest OCT segmentation datasets that includes manually corrected segmentation for 1470 B-scan slices covering both normal examples and various diseases with moderate to severe deformations. We also include a more comprehensive analysis on three datasets with different OCT instruments, comparing the proposed method with commercial OCT software solutions, as well as several state-of-the-art methods in literature.

III. PROPOSED METHOD

In this paper, we propose to combine OCT motion correction network with a graph-assisted 3D neural network for retinal layer segmentation. The proposed 3D segmentation pipeline is illustrated in Fig. 3. In the motion correction stage, a 2D segmentation method is first applied to the input OCT volume $\mathbf{V}$ slice-by-slice to obtain the binary segmentation $\mathbf{S}^{\text{bin}}$ for retinal and non-retinal regions. The motion correction network [37] then takes the 3D volume $\mathbf{V}$ and the extracted

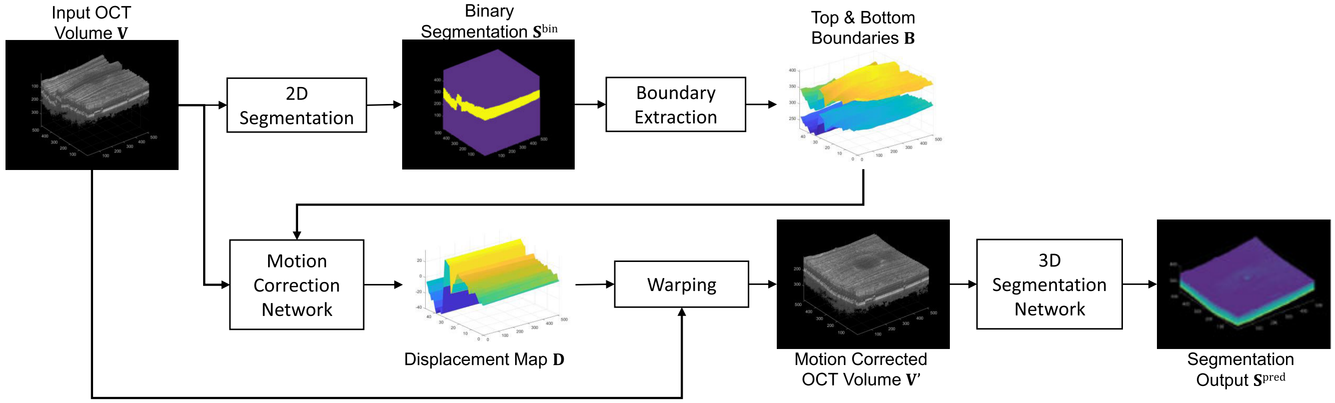

Input OCT Volume V
2D Segmentation
Binary Segmentation $\mathbf{S}^{\text{bin}}$
Boundary Extraction
Top & Bottom Boundaries B
Motion Correction Network
Displacement Map D
Warping
Motion Corrected OCT Volume V′
3D Segmentation Network
Segmentation Output $\mathbf{S}^{\text{pred}}$

FIGURE 3. **Proposed 3D OCT segmentation pipeline with motion correction.**

top (inner limiting membrane, ILM) and bottom (Bruch's membrane, BM) layer boundary as input, to predict a 2D displacement map $\mathbf{D}$ that compensates for axial motion. The original 3D volume $\mathbf{V}$ is warped using the 2D displacement map $\mathbf{D}$ to obtain the motion-corrected volume $\mathbf{V}'$. In the second stage, the specific retinal layers $\mathbf{S}^{\text{pred}}$ are classified using a segmentation network based on the 3D motion-corrected OCT volume $\mathbf{V}'$.

The axial motion correction network [37] takes the raw OCT volume $\mathbf{V} \in \mathbb{R}^{H\times W\times N}$ as input, where H, W, and N denotes the resolution along the Z, X, and Y axes. The BM and ILM boundaries are also included for improved performance. The output of the network is a 2D displacement map $\mathbf{D} \in \mathbb{R}^{W\times N}$, which compensates for Z directional motion in the 3D OCT volume. The architecture of the network is modified based on a residual U-Net, and the performance is verified for various diseases and resolutions [37]. After network prediction, the motion-corrected OCT volume $\mathbf{V}'$ can be obtained by warping the input volume $\mathbf{V}$ based on the predicted axial displacement map $\mathbf{D}$.

$$\mathbf{V}'(z,x,y) = \mathbf{V}(z-\mathbf{D}(x,y),x,y). \tag{1}$$

The main advantage of the graph pyramid structure is the enhanced global reasoning ability. This is achieved by deeper multi-resolution paths and the use of graph reasoning units (GRU). Each GRU include three branches for projection to node space, re-projection to feature space, and fusion of global features, where the two graph convolution blocks are used after projection to the node space. The standard convolution on image data with grid coordinates could be interpreted as a nearest neighbor graph. However, by utilizing the top projection branch of GRU, the input features could be projected to latent space using a learned projection matrix, which enables global reasoning over disjoint and distant areas. After projection to the latent node space, a graph is obtained where each node contains a feature state. Two graph convolutions are performed implemented by channel-wise and node-wise 1D convolutions. Finally, the graph is re-projected from node space to image space using a learned inverse projection matrix.

For comparison of the effect of 3D convolution, we train a 2D version of the segmentation network by removing the $1 \times 1 \times 3$ convolutions and replacing all the 3D operations with their 2D counterparts. We also use the 2D segmentation network to derive the top and bottom layer boundaries for the motion correction network.

The proposed 3D architecture includes 1.940M trainable parameters, and the 2D version includes 1.909M parameters. Both are reduced compared with MGU-Net which has 2.094M parameters, yet the graph pyramid structure can effectively improve the segmentation performance as demonstrated in the experimental result.

The segmentation network is trained using a hybrid loss function, which is a weighted sum of cross-entropy loss and Dice loss. Denoting the last convolution layer output as $\mathbf{x} \in \mathbb{R}^{H\times W\times M}$, the ground truth class label as $\mathbf{y} \in \mathbb{R}^{H\times W\times M}$, and the ground truth one-hot label as $\mathbf{S}^{\text{GT}}$. Note that we also include a valid mask $\mathbf{M} \in \mathbb{R}^{H\times W\times M}$ to exclude regions without annotation, where 1 denotes annotated pixels and 0 denotes otherwise. The masked cross-entropy loss can be expressed as

$$\mathcal{L}_{\text{CE}}(\mathbf{x},\mathbf{y}) = \frac{-1}{\sum_n \mathbf{M}_n}\sum_n\left(\log\frac{\exp \mathbf{x}_{\mathbf{y}_n,n}}{\sum_{k=0}^{K-1}\exp \mathbf{x}_{k,n}}\cdot \mathbf{M}_n\right), \tag{2}$$

where n spans the batch and spatial dimensions. Note that the cross-entropy loss could be implemented with better numeric stability by using the "log-sum-exp" trick, combining log-softmax activation with the negative log likelihood loss for the last convolution layer output $\mathbf{x}$.

The Dice loss is included to regularize the class-imbalance issue of each retinal layer, and emphasize retinal region (class $k = 1$ to $K-2$) over non-retinal regions ($k = 0$ or $K-1$). It is defined as one minus the soft Dice coefficient, which is a score between 0 and 1 characterizing the overlapping ratio between prediction and ground truth. The soft Dice

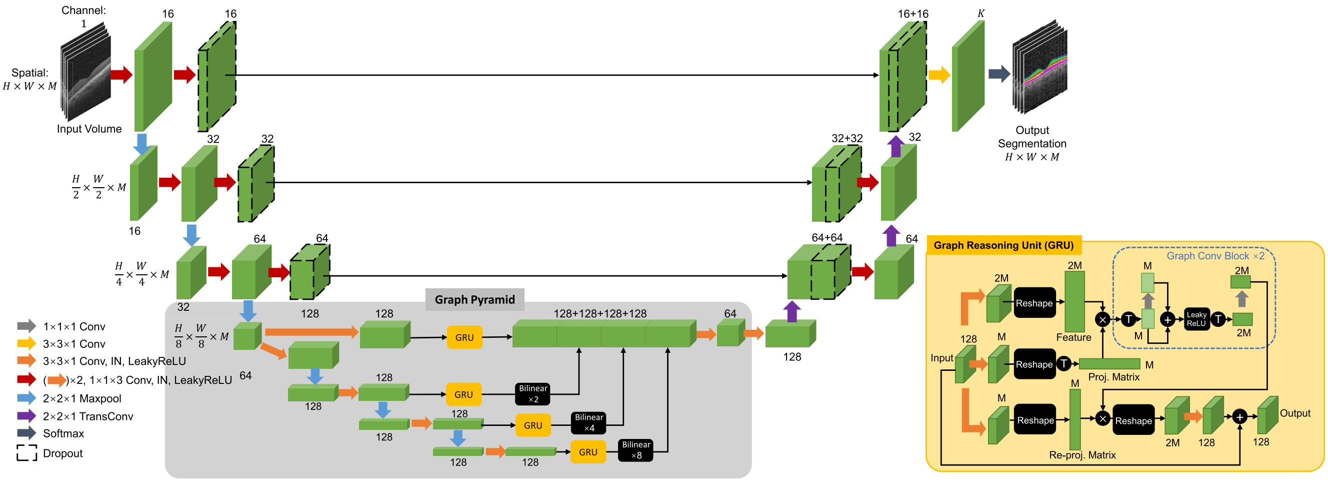


FIGURE 4. Proposed 3D OCT segmentation network with graph pyramid architecture. Here "IN" operation denotes Instance Normalization, "LReLU" denotes LeakyReLU activation, "T" in black circle denotes transpose, and "×" in black circle denotes matrix multiplication.

coefficient can be expressed as

$$\text{SoftDice}_k(\mathbf{x}, \mathbf{S}^{\text{GT}}) = \frac{2\sum_n \sigma(\mathbf{x}_{k,n})\mathbf{S}^{\text{GT}}_{k,n}\mathbf{M}_n}{\sum_n \sigma(\mathbf{x}_{k,n})\mathbf{M}_n + \sum_n \mathbf{S}^{\text{GT}}_{k,n}\mathbf{M}_n}, \tag{3}$$

where $\sigma(\cdot)$ denotes the softmax function along the channel dimension. Then the Dice loss for retinal layers is defined as

$$\mathcal{L}_{\text{Dice}}(\mathbf{x}, \mathbf{S}^{\text{GT}}) = 1 - \underset{k=1:K-2}{\text{mean}} \text{SoftDice}_k(\mathbf{x}, \mathbf{S}^{\text{GT}}). \tag{4}$$

Finally, the total loss is combined with weights $\lambda_{\text{CE}} = 1$, $\lambda_{\text{Dice}} = 2$,

$$\mathcal{L}_{\text{total}} = \lambda_{\text{CE}}\mathcal{L}_{\text{CE}} + \lambda_{\text{Dice}}\mathcal{L}_{\text{Dice}}. \tag{5}$$

Simulated shearing along the X axis is used besides standard data augmentation methods such as random cropping and horizontal flipping, which adds another degree of freedom to the image transformation. The method to generate simulated shearing is described in [37]. The random shearing is generated by an affine transformation with two Gaussian variables $a \sim N(0, \sqrt{2/W})$ and $b \sim N(0, 1)$,

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ a & 0 & 1 & b \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}. \tag{6}$$

Boundary extraction is a optional post-processing to convert the pixel-wise label prediction $\mathbf{S}^{\text{pred}}$ into segmentation boundaries $\mathbf{B}^{\text{pred}}$. The boundaries are detected, joined, and interpolated using the pseudo-code in Algorithm 1.

Algorithm 1 Pixel-Wise Label to Boundary

```
Pixel-wise label $\mathbf{S}^{\text{pred}} \in \mathbb{R}^{H \times W}$
K+1 boundaries $\mathbf{B}^{\text{pred}} \in \mathbb{R}^{(K+1) \times W}$
for $k = 0 : K$ do
    $k$-th edge $\mathbf{C}$ ($W$ lists)
    for $x = 0 : W - 1$ do
        $\mathbf{C}_{(x)} \leftarrow z$ s.t. $\mathbf{S}^{\text{pred}}_{(z,x)} = k + 1$ & $\mathbf{S}^{\text{pred}}_{(z-1,x)} \neq k + 1$
        if $x = 0$ then
            $\mathbf{B}^{\text{pred}}_{(k,x)} \leftarrow \min \mathbf{C}_{(x)}$
            $b_{\text{prev}} \leftarrow \mathbf{B}^{\text{pred}}_{(k,x)}$
        else if $\mathbf{C}_{(x)}$ is not empty then
            $\mathbf{B}^{\text{pred}}_{(k,x)} \leftarrow \arg\min_{\mathbf{C}_{(x)}} |\mathbf{C}_{(x)} - b_{\text{prev}}|$
            $b_{\text{prev}} \leftarrow \mathbf{B}^{\text{pred}}_{(k,x)}$
        end if
    end for
    Interpolate $\mathbf{B}^{\text{pred}}_{(k)}$ for missing $x$ values using b-spline
end for
```

IV. EXPERIMENTAL RESULT

In the experiment, we test and compare the segmentation performance of our proposed joint motion correction and 3D segmentation neural networks with several state-of-the-art conventional or deep learning methods.

A. DATASETS

The methods are evaluated on three different datasets: DME dataset [9], AMD and Control dataset [23], and our own dataset collected by Jacobs Retina Center (JRC). The comparison of the three dataset is summarized in Table 1.

We use the DME dataset [9] as a benchmark of the proposed method to compare with the state of the art methods in literature. The DME dataset [9] is one of the most widely used public datasets in literature [14], [17], [20]. The dataset includes 10 macular centered OCT volumes for patients with DME imaged by Heidelberg Spectralis OCT system after motion correction. The resolution for each volume is $496 \times 768 \times 61$, with voxel size ranging from $3.87 \times 11.07 \times 118\mu$m to $3.87 \times 11.59 \times 128\mu$m. 11 selected B-scans out of 61 B-scans in each volume (in total 110 B-scans)

TABLE 1. Dataset information.

Dataset	DME [5]	AMD & Control [10]	JRC
Source	Public, license unknown	Public, license unknown	Private, JRC IRB No. 120516
Number	10 OCT volumes (110 B-scans annotated)	269 AMD, 115 normal OCT volumes	190 OCT volumes (30 volumes annotated)
Resolution	512×768×61	512×1000×100	496×512×49
Equipment	Heidelberg Spectralis	Bioptigen	Heidelberg Spectralis
Motion	Corrected	Not corrected	Not corrected
Annotation	8 boundaries, sparse, center area, 2 graders	3 boundaries, dense, center area, graders number unknown	8 boundaries, dense, full area, 6 graders revised by 1 grader
Disease	DME	Normal, AMD	Normal, wet and dry AMD, DR, ERM, CRVO, retinal detachment, macular hole, chorioretinopathy

are manually annotated with 8 segmentation boundaries ($K = 9$ classes) in the central region. We follow the training and test division in other papers [14], [17], [20], where the first 55 images from subject 1 to 5 are used for training, and the last 55 images from subject 6 to 10 are used for testing.

We then use the AMD and control dataset [23] to evaluate the influence of the motion correction network in 3D segmentation on OCT volumes with real motion artifacts. The AMD and control dataset [23] is a public dataset with 384 macular centered OCT volumes from 269 patients with AMD and 115 normal control subjects. The OCT volumes are imaged by the Bioptigen system and mostly has resolution 512 × 1000 × 100, with some exceptions that has 82 B-scans. Manual annotations for 3 layer boundaries ($K = 4$ classes) are provided in a central circular region. Due to the different definition of the RPE-DC layer in AMD group [23], we only utilized normal control group for evaluation of the segmentation methods. The first 55 OCT volumes are used for training, and next 5 volumes are used for validation, and the last 55 volumes are used for testing.

Finally, the JRC dataset is used to compare the proposed method to OCT segmentation solutions clinically available to ophthalmologists in segmentation of retinal layers with various diseases. The JRC dataset contains 190 horizontal and vertical OCT volumes imaged with the Heidelberg Spectralis system [10], and 8 layers for 30 OCT volumes are manually corrected using the Heidelberg HEYEX software based on Heidelberg's segmentation result. The 6 graders are trained retinal MD fellows at the Jacobs Retina Center, and the annotations is revised by one grader. The OCT volumes without manual corrections are divided into 142 and 18 for training and validation using Heidelberg's segmentation as ground truth, and the 30 manually corrected volumes are divided into 15 and 15 for fine-tuning and testing. The resolution of the OCT volumes are 496 × 512 × 49 with size 1.9 × 5.8 × 5.8 mm^3. The dataset includes both normal subjects and patients with wet and dry AMD, nonproliferative diabetic retinopathy (NPDR), epi-retinal membrane (ERM), central retinal vein occlusion (CRVO), retinal detachment, macular hole, chorioretinopathy, and so on. The pathology and diagnosis are recorded for each OCT volume by the ophthalmologists at JRC.

B. SIMULATED MOTION FOR DME DATASET

For the AMD and control dataset and JRC dataset, we directly apply our motion correction approach on the original motion corrupted OCT volumes. Since the DME dataset [9] has been motion-corrected, we include simulated motion on the input OCT volumes to test the performance of our proposed motion correction approach. The simulated axial eye motion is generated using a similar method in [37], which is based on cumulative sum of Gaussian vector. We also verify the similarity of simulated motion and real eye motion by comparing their statistics in the AMD and control dataset and JRC dataset. Fig. 5 sub-figure (a) shows the histogram of motion amplitudes of the real and simulated motion vectors, and it can be observed that the real and simulated motion amplitudes follow a similar distribution. Fig. 5 sub-figure (b)-(e) respectively shows the normalized auto-correlation of 10 example motion vectors in the AMD and control dataset, JRC dataset, simulated motion, and Gaussian random vectors. The auto-correlation of real motion on both datasets are significantly different from Gaussian random vectors, and the auto-correlation of simulated motion resembles that of real motion.

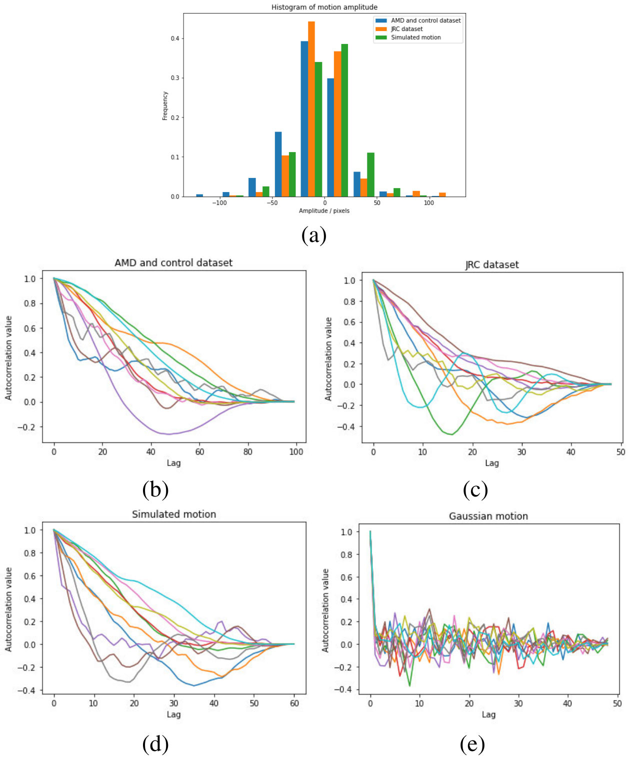
C. EVALUATION METRICS

The classification error and the Dice loss are used to evaluate the pixel-wise performance of each segmentation algorithm. Specifically, we present the overall error, the error of retinal layers, the averaged Dice loss for all layers, and the Dice loss for each layer. Denoting the predicted binary segmentation map with $\mathbf{S}^{\text{pred}}$, ground truth segmentation with $\mathbf{S}^{\text{GT}}$, valid mask with $\mathbf{M}$, and element-wise product with $\odot$. The Dice loss for the kth layer can be obtained by one minus the Dice coefficient

$$\mathcal{L}_{\text{Dice},k}(\mathbf{S}^{\text{pred}}, \mathbf{S}^{\text{GT}}) = 1 - \frac{2\sum \mathbf{S}_k^{\text{pred}} \odot \mathbf{S}_k^{\text{GT}} \odot \mathbf{M}}{\sum \mathbf{S}_k^{\text{pred}} \odot \mathbf{M} + \sum \mathbf{S}_k^{\text{GT}} \odot \mathbf{M}}, \tag{7}$$

and the averaged Dice loss for all retinal layers is

$$\mathcal{L}_{\text{Dice}}(\mathbf{S}^{\text{pred}}, \mathbf{S}^{\text{GT}}) = \underset{k=1:K-2}{\text{mean}} \mathcal{L}_{\text{Dice},k}(\mathbf{S}^{\text{pred}}, \mathbf{S}^{\text{GT}}). \tag{8}$$



(a)

(b) (c)

(d) (e)

FIGURE 5. Statistics of real and simulated eye motion. (a) Histogram of motion amplitudes, (b)-(e) auto-correlation of 10 example motion vectors in the AMD and control dataset, JRC dataset, simulated motion, and Gausian random vectors.

The pixel-wise error is calculated on the valid region given by $\mathbf{M}$, and non-retinal regions ($k = 0$ or $K - 1$) are merged into one class. The layer error is calculated based on retinal layers corresponding to class 1 to $K - 2$ in the ground truth label.

$$\text{Error}(\mathbf{S}^{\text{pred}}, \mathbf{S}^{\text{GT}}) = \frac{\sum(\mathbf{S}^{\text{pred}} \neq \mathbf{S}^{\text{GT}}) \odot \mathbf{M}}{\sum \mathbf{M}}. \tag{9}$$

After converting the pixel-wise predictions into layer boundaries using the proposed Algorithm 1, the mean absolute distance (MAD) is evaluated between the predicted and ground truth boundaries in the annotated region masked by $\mathbf{M}^{\text{b}} \in \mathbb{R}^{K \times W \times N}$,

$$\text{MAD}(\mathbf{B}^{\text{pred}}, \mathbf{B}^{\text{GT}}) = \frac{\sum |\mathbf{B}^{\text{pred}} - \mathbf{B}^{\text{GT}}| \odot \mathbf{M}^{\text{b}}}{\sum \mathbf{M}^{\text{b}}}. \tag{10}$$

D. IMPLEMENTATION

In the experiment, we compare the performance of our proposed 3D segmentation with 7 B-scans input (center ±3 neighboring B-scans) and the 2D version using a single B-scan input with or without OCT motion correction network. We compare with several conventional methods by Chiu et al. [9] and Rathke et al. [31], the OCTExplorer software [27], as well as deep learning method U-Net [24], RelayNet [14], MGU-Net [20], and the network proposed by He et al. [17].

Our proposed motion correction and segmentation networks are implemented in PyTorch. The motion correction network utilizes the pre-trained model in [37]. On the DME dataset, the 2D network is first trained on the first 55 images with expert 1's annotation as ground truth, using batch size 4 for 200 epochs with an initial learning rate of 10^{-3} and decayed to 10^{-4} after 100 epochs, using Adam optimizer with weight decay of 10^{-4}. Since the dataset is sparsely annnotated, we use the prediction of the 2D network as pseudo-ground truth for B-scan slices without manual annotations to obtain dense label for the 3D OCT volume. The 3D segmentation network is then trained based on 3D labels with batch size 1. On the AMD and control dataset, the segmentation networks are trained using batch size 4 for 15 epochs with an initial learning rate of 10^{-3} and decayed to 10^{-4} after 10 epochs. We first pre-train the model on the JRC training set using Heidelberg segmentation as ground truth for 15 epochs with initial learning rate 10^{-3} and weight decay of 10^{-4}, and then fine-tune on 15 OCT volumes with manual labels for 20 epochs with learning rate 5×10^{-4} and 10 epochs with learning rate 10^{-4} using weight decay of 10^{-3}.

The method by Chiu et al. [9] uses the predicted layer boundaries provided in the DME dataset, and the method by Rathke et al. [31] uses the original implementation in Matlab. The OCTExplorer [27] software version 3.8.0 is used. We include both the pre-trained PyTorch model on the DME dataset provided by the original authors of RelayNet [14], and also include our retrained model on all three datasets. The network by He et el. [17] uses the prediction results on the DME dataset provided by the authors. The U-Net [24] and MGU-Net [20] are trained in PyTorch using similar hyper parameters as our proposed segmentation network.

E. DME DATASET

The qualitative results of segmentation on the original motion-corrected DME dataset are shown in Fig. 6, where the first group (a) shows the pixel-wise segmentation and group (b) shows the layer boundaries. Sub-figure (1) shows the input B-scan cropped in the labeled retinal region, sub-figure (2) shows the ground truth manual label, and sub-figures (3) to (10) show the segmentation result of different segmentation methods. The results demonstrate that the conventional methods by OCTExplorer [27] and Rathke et al. [31] could not accurately segment the retinal layers for the B-scan with DME. The method by Chiu et al. [9] produces more accurate segmentation, while the RNFL (in blue) is thinner than the ground truth, and the OPL (in yellow) does not follow the shape of lesions in the B-scan. For deep learning methods, the boundaries between each class of the RelayNet [14] prediction is noisier compared with the MGU-Net [20] and our proposed networks. MGU-Net [20] yields mis-classification of the OPL, and He et al. [17] yields discontinuities denoted by yellow arrows. The segmentation result of our proposed networks with 2D or 3D input are both visually continuous, and our 3D network result in lower Dice loss and MAD.

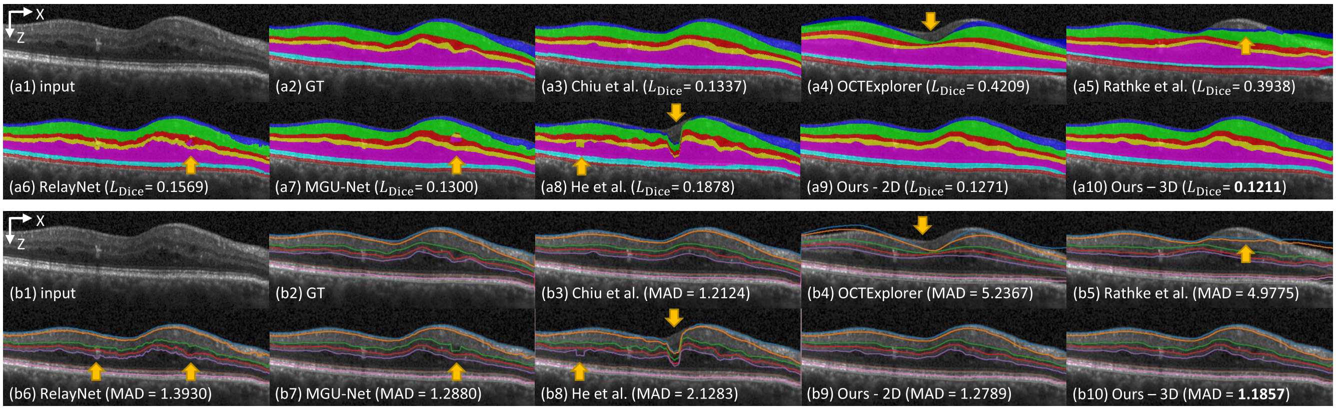

X
Z
(a1) input
(a2) GT
(a3) Chiu et al. (L_{Dice}= 0.1337)
(a4) OCTExplorer (L_{Dice}= 0.4209)
(a5) Rathke et al. (L_{Dice}= 0.3938)
(a6) RelayNet (L_{Dice}= 0.1569)
(a7) MGU-Net (L_{Dice}= 0.1300)
(a8) He et al. (L_{Dice}= 0.1878)
(a9) Ours - 2D (L_{Dice}= 0.1271)
(a10) Ours – 3D (L_{Dice}= **0.1211**)
X
Z
(b1) input
(b2) GT
(b3) Chiu et al. (MAD = 1.2124)
(b4) OCTExplorer (MAD = 5.2367)
(b5) Rathke et al. (MAD = 4.9775)
(b6) RelayNet (MAD = 1.3930)
(b7) MGU-Net (MAD = 1.2880)
(b8) He et al. (MAD = 2.1283)
(b9) Ours - 2D (MAD = 1.2789)
(b10) Ours – 3D (MAD = **1.1857**)

FIGURE 6. Qualitative results on the DME dataset [9]. Group (a) shows the pixel-wise prediction of each method with corresponding Dice loss, and group (b) shows the layer boundaries with mean absolute distance (MAD). (1) Input B-scan, (2) ground truth segmentation, (3) Chiu et al. [9], (4) Rathke et al. [31], (5) OCTExplorer [27], (6) RelayNet [14], (7) MGU-Net [20], (8) He et al. [17], (9) our proposed 2D network, (10) our proposed 3D network. Yellow arrows denote large segmentation errors.

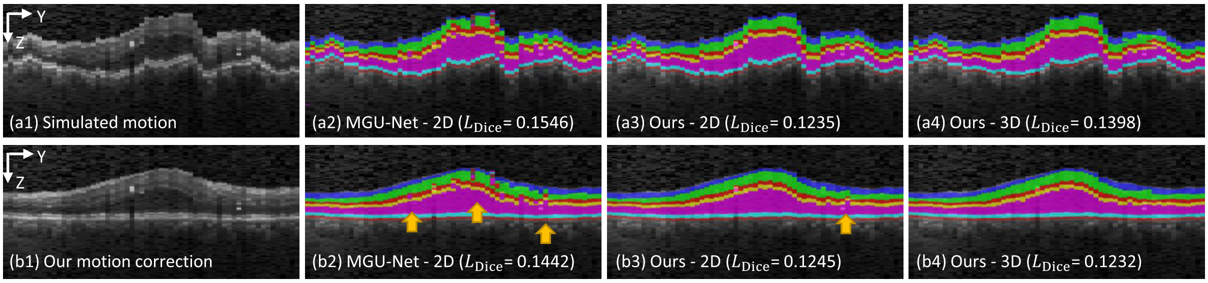

Y
Z
(a1) Simulated motion
(a2) MGU-Net - 2D (L_{Dice}= 0.1546)
(a3) Ours - 2D (L_{Dice}= 0.1235)
(a4) Ours - 3D (L_{Dice}= 0.1398)
Y
Z
(b1) Our motion correction
(b2) MGU-Net - 2D (L_{Dice}= 0.1442)
(b3) Ours - 2D (L_{Dice}= 0.1245)
(b4) Ours - 3D (L_{Dice}= 0.1232)

FIGURE 7. Qualitative comparison of 3D consistency on the DME dataset [9]. Group (a) shows results for OCT with simulate motion, group (b) shows results for motion-corrected OCT. (1) slow B-scan, (2)-(4) segmentation result of MGU-Net [20] and our proposed 2D or 3D network. Yellow arrows denote large segmentation errors.

TABLE 2. Comparison of pixel-wise label of different segmentation methods on the DME test dataset [9], where the best and the second best are denoted by bold text and underlined text, respectively.

					Dice Loss per Layer						
Data	Method	Error	Layer Error	Mean Dice Loss	RNFL	GCL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE
Original	Chiu et al. [9]	2.55%	13.46%	0.1616 (±0.063)	0.1490	0.1059	0.2439	0.2510	0.0694	0.1317	0.1805
	OCTExplorer [27]	9.64%	47.63%	0.5515 (±0.139)	0.5853	0.4200	0.6266	0.6625	0.2698	0.6453	0.6509
	Rathke et al. [31]	5.42%	29.82%	0.3279 (±0.090)	0.2935	0.2518	0.4101	0.4368	0.1687	0.4051	0.3290
	U-Net [24]	2.25%	10.63%	0.1360 (±0.062)	0.1131	0.0865	0.2441	0.2050	0.0585	0.1036	0.1414
	RelayNet [14]	3.18%	16.86%	0.1997 (±0.067)	0.2203	0.1412	0.2854	0.2689	0.0809	0.1732	0.2279
	RelayNet [14] (re-trained)	2.46%	12.61%	0.1441 (±0.055)	0.1295	0.1081	0.2248	0.2298	0.0823	0.1043	0.1301
	MGU-Net [20]	2.13%	11.46%	0.1359 (±0.054)	0.1356	0.1038	0.2118	0.2152	0.0565	0.1032	0.1249
	He et al. [17]	2.18%	10.88%	0.1393 (±0.058)	**0.1116**	0.0792	0.1855	0.2240	0.0580	0.1245	0.1924
	Ours (2D)	1.95%	10.26%	0.1238 (±0.048)	0.1176	0.0864	0.1850	0.1946	0.0526	0.1004	0.1296
	Ours (3D)	**1.80%**	**9.75%**	**0.1155** (±0.045)	0.1137	**0.0732**	**0.1665**	**0.1847**	**0.0485**	**0.0998**	**0.1218**
Simulated Motion	Ours (2D)	1.92%	10.06%	0.1216 (±0.048)	0.1165	0.0860	0.1855	0.1931	0.0513	**0.0984**	**0.1205**
	Ours (3D)	2.21%	12.04%	0.1401 (±0.048)	0.1370	0.0917	0.1830	0.2004	0.0578	0.1348	0.1760
	Ours (2D, motion corrected)	1.92%	10.15%	0.1219 (±0.048)	**0.1158**	0.0855	0.1848	0.1947	0.0525	0.0990	0.1211
	Ours (3D, motion corrected)	**1.85%**	**10.03%**	**0.1189** (±0.045)	0.1165	**0.0770**	**0.1727**	**0.1855**	**0.0495**	0.1069	0.1239

We also visualize the central slow B-scans in Fig. 7 to compare the 3D consistency of the MGU-Net and our proposed networks after applying simulated motion and our motion correction network. It can be observed that the simulated motion in sub-figure (a1) causes axial distortion to the slow B-scan in sub-figure (a1), and it can be effectively corrected by our motion correction network in sub-figure (b1). Overall, our proposed 3D network in sub-figure (b4) after motion correction achieves the best consistency, compare to MGU-Net [20] in sub-figures (b2) and our 2D network in sub-figure (b3) at the location denoted by yellow arrow. It demonstrates the joint motion correction and 3D segmentation networks can improve the performance of 3D consistency.

TABLE 3. Comparison of segmentation boundaries of different segmentation methods on the DME test dataset [9], where the best and the second best are denoted by bold text and underlined text, respectively.

			Mean Absolute Distance (MAD) per Layer							
Data	Method	MAD	ILM	RNFL	GCL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE
Original	Chiu et al. [9]	1.5723 (±0.475)	1.3146	1.6623	1.8948	2.1675	2.3025	0.9982	1.1213	1.1175
	OCTExplorer [27]	7.8175 (±0.705)	8.8859	8.4021	8.2109	7.8384	8.1318	6.7051	7.4995	6.8663
	Rathke et al. [31]	4.6272 (±1.381)	4.5134	5.7654	5.3873	5.9651	5.7223	2.3398	4.9436	2.3808
	U-Net [24]	1.7796 (±0.432)	2.0935	1.7927	2.2347	1.9086	2.3776	1.0859	1.3869	1.3569
	RelayNet [14]	2.0903 (±0.542)	1.6345	2.8419	2.4331	2.5968	2.5939	1.3682	1.6287	1.6250
	RelayNet [14] (re-trained)	2.0310 (±0.459)	1.7375	2.3206	2.4817	2.3928	2.6642	1.3617	1.5098	1.7797
	MGU-Net [20]	1.4934 (±0.440)	1.5402	1.9179	1.8086	1.8004	2.0251	0.8981	1.0423	0.9149
	He et al. [17]	1.3190 (±0.351)	**1.0348**	1.4022	1.3791	1.7694	1.9076	0.7938	1.1446	1.1209
	Ours (2D)	1.7860 (±0.283)	1.7129	2.0226	1.8373	2.0534	2.2302	1.3568	1.5514	1.5232
	Ours (3D)	**1.1624** (±0.307)	1.4368	**1.2911**	**1.2020**	**1.3595**	**1.6210**	**0.7439**	**0.8648**	**0.7799**
Simulated Motion	Ours (2D)	1.8274 (±0.297)	1.9250	2.0946	1.8363	2.0508	2.2693	1.3620	1.5491	1.5324
	Ours (3D)	1.6090 (±0.263)	1.3448	1.7890	1.5729	1.8232	2.0311	1.1383	1.6132	1.5595
	Ours (2D, motion corrected)	1.9163 (±0.301)	1.8915	2.1713	1.9412	2.1898	2.3955	1.4781	1.6505	1.6123
	Ours (3D, motion corrected)	**1.2040** (±0.289)	**1.3398**	**1.3465**	**1.2598**	**1.4134**	**1.6813**	**0.8154**	**0.9264**	**0.8498**

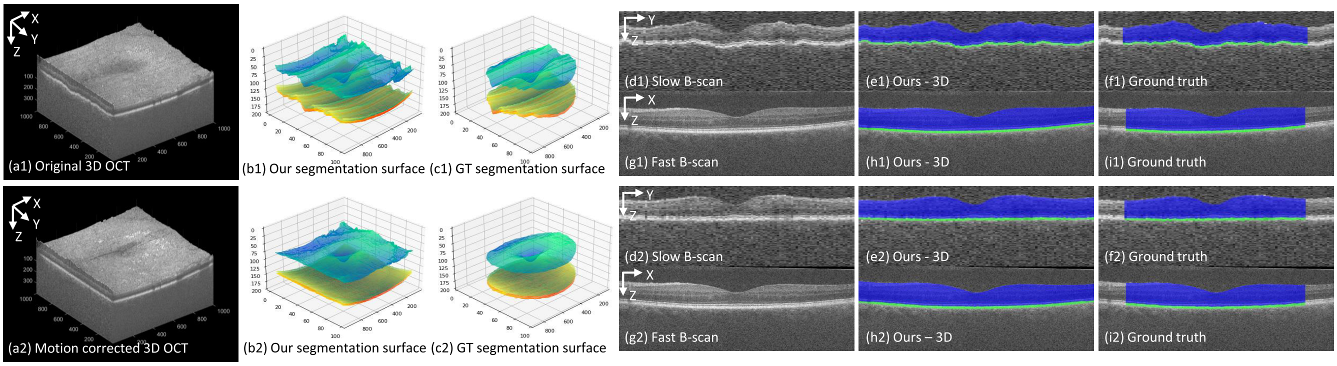


FIGURE 8. Qualitative results on the AMD and control dataset. Group 1 shows segmentation on the original OCT, group 2 shows segmentation on the motion-corrected OCT. (a) 3D OCT volume, (b) our segmentation surface, (c) ground truth (partially annotated) segmentation surface, (d)-(f) segmentation of the slow B-scan, (g)-(i) segmentation of the fast B-scan.

Quantitative evaluation for pixel-wise accuracy on the DME dataset [9] is shown in Table 2. The raw network output without boundary detection post-processing described in Algorithm 1 is used in this evaluation for deep learning methods, and the pixel-wise labels are derived for conventional methods using the predicted boundaries. The best performance in each column for original OCT and OCT with simulated motion is denoted by bold text, and the second best is denoted by blue text. We only include our proposed methods in the experiment with simulated motion in order to evaluate the performance of our motion correction network. On the original input, our proposed 3D segmentation network achieves the lowest error of 1.80%, layer accuracy of 9.75%, and average Dice loss of 0.1155. The Dice loss of our 3D network is also the lowest in the each retinal layer, except for the RNFL layer where it ranks as the third best result. For OCT with simulated motion, our proposed 3D segmentation network after motion correction achieves the lowest error and Dice coefficient, and note that the error of the 3D network increases without motion correction.

The mean average distance of layer boundaries are evaluated in Table 3. The proposed boundary detection algorithm is used to post-process the deep learning methods. Overall, the proposed 3D approach achieves the lowest average MAD at 1.1624 pixels, and the result by He et al. [17] achieves the second lowest average MAD at 1.3190 pixels. When comparing the the results on OCT with simulated motion, our 3D segmentation network with motion correction also achieves the lowest MAD with a improvement upon our 2D network and our 3D network without motion correction.

TABLE 4. Quantitative result of different methods on the AMD and control test dataset [23], where the best and the second best are denoted by bold and underlined text, respectively.

			Dice Loss per Layer	
Method	Error	Layer Error	RNFL-OS	RPE
Rathke et al. [31]	1.21%	5.50%	0.0292	0.3078
U-Net [24]	0.83%	2.31%	0.0135	0.1039
RelayNet [14]	1.87%	2.86%	0.0172	0.1126
MGU-Net [20]	0.56%	2.03%	0.0111	0.1061
Ours (2D, original)	0.49%	1.82%	0.0093	0.1022
Ours (3D, original)	0.48%	1.86%	0.0092	0.0990
Ours (2D, motion corrected)	0.44%	1.61%	0.0083	0.0944
Ours (3D, motion corrected)	**0.41%**	**1.56%**	**0.0074**	**0.0892**

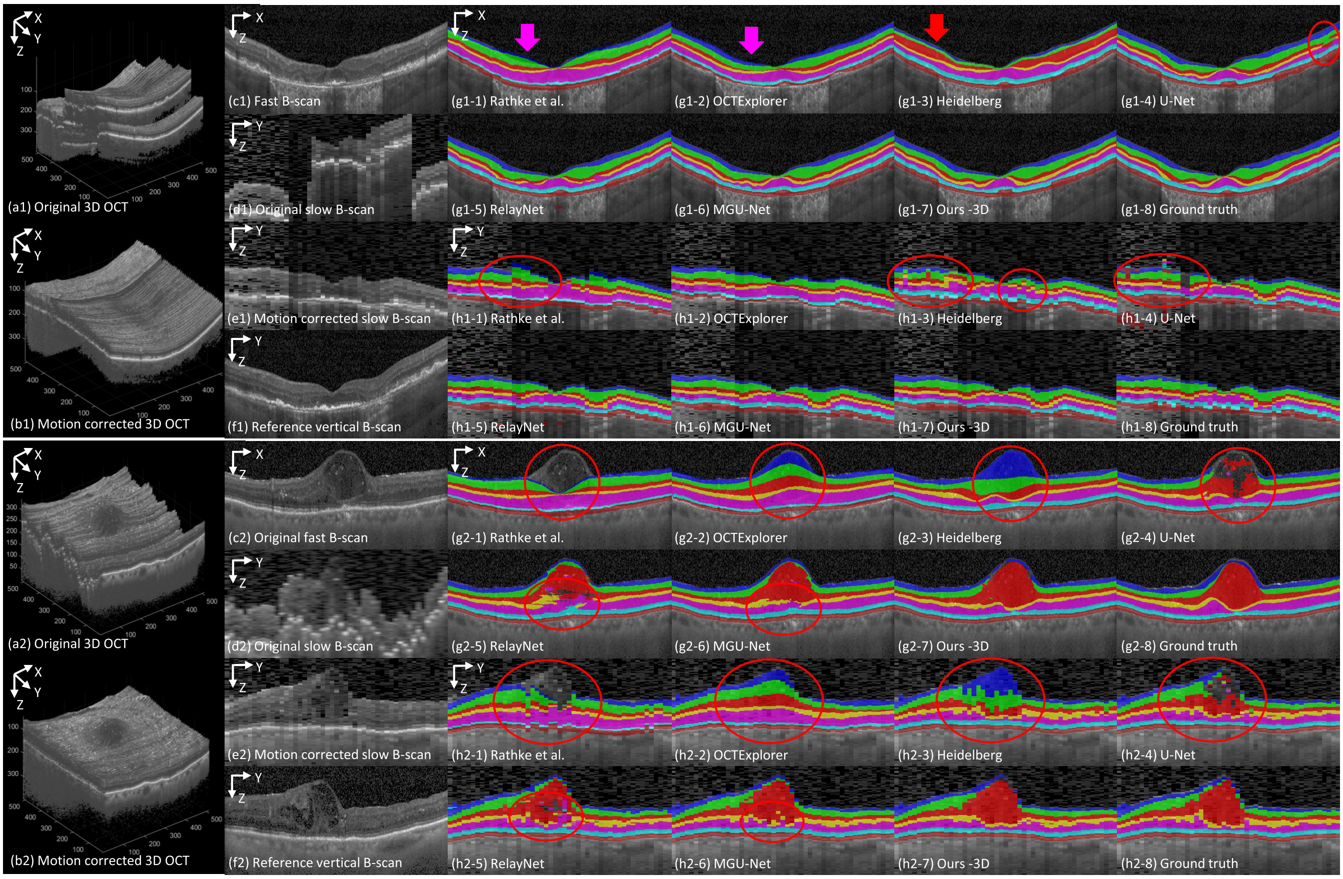


FIGURE 9. Qualitative results on the JRC dataset. Group (1) and (2) show two examples. (a) Original 3D OCT volume, (b) motion corrected 3D OCT volume, (c) fast B-scan, (d) original slow B-scan, (e) motion-corrected slow B-scan, (f) reference vertical B-scan, (g) segmentation on the fast B-scan using different methods, (h) segmentation on the slow B-scan using different methods. Pink arrows denote large error in OPL, red arrow denotes large error in GCL-IPL, and red circle denotes large segmentation error.

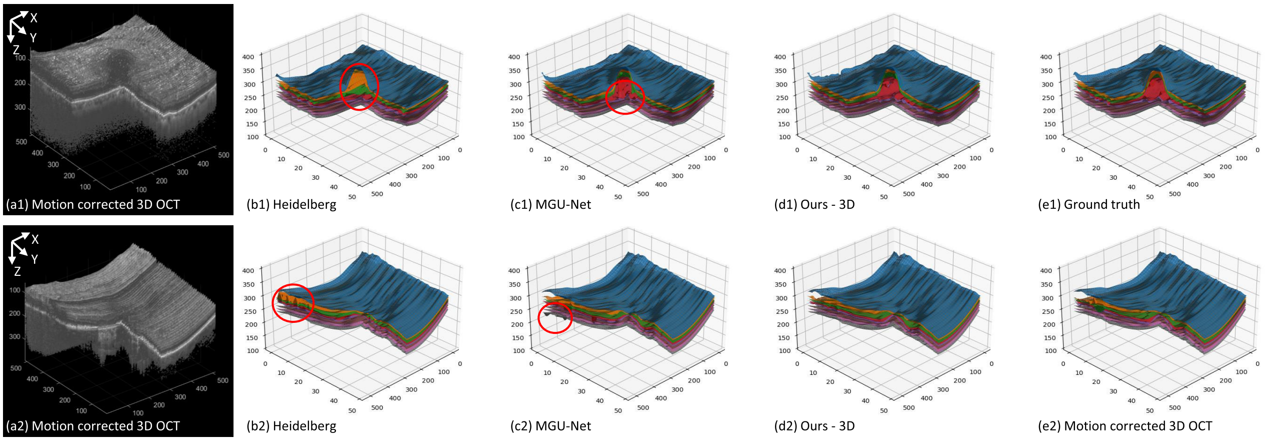


FIGURE 10. Visualization of segmentation result in 3D on the JRC dataset. Group (1) and (2) show two examples. (a) Motion corrected 3D OCT volume, (b)-(e) segmentation surfaces of Heidelberg, MGU-Net, our 3D network, and manual annotated ground truth. Red circle denotes large segmentation error.

F. AMD AND CONTROL DATASET

We use the AMD and control dataset to evaluate the influence of the motion correction network on real motion corrupted OCT volumes. We visualize one example OCT volume in Fig. 8, where segmentation on the original 3D OCT volume is shown in group (1), and segmentation on the motion-corrected OCT volume is shown in group (2). The 3D OCT is shown in sub-figure (a), the segmentation surface of our

TABLE 5. Quantitative result of different segmentation methods on the JRC test dataset, where the best and the second best are denoted by bold text and underlined text, respectively. Note that ground truth labels are corrected based on Heidelberg's result.

					Dice Loss per Layer						
Data	Method	Error	Layer Error	Dice Loss	RNFL	GCL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE
All	Rathke et al. [31]	4.51%	27.09%	0.2820 (±0.076)	0.2755	0.2190	0.3228	0.2976	0.2240	0.4376	0.1976
	OCTExplorer [27]	3.78%	22.11%	0.2291 (±0.063)	0.1983	0.1342	0.2195	0.3148	0.1657	0.2858	0.2855
	Heidelberg [10]	1.74%	10.96%	0.1070 (±0.028)	0.0972	0.1152	0.1519	0.1378	0.0807	0.0693	0.0966
	U-Net [24]	1.58%	9.34%	0.0930 (±0.024)	0.0855	0.0760	0.1217	0.1291	0.0566	0.0825	0.0998
	RelayNet [14]	1.32%	7.93%	0.0816 (±0.024)	0.0739	0.0675	0.1134	0.1220	0.0523	0.0686	0.0736
	MGU-Net [20]	1.25%	7.54%	0.0788 (±0.024)	**0.0702**	**0.0635**	0.1069	0.1215	**0.0508**	0.0715	0.0668
	Ours (2D, original)	1.24%	7.45%	0.0771 (±0.023)	0.0694	0.0647	**0.1043**	**0.1183**	0.0516	0.0657	0.0656
	Ours (3D, original)	1.49%	8.22%	0.0902 (±0.026)	0.0850	0.0838	0.1236	0.1329	0.0568	0.0659	0.0836
	Ours (3D, motion corrected)	**1.22%**	**7.13%**	**0.0766** (±0.024)	0.0709	0.0652	0.1055	0.1211	**0.0508**	**0.0628**	**0.0596**
Normal	Rathke et al. [31]	3.22%	19.67%	0.2077 (±0.086)	0.2267	0.1399	0.2007	0.2027	0.1819	0.3955	0.1063
	OCTExplorer [27]	2.99%	16.50%	0.1835 (±0.056)	0.2003	0.0928	0.1542	0.2499	0.1446	0.2637	0.1788
	Heidelberg [10]	**0.22%**	**1.26%**	**0.0138** (±0.004)	**0.0165**	**0.0100**	**0.0167**	**0.0213**	**0.0096**	**0.0097**	**0.0126**
	U-Net [24]	1.02%	5.38%	0.0612 (±0.021)	0.0505	0.0409	0.1010	0.0820	0.0360	0.0542	0.0639
	RelayNet [14]	0.87%	4.64%	0.0531 (±0.023)	0.0472	0.0404	0.0978	0.0753	0.0265	0.0359	0.0484
	MGU-Net [20]	0.81%	4.67%	0.0511 (±0.019)	0.0478	0.0408	0.0815	0.0767	0.0269	0.0372	0.0471
	Ours (2D, original)	0.78%	4.53%	0.0493 (±0.019)	0.0467	0.0404	0.0799	0.0742	0.0262	0.0341	0.0433
	Ours (3D, original)	0.90%	4.93%	0.0558 (±0.020)	0.0513	0.0449	0.0870	0.0837	0.0310	0.0370	0.0558
	Ours (3D, motion corrected)	1.22%	7.13%	0.0766 (±0.024)	0.0709	0.0652	0.1055	0.1211	0.0508	0.0628	0.0596
Moderate deformation	Rathke et al. [31]	3.48%	21.60%	0.2292 (±0.082)	0.2307	0.1734	0.2604	0.2380	0.1853	0.3983	0.1182
	OCTExplorer [27]	3.24%	19.01%	0.2025 (±0.058)	0.1919	0.1128	0.1883	0.2844	0.1429	0.2635	0.2338
	Heidelberg [10]	**0.55%**	**3.28%**	**0.0340** (±0.008)	**0.0431**	**0.0315**	**0.0410**	**0.0427**	**0.0200**	**0.0279**	**0.0315**
	U-Net [24]	1.14%	7.20%	0.0722 (±0.021)	0.0708	0.0597	0.0995	0.1011	0.0386	0.0638	0.0719
	RelayNet [14]	1.06%	6.51%	0.0675 (±0.022)	0.0677	0.0586	0.0998	0.0963	0.0350	0.0519	0.0631
	MGU-Net [20]	1.04%	6.43%	0.0669 (±0.021)	0.0653	0.0552	0.0985	0.0960	0.0348	0.0562	0.0625
	Ours (2D, original)	1.00%	6.12%	0.0642 (±0.021)	0.0647	0.0569	0.0959	0.0936	0.0347	0.0490	0.0547
	Ours (3D, original)	1.25%	6.87%	0.0771 (±0.025)	0.0788	0.0734	0.1131	0.1079	0.0401	0.0485	0.0776
	Ours (3D, motion corrected)	1.01%	6.09%	0.0654 (±0.023)	0.0662	0.0583	0.0997	0.0983	0.0355	0.0464	0.0531
Severe deformation	Rathke et al. [31]	7.48%	42.36%	0.4326 (±0.080)	0.3800	0.3618	0.5138	0.4904	0.3448	0.5596	0.3778
	OCTExplorer [27]	5.38%	31.43%	0.3097 (±0.088)	0.2088	0.2044	0.3142	0.4210	0.2404	0.3549	0.4245
	Heidelberg [10]	5.18%	32.01%	0.3116 (±0.090)	0.2378	0.3694	0.4445	0.4143	0.2772	0.1977	0.2406
	U-Net [24]	2.84%	15.82%	0.1545 (±0.033)	0.1246	0.1315	0.1754	0.2246	0.1237	0.1492	0.1526
	RelayNet [14]	2.14%	12.55%	0.1289 (±0.037)	0.0945	0.1022	0.1477	0.2088	0.1200	0.1313	0.0979
	MGU-Net [20]	1.95%	11.30%	0.1213 (±0.040)	0.0870	0.0946	0.1354	0.2068	0.1139	0.1316	0.0796
	Ours (2D, original)	2.00%	11.71%	0.1213 (±0.036)	**0.0859**	0.0956	0.1324	0.1997	0.1186	0.1276	0.0895
	Ours (3D, original)	2.34%	12.69%	0.1388 (±0.037)	0.1083	0.1289	0.1610	0.2197	0.1230	0.1294	0.1012
	Ours (3D, motion corrected)	**1.89%**	**10.53%**	**0.1157** (±0.037)	0.0872	**0.0933**	**0.1264**	**0.1965**	**0.1105**	**0.1229**	**0.0734**
Manually corrected area	Rathke et al. [31]	6.85%	38.90%	0.4002 (±0.075)	0.3697	0.3332	0.4707	0.4529	0.3159	0.5242	0.3350
	OCTExplorer [27]	5.10%	29.87%	0.2970 (±0.079)	0.2162	0.2025	0.3020	0.4072	0.2235	0.3368	0.3905
	Heidelberg [10]	4.72%	29.22%	0.2855 (±0.077)	0.2458	0.3379	0.4000	0.3727	0.2386	0.1829	0.2207
	U-Net [24]	2.56%	14.89%	0.1445 (±0.033)	0.1212	0.1268	0.1724	0.2114	0.1115	0.1356	0.1328
	RelayNet [14]	2.10%	12.41%	0.1268 (±0.036)	0.0990	0.1056	0.1528	0.2032	0.1097	0.1210	0.0961
	MGU-Net [20]	1.93%	11.37%	0.1203 (±0.038)	0.0927	0.0976	0.1419	0.2013	0.1044	0.1235	0.0808
	Ours (2D, original)	1.96%	11.60%	0.1195 (±0.035)	**0.0915**	0.0993	0.1388	0.1956	0.1077	0.1163	0.0874
	Ours (3D, original)	2.27%	12.45%	0.1351 (±0.036)	0.1125	0.1279	0.1632	0.2116	0.1120	0.1178	0.1005
	Ours (3D, motion corrected)	**1.85%**	**10.52%**	**0.1137** (±0.036)	0.0922	**0.0964**	**0.1324**	**0.1911**	**0.1006**	**0.1107**	**0.0728**
Normal	Rathke et al. [31]	3.72%	22.40%	0.2211 (±0.083)	0.2283	0.1519	0.2352	0.2259	0.1863	0.3990	0.1208
	OCTExplorer [27]	3.30%	17.98%	0.1968 (±0.052)	0.1690	0.1436	0.1806	0.2765	0.1511	0.2759	0.1808
	Heidelberg [10]	1.28%	7.14%	0.0794 (±0.030)	0.0592	0.0721	**0.1136**	0.1362	0.0564	0.0501	0.0685
	U-Net [24]	1.60%	7.56%	0.0914 (±0.036)	0.0561	0.0749	0.1614	0.1286	0.0666	0.0672	0.0850
	RelayNet [14]	1.28%	6.80%	0.0785 (±0.035)	0.0475	0.0687	0.1397	0.1239	0.0488	0.0526	0.0686
	MGU-Net [20]	1.18%	6.65%	0.0745 (±0.031)	0.0462	0.0662	0.1184	0.1253	0.0461	0.0534	0.0662
	Ours (2D, original)	1.17%	6.65%	0.0740 (±0.031)	0.0463	0.0677	0.1170	0.1273	0.0481	0.0501	0.0618
	Ours (3D, original)	1.21%	6.60%	0.0755 (±0.032)	0.0477	0.0682	0.1184	0.1299	0.0504	0.0479	0.0662
	Ours (3D, motion corrected)	**1.10%**	**6.11%**	**0.0696** (±0.032)	**0.0446**	**0.0654**	0.1145	**0.1218**	**0.0446**	**0.0407**	**0.0552**
Moderate deformation	Rathke et al. [31]	4.32%	25.80%	0.2771 (±0.079)	0.2770	0.2254	0.3183	0.3080	0.2224	0.4259	0.1624
	OCTExplorer [27]	3.78%	22.65%	0.2379 (±0.058)	0.2111	0.1660	0.2460	0.3390	0.1626	0.2723	0.2681
	Heidelberg [10]	1.96%	12.12%	0.1263 (±0.036)	0.1590	0.1407	0.1649	0.1612	0.0767	0.0893	0.0923
	U-Net [24]	1.66%	10.77%	0.1065 (±0.029)	0.1029	0.0993	0.1442	0.1521	0.0659	0.0882	0.0927
	RelayNet [14]	1.62%	10.14%	0.1042 (±0.032)	0.0963	0.1007	0.1491	0.1549	0.0658	0.0778	0.0844
	MGU-Net [20]	1.58%	9.95%	0.1029 (±0.031)	0.0958	**0.0936**	0.1454	0.1524	0.0645	0.0853	0.0830
	Ours (2D, original)	1.51%	9.51%	0.0987 (±0.032)	0.0947	0.0964	**0.1404**	0.1506	0.0637	0.0713	0.0736
	Ours (3D, original)	1.75%	10.07%	0.1103 (±0.034)	0.1083	0.1107	0.1561	0.1610	0.0691	0.0725	0.0944
	Ours (3D, motion corrected)	**1.48%**	**9.06%**	**0.0966** (±0.033)	**0.0946**	0.0949	0.1411	**0.1481**	**0.0622**	**0.0672**	**0.0683**
Severe deformation	Rathke et al. [31]	8.34%	45.96%	0.4695 (±0.081)	0.4221	0.3908	0.5499	0.5421	0.3766	0.5901	0.4150
	OCTExplorer [27]	5.88%	33.89%	0.3316 (±0.092)	0.2210	0.2219	0.3317	0.4501	0.2655	0.3803	0.4504
	Heidelberg [10]	6.34%	38.47%	0.3746 (±0.104)	0.2981	0.4481	0.5204	0.4947	0.3414	0.2420	0.2777
	U-Net [24]	3.08%	17.21%	0.1675 (±0.037)	0.1325	0.1415	0.1851	0.2487	0.1461	0.1702	0.1482
	RelayNet [14]	2.39%	13.77%	0.1425 (±0.043)	0.1030	0.1094	0.1550	0.2330	0.1436	0.1522	0.1011
	MGU-Net [20]	2.16%	12.27%	0.1337 (±0.047)	0.0940	0.1009	0.1413	0.2314	0.1359	0.1516	0.0805
	Ours (2D, original)	2.23%	12.84%	0.1342 (±0.042)	**0.0926**	0.1020	0.1390	0.2228	0.1421	0.1483	0.0930
	Ours (3D, original)	2.60%	13.87%	0.1526 (±0.042)	0.1178	0.1385	0.1680	0.2432	0.1457	0.1512	0.1035
	Ours (3D, motion corrected)	**2.07%**	**11.43%**	**0.1268** (±0.043)	0.0939	**0.0986**	**0.1293**	**0.2175**	**0.1309**	**0.1427**	**0.0748**

3D network is shown in sub-figure (b), and the partially annotated ground truth segmentation surface is shown in sub-figure (c). Sub-figures (d)-(f) show the central cross-section slow B-scan, our segmentation, and ground truth respectively, and sub-figures (g)-(i) show segmentation on the central fast B-scan. It can be observed that eye motion distorts the slow B-scan in sub-figures (a1), and causes jittered segmentation surface in (b1), (c1), (e1), and (f1). After applying the motion correction network, the segmentation surfaces are smooth along the slow-scanning axis and the slow B-scan provides better visualization.

The quantitative performance of several methods on the AMD and control dataset [23] are shown in Table 4, where our proposed 3D network after motion correction achieves the lowest error and Dice loss for each layer. It can also be observed that the performance of our 3D network degrades without the motion correction network, demonstrating the importance of motion correction on dataset with real motion artifacts.

G. JRC DATASET

We compare the proposed method to the clinically available solutions [10], [27] on the JRC dataset with various diseases. The qualitative results on the JRC dataset are shown in Fig. 9. Two examples with wet AMD and CRVO are illustrated in group (1) and group (2), respectively. Sub-figure (a) shows the 3D OCT volume with motion artifacts, sub-figure (b) shows the motion corrected 3D OCT volume, sub-figure (c) shows the slow B-scan with motion artifacts, sub-figure (d) shows the motion corrected B-scan, and sub-figure (e) shows a reference vertical B-scan imaged separately. Sub-figure (g) shows the segmentation of different methods on the fast B-scan, and sub-figure (h) shows the segmentation of different methods on the motion corrected slow B-scan. The conventional methods by Rathke et al. [31], Heidelberg [10], and OCTExplorer [2] yield significant segmentation errors of the OPL and GLC-IPL in example (1), as denoted by pink and red arrows. These methods also produce large errors denoted by red circles in example (2) compared to the ground truth in sub-figure (g2-8) and sub-figure (h2-8). This is because the conventional methods rely on graph prior designed for normal eyes or diseases with mild deformations, which could not generalize well for various diseases with large deformations. Deep learning methods including the U-Net [24], RelayNet [14], MGU-Net [20], and our proposed 3D network could produce segmentation with higher similarity to ground truth for both examples compared with conventional methods. However, the segmentation results of U-Net [24], RelayNet [14], and MGU-Net [20] yield mis-classifications as denoted by red circles in sub-figures (g2-4) to (g2-6), and they also yield and 3D inconsistency as denoted by red circles in sub-figures (h2-4) to (h2-6). The proposed 3D network is more accurate in the fast B-scans, and also yields better 3D consistency in the slow B-scans for both examples.

We also visualize the 3D segmentation surfaces of Heidelberg [10], MGU-Net [20], our 3D network, and ground truth in Fig. 10. One quarter of the OCT is cut to show the cross-section of the segmentation surfaces. It could be observed that Heidelberg and MGU-Net produce segmentation errors denoted in red circles, and the proposed method is the most similar to the ground truth.

The quantitative results are presented in Table 5, where the proposed 3D segmentation network is compared with different segmentation methods. Since the manual annotation is performed based on Heidelberg's segmentation, the quantitative results would be biased towards Heidelberg. Therefore we also report the error evaluated only on manually corrected areas, where at least 4 out of 8 layers in the ground truth differ from Heidelberg's segmentation. The OCT volumes in the test set are divided into three categories, including normal, moderate deformation, and severe deformation. The percentage of area that is manually corrected is 1.02% for normal, 11.41% for moderate, and 21.47% for severe deformation. On the entire test set, our proposed 3D segmentation network with motion correction yields the lowest error and Dice loss on average. When divided into three categories based on diseases, Heidelberg achieves the lowest error for normal and moderate deformation due to the bias of the ground truth. However, our proposed method outperforms Heidelberg segmentation by a large margin for the category of severe deformation, decreasing the layer error from 32.01% to 10.53%, and the average Dice loss from 0.3116 to 0.1157. For evaluation on the manually corrected area, the proposed 3D segmentation network with motion correction achieves the best performance overall and in each category of diseases. The results demonstrate a significant advantage of the combined motion correction and 3D segmentation network for a clinical dataset with various diseases.

V. CONCLUSION

In this paper, we proposed to combine motion correction and 3D OCT layer segmentation using a novel graph-inspired architecture, which led to promising improvement upon existing 2D segmentation methods with or without motion correction. We also collected one of the largest OCT segmentation dataset covering a variety of diseases. Experimental results demonstrated that the motion correction is essential to apply 3D segmentation, and combining motion correction with 3D segmentation achieved the best performance for three datasets compared to conventional and deep learning state-of-the-art methods. Specifically, the proposed network demonstrated a significant advantage over clinically available segmentation solutions for severe diseases. The diagnosis and evaluation of diseases with large deformation such as DME, wet AMD and CRVO would greatly benefit from the improved accuracy, which impacts tens of millions of patients.

One limitation of the proposed method is that the network could not identify retinal fluids in diseases like DME and

CRVO, where the retinal layers segmentation may have low confidence. In future work, the segmentation network can be extended to support segmentation of retinal fluid and other lesions. The proposed segmentation method could be used to generate more accurate OCT-A projection images, promote the analysis of layer thickness and vessel density, which is beneficial for diagnosing and monitoring retinal and systemic diseases.

ACKNOWLEDGMENT

The authors would like to thank Daniel Deussen, Fritz Kalaw, and Yassin Shaden from the Jacobs Retina Center for data collection.

REFERENCES

[1] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, and C. A. Puliafito, "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[2] M. D. Abramoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, 2010.

[3] J. G. Fujimoto, W. Drexler, J. S. Schuman, and C. K. Hitzenberger, "Optical coherence tomography (OCT) in ophthalmology: Introduction," *Opt. Exp.*, vol. 17, no. 5, pp. 3978–3979, 2009.

[4] A. London, I. Benhar, and M. Schwartz, "The retina as a window to the brain—From eye research to CNS disorders," *Nature Rev. Neurol.*, vol. 9, no. 1, pp. 44–53, Jan. 2013.

[5] A. Uchida, J. A. Pillai, R. Bermel, A. Bonner-Jackson, A. Rae-Grant, H. Fernandez, J. Bena, S. E. Jones, J. B. Leverenz, and S. K. Srivastava, "Outer retinal assessment using spectral-domain optical coherence tomography in patients with Alzheimer's and Parkinson's disease," *Investigative Ophthalmol. Vis. Sci.*, vol. 59, no. 7, pp. 2768–2777, 2018.

[6] A. Carass, A. Lang, M. Hauser, P. A. Calabresi, H. S. Ying, and J. L. Prince, "Multiple-object geometric deformable model for segmentation of macular OCT," *Biomed. Opt. Exp.*, vol. 5, no. 4, pp. 1062–1074, 2014.

[7] J. Novosel, G. Thepass, H. G. Lemij, J. F. de Boer, K. A. Vermeer, and L. J. van Vliet, "Loosely coupled level sets for simultaneous 3D retinal layer segmentation in optical coherence tomography," *Med. Image Anal.*, vol. 26, no. 1, pp. 146–158, Dec. 2015.

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

[9] S. J. Chiu, M. J. Allingham, P. S. Mettu, S. W. Cousins, J. A. Izatt, and S. Farsiu, "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," *Biomed. Opt. Exp.*, vol. 6, no. 4, pp. 1172–1194, 2015.

[10] M. M. Teussink, S. Donner, T. Otto, K. Williams, and A. Tafreshi, "State-of-the-art commercial spectral-domain and swept-source OCT technologies and their clinical applications in ophthalmology," *Heidelberg Eng.*, May 2019.

[11] J. L. Lauermann, A. K. Woetzel, M. Treder, M. Alnawaiseh, C. R. Clemens, N. Eter, and F. Alten, "Prevalences of segmentation errors and motion artifacts in OCT-angiography differ among retinal diseases," *Graefe's Arch. Clin. Experim. Ophthalmol.*, vol. 256, no. 10, pp. 1807–1816, Oct. 2018.

[12] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," *Biomed. Opt. Exp.*, vol. 8, no. 5, pp. 2732–2744, 2017.

[13] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, and D. Visentin, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature Med.*, vol. 24, no. 9, pp. 1342–1350, Sep. 2018.

[14] A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab, "ReLayNet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomed. Opt. Exp.*, vol. 8, no. 8, pp. 3627–3642, 2017.

[15] Y. He, A. Carass, Y. Yun, C. Zhao, B. M. Jedynak, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, "Towards topological correct segmentation of macular OCT from cascaded FCNS," in *Fetal, Infant and Ophthalmic Medical Image Analysis*. Berlin, Germany: Springer, 2017, pp. 202–209.

[16] Y. He, A. Carass, Y. Liu, B. M. Jedynak, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, "Deep learning based topology guaranteed surface and MME segmentation of multiple sclerosis subjects from retinal OCT," *Biomed. Opt. Exp.*, vol. 10, no. 10, pp. 5042–5058, 2019.

[17] Y. He, A. Carass, Y. Liu, B. M. Jedynak, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, "Fully convolutional boundary regression for retina OCT segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 120–128.

[18] Y. He, A. Carass, Y. Liu, B. M. Jedynak, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, "Structured layer surface segmentation for retina OCT using fully convolutional regression networks," *Med. Image Anal.*, vol. 68, Feb. 2021, Art. no. 101856.

[19] M. Pekala, N. Joshi, T. Y. A. Liu, N. M. Bressler, D. C. DeBuc, and P. Burlina, "Deep learning based retinal OCT segmentation," *Comput. Biol. Med.*, vol. 114, Nov. 2019, Art. no. 103445.

[20] J. Li, P. Jin, J. Zhu, H. Zou, X. Xu, M. Tang, M. Zhou, Y. Gan, J. He, and Y. Ling, "Multi-scale GCN-assisted two-stage network for joint segmentation of retinal layers and discs in peripapillary OCT images," *Biomed. Opt. Exp.*, vol. 12, no. 4, pp. 2204–2220, 2021.

[21] S. Mukherjee, T. De Silva, P. Grisso, H. Wiley, D. K. Tiarnan, A. T. Thavikulwat, E. Chew, and C. Cukras, "Retinal layer segmentation in optical coherence tomography (OCT) using a 3D deep-convolutional regression network for patients with age-related macular degeneration," *Biomed. Opt. Exp.*, vol. 13, no. 6, pp. 3195–3210, 2022.

[22] Y. He, A. Carass, Y. Liu, P. A. Calabresi, S. Saidha, and J. L. Prince, "Longitudinal deep network for consistent OCT layer segmentation," *Biomed. Opt. Exp.*, vol. 14, no. 5, pp. 1874–1893, 2023.

[23] S. Farsiu, S. J. Chiu, R. V. O'Connell, F. A. Folgar, E. Yuan, J. A. Izatt, and C. A. Toth, "Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography," *Ophthalmology*, vol. 121, no. 1, pp. 162–172, Jan. 2014.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[25] L. Sánchez Brea, D. Andrade De Jesus, M. F. Shirazi, M. Pircher, T. van Walsum, and S. Klein, "Review on retrospective procedures to correct retinal motion artefacts in OCT imaging," *Appl. Sci.*, vol. 9, no. 13, p. 2700, Jul. 2019.

[26] M. K. Garvin, M. D. Abramoff, R. Kardon, S. R. Russell, X. Wu, and M. Sonka, "Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-D graph search," *IEEE Trans. Med. Imag.*, vol. 27, no. 10, pp. 1495–1505, Oct. 2008.

[27] M. K. Garvin, M. D. Abramoff, X. Wu, S. R. Russell, T. L. Burns, and M. Sonka, "Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images," *IEEE Trans. Med. Imag.*, vol. 28, no. 9, pp. 1436–1447, Sep. 2009.

[28] H. Bogunović, F. Venhuizen, S. Klimscha, S. Apostolopoulos, A. Bab-Hadiashar, U. Bagci, M. F. Beg, L. Bekalo, Q. Chen, and C. Ciller, "RETOUCH: The retinal OCT fluid detection and segmentation benchmark and challenge," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1858–1874, Aug. 2019.

[29] D. Lu, M. Heisler, S. Lee, G. Ding, M. V. Sarunic, and M. Faisal Beg, "Retinal fluid segmentation and detection in optical coherence tomography images using fully convolutional neural network," 2017, *arXiv:1710.04778*.

[30] D. C. Fernández, H. M. Salinas, and C. A. Puliafito, "Automated detection of retinal layer structures on optical coherence tomography images," *Opt. Exp.*, vol. 13, no. 25, pp. 10200–10216, 2005.

[31] F. Rathke, S. Schmidt, and C. Schnörr, "Probabilistic intra-retinal layer segmentation in 3-D OCT images using global shape regularization," *Med. Image Anal.*, vol. 18, no. 5, pp. 781–794, Jul. 2014.

[32] S. J. Chiu, C. A. Toth, C. B. Rickman, J. A. Izatt, and S. Farsiu, "Automatic segmentation of closed-contour features in ophthalmic images using graph theory and dynamic programming," *Biomed. Opt. Exp.*, vol. 3, no. 5, pp. 1127–1140, 2012.

[33] Y. Chen, M. Rohrbach, Z. Yan, Y. Shuicheng, J. Feng, and Y. Kalantidis, "Graph-based global reasoning networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 433–442.

[34] B. Potsaid, I. Gorczynska, V. J. Srinivasan, Y. Chen, J. Jiang, A. Cable, and J. G. Fujimoto, "Ultrahigh speed spectral/Fourier domain OCT ophthalmic imaging at 70,000 to 312,500 axial scans per second," *Opt. Exp.*, vol. 16, no. 19, pp. 15149–15169, 2008.

[35] M. F. Kraus, B. Potsaid, M. A. Mayer, R. Bock, B. Baumann, J. J. Liu, J. Hornegger, and J. G. Fujimoto, ''Motion correction in optical coherence tomography volumes on a per A-scan basis using orthogonal scan patterns,'' *Biomed. Opt. Exp.*, vol. 3, no. 6, pp. 1182–1199, 2012.

[36] B. Antony, M. D. Abramoff, L. Tang, W. D. Ramdas, J. R. Vingerling, N. M. Jansonius, K. Lee, Y. H. Kwon, M. Sonka, and M. K. Garvin, ''Automated 3-D method for the correction of axial artifacts in spectral-domain optical coherence tomography images,'' *Biomed. Opt. Exp.*, vol. 2, no. 8, pp. 2403–2416, 2011.

[37] Y. Wang, A. Warter, M. Cavichini-Cordeiro, W. R. Freeman, D. G. Bartsch, T. Q. Nguyen, and C. An, ''Learning to correct axial motion in oct for 3D retinal imaging,'' in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 126–130.

[38] Y. Wang, C. Galang, W. R. Freeman, T. Q. Nguyen, and C. An, ''Joint motion correction and 3D segmentation with graph-assisted neural networks for retinal OCT,'' in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2022, pp. 766–770.

YIQIAN WANG received the B.S. degree in electrical engineering from the Beijing Institute of Technology, Beijing, China, in 2018, and the Ph.D. degree in electrical and computer engineering from the University of California, San Diego, in 2022. She is currently a Senior Engineer with Qualcomm Technologies Inc. Her research interests include medical image processing, signal processing, and deep learning.

CARLO GALANG received the M.D. degree from the University of Santo Tomas, Manila, Philippines, in 2014. He completed the Residency Training from the University of Santo Tomas Hospital, in 2020. He is currently a Retina Research Fellow with the Jacobs Retina Center, Shiley Eye Institute, University of California, San Diego. His research interests include retinal diseases, imaging, and surgeries.

WILLIAM R. FREEMAN is a Distinguished Professor of ophthalmology, the Director of the Jacobs Retina Center, and the Vice Chair of the Department of Ophthalmology, UCSD. He is a full time retina surgeon and also a researcher, who has held NIH grants for nearly 30 years. He works closely with the Imaging Group, Department of Ophthalmology, and the School of Engineering, UCSD. He has over 600 peer-reviewed publications.

ALEXANDRA WARTER received the M.D. degree from the University Institute of Medicine of the Hospital Italiano, Buenos Aires, Argentina, in 2015. She completed the Ophthalmology Residency from Centro de Ojos Quilmes, Buenos Aires, in 2020. She is currently a Clinical and Research Retina Fellow with the Jacobs Retina Center, Shiley Eye Institute, University of California, San Diego. Her research and clinical interests include macular degeneration, diabetic retinopathy, imaging of these and related diseases, and the areas of retina disorders. She is also focused on clinical trial methodology in retina.

ANNA HEINKE received the M.D. degree from the Jagiellonian University Medical College, Krakow, Poland, in 2013, and the Ph.D. degree in ophthalmology from the Medical University of Silesia, Katowice, Poland, in 2019. She completed the Ophthalmology Residency Training in Katowice and Germany (University Hospital Frankfurt am Main and Charite University Hospital in Berlin). She has been a European Board-Certified Ophthalmologist (FEBO), since 2022. She is currently a Retina Research Fellow with the Jacobs Retina Center, Shiley Eye Institute, University of California, San Diego. Her research interests include retinal diseases, multimodal retinal imaging, and the use of artificial intelligence in ophthalmology.

DIRK-UWE G. BARTSCH received the bachelor's degree from Technische Universitaet Darmstadt and the Ph.D. degree in bioengineering from the University of California, San Diego. He completed a postdoctoral fellowship with the University of California. He is an Associate Adjunct Professor and the Co-Director of the Jacobs Retina Center. His research is focused in retinal imaging, scanning laser imaging–confocal/non-confocal, optical coherence tomography (OCT), indocyanine green and fluorescein angiography, and tomographic reconstruction of the posterior pole in patients with various retina diseases, such as age-related macular degeneration, diabetes, and HIV-related complications.

TRUONG Q. NGUYEN (Fellow, IEEE) is currently a Professor with the ECE Department, UC San Diego. He is the coauthor (with Prof. Gilbert Strang) of a textbook *Wavelets & Filter Banks* (Wellesley-Cambridge Press, 1997), and the author of several MATLAB-based toolboxes on image compression, electrocardiogram compression, and filter bank design. He has over 400 publications. His current research interests are 3-D video processing and communications and their efficient implementation.

He received the IEEE Transactions on Signal Processing Paper Award (image and multidimensional processing area) for the paper he co-authored with Prof. P. P. Vaidyanathan on linear-phase perfect-reconstruction filter banks, in 1992. He received the NSF Career Award, in 1995. He is the Series Editor of *Digital Signal Processing* (Academic Press). He served as an Associate Editor for IEEE Transactions on Signal Processing, from 1994 to 1996; IEEE Signal Processing Letters, from 2001 to 2003; IEEE Transactions on Circuits and Systems, from 1996 to 1997 and from 2001 to 2004; and IEEE Transactions on Image Processing, from 2004 to 2005.

CHEOLHONG AN received the B.S. and M.S. degrees in electrical engineering from Pusan National University, Busan, South Korea, in 1996 and 1998, respectively, and the Ph.D. degree in electrical and computer engineering, in 2008. He is an Assistant Adjunct Professor of electrical and computer engineering with the University of California, San Diego. Earlier, he was with Samsung Electronics, South Korea, and Qualcomm, USA. His current research is focused on the medical image processing and the real-time bio image processing. His research interests are in 2-D and 3-D image processing with machine learning and sensor technology.