UC San Diego
UC San Diego Previously Published Works

Title

JOINT MOTION CORRECTION AND 3D SEGMENTATION WITH GRAPH-ASSISTED NEURAL NETWORKS FOR RETINAL OCT.

Permalink

https://escholarship.org/uc/item/0379r5c1

Authors

Wang, Yiqian
Galang, Carlo
Freeman, William
et al.

Publication Date

2022-10-01

DOI

10.1109/icip46576.2022.9898072

Peer reviewed

JOINT MOTION CORRECTION AND 3D SEGMENTATION WITH GRAPH-ASSISTED NEURAL NETWORKS FOR RETINAL OCT

Yiqian Wang[1], **Carlo Galang**[2], **William R. Freeman**[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, University of California, San Diego

Abstract

Optical Coherence Tomography (OCT) is a widely used non-invasive high resolution 3D imaging technique for biological tissues and plays an important role in ophthalmology. OCT retinal layer segmentation is a fundamental image processing step for OCT-Angiography projection, and disease analysis. A major problem in retinal imaging is the motion artifacts introduced by involuntary eye movements. In this paper, we propose neural networks that jointly correct eye motion and retinal layer segmentation utilizing 3D OCT information, so that the segmentation among neighboring B-scans would be consistent. The experimental results show both visual and quantitative improvements by combining motion correction and 3D OCT layer segmentation comparing to conventional and deep-learning based 2D OCT layer segmentation.

Index Terms—

Optical coherence tomography; deep learning; image segmentation; motion correction; retinal imaging

1. INTRODUCTION

Optical Coherence Tomography (OCT) is one of the most important imaging techniques in ophthalmology, which can image the cross-sections of the retina at μm-resolution [1]. OCT has been widely used for diagnosing and monitoring most retinal diseases [2], including age-related macular degeneration (AMD), diabetic macular edema (DME), glaucoma, and so on. Segmentation of retinal layers plays an important role in detecting changes in layer thickness and assessment of both retinal and systematic diseases [3]. In particular, the thickness of the retinal nerve fiber layer (RNFL) is frequently used for assessment of glaucoma, while the thickness of the entire retina is often used for assessment of DME, choroidal neovascularization (CNV), and macular hole [2].

Many works on OCT layer segmentation are based on the level set or graph methods [4, 5, 6, 7], yet segmentation error is prevalent with conventional segmentation approaches [8]. The use of deep learning is gradually popular in OCT segmentation, and convolutional neural networks have significantly improved the accuracy of layer segmentation [9, 10, 11, 12, 13]. In particular, the fully convolutional U-net [14], which enables image-to-image translation and includes multiscale analysis, has shown great success in many biomedical

image segmentation tasks. However, most deep learning methods are based on 2D images, which ignore 3D contextual information.

One major reason that prevents most methods from fully utilizing 3D contextual information is that involuntary eye motion causes misalignment artifacts between neighboring 2D raster B-scans in OCT imaging. In this paper, we propose neural networks that jointly correct eye motion and retinal layer segmentation utilizing 3D information, so that the segmentation among neighboring B-scans would be consistent. The performance of the proposed method is compared with several state-of-the-art methods. Experimental results show that motion correction and 3D contextual information enhance the accuracy of the OCT layer segmentation.

2. RELATED WORKS

Most earlier methods for OCT layer segmentation are based on 2D image, where segmentation labels or boundaries are computed using a single B-scan [4, 5, 6, 7, 15]. Level set-based methods were adopted to ensure topologically correct (i.e. in the correct order) layer boundaries [4, 5], but these methods were very computationally expensive with run-time up to hours per OCT volume. Another widely studied category of methods is based on graph theory [6, 16, 15], and can be combined with machine learning classifiers [7, 9]. However, as these methods work on 2D images, 3D contextual information is not utilized for segmentation.

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 as a pre-processing step to recover the motion-free 3D OCT volume. The first segmentation approach to utilize 3D information was proposed by Garvin et al. [17, 18], which could segment 7 boundaries for macular centered OCT scans. The approach first flattens the bottom surface of the retina to remove motion artifacts (along with retinal curvature) and then enhances 2D graph-based methods by additional 3D "feasibility" constraints. The feasibility constraints take advantage of 3D contextual information and enforce smoothness in neighboring surfaces and surface distance constraints. They demonstrate that their proposed method with 3D information could reduce segmentation failure compared to the 2D graph-based approach.

However, segmentation error was still prevalent in these conventional methods. Moreover, since they required manually selected features and extensive parameter tuning, it was difficult to generalize to different imaging instruments with various resolutions and appearances, or patient groups with different diseases. Recent deep learning-based methods led to a major improvement in segmentation accuracy [9, 10, 11, 12, 13]. RelayNet [19] is a deep learning approach for retinal layer segmentation, which is based on U-Net architecture and utilized weighted cross-entropy loss. Some works combined deep learning classifiers with conventional post-processing to obtain layer boundaries from pixel-wise prediction. Fang et al. [9] proposed a neural network and graph search in post-processing to segment 9 layer boundaries. Pekala et al. [13] proposed a dense U-net with Gaussian process regression as post-processing to segment 5 layer boundaries. Recent attempts have been made to

directly learn the boundaries using regression. He et al. [11] proposed a cascaded U-Net to segment layers with the topological constraints, which learns the thickness map of each layer. The network was later improved [12] by including two extra input channels containing X and Y coordinate information. MGU-Net [20] is the latest state-of-the-art method that combines U-Net with graph-convolution inspired global reasoning blocks, which achieves improved Dice coefficient and visual quality.

However, these neural networks only learn from 2D images without fully utilizing 3D information. Google DeepMind [10] proposed a 3D segmentation network taking 9 consecutive B-scans that could segment 15 classes of features to aid disease classification. The major limitation is that only two retinal layers can be identified by the segmentation network, namely the neurosensory retina and the RPE. Nevertheless, the network demonstrated that including 3D information would lead to improvement in OCT layer segmentation quality compared to 2D networks.

3. PROPOSED METHOD

In this paper, we proposed to combine an OCT motion correction network with a segmentation network using 3D B-scans. The proposed pipeline is shown in Fig. 1. In the first stage, motion artifacts in the input OCT volume $\mathbf{V}$ are corrected using the motion correction network proposed in [21] based on the 3D input volume and a rough 3-class segmentation of the entire retina using a 2D segmentation method. In the second stage, the detailed retinal layers $\mathbf{S}^{pred}$ are classified using a segmentation network with 3D input based on the motion-corrected OCT volume.

The motion correction network [21] is designed to predict a 2D displacement map $\mathbf{D}$ to remove axial motion artifacts from 3D OCT volume, and the top (inner limiting membrane, ILM) and bottom (Bruch's membrane, BM) segmentation boundaries of the retina where BM and ILM layers improve correction performance. The motion correction network is based on a modified U-Net with residual blocks, which can be applied to different resolutions. More information can be found in [21]. The OCT volume is then warped based on the predicted the 2D displacement map.

Next, the proposed network inputs N neighboring B-scans in the motion corrected OCT volume, and outputs the segmentation in K classes for the central B-scan as shown in Fig. 2. The architecture of the segmentation network features 4 down-sampling blocks, 4 up-sampling blocks, and a graph pyramid at the lowest resolution, inspired by the U-net [14] and MGU-Net [20]. The graph pyramid is comprised of 4 resolutions with graph reasoning units (GRU) [22]. The graph reasoning units include three branches for projection to node space, re-projection to feature space, and fusion of global features, where the two graph convolution blocks (GCB) are used after projection to the node space. The proposed architecture includes 1.909 M trainable parameters, which is reduced by 8% compared with MGU-Net which has 2.094M parameters, yet the graph pyramid structure effectively improves segmentation performance as demonstrated in the experimental result.

The loss function is a hybrid loss that combines the cross-entropy loss and a soft Dice loss. We denote the last convolution layer output as $\mathbf{x}$, the ground truth with integer class label as $\mathbf{y} \in [0, K]$, and the one-hot ground truth label as $\mathbf{S}^{\text{GT}}$. Note that we also include a valid mask $\mathbf{M}$, which is 1 for pixels with annotation and 0 for without annotation, so that regions without annotation are ignored in the loss calculation. The cross-entropy loss can be described as

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

where n spans the batch and spatial dimensions. Note that the cross-entropy loss integrates the Sigmoid activation for the last convolution layer output $\mathbf{x}$ using the "log-sum-exp" trick, which has better numeric stability.

The soft Dice loss is derived based on the Dice coefficient, which measures the overlapping ratio between predicted label and ground truth. The soft Dice coefficient is defined as

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

where $\sigma(\cdot)$ denotes the Sigmoid function. Then the Dice loss is defined as one minus the soft Dice coefficient averaged for the retinal layers in class $k = 1$ to $K - 2$ after excluding the top ($k = 0$) and bottom ($k = K - 1$) regions (non-retinal regions),

$$\mathscr{L}_{\text{Dice}}\left(\mathbf{x}, \mathbf{S}^{\text{GT}}\right) = 1 - \underset{k = 1 : K - 2}{\text{mean}} \text{SoftDice}_k\left(\mathbf{x}, \mathbf{S}^{\text{GT}}\right). \tag{3}$$

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

$$\mathscr{L}_{\text{total}} = \lambda_{\text{CE}} \mathscr{L}_{\text{CE}} + \lambda_{\text{Dice}} \mathscr{L}_{\text{Dice}}. \tag{4}$$

Besides the standard data augmentation techniques such as random cropping and horizontal flipping, we also include simulated shearing along the X axis to add another degree of freedom. The method to generate simulated shearing is described in [21].

4. EXPERIMENTAL RESULTS

In the experiment, we compare the segmentation performance of our proposed joint motion correction and 3D segmentation neural network with several state-of-the-art methods.

We test the performance of the proposed method on two public datasets: DME dataset [7] and the AMD and control dataset [23]. The DME dataset [7] is one of the most widely used public dataset in literature [19, 12, 20], which is comprised of 10 macular centered OCT volume for patients with DME. The dataset is captured by Heidelberg Spectralis system, and the OCT volumes have been motioncorrected. The resolution for each volume is $512 \times 768 \times 61$, and 11 selected B-scans in each volume (total 110 B-scans) have manually annotated 8 segmentation boundaries ($K = 9$ classes). We follow the training and test division in other

papers [19, 12], with the first 55 images from subject 1 to 5 is used for training, and the last 55 images from subject 6 to 10 are used for testing. The AMD and control dataset [23] is a public dataset with 384 macular centered OCT volumes from 269 AMD patients and 115 normal control subjects, captured by the Bioptigen system. The resolution of each OCT volume is $512 \times 1000 \times 100$, and annotations for 3 layer boundaries ($K = 4$ classes) are provided in a center 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. Since the DME dataset [7] has been motion corrected, we include simulated motion on the input OCT volumes to test the performance of our proposed motion correction approach as described in [21].

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

$$\mathscr{L}_{\text{Dice }k}\left(\mathbf{S}^{\text{pred}}, \mathbf{S}^{\text{GT}}\right) = 1 - \frac{2\sum \mathbf{S}_k^{\text{pred}} \odot \mathbf{S}_k^{\text{GT}} \odot \mathbf{M}}{\sum \mathbf{S}_k^{\text{pred}} \odot \mathbf{M} + \sum \mathbf{S}_k^{\text{GT}} \odot \mathbf{M}}. \tag{5}$$

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.

In the experiment, we compare the performance of our proposed segmentation with 5 B-scans input (5 channels, center ± 2 neighboring B-scans) versus a single B-scan input (1 channel) with or without our motion correction network. We compare with several conventional methods by Chiu et al. [7] and Rathke et al. [15], as well as deep learning method RelayNet [19] and MGU-Net [20]. Our proposed motion correction and segmentation networks are implemented in PyTorch. The motion correction network utilizes the pre-trained model in [21]. On the DME dataset, the segmentation networks are trained using batch size 1 for 150 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^{-3}. 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. The method by Chiu et al. [7] uses the predicted layer boundaries provided in their dataset, and the method by Rathke et al. [15] uses the original implementation in Matlab. The RelayNet [19] uses the pre-trained PyTorch model provided by the original authors. The MGU-Net is trained with the original implementation in PyTorch using similar hyper parameters as our segmentation network.

The qualitative results of segmentation on the original motion-corrected DME dataset are shown in Fig. 3, where sub-figure (a) shows the input B-scan cropped in the labeled retinal region, sub-figure (b) shows the ground truth segmentation label, and sub-figures

(c) to (h) show the segmentation result of different segmentation methods. The results demonstrate that the conventional method by Rathke et al. [15] could not accurately segment the retinal layers for the B-scan with DME. The method by Chiu et al. [7] 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 [19] prediction is much noisier compared with the MGU-Net [20] and our proposed networks. While MGU-Net [20] yields discontinuity at the OPL denoted by yellow arrows, segmentation of our proposed networks with 2D or 3D input are both continuous at the yellow arrow. We also compare the 3D consistency of the MGU-Net and our proposed networks with simulated motion and our motion correction network by showing the central cross-sectional B-scans in Fig. 4. It can be observed that the simulated motion in sub-figure (1a) is effectively corrected in sub-figure (2a). Overall, our proposed 3D network after motion correction achieves the best consistency, compare to our 2D network at the location denoted by yellow arrow. In sub-figures (2c) and (2e), the 3D version of both MGU-Net and our proposed network after motion correction generate smoother segmentation compared to the 2D versions in (2b) and (2d). When the input is motion corrupted in row (1), the 3D networks are not improved compared with the 2D versions. It demonstrates the motion correction can improve the performance of 3D segmentation.

Quantitative evaluation on the DME dataset [7] is shown in Table 1. The best performance in each column for original or simulated motion input is denoted by bold text, and the second best is denoted by blue text. On the original input, our proposed 3D segmentation network achieves the lowest error of 1.92%, layer error of 10.11%, and average Dice loss of 0.1218, which is reduced by 10% compared to the previous best method MGU-Net [20] at 0.1359. The Dice loss of our 3D network is also the lowest in the each retinal layer, except for the ISE and OS-RPE layer where it ranks as the second and third lowest respectively. For input with simulated motion, our proposed 3D segmentation network with motion correction also achieves the lowest error and average Dice loss overall. Note that the error of both 3D networks increase with simulated motion and no motion correction method compared with the 2D versions. The performance of several methods on the AMD and control dataset [23] are shown in Table 2, where our proposed network with 5 input channels achieves the lowest error and Dice loss after motion correction. 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.

5. CONCLUSION

In this paper, we proposed to combine motion correction and 3D OCT layer segmentation, which led to promising improvement upon existing 2D segmentation methods with or without motion correction. Experimental results demonstrated that the motion correction is required to apply 3D segmentation, and combining motion correction with 3D segmentation achieved the highest performance. In future work, the segmentation network can be regularized to perform topology-guaranteed segmentation, and can be extended to support retinal fluid segmentation. The proposed method will enable more accurate analysis of layer

thickness and vessel density, which is beneficial for diagnosing and monitoring retinal and systemic diseases.

Acknowledgments

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]. Fujimoto James G, Drexler Wolfgang, Schuman Joel S, and Hitzenberger Christoph K, "Optical coherence tomography (oct) in ophthalmology: introduction.," Optics express, vol. 17, no. 5, pp. 3978–3979, 2009. [PubMed: 19259239]

[4]. Carass Aaron, Lang Andrew, Hauser Matthew, Calabresi Peter A, Ying Howard S, and Prince Jerry L, "Multiple-object geometric deformable model for segmentation of macular oct," Biomedical optics express, vol. 5, no. 4, pp. 1062–1074, 2014. [PubMed: 24761289]

[5]. Novosel Jelena, Thepass Gijs, Lemij Hans G, de Boer Johannes F, Vermeer Koenraad A, and van Vliet Lucas J, "Loosely coupled level sets for simultaneous 3d retinal layer segmentation in optical coherence tomography," Medical image analysis, vol. 26, no. 1, pp. 146–158, 2015. [PubMed: 26401595]

[6]. Chiu Stephanie J, Li Xiao T, Nicholas Peter, Toth Cynthia A, Izatt Joseph A, and Farsiu Sina, "Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation,' Optics express, vol. 18, no. 18, pp. 19413–19428, 2010. [PubMed: 20940837]

[7]. Chiu Stephanie J, Allingham Michael J, Mettu Priyatham S, Cousins Scott W, Izatt Joseph A, and Farsiu Sina, "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," Biomedical optics express, vol. 6, no. 4, pp. 1172–1194, 2015. [PubMed: 25909003]

[8]. 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.

[9]. Fang Leyuan, Cunefare David, Wang Chong, Robyn H Guymer Shutao Li, and Farsiu Sina, "Automatic segmentation of nine retinal layer boundaries in oct images of non-exudative amd patients using deep learning and graph search," Biomedical optics express, vol. 8, no. 5, pp. 2732–2744, 2017. [PubMed: 28663902]

[10]. De Fauw Jeffrey, Ledsam Joseph R, Romera-Paredes Bernardino, Nikolov Stanislav, Tomasev Nenad, Blackwell Sam, Askham Harry, Glorot Xavier, O'Donoghue Brendan, Visentin Daniel, et al. , "Clinically applicable deep learning for diagnosis and referral in retinal disease," Nature medicine, vol. 24, no. 9, pp. 1342–1350, 2018.

[11]. He Yufan, Carass Aaron, Yun Yeyi, Zhao Can, Jedynak Bruno M, Solomon Sharon D, Saidha Shiv, Calabresi Peter A, and Prince Jerry L, "Towards topological correct segmentation of macular oct from cascaded fcns," in Fetal, Infant and Ophthalmic Medical Image Analysis, pp. 202–209. Springer, 2017.

[12]. He Yufan, Carass Aaron, Liu Yihao, Jedynak Bruno M, Solomon Sharon D, Saidha Shiv, Calabresi Peter A, and Prince Jerry L, "Structured layer surface segmentation for retina oct using fully convolutional regression networks,' Medical image analysis, vol. 68, pp. 101856, 2021. [PubMed: 33260113]

[13]. Pekala Mike, Joshi Neil, Liu TY Alvin, Bressler Neil M, DeBuc D Cabrera, and Burlina Philippe, "Deep learning based retinal oct segmentation," Computers in biology and medicine, vol. 114, pp. 103445, 2019. [PubMed: 31561100]

[14]. 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.

[15]. Rathke Fabian, Schmidt Stefan, and Schnörr Christoph, "Probabilistic intra-retinal layer segmentation in 3-d oct images using global shape regularization," Medical image analysis, vol. 18, no. 5, pp. 781–794, 2014. [PubMed: 24835184]

[16]. Chiu Stephanie J, Toth Cynthia A, Rickman Catherine Bowes, Izatt Joseph A, and Farsiu Sina, "Automatic segmentation of closed-contour features in ophthalmic images using graph theory and dynamic programming," Biomedical optics express, vol. 3, no. 5, pp. 1127–1140, 2012. [PubMed: 22567602]

[17]. Garvin Mona K, Abràmoff Michael D, Kardon Randy, Russell Stephen R, Wu Xiaodong, and Sonka Milan, "Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-d graph search," IEEE transactions on medical imaging, vol. 27, no. 10, pp. 1495–1505, 2008. [PubMed: 18815101]

[18]. Garvin Mona Kathryn, Abràmoff Michael David, Wu Xiaodong, Russell Stephen R, Burns Trudy L, and Sonka Milan, "Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images," IEEE transactions on medical imaging, vol. 28, no. 9, pp. 1436–1447, 2009. [PubMed: 19278927]

[19]. Roy Abhijit Guha, Conjeti Sailesh, Karri Sri Phani Krishna, Sheet Debdoot, Katouzian Amin, Wachinger Christian, and Navab Nassir, "Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks,' Biomedical optics express, vol. 8, no. 8, pp. 3627–3642, 2017. [PubMed: 28856040]

[20]. Li Jiaxuan, Jin Peiyao, Zhu Jianfeng, Zou Haidong, Xu Xun, Tang Min, Zhou Minwen, Gan Yu, He Jiangnan, Ling Yuye, et al. , "Multi-scale gcn-assisted two-stage network for joint segmentation of retinal layers and discs in peripapillary oct images," Biomedical Optics Express, vol. 12, no. 4, pp. 2204–2220, 2021. [PubMed: 33996224]

[21]. Wang Yiqian, Warter Alexandra, Cavichini-Cordeiro Melina, Freeman William R, Bartsch Dirk-Uwe G, Nguyen Truong Q, and An Cheolhong, "Learning to correct axial motion in oct for 3d retinal imaging," in 2021 IEEE International Conference on Image Processing (ICIP). IEEE, 2021, pp. 126–130.

[22]. Chen Yunpeng, Rohrbach Marcus, Yan Zhicheng, Shuicheng Yan, Feng Jiashi, and Kalantidis Yannis, "Graph-based global reasoning networks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 433–442.

[23]. Farsiu Sina, Chiu Stephanie J, O'Connell Rachelle V, Folgar Francisco A, Yuan Eric, Izatt Joseph A, Toth Cynthia A, Age-Related Eye Disease Study 2 Ancillary Spectral Domain Optical Coherence Tomography Study Group, et al. , "Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography," Ophthalmology, vol. 121, no. 1, pp. 162–172, 2014. [PubMed: 23993787]
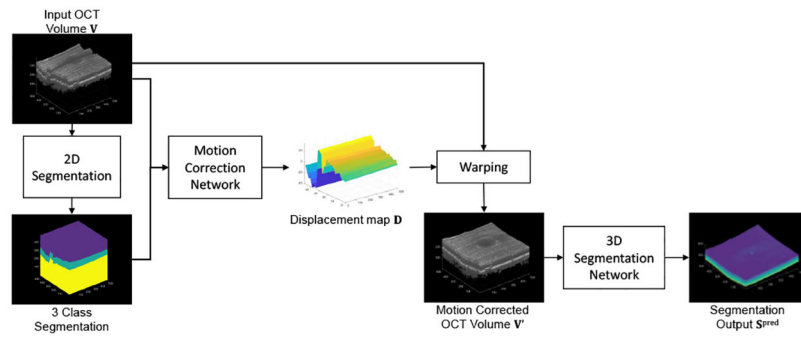
Fig. 1:
Proposed OCT motion correction and segmentation pipeline.
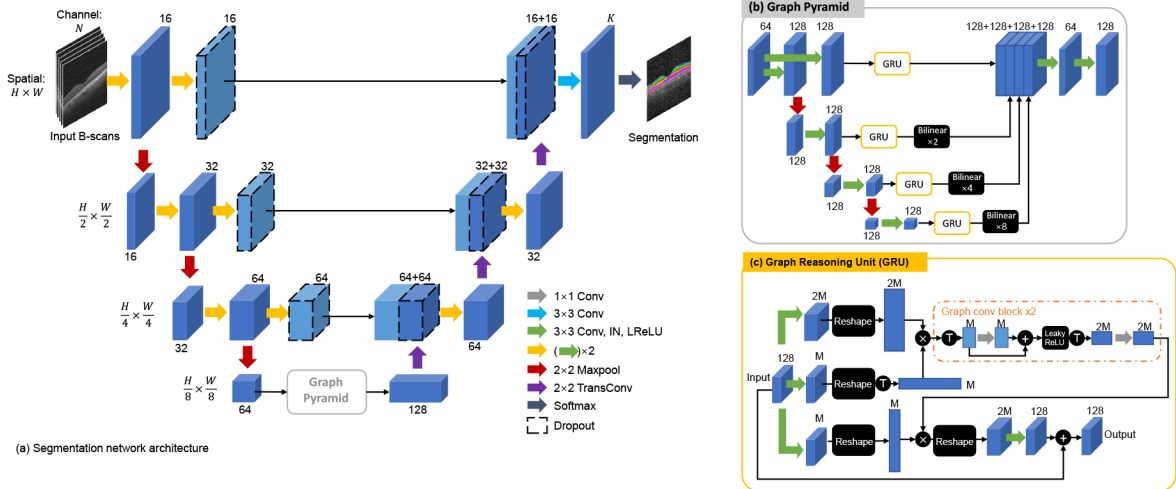
Fig. 2:
(a) Architecture of the proposed OCT segmentation network, (b) the graph pyramid, (c) the graph reasoning unit. Here "IN" operation denotes Instance Normalization, "T" in black circle denotes transpose, and "×" in black circle denotes matrix multiplication.
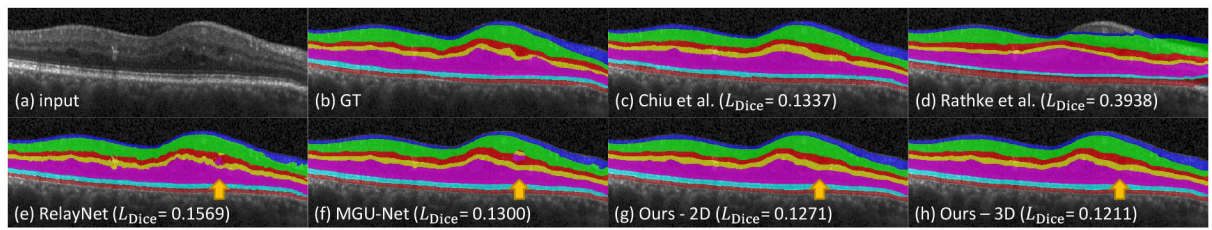
Fig. 3:
Qualitative results on the DME dataset [7]. (a) Input B-scan, (b) ground truth segmentation, (c) Chiu et al. [7], (d) Rathke et al. [15], (e) RelayNet [19], (f) MGU-Net [20], (g) our proposed 2D network, (h) our proposed 3D network.
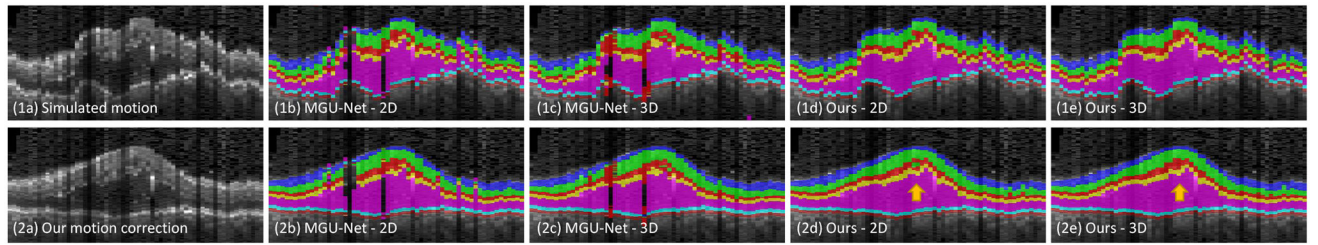
Fig. 4:
Qualitative comparison of 3D consistency on the DME dataset [7]. (1a) Cross-sectional B-scan for OCT volume with simulate motion, (b) ground truth segmentation, (b)-(e) segmentation result of MGU-Net [20] and our proposed network with 2D or 3D input.

Table 1:

Quantitative result of different segmentation methods on the last 55 B-scans from DME dataset [7] where the best and the second best are denoted by bold text and blue text, respectively.

Data	Method	Error	Layer Error	Mean Dice Loss	Dice Loss per Layer						
					RNFL	GCL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE
Original	Chiu et al. [7]	2.55%	13.46%	0.1616 (±0.063)	0.1490	0.1059	0.2439	0.2510	0.0694	0.1317	0.1805
	Rathkeetal. [15]	5.42%	29.82%	0.3279 (±0.090)	0.2935	0.2518	0.4101	0.4368	0.1687	0.4051	0.3290
	RelayNet [19]	3.18%	16.86%	0.1997 (±0.067)	0.2203	0.1412	0.2854	0.2689	0.0809	0.1732	0.2279
	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**
	MGU-Net (3D)	2.03%	10.65%	0.1308 (±0.050)	0.1164	0.0846	0.1917	0.2019	0.0549	0.1124	0.1534
	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)	**1.92%**	**10.11%**	**0.1218 (±0.047)**	**0.1098**	**0.0834**	**0.1850**	**0.1905**	**0.0524**	0.1010	0.1301
Simulated Motion	MGU-Net [20]	2.33%	12.36%	0.1478 (±0.059)	0.1479	0.1151	0.2380	0.2291	0.0623	0.1108	0.1317
	MGU-Net (3D)	2.66%	13.95%	0.1681 (±0.058)	0.1493	0.1125	0.2329	0.2390	0.0735	0.1618	0.2079
	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)	1.96%	10.25%	0.1243 (±0.050)	0.1123	0.0889	0.1955	0.1966	0.0524	0.1003	0.1241
Motion Corrected	MGU-Net [20]	2.22%	11.97%	0.1423 (±0.057)	0.1394	0.1096	0.2269	0.2219	0.0581	0.1099	0.1303
	MGU-Net (3D)	2.08%	10.68%	0.1326 (±0.053)	0.1167	0.0858	0.2041	0.2072	0.0555	0.1123	0.1470
	Ours (2D)	1.91%	10.08%	0.1214 (±0.047)	0.1158	0.0847	**0.1830**	0.1926	0.0514	0.1003	0.1218
	Ours (3D)	**1.89%**	**9.93%**	**0.1201 (±0.048)**	**0.1070**	**0.0838**	0.1890	**0.1888**	**0.0509**	0.1006	0.1210

Table 2:

Quantitative result of different methods on the last 55 volumes from AMD and control dataset [23], where the best and the second best are denoted by bold and blue text, respectively.

| Data | Method | Error | Layer Error | Dice Loss per Layer | |
				RNFL-OS	RPE
Original	MGU-Net [20]	0.56%	2.03%	0.0111	0.1061
	Ours (2D)	0.49%	1.82%	0.0093	0.1022
	Ours (3D)	0.53%	1.88%	0.0099	0.1083
Motion Corrected	MGU-Net [20]	0.45%	1.79%	0.0087	0.0976
	Ours (2D)	0.44%	1.61%	**0.0082**	0.0947
	Ours (3D)	**0.43%**	**1.60%**	**0.0082**	**0.0938**