A 3D Channel-Guided Layer Attention Network for Layer-Specific Structural Biomarker Analysis in Optical Coherence Tomography

1st Mane Pooja
Electronics and Electrical Engg.
IIT Guwahati
Assam, India
mpooja@iitg.ac.in

2nd Manish Bhatt
Electronics and Electrical Engg.
IIT Guwahati
Assam, India
manishb@iitg.ac.in

3rd Samarendra Dandapat
Electronics and Electrical Engg.
IIT Guwahati
Assam, India
samaren@iitg.ac.in

***Abstract*—Optical Coherence Tomography (OCT) provides high-resolution volumetric imaging of retinal layers, enabling quantitative analysis for early detection of diabetic retinopathy (DR). In this work, we propose a unified framework for automated retinal disease analysis that integrates volumetric segmentation and biomarker-based classification. A 3D encoder–decoder network enhanced with a novel Channel-Guided Layer Attention (CGLA) module is introduced for precise multi-class retinal layer segmentation. Following segmentation, layer-specific slabs corresponding to the Internal Limiting Membrane (ILM), Inner Plexiform Layer (IPL), and Outer Plexiform Layer (OPL) are extracted to compute quantitative layer-specific biomarkers, namely volume, thickness, and curvature, for the classification of normal and DR cases. Experimental results on the OCTA-500 dataset demonstrate that the proposed method achieves superior segmentation performance, with accuracies of 0.9819 (ILM), 0.9798 (IPL), and 0.9830 (OPL), compared to 3D UNet-based approaches, along with strong discriminative capability for disease classification, achieving a ROC-AUC of 84.70%.**

***Index Terms*—Optical Coherence Tomography, Retinal Layer Segmentation, 3D Deep Learning, Diabetic Retinopathy.**

I. Introduction

Optical Coherence Tomography (OCT) is a non-invasive imaging modality that provides high-resolution cross-sectional and volumetric visualization of retinal layers, enabling early detection and monitoring of retinal diseases such as diabetic retinopathy (DR), age-related macular degeneration, and diabetic macular edema [1], [2]. In particular, three-dimensional (3D) OCT volumes offer rich structural information for assessing retinal layer integrity. However, automated analysis remains challenging due to speckle noise, low inter-layer contrast, anatomical variability, and the difficulty of capturing subtle layer-specific structural changes. Retinal layer segmentation is a key step in quantitative OCT analysis, where thickness measurements between anatomical boundaries such as the Internal Limiting Membrane (ILM), Inner Plexiform Layer (IPL), and Outer Plexiform Layer (OPL) serve as important biomarkers [3]. Deep learning approaches, particularly UNet [4] and its variants, have significantly improved segmentation performance. Extensions such as ReLayNet [5] and attention-based models [6] further enhance feature representation. However, most methods operate on 2D B-scans, limiting volumetric consistency. While 3D CNN-based approaches improve spatial coherence [7], [8], they often lack explicit modeling of layer-oriented spatial dependencies, which is critical for capturing the structural continuity of retinal layers, especially under pathological conditions. Additionally, most existing works focus solely on segmentation, with limited integration of downstream clinical analysis.

Beyond segmentation, quantitative characterization of retinal layers using structural biomarkers such as thickness, volume, and morphological variations has shown strong correlation with disease progression [3], [9]. Layer-specific slab representations further enable localized analysis of anatomically meaningful regions, improving sensitivity to subtle pathological changes. However, the joint utilization of volumetric segmentation, slab extraction, and biomarker-driven classification remains underexplored.

To address these challenges, this work proposes a unified framework for automated OCT-based retinal disease analysis. A 3D encoder-decoder network enhanced with a novel Channel-Guided Layer Attention (CGLA) module is introduced for accurate multi-class retinal layer segmentation. The proposed attention module captures channel-wise feature importance along with direction-aware spatial dependencies, explicitly modeling horizontal intra-layer continuity and vertical inter-layer transitions. Subsequently, anatomically guided slab extraction is performed for key retinal layers, including ILM, IPL, and OPL, followed by the extraction of quantitative biomarkers such as volume, thickness, and curvature for the classification of normal and DR cases. The overview of the proposed framework is presented in Fig. 1, which enables accurate and interpretable analysis of retinal structures for automated DR diagnosis. Unlike prior studies, which have not investigated 3D layer thickness, volume, and curvature as biomarkers, this work systematically explores these features and demonstrates their potential for improved DR detection in OCT imaging.

The main contributions of this work are summarized as follows:

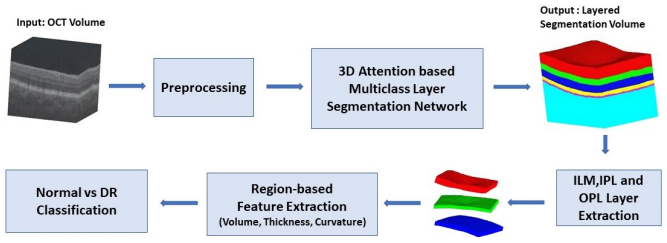

Fig. 1. Overview of the proposed framework. A 3D OCT volume is segmented using a CGLA-enhanced network to obtain multi-layer retinal maps. Layer-guided slab extraction enables computation of structural biomarkers, which are then used for classification of normal and diabetic retinopathy cases.

1. This work proposes new layer-specific structural biomarkers, namely thickness, volume, and curvature, to facilitate accurate and interpretable analysis of retinal structures for automated DR diagnosis
2. A novel Channel-Guided Layer Attention (CGLA) based 3D deep learning framework is presented for enhanced 3D retinal layer segmentation.
3. A unified pipeline integrating 3D segmentation, layer extraction, and biomarker-based disease classification is developed.

II. Methodology

A. Proposed Framework

This study presents a three-dimensional (3D) neural network for automated multi-class retinal layer segmentation from volumetric optical coherence tomography (OCT) B-scans. The network schematic is illustrated in Fig. 2(a). The proposed model follows an encoder-decoder architecture, where the encoder extracts hierarchical volumetric features, and the decoder reconstructs full-resolution segmentation by assigning voxel-wise class probabilities across multiple retinal layers.

The architecture consists of multiple encoder and decoder stages arranged symmetrically. At the input stage, volumetric OCT data is processed using a double convolution block comprising two $3 \times 3 \times 3$ convolutions, each followed by batch normalization and ReLU activation. Each convolutional block is augmented with a Channel-Guided Layer Attention (CGLA) module, which captures layer-specific feature representation. In the encoder, spatial resolution is progressively reduced using anisotropic max pooling with a kernel size of $1 \times 2 \times 2$, preserving depth resolution while aggregating contextual information along lateral dimensions. This design effectively maintains inter-slice continuity, which is essential for accurate modeling of retinal layer structures.

The decoder mirrors the encoder structure and performs feature reconstruction using transposed convolutions with a stride of $1 \times 2 \times 2$. Skip connections are employed to fuse encoder and decoder features at corresponding resolutions, enabling recovery of fine-grained spatial details. Each decoder block also incorporates the CGLA module to refine features during upsampling. Finally, a $1 \times 1 \times 1$ convolutional layer produces voxel-wise class scores, which are converted into probabilities using a softmax function for multi-class retinal layer prediction.

1) Channel-Guided Layer Attention (CGLA): The proposed Channel-Guided Layer Attention (CGLA) module, shown in Fig. 2(b), combines channel-wise global attention with direction-aware spatial attention to explicitly model horizontal intra-layer continuity and vertical inter-layer transitions, unlike conventional isotropic attention mechanisms. Given an input feature map $X \in \mathbb{R}^{C \times D \times H \times W}$, channel attention is first applied via global average pooling followed by two $1 \times 1 \times 1$ convolutions with ReLU and sigmoid activations to produce channel-wise weights that recalibrate feature responses. Subsequently, layer-oriented spatial attention is applied using depth-wise separable convolutions with anisotropic kernels $(1 \times k \times 1)$ and $(k \times 1 \times 1)$ to capture directional dependencies corresponding to intra-layer continuity and inter-layer transitions. The combined response is normalized and activated to generate a spatial attention map, refining feature representation. This sequential design emphasizes anatomically consistent structures and improves segmentation robustness in low-contrast and structurally disrupted regions.

2) Loss Function for Multi-Class Retinal Layer Segmentation: The network is trained using the focal loss [10] to mitigate class imbalance across multiple retinal layers in OCT segmentation. Unlike the conventional cross-entropy loss, focal loss introduces a modulating factor that reduces the contribution of well-classified voxels and emphasizes harder, misclassified examples. This property is particularly advantageous as certain retinal layers are underrepresented or more difficult to delineate. The focal loss for multi-class segmentation is defined as:

$$\mathcal{L}_{\text{focal}} = -\frac{1}{N} \sum_{i=1}^{N} \alpha_{y_i} \, (1 - p_{t,i})^{\gamma} \, \log(p_{t,i}), \tag{1}$$

where,

- N is the total number of voxels,
- $y_i \in \{1, \ldots, C\}$ denotes the ground-truth class label for voxel i,
- $p_{t,i} = p_{y_i}$ is the predicted probability for the true class,
- $\gamma > 0$ is the focusing parameter that controls the strength of down-weighting for easy examples,

In our implementation, the focal loss is employed with a focusing parameter $\gamma = 2.0$, enabling the network to concentrate on hard-to-classify voxels while diminishing the relative contribution of well-classified samples. The final training loss is computed by averaging the focal loss over all voxels within each 3D OCT volume.

The proposed network is optimized using Adam with an initial learning rate of 1×10^{-3}. Training is performed for up to 50 epochs with a batch size of 2, and early stopping with a patience of 8 epochs is employed based on validation loss.

B. Layer Feature Extraction and Classification

Quantitative retinal layer descriptors, including volume, mean thickness, and curvature (top boundary, bottom boundary, and thickness profile), are extracted from the ILM, IPL,

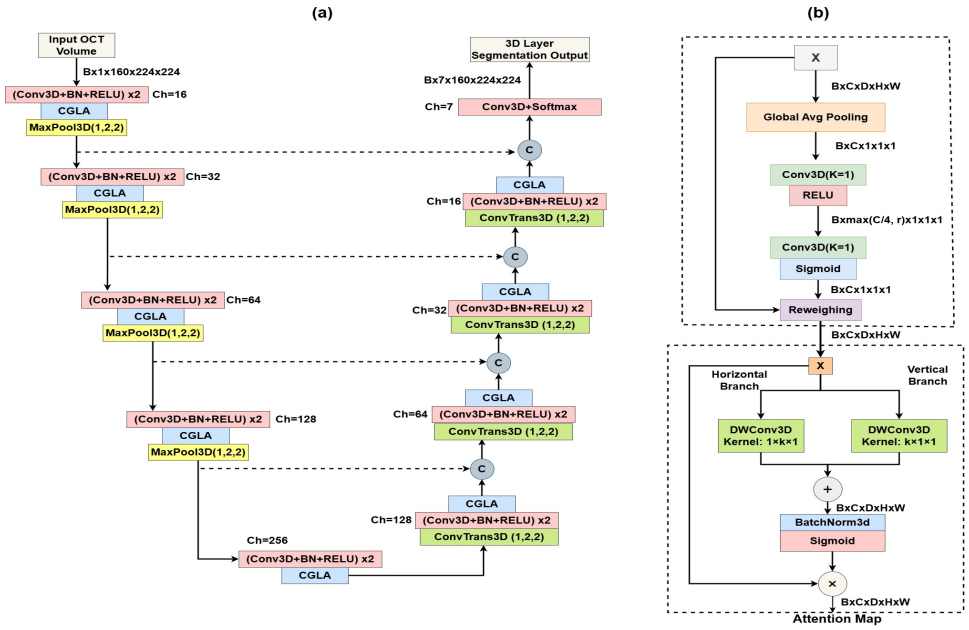

Fig. 2. (a) Architecture of the proposed Channel Guided Layer Attention (CGLA) enhanced neural network; (b) detailed illustration of the Channel Guided Layer Attention (CGLA) block.

and OPL layers using a region-based analysis. Each layer is partitioned into three regions to capture localized structural variations.

The key features are defined as:

$$V = \sum_{z,y,x} \mathbf{1}(z,y,x) \quad \kappa = \frac{|y''|}{(1+(y')^2)^{3/2}} \tag{2}$$

$$T = \frac{1}{N} \sum_{x} \left(y_{\text{bottom}}(x) - y_{\text{top}}(x) \right) \tag{3}$$

where V denotes layer volume (number of voxels), T represents mean layer thickness, y_{top} and y_{bottom} denote the top and bottom boundaries, and κ represents curvature.

The extracted features from all layers and regions are concatenated to form subject-wise feature vectors, which are paired with class labels (Normal and diabetic retinopathy (DR)). A logistic regression classifier with class-balanced weighting is employed for classification. Model performance is evaluated using stratified 5-fold cross-validation, reporting ROC-AUC.

C. Dataset and Preprocessing

This study utilizes the OCTA-500 dataset [11] consisting of OCT volumes and 3D layer ground-truth annotations. For each B-scan, region masks are generated by filling areas between annotated layer boundaries, with each region assigned a distinct label (Fig. 3). The 3×3 mm^2 subset was selected, representing the central macular region with a field of view of 3 mm $\times$ 3 mm and a depth of 2 mm. Each volume comprises 304 sequential B-scans of size 640×304 pixels, resulting in volumetric data of $304 \times 640 \times 304$ voxels. The dataset contains 189 volumes, including 160 Normal and 29 diabetic

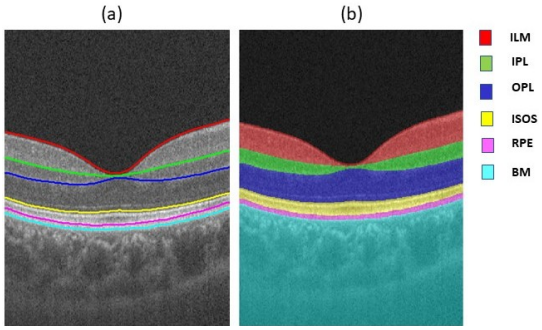

Fig. 3. (a) A single slice from an OCT volume with annotated layer boundaries; (b) a single slice from an OCT volume with annotated layer region boundaries.

retinopathy (DR) cases. Prior to augmentation, the data were divided into 113 training volumes (96 Normal and 17 DR) and 76 testing volumes (64 Normal and 12 DR). To improve model generalization, horizontal flipping was applied to the training set, increasing its size to 226 volumes. All volumes were preprocessed prior to training by applying intensity normalization, followed by cropping to $304 \times 304 \times 304$ to retain the central region of interest, and resizing to $160 \times 224 \times 224$ to meet the network input requirements.

III. Results and Discussion

A. Multi-class Layer Segmentation Performance

The effectiveness of the proposed Channel-Guided Layer Attention (CGLA) module is evaluated by comparing it with a standard 3D UNet (without attention) [4] and state-of-the-art attention-based variants, including Attention Gate [6] UNet (UNetA), UNet with Squeeze-and-Excitation (SE) [12]

TABLE I
COMPARISON OF 3D LAYER-WISE SEGMENTATION PERFORMANCE OF DIFFERENT MODELS ON THE OCTA-500 DATASET.

Layer	Metric	UNet	UNetA	UNetSE	UNetCBAM	Proposed
ILM	Dice	0.8483 ± 0.0732	0.8443 ± 0.0668	0.8269 ± 0.0760	0.8506 ± 0.0648	**0.9132 ± 0.0472**
	IoU	0.7426 ± 0.0959	0.7356 ± 0.0875	0.7111 ± 0.0975	0.7451 ± 0.0903	**0.8434 ± 0.0721**
	Precision	0.8264 ± 0.0832	0.8529 ± 0.0856	0.7944 ± 0.0845	0.8521 ± 0.0772	**0.9106 ± 0.0582**
	Recall	0.8739 ± 0.0726	0.8384 ± 0.0571	0.8649 ± 0.0751	0.8520 ± 0.0646	**0.9177 ± 0.0479**
	Accuracy	0.9675 ± 0.0167	0.9676 ± 0.0157	0.9625 ± 0.0176	0.9688 ± 0.0148	**0.9819 ± 0.0104**
IPL	Dice	0.7956 ± 0.1084	0.7977 ± 0.1018	0.7687 ± 0.0927	0.7824 ± 0.1103	**0.8475 ± 0.0832**
	IoU	0.6714 ± 0.1222	0.6736 ± 0.1208	0.6327 ± 0.1121	0.6544 ± 0.1322	**0.7432 ± 0.1096**
	Precision	0.7846 ± 0.1077	0.7681 ± 0.0988	0.7738 ± 0.0980	0.7790 ± 0.1076	**0.8562 ± 0.0821**
	Recall	0.8098 ± 0.1179	0.8322 ± 0.1136	0.7663 ± 0.0976	0.7886 ± 0.1194	**0.8411 ± 0.0917**
	Accuracy	0.9725 ± 0.0157	0.9722 ± 0.0151	0.9696 ± 0.0138	0.9710 ± 0.0166	**0.9798 ± 0.0125**
OPL	Dice	0.8824 ± 0.0852	0.8741 ± 0.0887	0.8715 ± 0.0781	0.8729 ± 0.0848	**0.9162 ± 0.0555**
	IoU	0.7981 ± 0.1128	0.7856 ± 0.1177	0.7796 ± 0.1059	0.7832 ± 0.1177	**0.8496 ± 0.0832**
	Precision	0.8695 ± 0.1034	0.8827 ± 0.1044	0.8459 ± 0.0916	0.8596 ± 0.1012	**0.9139 ± 0.0676**
	Recall	0.8988 ± 0.0737	0.8683 ± 0.0827	0.9021 ± 0.0747	0.8900 ± 0.0789	**0.9204 ± 0.0553**
	Accuracy	0.9758 ± 0.0173	0.9749 ± 0.0170	0.9734 ± 0.0154	0.9737 ± 0.0178	**0.9830 ± 0.0110**
ISOS	Dice	0.7972 ± 0.1576	0.7956 ± 0.1630	0.7895 ± 0.1531	0.7829 ± 0.1681	**0.8782 ± 0.0957**
	IoU	0.6852 ± 0.1741	0.6842 ± 0.1766	0.6739 ± 0.1740	0.6691 ± 0.1890	**0.7935 ± 0.1264**
	Precision	0.8010 ± 0.1658	0.7906 ± 0.1695	0.7775 ± 0.1579	0.8022 ± 0.1796	**0.8629 ± 0.1020**
	Recall	0.7956 ± 0.1540	0.8029 ± 0.1607	0.8035 ± 0.1510	0.7663 ± 0.1603	**0.8954 ± 0.0945**
	Accuracy	0.9823 ± 0.0129	0.9820 ± 0.0133	0.9812 ± 0.0129	0.9814 ± 0.0136	**0.9892 ± 0.0078**
RPE	Dice	0.6006 ± 0.1912	0.6023 ± 0.1844	0.5810 ± 0.1947	0.5952 ± 0.2085	**0.7536 ± 0.1537**
	IoU	0.4533 ± 0.1781	0.4530 ± 0.1689	0.4341 ± 0.1817	0.4521 ± 0.1939	**0.6249 ± 0.1652**
	Precision	0.6239 ± 0.1974	0.6276 ± 0.1963	0.6057 ± 0.2091	0.6041 ± 0.2058	**0.7573 ± 0.1601**
	Recall	0.5817 ± 0.1894	0.5811 ± 0.1775	0.5606 ± 0.1864	0.5897 ± 0.2140	**0.7518 ± 0.1528**
	Accuracy	0.9839 ± 0.0080	0.9838 ± 0.0081	0.9829 ± 0.0087	0.9834 ± 0.0086	**0.9895 ± 0.0072**
BM	Dice	0.9870 ± 0.0115	0.9868 ± 0.0121	0.9867 ± 0.0106	0.9864 ± 0.0139	**0.9925 ± 0.0085**
	IoU	0.9746 ± 0.0217	0.9743 ± 0.0227	0.9739 ± 0.0202	0.9734 ± 0.0259	**0.9852 ± 0.0161**
	Precision	0.9871 ± 0.0137	0.9878 ± 0.0119	0.9867 ± 0.0141	0.9843 ± 0.0207	**0.9930 ± 0.0107**
	Recall	0.9873 ± 0.0216	0.9864 ± 0.0237	0.9870 ± 0.0192	0.9889 ± 0.0183	**0.9920 ± 0.0116**
	Accuracy	0.9900 ± 0.0086	0.9899 ± 0.0089	0.9897 ± 0.0081	0.9893 ± 0.0111	**0.9942 ± 0.0065**

(UNetSE), and UNet with Convolutional Block Attention Modules (CBAM) [13] (UNetCBAM). As shown in Table I, the proposed CGLA-enhanced network consistently outperforms all baselines across all retinal layers, demonstrating the effectiveness of the proposed attention design. For ILM segmentation, the proposed model improves the Dice score by 6.49% over UNet and 8.63% over UNetSE, with corresponding IoU gains of 10.08% and 13.23%. In IPL, Dice improves by 5.19% compared to UNet and 7.88% compared to UNetSE, indicating enhanced sensitivity in intermediate layers. Similar improvements are observed for OPL and ISOS, with Dice gains of 3.38% and 8.10%, respectively. The most significant improvement is observed in RPE segmentation, where Dice increases by 15.30% over UNet, highlighting the ability of CGLA to handle low-contrast and ambiguous boundaries. Even for BM, where all models perform strongly, the proposed method achieves a consistent Dice improvement of 0.55%. Precision and recall exhibit similar trends, with recall gains of up to 11.38% in ISOS and 17.01% in RPE, underscoring improved sensitivity to thin and noisy boundaries. Overall, the proposed network achieves superior volumetric segmentation performance while maintaining structural consistency across all retinal layers.

Accurate 3D retinal OCT layer segmentation remains challenging due to low inter-layer contrast, subtle boundary transitions, and OCT-specific noise. Figure 4 presents a qualitative comparison of predicted layer maps and corresponding pixel-wise mis-segmentation across different models. The baseline UNet, UNetA, UNetSE, and UNetCBAM exhibit noticeable boundary leakage and depth-wise inconsistencies, particularly in the foveal region and around the ILM-IPL and IPL-OPL interfaces. In contrast, the proposed network produces fewer and more localized errors, primarily confined to thin boundary regions, while preserving structural continuity across layers.

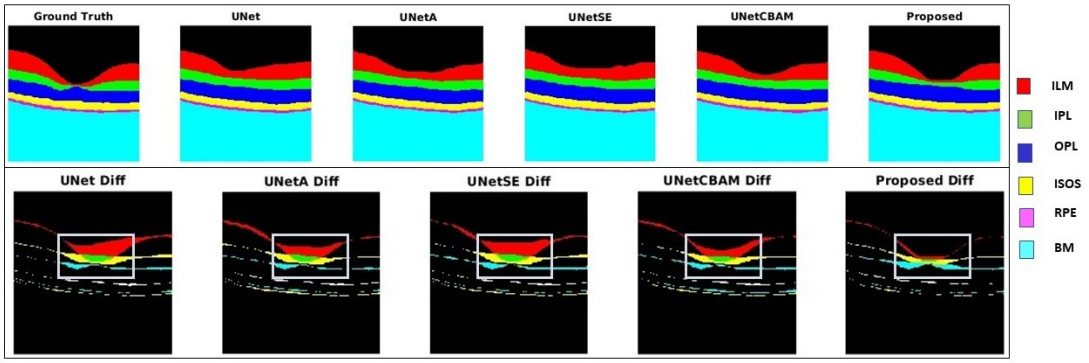

Fig. 4. Qualitative comparison of predicted layer produced by UNet, UNetA, UNetSE, UNetCBAM, and the proposed network with the ground truth (GT) shown in the first row. The second row illustrates pixel-wise mis-segmentation maps between the GT and the corresponding predictions for UNet, UNetA, UNetSE, UNetCBAM, and the proposed network.

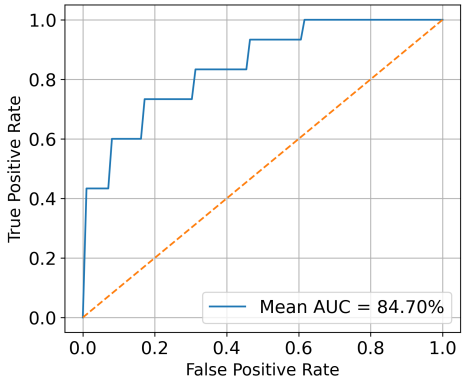

Fig. 5. ROC curve illustrating logistic regression classification of Normal vs. DR eyes using 3D layer structural features.

This demonstrates improved robustness to noise and enhanced anatomical consistency.

B. FAZ Feature-Based Classification

Structural features extracted from segmented 3D retinal layer volumes, including voxel-based volume, thickness, and curvature descriptors, are used for diabetic retinopathy (DR) classification. As shown in Fig. 5, the logistic regression classifier achieves strong discriminative performance, with an AUC-ROC of 84.70%, indicating an effective sensitivity–specificity trade-off.

IV. Conclusion

This work presents a CGLA-enhanced 3D network for automated retinal layer segmentation from OCT volumes, improving structural consistency and segmentation accuracy across all layers. The proposed attention mechanism effectively captures intra-layer continuity and inter-layer transitions, enabling reliable delineation of challenging boundaries. The study demonstrates that accurate volumetric layer segmentation facilitates the extraction of meaningful structural descriptors, including volume, thickness, and curvature. The classifier achieved reliable discrimination between normal and diseased eyes, highlighting the diagnostic relevance of the extracted features. These results indicate that the proposed framework enhances OCT-based analysis by enabling robust volumetric characterization of retinal morphology, supporting improved disease assessment.

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, C. A. Puliafito *et al.*, "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[2] J. G. Fujimoto, "Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy," *Neuroradiology*, vol. 50, no. 9, pp. 711–715, 2008.

[3] F. M. Heussen *et al.*, "Measurement of retinal thickness in spectral-domain oct," *Ophthalmology*, vol. 115, no. 7, pp. 1236–1241, 2008.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.

[5] A. Roy, S. Conjeti, S. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab, "Relaynet: Retinal layer and fluid segmentation of macular oct images using fully convolutional networks," *Biomedical Optics Express*, vol. 8, no. 8, pp. 3627–3642, 2017.

[6] O. Oktay, J. Schlemper, L. Le Folgoc *et al.*, "Attention u-net: Learning where to look for the pancreas," in *Medical Imaging with Deep Learning (MIDL)*, 2018.

[7] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "3d fully convolutional network for retinal oct layer segmentation," *IEEE Access*, vol. 7, pp. 19 205–19 216, 2019.

[8] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in oct images using a 3d deep learning method," in *SPIE Medical Imaging*, 2019.

[9] J. Son *et al.*, "Retinal layer thickness changes in diabetic retinopathy," *Investigative Ophthalmology & Visual Science*, vol. 55, no. 3, pp. 1632–1640, 2014.

[10] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.

[11] M. Li, S. Yuan, and Q. Chen, "Octa-500," 2019. [Online]. Available: https://dx.doi.org/10.1016/j.media.2024.103092

[12] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[13] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.