Biomedical Signal Processing and Control 121 (2026) 110250

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

DLSANet: A dual-path learnable structure-prior attention network for retinal layer segmentation

Enyu Liu [a,b], Muhao Xu [a,b], Haohua Yang [a,b], Bingcan Yan [a,b], Hua Wei [a,b], Weiye Song [a,b,*]

[a] School of Mechanical Engineering, Shandong University, Jinan, 250061, China
[b] Key Laboratory of High Efficiency and Clean Mechanical Manufacture of Ministry of Education, Shandong University, Jinan, 250061, China

ARTICLE INFO

Keywords:
Retinal layer segmentation
OCT
Dual-path attention network
Learnable structure prior

ABSTRACT

Automatic segmentation of retinal layers in optical coherence tomography (OCT) images is crucial for the early diagnosis and monitoring of ophthalmic diseases. However, existing methods face two fundamental challenges: the structural incoherence of retinal layer segmentation results due to inadequate modeling of anatomical priors and their susceptibility to speckle noise inherent in OCT imaging. To address these limitations, we propose the Dual-path Learnable Structure-prior Attention Network (DLSANet), which features a novel DLSA module as its core innovation. This module integrates two synergistic components: the Prior Modeling and Guidance component and the Multi-domain Feature Enhancement component. The Prior Modeling and Guidance component explicitly learns constraints derived from retinal anatomical priors through a dual-path structural pattern design; it captures globally consistent structural patterns via the learnable structure branch and adapts to local structural deformations in pathological regions through the deformable structure branch, thereby improving structural consistency during feature fusion. The Multi-domain Feature Enhancement component mitigates speckle noise through frequency-spatial joint processing while preserving fine structural details. Extensive evaluations on a self-collected dataset, Retina500, comprising 500 healthy samples, as well as two public datasets, DME with 110 samples and Glaucoma with 244 samples, demonstrate that DLSANet achieves state-of-the-art performance with mean Intersection over Union (mIoU) values of 82.22%, 70.63%, and 69.08%, respectively, while maintaining superior Accuracy (Acc) and mean Pixel Accuracy (mPA). Our code is available at: https://github.com/ley2024/DLSANet

1. Introduction

Retinal diseases constitute a significant and growing global health challenge, with prevalence rates increasing markedly in recent decades [1]. Epidemiological studies document a clear correlation between modern environmental factors and the rising incidence of conditions such as glaucoma and diabetic retinopathy [2]. The clinical significance of retinal layer morphology is well established in ophthalmology, where specific structural alterations serve as definitive biomarkers for early disease detection [3]. In glaucoma pathogenesis, progressive thinning of the retinal nerve fiber layer represents a characteristic pathological change [4], while diabetic macular edema manifests as abnormal thickening of the macular region [5]. Consequently, precise quantification of retinal layer architecture and monitoring of morphological dynamics provide not only critical metrics for early screening but also enable objective assessment of disease progression and therapeutic outcomes [6]. This analytical approach substantially enhances the standardization and efficacy of ophthalmic diagnostic practices [7].

OCT is established as a clinical gold standard for in vivo retinal layer assessment, owing to its high-resolution and non-invasive imaging capabilities [8]. The structural prior of the retina is fundamentally rooted in its intrinsic physical and optical properties. The higher cell density observed in specific retinal layers results in increased stiffness, as evidenced by an elevated Brillouin modulus, which creates distinct interlayer boundaries [9]. Additionally, the characteristic variation in refractive index enhances the capability of OCT to delineate retinal layers through light interaction [10]. These intrinsic properties collectively establish the structural prior, characterized by ordered layering, stable spacing, and continuity. This alignment with established concepts in medical image segmentation addresses the shortcomings of physical realism in data-driven models [11]. However,

* Corresponding author.
E-mail addresses: 202534541@mail.sdu.edu.cn (E. Liu), ujnmhxu@hotmail.com (M. Xu), 202534560@mail.sdu.edu.cn (H. Yang), 202534584@mail.sdu.edu.cn (B. Yan), weihua@sdu.edu.cn (H. Wei), songweiye@sdu.edu.cn (W. Song).

https://doi.org/10.1016/j.bspc.2026.110250
Received 2 February 2026; Received in revised form 22 March 2026; Accepted 3 April 2026
Available online 11 April 2026

manual segmentation of OCT images by ophthalmologists remains the conventional practice, which is labor-intensive, time-consuming, and prone to inter-observer variability, severely limiting large-scale screening and longitudinal studies [12]. To address these drawbacks, deep learning-based automated segmentation methods have emerged as a promising alternative, with encoder–decoder architectures – exemplified by the generic U-net [13] and its retinal OCT-specialized variant ReLayNet [14] – demonstrating significant potential in enhancing segmentation efficiency and consistency through an end-to-end learning paradigm.

Although deep learning has significantly advanced automated retinal layer segmentation in OCT, achieving structurally coherent results and robust performance against speckle noise remains a challenge [15]. The retina exhibits a strictly ordered layered architecture, where the spatial position and continuity of each layer adhere to well-defined anatomical rules [16]. Recent advancements in technologies, such as dilated convolutions and context aggregation modules, have effectively mitigated the limitations of convolutional-based models in capturing long-range dependencies [17,18]. However, due to the incomplete learning of the anatomical structure of the retina during the feature learning phase, the segmentation results may still exhibit discontinuities and misalignments between layers, as well as anatomically inconsistent cross-layer overlaps [19]. Furthermore, speckle noise, an intrinsic characteristic of OCT imaging, often degrades the contrast between retinal layers and obscures boundary delineation [20]. Previous studies have incorporated anatomical priors through topology-preserving post-processing, shape-regularized loss terms, or additional structural super-vision [21,22]. While these approaches can enhance final segmentation maps, they are typically applied post feature extraction, meaning the model does not genuinely learn anatomical organization during representation learning and remains vulnerable to structural ambiguity and noise interference [23].

To address these limitations, we propose the Dual-path Learnable Structure-prior Attention Network (DLSANet). The core idea is to strategically embed learnable structural patterns guided by inherent retinal anatomical priors directly into the decoding process. Since anatomical constraints are crucial for fusing multi-scale features to restore segmentation details, this design enables the model to maintain retinal layer continuity and ordering during feature reconstruction, rather than relying solely on post-hoc correction. Specifically, the Dual-path Learnable Structure-prior Attention (DLSA) module consists of two synergistic components tailored to address the core challenges. First, the Prior Modeling and Guidance component utilizes a dual-path generator to adaptively balance global anatomical consistency with local structural adaptability. Given the retina's strictly ordered hierarchy in healthy regions and frequent local deformations in pathological areas, this design explicitly encodes topological constraints and hierarchical layer relationships derived from anatomical knowledge, ensuring that structural knowledge guides feature fusion even when boundary cues are weak. Second, the Multi-domain Feature Enhancement unit effectively refines spatial and frequency-domain representations to suppress boundary blurring induced by speckle noise in OCT imaging. This optimization is particularly valuable because speckle noise inherently degrades inter-layer contrast, which facilitates the precise delineation of thin retinal layers. By integrating prior modeling and feature enhancement into an end-to-end trainable pipeline, DLSANet achieves anatomically consistent and noise-resilient retinal layer segmentation with only 4.1M parameters, thereby supporting practical clinical deployment.

In summary, our contributions are:

- We propose DLSANet, a retinal layer segmentation framework that embeds learnable structural patterns guided by retinal anatomical priors directly into the decoding stage to strengthen structural coherence during representation learning.
- We design the DLSA module, which integrates a Prior Modeling and Guidance component and a Multi-domain Feature Enhancement unit to enforce anatomical consistency and robustness to noise and artifacts.
- DLSANet achieves state-of-the-art performance with only 4.1M parameters across healthy and pathological datasets, demonstrating strong potential for clinical deployment.

2. Related works

This section systematically examines pivotal advancements in retinal OCT layer segmentation, with a focused analysis organized along four principal research trajectories: traditional feature-engineering approaches, segmentation models based on deep learning, and structure-prior-guided segmentation strategies, and alternative structural modeling paradigms.

2.1. Traditional segmentation methods

Prior to the advent of deep learning, retinal OCT layer segmentation primarily relied on hand-crafted feature extraction pipelines. At the feature design level, researchers developed various low-level image descriptors to characterize layer boundary properties, including gradient-based edge detectors, texture statistic-based regional descriptors, and feature enhancement methods incorporating prior knowledge. From an optimization perspective, Garvin et al. formulated the multi-surface detection problem as a graph search task, integrating boundary strength and inter-layer thickness constraints into a cost function [24]. Chiu et al. conversely, combined kernel regression with graph theory and dynamic programming to improve segmentation efficiency while maintaining the accuracy of retinal layer and fluid region segmentation [25]. Furthermore, Mishra et al. and Yazdanpanah et al. proposed targeted solutions for pathological regions with severe structural distortions, respectively employing a two-step kernel-based optimization scheme integrated with dynamic programming and a multi-phase edge-free active contour method incorporating a circular shape prior and contextual adaptive weights, respectively [26,27]. These traditional methods achieved segmentation accuracy comparable to expert annotations in healthy retinal images, laying a crucial foundation for subsequent research [28]. However, their heavy reliance on hand-crafted features and intricate parameter tuning significantly constrained their generalization capability when confronted with pathological structural distortions, discontinuities, or blurred boundaries [29].

2.2. Deep learning-based segmentation models

The introduction of the encoder–decoder architecture marked a significant turning point, leading to substantial improvements in retinal layer segmentation. The U-net framework, proposed by Ronneberger et al. established the paradigm for end-to-end segmentation through its unique skip connections, which effectively recover spatial detail in the decoder path while capturing abstract semantic features in the encoder [13]. Subsequently, extensive research has been conducted to address key challenges, including feature representation capacity, receptive field expansion, and computational efficiency. He et al. alleviated the vanishing gradient problem in deep networks by introducing residual connections, thereby enhancing feature propagation efficiency [30]. Oktay et al. designed an attention gating mechanism that enables the model to autonomously focus on regions relevant to the segmentation task, effectively suppressing interference from irrelevant backgrounds [31]. To overcome the limitations of convolutional operations in capturing long-range dependencies, Chen et al. pioneered the integration of Transformer modules into the U-net architecture, establishing global contextual relationships via self-attention mechanisms [32]. Building upon this, Huang et al. proposed a PolarFormer, a Transformer-based model, which significantly improved

the capacity for segmenting multi-class vulnerable plaques in intravascular pathological OCT images through Polar-Attention-driven radial hierarchical feature interaction [33]. Concurrently, model efficiency has garnered widespread attention. Various lightweight Transformer architectures are progressively being applied to retinal OCT image analysis tasks, effectively balancing global contextual modeling capabilities with computational efficiency [34,35]. He et al. proposed a lightweight network named LightReSeg that adopts lightweight design strategies to minimize parameter scale while achieving state-of-the-art segmentation accuracy on pathological OCT images including diabetic macular edema and glaucoma [36]. He et al. further explored data-level innovation by integrating multispectral information (MSI) into retinal layer segmentation, developing a dedicated MSI encoder to fuse spectral and structural features [37]. Their work demonstrated consistent accuracy improvements across multiple deep learning frameworks and spectral ranges, effectively mitigating intra-class errors and inter-layer boundary ambiguity caused by over-reliance on single structural information. However, despite their exceptional performance in pixel-wise accuracy, these data-driven approaches still face challenges in ensuring topological continuity and anatomical plausibility in their outputs, primarily due to the lack of explicit modeling of the retina's inherent anatomical structure.

2.3. Structure-prior-guided segmentation strategies

The incorporation of structural priors has been pivotal in advancing retinal layer segmentation performance. Ho et al. proposed LiteMamba-Bound, a lightweight variant of the Mamba architecture. Its innovative Normalized Active Contour (NAC) loss achieves superior segmentation performance in medical image segmentation compared to traditional boundary-aware losses by enhancing the contrast between foreground and background while constraining the smoothness of boundaries [38, 39]. Fu et al. introduced a multi-context network, which processes local details and global contextual information through parallel pathways, demonstrating the efficacy of multi-path architectures in glaucoma screening [40]. Fazekas et al. proposed SD-LayerNet, which pioneers disentangled representation learning for retinal OCT layer segmentation. A fully differentiable topological engine transforms one-dimensional surface position regression into two-dimensional segmentation maps, demonstrating the efficacy of disentangled representations with anatomical priors in enhancing segmentation data efficiency [41]. Building on this foundation, the group introduced the semi-supervised SD-RetinaNet in 2025, which incorporates a differentiable biomarker topology engine and anatomical prior losses for topological consistency, facilitating bidirectional learning between layers and lesions [23]. Recent trends have further solidified this direction, with growing emphasis on embedding structural constraints directly into network architectures. This includes the development of boundary-aware mechanisms that explicitly guide feature learning toward anatomically consistent edges, and the use of adaptive sampling techniques to better model the biological variations in retinal layer morphology [41]. These studies collectively affirm that translating anatomical priors into learnable constraint mechanisms is a highly effective pathway for boosting model performance. Nevertheless, prevailing methods still exhibit shortcomings in achieving the end-to-end collaborative learning of local boundary details and global topological constraints, as well as in maintaining the stability of these priors in noisy environments. These identified limitations precisely define the innovation space explored in this work.

2.4. Alternative structural modeling paradigms

Beyond the explicit embedding of structural priors within deep networks, several alternative modeling paradigms have been explored to address anatomical consistency and structural continuity in medical image segmentation. For instance, Gaussian Process Regression (GPR) models spatial correlations through kernel functions and inherently provides uncertainty estimation, offering theoretical advantages for continuous boundary modeling [42]. Graph Neural Networks (GNNs) explicitly encode inter-layer relationships as node-edge structures, enabling the direct modeling of topological dependencies [43,44]. Ensemble strategies leverage complementary predictions from multiple models and have demonstrated improved robustness against imaging noise and artifacts in various medical imaging tasks [45]. In parallel, structural modeling has increasingly extended toward continuous geometric and dynamical system formulations. Differentiable rendering maintains differentiability in geometric optimization processes, providing a potential mechanism for enforcing shape consistency within end-to-end frameworks [46,47]. Neural Ordinary Differential Equations (Neural ODEs) formulate feature evolution as a continuous-time dynamical system, offering a mathematical perspective for modeling smooth anatomical trajectories [48]. Nevertheless, in high-resolution pixel-wise OCT segmentation, these approaches may encounter practical challenges in computational scalability, numerical stability, and integration with dense prediction architectures. Consequently, achieving a unified framework that preserves structural expressiveness while maintaining computational efficiency and end-to-end trainability remains an open research problem.

3. Proposed method

3.1. Framework overview

We propose DLSANet, a U-shaped encoder–decoder architecture specifically designed for retinal OCT image layer segmentation to address structural incoherence from insufficient anatomical prior modeling and boundary blurring induced by OCT image noise. As illustrated in Fig. 1, the network takes single-channel OCT images as input extracts multi-scale hierarchical features via the encoder enhances global contextual correlations through a Transformer bottleneck layer realizes structured feature fusion and resolution restoration in the decoder with integrated DLSA modules and ultimately outputs pixel-wise retinal layer segmentation maps.

The core innovation lies in the decoder, where our proposed DLSA module guides multi-scale feature fusion. The DLSA module systematically addresses the two key challenges identified in Section 1: (1) it resolves structural incoherence through the Prior Modeling and Guidance component that learns and enforces anatomical constraints, and (2) it enhances robustness to noise and artifacts through the Multi-domain Feature Enhancement component that suppresses degradations while preserving structural details.

Designed for clinical practicality, the network employs computationally optimized operations throughout. Section 3.2 details the encoder–decoder architecture, while Section 3.3 provides the complete DLSA formulation.

3.2. Multi-scale encoder and cross-scale decoder architecture

3.2.1. Multi-scale encoder

The encoder employs a 4-level downsampling architecture to extract multi-scale feature representations with each level integrating a Contracting Block and an AttentionDownsample module whose synergistic design progressively increases the feature channel dimension while reducing spatial resolution to preserve critical structural details. For deep local feature extraction the Contracting Block consists of two stacked sequential stages each comprising 3×3 convolution, BatchNorm normalization (BN) and ReLU activation where the input feature of the Contracting Block denoted as X undergoes two consecutive rounds of this integrated process starting with 3×3 convolution followed by BatchNorm normalization and ReLU activation to generate the final output feature at the kth encoder level referred to as F_k. The AttentionDownsample module generates attention weights via 1×1 convolution to adaptively weight input features before performing downsampling

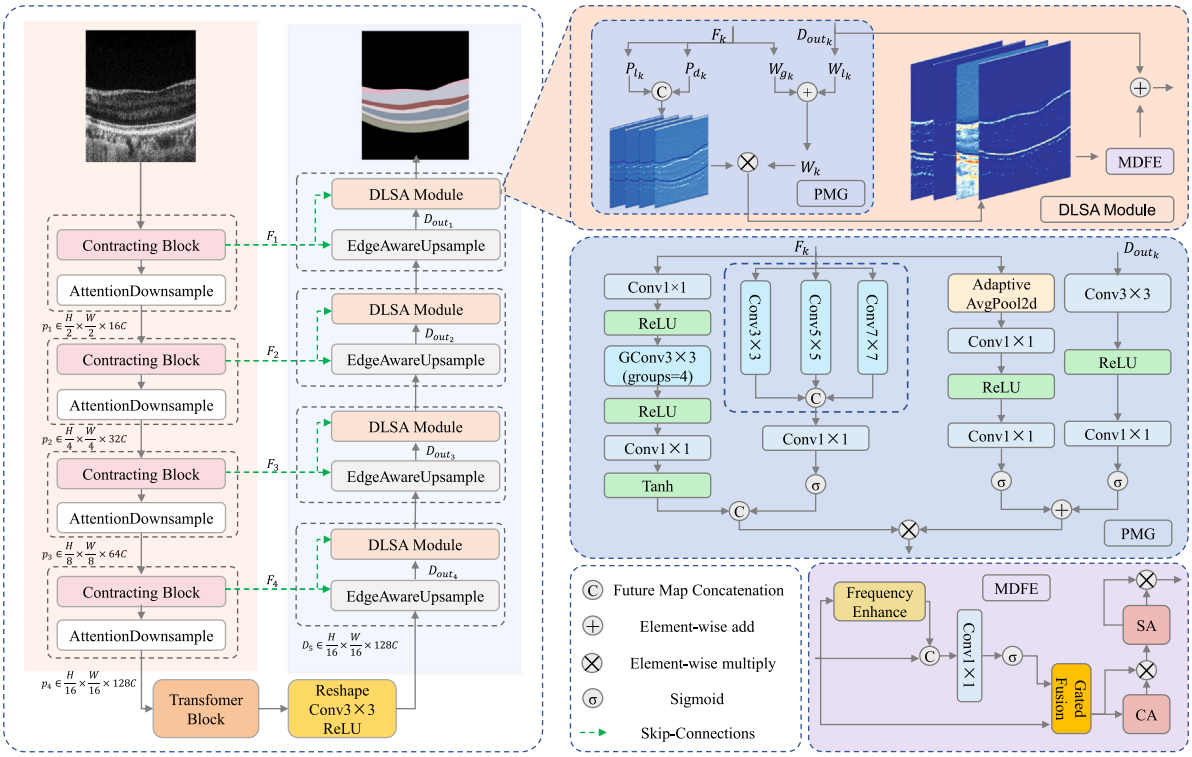

Fig. 1. Architecture of the proposed DLSANet. The encoder employs a four-level downsampling structure with Contracting Blocks and AttentionDownsample modules for multi-scale feature extraction and connects to a Transformer bottleneck layer for global contextual enhancement. The decoder integrates edge-aware upsampling for resolution restoration and embeds core DLSA modules at each decoding stage. These DLSA modules fuse encoder skip-connected features and decoder upsampled features via Prior Modeling and Guidance component and Multi-domain Feature Enhancement to maintain structural coherence and suppress noise.

using 3 × 3 convolution with a stride of 2 to reduce spatial dimensions while enhancing semantically salient regions with F_k directly fed into the module to produce the output denoted as p_k. The operation is formulated as:

$$p_k = Down\left(F_k \otimes Attn(F_k)\right) \tag{1}$$

where $Attn(\cdot)$ denotes the attention weight generation function that consists of 1 × 1 convolution and Sigmoid activation, $Down(\cdot)$ represents feature transformation integrating 3 × 3 strided convolution, BN, and ReLU activation, and ⊗ signifies element-wise multiplication. Ultimately, the encoder outputs four multi-scale feature maps denoted as $F_k \in \mathbb{R}^{16\times 2^{k-1}\times H/2^{k-1}\times W/2^{k-1}}$. Afterward, the final downsampling feature $p_4 \in \mathbb{R}^{128\times H/16\times W/16}$ is fed into the Transformer bottleneck layer for global contextual modeling, yielding refined features F_T that are further processed via bottleneck convolution to obtain $B_4 \in \mathbb{R}^{256\times H/16\times W/16}$. In our implementation, the Transformer bottleneck comprises three stacked Transformer encoder layers, each characterized by an embedding dimension of 128 and eight attention heads. Initially, the encoder feature map is flattened into a sequence of tokens and subsequently projected into the embedding space via a linear layer. To maintain spatial positional information, learnable positional embeddings are added.

3.2.2. Cross-scale decoder

The decoder progressively restores spatial resolution via the EdgeAwareUpsample module, fuses features from corresponding encoder levels through skip connections, and embeds DLSA modules at each decoding stage for structured feature calibration and enhancement. The EdgeAwareUpsample module incorporates PixelShuffle to boost resolution and generates an attention mask via an edge detection branch, facilitating accurate fusion of encoder features and decoder upsampled features. The operation is formulated as:

$$D_{up_k} = PixelShuffle\left(Conv_{3\times3}(D_k)\right) \tag{2}$$

$$M_{edge_k} = \sigma\left(Conv_{1\times1}\left(ReLU\left(Conv_{3\times3}(F_k)\right)\right)\right) \tag{3}$$

$$D_{out_k} = Conv_{3\times3}\left(Concat(D_{up_k}, F_k \otimes M_{edge_k})\right) \tag{4}$$

where D_k denotes the input feature of the kth decoder stage, $Concat(\cdot)$ signifies channel-wise concatenation, σ represents the Sigmoid activation function, and $M_{edge,k}$ denotes the edge attention mask for the kth decoder stage. Through 4-level upsampling (one per stage k) and DLSA module refinement, the decoder progressively generates enhanced feature maps $\{F_{M,k}\}_{k=1}^{4}$. A final 1 × 1 convolution projects the highest-resolution feature $F_{M,4}$ (from the 4th decoder stage) to the segmentation map $S \in \mathbb{R}^{C\times H\times W}$, where C denotes the number of retinal layer classes.

3.3. DLSA module

The DLSA module is the core innovative component of DLSANet deployed at each decoder level taking encoder skip-connected feature F_k and previous decoder level upsampled feature D_{out_k} as inputs to achieve structurally consistent and noise-robust feature fusion through the synergy of prior modeling and feature enhancement.

3.3.1. Prior modeling and guidance component

The Prior Modeling and Guidance component integrates constraints derived from retinal anatomical priors into feature learning through dual-path structural pattern generation and adaptive weight modulation, effectively addressing structural incoherence. The learnable structure branch leverages 1D convolution to derive globally consistent anatomical constraints based on the channel-wise mean grayscale projection map G_k of encoder feature F_k. This design adapts to the retina's strictly ordered layered architecture along the depth direction enabling efficient modeling of cross-channel hierarchical correlations without

spatial redundancy. The resulting learnable structure prior P_{l_k} is tailored to the kth decoder stage consistent with the hierarchical definition of k. Meanwhile the deformable structure branch captures local boundary patterns through multi-scale convolution adapting to lesional structural deformations to yield the deformable structure prior P_{d_k}. The operations are formulated as:

$$G_k = \frac{1}{N}\sum_{i=1}^{N} F_k \tag{5}$$

$$P_{l_k} = \tanh(f_l(G_k)) \tag{6}$$

$$P_{d_k} = \sigma\left(f_d(G_k)\right) \tag{7}$$

where G_k denotes the channel-wise mean grayscale map of encoder feature F_k, i denotes the feature channel index, and N is the total number of feature channels. $f_l(\cdot)$ comprises three convolutional layers: a 1×1 convolution, a grouped convolution (GConv3 $\times$ 3 with groups set to 4), and another 1×1 convolution. These layers function as the feature transformation network for the Learnable Anatomical Structure Prior Module. $f_d(\cdot)$ acts as a multi-scale convolution fusion network for the Deformable Boundary-Adaptive Structure Prior Module, employing parallel convolutional branches with kernel sizes of 3×3, 5×5, and 7×7 that operate on the same input feature map for feature extraction and fusion. The hyperbolic tangent activation $\tanh(\cdot)$ projects prior values onto the range $[-1, 1]$, while σ denotes the Sigmoid activation function. The dual-path priors are concatenated along the channel dimension, and a 1×1 convolution is applied for feature fusion and dimensional alignment to generate the final structure prior map P_k, formulated as:

$$P_k = \sigma\left(Conv_{1\times1}\left(Concat(P_{l_k}, P_{d_k})\right)\right) \tag{8}$$

A global context path and a local detail path are designed to generate adaptive weight maps that balance global structural constraints and local semantic features: the global weight W_{g_k} captures channel-wise global importance via Global Average Pooling and fully connected layers, while the local weight W_{l_k} enhances local semantic structures of decoder features through convolutional layers. This is formulated as:

$$W_{g_k} = \sigma\left(FC\left(ReLU\left(FC\left(GAP(F_k)\right)\right)\right)\right) \tag{9}$$

$$W_{l_k} = \sigma\left(f_w(D_{out_k})\right) \tag{10}$$

The global and local weights are averaged fused and act synergistically with the structure prior on encoder features for structured modulation formulated as:

$$W_k = \frac{W_{g_k} + W_{l_k}}{2} \tag{11}$$

$$E'_k = F_k \otimes \left(1 + \gamma \cdot Mean_c(P_k \otimes W_k)\right) \tag{12}$$

where γ is a learnable scaling parameter, which is initialized at 1.2 and optimized through backpropagation during the training process. $\mathrm{Mean}_c(\cdot)$ denotes channel-wise averaging to generate a single-channel attention map for anatomically salient region enhancement and noise suppression.

3.3.2. Multi-domain feature enhancement

The Multi-domain Feature Enhancement suppresses speckle noise and artifacts in OCT images via frequency spatial collaborative enhancement while preserving fine structural details. It targets the inherent characteristic of OCT speckle noise which concentrates in high frequency bands and partially overlaps with valid signals. Frequency domain refinement is realized via amplitude modulation with 1×1 convolution avoiding complex FFT transformations and reducing computational overhead. The modulation network $f_a(\cdot)$ generates adaptive band specific weights via Sigmoid activation emphasizing signal dominant frequency bands and suppressing noise dominant ones to achieve efficient frequency domain purification without explicit spectrum decomposition. This is formulated as:

$$E'_{freq_k} = f_a(\cdot)(E'_k) \tag{13}$$

where $f_a(\cdot)$ refines the input feature E'_k into E'_{freq_k}, a frequency-domain enhanced feature. An adaptive gating mechanism fuses frequency-enhanced features with original spatial features and preserves initial feature information via residual connections. This fusion process is formulated as:

$$G_{g_k} = \sigma\left(Conv_{1\times1}\left(Concat(E'_{freq_k}, E'_k)\right)\right) \tag{14}$$

$$E'_{fuse_k} = G_{g_k} \otimes E'_{freq_k} + (1 - G_{g_k}) \otimes E'_k \tag{15}$$

$$F_{out_k} = E'_{fuse_k} + \alpha \cdot D_{out_k} \tag{16}$$

where G_{g_k} denotes the fusion gating weight, α is a learnable residual parameter, which is initialized at 0.5 and optimized through backpropagation during the training process. F_{out_k} signifies the enhanced feature output of the DLSA module, serving for upsampling and fusion in the subsequent decoder stage.

4. Experiments

This section presents a comprehensive evaluation of our proposed DLSANet against several state-of-the-art methods for retinal layer segmentation. The experiments are designed to assess model performance across diverse conditions: (i) two public datasets featuring pathological cases—a Glaucoma dataset and a Diabetic Macular Edema (DME) dataset; and (ii) a privately collected dataset, Retina500, which comprises OCT scans from 500 healthy human eyes. Furthermore, an extensive ablation study is conducted on Retina500 to isolate and systematically validate the contribution of each key component within our proposed DLSA module.

4.1. Datasets

4.1.1. Retina500

We independently designed and developed the Retina500 dataset using a customized visible and near-infrared OCT system. A broadband laser output is generated by a SuperK supercontinuum laser (NKT Photonics, Birkerød, Denmark). This output is separated into visible and near-infrared (NIR) beams via a dichroic mirror (DM1) with a cut-off wavelength of 650 nm. The visible light is polarized using a polarization beam splitter (PBS) and subsequently expanded through a pair of prisms. Polarization controllers (PCs) are used to optimize the polarization state for interference efficiency. A specific band within the visible spectrum is selected using a slit aperture and then redirected by a mirror (M). The NIR beam is further separated by a second dichroic mirror (DM2) with a cut-off wavelength of 900 nm and filtered to a bandwidth of 800–875 nm using edge filters. These spectral components are combined via a custom wavelength division multiplexer (WDM) and directed into an optical fiber coupler (TW670R2A2, Thorlabs, Newton, NJ, USA). In the sample arm, the beam is collimated using a 6 mm lens (CL), corrected with an achromatizing lens (AL), scanned via galvanometer mirrors, and focused onto the pupil through a 2:1 telescope, achieving a 2 mm beam diameter at the cornea. The reference arm consists of collimated light reflected back toward the detector. Dispersion compensation in the sample arm is achieved using BK7 glass plates (DC), while a variable neutral density (ND) filter regulates the light intensity. Additional dispersion matching is performed using a water cuvette. Light returning from both arms is recombined in the fiber coupler and then split into two spectrometers by another WDM. Each spectrometer is equipped with a line-scan camera (spl2048140 km, Basler, Ahrensburg, Germany), capturing spectral ranges of 535–600 nm and 780–880 nm, respectively. The spectrometer

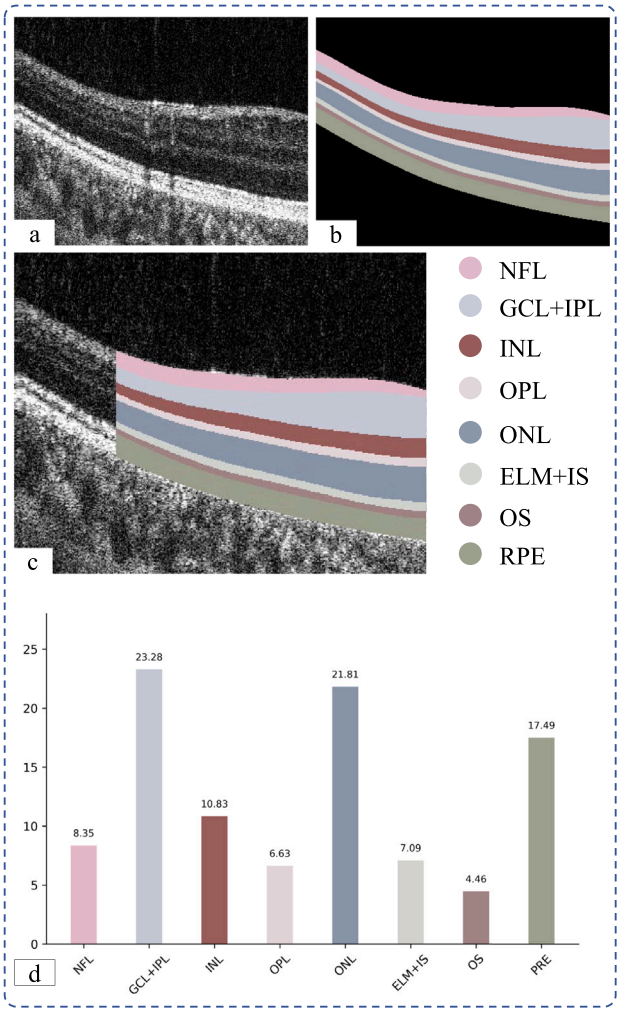

Fig. 2. OCT B-scan images of retinal layers and their annotations from the Retina500 dataset (healthy human eyes). (a) Original retinal B-scan image; (b) Ground-truth annotation map; (c) Eight classes of annotations for retinal layers, namely, NFL, GCL + IPL, INL, OPL, ONL, ELM + IS, OS, and RPE, and other areas annotated as background. (d)The average percentage of pixels in each retinal layer in the Retina500 dataset among all retinal layers (excluding background).

converts the optical signal into an electrical signal, which is processed by a computer to reconstruct the final OCT image. This study employed single-channel near-infrared images (780–880 nm) for model training, validation, and inference, aiming to accurately assess the segmentation performance of the proposed algorithm.

For the annotation of retinal layers, we utilized the advanced graphics software Inkscape (v1.1.2) under the supervision of experienced ophthalmologists to precisely delineate eight distinct retinal layers: Nerve Fiber Layer (NFL), Ganglion Cell Layer + Inner Plexiform Layer (GCL+IPL), Inner Nuclear Layer (INL), Outer Plexiform Layer (OPL), Outer Nuclear Layer (ONL), External Limiting Membrane + Inner Segments (ELM+IS), Outer Segments (OS) and Retinal Pigment Epithelium (RPE). Each layer was assigned a unique color for visual distinction, with non-layer regions marked in black as background. Prior to finalization, each annotated image underwent thorough review and validation by our clinical ophthalmology team to ensure annotation accuracy and reliability, with representative examples of these annotated retinal layers shown in Fig. 2. The finalized images were exported in PNG format with a resolution of 480 × 400 pixels. To support effective training and evaluation of machine learning models, the dataset was partitioned into a training set of 400 images, a validation set of 50 images, and a test set of 50 images. To prevent data leakage, the partitioning was conducted at the patient level, ensuring that all B-scans from a single subject were allocated to the same subset.

4.1.2. DME

This dataset is curated by Chiu et al. via the Duke Enterprise Data Unified Content Explorer search engine, which is used to retrospectively identify subjects with DME at the Duke Eye Center [25]. It comprises 110 OCT B-scans acquired from 10 patients with DME, with each image having a resolution of 496 × 768 pixels. Each B-scan is manually annotated to delineate 9 retinal layers. The dataset is partitioned into training, validation, and test sets, containing 88, 11, and 11 images, respectively. All scans are centered on the macula.

4.1.3. Glaucoma

This dataset is collected from 61 distinct subjects [49]. The Ophthalmology Department of Shanghai General Hospital acquires 12 radial OCT B-scans for each subject using a DRI OCT-1 Atlantis device. All images are captured in the optic nerve head region, with a field of view of 20.48 mm × 7.94 mm. Under the supervision of a glaucoma specialist, two ophthalmologists manually annotate these images to identify the optic disc and nine retinal layers. The images have a size of 1024 × 992 pixels, and the dataset is divided into training, validation, and test sets, consisting of 148, 48, and 48 images, respectively.

4.2. Performance metrics

To consider the class imbalance problem in the dataset, we quantitatively evaluate the segmentation performance using the Dice score, mIoU, Acc, and mPA. They are computed using the following formulas:

$$Dice = \frac{2TP}{2TP + FP + FN}. \tag{17}$$

$$mIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{TP}{FN + FP + TP}. \tag{18}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}. \tag{19}$$

$$mPA = \frac{1}{k+1} \sum_{i=0}^{k} \frac{TP}{TP + TN + FP + FN}. \tag{20}$$

where TP, FP, TN, and FN represent True Positives, False Positives, True Negatives, and False Negatives respectively. k represents the number of categories. In addition, we count the number of parameters for compared approaches to represent their computation complexity.

Additionally, we evaluate the boundary accuracy of the segmentation results using the 95th percentile Hausdorff distance (HD_{95}). HD_{95} quantifies the 95th percentile of the distances between the predicted boundaries and the ground-truth boundaries, thereby mitigating the influence of extreme outliers. It is defined as follows:

$$HD_{95}(P, G) = \max\left\{P_{95}(d(P, G)), P_{95}(d(G, P))\right\} \tag{21}$$

where $d(P, G)$ denotes the set of distances from each point in P to its nearest neighbor in G, and $P_{95}(\cdot)$ represents the 95th percentile operator.

Table 1
Multiple approaches to multiple metrics (Dice score (%), mIoU (%), Acc(%), mPA(%), HD_{95}(pixel)) evaluation on the Retina500 dataset.

Method	NFL	GCL+IPL	INL	OPL	ONL	ELM+IS	OS	RPE	mIoU	Acc	mPA	$HD_{95}\downarrow$
ReLayNet [14]	87.25	94.23	90.21	82.27	94.78	91.64	80.02	92.40	80.72	91.07	96.69	4.40
Attention_Unet [31]	88.28	94.77	**91.68**	82.72	94.22	92.34	78.75	92.29	81.22	91.37	96.79	3.12
BiSeNet [50]	88.46	95.00	91.29	82.52	94.44	92.28	79.23	92.84	81.43	91.67	96.84	3.28
U-net [13]	88.59	94.46	90.17	80.38	93.85	92.53	76.34	91.52	79.87	90.53	96.57	6.06
OS_MGU [49]	88.25	93.85	89.60	81.86	94.34	92.55	80.35	93.25	80.97	90.67	96.78	3.18
DFANet [51]	87.03	94.02	89.99	80.59	93.87	91.04	77.55	91.81	79.40	90.05	96.49	5.82
TransUnet [32]	89.02	94.62	90.65	82.21	94.30	92.45	78.46	92.05	80.96	91.12	96.75	3.19
EMV-Net [52]	88.41	94.70	91.16	82.41	94.37	**92.76**	80.22	92.90	81.57	91.48	96.85	3.00
LightReSeg [36]	88.19	95.01	90.99	83.03	94.55	92.03	80.97	**93.47**	81.81	91.60	96.91	3.19
DLSANet(Ours)	**89.25**	**95.25**	91.16	**83.15**	**94.87**	92.41	**81.07**	93.05	**82.22**	**91.87**	**96.95**	**2.87**

Table 2
Multiple approaches to multiple metrics (Dice score (%), mIoU (%), Acc(%), mPA(%), HD_{95}(pixel)) evaluation on the DME dataset.

Method	NFL	GCL+IPL	INL	OPL	ONL	ELM+IS	OS	RPE	mIoU	Acc	mPA	$HD_{95}\downarrow$
ReLayNet [14]	81.26	93.63	80.85	**78.12**	87.19	87.32	86.33	49.12	68.96	79.41	95.86	25.17
Attention_Unet [31]	81.94	93.48	79.81	77.98	87.29	87.16	86.39	50.78	69.03	79.73	95.87	23.88
BiSeNet [50]	80.92	92.89	78.59	75.25	86.91	86.74	85.85	58.02	68.66	79.64	95.65	21.81
U-net [13]	81.81	93.57	79.82	77.70	86.68	**87.91**	86.48	51.48	69.11	79.27	95.77	29.95
OS_MGU [49]	80.07	91.98	77.82	74.56	85.85	86.72	85.76	49.75	66.87	78.08	95.27	32.10
DFANet [51]	76.69	90.95	74.69	73.29	85.45	85.19	85.12	46.01	64.48	75.30	95.05	29.06
TransUnet [32]	**82.23**	93.39	78.50	77.34	87.49	87.31	86.43	57.73	69.63	80.04	95.85	26.06
EMV-Net [52]	81.43	93.09	79.49	76.81	86.38	86.88	86.43	52.09	68.53	79.61	95.70	30.80
LightReSeg [36]	81.30	93.31	78.10	76.47	86.52	86.52	85.73	52.69	68.15	78.83	95.54	29.58
DLSANet(Ours)	82.13	**93.71**	**81.19**	76.94	**87.68**	87.85	**87.15**	**59.85**	**70.63**	**81.04**	**96.05**	**20.36**

Table 3
Multiple approaches to multiple metrics (Dice score (%), mIoU (%), Acc(%), mPA(%), HD_{95}(pixel)) evaluation on the Glaucoma dataset.

Method	NFL	GCL	IPL	INL	OPL	ONL	IN+OS	RPE	Choroid	OD	mIoU	Acc	mPA	$HD_{95}\downarrow$
ReLayNet [14]	80.35	66.62	71.77	76.47	79.71	90.71	85.57	81.68	88.29	77.83	67.16	79.71	96.66	29.71
Attention_Unet [31]	81.83	63.62	72.05	75.86	**79.98**	89.16	85.76	82.50	89.64	**84.28**	67.98	80.87	97.05	23.84
BiSeNet [50]	79.73	64.36	68.96	72.68	76.66	88.21	83.22	78.95	86.94	83.09	64.90	79.11	96.65	21.76
U-net [13]	82.03	66.40	71.54	74.32	78.81	90.44	86.32	82.33	89.59	82.34	67.85	80.10	96.98	26.76
OS_MGU [49]	80.91	**69.11**	72.23	74.73	75.28	89.85	85.85	82.13	89.08	82.90	67.49	81.38	96.86	21.47
DFANet [51]	81.86	64.58	70.82	75.66	79.69	89.89	85.49	81.90	89.14	82.73	67.55	80.30	96.96	26.08
TransUnet [32]	81.29	64.27	65.79	69.93	77.84	90.20	85.46	81.98	88.93	83.63	66.05	78.96	96.86	21.81
EMV-Net [52]	82.15	67.06	**72.43**	76.36	77.02	89.81	85.03	82.00	88.53	83.09	67.68	79.32	96.97	22.41
LightReSeg [36]	80.91	64.10	67.40	73.86	78.10	89.63	85.27	81.79	88.25	82.18	66.21	78.77	96.76	23.83
DLSANet(Ours)	**82.24**	69.02	72.04	**76.97**	79.95	**90.79**	**85.92**	**82.72**	**90.23**	83.45	**69.08**	**81.46**	**97.12**	**20.94**

4.3. Implementation details

DLSANet is implemented using PyTorch and trained with the Adam optimizer, with cross-entropy loss adopted as the objective function. The initial learning rate is set to 0.001 and gradually halved every 40 epochs. The batch size is set to 2, and the model is trained for 200 epochs until convergence. Early stopping, based on the validation loss, is employed to avoid overfitting. All reported results are obtained from a single training run with a fixed random initialization. Data augmentation is applied to all three datasets, including horizontal flipping with P = 0.5, random central rotation within ±20° with P = 0.5, median and motion blur processing with P = 0.5, random addition of Gaussian noise, and random adjustments of brightness and contrast with P = 0.5. All experiments are conducted on an NVIDIA A100 GPU.

4.4. Comparison

We compare DLSANet with the state-of-the-art approaches including ReLayNet [14], Attention_Unet [31], BiSeNet [50], U-net [13], OS_MGU [49], DFANet [51], TransUnet [32], EMV-Net [52] and LightReSeg [36]. We report the results on the above-mentioned three datasets.

4.4.1. Quantitative analysis

Extensive comparative evaluations with state-of-the-art methods are performed on the Retina500 dataset, where quantitative outcomes are presented in Table 1 . We can see that our approach DLSANet scores the best performance in both mIoU, Acc and mPA metrics, with values of 82.22%, 91.87%, and 96.95%, respectively. This establishes a consistent and notable advantage over the second-best approach LightReSeg. At the layer-wise level, our approach achieves the highest Dice score in the NFL layer, GCL/IPL, OPL, ONL, and OS layers. It also maintains highly competitive performance on the remaining layers (INL, ELM/IS, RPE), with results within 0.52% of the best-performing methods, underscoring its comprehensive robustness. Beyond regional metrics, we further evaluate the accuracy of segmentation boundary localization using the HD_{95} metric. As shown in Table 1, the proposed DLSANet outperforms other methods by achieving the lowest distance error on the HD_{95} metric, which indicates a higher precision in localizing retinal layer boundaries. We further perform a statistical significance test, using the Wilcoxon rank sum test, to compare the Dice Score performance of different methods on each layer. When comparing our approach with LightReSeg, we observe a Pvalue of 0.034214 ($p < 0.05$), indicating a statistically significant difference. Similar statistically significant differences are observed, with Pvalues of 0.028758 ($p < 0.05$), 0.016986 ($p < 0.05$), and 0.014990 ($p < 0.05$) respectively when comparing our method with U-net, EMV-Net, and TransUnet. This performance gap stems from several limitations in existing methods. First, architectures like U-net and RelayNet employ a decoupled paradigm where feature learning is separate from topological modeling, leading to anatomically implausible segmentations in ambiguous regions like the OPL and OS. Second, while TransUnet captures long-range context, its computational complexity and lack of a dedicated anatomical prior limit its precision on fine layers. Finally, lightweight models like EMV-Net and LightReSeg, despite their deployability, exhibit inconsistent

performance across different layers, revealing their limited capacity for enforcing global topological constraints.

We evaluated our method on the DME dataset, as shown in Table 2. We achieved the best performance across all four metrics: mIoU, mPA, Acc and HD_{95}. Under the Dice score metric, we ranked first across five layers (GCL+IPL, INL, ONL, OS and RPE) and achieved strong results on several other layers. We further evaluated our method on the glaucoma dataset, as shown in Table 3. DLSANet achieved the best overall results across mIoU, mPA, Acc and HD_{95}. Additionally, our method demonstrated superior performance in Dice score across most layers. We further perform the statistical significance test by using the Wilcoxon rank sum test. For example, when comparing our method with LightReSeg on the same domain, we observe a Pvalue of 0.038975 ($p < 0.05$), indicating a statistically significant difference.

In our findings, we observe that certain layers, specifically the GCL+IPL and ONL, exhibit significantly higher accuracy compared to other layers, as illustrated in Table 1. We posit that class imbalance is the primary contributing factor, as the number of samples in specific classes is substantially larger than in others. This disparity likely enables the model to learn more effectively for those classes during training, resulting in elevated accuracy during evaluation. Fig. 2(d) indicates that the average pixel proportion of the GCL+IPL and ONL layers is relatively high, corroborating the higher accuracy presented in Table 1 for these layers. Furthermore, variations in shape, color, texture, and other characteristics among different layers may facilitate easier segmentation for certain target categories while complicating it for others. In conclusion, we advocate that addressing class imbalance is the most effective strategy to mitigate this issue.

4.4.2. Qualitative analysis

We conduct qualitative analysis on a challenging retinal B-scan degraded by speckle noise, as presented in Fig. 3. As shown in Fig. 3(l), our proposed method achieves a more accurate and continuous segmentation across all layers. The errors observed in other methods can be categorized into two types. Intra-class errors, primarily misclassifications within the same layer, are evident in the regions marked by white dashed lines in Fig. 3(f), (i), and (j). These errors occur because the blood flow artifacts obscure the layered information of the INL and OPL layers. Inter-class errors, indicated by red dashed lines in Fig. 3(c), (e), (g), (h), (j), and (k), arise from the speckle noise which reduces the contrast of boundaries, leading to the inaccurate identification of category borders. Although Attention_Unet in Fig. 3(d) shows reasonable performance, it still exhibits inaccuracies at the OPL and ONL layer boundaries, as highlighted by the yellow arrow. In contrast, our approach produces a segmentation that is not only accurate and continuous but also demonstrates superior robustness to these common imaging disturbances.

We present segmentation prediction maps from multiple mainstream segmentation methods on the DME dataset, as shown in Fig. 4. As shown in Fig. 4(j), our approach achieves much better segmentation performance. The vast majority of methods exhibit varying degrees of intra-class errors in the lesion fluid layer, as shown in Fig. 4(c), (d), (f-k). Although BiSeNet in Fig. 4(e) demonstrates no intra-class errors, the segmentation between the lesion fluid layer and the boundary layer remains unclear, resulting in significant inter-class errors. However, our method not only achieves effective segmentation of the retinal layer but also outperforms the other methods in the segmentation of the lesion fluid layer. These results highlight the robustness and precision of our model in the challenging task of retinal layer segmentation under disease conditions.

For the retinal optic nerve head region, comparative prediction plots of several mainstream segmentation approaches on a single B-scan image are illustrated in Fig. 5. As shown in Fig. 5(j), our approach shows a better segmentation performance. As previously mentioned, segmentation errors are divided into intra-class errors, represented by

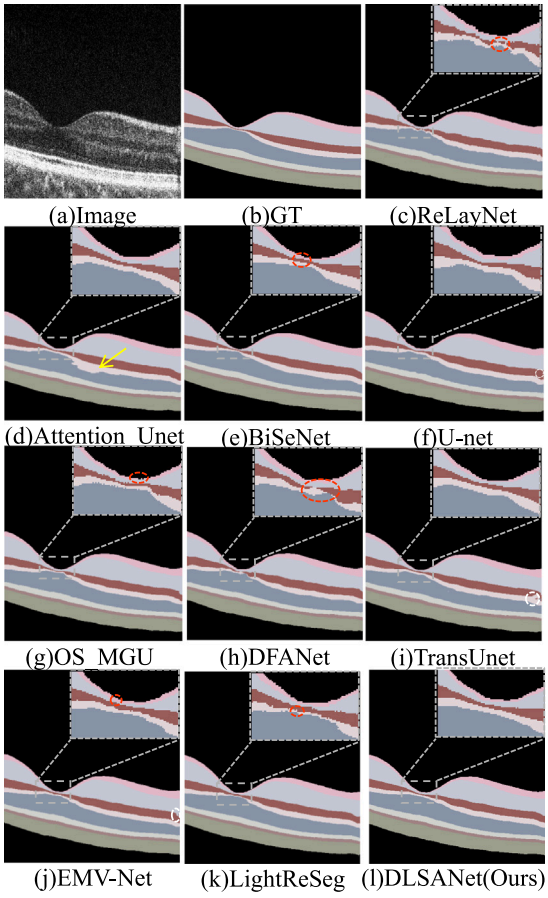

Fig. 3. Comparison of segmentation prediction maps of mainstream approaches on the Retina500 dataset. (a) Original image. (b) Ground truth (c) Prediction map of ReLayNet. (d) Prediction map of Attention_Unet. (e) Prediction map of BiSeNet. (f) Prediction map of U-net. (g) Prediction map of OS_MGU. (h) Prediction map of DFANet. (i) Prediction map of TransUnet. (j) Prediction map of EMV-Net. (k) Prediction map of LightReSeg. (i) Prediction map of DLSANet.

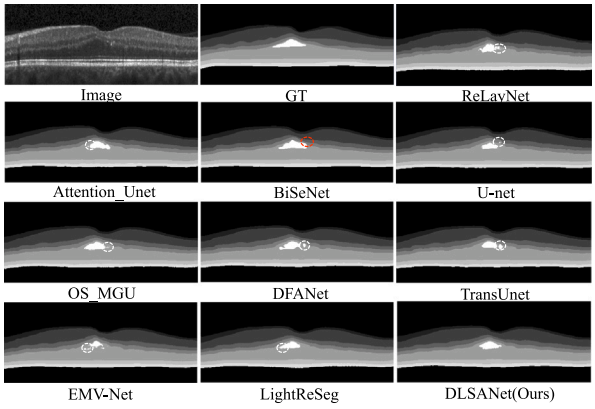

Fig. 4. Comparison of segmentation prediction maps of mainstream approaches on the DME dataset. (a) Original image. (b) Ground truth (c) Prediction map of ReLayNet. (d) Prediction map of Attention_Unet. (e) Prediction map of BiSeNet. (f) Prediction map of U-net. (g) Prediction map of OS_MGU. (h) Prediction map of DFANet. (i) Prediction map of TransUnet. (j) Prediction map of EMV-Net. (k) Prediction map of LightReSeg. (i) Prediction map of DLSANet.

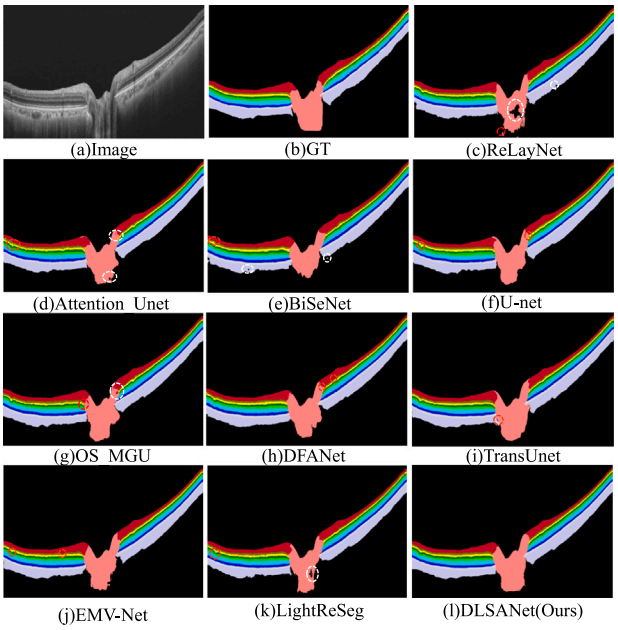

Fig. 5. Comparison of segmentation prediction maps of mainstream approaches on the Glaucoma dataset. (a) Original image. (b) Ground truth (c) Prediction map of ReLayNet. (d) Prediction map of Attention_Unet. (e) Prediction map of BiSeNet. (f) Prediction map of U-net. (g) Prediction map of OS_MGU. (h) Prediction map of DFANet. (i) Prediction map of TransUnet. (j) Prediction map of EMV-Net. (k) Prediction map of LightReSeg. (i) Prediction map of DLSANet.

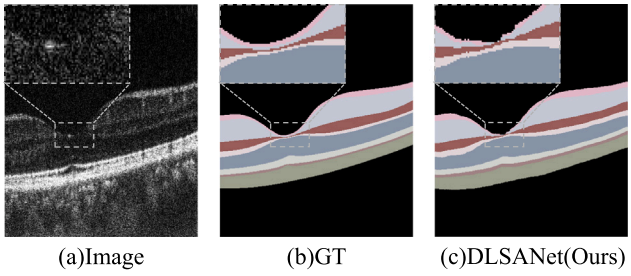

Fig. 6. Example of Segmentation Under Extreme Speckle Noise Conditions. (a) Original image. (b) Ground truth. (i) Prediction map of DLSANet.

white dashed lines as shown in Fig. 5(c), (d), (e), (g) and (k), and inter-class errors, indicated by red dashed lines in Fig. 5(c-k). In contrast, our approach not only has no problem with segmentation category error but also the segmented boundary is smoother than other approaches.

To evaluate the performance of the proposed method comprehensively, we present an additional segmentation example under extreme noise conditions. As illustrated in Fig. 6(a), this OCT image is severely affected by speckle noise and exhibits extremely low inter-layer contrast, resulting in blurred or even indistinguishable retinal layer boundaries. In this scenario, while DLSANet manages to preserve the basic layer structure in most regions, significant segmentation errors still occur in localized areas, as depicted in Fig. 6(c). This failure case further underscores the limitations of current methods when addressing extreme imaging conditions. Future research will investigate more robust feature modeling strategies to enhance the model's segmentation stability under challenging noise conditions.

4.5. Parameter scale and inference time statistics

We summarize the parameter scales and inference time of all competing methods across the Retina500, DME, and Glaucoma datasets in Table 4. DLSANet achieves state-of-the-art segmentation performance while maintaining a compact parameter footprint and efficient inference speed that are competitive with existing methods in the field. Its balanced performance in parameter scale and inference efficiency aligns with the core requirements of clinical application scenarios especially for portable OCT systems where both resource constraints and performance reliability are critical. The statistical results in the table confirm that DLSANet avoids the excessive parameter overhead of heavyweight models without sacrificing the segmentation accuracy necessary for practical use. This balance between model complexity and computational efficiency underscores its potential for real-world deployment while providing a benchmark for subsequent research on lightweight medical image segmentation architectures.

Table 4
Statistical inference speed of different methods on three datasets, and the best performance on each dataset is bolded.

Method	Params	Retina500	DME	Glaucoma
ReLayNet [14]	0.7M	0.02s	0.03s	0.12s
Attention_Unet [31]	34.8M	0.02s	0.04s	0.18s
BiSeNet [50]	13.1M	0.01s	0.03s	0.11s
U-net [13]	34.5M	0.02s	0.06s	0.16s
OS_MGU [49]	2.0M	0.02s	0.04s	0.14s
DFANet [51]	2.1M	0.02s	0.04s	0.13s
TransUnet [32]	105.7M	0.03s	0.05s	0.23s
EMV-Net [52]	1.9M	0.05s	0.07s	0.21s
LightReSeg [36]	3.3M	0.02s	0.04s	0.15s
DLSANet(Ours)	4.1M	0.02s	0.04s	0.14s

4.6. Ablation study

We conduct ablation experiments on the Retina500 dataset to evaluate the contribution of each component within the DLSA module to the overall segmentation performance. The ablated variants are defined as follows. Base denotes the U-shaped backbone. Base_PMG and Base_MDFE correspond to the Base model with the PMG component and MDFE component added, respectively. To further analyze the internal mechanisms of the PMG component, we construct two mechanism-level variants: Base_PMG-Prior and Base_PMG-Weight. Specifically, these variants integrate incomplete versions of the PMG component into the Base model. Base_PMG-Prior retains solely the dual-path structure prior generation mechanism of the PMG component, while Base_PMG-Weight retains exclusively the adaptive weight modulation mechanism. The comprehensive model, Base_all, signifies the final version of the proposed DLSANet. As shown in Table 5, Base_all demonstrates a significant improvement over Base, with increases of +1.19% in mIoU, +0.78% in Acc, and +0.18% in mPA. This indicates a clear progressive enhancement as various components of the DLSA module are introduced. Specifically, models that incorporate any single component consistently outperform the Base model across Average_Dice, mIoU, and mPA metrics. Compared to Base, Base_PMG-Prior achieves notable enhancements in Dice, mIoU, and mPA, suggesting that the dual-path structure prior generation mechanism effectively enhances structural consistency. In contrast, Base_PMG-Weight shows only marginal improvements, indicating that adaptive weight modulation alone has limited impact without the constraints of structural priors. When both mechanisms are integrated, the complete PMG component consistently outperforms its individual sub-mechanisms. Furthermore, Table 5 demonstrates that as the PMG and MDFE components are progressively integrated, the model size experiences a moderate increase, yet it remains within a relatively lightweight range overall. These modules substantially enhance segmentation performance while contributing only a limited number of parameters.

The qualitative results presented in Fig. 7 further substantiate these findings. As illustrated in the magnified region of Fig. 7(c), the Base

Table 5
The proposed approach performs multiple metrics evaluation of ablation experiments on the Retina500 dataset.

Method	Avg Dice	mIoU	Acc	mPA	Params
Base	89.25	81.03	91.09	96.77	2.5*M*
Base_PMG-Prior	89.75	81.41	91.17	96.83	3.4*M*
Base_PMG-Weight	89.72	81.37	91.17	96.85	3.5*M*
Base_PMG	89.82	81.52	91.00	96.89	4.1*M*
Base_MDFE	89.91	81.34	91.18	96.88	4.0*M*
Base_all(DLSANet)	**90.24**	**82.22**	**91.87**	**96.95**	4.1*M*

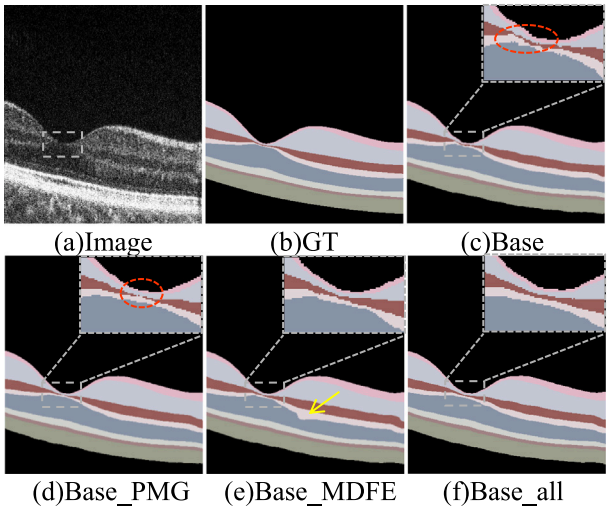

Fig. 7. Prediction images of our method under different ablation settings. (a) Original image. (b) Ground truth. (c) Base framework (d) Base framework plus PMG component (e) Base framework plus MDFE component (f) Base framework plus full DLSA module.

model demonstrates significant inter-class errors at the boundaries of the GCL+IPL, INL, and OPL layers, accompanied by layer discontinuity within the OPL layer. Upon the integration of the PMG component, as depicted in Fig. 7(d), the OPL discontinuity is mitigated; however, some inter-class errors persist due to diminished inter-layer contrast resulting from speckle noise. The introduction of the MDFE component, shown in Fig. 7(e), effectively suppresses speckle noise and eliminates inter-class errors in the magnified region, although noticeable layer misalignment is observed in the OPL layer (indicated by the yellow arrow). Following the incorporation of the complete DLSA module, the results in Fig. 7(f) indicate a substantial reduction in both layer discontinuities and misalignments. Although the introduction of the PMG and MDFE components slightly increases the model size, the overall architecture remains relatively lightweight. This indicates that the proposed modules yield substantial performance gains with only a modest increase in parameters.

To further interpret the role of different modules, we use a modified version of Grad-CAM [53] to visualize the feature activations in different layers of the model. Specifically, we extract feature activation heatmaps for different retinal layers across various model architectures. Comparing Fig. 8(e–h) reveals that the Base model exhibits significant segmentation attention deficiencies in the OPL, ONL, ELM+IS, and OS layers, as indicated by the white arrows. Fig. 8(i–l) demonstrate that after incorporating the PMG component into the Base model, such segmentation attention deficiencies are effectively alleviated while the absence of the MDFE component results in a lack of speckle noise suppression, leading to insufficient focus on regions with reduced contrast such as the boundaries between the OPL and INL layers in Fig. 8(i) and between the ONL and ELM+IS layers in Fig. 8(j). Fig. 8(m–p) show that adding the MDFE component to the Base model

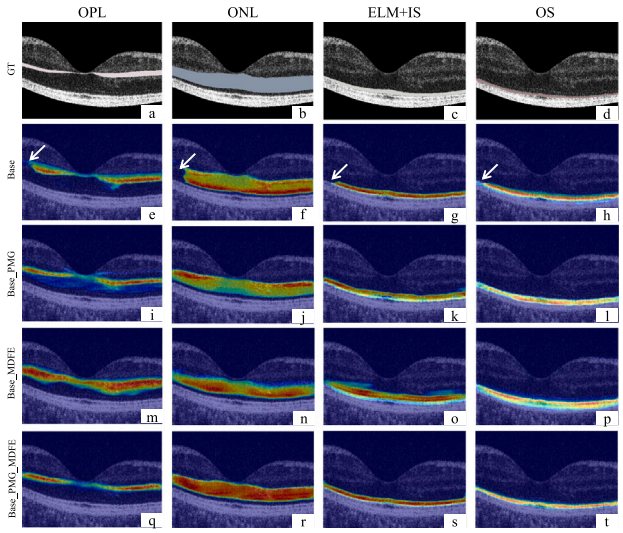

Fig. 8. Heatmap for different retinal layers at different model structures and at different positions of the model, (a–d) for the single category prediction, (e–h) for the Base structure, (i–l) for the Base_PMG structure, (m–p) for the Base_MDFE structure, and (q–t) for the Base_all(Our) structure. The white arrow highlights the segmentation attention deficiencies in the graph.

strengthens speckle noise suppression and significantly improves the focus on the aforementioned regions with reduced contrast, yet the lack of the PMG component causes significant layer misalignment in segmentation attention as illustrated at the boundaries between the OPL and ONL layers in Fig. 8(m) and between the ELM+IS and ONL layers in Fig. 8(o). Fig. 8(q–t) illustrate that after integrating the full DLSA module into the Base model, the segmentation attention is fully concentrated on the regions corresponding to the retinal layers, further demonstrating the effectiveness of our DLSA module in enhancing segmentation attention. Overall, the visualization of deep feature activation provides valuable insights into the model's decision-making process, enabling the model to leverage multi-level information for more robust segmentation results.

5. Limitation

Although the method described in this paper demonstrates commendable retinal segmentation performance across multiple datasets, several issues necessitate further investigation. First, the model may still exhibit segmentation errors in regions characterized by extreme noise interference or complex pathological structures; future research should incorporate more detailed pathological analyses to address this concern. Second, the inference times reported herein were measured on an NVIDIA A100 GPU and were primarily intended to compare computational efficiency among various methods; however, the actual runtime efficiency on resource-constrained hardware platforms remains to be assessed. Furthermore, variations in imaging across different OCT devices and the challenge of physician verification within clinical workflows warrant additional exploration in subsequent research.

6. Conclusion

We propose DLSANet to address two core challenges in retinal layer segmentation structural incoherence from insufficient explicit anatomical constraints and boundary blurring caused by OCT speckle noise reducing inter-layer contrast. The core DLSA module incorporates learnable structural patterns, guided by retinal anatomical priors, directly into the decoding process. This approach ensures structural continuity and the orderly arrangement of segmentation results, even under weak

boundary cues. Additionally, it synergistically refines spatial-frequency representations to suppress noise and preserve fine-grained boundaries. Experiments across Retina500, DME and Glaucoma datasets show DL-SANet achieves state-of-the-art mIoU Acc and mPA with only 4.1M parameters outperforming existing methods in accurate retinal layer segmentation. This lightweight framework offers a clinically translatable solution to standardize ophthalmic diagnostic workflows for large scale screening and longitudinal disease monitoring. Future work will expand multi center multi device datasets to enhance domain generalization and integrate uncertainty quantification to better support clinical decision making.

CRediT authorship contribution statement

Enyu Liu: Writing – original draft, Software, Methodology, Conceptualization. **Muhao Xu:** Writing – review & editing. **Haohua Yang:** Writing – review & editing. **Bingcan Yan:** Writing – review & editing. **Hua Wei:** Writing – review & editing. **Weiye Song:** Writing – review & editing, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62205181, in part by the Natural Science Foundation of Shandong Province under Grant ZR2022QF017, in part by the Shandong Province Outstanding Youth Science Fund Project (Overseas) under Grant 2023HWYQ-023, in part by the Taishan Scholar Foundation of Shandong Province under Grant tsqn202211038, and in part by the Key R&D Program of Shandong Province under Grant 2024CXGC010106.

Data availability

The authors do not have permission to share data.

References

[1] R. Bourne, J.D. Steinmetz, S. Flaxman, P.S. Briant, H.R. Taylor, S. Resnikoff, R.J. Casson, A. Abdoli, E. Abu-Gharbieh, A. Afshin, et al., Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study, Lancet Glob. Health 9 (2) (2021) e130–e143.

[2] A. Ha, Y.J. Lee, M. Lee, S.R. Shim, Y.K. Kim, Digital screen time and myopia: a systematic review and dose-response meta-analysis, JAMA Netw. Open 8 (2) (2025) e2460026–e2460026.

[3] U. Schmidt-Erfurth, J. Garcia-Arumi, F. Bandello, K. Berg, U. Chakravarthy, B.S. Gerendas, J. Jonas, M. Larsen, R. Tadayoni, A. Loewenstein, Guidelines for the management of diabetic macular edema by the European Society of Retina Specialists (EURETINA), Ophthalmologica 237 (4) (2017) 185–222.

[4] R.N. Weinreb, P.T. Khaw, Primary open-angle glaucoma, Lancet 363 (9422) (2004) 1711–1720.

[5] Early Treatment Diabetic Retinopathy Study Research Group, et al., Classification of diabetic retinopathy from fluorescein angiograms: ETDRS report number 11, Ophthalmology 98 (5) (1991) 807–822.

[6] M. Sahoo, S. Pal, M. Mitra, Automatic segmentation of accumulated fluid inside the retinal layers from optical coherence tomography images, Measurement 101 (2017) 138–144.

[7] M.D. Abràmoff, J.M. Reinhardt, S.R. Russell, J.C. Folk, V.B. Mahajan, M. Niemeijer, G. Quellec, Automated early detection of diabetic retinopathy, Ophthalmology 117 (6) (2010) 1147–1154.

[8] J.G. Fujimoto, C. Pitris, S.A. Boppart, M.E. Brezinski, Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy, Neoplasia 2 (1–2) (2000) 9–25.

[9] Y.S. Ambekar, M. Singh, G. Scarcelli, E.M. Rueda, B.M. Hall, R.A. Poché, K.V. Larin, Characterization of retinal biomechanical properties using Brillouin microscopy, J. Biomed. Opt. 25 (9) (2020) 090502–090502.

[10] P. Chauhan, A.M. Kho, P. FitzGerald, B. Shibata, V.J. Srinivasan, Subcellular comparison of visible-light optical coherence tomography and electron microscopy in the mouse outer retina, Investig. Ophthalmol. Vis. Sci. 63 (9) (2022) 10–10.

[11] D. Hein, A. Bozorgpour, D. Merhof, G. Wang, Physics-inspired generative models in medical imaging: A review, 2024, arXiv preprint arXiv:2407.10856.

[12] S. Farsiu, S.J. Chiu, R.V. O'Connell, F.A. Folgar, E. Yuan, J.A. Izatt, C.A. Toth, 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 121 (1) (2014) 162–172.

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

[14] A.G. Roy, S. Conjeti, S.P.K. Karri, D. Sheet, A. Katouzian, C. Wachinger, N. Navab, ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks, Biomed. Opt. Express 8 (8) (2017) 3627–3642.

[15] D. Cabrera Fernández, H.M. Salinas, C.A. Puliafito, Automated detection of retinal layer structures on optical coherence tomography images, Opt. Express 13 (25) (2005) 10200–10216.

[16] A.J. Green, S. McQuaid, S.L. Hauser, I.V. Allen, R. Lyness, Ocular pathology in multiple sclerosis: retinal atrophy and inflammation irrespective of disease duration, Brain 133 (6) (2010) 1591–1601.

[17] H. Huang, Z. Shang, C. Yu, FRD-Net: a full-resolution dilated convolution network for retinal vessel segmentation, Biomed. Opt. Express 15 (5) (2024) 3344–3365.

[18] S. Yadav, S. Mandal, R. Murugan, T. Goel, T. Ahmed, Segmentation and visualization of retinal detachment lesions through retinal fundus images, Biomed. Signal Process. Control. 96 (2024) 106627.

[19] P. Mani, N. Ramachandran, P. Naveen, P.V. Ramesh, Automated segmentation of retinal layers in optical coherence tomography images using Xception70 feature extraction, Appl. Soft Comput. 167 (2024) 112414.

[20] L. Fang, S. Li, D. Cunefare, S. Farsiu, Segmentation based sparse reconstruction of optical coherence tomography images, IEEE Trans. Med. Imaging 36 (2) (2016) 407–421.

[21] N. Shiee, P.-L. Bazin, A. Ozturk, D.S. Reich, P.A. Calabresi, D.L. Pham, A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions, NeuroImage 49 (2) (2010) 1524–1535.

[22] K. Chen, T. Qin, V.H.-F. Lee, H. Yan, H. Li, Learning robust shape regularization for generalizable medical image segmentation, IEEE Trans. Med. Imaging 43 (7) (2024) 2693–2706.

[23] B. Fazekas, G. Aresta, P. Seeböck, J. Mai, U. Schmidt-Erfurth, H. Bogunović, SD-RetinaNet: Topologically constrained semi-supervised retinal lesion and layer segmentation in OCT, IEEE Trans. Med. Imaging (2025).

[24] M.K. Garvin, M.D. Abramoff, X. Wu, S.R. Russell, T.L. Burns, M. Sonka, Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images, IEEE Trans. Med. Imaging 28 (9) (2009) 1436–1447.

[25] S.J. Chiu, M.J. Allingham, P.S. Mettu, S.W. Cousins, J.A. Izatt, S. Farsiu, Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema, Biomed. Opt. Express 6 (4) (2015) 1172–1194.

[26] A. Mishra, A. Wong, K. Bizheva, D.A. Clausi, Intra-retinal layer segmentation in optical coherence tomography images, Opt. Express 17 (26) (2009) 23719–23728.

[27] A. Yazdanpanah, G. Hamarneh, B. Smith, M. Sarunic, Intra-retinal layer segmentation in optical coherence tomography using an active contour approach, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2009, pp. 649–656.

[28] D.A. Michael, M.K. Garvin, M. Sonka, Retinal imaging and image analysis, IEEE Rev. Biomed. Eng. 3 (169–208) (2010) 1.

[29] R. Yu, Y. Zhang, Y. Tian, Z. Liu, X. Li, J. Gao, CP-UNet: Contour-based probabilistic model for medical ultrasound images segmentation, 2024, arXiv preprint arXiv:2411.14250.

[30] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[31] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention u-net: Learning where to look for the pancreas, 2018, arXiv preprint arXiv:1804.03999.

[32] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, Y. Zhou, Transunet: Transformers make strong encoders for medical image segmentation, 2021, arXiv preprint arXiv:2102.04306.

[33] Z. Huang, J. Sun, Y. Shao, Z. Wang, S. Wang, Q. Li, J. Li, Q. Yu, PolarFormer: a transformer-based method for multi-lesion segmentation in intravascular OCT, IEEE Trans. Med. Imaging 43 (12) (2024) 4190–4199.

[34] P. Mani, N. Ramachandran, P. Naveen, P.V. Ramesh, An enhanced lightweight transformer-based framework for accurate retinal disease classification from OCT images, J. Opt. (2025) 1–20.

[35] J.R. Clough, N. Byrne, I. Oksuz, V.A. Zimmer, J.A. Schnabel, A.P. King, A topological loss function for deep-learning based image segmentation using persistent homology, IEEE Trans. Pattern Anal. Mach. Intell. 44 (12) (2020) 8766–8778.
[36] X. He, W. Song, Y. Wang, F. Poiesi, J. Yi, M. Desai, Q. Xu, K. Yang, Y. Wan, Lightweight retinal layer segmentation with global reasoning, IEEE Trans. Instrum. Meas. 73 (2024) 1–14.
[37] X. He, F. Wu, K. Hu, L. Cui, W. Song, Y. Wan, Fusing multispectral information for retinal layer segmentation, npj Digit. Med. 8 (1) (2025) 39.
[38] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, I.B. Ayed, Boundary loss for highly unbalanced segmentation, in: International Conference on Medical Imaging with Deep Learning, PMLR, 2019, pp. 285–296.
[39] Q.-H. Ho, T.-T. Tran, V.-T. Pham, et al., LiteMamba-bound: A lightweight Mamba-based model with boundary-aware and normalized active contour loss for skin lesion segmentation, Methods 235 (2025) 10–25.
[40] H. Fu, Y. Xu, S. Lin, D.W.K. Wong, B. Mani, M. Mahesh, T. Aung, J. Liu, Multi-context deep network for angle-closure glaucoma screening in anterior segment OCT, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 356–363.
[41] B. Fazekas, G. Aresta, D. Lachinov, S. Riedl, J. Mai, U. Schmidt-Erfurth, H. Bogunović, SD-LayerNet: Semi-supervised retinal layer segmentation in OCT using disentangled representation with anatomical priors, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2022, pp. 320–329.
[42] M. Pekala, N. Joshi, T.A. Liu, N.M. Bressler, D.C. DeBuc, P. Burlina, Deep learning based retinal OCT segmentation, Comput. Biol. Med. 114 (2019) 103445.
[43] Y. Shen, J. Li, W. Zhu, K. Yu, M. Wang, Y. Peng, Y. Zhou, L. Guan, X. Chen, Graph attention u-net for retinal layer surface detection and choroid neovascularization segmentation in oct images, IEEE Trans. Med. Imaging 42 (11) (2023) 3140–3154.
[44] H. Xu, Y. Wu, G2ViT: Graph neural network-guided vision transformer enhanced network for retinal vessel and coronary angiograph segmentation, Neural Netw. 176 (2024) 106356.
[45] M. Moradi, Y. Chen, X. Du, J.M. Seddon, Deep ensemble learning for automated non-advanced AMD classification using optimized retinal layer segmentation and SD-OCT scans, Comput. Biol. Med. 154 (2023) 106512.
[46] S. Li, Y. Feng, H. Xu, Y. Miao, Z. Lin, H. Liu, Y. Xu, F. Li, CAENet: Contrast adaptively enhanced network for medical image segmentation based on a differentiable pooling function, Comput. Biol. Med. 167 (2023) 107578.
[47] H. Xie, W. Xu, Y.X. Wang, X. Wu, Deep learning network with differentiable dynamic programming for retina OCT surface segmentation, Biomed. Opt. Express 14 (7) (2023) 3190–3202.
[48] J. Hu, C. Yu, Z. Yi, H. Zhang, Enhancing robustness of medical image segmentation model with neural memory ordinary differential equation, Int. J. Neural Syst. 33 (12) (2023) 2350060.
[49] J. Li, P. Jin, J. Zhu, H. Zou, X. Xu, M. Tang, M. Zhou, Y. Gan, J. He, Y. Ling, et al., Multi-scale GCN-assisted two-stage network for joint segmentation of retinal layers and discs in peripapillary OCT images, Biomed. Opt. Express 12 (4) (2021) 2204–2220.
[50] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, N. Sang, Bisenet: Bilateral segmentation network for real-time semantic segmentation, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 325–341.
[51] H. Li, P. Xiong, H. Fan, J. Sun, Dfanet: Deep feature aggregation for real-time semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 9522–9531.
[52] X. He, Y. Wang, F. Poiesi, W. Song, Q. Xu, Z. Feng, Y. Wan, Exploiting multi-granularity visual features for retinal layer segmentation in human eyes, Front. Bioeng. Biotechnol. 11 (2023) 1191803.
[53] K. Vinogradova, A. Dibrov, G. Myers, Towards interpretable semantic segmentation via gradient-weighted class activation mapping (student abstract), in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, (no. 10) 2020, pp. 13943–13944.