

Retinal OCT image segmentation with deep learning: A review of advances, datasets, and evaluation metrics[☆]

Huihong Zhang ^{a,b,1}, Bing Yang ^b, Sanqian Li ^b, Xiaoqing Zhang ^b, Xiaoling Li ^b, Tianhang Liu ^b, Risa Higashita ^c, Jiang Liu ^{b,c,d,*}

^a Harbin Institute of Technology, No. 92 West Dazhi Street, Nangang District, Harbin, 150001, Heilongjiang, China

^b Department of Computer Science and Engineering, Southern University of Science and Technology, 1088 Xueyuan Avenue, Shenzhen, 518055, Guangdong, China

^c Research Institute of Trustworthy Autonomous Systems, Southern University of Science and Technology, 1088 Xueyuan Avenue, Shenzhen, 518055, Guangdong, China

^d University of Nottingham Ningbo China, 199 Taikang East Road, 315100, Ningbo, China



ARTICLE INFO

Keywords:

Optical coherence tomography
Deep learning
Review
Ophthalmic imaging
Anatomical structures segmentation
Retinal layer segmentation
Lesion segmentation
Datasets
Evaluation metrics

ABSTRACT

Optical coherence tomography (OCT) is a widely used imaging technology in ophthalmic clinical practice, providing non-invasive access to high-resolution retinal images. Segmentation of anatomical structures and pathological lesions in retinal OCT images, directly impacts clinical decisions. While commercial OCT devices segment multiple retinal layers in healthy eyes, their performance degrades severely under pathological conditions. In recent years, the rapid advancements in deep learning have significantly driven research in OCT image segmentation. This review provides a comprehensive overview of the latest developments in deep learning-based segmentation methods for retinal OCT images. Additionally, it summarizes the medical significance, publicly available datasets, and commonly used evaluation metrics in this field. The review also discusses the current challenges faced by the research community and highlights potential future directions.

Contents

1. Introduction	2
2. OCT imaging and characteristics	3
2.1. Principles and evolution of OCT technology	3
2.2. Protocols of retinal OCT imaging	3
2.3. Normal retinal anatomy and OCT characteristics	4
2.4. Pathological retinal lesions and OCT characteristics	4
3. Public datasets	4
3.1. DUKE dataset	4
3.2. Cell/Kermany dataset	5
3.3. HC-MS dataset	5
3.4. AI challenger/RESC dataset	6
3.5. RETOUCH dataset	6
3.6. OCTA-500 dataset	6
3.7. GOALS dataset	6
4. Evaluation metrics	7
4.1. Region-based metrics	7
4.2. Contour-based metrics	7
4.3. Pixel error-based metrics	8
4.4. Confusion matrix-based metrics	8
4.5. Biomarker-based metrics	8
5. Anatomical structures segmentation	8

[☆] This work was supported in part by National Natural Science Foundation of China (No. 82272086).

* Corresponding author.

E-mail address: liuj@sustech.edu.cn (J. Liu).

5.1.	Full supervision learning methods.....	8
5.1.1.	FCN/UNet and their variants.....	8
5.1.2.	Multi-scale feature fusion.....	9
5.1.3.	Prior knowledge incorporation.....	9
5.1.4.	Uncertainty learning.....	9
5.1.5.	Attention mechanisms.....	10
5.1.6.	Dimension hybrid fusion.....	10
5.2.	Limited supervision learning methods.....	10
5.2.1.	Semi-supervised learning.....	10
5.3.	Summary of retinal OCT anatomical structures segmentation.....	11
6.	Pathological lesions segmentation	12
6.1.	Full supervision learning methods.....	12
6.1.1.	UNet and its variants.....	12
6.1.2.	Multi-scale feature fusion.....	13
6.1.3.	Uncertainty learning.....	13
6.1.4.	Attention mechanisms.....	13
6.1.5.	Multi-task learning	13
6.2.	Limited supervision learning methods.....	14
6.2.1.	Semi-supervised learning.....	14
6.2.2.	Weakly supervised learning.....	14
6.2.3.	Unsupervised learning	15
6.3.	Summary of retinal OCT pathological lesions segmentation.....	15
7.	Discussion	15
7.1.	Challenges	15
7.2.	Future directions	16
8.	Conclusion	17
	CRediT authorship contribution statement	17
	Declaration of competing interest.....	17
	Data availability	17
	References.....	17

1. Introduction

Optical coherence tomography (OCT) is a non-invasive, high-resolution, cross-sectional imaging technology that is capable of capturing the internal microstructure of biological tissues (Huang et al., 1991). Since its inception in 1991, OCT has been widely adopted in various clinical applications. In ophthalmology, especially for ocular disease diagnosis, OCT has become a standard tool for the non-invasive capture of both anatomical and pathological structures (Schuman et al., 2024).

Anatomical structures and pathological lesion segmentation are fundamental tasks, providing crucial biomarkers for the diagnosis of ocular diseases. For instance, retinal thickness serves as a diagnostic indicator for macular edema, while the size and location of edema are critical for assessing disease progression and formulating appropriate treatment plans. Since manual segmentation is time-consuming and labor-intensive, most commercial OCT devices offer automatic retinal layer segmentation with satisfactory accuracy in normal eyes. However, these devices typically provide a limited number of distinguishable layers, and the segmentation performance decreases in the presence of speckle noise, artifacts, or lesions. Consequently, there is a pressing need for more effective automated OCT image segmentation algorithms.

OCT image segmentation presents several significant challenges: (1) Reduced resolution of deeper tissues due to diminishing light intensity with depth, makes it difficult to distinguish between histologically distinct sub-layers within the retinal and choroidal microarchitecture. (2) Speckle noise inherent in OCT images, which can obscure important details and complicate the segmentation process. (3) Variability in retinal lesions, which differ in shape, size, and location, making it challenging to recognize their patterns accurately.

In the face of these challenges, early OCT segmentation algorithms typically relied on thresholding, edge detection, region-based, and graph theory methods, utilizing manually extracted features such as intensity, thickness, texture, and shape for decision-making (Kafieh

et al., 2013). These methods offer high interpretability but depend heavily on manually extracted features, resulting in limited accuracy and generalization performance in complex scenarios.

In recent years, deep learning (DL) has attracted considerable attention, with convolutional neural networks (CNNs) being among the most successful and widely used architectures for visual tasks. CNNs can automatically extract high-level features from large datasets, enabling them to achieve high accuracy, robust generalization, and rapid inference (LeCun et al., 2015). Currently, CNN-based deep learning methods have become the dominant approach in medical image analysis (Siddique et al., 2021; Wang et al., 2022a), including OCT image segmentation.

Several review studies have been conducted on deep learning algorithms for OCT image segmentation (Pekala et al., 2019a; Ran et al., 2021; Alizadeh Eghtedar, 2022; Viedma et al., 2022). While these works provide essential contributions towards OCT segmentation research, three critical gaps limit their comprehensive utility. First, prior efforts predominantly focus on anatomical layer segmentation (e.g., retinal layers by Pekala et al., 2019a, choroidal layers by Alizadeh Eghtedar, 2022), with limited discussion on pathological lesion characterization crucial for diseases like diabetic macular edema. Second, current taxonomies largely follow conventional architectural classifications (e.g., CNN vs. Encoder-decoder by Viedma et al., 2022), with limited emphasis on innovations addressing domain-specific challenges like annotation scarcity. Thirdly, despite the rapid development of advanced techniques such as uncertainty learning and attention mechanisms in medical imaging analysis, their applications in OCT segmentation remain underexplored in existing reviews.

This review provides a comprehensive summary of recent DL-based OCT image segmentation methods, and discusses potential challenges and future research directions. We conducted a literature search using the keywords “OCT” and “segmentation” across influential journals/conferences in the fields of deep learning, computer vision, medical image analysis, ophthalmology, and optical imaging. After filtering out irrelevant content, 60 DL-based OCT image segmentation papers were selected for review. Fig. 1 shows the number of publications over

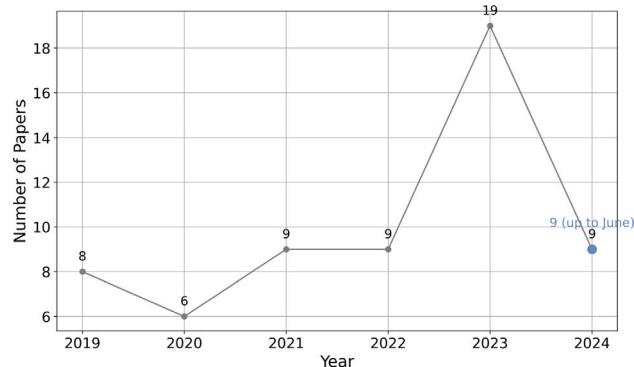


Fig. 1. Number of reviewed DL-based OCT segmentation papers (2019–2024).

time, while Fig. 2 illustrates the source distribution of these papers. Additionally, the structure of this review is outlined in Fig. 3. The main contributions of this review are summarized as follows:

(1) Recent Advances: This review presents an analysis of 60 recent state-of-the-art deep learning studies (from January 2019 to June 2024) applied to OCT image segmentation. By concentrating on recent developments, it provides a current and comprehensive overview of the field.

(2) Systematic Overview: The review systematically highlights the diversity of supervision methods applied in this task, including fully supervised, semi-supervised, weakly supervised, and unsupervised approaches. We introduce an innovation-driven taxonomy that categorizes methods based on their core technical strategies (e.g., multi-scale feature fusion, prior knowledge incorporation, uncertainty learning, attention mechanisms, dimension hybrid fusion, multi-task learning) rather than backbone architectures. Additionally, the review bridges the technical-clinical knowledge gap by introducing retinal OCT imaging, with detailed descriptions of key anatomical structures and pathological lesions.

(3) Comprehensive Resources: The review systematically summarizes 20 publicly available datasets and 15 evaluation metrics used in OCT image segmentation. We further provide implementable code templates for key algorithms via an open repository¹, accompanied by evaluation metrics and public datasets for performance benchmarking.

(4) Challenges and Future Directions: This review identifies and analyzes the ongoing challenges in dataset construction, algorithmic performance, and evaluation standardization present in current studies. Furthermore, it suggests possible future directions to address these challenges, providing insights into possible avenues for further exploration in the field.

2. OCT imaging and characteristics

2.1. Principles and evolution of OCT technology

Optical coherence tomography was first demonstrated by Huang et al. in 1991 for biological tissue imaging (Huang et al., 1991), using low-coherence interferometry to achieve cross-sectional microstructural visualization. As shown in Fig. 4, the light from the light source is split into two beams, directing them to the reference mirror and the sample, respectively. The light projected onto the sample is reflected by anatomical structures at different depths, interferes with the reference beam in the optical fiber coupler, and is then recorded by the detector for further processing (Aumann et al., 2019).

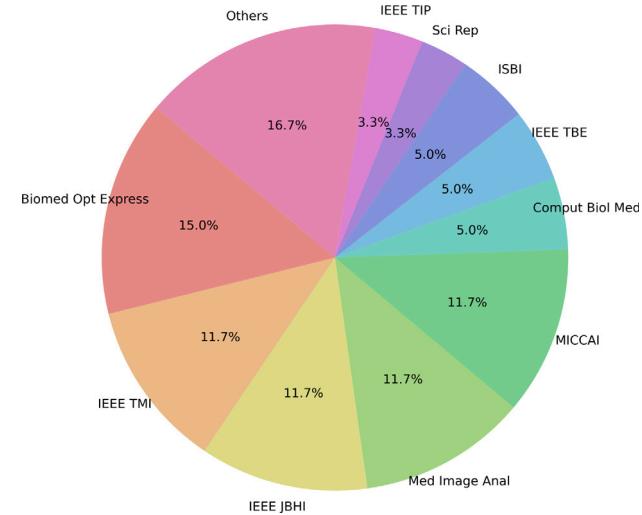


Fig. 2. Distribution of reviewed DL-based OCT segmentation papers by journal/conference.

Following the initial time-domain OCT (TD-OCT) achieving 10–20 μm axial resolution at 400 A-scans/s, OCT technology has undergone two transformative advancements. Spectral-domain OCT (SD-OCT) replaced mechanical depth scanning with spectrometer-based Fourier-domain detection, enhancing axial resolution to 3–7 μm and imaging speed to 20,000–100,000 A-scans/s, enabling systems like Carl Zeiss Cirrus HD-OCT. Swept-source OCT (SS-OCT) further advanced performance through wavelength-tunable lasers, achieving unprecedented acquisition rates (400,000 A-scans/s) and penetration depths (2.6–3.0 mm), exemplified by Topcon DRI Triton for retinal and choroidal imaging² (Zappi et al., 2023). In the current retinal OCT market, Heidelberg Engineering, Carl Zeiss Meditec, Optovue, and Topcon collectively maintain leadership through advanced OCT implementations (Global Growth Insights, 2024).

2.2. Protocols of retinal OCT imaging

In clinical OCT examinations for ocular diseases, the imaging protocols prioritize the macular region and optic nerve head (ONH). Macular-centered OCT is predominantly employed to visualize foveal microstructures and pathological lesions, making it the clinical benchmark for tracking various diseases (e.g., diabetic macular edema and age-related macular degeneration). Given its demonstrated clinical utility, macular-centered OCT has been widely used in ophthalmic diagnosis and research. ONH-centered OCT typically assesses pathological alterations of the optic disc (e.g., optic disc edema and optic nerve atrophy), proving particularly effective in the diagnosis of glaucoma.

The selection of scanning protocols fundamentally depends on the required spatial resolution characteristics. As demonstrated in the macular cube scan of Fig. 5, **A-line** (or A-scan) is a one-dimensional image parallel to the direction of incident light. **B-scan** is a two-dimensional image, formed by combining multiple A-scans perpendicular to the surface of the retina, enabling visualization of intraretinal layer architecture. **C-scan** or **en face** is a two-dimensional image parallel to the surface of the retina, revealing vascular patterns and focal pathologies within specific layers. Clinically, B-scans are the preferred modality for retinal imaging due to their superior axial resolution, enabling precise

² Although contemporary studies have begun to use the expression 'posterior segment OCT' to acknowledge its choroidal visualization capabilities, major textbooks and literature still employ the traditional term 'retinal OCT'. We follow this convention.

¹ https://github.com/ZhangHH233/Retinal_OCT_Image_Segmentation_via_Deep_Learning

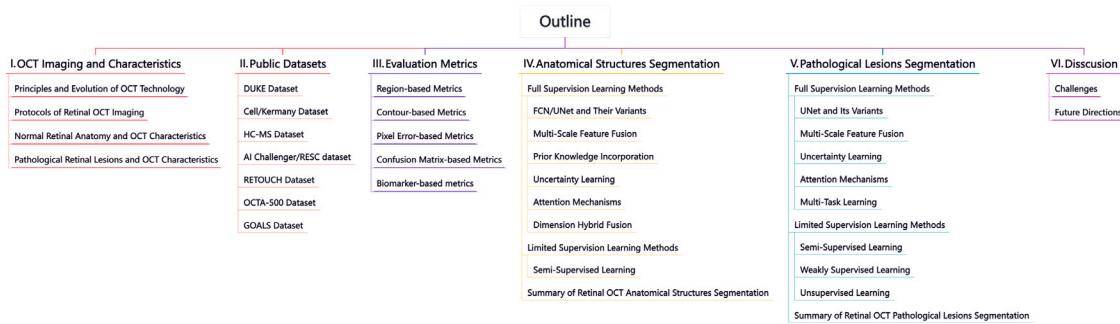


Fig. 3. Outline of the Main Contents.

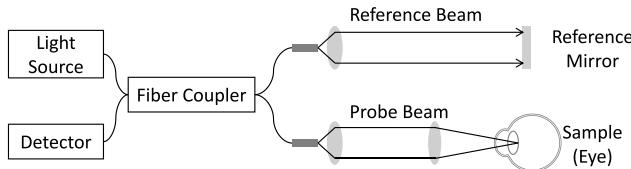


Fig. 4. Working principle of OCT.

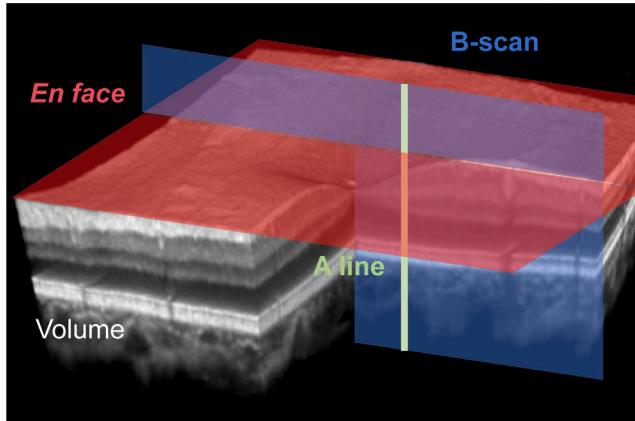


Fig. 5. Illustration of macular OCT volume and cross sections.

delineation of intraretinal layers essential for early-stage pathology identification.

OCT imaging demonstrates tissue-specific optical reflectivity. In grayscale retinal OCT, the structure and water content of tissues are the main factors influencing image quality. As shown in Fig. 6, tissues arranged in horizontal layers (perpendicular to the direction of incidence) reflect light strongly, appearing as bright areas in grayscale OCT images; in contrast, tissues with vertical structures or higher water content appear darker. On the other hand, imaging of deeper tissues is affected by light absorption and scattering, resulting in lower resolution (Aumann et al., 2019).

2.3. Normal retinal anatomy and OCT characteristics

In anatomical classification, the retinal compartment extends axially from the nerve fiber layer (NFL) to the retinal pigment epithelium (RPE), principally responsible for phototransduction cascade and neuromotor signal integration. The choroidal compartment is defined between Bruch's membrane (BM) and the choroidal scleral junction (CSJ), fulfilling critical roles in oxygen-nutrient support and metabolic homeostasis through its vascular networks. The microarchitecture of

both retinal and choroidal tissues comprises histologically distinct sub-layers (Fig. 6(a)). OCT B-scans distinguish these sub-layers with micron-scale resolution. Fig. 6(b) illustrates the International Nomenclature for OCT consensus-based stratification system of retinal and choroidal sublayers (Staurenghi et al., 2014). This stratification capability enables distinct clinical applications: (1) The combination of morphological analysis of layer-specific architectures and quantitative measurement of diagnostic biomarkers (e.g., evaluation of the retinal layer structure integrity and the thinning of the NFL for glaucoma diagnosis), and (2) Pathological categorization through anatomical layer referencing (e.g., differentiating intraretinal fluid, subretinal fluid, and pigment epithelial detachment in macular edema).

2.4. Pathological retinal lesions and OCT characteristics

Retinal OCT facilitates comprehensive disease diagnosis and grading through two complementary mechanisms: (1) the optical reflection characteristics of lesions, and (2) abnormal changes in anatomical structures (Guanglu and Wenbin, 2009). Fig. 7 presents representative examples of retinal OCT images depicting common lesions.

From the reflectivity perspective, pathological lesions such as drusen deposition, choroidal neovascularization (CNV), and hyperreflective foci (HRF) manifest as hyperreflective signals on OCT imaging. Conversely, lesions including intraretinal fluid (IRF), subretinal fluid (SRF), pigment epithelial detachment (PED), and cystoid macular edema (CME) exhibit hyporeflective appearances.

From the structural perspective, distinct pathologies demonstrate characteristic alterations: the advanced non-neovascular age-related macular degeneration (AMD) exhibits geographic atrophy (GA); glaucoma can lead to NFL and GC-IPL thinning; DME causes retinal thickening through edema; while traumatic injury/pathologic myopia may induce the formation of macular holes (MH).

3. Public datasets

Deep learning algorithms require extensive datasets for model training and performance testing, making access to public datasets crucial. This section summarizes 20 publicly available retinal OCT datasets published between 2011 and 2023 in Table 1, providing detailed descriptions of the most commonly used datasets. Fig. 8 displays examples from some of these datasets.

3.1. DUKE dataset

The DUKE dataset is one of the earliest publicly available retinal OCT datasets, released by DUKE University. Since 2012, it has included four sub-datasets: DUKE-AMD (Chiu et al., 2012), DUKE-WLOA (Farsiu et al., 2014), DUKE-DME (Srinivasan et al., 2014), and DUKE-BOE (Chiu et al., 2015).

DUKE-AMD: This dataset was derived from the "Age-Related Eye Disease Study 2 Ancillary Spectral Domain Optical Coherence Tomography" (A2 A SD-OCT) project. Volumetric scans were acquired using

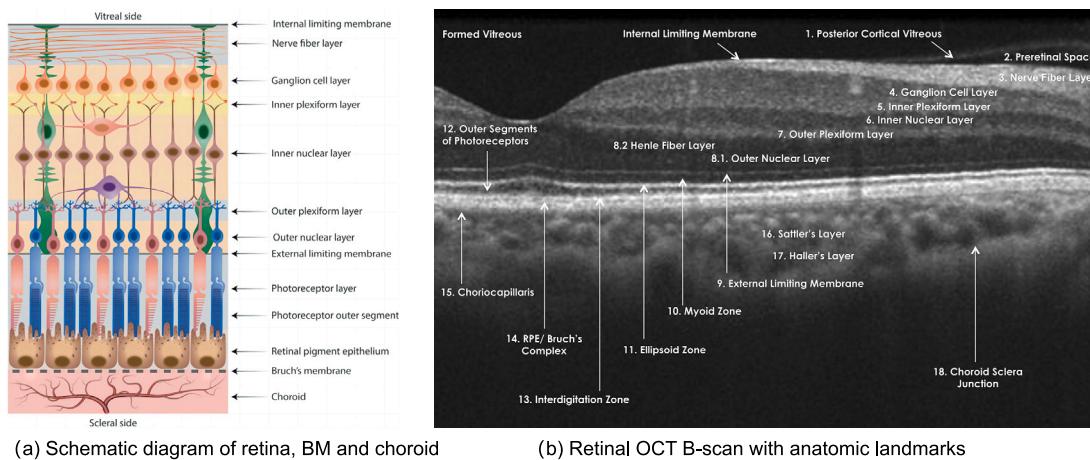


Fig. 6. Anatomy of healthy retina and choroid. (a) is a schematic diagram (Ferrara et al., 2021), and (b) is a retinal OCT B-scan (Staurenghi et al., 2014). From Posterior cortical vitreous to choroidal scleral junction, the full names and abbreviations of these anatomic landmarks are as follows: 1. Posterior cortical vitreous (PCV), 2. Pre-retinal space (PRS), 3. Nerve fiber layer (NFL), 4. Ganglion cell layer (GCL), 5. Inner plexiform layer (IPL), 6. Inner nuclear layer (INL), 7. Outer plexiform layer (OPL), 8. Inner half: Henle's nerve fiber layer; outer half: outer nuclear layer, 9. External limiting membrane (ELM), 10. Myoid zone of the photoreceptors (MZP), 11. Ellipsoid zone of the photoreceptors (EZP), 12. Outer segments of the photoreceptors (OSP), 13. Interdigitation Zone (IZ), 14. Retinal pigment epithelium (RPE)/Bruch's membrane (BM) complex, 15. Choriocapillaris, 16. Sattler's layer, 17. Haller's layer, 18. Choroidal scleral juncture (CSJ) (Staurenghi et al., 2014).

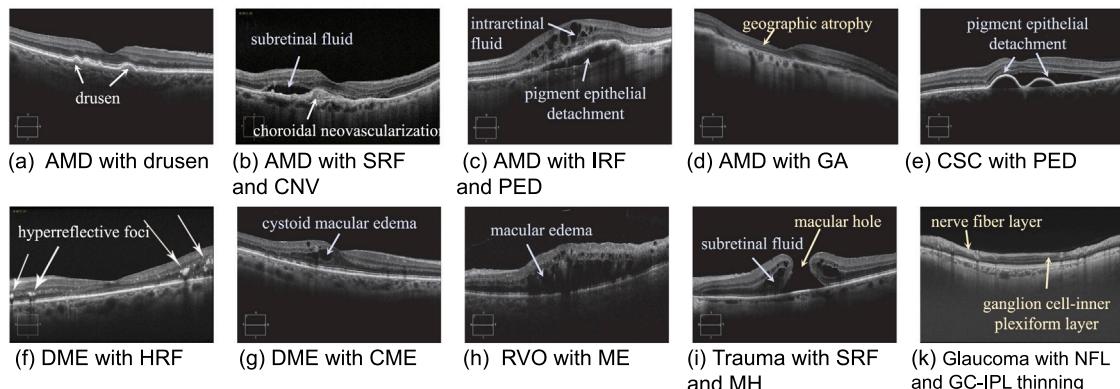


Fig. 7. OCT imaging features of major retinal disorders: Age-related Macular Degeneration (AMD), Central Serous Chorioretinopathy (CSC), Diabetic Macular Edema (DME), Retinal Vein Occlusion (RVO), traumatic injury, and glaucoma. Panels (a)–(i) demonstrate macular-centered OCT scans with distinct pathological presentations (Schuman et al., 2024), while panel (j) illustrates ONH-centered OCT findings in glaucoma (Fang et al., 2022). Color-coded annotations primarily correspond to dominant imaging signatures: white (hyperreflective lesions), blue (hyporeflective lesions), and yellow (predominant structural disruption). Note that some pathologies may exhibit overlapping features.

SD-OCT imaging systems from Bioptigen, Inc. (Research Triangle Park, NC) located at the four clinic sites. It includes OCT images from 20 AMD patients, with annotated retinal layer boundaries (Chiu et al., 2012).

DUKE-WLOA: This dataset was also derived from the A2 A SD-OCT project. It includes OCT data from 269 AMD patients and 115 healthy participants, along with corresponding retinal annotations (Far-siu et al., 2014).

DUKE-DME: This dataset includes OCT images from 15 healthy participants, 15 AMD patients, and 15 DME patients. It was designed for disease diagnosis (classification) tasks (Srinivasan et al., 2014).

DUKE-BOE: The DUKE-BOE dataset includes OCT images from 10 DME patients. The lateral and azimuthal resolutions range from 10.94 to 11.98 $\mu\text{m}/\text{pixel}$ and 118 to 128 $\mu\text{m}/\text{pixel}$, respectively. Annotations cover retinal and DME lesion segmentation (Chiu et al., 2015).

3.2. Cell/Kermany dataset

The Cell dataset is a classification dataset published by Kermany et al. (2018). Data collection occurred between 2013 and 2017 at the Shiley Eye Institute of the University of California, San Diego; the California Retinal Research Foundation; Medical Center Ophthalmology Associates; the Shanghai First People's Hospital; and the Beijing

Tongren Eye Center, using Spectralis OCT devices (Heidelberg Engineering, Heidelberg, Germany). Initially, 207,130 OCT images were collected. After quality screening, 108,312 images from 4,686 patients were retained, including 37,206 CNV images, 11,349 DME images, 8,617 Drusen images, and 51,140 healthy images. Since the Cell dataset was released, ongoing efforts have been made to clean and correct the classification annotations, resulting in the creation of several smaller subsets.

3.3. HC-MS dataset

The HC-MS dataset includes OCT images of 35 right eyes collected by He et al. (2019) at Johns Hopkins Hospital in the United States (He et al., 2019b). The dataset comprises 14 healthy subjects and 21 multiple sclerosis (MS) patients. The imaging was performed using a Spectralis OCT device (Heidelberg Engineering, Heidelberg, Germany). The B-scan resolution varied slightly between subjects. The lateral resolution had a mean of 5.8 μm across all subjects, while the axial resolution was 3.9 μm . The imaging area was approximately 6 \times 6 mm^2 . Annotations in this dataset include binary labels for healthy/MS classification and segmentation labels for eight retinal layers (NFL, GC-IPL, INL, OPL, ONL, IS, OS, RPE).

Table 1

Overview of publicly available retinal OCT datasets (H: Healthy subjects; DR: Diabetic Retinopathy; G: Glaucoma; FAZ: Foveal Avascular Zone; RL: Retinal Layer; CL: Choroidal Layer).

Reference	Dataset	Year	Data volume	Disease	Label	Device
Chiu et al. (2012)	DUKE-AMD	2012	20 volumes	20 AMD	RL	Bioptigen SD-OCT
Farsiu et al. (2014)	DUKE-WLOA	2014	38800 B-scans	269 AMD, 115 H.	RL	Bioptigen SD-OCT
Srinivasan et al. (2014)	DUKE-DME	2014	45 volumes	15 Dry AMD, 15 DME, 15 H.	Disease Categories	Spectralis SD-OCT
Chiu et al. (2015)	DUKE-BOE	2015	6 volumes	6 DME	RL, Fluids	Spectralis SD-OCT
Tian et al. (2015)	HEG	2015	10 × 61 B-scans	H.,	RL	Spectralis SD-OCT
Mahmudi et al. (2014)	SPIE	2014	19 volumes	19 H.	—	Topcon
Tian et al. (2016)	MIAMI	2016	50 B-scans	10 mild DR	RL	Spectralis SD-OCT
Wu et al. (2016)	OPTIMA	2016	30 B-scans	IRF	IRF	Zeiss Cirrus, Heidelberg Spectralis, Topcon 3D 2000, Nidek RS3000
Kashefpur et al. (2017)	Isfahan MISP	2017	—	—	RL, Vessel, Noise	Topcon 3D OCT1000
Kermany et al. (2018)	Cell	2018	108,312 B-scans	H., CNV, DME, Drusen	Disease Categories	Spectralis OCT
Hassan et al. (2018)	BIOMISA	2018	2497 B-scans, 19 C-scans, 64 fundus scans	14 AMD, 13 ME, 50 H., 26 G.	RL, Fluids	TOPCON 3D OCT 2000
He et al. (2019b)	HC-MS	2019	49 B-scans	14 H., 21 MS	RL	Spectralis OCT
Maetschke et al. (2019)	Zenodo	2019	1100 B-scans	847 G., 263 H.	Disease Categories	Cirrus SD-OCT
Hu et al. (2019)	AI-challenger/RESC	2019	100 volumes	SRF, PED	Edema, Disease Categories	Zeiss SD-OCT
Bogunović et al. (2019)	RETOUCH	2019	112 volumes (1134 B-scans)	IRF, SRF, PED	Edema	Cirrus, Nidek, Spectralis
Gholami et al. (2020)	OCTID	2020	470 B-scans	206 H., 102 MH, 55 AMD, 107 DR	RL, Disease Categories	Cirrus HD-OCT
Raja et al. (2020)	RAJA-Glaucoma	2020	50 OCT volumes, fundus scans	18 H., 32 G.	RL, Disease Categories	TOPCON 3D OCT-1000
Li et al. (2020a)	OCTA-500	2020	500 volumes	H., AMD, DR, CNV, CSC, RVO	RL, Vessel, FAZ, Disease Categories	RTVue-XR SD-OCT
Melinščak et al. (2021)	AROI	2021	1136 B-scans	24 AMD	RL, Fluids	Cirrus HD-OCT
Fang et al. (2022)	GOALS	2022	300 B-scans	53 H., 13 G.	RL, CL, Disease Categories	TOPCON DRI Swept Source OCT

Download link and more information available at <https://github.com/ZhangHH233/Public-available-retinal-OCT-datasets>.

3.4. AI challenger/RESC dataset

The dataset was sourced from the “Artificial Intelligence Challenger” competition platform (Hu et al., 2019). It includes OCT images of patients with Subretinal Fluid (SRF) and Pigment Epithelial Detachment (PED), with pixel-level annotations for the lesions. The dataset is divided into training, validation, and testing sets, comprising 70, 15, and 15 cases, respectively. Each case consists of 128 slices, each with a resolution of 512 × 1024 pixels. Note that annotations are available only for the training and validation sets.

3.5. RETOUCH dataset

The RETOUCH dataset originated from the 2017 MICCAI conference and focuses on the segmentation challenge of retinal OCT edema. The goal was to detect and segment various types of fluid lesions in macula-centered OCT data from different manufacturers (Bogunović et al., 2019). The devices used include Cirrus HD-OCT (Zeiss Meditec), Spectralis (Heidelberg Engineering), and T-1000/T-2000 (Topcon). Each device has different resolution parameters, and the dataset roughly balances the data from these devices. The data were collected from 112 subjects, with half having macular edema secondary to AMD and the

other half having edema secondary to RVO. The dataset provides category and segmentation annotations for three types of edema lesions: Intraretinal Fluid (IRF), Subretinal Fluid (SRF), and Pigment Epithelial Detachment (PED).

3.6. OCTA-500 dataset

The OCTA-500 dataset, published by the research group at Nanjing University, is a comprehensive dataset that includes 316 Optical Coherence Tomography Angiography (OCTA) volumes and corresponding OCT volumes from 293 subjects with AMD, RVO, CSC, CNV, DR, and other conditions (Li et al., 2020a). The dataset includes extensive image and annotation data. Specifically, it comprises two fields of view, two modes (OCT/OCTA volumes), six projection types, four text labels, and seven segmentation labels.

3.7. GOALS dataset

The GOALS dataset was released at the 2022 MICCAI OMIA workshop for Glaucoma OCT Analysis and Layer Segmentation (GOALS) challenge, containing 300 circum papillary OCT images (Fang et al., 2022). These images were obtained using the TOPCON DRI Swept

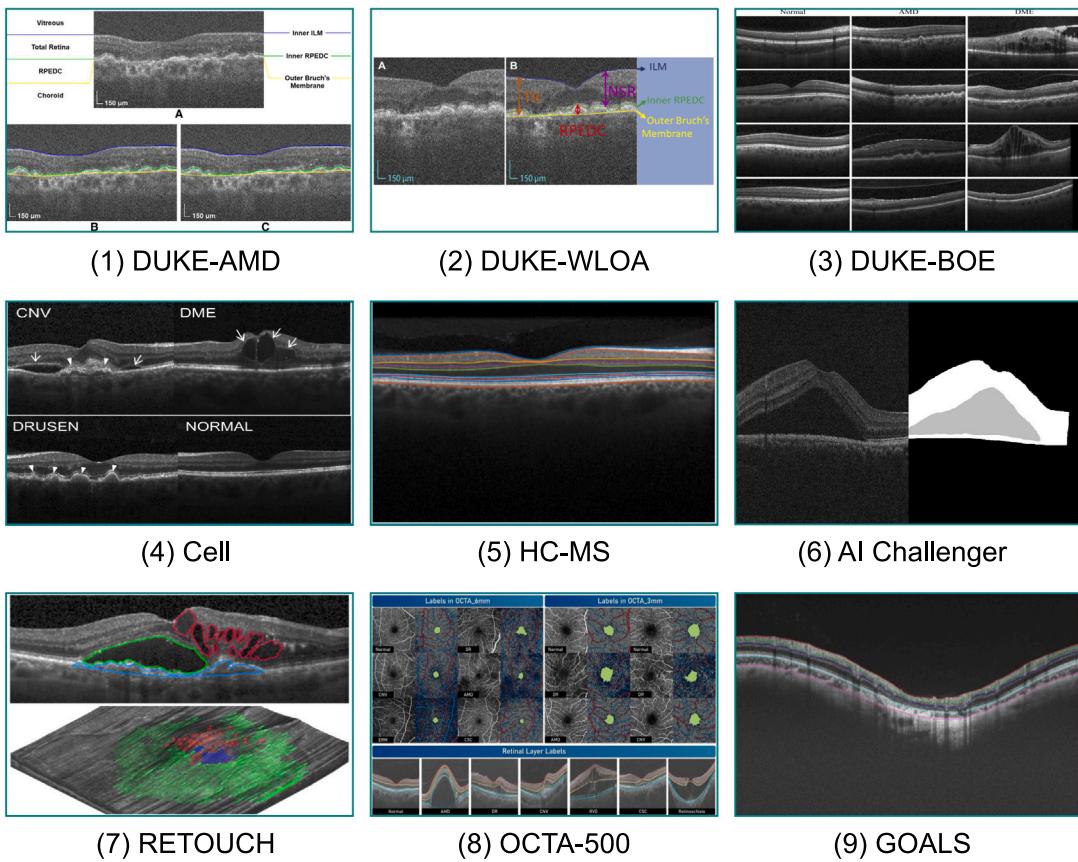


Fig. 8. Illustrations of retinal OCT samples from public datasets.

Source OCT at Zhongshan Ophthalmic Center, saved in PNG format with a resolution of 1100×800 pixels. It provides classification labels (glaucoma/healthy) and segmentation labels for the NFL, GC-IPL, and choroidal layers.

4. Evaluation metrics

Segmentation evaluation typically measures the accuracy of pixel classification and the correctness of localization, aiming to fully assess the similarity between the predictions and ground truth (Müller et al., 2022). In the field of OCT image segmentation, a large variety of evaluation metrics have been used.

4.1. Region-based metrics

The Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) are the most commonly used evaluation metrics in 2D medical image segmentation tasks. They assess the overlap between the predicted and ground truth foreground regions. Their formulas are as follows:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (1)$$

$$IoU = \frac{|X \cap Y|}{|X \cup Y|} \quad (2)$$

where X is the set of pixels predicted as foreground, and Y is the set of pixels in the ground truth foreground.

Although both DSC and IoU assess segmentation accuracy based on overlap, DSC is more sensitive to small structures, making it particularly well-suited for evaluating lesion segmentation.

4.2. Contour-based metrics

Accurate contour prediction is crucial for complex and fine-grained segmentation tasks. Hausdorff Distance (HD) serves as a distance metric to measure the similarity between two sets, commonly employed for comparing the similarity of shapes or contours. The calculation of HD is as follows:

$$HD = \max \left\{ \sup_{x \in P} \inf_{y \in G} d(x, y), \sup_{y \in G} \inf_{x \in P} d(x, y) \right\} \quad (3)$$

Where P and G denote the sets of pixel points on the Predicted contour and Ground truth, respectively. \sup and \inf denote Supremum and Infimum, respectively. $d(\cdot)$ representing a distance function (e.g., Euclidean distance).

Since HD considers only the maximum distance between points, it is sensitive to outliers. Therefore, a more commonly used improved version is the 95% Hausdorff Distance (HD95), which introduces confidence intervals. To calculate HD95:

1. Sort the combined list of distance values in ascending order.
2. Identify the value at the 95th percentile of the sorted list. The HD95 is the value below which 95% of the distance values fall.

Average Symmetric Surface Distance (ASSD) is a metric used to measure the average distance between the surfaces, measuring the distance in both directions (from set P to set G and vice versa). ASSD provides a symmetric measure of surface similarity, with its calculation formula as follows:

$$ASSD = \frac{1}{|P|} \sum_{x \in P} \min_{y \in G} d(x, y) + \frac{1}{|G|} \sum_{y \in G} \min_{x \in P} d(x, y) \quad (4)$$

Mean Absolute Difference (MAD) typically refers to the average of the absolute differences between corresponding elements of two

sets of values, which is often used to measure the error in boundary positioning.

$$MAD = \frac{1}{|N|} \sum_{i=1}^N |P_i - G_i| \quad (5)$$

Among these contour assessment metrics, HD95 and ASSD are favored for their robustness and reliability, as they provide balanced error estimates while minimizing the impact of outliers. However, MAD is also widely used due to its computational simplicity and straightforward interpretation.

4.3. Pixel error-based metrics

Based on pixel-error statistics, metrics such as Mean Squared Error (**MSE**), and Root Mean Squared Error (**RMSE**) are also used. Here, x_i and y_i represent the predicted and true values for the i th sample, respectively, and n denotes the number of samples. The calculation formulas for these metrics are as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2 \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (7)$$

In medical image segmentation, MSE and RMSE are occasionally used to assess boundary errors. However, they are used less frequently compared to DSC and MAD, and are rarely employed as the sole metrics.

4.4. Confusion matrix-based metrics

Classification metrics based on the confusion matrix, such as Sensitivity, Specificity, Accuracy, and Area Under the Receiver Operating Characteristic Curve are also used for evaluating segmentation performance.

Accuracy (**Acc**) measures the proportion of correctly predicted samples out of the total samples. The formulas for these metrics are as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

Sensitivity (**Sens**), also known as **Recall** or True Positive Rate (**TPR**), focuses on the model's ability to detect true positives.

$$Sens = \frac{TP}{TP + FN} \quad (9)$$

Precision (**Pre**), also known as Positive Predictive Value (**PPV**), focuses on the accuracy of the positive predictions.

$$Pre = \frac{TP}{TP + FP} \quad (10)$$

Specificity (**Spec**), also known as True Negative Rate (**TNR**), assesses the model's ability to correctly identify true negatives.

$$Spec = \frac{TN}{TN + FP} \quad (11)$$

In these formulas:

- TP (True Positives): correctly predicted positive cases.
- FN (False Negatives): actual positive cases incorrectly predicted as negative.
- TN (True Negatives): correctly predicted negative cases.
- FP (False Positives): actual negative cases incorrectly predicted as positive.

Area Under the Receiver Operating Characteristic Curve (**AUC**) measures the diagnostic ability of a classifier by visualizing its performance at different discrimination thresholds. It can be calculated in two stages: (1) Plot a Receiver Operating Characteristic (ROC) curve of TPR vs. FPR

at various threshold settings. (2) Calculate the AUC, which can be done using numerical methods such as the trapezoidal rule (Müller et al., 2022).

In medical image segmentation, the confusion matrix-based metrics of Acc, Sens, Pre, and Spec are often used in conjunction to provide a comprehensive evaluation of the model's performance. AUC, which was once widely used for evaluating classification models, is less used in the current segmentation model evaluation.

4.5. Biomarker-based metrics

Common biomarker-based metrics that involve OCT image analysis include retinal/choroidal Thickness Difference (**TD**) and similarity of Vascularity Index (**VI**), among others.

$$TD = \frac{1}{n} \sum_{i=1}^n d(t_i, \hat{t}_i) \quad (12)$$

where t is the true thickness measurement at the i th location, \hat{t}_i is the predicted thickness.

$$VI = \frac{\text{Area occupied by blood vessels}}{\text{Total area of interest}} \quad (13)$$

Since the goal of OCT tissue and lesion segmentation is to support diagnosis, biomarker-based metrics theoretically align well with this task. However, based on the literature reviewed in this study, their use is not widespread. This may be due to researchers from computer science backgrounds lacking sufficient medical knowledge. This review offers a comprehensive overview of the medical background and evaluation metrics for tissue and lesion segmentation, intending to enhance their use in algorithm development and evaluation, and promoting their integration into ophthalmic clinical applications.

5. Anatomical structures segmentation

Retinal and choroidal layer segmentation is a fundamental task for the morphological and quantitative analysis of anatomical structures, providing key biomarkers for diagnosing various diseases. This section reviews the literature regarding retinal and choroidal layer segmentation. Based on the data and annotations used for model training, these studies are categorized into fully supervised learning methods and limited supervision learning methods.

5.1. Full supervision learning methods

When trained using a fully supervised learning approach, each training sample consists of an input image paired with corresponding pixel-level labels. The model learns to predict by establishing a mapping between the input and the labels.

5.1.1. FCN/UNet and their variants

Fully Convolutional Network (FCN) (Long et al., 2015) and UNet (Ronneberger et al., 2015) are commonly used as baselines for segmentation tasks. FCN extends traditional CNN classification networks (Krizhevsky et al., 2012) by incorporating deconvolution layers, enabling pixel-level classification for image segmentation. UNet, on the other hand, is built on an encoder-decoder architecture with skip connections. This architecture enables the extraction of high-level semantic features while preserving low-dimensional spatial information. Due to their effectiveness in image segmentation tasks, many studies on retinal layer segmentation adopt FCNs and UNets as backbone architectures. For instance, Pekala et al. (2019) utilized a Deep Dense FCN (DenseNet-FCN) as the backbone network to predict the boundaries of four sub-layers in the retina, and BM boundary in the choroid from OCT images (Pekala et al., 2019b). Subsequently, a Gaussian filter was used for post-processing. Gende et al. (2023) employed a UNet-based method to automatically segment retinal layers from OCT

images of patients with the most prevalent neurodegenerative diseases, including Alzheimer's disease, Parkinson's disease, multiple sclerosis, and essential tremor³, as well as healthy control subjects (Gende et al., 2023).

5.1.2. Multi-scale feature fusion

Several studies have leveraged multi-scale feature extraction and fusion to robustly segment targets across varying scales. By utilizing atrous convolutions to enlarge the receptive field without sacrificing spatial resolution, these methods generate multi-scale feature maps while preserving fine spatial details (Gu et al., 2019; Li et al., 2020b). For instance, Gu et al. (2019) introduced a Context Encoding Network (CE-Net) for retinal/choroidal layer segmentation (Gu et al., 2019). This work integrates the advantages of dense atrous convolution blocks and residual multi-kernel pooling, enabling the extraction of contextual information over multiple scales. Li et al. (2020b) proposed a DeepRetina network that utilizes an Atrous Spatial Pyramid Pooling (ASPP) (Chen et al., 2017) to extract multi-scale information and facilitate retinal layer segmentation.

5.1.3. Prior knowledge incorporation

Anatomical priors, such as the shape and arrangement of retinal and choroidal layers, have been widely used in OCT segmentation studies. Such prior knowledge enhances the model's ability to distinguish different structures, resulting in more accurate and realistic segmentation predictions. Based on how prior knowledge is incorporated into the algorithms, these studies can be categorized into four main areas: feature learning, model design, regularization, and post-processing.

Researchers have explored incorporating feature descriptors (Ruan et al., 2019; Cansiz et al., 2023) and target shape masks (Wang et al., 2024) during network training to implement prior knowledge-guided feature learning methods. For feature descriptors-based priors, Ruan et al. (2019) presented a two-stage segmentation method that uses retinal thickness as prior knowledge (Ruan et al., 2019). In the first stage, an FCN was used to learn the characteristics of specific retinal layers and extract five boundaries as initial values. In the second stage, a multiphase level set evolution guided by retinal thickness priors was performed to identify nine sub-layer boundaries. Similarly, Cansiz et al. (2023) developed FourierNet, a shape-preserving network for segmenting the Henle's Fiber Layer (HFL) (Cansiz et al., 2023). The FourierNet utilized a set of Fourier descriptors to represent the HFL shape prior and defined a regression task to learn these descriptors. A cascade network then performed parallel learning of descriptor regression and classification, feeding the Fourier descriptors and the original image into an encoder-decoder network to produce the HFL segmentation results. For mask-based priors, Wang et al. proposed PGKD-Net, which effectively leverages retinal structural information to enhance choroidal region feature extraction (Wang et al., 2024). The PGKD-Net consists of a prior-mask guided network that uses the region below BM as a prior mask to eliminate background noise, and a knowledge-diffusive network to adaptively aggregate complementary information from deep and shallow features for refined segmentation.

Another approach involves integrating prior knowledge into the model design by embedding modules that focus on specific regions and features, thereby enhancing the model's robustness to specific scenarios or diseases (Li et al., 2021; He et al., 2021, 2023a). For instance, Li et al. (2021) introduced a graph convolutional network to simultaneously label the nine retinal layers and the optic disc in peripapillary OCT images (Li et al., 2021). Based on the disc/non-disc distribution and the layered structure of the non-disc region in peripapillary OCT images, this method constructs a graph-based representation as an anatomical

prior. This prior is then incorporated between the encoder and decoder of the UNet to address segmentation challenges related to retinal layer thickness variations. Based on their prior work (He et al., 2019a), He et al. (2021) utilized a ResUNet architecture to model the positions of eight retinal sub-layer surfaces, achieving smooth and continuous surfaces with correct topological structure (He et al., 2021). In 2023, they further introduced a longitudinal deep network for consistent OCT layer segmentation, revealing mild thickness changes in multiple sclerosis patients (He et al., 2023a). This approach involves a longitudinal network to extract features from a series of 2D OCT B-scans and a novel longitudinal fusion network to incorporate longitudinal priors.

Additionally, some studies incorporate prior knowledge into regularization by designing loss functions based on prior information, such as boundary smoothness and the sequential arrangement of retinal layers, guiding the model toward topologically accurate predictions (He et al., 2019a; Zhang et al., 2020a; Fazekas et al., 2022). For instance, He et al. (2019) developed a topological constraint module to ensure the continuity of the segmentation surfaces (He et al., 2019a). Zhang et al. (2020) incorporated choroidal thickness distribution as a prior and proposed a thickness regularization module (Zhang et al., 2020a). Fazekas et al. (2022) used anatomical priors to enhance retinal layer segmentation in OCT (Fazekas et al., 2022). They adopted positional relationships, boundary continuity, and boundary shapes to construct a series of regularization terms that are added to the loss function.

Furthermore, prior knowledge has been employed in post-processing to refine segmentation results after the model's initial predictions, correcting potential errors and enhancing accuracy. For example, Liu et al. (2024b) proposed a boundary-repairing dual-path network for retinal layer segmentation in OCT images with PED. This network employs a feature interaction fusion module to enforce boundary shape constraints and a layer boundary repair module with contrastive loss to boost confidence in blurred boundary regions. Additionally, a novel bilateral threshold distance map is integrated with a topology correction module to ensure topology-preserving segmentation results.

Overall, the use of anatomical priors is well-suited to the layer segmentation task, as it reflects the inherent structure of the retinal and choroidal layers. However, if the priors are inaccurate or overly strict, they may lead to suboptimal performance by constraining the model's ability to adapt to variations in the data.

5.1.4. Uncertainty learning

Uncertainty learning holds significant value in clinical applications. By quantifying the confidence of model predictions, uncertainty learning can help clinicians assess whether a segmentation result is reliable or requires further review. There are two primary types of uncertainty: aleatoric and epistemic uncertainties (Gawlikowski et al., 2023; Zou et al., 2023). Aleatoric uncertainty originates from inherent data noise, including imaging artifacts (e.g., OCT speckle noise) and annotation inconsistencies (e.g., inter-observer variability in layer boundary delineation). This irreducible uncertainty persists even with expanded datasets. Epistemic uncertainty arises from model limitations, such as inadequate training data coverage or suboptimal network architectures. This uncertainty type can be mitigated through model refinement.

Bayesian Neural Networks (BNNs) probabilistically model network parameters to simultaneously capture both uncertainty types. These networks quantify the distribution of possible outputs for a given input through parametric modeling. For instance, Orlando et al. (2019) used Bayesian UNet for segmenting the photoreceptor layer in pathologies OCT images with AMD and DR (Orlando et al., 2019). The method implements Monte Carlo sampling with random dropout during inference to generate multiple segmentation instances, with averaged instances used as predictions and inter-sample standard deviations used as pixel-wise epistemic uncertainty maps.

Deterministic approaches primarily address aleatoric uncertainty by directly estimating data-dependent uncertainty from model outputs

³ While ET's classification as a neurodegenerative disorder remains debated (Thanvi et al., 2006; Louis, 2010), its inclusion here follows the original study's experimental design (Gende et al., 2023).

or intermediate features, leveraging computationally efficient single-pass inference to accommodate real-time clinical applications. Building on their earlier work SD-LayerNet (Fazekas et al., 2022), Fazekas et al. (2023) incorporated deterministic-based uncertainty estimation to guide a post-processing module for choroidal BM boundary localization (Fazekas et al., 2023). Specifically, this method in the first stage utilizes an attention UNet to output a probability density function of the BM location. In the second stage, it estimates the A-scan-wise uncertainty of the segmentation output and interpolates highly uncertain A-scans using thin plate spline functions. Compared to Bayesian methods that explicitly model parameter uncertainty, the deterministic method based on distance measurements has the advantage of simplicity and computational efficiency.

These methods identify potential diseases or segmentation errors by evaluating uncertainties at the pixel level. Based on the uncertainty maps, targeted post-processing and adjustments to prediction results can be performed, thereby enhancing the accuracy of segmentation.

5.1.5. Attention mechanisms

The attention mechanism is a technique that adaptively assigns weights to different regions of the feature maps, enabling neural networks to focus more on important areas relevant to the task (Vaswani et al., 2017). In early image segmentation work, spatial and channel attention mechanisms were widely used. However, with the development of Vision Transformer (ViT) technology (Dosovitskiy et al., 2020), recent studies have predominantly combined Transformer and CNN architectures to achieve a complementary balance between long-range dependencies and local feature focus (Tan et al., 2023; Cao et al., 2024; Yan et al., 2024).

One strategy is to embed the Transformer into the CNN construct layer by layer. For instance, Tan et al. (2023) developed a Tightly Combined Cross-convolution and Transformer with Boundary regression and feature Polarization (TCCT-BP) method for retinal layer segmentation (Tan et al., 2023). This method adopts a hybrid architecture of CNN and a lightweight Transformer to improve the perception of retinal layers. Besides, a feature grouping and sampling method alongside a corresponding polarization loss function were designed to enhance the differentiation of feature vectors across different retinal layers. While this tightly coupled strategy delivers richer feature representations and more refined feature fusion, it also incurs significant computational costs.

Another alternative strategy involves adding a Transformer module at the bottom of the UNet, achieving a balance between performance and computational efficiency (Cao et al., 2024; Yan et al., 2024). For example, Cao et al. (2024) proposed a self-attention CNN for retinal layer segmentation in OCT. This work introduces a self-attentive module in the vertical direction at the bottom of the U-shaped network and incorporates an attention mechanism in the skip connections and up-sampling stages (Cao et al., 2024). In this way, the model achieves complementarity between local details and context information. Yan et al. (2024) introduced a Choroidal Layer Analysis (CLA) network to jointly segment choroidal sub-layers and choroidal vessels in OCT images (Yan et al., 2024). CLA is a two-stage architecture, with an ambiguous boundary attention block in the first stage to identify the inconspicuous boundaries of choroidal sub-layers, and a Transformer block in the second stage to enhance choroidal vessel segmentation.

5.1.6. Dimension hybrid fusion

Some studies have explored innovative methods by integrating 1-D A-line (Tran et al., 2020), 2.5-D (sequences of B-scans) (Cahyo et al., 2021) and 3-D volumetric (Liu et al., 2021) information into the OCT segmentation tasks. These integrations provide the model with supplementary contextual information through dimensional hybrid and enhance the overall performance. For instance, Tran et al. (2020) restructured OCT segmentation tasks using natural language

processing methods (Tran et al., 2020). This method treats OCT B-scans as sequences of A-lines, transforming the image segmentation task into a Natural Language Processing (NLP) problem. In the context of 2.5-D data, Cahyo et al. (2021) proposed a method for choroidal segmentation in 3D OCT based on multi-task learning (Cahyo et al., 2021). The method considers OCT volumes as a sequence of B-scans and utilizes spatial information from adjacent cross-sectional slices to reconstruct the central slice, thereby learning additional contextual information. Subsequently, the extracted features are integrated with a UNet-based segmentation architecture. Additionally, Liu et al. (2021) introduced a hybrid 2D-3D CNN for obtaining continuous 3D retinal layer surfaces from OCT (Liu et al., 2021). Initially, each B-scan is used to extract 2D features. These 2D features are then input into two 3D decoders to generate alignment displacement fields and retinal segmentation results. The method was tested on the DUKE-WLOA public dataset, demonstrating superior layer segmentation accuracy and boundary continuity compared to contemporary 2D methods at that time.

5.2. Limited supervision learning methods

The lack of labeled data is a common difficulty in medical image segmentation scenarios, as annotations require the expertise of trained professionals and are often time-consuming to produce. Through the investigation, we observed that researchers primarily utilize semi-supervised learning to mitigate the scarcity of labeled data in OCT anatomical structures segmentation.

5.2.1. Semi-supervised learning

In semi-supervised learning, two sample sets are involved: labeled samples and unlabeled samples. The goal is to leverage the abundant unlabeled data to improve model performance, particularly in cases where acquiring labeled data is costly or time-consuming (Cheplygina et al., 2019).

A common semi-supervised approach is label propagation via self-training, which follows several key steps. First, a model is trained using the labeled data; next, the trained model generates pseudo-labels for the unlabeled samples. These pseudo-labeled samples, or a subset of them, are then added to the training set. This process can be repeated iteratively. For example, Zhang et al. (2023) proposed a semi-supervised method for choroidal segmentation tasks using prior knowledge and uncertainty for pseudo-label filtering (Zhang et al., 2023b). The method builds upon the standard process by enhancing the filtering of candidate labels, using confidence scores from the pre-trained model and their alignment with the true label distribution. High-confidence candidate labels are assigned different weights as pseudo-labels, improving the overall labeling accuracy. Building on their previous work (Liu et al., 2021), Liu (2024) introduced a label propagation strategy for extending sparse annotations in semi-supervised OCT segmentation (Liu et al., 2024c). This method employs a global coherence loss to ensure consistent layer surfaces both within and across B-scans based on gradient information. By exploiting the smoothness property across B-scans, the segmentation model can be trained semi-supervised with only a fraction of the B-scans annotated within each OCT volume.

Another widely adopted strategy involves leveraging unlabeled data to better assess data distributions. This approach typically incorporates perturbations either at the input or feature extraction stage and utilizes consistency regularization at the output stage. For instance, Yang et al. (2022) proposed a robust semi-supervised layer segmentation network to alleviate segmentation errors on abnormal retinas (Yang et al., 2022). This method extracts lesion features from labeled images with a disease-balanced distribution and supplements layer structural information from unlabeled images. For the unlabeled images, the cross-consistency training is applied on multiple decoders to enhance encoder representation. Similarly, Lu et al. (2022) introduced

Table 2

Full supervision learning methods for anatomical structures segmentation (OD: optic disc; MATD: mean absolute thickness difference; CE: counter error).

Reference	Year	Backbone	Disease	Dataset	Target	Metrics
Baselines						
Pekala et al. (2019b)	2019	FCN	DR	MIAMI	RL (4 sub-layers); CL (BM)	MAD
Gende et al. (2023)	2023	UNet	ND	private	RL (2 sub-layers)	Pre, Sens, DSC, MAE
Multi-Scale						
Gu et al. (2019)	2019	UNet	H.	private	RL (9 sub-layers), CL	MAE
Li et al. (2020b)	2020	UNet	H., AMD, DR	private	RL (10 sub-layers)	IoU, Sens, GPU average time
Prior Knowledge						
Ruan et al. (2019)	2019	FCN	H., CSC	private	RL (9 sub-layers)	MAD, MATD
He et al. (2019a)	2019	UNet	H., DME	HC-MS, DUKE-DME	RL (8 sub-layers), CL (BM)	MAD, DSC
Zhang et al. (2020a)	2019	UNet	H.	private	CL	DSC, IoU, TD
Li et al. (2021)	2021	GCN	DME; High Myopia, Peripapillary Atrophy, Cataract	DUKE-BOE, private	RL (9 sub-layers), OD	DSC, Acc
He et al. (2021)	2021	ResUNet	H., MS, DME	HC-MS, DUKE-DME	RL (8 sub-layers), CL (BM)	MAD, RMSE
Fazekas et al. (2022)	2022	UNet	AMD	private	RL	RMSE
Cansiz et al. (2023)	2023	UNet	H., MS	HC-MS	RL (HFL)	MAE
He et al. (2023a)	2023	UNet	–	private	RL (8 sub-layers)	TD, MAD, RMSE
Wang et al. (2024)	2024	UNet	High Myopia	private	CL	DSC, IoU, Acc, Pre, HD
Liu et al. (2024b)	2024	UNet	PED	OCTA-500, DUKE-BOE, private	RL (5 sub-layers)	DSC, MAD, HD95
Uncertainty						
Orlando et al. (2019)	2019	UNet	AMD, DME, RVO	private	RL (PCL)	AUC, DSC
Fazekas et al. (2023)	2023	Att-UNet	AMD	private	CL (BM)	MAE, RMSE
Attention						
Tan et al. (2023)	2023	CNN, Transformer	H., MS, DME, G., ect	DUKE-DME, HEG, HC-MS, GOALS	RL (2-8 sub-layers), CL, Fluid	DSC, IoU, HD, ED
Cao et al. (2024)	2024	UNet	H., DME; Cataract, High Myopia, Peripapillary Atrophy	DUKE-DME, private	RL, CL, OD	DSC, Acc
Yan et al. (2024)	2024	UNet	H., High Myopia	private	CL (3 sub-layers), Vessels	AUSDE, ATD
Dimension hybrid						
Tran et al. (2020)	2020	RCNN	H., AMD, DME	DUKE-WLOA, DUKE-BOE	RL (7 sub-layers)	DSC, MAD, MAD-LT
Cahyo et al. (2021)	2021	UNet	High Myopia	private	CL	IoU, DSC, Structural Similarity Index
Liu et al. (2021)	2021	UNet	H., AMD	DUKE-WLOA	RL (3 sub-layers)	MAD

a semi-supervised retinal segmentation framework that consists of a shared encoder and two different decoder branches (Lu et al., 2022). These branches utilize augmentation-consistent supervision modules and scale-transform-consistent supervision modules to enhance their generalization capability. Moreover, cross-consistency is computed as new supervision after feature perturbation between the two branches to further regularize their advantages. In 2023, Lu et al. advanced this approach with a Boundary-Enhanced Semi-supervised Network (BE-SemiNet) (Lu et al., 2023). This network incorporates an auxiliary distance map regression task, designing a Unilaterally Truncated Distance Map (UTDM) to address the class imbalance and enhance boundary detection. Task-level and data-level consistency regularization are then applied to unlabeled data for both pixel-wise segmentation and UTDM regression branches. This enriches the diversity of unsupervised

information and improves regularization effects. Moreover, pseudo supervision is used to ensure consistency in task prediction spaces and effectively expand the training set with labeled data.

5.3. Summary of retinal OCT anatomical structures segmentation

In OCT image anatomical structure segmentation, the majority of studies (77.8%) involved retinal layer segmentation, a smaller percentage (51.9%) have addressed choroidal layer segmentation, and only a few have considered vessels or other structures. As shown in Fig. 9, most of these studies rely on private datasets, with the HC-MS and DUKE datasets being the most frequently utilized public datasets.

As indicated in Tables 2 and 3, UNet, FCN, and their derivatives are the most widely adopted backbone architecture for anatomical structure segmentation. Low contrast and blurred boundaries are

Table 3
Limited supervision methods for anatomical structures segmentation.

Reference	Year	Backbone	Disease	Dataset	Target	Metrics
Semi-supervised						
Yang et al. (2022)	2022	UNet	H., AMD, CNV, DR, OA, CSC, etc	private	RL (5 sub-layers), CL (BM)	DSC, MAD
Lu et al. (2022)	2022	UNet	H., MC	HC-MS	RL (8 sub-layers)	DSC
Lu et al. (2023)	2023	UNet	H., MS; AMD	HC-MS, DUKE-WLOA	RL	DSC, MAD
Zhang et al. (2023b)	2023	UNet	H., G.,	private, GOALS	CL	DSC
Liu et al. (2024c)	2024	UNet	H., AMD, DME, MS	DUKE-DME, DUKE-WLOA, HC-MS	RL, CL (BM)	MAD, HD95, Normalized cross-correlation

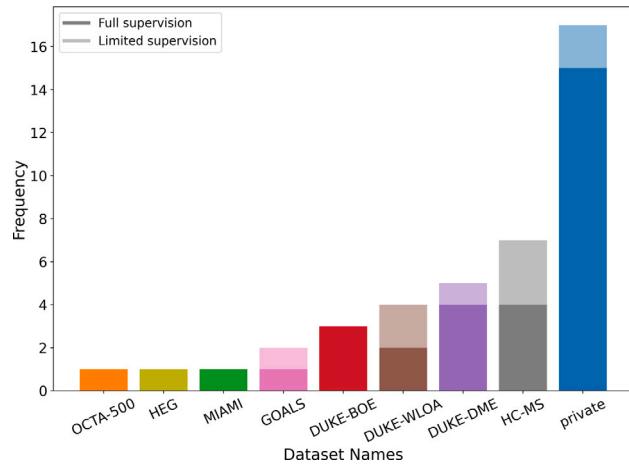


Fig. 9. OCT datasets used for anatomical structures segmentation.

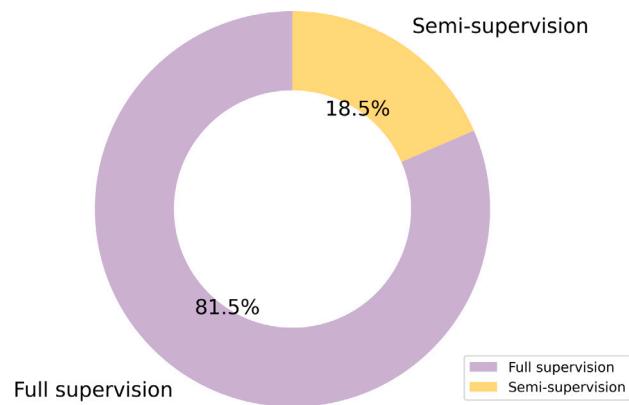


Fig. 10. Supervision strategy distribution for OCT anatomical structure segmentation demonstrates predominant use of full supervision (81.5%), with limited supervision approaches exclusively employing semi-supervised learning.

primary factors that impact segmentation performance. To address these challenges, many studies have integrated prior knowledge to ensure the correct topological structure of the predictions. Additionally, multi-scale feature fusion, uncertainty learning, attention mechanisms, and dimension hybrid fusion have been employed to enhance feature representation.

In terms of supervision methods, the majority (81.5%) of studies employ fully supervised learning, while a small portion (18.5%) utilize semi-supervised learning to mitigate the dependence on large-scale annotated data (see Fig. 10).

6. Pathological lesions segmentation

Accurate lesion segmentation in OCT images is critical for diagnosing and monitoring various ocular conditions. Traditionally, lesion segmentation in OCT images refers to identifying and localizing specific types of lesions. Recently, however, some studies have adopted anomaly detection methods to segment lesion regions without differentiating between specific lesion categories.

This section will provide a comprehensive review of both supervised and limited supervision methods employed in OCT lesion segmentation.

6.1. Full supervision learning methods

In addition to FCN and UNet, Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) and Transformers (Dosovitskiy et al., 2020) are also commonly employed in lesion segmentation tasks. GANs are particularly notable for their ability to enhance segmentation results by generating realistic outputs through a competitive process between generators and discriminators. They can be used to refine segmentation predictions, generate pseudo-labels for unlabeled data, and achieve data alignment across multiple domains, etc. Transformers, on the other hand, are adept at capturing global information and overcoming the limitations of CNNs in modeling long-range dependencies.

Similar to anatomical structures segmentation tasks, lesion segmentation also benefits from techniques such as multi-scale feature fusion, uncertainty learning, and attention mechanisms. Additionally, multi-task learning approaches, which combine lesion segmentation with tasks such as anatomy segmentation or lesion classification, are widely employed. These methods enhance model feature representation and improve segmentation accuracy by incorporating diverse sources of information and refining the focus on relevant features.

6.1.1. UNet and its variants

Due to the strong feature extraction capabilities of the U-shaped architecture, UNet and its variants excel in both anatomical structures segmentation and lesion segmentation tasks. For instance, Lu et al. (2019) proposed a segmentation and detection method for multi-class fluid lesions (Lu et al., 2019) based on the UNet structure. The work first employed the graph cut algorithm for intensity segmentation and retinal layer segmentation in OCT images. Subsequently, an UNet network was trained to identify and label fluid pixels. Finally, a random forest classifier was applied to the segmented fluid regions to detect and correct mislabeled fluid areas. Additionally, Lachinov et al. (2021) proposed a low-dimensional segmentation approach tailored to generate segmentation masks of lower dimensions than the input images (Lachinov et al., 2021). This method introduced a novel projective skip connection structure to establish stronger connections between the encoder and decoder. The approach demonstrated good performance in geographic atrophy (GA) lesions and retinal blood vessel segmentation tasks.

6.1.2. Multi-scale feature fusion

Retinal lesions exhibit significant variability in both scale and distribution. By employing multi-scale feature fusion, the model can integrate features at different scales, enabling the capture of lesion characteristics across various shapes and sizes, thereby enhancing segmentation performance.

Spatial Pyramid Pooling (SPP) is a commonly used technique for multi-scale feature extraction (Lazebnik et al., 2006). It applies pooling operations at multiple grid sizes to capture spatial information at different scales. The pooled features are then fused into a fixed-size representation, enabling the model to effectively handle objects of varying sizes and spatial locations. Xing et al. (2022) proposed an architecture for segmenting three types of fluid lesions in OCT images, improving segmentation performance by incorporating attention and shape prior information (Xing et al., 2022). In this work, attention gates and SPP modules were employed to enhance the network's ability to extract multi-scale features, addressing significant variations in the position, size, and shape of pathological fluid lesions in OCT images.

With the development of Atrous Spatial Pyramid Pooling (ASPP) (Chen et al., 2017), the technique extends SPP by utilizing atrous convolutions instead of pooling operations. Atrous convolutions enlarge the receptive field without reducing spatial resolution, allowing ASPP to produce multi-scale feature maps while preserving fine spatial details. So far, ASPP has been widely adopted in OCT lesion segmentation for its ability to handle lesions of varying sizes and shapes (Hu et al., 2019; Yang et al., 2020; Sappa et al., 2021). For instance, Hu et al. (2019) proposed a Deep Neural Network (DNNs) combined with ASPP for segmenting SRF and PED lesions (Hu et al., 2019). Initially, DNNs extract lesion-related features, which are then input into ASPP, utilizing multiple atrous convolutions with different receptive fields to extract features at various scales. Building upon ASPP, this work introduced a new module called stochastic ASPP (sASPP) to address co-adaptation issues among different scale convolution kernels. Similarly, Yang et al. (2020) and Sappa et al. (2021) integrated improved ASPP into neural networks to extract multi-scale features, achieving excellent performance in various fluid lesion segmentation tasks.

Recent studies have also utilized self-attention mechanisms to integrate global and multi-scale information. For instance, Yao et al. (2022) introduced GD-Net, a neural network designed to segment hyper-reflective foci (HRF) in OCT images with DR (Yao et al., 2022). GD-Net employs a global information fusion module to aggregate global semantic information. Rasti et al. (2023) presented RetiFluidNet, a multi-class retinal fluids segmentation method (Rasti et al., 2023). The RetiFluidNet integrates several modules, including a self-adaptive dual-attention module that captures relevant contextual information across spatial and channel domains, and self-adaptive attention-based skip connection paths that learn global and multi-semantic contextual features. Additionally, the method employs multi-scale local losses, such as dice loss components for texture-specific enhancement and connectivity-based loss components for edge-specific enhancement.

6.1.3. Uncertainty learning

Pathological retinal OCT images exhibit characteristics like imaging noise and indistinct lesion boundaries, introducing aleatoric uncertainty. Additionally, pathological datasets typically exhibit limited volume, significant variations in lesion features, and imbalanced distributions, which contribute to epistemic uncertainty during model training. These factors may reduce the model's confidence in handling ambiguous edges and rare lesions, leading to inaccurate segmentation results in these scenarios. To enhance the confidence and accuracy of prediction, several uncertainty learning-based approaches have been proposed, including Bayesian-based methods (Seeböck et al., 2020) and ensemble-based approaches (Seeböck et al., 2024; Liu et al., 2024a). These methods are typically implemented through a two-stage architecture: (1) an uncertainty quantification stage, where intermediate predictions are analyzed to systematically identify potential anatomical

structures and pathological lesions; and (2) a refinement stage, in which uncertainty maps guide precise segmentation of pathological lesions.

In 2020, Seeböck et al. proposed a Bayesian UNet method for anomaly detection and segmentation in limited supervision scenarios (Seeböck et al., 2020). This method uses weak labels of healthy anatomical structures during training and employs Monte Carlo dropout during inference to capture the model's epistemic uncertainty estimates. Finally, the appearance of uncertainty maps is transformed into smooth blob-shaped anomaly segmentation results. In Seeböck et al. (2024), Seeböck et al. (2024) advanced this work by using anomaly maps predicted by a pretrained Bayesian UNet as weak context supervision to train a multi-class lesion segmentation model.

For ensemble-based approaches, Liu et al. (2024) developed a Semantic Uncertainty Guided Cross-Transformer (SuGCTNet) for simultaneous segmentation of multi-class macular edema (Liu et al., 2024a). SuGCTNet follows a coarse-to-fine framework, where the coarse branch quantifies semantic uncertainty and generates initial segmentation, while the refine branch leverages both semantic uncertainty and multi-scale features to produce accurate lesion segmentation results.

Among these methods, Bayesian networks simultaneously address aleatory and epistemic uncertainties, making them highly suitable for handling noisy and complex data distributions in OCT lesion segmentation scenarios. Ensemble-based methods identify and focus on fuzzy and noisy regions by quantifying aleatory uncertainty, thereby enhancing boundary clarity and robustness.

6.1.4. Attention mechanisms

Researchers have also introduced attention mechanisms into lesion segmentation tasks in OCT images. Wang et al. (2022) addressed the challenges of large variations in size and shape and blurry boundaries of drusen lesions by proposing a novel Multi-scale Transformer Global Attention Network (MsTGANet) (Wang et al., 2022b). MsTGANet introduces a multi-scale Transformer non-local module embedded at the top of the encoder to capture long-range dependencies and multi-scale non-local features from different layers of the encoder. Additionally, they proposed a multi-semantic global channel and spatial joint attention module to enhance the model's ability to learn multi-semantic global contextual information. To mitigate the problem of limited labeled data, this work also introduced a semi-supervised version of MsTGANet (Semi-MsTGANet) that utilizes pseudo-labeling data augmentation strategies to enhance segmentation performance further.

6.1.5. Multi-task learning

Multi-Task Learning (MTL) is a paradigm that leverages the shared information among multiple related tasks to improve generalization performance across all tasks (Zhang and Yang, 2022). Based on whether the feature representations between tasks are consistent, MTL can be categorized into heterogeneous and homogeneous types.

In the context of OCT lesion segmentation, a common example of heterogeneous MTL is lesion segmentation-classification. Since the distribution and shape of lesions are closely tied to both disease diagnosis and lesion type, jointly learned models can share low-level spatial information and mid-level semantic information. This enables the model to capture complementary features across tasks, enhancing the performance of each task. For instance, Diao et al. (2023) considered a classification and segmentation method for AMD-related lesions (Diao et al., 2023). Initially, they introduced a complementary mask-guided CNN to classify OCT B-scans with drusen or CNV from normal ones, where the guiding mask is generated by an auxiliary segmentation task. Subsequently, a class activation map-guided UNet was adopted to segment drusen and CNV lesions.

For homogeneous MTL, typical applications include multi-class lesion segmentation (Shi et al., 2023) and joint lesion-anatomy segmentation (Mishra et al., 2020; Morelle et al., 2023; Shen et al., 2023a). In joint segmentation tasks, shared weights and feature representations across tasks help reduce overfitting, particularly in scenarios with

limited data. Moreover, related MTL tasks allow the model to capture richer contextual information, enhancing the accuracy of both tasks. For example, Shi et al. (2023) proposed a multi-class segmentation network to accurately segment retinal detachment and retinoschisis in OCT images of high myopia patients (Shi et al., 2023). Leveraging domain knowledge, they designed a three-class segmentation path and a five-class segmentation path. The outputs of these paths are integrated through additional decision fusion layers, resulting in improved segmentation performance on multiple lesions through a complementary approach. In OCT lesion-anatomy joint segmentation tasks, the interaction between lesion localization and anatomy segmentation further improves model performance, making this design widely adopted. For instance, Mishra et al. (2020) jointly segment 11 retinal sub-layers, drusen, and RPD lesions in SD-OCT images (Mishra et al., 2020). Morelle et al. (2023) first predicted retinal layer positions and then used them as a reference to segment drusen (Morelle et al., 2023). Shen et al. (2023) incorporated a graph attention encoder into the UNet to preserve the correct topological structure between retinals and CNV lesions in the segmentation results (Shen et al., 2023a).

6.2. Limited supervision learning methods

Acquiring a sufficient amount of annotated OCT pathology images for fully supervised training presents significant challenges. Rare or advanced lesions are particularly scarce, making it difficult to form balanced and comprehensive OCT lesion datasets. Moreover, the annotation process is time-consuming and requires highly trained experts. These factors contribute to the scarcity of labeled data in OCT lesion segmentation tasks. The lack of well-labeled data motivates approaches that extend beyond traditional supervised learning by utilizing available data more efficiently. These approaches include semi-supervised learning, which leverages a small amount of labeled data alongside a larger pool of unlabeled data; weakly supervised learning, which uses coarse or incomplete annotations; and unsupervised learning, which operates without any labeled data.

6.2.1. Semi-supervised learning

Like anatomical structure segmentation, lesion segmentation also utilizes semi-supervised learning to reduce reliance on annotated data. These approaches can be broadly categorized into label propagation, consistency learning, and contrastive learning strategies.

In the context of label propagation methods, Zhang et al. (2023) proposed a label propagation approach via self-supervised learning (Zhang et al., 2023a). This approach first uses self-supervised generative learning to pre-train a Transformer for learning OCT image representations. A segmentation model is then constructed using this Transformer as the encoder and a CNN as the decoder, trained on open-access data. The segmentation model generates pseudo-labels for the target dataset, using a greedy approximation for the k-center problem to select the most useful labels. By incorporating these pseudo-labels into model training, the approach achieves performance comparable to fully-supervised methods while using only about 10% of the actual annotations.

For consistency-based methods, Liu et al. (2021) proposed a semi-supervised learning-based framework to identify small fluids (Liu and Wang, 2021). The framework consists of a teacher network and a student network, each composed of an encoder and three decoders. These three decoders predict probability maps, contour maps, and distance maps, which are then fused to obtain the segmented results of lesions. Wang et al. (2023) addressed the complexity of macular holes (MH) and cystoid macular edema (CME) pathology features and the lack of pixel-level annotated data by proposing a novel self-guided optimization semi-supervised segmentation method (Semi-SGO) (Wang et al., 2023). The proposed SGO method utilizes a novel Dual-Decoder Dual-Task FCN (D3T-FCN) to enhance the model's ability to learn complex pathological features and alleviate the feature learning bias

that may occur in UNet architecture. Moreover, this work introduces knowledge distillation techniques and further designs a semi-supervised version of SGO, which leverages unlabeled data to further enhance segmentation accuracy.

Some methods integrate both self-supervised and consistency strategies. For instance, Shen et al. (2023) presented a Transformation-Consistent with Uncertainty and Self-deep supervision (TCUS) framework for semi-supervised OCT lesion segmentation (Shen et al., 2023b). The TCUS adopts an uncertainty-guided transformation-consistent strategy, allowing the student network to gradually learn from meaningful and reliable targets by utilizing the uncertainty information from the teacher network. Additionally, self-deep supervision is employed to capture multi-scale information, improving segmentation accuracy for lesions of various sizes and shapes.

Gomariz et al. (2022) introduced a semi-supervised contrastive learning framework for accurately segmenting fluid lesions in 3D OCT images (Gomariz et al., 2022). The framework employs an augmentation strategy that exploits the expected similarity between adjacent slices in the 3D volume, and introduces channel-wise aggregation for contrastive feature map projection.

6.2.2. Weakly supervised learning

Weakly supervised methods utilize varying levels of supervision, such as bounding boxes, scribbles, points, and image-level labels (Zhang et al., 2020b). Among these, image and point-level labels, which have the lowest annotation cost, are widely employed for lesion segmentation tasks.

Methods utilizing image-level labels often create pseudo-labels with spatial information through weak labeling techniques. The pseudo labels are typically generated using attention maps (Ma et al., 2020), anomaly maps (Tao et al., 2023; Wang et al., 2021; Yang et al., 2024), or other techniques (Xing et al., 2021). For attention maps, Ma et al. (2020) proposed a segmentation method that generates high-resolution class activation maps from image-level category labels (Ma et al., 2020). For the anomaly maps, Tao et al. (2023) developed a LAGAN network for fluid lesion segmentation using heatmap pseudo-labels (Tao et al., 2023). The approach first trains an image classifier using image-level labels to distinguish normal from abnormal images, then generates object heatmaps for each training sample via model visualization tools based on the trained classifier. These heatmaps are subsequently used as pseudo-labels to train a GAN-based image generator for fluid lesion segmentation. Similarly, Wang et al. (2021) utilized a CycleGAN for lesion segmentation in full-width OCT images (Wang et al., 2021). Their model reconstructed normal anatomical structures from abnormal images and identified lesions by analyzing the differences between input and reconstructed outputs. Also, Yang et al. (2024) construct anomaly-discriminative representations to incorporate anomalies into training (Yang et al., 2024). For other methods, Xing et al. (2021) detected NDR in central serous CSC patients using a two-stage framework (Xing et al., 2021). The first stage generates pixel-level pseudo-labels using CNNs and level sets, while the second stage applies an active contour loss function to train the segmentation model for accurate lesion detection.

Point-level annotations involve marking a single point on each object. He et al. (2022) proposed an Intra- and inter-Slice Contrastive Learning Network (ISCLNet) for OCT fluid segmentation (He et al., 2022). The ISCLNet employs intra-slice contrastive learning to differentiate between fluid and background patches, and enforces similarity by learning inter-slice relationships between features and predictive probability maps. This work improves the accuracy of OCT image fluid lesion segmentation in weakly supervised scenarios with point-level annotations.

6.2.3. Unsupervised learning

Unsupervised Learning seeks to learn representations and extract underlying semantic information without any label supervision, typically by leveraging task-agnostic unlabeled data (Chen et al., 2022). Common strategies in unsupervised learning include domain adaptation, clustering, and pretext tasks.

Unsupervised domain adaptation techniques have been extensively applied to this task, employing various alignment strategies to address domain shifts and improve performance across different domains (Song et al., 2023; He et al., 2023b; Liao et al., 2024). For instance, Song et al. (2023) introduced MDT-Net, a multi-domain transfer network with perceptual supervision (Song et al., 2023). This model adopts a unified encoder-decoder network alongside multiple domain-specific transfer modules, effectively disentangling feature representations of anatomical content from domain variance. He et al. (2023) developed the Structure-Guided Cross-Attention Network (SCAN) for cross-domain OCT fluid segmentation, utilizing retinal layer structures to aid domain alignment (He et al., 2023b). A cross-attention module is integrated to assess the correlation between layer-specific and fluid-specific features, enabling the model to focus on highly relevant regions during domain alignment. Additionally, Liao et al. (2024) proposed an Orthogonal Style Space Reparameterization (OSSR) method to achieve domain adaptation between OCT images obtained from different devices (Liao et al., 2024). The OSSR employs a Graph Adversarial Network to align the generated samples with those from the source domain. In addition, a graph semantic alignment module is employed to further align features with the same label based on semantic features in the graph semantic space.

Clustering automatically discovers latent structures in the data. Yuan et al. (2022) proposed an unsupervised segmentation framework combining prior knowledge and clustering concepts (Yuan et al., 2022). Observing that small lesions with intricate texture details are more apparent in smaller-scale images, while larger lesions are more detectable in larger-scale images, the authors developed a two-stage approach combining image-level clustering with pixel-level segmentation. Additionally, scale-invariant regularization and multi-scale class activation map fusion strategies are employed to capture multi-scale information effectively across both stages of the method.

Pretext tasks are designed to create pseudo-supervised learning objectives by generating pseudo-labels, allowing the model to learn meaningful representations from unlabeled data. Li et al. (2023) proposed a universal unsupervised anomaly detection framework (SSL-AnoVAE) (Li et al., 2023). The framework introduces an unsupervised anomaly staging method to better understand the severity of the retinal disease. Besides, a self-supervised learning module is trained with “free” labels derived from transformations of unannotated raw images.

6.3. Summary of retinal OCT pathological lesions segmentation

In OCT image lesion segmentation, most studies have focused on multiple types of fluid/edema lesions, with drusen and CNV being the next most commonly studied. Similar to anatomical segmentation tasks, the majority of these studies rely on private datasets, while the RETOUCH and Cell datasets are the most frequently utilized public datasets (Fig. 11).

In addition to UNet, researchers have also adopted GANs and Transformers as backbone architectures for lesion segmentation (see Tables 4 and 5). The diversity of lesions in terms of shape, size, and location makes it difficult to directly construct prior knowledge of morphology and distribution, as is often done with anatomical structure segmentation. To address this, many studies have incorporated multi-scale feature fusion to better capture features of lesions of varying sizes. Additionally, multi-task learning strategies, such as multi-lesion segmentation, lesion-anatomy joint segmentation, and lesion classification-segmentation, have been widely adopted. These approaches facilitate

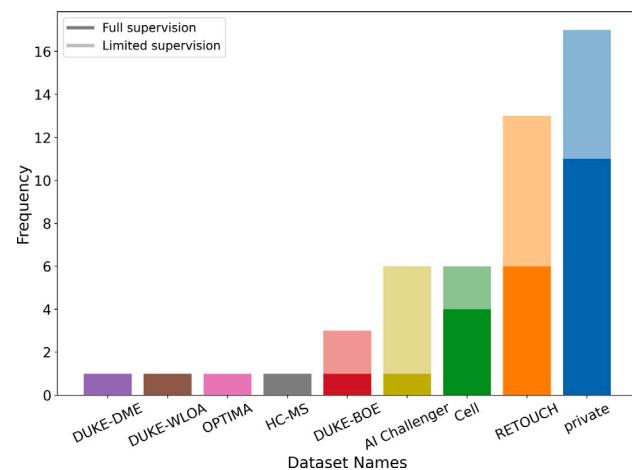


Fig. 11. OCT datasets used for lesion segmentation.

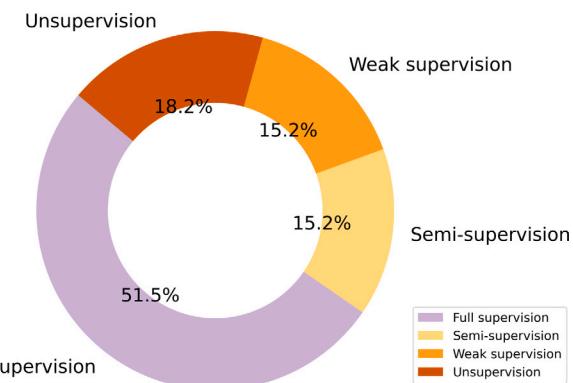


Fig. 12. Supervision strategy distribution for lesion segmentation reveals balanced adoption between full (51.5%) and limited supervision (48.5%), with the latter containing semi-supervised (31.2%), weakly-supervised (31.2%), and unsupervised (37.5%) methods.

better extraction of contextual features and enhance overall performance. Furthermore, uncertainty learning and attentional mechanisms were used to improve the attention to the abnormal regions.

Regarding supervision methods, the ratio of fully supervised to limited supervision approaches was nearly balanced at 17:16. Among the limited supervision approaches, the proportions of studies employing semi-supervised, weakly supervised, and unsupervised methods were comparable (see Fig. 12). Compared to OCT anatomical structures segmentation, lesion segmentation poses greater challenges in data acquisition and annotation, which may explain the broader adoption of limited supervision methods in this task.

7. Discussion

7.1. Challenges

This review comprehensively examines recent advancements in deep learning-based retinal OCT image segmentation. While notable progress has been achieved in anatomical structures and pathological lesions segmentation tasks, several critical limitations persist:

- The retinal OCT datasets face challenges of data scarcity and annotation difficulty. First, the acquisition of high-quality OCT data is challenging, particularly for rare diseases and complex cases, leading to insufficient data diversity and imbalanced category

Table 4

Full supervision learning methods for pathological lesions (NRD: Neurosensory Retinal Detachment; RD: Retinal Detachment; RS: Retinoschisis).

Reference	Year	Backbone	Dataset	Target	Metrics
Baselines					
Lu et al. (2019)	2019	UNet	RETOUCH, Cell	IRF, SRF, PED	DSC, AUC
Lachinov et al. (2021)	2021	UNet	private	GA, Retinal Vessels	DSC, HD95
Multi-Scale					
Hu et al. (2019)	2019	UNet	AI Challenger	Fluid	DSC
Yang et al. (2020)	2020	UNet	private	NRD	DSC
Sappa et al. (2021)	2021	UNet	private	IRF, SRF, PED	DSC, IoU
Yao et al. (2022)	2022	UNet	private	HRF	DSC, IoU, Recall, Pre
Xing et al. (2022)	2022	UNet	RETOUCH, private	Fluids	DSC, AVD
Rasti et al. (2023)	2023	UNet	DUKE-BOE, OPTIMA, RETOUCH	Fluids	DSC, Acc
Uncertainty					
Seeböck et al. (2020)	2020	Bayesian UNet	private	Anomaly	DSC, Recall, Pre
Seeböck et al. (2024)	2024	Bayesian UNet	private, RETOUCH, Cell	IRC, SHRF, SRF, HRF, Pseudodrusen, PED	DSC, Pre, Recall
Liu et al. (2024a)	2024	UNet, Transformer	RETOUCH	IRF, SRF, PED	DSC
Attention					
Wang et al. (2022b)	2022	GAN	Cell	Drusen	DSC, IoU, Pre
Multi-Task					
Mishra et al. (2020)	2020	FCN	private	RL (11 sub-layers), Drusen, RPD	ME, MAD
Diao et al. (2023)	2023	UNet	Cell, RETOUCH, DUKE-DME	Drusen, CNV	DSC, IoU, Sen, Spe, Acc
Morelle et al. (2023)	2023	Transformer	private, DUKE-WLOA	RL (RPE), CL (BM), drusen	DSC, MAE
Shen et al. (2023a)	2023	UNet	HC-MS, private	CNV	DSC, ACC, Pre, Sens, Spec
Shi et al. (2023)	2023	UNet	private	RD, RS	DSC, IoU, Sens, Spec

Table 5

Limited supervision methods for pathological lesions (SHRM: Subretinal Hyperreflective Material; AAD: Absolute Area Difference; CC: Correlation Coefficient, TPVF: True Positive Volume Fraction).

Reference	Year	Backbone	Dataset	Target	Metrics
Semi-supervised					
Liu and Wang (2021)	2021	UNet	RETOUCH	IRF, SRF, PED	DSC, AVD
Gomariz et al. (2022)	2022	UNet	private	IRF, SRF, PED, SHRM	DSC
Wang et al. (2023)	2023	FCN	private	MH, CME	IoU, DSC, Sens
Zhang et al. (2023a)	2023	Transformer, CNN	RETOUCH, private	SRF, PED, CL	DSC
Shen et al. (2023b)	2023	UNet	Cell, private	DME, Drusen; CNV, CSC	DSC, Acc, Sens, Spe
Weakly Supervised					
Ma et al. (2020)	2020	Att-UNet	private	GA	DSC, IoU, AAD, CC, AUC
Wang et al. (2021)	2021	Cycle-GAN	Cell, DUKE-BOE	DME, CNV	Pre, Sens, IoU, Spe, AUC
Xing et al. (2021)	2021	CNN	private	NRD	Pre, DSC, Pre, TPVF
He et al. (2022)	2022	UNet	AI Challenger, RETOUCH	IRF, SRF, PED	DSC
Tao et al. (2023)	2023	GAN	AI Challenger, RETOUCH	Edema	DSC, IoU, FNR, FPR
Unsupervised					
Yuan et al. (2022)	2022	GAN	AI Challenger	Edema	DSC, Pre, Recall, ASSD
Li et al. (2023)	2023	UNet	AI Challenger	Edema	AUC, Acc, Spe, Sen
Song et al. (2023)	2023	UNet	RETOUCH	Fluids	DSC
He et al. (2023b)	2023	GAN	RETOUCH	Fluids	DSC
Yang et al. (2024)	2024	GAN	AI Challenger, DUKE-BOE	Anomaly	DSC, IoU
Liao et al. (2024)	2024	GAN	RETOUCH	Fluids, Drusen	DSC

distribution. Second, annotating retinal OCT images requires professional ophthalmologists, which is time-consuming and labor-intensive. Consequently, existing datasets are limited in scale and exhibit inconsistent annotation quality.

- The current retinal OCT segmentation **algorithms** exhibit limitations in noise robustness, performance on complex pathologies, and model generalization. First, OCT-specific artifacts, such as speckle noise and motion artifacts, significantly degrade segmentation accuracy, particularly when resolving low-contrast choroidal structures and small pathological features like hyperreflective foci. Second, the inherent variability in lesion morphology across different diseases and progression stages, along with changes in retinal layer anatomy (as described in Section 2.4), fundamentally limit the effectiveness of automatic segmentation methods in handling these complex scenarios. Third, despite advancements in domain adaptation techniques, the clinical reliability of cross-device performance and generalization under varying imaging protocols remains inadequately validated, raising concerns about real-world deployment feasibility.
- The current model **evaluation** exhibits inadequacies in the consistency of experimental settings and the comprehensiveness of

evaluation metrics. First, divergent experimental paradigms, ranging from data sources to preprocessing techniques, hinder the comparative analysis of segmentation performance across studies. Second, although conventional metrics such as the DSC and MAD are widely used for performance reporting, the field lacks standardized assessments of clinically relevant metrics (e.g., error propagation in retinal thickness measurements) and systematic evaluations of cross-device generalizability and noise resilience.

7.2. Future directions

- Dataset** expansion and efficient utilization are essential for robust model development and evaluation. Constructing large-scale, multi-center datasets with standardized annotations is crucial, while synthetic data generation techniques such as GANs and diffusion models can enhance the volume and diversity of training data to mitigate data scarcity. Additionally, semi-supervised, weakly supervised, self-supervised, and transfer learning methods facilitate the effective extraction of valuable features from limited or incomplete annotations.

- The integration of advanced architectures, model generalization, and multimodal data fusion is likely to constitute critical directions for future **algorithm** development. First, by leveraging advanced visual transformers and pre-trained foundation models based on multimodal ophthalmic datasets, novel algorithms can improve the segmentation in complex anatomical structures and pathological lesions. Second, to enhance model robustness, potential approaches may include integrating data augmentation, domain generalization, and federated learning, thereby ensuring consistent performance across imaging devices and patient populations. Third, integrating multimodal imaging (e.g., fundus photography, and OCT angiography) enhances segmentation accuracy by leveraging complementary characteristics.
- The establishment of standardized **evaluation** frameworks combined with incentivized algorithm open-sourcing will advance this field through two synergistic mechanisms. First, establish a consistent evaluation framework with unified experimental settings and comprehensive metrics. Second, promote community-wide sharing of open-source algorithms and pre-trained models. This integrated approach facilitates objective performance benchmarking and ensures methodological reproducibility across studies.

8. Conclusion

This review provides a comprehensive overview of recent advances in deep learning-based retinal OCT image segmentation, with systematic analysis of state-of-the-art methods and their technical evolution from 2019 to 2024. It also provides a practical framework to assist cross-disciplinary researchers, such as those in medical imaging and biomedical engineering, in bridging the gap between retinal imaging and algorithm design. Additionally, it synthesizes information on publicly available datasets and evaluation metrics, providing crucial resources for researchers to select appropriate datasets and understand performance standards. This review further discusses the challenges and potential future directions in dataset construction, algorithmic performance, and evaluation standardization within the current studies.

CRediT authorship contribution statement

Huihong Zhang: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Bing Yang:** Writing – review & editing, Validation. **Sanqian Li:** Writing – review & editing. **Xiaoqing Zhang:** Writing – review & editing. **Xiaoling Li:** Writing – review & editing. **Tianhang Liu:** Writing – review & editing. **Risa Higashita:** Supervision, Project administration. **Jiang Liu:** Supervision, 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.

Data availability

No data was used for the research described in the article.

References

- Alizadeh Eghetedar, R., 2022. An update on choroidal layer segmentation methods in optical coherence tomography images: a review. *J. Biomed. Phys. Eng.* <http://dx.doi.org/10.31661/jbpe.v0i0.1234>.
- Aumann, S., Donner, S., Fischer, J., Müller, F., 2019. Optical coherence tomography (OCT): principle and technical realization. *High Resolut. Imaging Microsc. Ophthalmology: New Front. Biomed. Opt.* 59–85.
- Bogunović, H., Venhuizen, F., Klimscha, S., Apostopoulos, S., Bab-Hadiashar, A., Bagci, U., Beg, M.F., Bekalo, L., Chen, Q., Ciller, C., et al., 2019. RETOUCH: The retinal OCT fluid detection and segmentation benchmark and challenge. *IEEE Trans. Med. Imaging* 38 (8), 1858–1874.
- Cahyo, D.A., Yow, A.P., Saw, S.-M., Ang, M., Girard, M., Schmetterer, L., Wong, D., 2021. Multi-task learning approach for volumetric segmentation and reconstruction in 3D OCT images. *Biomed. Opt. Express* 12 (12), 7348–7360.
- Cansiz, S., Kesim, C., Bektas, S.N., Kulali, Z., Hasaneisoglu, M., Gunduz-Demir, C., 2023. FourierNet: Shape-preserving network for henle's fiber layer segmentation in optical coherence tomography images. *IEEE J. Biomed. Heal. Informatics* 27 (2), 1036–1047. <http://dx.doi.org/10.1109/JBHI.2022.3225425>.
- Cao, G., Wu, Y., Peng, Z., Zhou, Z., Dai, C., 2024. Self-attention CNN for retinal layer segmentation in OCT. *Biomed. Opt. Express* 15 (3), 1605–1617.
- Chen, Y., Mancini, M., Zhu, X., Akata, Z., 2022. Semi-supervised and unsupervised deep visual learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 46 (3), 1327–1347.
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (4), 834–848.
- Cheplygina, V., de Bruijne, M., Pluim, J.P., 2019. Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Med. Image Anal.* 54, 280–296. <http://dx.doi.org/10.1016/j.media.2019.03.009>, URL <https://www.sciencedirect.com/science/article/pii/S1361841518307588>.
- Chiu, S.J., Allingham, M.J., Mettu, P.S., Cousins, S.W., Izatt, J.A., Farsiu, S., 2015. Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. *Biomed. Opt. Express* 6 (4), 1172–1194.
- Chiu, S.J., Izatt, J.A., O'Connell, R.V., Winter, K.P., Toth, C.A., Farsiu, S., 2012. Validated automatic segmentation of AMD pathology including drusen and geographic atrophy in SD-OCT images. *Investig. Ophthalmol. Vis. Sci.* 53 (1), 53–61.
- Diao, S., Su, J., Yang, C., Zhu, W., Xiang, D., Chen, X., Peng, Q., Shi, F., 2023. Classification and segmentation of OCT images for age-related macular degeneration based on dual guidance networks. *Biomed. Signal Process. Control.* 84, 104810.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *ArXiv: Computer Vision and Pattern Recognition*, ArXiv: Computer Vision and Pattern Recognition.
- Fang, H., Li, F., Fu, H., Wu, J., Zhang, X., Xu, Y., 2022. Dataset and evaluation algorithm design for goals challenge. In: International Workshop on Ophthalmic Medical Image Analysis. Springer, pp. 135–142.
- Farsiu, S., Chiu, S.J., O'Connell, R.V., Folgar, F.A., Yuan, E., Izatt, J.A., Toth, C.A., Age-Related Eye Disease Study 2 Ancillary Spectral Domain Optical Coherence Tomography Study Group, et al., 2014. Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography. *Ophthalmology* 121 (1), 162–172.
- Fazekas, B., Aresta, G., Lachinov, D., Riedl, S., Mai, J., Schmidt-Erfurth, U., Bogunović, H., 2022. SD-LayerNet: Semi-supervised retinal layer segmentation in OCT using disentangled representation with anatomical priors. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*. Springer Nature Switzerland, Cham, pp. 320–329. http://dx.doi.org/10.1007/978-3-031-16452-1_31.
- Fazekas, B., Lachinov, D., Aresta, G., Mai, J., Schmidt-Erfurth, U., Bogunović, H., 2023. Segmentation of Bruch's membrane in retinal OCT with AMD using anatomical priors and uncertainty quantification. *IEEE J. Biomed. Heal. Informatics* 27 (1), 41–52. <http://dx.doi.org/10.1109/JBHI.2022.3217962>.
- Ferrara, M., Lugano, G., Sandinha, M.T., Kearns, V.R., Geraghty, B., Steel, D.H., 2021. Biomechanical properties of retina and choroid: a comprehensive review of techniques and translational relevance. *Eye* 35 (7), 1818–1832.
- Gawlikowski, J., Tassi, C.R.N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R., et al., 2023. A survey of uncertainty in deep neural networks. *Artif. Intell. Rev.* 56 (Suppl 1), 1513–1589.
- Gende, M., Mallen, V., de Moura, J., Cordón, B., García-Martin, E., Sánchez, C.I., Novo, J., Ortega, M., 2023. Automatic segmentation of retinal layers in multiple neurodegenerative disorder scenarios. *IEEE J. Biomed. Heal. Informatics* 27 (11), 5483–5494. <http://dx.doi.org/10.1109/JBHI.2023.3313392>.
- Gholami, P., Roy, P., Parthasarathy, M.K., Lakshminarayanan, V., 2020. OCTID: Optical coherence tomography image database. *Comput. Electr. Eng.* 81, 106532.
- Global Growth Insights, 2024. Optical Coherence Tomography for Ophthalmology Market Analysis. (GGI107498), Global Growth Insights, URL <https://www.globalgrowthinsights.com/market-reports/optical-coherence-tomography-for-ophthalmology-market-107498>.

- Gomariz, A., Lu, H., Li, Y.Y., Albrecht, T., Maunz, A., Benmansour, F., Valcarcel, A.M., Luu, J., Ferrara, D., Goksel, O., 2022. Unsupervised domain adaptation with contrastive learning for OCT segmentation. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI. Springer, pp. 351–361.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014. December 8–13 2014, Montreal, Quebec, Canada, pp. 2672–2680. http://dx.doi.org/10.3156/jsoft.29.5_177_2.
- Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J., 2019. CE-net: Context encoder network for 2D medical image segmentation. IEEE Trans. Med. Imaging 38 (10), 2281–2292. <http://dx.doi.org/10.1109/TMI.2019.2903562>.
- Guangli, W., Wenbin, W., 2009. *Atlas of Fundus Imaging with Optical Coherence Tomography*. Beijing Science and Technology Publishing House, p. 351.
- Hassan, T., Akram, M.U., Masood, M.F., Yasin, U., 2018. BIOMISA retinal image database for macular and ocular syndromes. In: Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings 15. Springer, pp. 695–705.
- He, Y., Carass, A., Liu, Y., Calabresi, P.A., Saidha, S., Prince, J.L., 2023a. Longitudinal deep network for consistent OCT layer segmentation. Biomed. Opt. Express 14 (5), 1874–1893.
- He, Y., Carass, A., Liu, Y., Jedynak, B.M., Solomon, S.D., Saidha, S., Calabresi, P.A., Prince, J.L., 2019a. Fully convolutional boundary regression for retina OCT segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Springer International Publishing, Cham, pp. 120–128. http://dx.doi.org/10.1007/978-3-030-32239-7_14.
- He, Y., Carass, A., Liu, Y., Jedynak, B.M., Solomon, S.D., Saidha, S., Calabresi, P.A., Prince, J.L., 2021. Structured layer surface segmentation for retina OCT using fully convolutional regression networks. Med. Image Anal. 68, 101856. <http://dx.doi.org/10.1016/j.media.2020.101856>, URL <https://www.sciencedirect.com/science/article/pii/S1361841520302206>.
- He, Y., Carass, A., Solomon, S.D., Saidha, S., Calabresi, P.A., Prince, J.L., 2019b. Retinal layer parcellation of optical coherence tomography images: Data resource for multiple sclerosis and healthy controls. Data Brief 22, 601–604.
- He, X., Fang, L., Tan, M., Chen, X., 2022. Intra-and inter-slice contrastive learning for point supervised oct fluid segmentation. IEEE Trans. Image Process. 31, 1870–1881.
- He, X., Zhong, Z., Fang, L., He, M., Sebe, N., 2023b. Structure-guided cross-attention network for cross-domain OCT fluid segmentation. IEEE Trans. Image Process. 32, 309–320. <http://dx.doi.org/10.1109/TIP.2022.3228163>.
- Hu, J., Chen, Y., Yi, Z., 2019. Automated segmentation of macular edema in OCT using deep neural networks. Med. Image Anal. 55, 216–227. <http://dx.doi.org/10.1016/j.media.2019.05.002>, URL <https://www.sciencedirect.com/science/article/pii/S1361841519300386>.
- Huang, D., Swanson, E.A., Lin, C.P., Schuman, J.S., Stinson, W.G., Chang, W., Hee, M.R., Flotte, T., Gregory, K., Puliafito, C.A., et al., 1991. Optical coherence tomography. Science 254 (5035), 1178–1181.
- Kafieh, R., Rabbani, H., Kermani, S., 2013. A review of algorithms for segmentation of optical coherence tomography from retina. J. Med. Signals & Sensors 3 (1), 45–60.
- Kashefpur, M., Kafieh, R., Jorjandi, S., Golmohammadi, H., Khodabande, Z., Abbasi, M., Teifuri, N., Fakharzadeh, A.A., Kashefpoor, M., Rabbani, H., 2017. Isfahan MIS dataset. J. Med. Signals & Sensors 7 (1), 43–48.
- Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., et al., 2018. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172 (5), 1122–1131.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. Commun. ACM 60 (6), 84–90. <http://dx.doi.org/10.1145/3065386>.
- Lachinov, D., Seeböck, P., Mai, J., Goldbach, F., Schmidt-Erfurth, U., Bogunovic, H., 2021. Projective skip-connections for segmentation along a subset of dimensions in retinal OCT. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Springer International Publishing, Cham, pp. 431–441. http://dx.doi.org/10.1007/978-3-030-87193-2_41.
- Lazebnik, S., Schmid, C., Ponce, J., 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2. CVPR'06, <http://dx.doi.org/10.1109/cvpr.2006.68>.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.
- Li, M., Chen, Y., Ji, Z., Xie, K., Yuan, S., Chen, Q., Li, S., 2020a. Image projection network: 3D to 2D image segmentation in OCTA images. IEEE Trans. Med. Imaging 39 (11), 3343–3354.
- Li, J., Jin, P., Zhu, J., Zou, H., Xu, X., Tang, M., Zhou, M., Gan, Y., He, J., Ling, Y., et al., 2021. Multi-scale GCN-assisted two-stage network for joint segmentation of retinal layers and discs in peripapillary OCT images. Biomed. Opt. Express 12 (4), 2204–2220.
- Li, Y., Lao, Q., Kang, Q., Jiang, Z., Du, S., Zhang, S., Li, K., 2023. Self-supervised anomaly detection, staging and segmentation for retinal images. Med. Image Anal. 87, 102805.
- Li, Q., Li, S., He, Z., Guan, H., Chen, R., Xu, Y., Wang, T., Qi, S., Mei, J., Wang, W., 2020b. DeepRetina: Layer Segmentation of Retina in OCT Images Using Deep Learning. Transl. Vis. Sci. Technol. 9 (2), 61. <http://dx.doi.org/10.1167/tvst.9.2.61>, arXiv:https://arvojournals.org/arvo/content_public/journal/tvst/938366/i2164-2591-9-2-61_1607511436.37786.pdf.
- Liao, S., Peng, T., Chen, H., Lin, T., Zhu, W., Shi, F., Chen, X., Xiang, D., 2024. Dual-spatial domain generalization for fundus lesion segmentation in unseen manufacturer's OCT images. IEEE Trans. Biomed. Eng..
- Liu, H., Gao, W., Yang, L., Wu, D., Zhao, D., Chen, K., Liu, J., Ye, Y., Xu, R.X., Sun, M., 2024a. Semantic uncertainty guided cross-transformer for enhanced macular edema segmentation in OCT images. Comput. Biol. Med. 174, 108458.
- Liu, X., Li, X., Zhang, Y., Wang, M., Yao, J., Tang, J., 2024b. Boundary-repairing dual-path network for retinal layer segmentation in OCT image with pigment epithelial detachment. J. Imaging Informatics Med. 1–30.
- Liu, X., Wang, S., 2021. Uncertainty-aware semi-supervised framework for automatic segmentation of macular edema in OCT images. In: 2021 IEEE 18th International Symposium on Biomedical Imaging. ISBI, IEEE, pp. 1453–1456.
- Liu, H., Wei, D., Lu, D., Li, Y., Ma, K., Wang, L., Zheng, Y., 2021. Simultaneous alignment and surface regression using hybrid 2D-3D networks for 3D coherent layer segmentation of retina OCT images. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Springer International Publishing, Cham, pp. 108–118. http://dx.doi.org/10.1007/978-3-030-87237-3_11.
- Liu, H., Wei, D., Lu, D., Tang, X., Wang, L., Zheng, Y., 2024c. Simultaneous alignment and surface regression using hybrid 2D-3D networks for 3D coherent layer segmentation of retinal OCT images with full and sparse annotations. Med. Image Anal. 91, 103019.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, <http://dx.doi.org/10.1109/cvpr.2015.7298965>.
- Louis, E.D., 2010. Essential tremor as a neuropsychiatric disorder. J. Neurol. Sci. 289 (1–2), 144–148.
- Lu, D., Heisler, M., Lee, S., Ding, G.W., Navajas, E., Sarunic, M.V., Beg, M.F., 2019. Deep-learning based multiclass retinal fluid segmentation and detection in optical coherence tomography images using a fully convolutional neural network. Med. Image Anal. 54, 100–110. <http://dx.doi.org/10.1016/j.media.2019.02.011>, URL <https://www.sciencedirect.com/science/article/pii/S1361841519300167>.
- Lu, Y., Shen, Y., Xing, X., Meng, M.Q.-H., 2022. Multiple consistency supervision based semi-supervised OCT segmentation using very limited annotations. In: 2022 International Conference on Robotics and Automation. ICRA, pp. 8483–8489. <http://dx.doi.org/10.1109/ICRA46639.2022.9812447>.
- Lu, Y., Shen, Y., Xing, X., Ye, C., Meng, M.Q.-H., 2023. Boundary-enhanced semi-supervised retinal layer segmentation in optical coherence tomography images using fewer labels. Comput. Med. Imaging Graph. 105, 102199.
- Ma, X., Ji, Z., Niu, S., Leng, T., Rubin, D., Chen, Q., 2020. MS-CAM: Multi-scale class activation maps for weakly-supervised segmentation of geographic atrophy lesions in SD-OCT images. IEEE J. Biomed. Heal. Informatics 24, 3443–3455. <http://dx.doi.org/10.1109/JBHI.2020.2999588>, URL <https://api.semanticscholar.org/CorpusID:220979208>.
- Maetschke, S., Antony, B., Ishikawa, H., Wollstein, G., Schuman, J., Garnavi, R., 2019. A feature agnostic approach for glaucoma detection in OCT volumes. PLoS One 14 (7), e0219126.
- Mahmudi, T., Kafieh, R., Rabbani, H., Akhlagi, M., et al., 2014. Comparison of macular OCTs in right and left eyes of normal people. In: Medical Imaging 2014: Biomedical Applications in Molecular, Structural, and Functional Imaging, vol. 9038, SPIE, pp. 472–477.
- Melinšćak, M., Radmilović, M., Vatavuk, Z., Lončarić, S., 2021. Annotated retinal optical coherence tomography images (AROI) database for joint retinal layer and fluid segmentation. Automatika: Časopis Za Autom. Mjer. Elektron. Računarstvo I Komun. 62 (3–4), 375–385.
- Mishra, Z., Ganegoda, A., Selicha, J., Wang, Z., Sadda, S.R., Hu, Z.J., 2020. Automated retinal layer segmentation using graph-based algorithm incorporating deep-learning-derived information. Sci. Rep. 10 (9541), <http://dx.doi.org/10.1038/s41598-020-66355-5>, URL <https://api.semanticscholar.org/CorpusID:219589284>.
- Morelle, O., Wintergerst, M.W., Finger, R.P., Schultz, T., 2023. Accurate drusen segmentation in optical coherence tomography via order-constrained regression of retinal layer heights. Sci. Rep. 13 (1), 8162.
- Müller, D., Soto-Rey, I., Kramer, F., 2022. Towards a guideline for evaluation metrics in medical image segmentation. BMC Res. Notes 15 (1), 210.
- Orlando, J.I., Seeböck, P., Bogunović, H., Klimscha, S., Grechenig, C., Waldstein, S., Gerendas, B.S., Schmidt-Erfurth, U., 2019. U2-net: A Bayesian U-net model with epistemic uncertainty feedback for photoreceptor layer segmentation in pathological OCT scans. In: 2019 IEEE 16th International Symposium on Biomedical Imaging. ISBI 2019, pp. 1441–1445. <http://dx.doi.org/10.1109/ISBI.2019.8759581>.
- Pekala, M., Joshi, N., Liu, T.Y.A., Bressler, N.M., Cabrera DeBuc, D., Burlina, P., 2019a. OCT segmentation via deep learning: A review of recent work. In: Lecture Notes in Computer Science, Computer Vision – ACCV 2018 Workshops. pp. 316–322. http://dx.doi.org/10.1007/978-3-030-21074-8_27.

- Pekala, M., Joshi, N., Liu, T.A., Bressler, N., DeBuc, D.C., Burlina, P., 2019b. Deep learning based retinal OCT segmentation. *Comput. Biol. Med.* 114, 103445. <http://dx.doi.org/10.1016/j.combiomed.2019.103445>, URL <https://www.sciencedirect.com/science/article/pii/S0010482519303221>.
- Raja, H., Akram, M.U., Shaukat, A., Khan, S.A., Alghamdi, N., Khawaja, S.G., Nazir, N., 2020. Extraction of retinal layers through convolution neural network (CNN) in an OCT image for glaucoma diagnosis. *J. Digit. Imaging* 33, 1428–1442.
- Ran, A.R., Tham, C.C., Chan, P.P., Cheng, C.-Y., Tham, Y.-C., Rim, T.H., Cheung, C.Y., 2021. Deep learning in glaucoma with optical coherence tomography: a review. *Eye* 35 (1), 188–201.
- Rasti, R., Biglari, A., Rezapourian, M., Yang, Z., Farsiu, S., 2023. RetiFluidNet: A self-adaptive and multi-attention deep convolutional network for retinal OCT fluid segmentation. *IEEE Trans. Med. Imaging* 42 (5), 1413–1423. <http://dx.doi.org/10.1109/TMI.2022.3228285>.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: *Lecture Notes in Computer Science, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. pp. 234–241. http://dx.doi.org/10.1007/978-3-319-24574-4_28.
- Ruan, Y., Xue, J., Li, T., Liu, D., Lu, H., Chen, M., Liu, T., Niu, S., Li, D., 2019. Multi-phase level set algorithm based on fully convolutional networks (FCN-MLS) for retinal layer segmentation in SD-OCT images with central serous chorioretinopathy (CSC). *Biomed. Opt. Express* 10 (8), 3987–4002. <http://dx.doi.org/10.1364/BOE.10.003987>, URL <https://opg.optica.org/boe/abstract.cfm?URI=boe-10-8-3987>.
- Sappa, L.B., Okuwobi, I.P., Li, M., Zhang, Y., Xie, S., Yuan, S., Chen, Q., 2021. RetiFluidNet: Retinal fluid segmentation for SD-OCT images using convolutional neural network. *J. Digit. Imaging* 34, 691–704. <http://dx.doi.org/10.1007/s10278-021-00459-w>, URL <https://api.semanticscholar.org/CorpusID:235307208>.
- Schuman, J.S., Fujimoto, J.G., Duke, J., Ishikawa, H., 2024. *Optical Coherence Tomography of Ocular Diseases*. CRC Press.
- Seeböck, P., Orlando, J.I., Michl, M., Mai, J., Schmidt-Erfurth, U., Bogunović, H., 2024. Anomaly guided segmentation: Introducing semantic context for lesion segmentation in retinal OCT using weak context supervision from anomaly detection. *Med. Image Anal.* 93, 103104.
- Seeböck, P., Orlando, J.I., Schlegl, T., Waldstein, S.M., Bogunović, H., Klimscha, S., Langs, G., Schmidt-Erfurth, U., 2020. Exploiting epistemic uncertainty of anatomy segmentation for anomaly detection in retinal OCT. *IEEE Trans. Med. Imaging* 39 (1), 87–98. <http://dx.doi.org/10.1109/TMI.2019.2919951>.
- Shen, Y., Li, J., Zhu, W., Yu, K., Wang, M., Peng, Y., Zhou, Y., Guan, L., Chen, X., 2023a. Graph attention U-net for retinal layer surface detection and choroid neovascularization segmentation in OCT images. *IEEE Trans. Med. Imaging* 42 (11), 3140–3154. <http://dx.doi.org/10.1109/TMI.2023.3240757>.
- Shen, H., Yang, Q., Chen, Z., Ye, Z., Dai, P., Duan, X., 2023b. Semi-supervised OCT lesion segmentation via transformation-consistent with uncertainty and self-deep supervision. *Biomed. Opt. Express* 14 (7), 3828–3840.
- Shi, F., Yang, C., Jiang, Q., Zhu, W., Xu, X., Chen, X., Fan, Y., 2023. Segmentation of retinal detachment and retinoschisis in OCT images based on complementary multi-class segmentation networks. *Phys. Med. Biol.* 68 (11), 115019.
- Siddique, N., Paheging, S., Elkin, C.P., Devabhaktuni, V., 2021. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* 9, 82031–82057.
- Song, W., Fotedar, G., Tajbakhsh, N., Zhou, Z., He, L., Ding, X., 2023. MDT-net: Multi-domain transfer by perceptual supervision for unpaired images in OCT scan. In: *2023 IEEE 20th International Symposium on Biomedical Imaging*. ISBI, pp. 1–5. <http://dx.doi.org/10.1109/ISBI53787.2023.10230361>.
- Srinivasan, P.P., Kim, L.A., Mettu, P.S., Cousins, S.W., Comer, G.M., Izatt, J.A., Farsiu, S., 2014. Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomed. Opt. Express* 5 (10), 3568–3577.
- Staurenghi, G., Sadda, S., Chakravarthy, U., Spaide, R.F., et al., 2014. Proposed lexicon for anatomic landmarks in normal posterior segment spectral-domain optical coherence tomography: the IN• OCT consensus. *Ophthalmology* 121 (8), 1572–1578.
- Tan, Y., Shen, W.-D., Wu, M.-Y., Liu, G.-N., Zhao, S.-X., Chen, Y., Yang, K.-F., Li, Y.-J., 2023. Retinal layer segmentation in OCT images with boundary regression and feature polarization. *IEEE Trans. Med. Imaging*.
- Tao, Y., Ma, X., Zhang, Y., Huang, K., Ji, Z., Fan, W., Yuan, S., Chen, Q., 2023. LAGAN: Lesion-aware generative adversarial networks for edema area segmentation in SD-OCT images. *IEEE J. Biomed. Heal. Informatics* 27 (5), 2432–2443.
- Thanvi, B., Lo, N., Robinson, T., 2006. Essential tremor—the most common movement disorder in older people. *Age Ageing* 35 (4), 344–349.
- Tian, J., Varga, B., Somfai, G.M., Lee, W.-H., Smiddy, W.E., Cabrera DeBuc, D., 2015. Real-time automatic segmentation of optical coherence tomography volume data of the macular region. *PloS One* 10 (8), e0133908.
- Tian, J., Varga, B., Tatrai, E., Fanni, P., Somfai, G.M., Smiddy, W.E., DeBuc, D.C., 2016. Performance evaluation of automated segmentation software on optical coherence tomography volume data. *J. Biophotonics* 9 (5), 478–489.
- Tran, A., Weiss, J., Albarqouni, S., Faghi Roohi, S., Navab, N., 2020. Retinal layer segmentation reformulated as OCT language processing. In: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L. (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Springer International Publishing, Cham, pp. 694–703. http://dx.doi.org/10.1007/978-3-030-59722-1_67.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30.
- Viedma, I., Alonso-Caneiro, D., Read, S., Collins, M., 2022. Deep learning in retinal optical coherence tomography (OCT): A comprehensive survey. *Neurocomputing* 507, 247–264. <http://dx.doi.org/10.1016/j.neucom.2022.08.021>.
- Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., Nandi, A.K., 2022a. Medical image segmentation using deep learning: A survey. *IET Image Process.* 16 (5), 1243–1267.
- Wang, J., Li, W., Chen, Y., Fang, W., Kong, W., He, Y., Shi, G., 2021. Weakly supervised anomaly segmentation in retinal OCT images using an adversarial learning approach. *Biomed. Opt. Express* 12 8, 4713–4729. <http://dx.doi.org/10.1364/boe.426803>, URL <https://api.semanticscholar.org/CorpusID:237492393>.
- Wang, M., Lin, T., Peng, Y., Zhu, W., Zhou, Y., Shi, F., Yu, K., Meng, Q., Liu, Y., Chen, Z., et al., 2023. Self-guided optimization semi-supervised method for joint segmentation of macular hole and cystoid macular edema in retinal OCT images. *IEEE Trans. Biomed. Eng.* 70 (7), 2013–2024.
- Wang, Y., Yang, Z., Liu, X., Li, Z., Wu, C., Wang, Y., Jin, K., Chen, D., Jia, G., Chen, X., et al., 2024. PGKD-Net: Prior-guided and knowledge diffusive network for choroid segmentation. *Artif. Intell. Med.* 150, 102837.
- Wang, M., Zhu, W., Shi, F., Su, J., Chen, H., Yu, K., Zhou, Y., Peng, Y., Chen, Z., Chen, X., 2022b. MsTGANet: Automatic drusen segmentation from retinal OCT images. *IEEE Trans. Med. Imaging* 41 (2), 394–406. <http://dx.doi.org/10.1109/TMI.2021.3112716>.
- Wu, J., Philip, A.-M., Podkowinski, D., Gerendas, B.S., Langs, G., Simader, C., Waldstein, S.M., Schmidt-Erfurth, U.M., 2016. Multivendor spectral-domain optical coherence tomography dataset, observer annotation performance evaluation, and standardized evaluation framework for intraretinal cystoid fluid segmentation. *J. Ophthalmol.* 2016 (1), 3898750.
- Xing, G., Chen, L., Wang, H., Zhang, J., Sun, D., Xu, F., Lei, J., Xu, X., 2022. Multi-scale pathological fluid segmentation in OCT with a novel curvature loss in convolutional neural network. *IEEE Trans. Med. Imaging* 41 (6), 1547–1559. <http://dx.doi.org/10.1109/TMI.2022.3142048>.
- Xing, R., Niu, S., Gao, X., Liu, T., Fan, W., Chen, Y., 2021. Weakly supervised serous retinal detachment segmentation in SD-OCT images by two-stage learning. *Biomed. Opt. Express* 12 (4), 2312–2327.
- Yan, Q., Ma, Y., Wu, W., Mou, L., Huang, W., Cheng, J., Zhao, Y., 2024. Choroidal layer analysis in OCT images via ambiguous boundary-aware attention. *Comput. Biol. Med.* 175, 108386.
- Yang, J., Ji, Z., Niu, S., Chen, Q., Yuan, S., Fan, W., 2020. RMPPNet: residual multiple pyramid pooling network for subretinal fluid segmentation in SD-OCT images. *OSA Contin.* 3 (7), 1751–1769. <http://dx.doi.org/10.1364/OSAC.387102>, URL <https://opg.optica.org/osac/abstract.cfm?URI=osac-3-7-1751>.
- Yang, J., Mehta, N., Demirci, G., Hu, X., Ramakrishnan, M.S., Naguib, M., Chen, C., Tsai, C.-L., 2024. Anomaly-guided weakly supervised lesion segmentation on retinal oct images. *Med. Image Anal.* 94, 103139.
- Yang, J., Tao, Y., Xu, Q., Zhang, Y., Ma, X., Yuan, S., Chen, Q., 2022. Self-supervised sequence recovery for semi-supervised retinal layer segmentation. *IEEE J. Biomed. Heal. Informatics* 26 (8), 3872–3883. <http://dx.doi.org/10.1109/JBHI.2022.3166778>.
- Yao, C., Wang, M., Zhu, W., Huang, H., Shi, F., Chen, Z., Wang, L., Wang, T., Zhou, Y., Peng, Y., Zhu, L., Chen, H., Chen, X., 2022. Joint segmentation of multi-class hyper-reflective foci in retinal optical coherence tomography images. *IEEE Trans. Biomed. Eng.* 69 (4), 1349–1358. <http://dx.doi.org/10.1109/TBME.2021.3115552>.
- Yuan, W., Lu, D., Wei, D., Ning, M., Zheng, Y., 2022. Multiscale unsupervised Retinal Edema Area segmentation in OCT images. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*. Springer Nature Switzerland, Cham, pp. 667–676. http://dx.doi.org/10.1007/978-3-031-16434-7_64.
- Zeppieri, M., Marsili, S., Enaholo, E.S., Shuaibu, A.O., Uwagboe, N., Salati, C., Spadea, L., Musa, M., 2023. Optical coherence tomography (OCT): a brief look at the uses and technological evolution of ophthalmology. *Medicina* 59 (12), 2114.
- Zhang, Y., Yang, Q., 2022. A survey on multi-task learning. *IEEE Trans. Knowl. Data Eng.* 34 (12), 5586–5609. <http://dx.doi.org/10.1109/TKDE.2021.3070203>.
- Zhang, H., Yang, J., Zheng, C., Zhao, S., Zhang, A., 2023a. Annotation-efficient learning for OCT segmentation. *Biomed. Opt. Express* 14 (7), 3294–3307.
- Zhang, H., Yang, J., Zhou, K., Li, F., Hu, Y., Zhao, Y., Zheng, C., Zhang, X., Liu, J., 2020a. Automatic segmentation and visualization of choroid in OCT with knowledge infused deep learning. *IEEE J. Biomed. Heal. Informatics* 24 (12), 3408–3420. <http://dx.doi.org/10.1109/JBHI.2020.3023144>.
- Zhang, H., Zhang, X., Zhang, Y., Higashita, R., Liu, J., 2023b. Prior-SSL: A thickness distribution prior and uncertainty guided semi-supervised learning method for choroidal segmentation in OCT images. In: Iliadis, L., Papaleonidas, A., Angelov, P., Jayne, C. (Eds.), *Artificial Neural Networks and Machine Learning – ICANN 2023*. Springer Nature Switzerland, Cham, pp. 570–581. http://dx.doi.org/10.1007/978-3-031-44210-0_46.
- Zhang, M., Zhou, Y., Zhao, J., Man, Y., Liu, B., Yao, R., 2020b. A survey of semi- and weakly supervised semantic segmentation of images. *Artif. Intell. Rev.* 53, 4259–4288.
- Zou, K., Chen, Z., Yuan, X., Shen, X., Wang, M., Fu, H., 2023. A review of uncertainty estimation and its application in medical imaging. *Meta-Radiol.* 1 (1), 100003. <http://dx.doi.org/10.1016/j.metrad.2023.100003>.