PC-Seg: Progressive Cross-View Consistency for 3D OCT Segmentation from Sparse 2D Annotations

Tsubasa Konno[1 [0009−0005−3069−8627]], Takahiro Ninomiya[2,3,4 [0000−0002−2214−6698]], Yukun Zhou[3,4,5,6 [0000−0002−0840−6422]], Koichi Ito[1 [0000−0001−7431−7105]], Siegfried K. Wagner[3,4 [0000−0003−4915−4353]], Yiqun Lin[3,4,5,6 [0000−0002−7697−0842]], Pearse A. Keane[3,4 [0000−0002−9239−745X]], Toru Nakazawa[2 [0000−0002−5591−4155]], and Takafumi Aoki[1 [0000−0001−8308−2416]]

[1] Graduate School of Information Sciences, Tohoku University, Japan
konno@aoki.ecei.tohoku.ac.jp
[2] Department of Ophthalmology, Graduate School of Medicine, Tohoku University, Japan
[3] Institute of Ophthalmology, University College London, UK
[4] NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust, UK
[5] UCL Hawkes Institute, University College London, UK
[6] Department of Computer Science, University College London, UK

Abstract. Volumetric segmentation of Optical Coherence Tomography (OCT) images is essential for diagnosing ocular diseases but requires labor-intensive voxel-wise annotations. While semi-supervised learning (SSL) can reduce annotation costs, most existing methods process data slice-by-slice, failing to exploit the inherent 3D spatial context. In this paper, we propose PC-Seg (Progressive Cross-view Segmentation), a novel curriculum learning framework that lifts sparse 2D annotations to high-precision 3D segmentation models. Unlike conventional multi-view approaches, our method employs a single 2D model to learn cross-view consistency from both standard B-scans and orthogonal slices, generating reliable volumetric pseudo-labels. These labels are then distilled into a 3D model, followed by a co-training phase where 2D and 3D models mutually refine each other through ensemble pseudo-labeling. Experiments on the MSHC dataset and the Duke DME dataset demonstrate that our method achieves accuracy comparable to fully supervised learning using only about 0.7% of the labeled data, outperforming both state-of-the-art semi-supervised and retinal layer segmentation methods. Our code is publicly available at `https://github.com/gsisaoki/pc-seg-official`.

Keywords: OCT · retina · segmentation · semi-supervised learning · curriculum learning

1 Introduction

Retinal layer thickness in Optical Coherence Tomography (OCT) serves as a critical biomarker for ocular diseases such as glaucoma and diabetic retinopa-

thy [24]. While deep learning has achieved high segmentation accuracy [11,15,24], it typically demands large-scale pixel-wise annotations, which are labor-intensive to obtain [23]. Semi-Supervised Learning (SSL) approaches aim to mitigate annotation costs [8, 17–19, 21, 23]; however, most existing methods are restricted to 2D slice-wise processing, failing to capture the inherent 3D spatial context.

In other 3D medical imaging domains, such as CT and MRI, 3D context is effectively captured via multi-view co-training among 3D networks [2, 27, 29] or cross-teaching between 2D and 3D networks [3, 25]. However, OCT volumes exhibit strong anisotropy, and sparse annotations are typically limited to standard B-scans [7–9, 22]. Consequently, applying methods [2, 3, 25, 27, 29] that utilize annotations on orthogonal planes or dense voxel-wise labels requires additional manual annotation. Moreover, simultaneously training multiple networks from scratch is computationally expensive and highly susceptible to overfitting when annotations are extremely sparse (e.g., only a few 2D slices) [3, 25, 31]. Therefore, a dimension-efficient approach designed to handle OCT anisotropy and data scarcity is highly desired.

To address the above issues, we propose Progressive Cross-view Segmentation (*PC-Seg*), a novel framework that lifts sparse 2D annotations to high-precision 3D segmentation by a five-stage curriculum learning strategy. Unlike conventional co-training that often suffers from confirmation bias due to early inaccurate predictions, PC-Seg begins by training a single 2D model to sequentially learn from standard and orthogonal B-scans. By enforcing cross-view consistency, the model accommodates OCT anisotropy and generates reliable volumetric pseudo-labels, which are then distilled into a 3D model. Finally, the 2D and 3D models mutually refine each other through ensemble pseudo-labeling. Experiments on two public datasets demonstrate that PC-Seg achieves accuracy comparable to fully supervised learning using only about 0.7% of the labeled data, outperforming both state-of-the-art semi-supervised and retinal layer segmentation methods.

2 Methods

This section introduces PC-Seg, which lifts sparse 2D annotations to a 3D segmentation model using a five-stage curriculum learning strategy. We describe the baseline model and SSL framework, followed by the details of the proposed progressive training stages.

2.1 Baseline Architecture and SSL Framework

The five-stage curriculum learning of PC-Seg is a model-agnostic framework. In this paper, we adopt the standard Residual U-Net (ResUNet) as the baseline model. In our preliminary evaluations, this CNN-based architecture demonstrated superior segmentation accuracy with significantly fewer parameters compared to Vision Transformer-based architectures, such as SegFormer [28], Swin-UNet [6], and UNETR [10]. Therefore, we employ one 2D ResUNet and one

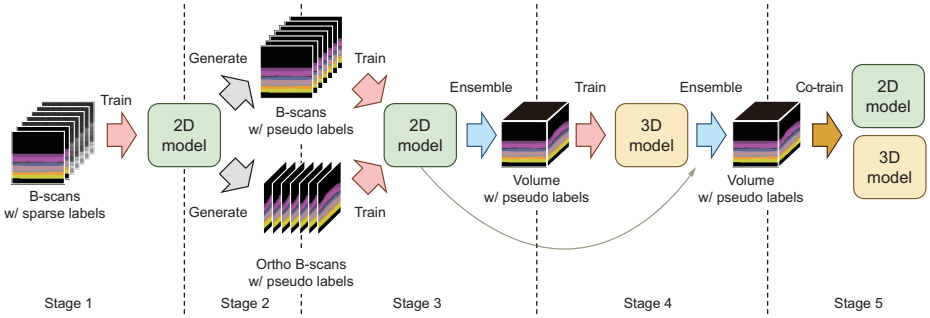

Fig. 1. Overview of the five-stage curriculum learning in PC-Seg. The 2D model trained in Stage 3 is utilized again in Stage 5 for the final ensemble. The "Train" arrows denote optimization via the baseline SSL framework (incorporating weak/strong/feature perturbations and cross-teaching) in Sect. 2.1.

3D ResUNet (both with depth=4 and base_channels=32) for all experiments in this paper, which are trained according to the curriculum described later. As the baseline SSL framework, we adopt UniMatch [30], a weak-to-strong consistency regularization method. UniMatch first generates a reliable pseudo-label from the prediction of an unlabeled image with weak perturbations by extracting pixels with confidence above a threshold τ. Then, an unsupervised consistency loss L_{unsup} is calculated so that multiple predictions under strong spatial and feature-level perturbations align with this pseudo-label. The overall loss function L_{SSL} is given as the weighted sum of the supervised loss L_{sup} for labeled data and the consistency loss. Both losses are computed using cross-entropy loss. In each stage of our proposed method, L_{SSL} is optimized when training with a mix of labeled and unlabeled data.

In the proposed method, instead of generating pseudo-labels from weakly perturbed inputs of the same view, we derive them from the predictions of other orthogonal views within the same volume. We explicitly enforce cross-view consistency by training the model to align its predictions under various perturbations with these cross-view pseudo-labels.

2.2 Progressive Cross-view Segmentation (PC-Seg)

Fig. 1 illustrates the overview of our proposed five-stage curriculum learning strategy, *PC-Seg*. PC-Seg efficiently trains a 3D segmentation model while preventing error accumulation in pseudo-labels by progressively expanding the target dimensions and views. Unlike standard training strategies that require high computational costs by training 3D models directly from scratch, our method optimizes resource allocation by refining pseudo-labels using a lightweight 2D model before introducing the 3D model. In this curriculum, each stage proceeds for fixed epochs: [10, 5, 5, 50, 50] for MSHC and [20, 10, 10, 40, 120] for Duke DME in the experiments. From Stage 2, pseudo-labels for the training set are

dynamically updated using the model checkpoint that yields the highest validation accuracy. This strategy mitigates error accumulation and progressively refines the pseudo-label quality.

Stage 1 (2D Warm-up): We train a 2D model using labeled and unlabeled B-scan images based on the UniMatch SSL framework. At the end of this stage, the trained model predicts segmentation masks for all B-scans, which are retained as initial pseudo-labels for the orthogonal (Ortho) B-scans for Stage 2.

Stage 2 (Adaptation to Orthogonal Slices): Unlabeled Ortho B-scan images are introduced into the training in addition to the B-scan images. For the Ortho images, we utilize the pseudo-labels generated in Stage 1. From Stage 2, unlike the standard UniMatch, we compute the consistency loss against pseudo-labels for predictions under all four (weak, feature-level, and two strong) perturbations. At the end of this stage, we predict all Ortho B-scans using the weights with the highest accuracy to generate B-scan pseudo-labels for Stage 3.

Stage 3 (Orthogonal Cross-teaching): We employ a cross-teaching strategy between orthogonal slices, where pseudo-labels for each plane are mutually generated from the predictions of its orthogonal counterpart. At the end of this stage, we stack 2D probability maps with simple coordinate permutation to generate high-quality 3D pseudo-labels.

Stage 4 (Knowledge Distillation to 3D Model): We conduct semi-supervised learning of the 3D model on sparsely labeled and unlabeled volumes, guided by the 3D pseudo-labels generated in Stage 3. During this stage, for sparsely labeled volumes, the supervised loss is computed using ground-truth labels for annotated slices and 3D pseudo-labels for unannotated ones. For unlabeled volumes, the consistency loss is optimized to align the model predictions with the generated pseudo-labels. At the end of this stage, we ensemble the two-plane predictions from the 2D model and the prediction from the 3D model to generate further refined 3D pseudo-labels.

Stage 5 (2D/3D Co-training): Using the highly accurate 3D pseudo-labels generated in Stage 4, we alternately train the 2D and 3D models. The ensemble pseudo-labels are dynamically updated using predictions from both models, enabling a joint optimization process that continuously improves the representations of both models.

3 Experiments

This section presents the experiments evaluating the effectiveness of the proposed method in retinal layer segmentation.

3.1 Datasets and Evaluation Metrics

We evaluated our proposed method on two public OCT datasets: the OCT MS and Healthy Control (MSHC) dataset [12] and the Duke DME dataset [7]. Following previous works [11,18,23], retinal flattening preprocessing [16] and intensity normalization were applied to all volumes.

MSHC Dataset: This dataset contains 35 volumes from 14 healthy controls and 21 patients with multiple sclerosis with 49 B-scans of $496 \times 1,024$ pixels per volume, annotated for nine layer boundaries. Following [23], we split the subjects into 12 for training, 3 for validation, and 20 for testing. For preprocessing, the images were flattened to a height of 128 pixels and then divided into patches of 128×128 pixels with a horizontal overlap of 64 pixels, yielding a total of 8,820 training patches. In the semi-supervised setting, we randomly selected 6, 30, or 60 patches from the training set as labeled data (approx. 0.07%, 0.3%, and 0.7%, respectively), treating the rest as unlabeled. During inference, predictions are performed using a sliding window of $128 \times 128 \times 48$ voxels with a $1/3$ overlap, and the outputs are averaged in the overlapping regions.
Duke DME Dataset: This dataset consists of 10 volumes from patients with diabetic macular edema (DME), each containing 61 B-scans of 496×768 pixels. Annotations for eight layer boundaries and fluid regions are provided, limited to a central region of approximately 500 pixels in width. Following [5], this dataset was divided into 6 volumes for training, 2 for validation, and 2 for testing. The volumes were flattened and randomly cropped to inputs of $224 \times 224 \times 48$ voxels. In the semi-supervised setting, we utilized the 66 labeled B-scans (11 slices per volume), supplemented by the remaining 300 unlabeled B-scans (50 slices per volume) from the training subjects. Similar to the MSHC dataset, inference is performed using a sliding window of $224 \times 224 \times 48$ voxels with a $1/3$ overlap.
Evaluation Metrics: We employed the F-score (Dice score) to evaluate segmentation accuracy. The mean F-score was calculated across all classes including the background for the MSHC dataset, and excluding the background for the Duke DME dataset. During inference, we applied a sliding window approach with a $1/3$ overlap and averaged the predictions.

3.2 Implementation Details

The models were trained using the AdamW optimizer [20] with an initial learning rate of 5.0×10^{-5}. The batch size was set to 2, consisting of one labeled and one unlabeled image (or volume) for the semi-supervised setting. Regarding the UniMatch perturbations, the weak perturbation A^w for labeled data utilized random cropping (100%), horizontal flipping (50%), vertical scaling (80%), and two OCT-specific augmentations [14] (80% each): Formula-Driven Data Augmentation (FDDA) and Partial Retinal Layer Copying (PRLC). For unlabeled data, A^w consisted only of random cropping (100%) and horizontal flipping (50%), while the feature-level perturbation A^{fp} applied dropout with a rate of 0.5 (100%). The strong perturbations A^{s1} and A^{s2} were generated independently using color jitter (80%), blur (20%), vertical scaling (80%), FDDA (80%), and PRLC (80%). The confidence threshold τ was empirically optimized and initially set to 0.95. However, pathological lesion classes typically exhibit lower prediction confidence, resulting in insufficient learning on unlabeled data. To address this issue, we dynamically lowered the threshold specifically for lesion classes to 0.5 during Stages 4 and 5. This adjustment is highly effective because the ensem-

Table 1. Ablation study of the proposed five-stage curriculum learning on the MSHC dataset. Values represent F-scores for each data type.

Method	Stage	2D model		3D model	Ensemble
		B-scan	Ortho	Volume	
Proposed	1	0.8744	—	—	—
	2	0.8917	0.8959	—	—
	3	0.8952	0.8969	—	—
	4	—	—	0.8993	—
	5	0.9016	0.9027	0.9020	**0.9047**
Only Stage 3	3	0.8713	0.8748	—	—
w/o Ortho B-scan	5	0.8806	—	0.8854	0.8857

ble predictions in these stages are inherently more robust and stable, effectively preventing the inclusion of erroneous pseudo-labels even at a lower threshold.

3.3 Experimental Results

In this section, we present the evaluation results of our proposed method using the MSHC and Duke DME datasets.

Ablation Study In this section, we conduct an ablation study to evaluate the effectiveness of each component of our proposed method using 6 labeled B-scan images of the MSHC dataset. Table 1 presents the performance evolution across the five stages of our curriculum learning. For comparison, it also includes baseline methods that skip specific progressive steps, denoted as "Only Stage 3" and "w/o Ortho B-scan". Compared to Stage 1, which uses only B-scans, Stage 2 introduces unlabeled Ortho B-scans and achieves a significant improvement in the F-score for B-scans. This result indicates that Ortho B-scans effectively serve as powerful additional unlabeled data. Furthermore, Stage 3 introduces cross-teaching to mutually generate pseudo-labels between orthogonal planes. This cross-teaching improves the accuracy of Ortho B-scans and narrows the performance gap between the two planes, resulting in a spatially consistent 2D model. In Stage 4, training the 3D model using pseudo-labels generated from the ensemble predictions of the 2D model yields higher accuracy than the standalone 2D model. Subsequently, applying 2D/3D co-training in Stage 5 facilitates the mutual acquisition of feature representations, allowing the final ensemble prediction to reach an F-score of 0.9047. This result exhibits a significant performance gain of approximately 0.03 compared to Stage 1.

To validate the necessity of our proposed progressive curriculum strategy, we compared it against settings that omit the progressive steps. The "Only Stage 3" method, which initiates cross-teaching without sufficient prior learning on B-scans, suffers from accuracy degradation compared to our Stage 3 due to error accumulation caused by low-quality initial pseudo-labels. Similarly, the "w/o

Table 2. Quantitative comparison on the flattened MSHC dataset. Values represent F-scores. For the most practical setting (60 labels), we report the mean ± standard deviation across 7 independent trials. Results for [19], [8], and [23] are cited from [23].

Method	# of labeled images			
	6	30	60	Full (8,820)
SGNet [19]	0.79	0.79	0.79	—
SD-LayerNet [8]	0.82	0.84	0.86	—
GOctSeg [23]	0.87	0.86	0.88	—
Structured-Layer [11] (Full)	—	—	—	0.9193
1D+2D U-Net [15] (Full)	—	—	—	0.9198
SemiVL [13]	0.8103	0.8919	0.9109±0.0027	0.9185
2D ResUNet w/ UM	0.8744	0.9079	0.9123±0.0029	0.9168
3D ResUNet	0.6822	0.9069	0.9086±0.0016	0.9176
3D ResUNet w/ UM	0.7843	0.9112	0.9084±0.0026	—
Proposed-2D (B-scan)	0.9016	0.9129	0.9152±0.0011	0.9164
Proposed-2D (Ortho)	0.9027	0.9128	0.9141±0.0011	0.9164
Proposed-3D (Volume)	0.9020	0.9133	0.9146±0.0019	0.9176
Proposed-Ensemble	**0.9047**	**0.9157**	**0.9175±0.0012**	**0.9212**

Ortho B-scan" method, which directly transitions to 3D model training without utilizing Ortho B-scans, also results in a drop in final accuracy. These results demonstrate that our learning strategy of progressively expanding dimensions and views through a sequence of B-scans, Ortho B-scans, and volumes is essential for optimizing the 3D model while maintaining high-quality pseudo-labels.

Comparison with State-of-the-Art Methods We compare PC-Seg against three groups of baselines: (i) Domain-specific supervised methods: Structured-Layer [11], 1D+2D U-Net [15], TCCT [26], SA CNN [4], and GD-Net [5]; (ii) Domain-specific semi-supervised methods: SGNet [19], SD-LayerNet [8], GOctSeg [23], and SASR [18]; and (iii) General semi-supervised frameworks: SemiVL [13], UniMatch (UM) [30] with ResUNet backbones, and DyCON [1]; Regarding our method, "Proposed-2D (B-scan)" and "Proposed-2D (Ortho)" denote the F-scores of the 2D model (Stage 5) on B-scan slices and Ortho B-scan slices, respectively. "Proposed-3D (Volume)" denotes F-score of the 3D model (Stage 5) on volumes, and "Proposed-Ensemble" denotes F-score of the ensemble of these three predictions.

Table 2 summarizes the quantitative comparison on the MSHC dataset. Our proposed method consistently outperforms existing baselines across all settings, showing statistically significant improvements for 60 labels across 7 trials ($p = 0.0156$, Wilcoxon signed-rank test). In particular, with only 60 labeled images (about 0.7% of the training data), PC-Seg achieves an F-score comparable to fully supervised methods (e.g., Structured-Layer [11] and 1D+2D U-Net [15]) trained on all 8,820 images.

Table 3. Quantitative comparison on the flattened Duke DME dataset. Values represent F-scores. Results from TCCT [26] to GD-Net [5] are cited from [5].

Method	ILM	NFL -IPL	INL	OPL	ONL -ISM	ISE	OS -RPE	Fluid	Mean
TCCT [26]	**0.88**	0.91	0.79	**0.79**	0.90	0.90	0.87	0.648	0.836
SA CNN [4]	**0.88**	0.91	0.78	0.78	0.90	**0.91**	**0.89**	0.536	0.823
GD-Net [5]	0.87	0.91	0.79	**0.79**	**0.91**	**0.91**	**0.89**	0.641	0.839
2D ResUNet w/ UM	**0.88**	0.91	0.80	**0.79**	0.90	**0.91**	0.86	0.635	0.835
SemiVL [13]	0.86	0.90	0.78	0.78	0.89	0.90	0.84	0.590	0.818
3D ResUNet w/ UM	0.83	0.85	0.73	0.71	0.89	0.87	0.70	0.670	0.780
SASR [18]	0.87	0.91	0.78	0.77	0.88	0.90	0.88	0.459	0.807
DyCON [1]	0.79	0.85	0.71	0.69	0.89	0.90	0.87	0.466	0.771
Proposed-2D (B-scan)	**0.88**	**0.92**	**0.81**	0.77	0.90	**0.91**	0.88	0.709	0.847
Proposed-2D (Ortho)	0.87	0.90	0.77	0.72	0.90	0.90	0.88	0.726	0.833
Proposed-3D (Volume)	0.87	0.90	0.77	0.74	0.90	0.90	0.88	0.736	0.838
Proposed-Ensemble	**0.88**	0.91	0.80	0.76	0.90	0.90	0.88	**0.742**	**0.848**

Table 3 summarizes the results on the Duke DME dataset. Following [5], we used 66 labeled B-scans (11 slices $\times$ 6 subjects) for training. This dataset is challenging due to the scarcity of labels and the high diversity of fluid lesions. While conventional methods struggle with the Fluid class (F-scores around 0.45–0.67), our method achieves a significant improvement, reaching an F-score of 0.742 in the ensemble prediction. This result demonstrates that incorporating 3D spatial context is crucial for accurately identifying complex pathological structures. In contrast, 3D semi-supervised methods (3D ResUNet w/ UM, SASR [18], DyCON [1]) fail to improve performance significantly under sparse label conditions. This result demonstrates that our progressive curriculum strategy effectively bridges the gap between 2D and 3D learning, enabling high-performance 3D segmentation even with extremely limited annotations.

Fig. 2 shows the qualitative comparison on the Duke DME dataset. Existing semi-supervised methods like SemiVL [13] and SASR [18] fail to detect the Fluid region (yellow) or under-segment it. Our "Proposed-2D (B-scan)" detects the lesion but lacks completeness. "Proposed-2D (Ortho)" captures the lesion better but suffers from inter-slice inconsistency, resulting in striping artifacts. However, "Proposed-3D (Volume)" effectively utilizes 3D context to smooth out these artifacts, and the final "Proposed-Ensemble" achieves the highest segmentation fidelity, accurately preserving both layer boundaries and lesion shapes.

4 Conclusion

In this paper, we proposed PC-Seg, a progressive cross-view consistency framework for semi-supervised 3D OCT segmentation from sparse 2D annotations. Although the multi-stage pipeline introduces implementation complexity, our

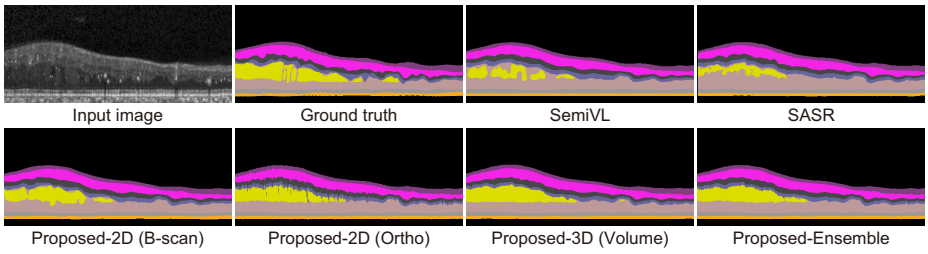

Fig. 2. Qualitative comparison of segmentation results on the Duke DME dataset.

curriculum strategy effectively mitigates error accumulation and bridges the performance gap between 2D and 3D models. Experiments on the MSHC dataset demonstrated that PC-Seg achieves accuracy comparable to fully supervised methods using only 0.7% of the labeled data. Furthermore, on the Duke DME dataset, PC-Seg outperformed state-of-the-art semi-supervised methods, particularly in detecting complex fluid lesions.

Acknowledgments. This work was supported in part by JSPS KAKENHI Grant Numbers 23H00463, 25K03131, and 25KJ0629, and the WISE program for Artificial Intelligence and Electronics in Tohoku University.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Assefa, M., Naseer, M., Ganapathi, I.I., Ali, S.S., Seghier, M.L., Werghi, N.: DyCON: Dynamic uncertainty-aware consistency and contrastive learning for semi-supervised medical image segmentation. Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition pp. 30850–30860 (Jun 2025)
2. Cai, H., Wang, R., Zhang, S., Wang, Y., Li, Y., Huang, X.: Orthogonal annotation benefits barely-supervised medical image segmentation. Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition pp. 3302–3311 (Jun 2023)
3. Cai, H., Qi, L., Yu, Q., Shi, Y., Gao, Y.: 3D medical image segmentation with sparse annotation via cross-teaching between 3D and 2D networks. Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention **14222**, 614–624 (Oct 2023)
4. Cao, G., Wu, Y., Peng, Z., Zhou, Z., Dai, C.: Self-attention CNN for retinal layer segmentation in OCT. Biomed. Opt. Express **15**(3), 1605–1617 (Mar 2024)
5. Cao, G., Zhou, Z., Wu, Y., Peng, Z., Yan, R., Zhang, Y., Jiang, B.: GCN-enhanced spatial-spectral dual-encoder network for simultaneous segmentation of retinal layers and fluid in OCT images. Biomedical Signal Processing and Control **98**, 106702 (Dec 2024)
6. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-Unet: Unet-like pure transformer for medical image segmentation. Proc. European Conf. Computer Vision Workshops **13803**, 205–218 (Oct 2023)

7. Chiu, S.J., Allingham, M.A., Mettu, P.S., Cousins, S.W., Izatt, J.A., Farsiu, S.: Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. Biomed. Opt. Express **6**(4), 1172–1194 (Apr 2015)
8. Fazekas, B., Aresta, G., Lachinov, D., Riedl, S., Mai, J., Schmidt-Erfurth, U., Bogunović, H.: SD-LayerNet: Semi-supervised retinal layer segmentation in OCT using disentangled representation with anatomical priors. Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention pp. 320–329 (Sep 2022)
9. Fazekas, B., Aresta, G., Seeböck, P., Mai, J., Schmidt-Erfurth, U., Bogunović, H.: SD-RetinaNet: Topologically constrained semi-supervised retinal lesion and layer segmentation in OCT. IEEE Trans. Med. Imaging pp. 1–15 (Oct 2025)
10. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: UNETR: Transformers for 3D medical image segmentation. Proc. IEEE/CVF Winter Conf. Applications of Computer Vision pp. 574–584 (Jan 2022)
11. He, Y., Carass, A., Liu, Y., Jedynak, B.M., Solomon, S.D., Saidha, S., Calabresi, P.A., Prince, J.L.: Structured layer surface segmentation for retina OCT using fully convolutional regression networks. Medical Image Analysis **68**, 101856 (Feb 2021)
12. He, Y., Carass, A., Solomon, S.D., Saidha, S., Calabresi, P.A., Prince, J.L.: Retinal layer parcellation of optical coherence tomography images: Data resource for multiple sclerosis and healthy controls. Data Brief **22**, 601–604 (Feb 2018)
13. Hoyer, L., Tan, D.J., Naeem, M.F., Gool, L.V., Tombari, F.: SemiVL: Semi-supervised semantic segmentation with vision-language guidance. Proc. European Conf. Computer Vision pp. 257–275 (Oct 2024)
14. Konno, T., Ninomiya, T., Miura, K., Ito, K., Himori, N., Sharma, P., Nakazawa, T., Aoki, T.: Formula-driven data augmentation and partial retinal layer copying for retinal layer segmentation. Proc. Ophthalmic Medical Image Analysis Workshop on MICCAI pp. 136–145 (Oct 2024)
15. Konno, T., Ninomiya, T., Miura, K., Ito, K., Himori, N., Sharma, P., Nakazawa, T., Aoki, T.: Retinal layer segmentation using 1D+2D U-Net from OCT images. Proc. IEEE Int'l Symp. Biomedical Imaging pp. 1–5 (May 2024)
16. Lang, A., Carass, A., Hauser, M., Sotirchos, E.S., Calabresi, P.A., Ying, H.S., Prince, J.L.: Retinal layer segmentation of macular OCT images using boundary classification. Biomed. Opt. Express **4**(7), 1133–1152 (Jul 2013)
17. Li, X., Fang, H., Liu, M., Xu, Y., Duan, L.: Diffusion-enhanced transformation consistency learning for retinal image segmentation. Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention pp. 221–231 (Oct 2024)
18. Liu, H., Wei, D., Lu, D., Tang, X., Wang, L., Zheng, Y.: Simultaneous alignment and surface regression using hybrid 2D–3D networks for 3D coherent layer segmentation of retinal OCT images with full and sparse annotations. Medical Image Analysis **91**(103019), 1–14 (Jan 2024)
19. Liu, X., Cao, J., Fu, T., Pan, Z., Hu, W., Zhang, K., Liu, J.: Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning. IEEE Access **7**, 3046–3061 (Dec 2018)
20. Loshchilov, L., Hutter, F.: Decoupled weight decay regularization. Proc. Int'l Conf. Learning Representations pp. 1–10 (May 2019)
21. Lu, Y., Shen, Y., Xing, X., Meng, M.H.: Multiple consistency supervision based semi-supervised OCT segmentation using very limited annotations. IEEE Int'l Conf. Robotics and Automation pp. 8483–8489 (May 2022)
22. Melinščak, M., Radmilović, M., Vatavuk, Z., Lončarić, S.: Annotated retinal optical coherence tomography images (AROI) database for joint retinal layer and fluid segmentation. Automatika **62**(3-4), 375–385 (Aug 2021)

23. Ong, C.Z.L., Ali, A.A.B., Rajapakse, J.C.: Generative adversarial learning for semi-supervised retinal layer segmentation in OCT images. Proc. IEEE EMBS Int. Conf. Biomedical and Health Informatics pp. 1–8 (Nov 2024)
24. Roy, A., Conjeti, S., Karri, S., Sheet, D., Katouzian, A., Wachinger, C., Navab, N.: ReLayNet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. Biomed. Opt. Express **8**(8), 3627–3642 (Aug 2017)
25. Su, H., Liu, Z., Yao, L., Li, S., Lin, H., Chen, G., Chen, X., Lei, H., Lei, B.: Sparsely annotated medical image segmentation via cross-SAM of 3D and 2D networks. Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention **15970**, 530–540 (Oct 2025)
26. Tan, Y., Shen, W.D., Wu, M.Y., Liu, G.N., Zhao, S.X., Chen, Y., Yang, K.F., Li, Y.J.: Retinal layer segmentation in OCT images with boundary regression and feature polarization. IEEE Trans. Medical Imaging **43**(2), 686–700 (Feb 2024)
27. Xia, Y., Liu, F., Yang, D., Cai, J., Yu, L., Zhu, Z., Xu, D., Yuille, A., Roth, H.: 3D semi-supervised learning with uncertainty-aware multi-view co-training. Proc. IEEE/CVF Winter Conf. Applications of Computer Vision pp. 3646–3655 (Mar 2020)
28. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. Proc. Advances in Neural Information Processing Systems **34**, 12077–12090 (Dec 2021)
29. Yan, K., Cai, Q., Zhang, F., Cao, Z., Liu, Z.: SGTC: Semantic-guided triplet co-training for sparsely annotated semi-supervised medical image segmentation. Proc. AAAI Conf. Artificial Intelligence **39**(9), 9112–9120 (2025)
30. Yang, L., Qi, L., Feng, L., Zhang, W., Shi, Y.: Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition pp. 7236–7246 (Jun 2023)
31. Zhang, Y., Liu, J., Liao, B., Wang, Q.: Multi-task semi-supervised 3D medical image segmentation based on CNN and self-attention. Proc. Int'l Conf. Health Big Data and Intelligent Healthcare pp. 233–237 (Dec 2024)