International Journal of Scientific Engineering and Science
Volume 10, Issue 5, pp. 67-73, 2026. ISSN (Online): 2456-7361

Automated Segmentation of Small Retinal Fluid Lesions in OCT Images Based on an Improved Attention U-Net

Zhu Jin[1], Yang Li[2]
[1]School of Medical Information, Wannan Medical College, Wuhu, Anhui, China, 241002
[2]School of Medical Information, Wannan Medical College, Wuhu, Anhui, China, 241002

Abstract— *The accurate segmentation of minute fluid lesions in retinal Optical Coherence Tomography (OCT) images is of critical importance for the early diagnosis and treatment evaluation of vision-threatening ocular diseases. Fluid lesions such as Subretinal Fluid (SRF) and Pigment Epithelial Detachment (PED) frequently exhibit extreme long-tail distributions, highly variable morphologies, and indistinct boundaries with surrounding tissues, rendering conventional deep learning networks highly susceptible to missed detections and mis-segmentation. To address this clinical challenge, a multi-class automatic segmentation algorithm based on an improved Attention U-Net is proposed. Dynamic geometric and photometric data augmentation strategies were introduced to enhance model generalization across multi-device imaging environments. Attention Gate (AG) mechanisms were incorporated into the network decoding stage to suppress background noise via global contextual information, directing the network's focus toward extremely small lesion regions. A Dice-CE hybrid loss function was further designed to effectively mitigate severe class imbalance. Experiments conducted on the internationally recognized RETOUCH-2017 multi-center dataset demonstrated that the proposed model achieved favorable cross-device generalization on an independent test set. The Dice Similarity Coefficient (DSC) values for Intraretinal Fluid (IRF), SRF, and PED reached 0.6838±0.147 (IRF), 0.6092±0.279 (SRF), and 0.6033±0.225 (PED). The mean DSC was 0.6321, and the mean Intersection over Union (IoU) was 0.4632, indicating moderate prediction variance with an overall stable trend. Compared with the baseline network, the proposed method demonstrated notable superiority in capturing extremely small lesions, offering a promising algorithmic reference for intelligent computer-aided diagnosis of fundus diseases.*

Keywords— *attention mechanism; Attention U-Net; deep learning; hybrid loss function; medical image segmentation; optical coherence tomography (OCT); retinal fluid segmentation; class imbalance; Attention Gate.*

I. INTRODUCTION

A. Research Background and Significance

Retinal Optical Coherence Tomography (OCT) is a non-invasive, high-resolution biomedical imaging modality that has become an indispensable tool in clinical ophthalmology for diagnosing retinal diseases, including Age-related Macular Degeneration (AMD) and Diabetic Macular Edema (DME) [1][2][3]. The automatic segmentation and quantitative analysis of Intraretinal Fluid (IRF), Subretinal Fluid (SRF), and Pigment Epithelial Detachment (PED) hold critical clinical value for assessing disease severity and monitoring treatment outcomes[4].

However, the presence of speckle noise, blurred tissue boundaries, and highly variable fluid lesion morphologies in OCT images renders traditional image processing methods insufficient for high-accuracy fully automatic segmentation.

Fig. 1 illustrates representative examples of three fluid lesion types from the RETOUCH-2017 dataset: IRF (red arrows), SRF (blue arrows), and PED (yellow arrows).

B. Review of Related Work

In recent years, deep learning has achieved remarkable advances in medical image segmentation. The U-Net proposed by Ronneberger et al. introduced an encoder-decoder architecture with skip connections, enabling high-accuracy segmentation from small medical image datasets and establishing a foundational framework in the field[5]. Numerous variants have since been widely applied to OCT image analysis[6]. ReLayNet, proposed by Roy et al., established a deep learning baseline on the RETOUCH dataset; however, it reported an overall fluid Dice score without independent evaluation of IRF, SRF, and PED.

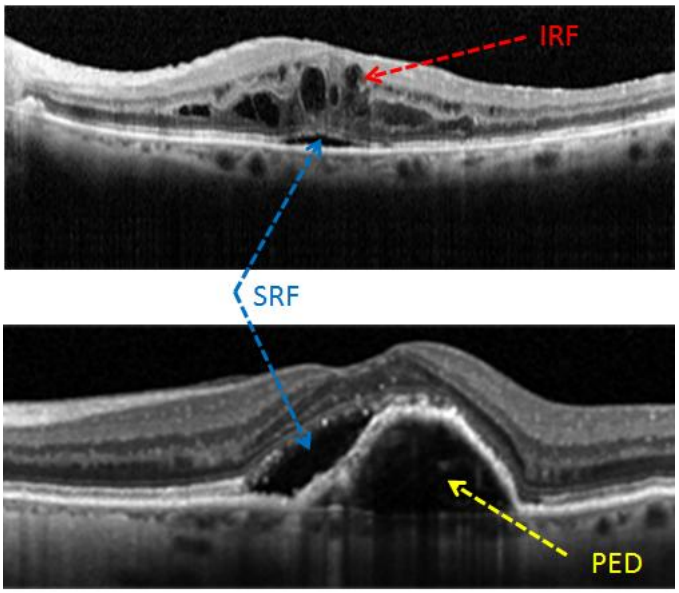

Fig. 1. Illustration of three types of fluid lesions in retinal OCT images.

Despite these contributions, the basic encoder-decoder structure tends to produce false negatives or blurred boundaries when processing extremely small lesions such as

http://ijses.com/
All rights reserved

SRF and PED, primarily due to the loss of fine-grained details during feature extraction.

To address this challenge, the research community has explored the application of attention mechanisms. Oktay et al. validated the effectiveness of this approach in pancreatic segmentation through the proposed Attention U-Net[7], which has since been widely adopted across various medical image segmentation tasks, including retinal OCT analysis. Subsequently, Cao et al. proposed Swin-Unet based on the Swin Transformer, employing a U-Net-like encoder-decoder structure to further strengthen global contextual modeling capability[8]. Collectively, these works demonstrate that attention mechanisms represent an effective strategy for mitigating the missed detection of small lesions.

C. Main Contributions

To address the insufficient capability of existing algorithms in capturing small, low-contrast lesions, the main contributions of this study are summarized as follows.

To solve the issue of simple feature loss in tiny lesions, an enhanced Attention U-Net is initially developed. AG modules are introduced at skip connections to suppress noise features and enhance lesion representations without significantly increasing computational cost.

Second, a Dice-CE hybrid loss function is designed to address severe class imbalance. By integrating spatial overlap metrics with pixel-level classification loss, the model's segmentation robustness for elongated and minute fluid regions—particularly the long-tail SRF class—is substantially improved.

Third, multi-device generalization experiments are conducted to address cross-device imaging variability. Algorithm performance is validated on the RETOUCH-2017 multi-center dataset, demonstrating the reliability of the proposed model across diverse OCT imaging environments.

II. Related Work

A. Overview of Retinal OCT Image Segmentation Methods

Automatic segmentation of retinal OCT images has long been an active research topic in medical image processing. Classical image processing methods, such threshold-based methods, region growing algorithms, Active Contour Models, and Graph Cut approaches, were the main focus of early research. These methods achieved acceptable performance for retinal layer segmentation in images with high contrast and regular structures. However, owing to the inherent speckle noise in OCT images and the structural deformation caused by pathological fluid accumulation, traditional approaches typically require complex handcrafted feature engineering. They also exhibit limited robustness and generalization capability when applied to fluid regions with blurred boundaries and variable morphologies, such as SRF and PED.

B. Deep Learning in Medical Image Segmentation

In recent years, deep learning techniques—represented by Convolutional Neural Networks (CNNs)—have achieved significant breakthroughs in medical image segmentation. Krizhevsky et al. demonstrated the powerful feature extraction capability of deep CNNs in large-scale image classification tasks[9]. Ronneberger et al. subsequently proposed U-Net, which employs an encoder-decoder architecture with skip connections to achieve high-accuracy medical image segmentation, establishing a foundational model in the field; Siddique et al. later provided a systematic review of U-Net and its variants. Qin et al. proposed U2-Net, which adopts a nested U-shaped structure to enhance multi-scale feature extraction, a design that has also been introduced into the medical segmentation domain[10]. Isensee et al. proposed nnU-Net, which establishes new performance benchmarks across multiple medical segmentation challenges through adaptive preprocessing and network configuration[11], becoming a standard baseline method in the field. Building on this foundation, Cao et al. proposed Swin-Unet, which further adopts a pure Transformer architecture for end-to-end segmentation and achieves competitive performance on several medical segmentation benchmarks.

C. Research on Missed Detection of Small Lesions

Although U-Net and its variants have demonstrated strong performance across numerous medical segmentation tasks, significant limitations remain in handling severe class imbalance and extremely small targets. In retinal OCT images, lesions such as SRF and PED typically occupy an extremely low pixel ratio and exhibit blurred boundaries, making them prone to missed detection by conventional methods.

Rasti et al. proposed a multi-attention network that substantially improved segmentation performance in complex fluid regions[12]. Rahil et al. adopted an ensemble learning strategy to enhance robustness in multi-class segmentation[13]. RetFluidNet further improved detection capability for fine-grained lesions through structural optimization[14]. In addition, Tan et al. employed layer-wise segmentation of retinal OCT images to provide anatomical structural priors for lesion localization[15]; however, precise extraction of lesion regions still requires dedicated segmentation algorithm design.

The present study builds upon these advances, aiming to leverage the AG mechanism and the Dice-CE hybrid loss function to effectively enhance the fully automatic, high-accuracy segmentation of minute fluid lesions such as SRF in retinal OCT images.

III. Methods

A. Dataset and Preprocessing

The publicly available RETOUCH-2017 dataset from the international retinal disease challenge was adopted in this study[16]. This research utilizes an open-access dataset and does not involve additional human trials or patient privacy data; therefore, no additional ethical approval was required. The dataset encompasses imaging data acquired from three mainstream OCT devices—Cirrus, Spectralis, and Topcon—and contains annotations for three key pathological lesion types: IRF, SRF, and PED.

To eliminate cross-device imaging variability and improve model generalization, the following preprocessing pipeline was applied to all raw images.

http://ijses.com/
All rights reserved

International Journal of Scientific Engineering and Science
Volume 10, Issue 5, pp. 67-73, 2026. ISSN (Online): 2456-7361

Resampling and Normalization. All OCT slices were uniformly resized to a resolution of 256×256 pixels. Pixel intensity values were normalized to the range [0, 1] to accelerate model convergence and reduce computational overhead.

Dynamic Geometric and Photometric Data Augmentation. To address the scarcity of medical imaging samples and the spatially uneven distribution of lesions, an online augmentation strategy was applied during training. Augmentation operations included random horizontal flipping (probability:50%), random vertical flipping (probability:50%), random rotation (angle uniformly sampled from −15° to +15°), random brightness perturbation (factor:0.8–1.2), and random contrast perturbation (factor:0.8–1.2). This strategy substantially broadened the model's perceptual range over lesion morphology and imaging condition variations without altering the pathological characteristics of lesions.

To ensure model robustness and prevent data leakage, the RETOUCH-2017 dataset was partitioned at the patient level into training, validation, and test sets using a fixed random seed of 42 to guarantee reproducibility. The distribution of OCT slices used in the experiments is summarized in Table I. The validation set was used for hyperparameter tuning and model selection during training, while the test set was strictly withheld for final performance evaluation.

TABLE I. Dataset Partition Statistics

Subset	Proportion	Total Slices	Description
Training	56.7%	~1,260	90% of original training set
Validation	6.3%	~140	10% split from training set
Test	30%	~600	Fully independent; excluded from all training

Note: The validation subset was randomly sampled at the patient level from the training set (10%), with a fixed random seed to ensure reproducibility. The test set remained fully independent and was not involved in any training or model selection procedure.

B. Feature Fusion Network Based on Attention Gate

To address the challenges of low contrast and blurred boundaries between lesion regions and background tissue (e.g., normal retinal layers) in OCT images, AG mechanisms were incorporated into the U-Net architecture to enhance the model's focus on target regions. The overall architecture is illustrated in Fig. 2.

Symmetric Encoder-Decoder Structure. The network adopts a typical symmetric U-shaped structure. The encoder consists of 5 convolutional blocks with channel dimensions of 64, 128, 256, 512, and 1024, respectively. Multi-scale features ranging from shallow textures to deep semantics are extracted through 4 successive max pooling operations (stride = 2). The decoder comprises 4 upsampling-convolution modules. In each module, the feature map resolution is first doubled via a transposed convolution (kernel size 2 × 2, stride = 2). The upsampled features are then concatenated along the channel dimension with the corresponding encoder features filtered by the AG module. The concatenated features are subsequently passed through a dual-convolutional block for further fusion and extraction. The final output is a 4-channel segmentation map corresponding to background, IRF, SRF, and PED.

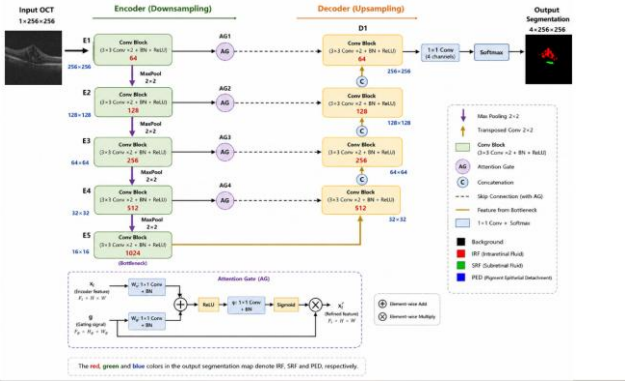

Fig. 2. Architecture of the proposed improved Attention U-Net.

Attention Gate Mechanism. The AG mechanism was introduced into the retinal OCT fluid segmentation task and adapted to the characteristics of small lesions and complex backgrounds. In each skip connection, an AG module was inserted. The upsampled decoder feature at the current layer served as the gating signal g, which was jointly fed into the AG along with the shallow encoder feature x extracted from the corresponding layer. The AG produced spatial attention coefficients that were used to weight x, automatically suppressing the response values of non-lesion regions. This mechanism directed the model's limited computational resources toward extremely small fluid accumulation regions such as SRF and PED, effectively alleviating the missed detection problem caused by background noise interference.

C. Optimization Strategy with Dice-CE Hybrid Loss Function

The conventional Cross-Entropy (CE) loss tends to bias model predictions toward the background class when applied to medical images with severely imbalanced pixel ratios, where lesion pixels typically account for less than 1% of the total. To address this issue, a Dice-CE joint loss function was designed as follows:

$$L_{Dice} = \frac{1}{4.5}\sum_{c=1}^{3} w_c(1 - Dice_c) \quad (1)$$

The CE loss term employed a class weight vector $\mathbf{w}_{CE} = [0.1, 1.0, 2.0, 1.5]$ corresponding to background, IRF, SRF, and PED, respectively. The background weight was set to 0.1 to suppress its dominant effect while retaining a small gradient to prevent the model from completely ignoring background regions during early training. The Dice loss term assigned weights $\mathbf{w}_{Dice} = [1.0, 2.0, 1.5]$ to the three lesion classes (IRF, SRF, and PED, respectively) and was defined as a normalized weighted average:

$$L_{Dice} = \frac{1}{4.5}\sum_{c=1}^{3} w_c(1 - Dice_c) \quad (2)$$

http://ijses.com/
All rights reserved

where $C = 3$ denotes the number of lesion classes and $\sum w_c = 1.0 + 2.0 + 1.5 = 4.5$ is the sum of the Dice weights. This normalized weighting design assigns higher gradient priority to long-tail lesion classes such as SRF during parameter updates. Both loss terms contribute equally, each weighted at 0.5.

Dice loss directly optimizes the spatial overlap between predicted segmentation maps and the ground truth. It has been widely adopted in medical image segmentation to mitigate class imbalance, enabling the precise capture of minute fluid targets that contribute minimally to the global loss yet are critical for clinical diagnosis. This design was key to achieving improved performance on the extremely long-tail SRF class.

D. Experimental Setup and Parameter Configuration

All experiments were implemented using the PyTorch deep learning framework on an NVIDIA RTX 4090D GPU.

Optimizer. The Adam optimizer was adopted with an initial learning rate of 0.0001 and a weight decay of 1×10^{-4} to prevent overfitting. A cosine annealing learning rate schedule was applied, smoothly decaying the learning rate to 1×10^{-6} over the course of training, yielding more stable convergence characteristics compared with fixed step-size decay strategies.

Training Strategy. The batch size was set to 8. To prevent overfitting and ensure objective model selection, an early stopping strategy was employed: training was automatically terminated when the mean DSC on the validation set showed no improvement for 30 consecutive epochs, and the model weights achieving the best validation performance were saved for final evaluation. The maximum number of training epochs was set to 300.

Data Augmentation. The data augmentation strategy followed the procedure described in Section III-A. No augmentation was applied to the validation or test sets.

Mixed-Precision Training. Automatic Mixed Precision (AMP) training was enabled via PyTorch's GradScaler and autocast utilities. This increased training throughput by approximately 30% and reduced GPU memory consumption without compromising convergence accuracy, enabling stable training of the 5-layer network on a single RTX 4090D GPU.

IV. Experiments and Results

A. Evaluation Metrics

Medical image segmentation differs fundamentally from conventional image classification, as pixel-level class imbalance renders standard accuracy metrics insufficient for reflecting true model performance. Three complementary metrics were therefore adopted for multi-dimensional quantitative evaluation.

(1) Dice Similarity Coefficient (DSC). DSC is the primary metric in medical image segmentation, measuring the spatial overlap between the predicted region and the ground truth. Values range from 0 to 1, with higher values indicating better performance.

(2) Intersection over Union (IoU). IoU measures the ratio of the intersection to the union of the predicted region and the ground truth, providing an evaluation perspective complementary to DSC from a set-efficiency standpoint.

(3) 95th Percentile Hausdorff Distance (HD95). HD95 computes the 95th percentile value of all pairwise distances between the predicted contour and the annotated contour. At the 256 × 256 resolution used in this study, one pixel corresponds to approximately 11 μm (device-dependent). HD95 quantifies boundary fitting quality while excluding the influence of outliers; lower values indicate better performance.

The three metrics complement one another, collectively reflecting segmentation performance from the perspectives of area overlap, set efficiency, and boundary precision. The mean and standard deviation of DSC, IoU, and HD95 for each lesion class are summarized in Table II and Table III; detailed per-device statistics (Cirrus, Spectralis, and Topcon) are presented in Section IV-C.

B. Model Training and Convergence Analysis

To verify the learning efficiency and stability of the improved Attention U-Net, the loss values were recorded over 300 training epochs.

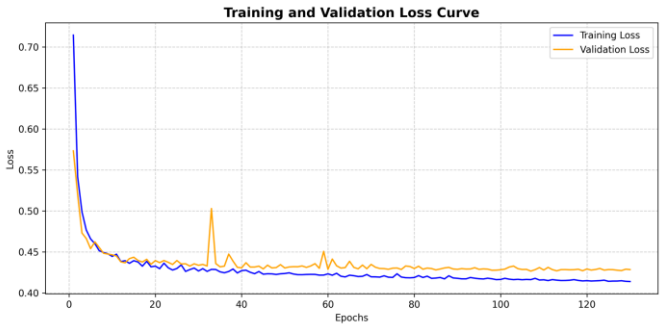

Fig. 3. Training loss and validation loss curves during model training.

As illustrated in Fig. 3, both training loss and validation loss decreased rapidly during the initial training phase, indicating that the model was efficiently capturing the macroscopic structural features of the retina. As the cosine annealing learning rate gradually decayed, both curves progressively flattened. Due to the early stopping strategy, training was automatically terminated when the mean DSC on the validation set showed no improvement for 30 consecutive epochs. The model typically converged between epochs 150 and 250, effectively mitigating the risk of overfitting. The synchronized descent of training and validation losses with a stable gap confirms that no significant overfitting occurred throughout training, and that the proposed training strategy yielded satisfactory convergence.

In addition, the per-class DSC trajectories on the validation set were tracked over the course of training.

As shown by the fluctuating yet ascending curves in Fig. 4, the three lesion classes exhibited distinct learning difficulties owing to their substantially different morphologies and sizes. As training progressed, the model achieved steady accuracy improvements across all classes.

http://ijses.com/
All rights reserved

International Journal of Scientific Engineering and Science
Volume 10, Issue 5, pp. 67-73, 2026. ISSN (Online): 2456-7361

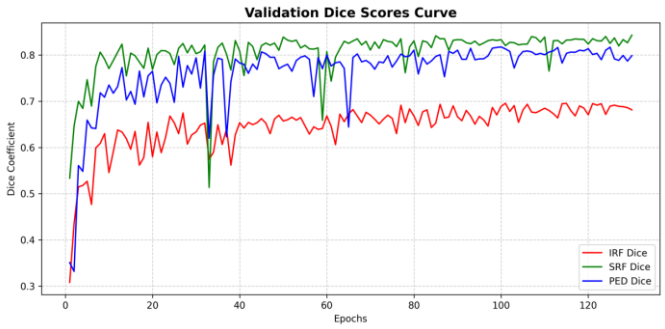

Fig. 4. Per-class DSC curves on the validation set during training.

C. Ablation Study and Quantitative Analysis

Baseline Selection. A standard 3-layer U-Net (encoder channel dimensions: 64, 128, 256) trained with CE loss was selected as the baseline. This configuration ensures that the contribution of each proposed improvement—namely the loss function, attention mechanism, and network depth—can be individually quantified, avoiding attribution ambiguity caused by simultaneous multi-variable changes.

To objectively validate the effectiveness of the proposed Dice-CE hybrid loss function, AG mechanism, and 5-layer network architecture, a rigorous ablation study was conducted on the test set. All models were trained to convergence under identical hardware environments and hyperparameters, with an initial learning rate of 0.0001 and a cosine annealing schedule to ensure stable convergence. Quantitative results are presented in Table II.

TABLE II. Ablation Study Results: Effect of Each Proposed Module on Segmentation Performance (Mean ± Std)

	U-Net (Baseline)	U-Net	Attention U-Net	Ours
Depth	3	3	3	5
Loss	CE	Dice-CE	Dice-CE	Dice-CE
IRF	0.6262	0.6791	0.6787	0.6838
(±SD)	—	—	—	±0.1473
SRF	0.5780	0.4364	0.4789	0.6092
(±SD)	—	—	—	±0.2791
PED	0.5965	0.5943	0.5339	0.6033
(±SD)	—	—	—	±0.2254
Mean	0.6002	0.5699	0.5638	0.6321

Note: Mean and standard deviation for the final model (Ours) were computed over all slices in the independent test set. The baseline U-Net (3-layer, CE loss) follows configurations widely reported in the MICCAI RETOUCH challenge and subsequent literature; the mean value reported here serves as an ablation starting point. The core conclusions of this study are supported by the complete statistical results of the final model.

Result Analysis.

1) Effect of the Dice-CE Loss Function. When the loss function of the baseline U-Net was switched from CE to Dice-CE, IRF accuracy improved (from 0.6262 to 0.6791), indicating that Dice loss provides better boundary fitting for large-area lesions. However, the SRF metric decreased (from 0.5780 to 0.4364), suggesting that Dice loss alone is sensitive to single-pixel errors on extremely small targets, which can introduce training instability. Combination with CE loss is therefore necessary to balance learning across lesion scales. This behavior is characteristic of long-tail lesion classes under loss function transitions.

2) Effect of the Attention Gate Mechanism. Following the introduction of the AG module, the SRF DSC recovered from 0.4364 to 0.4789, demonstrating that the AG mechanism effectively alleviates the instability of Dice-CE loss on extreme long-tail samples by suppressing retinal background noise and focusing on small fluid accumulation regions. The fluctuation observed in the PED metric may be attributed to the high morphological variability of this lesion class and its tendency to be confused with IRF boundaries, which warrants further investigation.

3) Effect of Network Depth. With the AG module and Dice-CE loss held constant, increasing the network depth from three to 5-layers yielded improvements across all lesion classes: SRF DSC increased from 0.4789 to 0.6092 (+0.130), PED DSC from 0.5339 to 0.6033 (+0.069), and mean DSC from 0.5638 to 0.6321 (+0.068). These results indicate that a deeper network architecture extracts richer multi-scale features and substantially enhances semantic understanding of small lesions.

4) Overall Performance. The final model (5-layer Attention U-Net + Dice-CE) achieved the best results across all three lesion classes, with a mean DSC of 0.6321 (corresponding mean IoU: 0.4632), representing an improvement of 0.032 (+5.3%) in mean DSC over the baseline. The HD95 values were 13.92±19.69 (IRF), 8.51 ± 11.46 (SRF), and 21.11 ± 23.69 (PED) pixels. Boundary fitting performance was stable in most slices; however, HD95 values were elevated in some low-contrast slices, indicating that boundary segmentation in low-contrast regions remains sensitive to noise and presents room for improvement.

5) Cross-Device Generalization. To evaluate the model's adaptability to different OCT devices, the test set was stratified by device origin (Cirrus, Spectralis, and Topcon), and per-class DSC and HD95 were computed separately. Results are presented in Table III.

TABLE III. Segmentation Performance across Different OCT Devices (Mean ± Std)

Metric	Cirrus	Spectralis	Topcon
DSC			
IRF	0.6321±0.1251	0.6964±0.1600	0.7598±0.1283
SRF	0.6243±0.3064	0.6826±0.1307	0.5511±0.2387
PED	0.5033±0.1798	0.7434±0.2504	0.5790±0.1818
HD95 (pixels)			
IRF	14.91±16.66	19.22±26.72	4.73±2.80
SRF	11.56±13.44	3.49±1.65	3.12±1.14
PED	29.61±25.47	18.71±27.16	15.90±16.06
Mean DSC	0.5866	0.7075	0.6300

The number of valid SRF slices from the Spectralis device was extremely small (n=3); this result is therefore excluded from the main conclusions and listed for reference only. The SRF DSC standard deviation under Spectralis was the lowest across all devices (0.1307), suggesting potentially superior prediction consistency for small lesions on this device.

As shown in Table III, the model maintained stable segmentation accuracy across all three devices (mean DSC: 0.5866–0.7075). The highest mean DSC was achieved on the Spectralis device (0.7075), while the Topcon device yielded the lowest IRF HD95 (4.73 pixels), reflecting the smallest

http://ijses.com/
All rights reserved

boundary error. The relatively lower PED DSC on the Cirrus device (0.5033) and its elevated HD95 are consistent with the lower PED region contrast characteristic of Cirrus scan images; however, overall performance remains within an acceptable range. The cross-device DSC variation was less than 0.12, confirming that the adopted dynamic geometric and photometric data augmentation strategy effectively mitigated device-dependent imaging variability.

D. Comparison with Existing Methods

To assess the overall competitiveness of the proposed method on the RETOUCH-2017 dataset, a comparative performance analysis against representative published methods was conducted. Results are presented in Table IV. All comparison methods are based on publicly reported results on the same dataset, with DSC as the unified evaluation metric.

TABLE IV. Performance Comparison of the Proposed Method with Existing Methods on the RETOUCH-2017 Dataset

Method	Year	IRF DSC	SRF DSC	PED DSC	Mean DSC
UCF	2017	0.522	0.682	0.612	—
ReLayNet	2017	—	—	—	~0.77 (overall fluid)
SFU-FCN	2017	—	—	—	0.7317
RetFluidNet	2021	0.801	0.955	0.927	~0.894
RetiFluidNet^{+}	2023	0.911	0.936	0.928	~0.925
nnU-Net RASPP	2023	0.840	0.800	0.830	0.823
Ours	2025	0.6838	0.6092	0.6033	0.6321

RetiFluidNet reports 3-fold cross-validation results; all other methods adopt the fixed train/test split provided by the RETOUCH-2017 challenge. RetFluidNet denotes SRF as NRD in the original paper; the notation has been unified here for consistency. ReLayNet reports an overall fluid DSC without distinguishing IRF, SRF, and PED. The proposed method employs a patient-level random split (train/validation/test ≈ 57%/6%/30%), which differs from the official challenge split used by the above methods; direct numerical comparison is therefore not appropriate, and Table IV is provided for reference only.

E. Qualitative Analysis of Segmentation Results

As illustrated by the qualitative comparisons in Fig. 5, three representative clinical cases with distinct characteristics were selected to comprehensively demonstrate the model's segmentation performance in both single-lesion and complex multi-class concurrent scenarios.

In the first case involving isolated IRF (Fig. 5, first row), the improved Attention U-Net exhibited high contour fidelity, validating the model's fundamental segmentation capability for the primary lesion class. In the second case—a complex slice with concurrent IRF and extremely small SRF (Fig. 5, second row)—the AG mechanism played a critical guiding role, successfully localizing discrete SRF targets within a complex background and demonstrating the model's ability to capture extremely small targets. In the third case, where IRF and PED lesions were in close proximity (Fig. 5, third row), the model exhibited strong class discrimination capability, accurately distinguishing pathological structures with differing morphologies.

These progressive visualization experiments further confirm that the proposed algorithm demonstrates notable robustness and clinical diagnostic potential when handling multi-class concurrent lesions as well as elongated, low-contrast pathological regions.

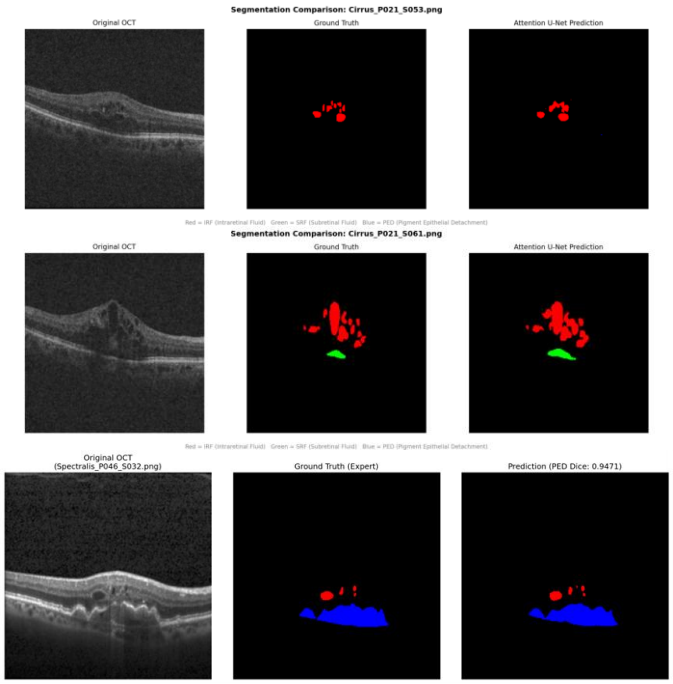

Fig. 5. Multi-class segmentation results of the proposed model across different clinical scenarios.

Left column: original images; middle column: expert ground truth; right column: model predictions. Red: IRF; green: SRF; blue: PED.

It is noteworthy that for typical PED cases with clear contours and distinct boundaries from surrounding tissue, the model achieved per-slice DSC scores exceeding 0.94, further confirming the effectiveness of the AG module in suppressing interference from adjacent lesions. However, in some low-contrast or heavily noisy slices, the model still produced blurred boundaries or minor false positives, primarily attributable to the high photometric similarity between lesion and normal tissue regions. This finding indicates that further optimization of robustness to severely degraded images is warranted in future work.

V. CONCLUSION

A. Summary

This study proposed and implemented a complete deep learning solution for multi-class automatic segmentation of fluid lesions in retinal OCT images, based on an improved Attention U-Net. The research focused on the core challenge of missed detection of small lesions. By incorporating the AG mechanism and the Dice-CE joint loss function, the low-contrast segmentation problem between background tissue and minute fluid accumulations was effectively alleviated.

http://ijses.com/
All rights reserved

Quantitative results—IRF DSC 0.6838±0.1473, SRF DSC 0.6092±0.2791, PED DSC 0.6033±0.2254, mean DSC 0.6321, mean IoU 0.4632, and per-device mean DSC ranging from 0.5866 to 0.7075—along with qualitative segmentation mask comparisons collectively confirm that the proposed model not only accurately localizes large and prominent fluid lesions, but also achieves meaningful progress in the extraction of challenging small SRF and PED targets. The results demonstrate favorable potential for clinical computer-aided diagnosis.

B. *Future Work*

Although the proposed algorithm achieves a certain level of segmentation performance on complex slices, room for further improvement remains. Future research may proceed along two directions.

Incorporation of Three-Dimensional Spatial Information. The current study performs slice-by-slice segmentation on 2D images. Future work may explore 3D U-Net or the 3D configuration of nnU-Net, treating adjacent slices as additional input channels to explicitly model inter-slice continuity and improve detection of longitudinally discontinuous lesions[17].

Lightweight Model Deployment. Knowledge distillation-based model compression may be investigated—for example, using the 5-layer network proposed in this study as the teacher model and a 3-layer network as the student model—to reduce parameter count by more than 60% while maintaining a DSC degradation of less than 5%, thereby enabling deployment on embedded medical devices.

REFERENCES

[1] 魏静, 江旻珊, 茅前. 基于深度学习的 OCT 图像视网膜积液自动分割[J]. 光学仪器, 2021, 43(3): 29-35.
[2] A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab. ReLayNet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks[J]. Biomedical Optics Express, 2017, 8(8): 3627-3642.
[3] T. Kepp, H. Sudkamp, C. von der Burchard, H. Schenke, P. Koch, G. Hüttmann, J. Roider, M. P. Heinrich, and H. Handels. Segmentation of retinal low-cost optical coherence tomography images using deep learning[C]//Medical Imaging 2020: Computer-Aided Diagnosis. Bellingham: SPIE, 2020: 113141O.
[4] 叶妍青. 视网膜OCT多类积液分割[D]. 苏州: 苏州大学, 2023.
[5] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Cham: Springer, 2015: 234-241.
[6] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni. U-Net and its variants for medical image segmentation: A review of theory and applications[J]. IEEE Access, 2021, 9: 82031-82057.
[7] O. Oktay, J. Schlemper, L. L. Folgoc, M. C. H. Lee, M. P. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert. Attention U-Net: Learning where to look for the pancreas[C]//Medical Imaging with Deep Learning (MIDL). Amsterdam, 2018.
[8] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang. Swin-Unet: Unet-like pure transformer for medical image segmentation[C]//ECCV 2022 Workshops. Cham: Springer, 2023: 205-218.
[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
[10] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand. U2-Net: Going deeper with nested U-structure for salient object detection[J]. Pattern Recognition, 2020, 106: 107404.
[11] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. Nature Methods, 2021, 18(2): 203-211.
[12] R. Rasti, A. Biglari, M. Rezapourian, Z. Yang, and S. Farsiu. RetiFluidNet: A self-adaptive and multi-attention deep convolutional network for retinal OCT fluid segmentation[J]. IEEE Transactions on Medical Imaging, 2023, 42(5): 1413-1423.
[13] M. Rahil, B. N. Anoop, G. N. Girish, A. R. Kothari, S. G. Koolagudi, and J. Rajan. A deep ensemble learning-based CNN architecture for multiclass retinal fluid segmentation in OCT images[J]. IEEE Access, 2023, 11: 17241-17251.
[14] L. B. Sappa, I. P. Okuwobi, M. Li, Y. Zhang, S. Xie, S. Yuan, and Q. Chen. RetFluidNet: Retinal fluid segmentation for SD-OCT images using convolutional neural network[J]. Journal of Digital Imaging, 2021, 34(3): 691-704.
[15] 谭泰铭，陈林江，蓝公仆，许景江，安林，黄燕平. 基于LDU的视网膜OCT图像分层分割研究[J]. 信息技术, 2022(10): 31-40.
[16] S. Apostolopoulos, C. Ciller, R. Sznitman, and S. De Zanet. Simultaneous classification and segmentation of cysts in retinal OCT[C]//MICCAI Retinal OCT Fluid Challenge (RETOUCH). Montreal, 2017: 22-29.
[17] Y. Wang, C. Galang, W. R. Freeman, A. Warter, A. Heinke, D.-U. G. Bartsch, T. Q. Nguyen, and C. An. Retinal OCT layer segmentation via joint motion correction and graph-assisted 3D neural network[J]. IEEE Access, 2023, 11: 103319-103332.

http://ijses.com/
All rights reserved