



A method for recovering shadow retinal OCT images by the auxiliary gradient branch model

Yukuan Wang^a, Wei Yi, Hua Wei, Yi Wan, Zhengkai Yao, Xiaoqi Lu, Peng Sheng, Xianchun Jiang, Weiye Song^{*}

School of Mechanical Engineering, Shandong University, Jinan 250100, PR China

ARTICLE INFO

Keywords:

Retinal OCT images
Neural network
Shadow segmentation
Shadow recovery

ABSTRACT

Optical coherence tomography has become the gold standard for diagnosing ophthalmic diseases. It obtains the retinal structure of the eye through noninvasive measurements. However, the absorption of incident light by blood flow in blood vessels makes shadow areas appear in B-Scan images at the vessel locations, which seriously affects the accuracy of retinal layer segmentation and the observation of small lesions. In this research, a two-stage neural network is proposed, which is established by auxiliary training head and a global attention mechanism between shadowed and non-shadowed areas. In the task of segmenting shadow areas, this work achieves that the segmentation mIoU of retinal B-Scan images is 79.07%. Based on the segmentation mask, the shadow area is accurately recovered. The SSIM of the Sobel edge image is used to evaluate the effect of the recovered area. The proposed method achieves an SSIM_{area} of 98.2% between the shadow areas after and before recovery, and the average pixel brightness is improved by 36.7. Compared to all existing methods, this work introduces model constraints to ensure the structural invariance of the original image for the first time, resulting in an improvement of SIMM_{total} from 47% to 99.9%. Finally, the proposed method is an improvement of clinical retinal layer segmentation accuracy by comparing the effect of RPE-Choroid layer segmentation of four types of retinal B-Scan images.

1. Introduction

Optical coherence tomography (OCT) is a non-invasive measurement method that allows the non-destructive detection of tissues below the surface of an object [1]. OCT is widely used in the diagnosis of ophthalmic diseases because of retinal structures cannot be directly measured [2]. Lots of research has demonstrated that glaucoma [3], choroidal neovascularization [4] and age-related macular degeneration [5] are among the diseases that have a direct link both retinal nerve fiber layer thickness and changes of the retinal layers structure. So that, both retinal layer thickness and retinal B-Scan structure are recognized as the gold standard for the diagnosis of ocular diseases [6,7].

Retinal layer thickness measurement and retinal structure changes only can be characterized by B-Scan images of OCT. Thus, how to obtain high-quality, more information, and low-noise B-Scan images is the key to finding ocular diseases. The incident light is absorbed by the blood flow in the blood vessels, which creates shadow areas in the retinal B-Scan image and even loss of structural information [8,9]. The presence

of vascular shadows makes it more difficult to observe small lesions at the bottom of blood vessels and miss the opportunity for early detection.

Due to the difficulty in obtaining shadow-free retinal images for labeling, and since existing recovery methods generally employ generative redrawing methods for shadow image recovery, the recovery results lack credibility in clinical applications. Thus this, we propose a two-stage neural network method for recovering B-Scan images: The first stage involves developing the OSSNet model for segmentation of shadowed areas in B-Scan images, which utilizes a feature map skip-connection method similar to U-Net and auxiliary training head. Both can provide the backbone network with more structural information to enhance the accuracy of shadow segmentation. The second stage involves developing the OSRNet model for recovering shadowed areas in B-Scan images based on a binary mask mechanism, which achieves precise recovery of shadowed areas through the integration of a global receptive field based on the SW-Transform architecture and auxiliary training head. We propose a transfer learning strategy for recovering shadowed areas by learning the distribution patterns of shadows in

^{*} Corresponding author.

E-mail address: songweiye@sdu.edu.cn (W. Song).

<https://doi.org/10.1016/j.bspc.2026.111302>

Received 10 November 2025; Received in revised form 9 July 2026; Accepted 11 August 2026

Available online 17 August 2026

1746-8094/© 2026 Published by Elsevier Ltd.

natural images. The model can incorporate prior knowledge into B-Scan image analysis, and it can achieve the aim of recovering the shadowed areas without altering the original structural information in the images (as shown in Fig. 1).

The main contributions of this research are summarized as follows:

1. A shadow segmentation detection dataset is established based on the OCT2017 dataset, containing 1440 B-scan images across four types of retina images.
2. A shadow segmentation model, OSSNet, is proposed based on auxiliary training heads, which improves segmentation accuracy while minimizing interference with non-shadowed areas.
3. A method for generating artificial shadows in B-scan images is proposed, simulating the shape characteristics of shadowed areas. And then, we establish an artificial B-Scan shadow dataset with varying degrees of shadow coverage.
4. This work proposes a shadow recovery model, OSRNet, for B-scan images based on a mask mechanism. It captures shadow distribution patterns through transfer learning, and preserves image structural consistency.
5. We verify the improvement in segmentation accuracy of the RPE-Choroid layer of the retina before and after shadow recovery.

2. Related work

At present, there are two ways to recover medical image artifact for improving image quality: The first way is to optimize the hardware optical path for recovering the shadow area. For example [10], designs an optical system, that changes the second channel to incorporate into a typical ICI system, to remove the self-coherent image in coherent imaging and enhance the image quality. Another way involves reconstructing the B-Scan image through image processing, using statistical features of image pixels or a deep learning neural network for direct image handling. Numerous researchers have employed mathematical analysis methods to recover artifacts in medical images, such as mathematical interpolation compensation for image [11,12], constructing rank-1 auxiliary matrices [13], and canonical correlation analysis [14], process every pixel in the image when dealing with artifacts. It leads to increasing noise while reducing artifacts, and lacking precision and robustness. Consequently, deep learning methods have been widely applied in medical image artifact recovery. Many researchers have used generative network models for image recover, just like studies based on GAN(Generative Adversarial Networks) have shown that generators exhibit strong generalization in artifact recovery tasks, continuously improving image quality through the training of generators and discriminators [15,16]. Another approach involves designing network structures to encode medical images and extract deeper feature relationships, which can generate the large missing artifact areas in the images [17]. For example [18], developed a B-Scan shadow image detection model to remove images containing shadows from the database, thus enhancing overall image quality. Ref. [19] denoised and

recovered the shadowed areas of B-Scan images using a GAN(Generative Adversarial Network) model. The above methods have explored ways to enhance the quality of B-Scan images.

Then, some researchers have achieved shadow area recovery through variations of the GAN model, where the discriminator cannot distinguish the recovery of shadows generated by generator in the image [20,21]. The common issue with GAN-based recovery methods is that all pixels in the entire image are processed, which cannot guarantee image structure invariance, a crucial aspect of medical images. To address this issue, researchers have established binary masks during input to precisely control the attention area of the recovery model, which can ensure image structure invariance during the recovery process [22]. In the field of medical image artifact recovery, inspired by natural image models, researchers have utilized GAN networks to achieve artifact recovery. However, these models also process all pixels in the entire image, sometimes altering the image structure and generating incorrect information.

To avoid generating error structural information and enhance the smoothness of pixel gradients between shaded and non-shaded regions. The binary mask mechanism is established for the first time to achieve precise control over the recovery area in B-Scan images. It leverages the transformer mechanism and auxiliary training heads to efficiently transmit information between shaded and non-shaded areas. The binary mask is output through a high-precision segmentation model.

2.1. Proposed method

This section will detail the proposed two-stage recovery model, which is designed to identify shadow pixels and recover the brightness of the shadow areas. The model framework consists of three components, as illustrated in Fig. 2:

- OSSNet(Image Segmentation Model): It achieves pixel-level classification through supervised training to ensure the accuracy of the recovery area.
- OSRNet(Image Recovery Model): It enhances regional constraints by introducing binary masks and combines a global attention mechanism to ensure the accuracy of the recovery process.
- Transfer Training Strategy: Shadow distribution patterns from natural images are used in secondary training on B-Scan images, and the model's ability to perceive shadows is significantly enhanced.

Through the collaboration of these five components, the model can identify shadow areas, perceives the relationship between shadow and non-shadow areas, constrains the recovery area, leverages auxiliary multi-head information to improve accuracy, and recover the shadow areas.

2.2. Shadow segmentation model

Numerous researchers in the field of medical image segmentation

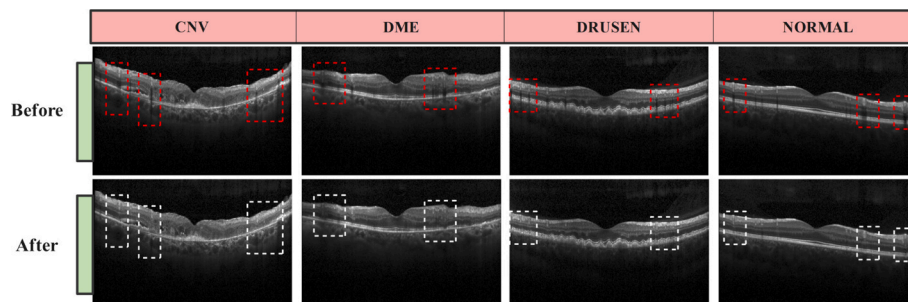


Fig. 1. Comparison of retina B-scan images recovery results. The red rectangular boxes in the “Before” rows indicate areas before recovery, and the white rectangular boxes in the “After” rows indicate areas after recovery.

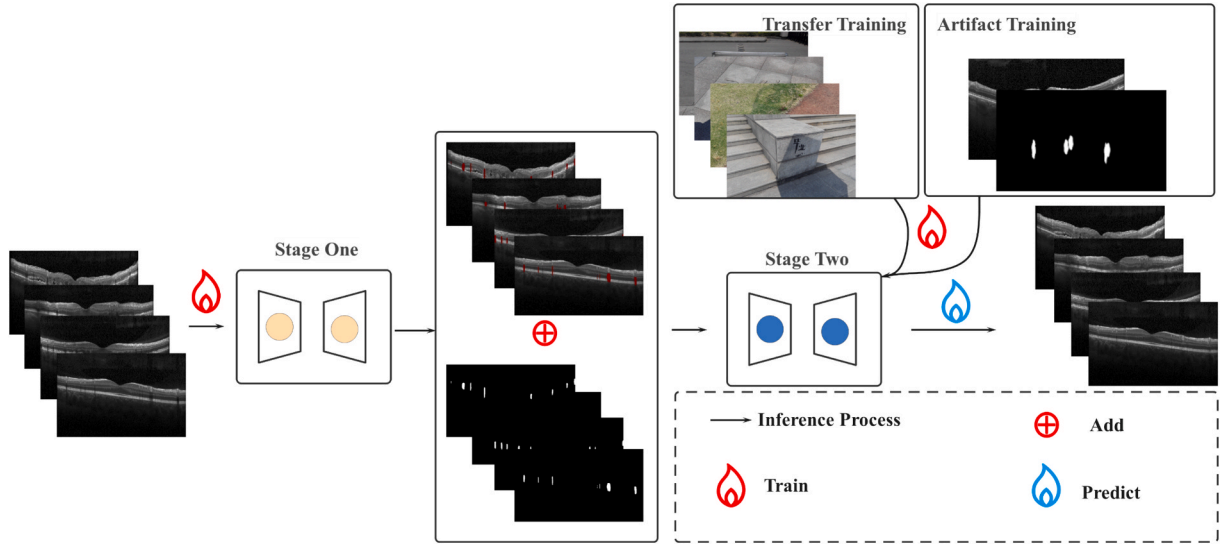


Fig. 2. Flowchart of the Two-Stage Recovery Model. In the figure, Stage One represents the OSSNet, which is the B-Scan shadow segmentation model. Stage Two represents the OSRNet, which is the B-Scan shadow area recovery model.

have demonstrated that full-size feature map fusion methods similar to the U-Net structure are highly effective [23–27]. Based on the U-Net concept, this work constructs the OSSNet segmentation model (as shown in Fig. 3) using Resnet50 as the backbone network combining features at various levels. During the forward propagation of the model, the original image $X \in R^{H \times W \times 3}$ (where H and W represent the height and width of the image) undergoes downsampling and channel expansion, and the resulting feature maps are passed into the Res-Blocks. Each Block contains BatchNorm normalization, $Conv_{3 \times 3}$ feature extraction, $Conv_{1 \times 1}$ data dimensionality reduction, residual connections, and ReLU activation. The backbone network undergoes four downsampling operations during forward propagation, and generate five feature maps of different scales. During each downsampling, the width and height of the feature map become half of the previous layer's feature map, and the number of channels is doubled along the depth direction of the feature map. As the

size of the feature maps decreases, the receptive field increases, which allows the observation of more feature information between pixels.

The five feature maps are concatenated with the feature maps on the prediction branches through upsampling, in order to enhance the feature fusion capability of the prediction branches. Lots of researchers have found that the fusion of features at different scales can improve model performance. So, the segmentation accuracy can be enhanced by combining multi-scale features [28–30]. In this research, the model utilizes Silence layers to preserve the initial feature map information, CBLiner layers to split multi-scale feature map information, and CBFuse layers to fuse multi-scale feature map information through programmable gradients [31].

Specifically, the Silence module is instantiated as a non-linear transformation layer designed to execute localized feature extraction. By performing high-level manifold mapping on the current feature

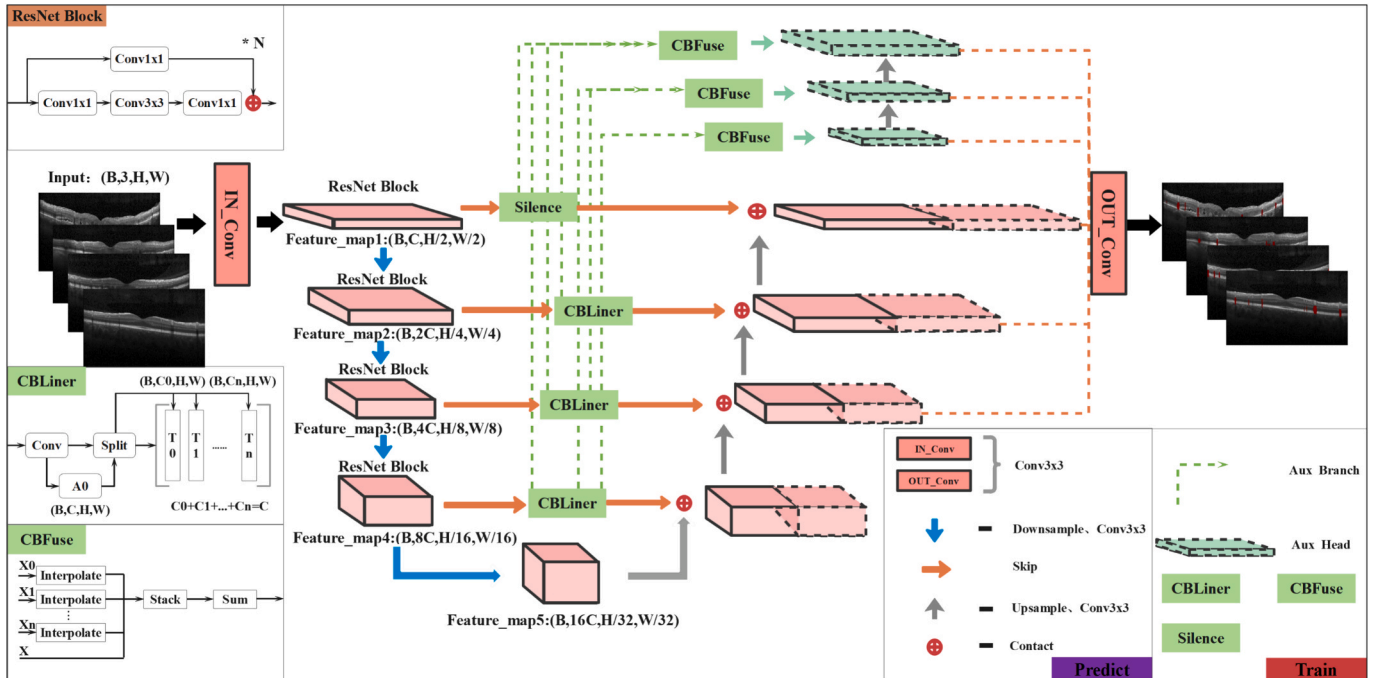


Fig. 3. Structure of the Shadow Segmentation Model.

representation, it effectively enhances the model's selective activation, thereby precisely preserving the structurally salient features within the target regions of interest. The mathematical formulation is defined as follows:

$$X_{\text{silence}} = \text{Silu}(\text{BatchNorm}(\text{Conv}(X))) \quad (1)$$

To address spatially confined retinal vessel shadows and fortify feature representation in highly attenuated, low-signal zones, we propose the multi-scale splitting CBLinear module. By factorizing intermediate features across substantial architectural spans, this module enables shallow layers to adaptively perceive multi-level contextual variations, thereby synergistically bolstering hierarchical semantic and structural encoding. It is formalized as follows:

$$X_{\text{fused}} = \text{Conv}(X) + A_0(\text{Conv}(X)) \quad (2)$$

where A_0 represents the 1×1 convolution with SiLU activation, tasked with performing non-linear manifold mapping on the feature maps generated by the 3×3 convolution. Thereafter, the fused feature map is factorized along the channel dimension, and the resulting sub-features are progressively cascaded into the antecedent feature layers, as expressed below:

$$\{X_0, X_1, \dots, X_n\} = \text{Split}(X_{\text{fused}}) \quad (3)$$

$$\sum_{i=0}^n C_i = C_0 + C_1 + \dots + C_n = C \quad (4)$$

Upon obtaining the activated feature maps, to eliminate the spatial misalignment across heterogeneous feature hierarchies, the CBFuse module first employs an interpolation algorithm to adaptively rescale all auxiliary feature maps to the target canonical resolution:

$$\hat{X}_i = \text{Interp}(X_i, \text{size} = (H_i, W_i)), \quad \text{for } i = 0, 1, \dots, n \quad (5)$$

where H_i and W_i denote the canonical spatial resolution of the current feature hierarchy. To eliminate scale discrepancies, all aligned auxiliary features are aggregated with the baseline feature via channel-wise concatenation. This architecture establishes a cross-scale contextual synergy mechanism, enhancing the model's capacity to characterize both macro-scale geometric topologies and micro-scale boundary details.

$$Y = \text{Sum}(\text{Stack}(\hat{X}_0, \hat{X}_1, \dots, \hat{X}_n, X)) \quad (6)$$

Ultimately, the multi-scale auxiliary detection heads are to achieve efficient gradient backpropagation during the training process. During the prediction stage, the model can achieve high-precision and high-speed predictions by removing the auxiliary training branches and redundant detection heads.

During the model training process, the label image is a multi-channel image with the same width and height as the input image $X_{\text{in}} \in R^{H \times W \times 3}$, and the number of channels corresponds to the number of classes. In this research, the number of classes is 2, with one class representing the background and the other representing the shadow area. Therefore, the model output and label $(X_{\text{out}}, X_{\text{label}}) \in R^{H \times W \times 2}$, and the loss between them is calculated as follows:

$$\begin{cases} l_s(x, y) = \{l_1, \dots, l_N\}^T \\ l_n = -w_{y_n} \log \frac{\exp(x_{n,y_n})}{\sum_{c=1}^C \exp(x_{n,c})} \end{cases} \quad (7)$$

where x is the input value, y is the label value, w represents the weight, C represents the number of classes, and N represents the total number of batches per iteration. OSSNet uses cross-entropy loss as the classification loss for binary classification. Due to the small area occupied by the shadow area, which is generally less than 5% of the total image, we establish Focal-Loss and Dice-loss on binary classification to balance the pixel difference between positive and negative samples. The calculation

formular are as follows:

$$\begin{cases} Fl = -[(1 - pt)^\gamma] \log(pt) \\ Dl = 1 - \frac{2|x \cap y| + \xi}{|x| + |y| + \xi} \end{cases} \quad (8)$$

where pt is $\exp^{-\gamma s}$, and ξ is a constant of $1e-5$ to prevent the denominator from being zero. Two auxiliary training heads were added in the model. Therefore, the total loss of the model can be expressed as follows:

$$\mathcal{L}_{\text{oss}} = \sum_{i=1}^3 (Fl_i + \text{Aux}_i.Fl_i) + (Dl_1 + \text{Aux}_1.Dl_1) \quad (9)$$

where $\text{Aux}_i.Fl_i$ represents the Focal-Loss for the auxiliary training head, and $\text{Aux}_1.Dl_1$ represents the Dice-loss for the auxiliary training head. By integrating multi-level features through multiple auxiliary training heads and programmable gradient branches, it can enhance the training stability with severely imbalanced positive and negative sample ratios, and improve the model's overall perception capabilities. By adding Dice-loss to the final output layer, the model can be controlled to reduce the probability of false detections around the pixels in the shadow area, which can improve segmentation precision.

2.3. Shadow recovery model

The shadow recovery model requires a larger receptive field, even a global receptive field. The recovery process not only aims to increase the average brightness of the shadow area but also enhances the smoothness of pixel gradients by utilizing external features from other regions in the image. Early retinal diseases often manifest as small hole regions in images, which can easily be confused with shadow parts, thereby affecting diagnosis [32]. Therefore, the model needs to possess the ability to perceive global structures, communicate with surrounding pixel information, and interact with information at different scales. Research based on the Vision Transformer (ViT) model architecture has demonstrated that Q (query vectors), K (key vectors), and V (value vectors) in image encoding can effectively acquire global perception capabilities [33,34]. Due to the high computational complexity of ViT and its lack of multi-scale interactive information, this work proposes the OSRNet model by utilizing window multi-head self-attention (W-MSA) and shifted window multi-head self-attention (SW-MSA), to achieve feature interaction at different scales (as shown in Fig. 4). Then, the model utilizes auxiliary training heads achieving efficient reverse gradient propagation through the splitting and recombination of information across different dimensions of image encoding vectors. During model training, auxiliary training heads are once again introduced to continuously correct the gradient descent direction of the prediction head.

Based on the concept of ShadowFormer [22], the model framework adopts an encoder-decoder structure. The input image to the encoder part is a combination of the original image and a binary mask aiming to constrain the model. The input image undergoes dimensionality transformation in the first layer, encoding the image into feature vectors. The second and third layers consist of a attention module for more information interaction, followed by downsample to increase the model's receptive field. The channel attention module incorporates channel normalization, channel attention mechanisms, dropout strategies, and residual connections. These elements help prevent overfitting while ensuring that information is not lost. The fourth, fifth and sixth layers consist of interaction modules based on W-MSA and SW-MSA for shadow and non-shadow area. These layers further enhance the model's ability to extract features from different scale feature maps and global information by intra-window self-attention and inter-window cross-attention. The decoder is the reverse upsample process of the encoder, with information transfer achieved through feature map matching. Finally, the output encoding layer converts the image feature vectors back into a two-dimensional image. The encoder-decoder structure also involves a

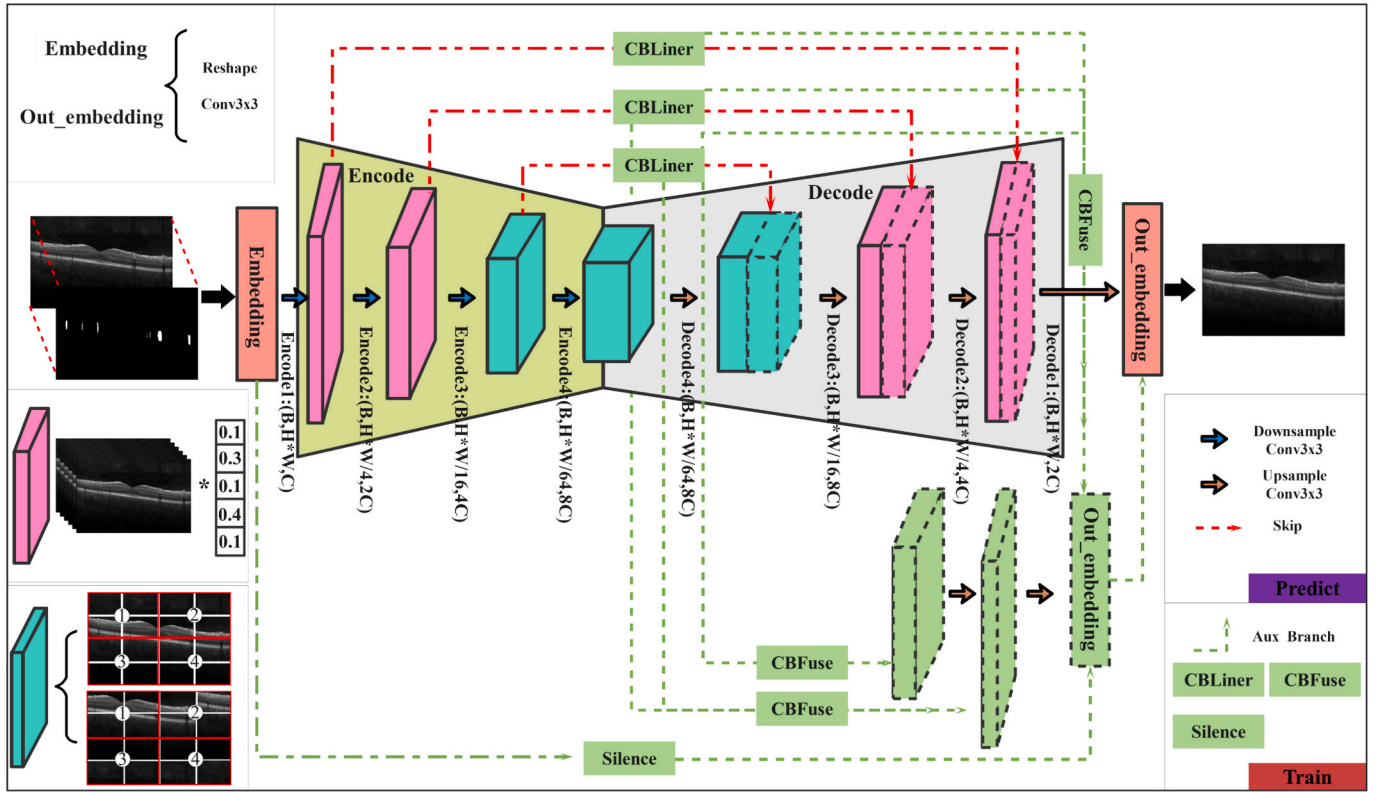


Fig. 4. Shadow recovery model. The CBLiner, CBFuse, and Silence layers are maintaining the same structure as the shadow segmentation model.

process of forward and backward propagation of feature maps. Therefore, we use the auxiliary training heads to achieve multi-level feature interconnection. This mechanism fully leverages the encoder and decoder's ability to understand global information in the image during the training phase. By utilizing auxiliary training heads and branches, the model can learn global and local information efficiently, and enhance generalization capabilities.

During the model training process, the label image is a multi-channel image with the same width and height as the input image. Therefore, the model output and label $\langle X_{out}, X_{label} \rangle \in R^{H \times W \times 3}$, and the loss between them is calculated as follows:

$$\mathcal{L}_s(x, y) = \sum_{n=1}^N \frac{\sqrt{(x_n - y_n)^2 + \xi}}{N} \quad (10)$$

where N represents the total number of batches per iteration, where x_n and y_n denote the model's output and label values for the N th batch. Due to the inclusion of an auxiliary training head, the total loss calculation formula for the model is as follows:

$$\mathcal{L}_{OSR}(x, y) = \sum_{k=1}^K [\mathcal{L}_s(x, y) + Aux_ \mathcal{L}_s(x, y)] \quad (11)$$

where K represents the number of auxiliary training heads, here $K = 1$. $Aux_ \mathcal{L}_s(x, y)$ denotes the model loss corresponding to the auxiliary training head.

2.4. Train strategy

2.4.1. Segmentation model train strategy

This research trains a blood vessel shadow segmentation model on 1440 ocular images across four categories to enhance its robustness for B-scan images. The dataset is partitioned into a training set and a validation set at an approximate ratio of 7:3. Additionally, negative background images containing no shadow areas are explicitly incorporated

into the dataset to improve the model's generalization capability and mitigate false positives. Once trained, this model is deployed to predict and provide precise shadow masks for the downstream shadow recovery task on B-scan images.

2.4.2. Shadow recovery model train strategy

The recovery of B-scan shadow areas is accomplished using a shadow recovery model optimized via a two-step training strategy: generic pretraining and domain-specific finetuning. Both stages require shadow masks to guide the illumination restoration process.

Step 1 (Generic Shadow Recovery Pretraining): The recovery model is pretrained from scratch on 1870 real-world natural image pairs from the public ISTD dataset. This stage utilizes the paired shadow masks provided by the dataset, enabling the network to learn fundamental pixel interactions and illumination compensation mechanisms between shadow and non-shadow regions in the real world. Step 2 (Domain-Specific Finetuning on Simulated Shadows): Based on the observation of actual B-scan images, shadow regions typically manifest as thin, vertical stripes with irregular, curved boundaries rather than straight rectangular shapes due to the interference of adjacent tissue structures. Therefore, an artificial shadow simulation method is proposed to construct thin stripes with irregular curved boundaries according to the following formula. The model inherits the pretrained weights from Step 1 and is finetuned on this synthesized paired B-scan dataset using the generated simulation masks, thereby successfully transferring the generalized restoration capability from natural scenes to achieve precise shadow removal in medical B-scan images.

$$\begin{cases} S_{contour} = \{(x_1, y_1), \dots, (x_N, y_N)\} \\ x_i = cx + [rx + A \sin(f\theta_i)] \sin(\theta_i) \\ y_i = cy + [ry + A \cos(f\theta_i)] \cos(\theta_i) \\ \theta_i = \frac{2\pi i}{N}, i \in 0, N+1 \end{cases} \quad (12)$$

where S_{contur} is the set of generated contour points, (x_i, y_i) is a specific generated contour point, A is a random number controlling the amplitude of the contour, f is a random number controlling the rate of contour variation, (cx, cy) is the randomly initialized center point, (rx, ry) is the fluctuation radius. The comparison of the generated region contours is shown in Fig. 5. Then, different brightness values are assigned to the generated regions. Gaussian blur is applied to the filled regions to simulate the unevenness of brightness, as shown in the following formula:

$$\begin{cases} S_{area} = G_{kernel} \odot X_{mask} \\ X_{in} = (0 \ 1) S_{area} \cdot X_{Original} \end{cases} \quad (13)$$

where S_{area} denotes the unevenly shaded filled area, G_{kernel} denotes the Gaussian kernel, X_{mask} denotes the uniformly shaded area, $X_{Original}$ denotes the original image data. We train the model to recover the structure of B-Scan images with the artificial shadow.

2.5. Train setup

The experiments presented in this study were conducted using a server equipped with two A100 (40G) GPUs, running a Linux operating system. The implementation was performed in Python(3.8.16), utilizing the PyTorch(1.10.1) deep learning framework. Details regarding the training hyperparameters for the two-stage model can be found in Table 1. To minimize the impact of random initialization on the model's performance, each experiment was repeated multiple times with a fixed random seed of 1234, and the results were averaged to provide a more robust evaluation. In this research, OSSNet is trained for 200 epochs with input images of size 512×512 and each epoch involves a batch size of 16. The initial learning rate (Init_lr) is set to $1e-4$, and the minimum learning rate (Min_lr) is calculated as the Init_lr multiplied by 0.01. During the training process, Adam is selected as the model optimizer, with a momentum of 0.9 and a weight decay of 0.01. The learning rate decay schedule employed is cosine annealing. The training duration for OSRNet is 400 epochs, with input images sized at 512×512 and each epoch processes a batch size of 16. During the training process, Adamw is chosen as the model optimizer, with a weight decay of 0.02 and the initial learning rate is set to $2e-4$.

2.6. Dataset

The ocular shadow area segmentation dataset constructed in this research is based on the OCT2017 dataset. We selected 1440 high-quality images as shadow segmentation training and testing samples. Then, we employed the natural shadow images from the ISTD dataset for transfer learning. Meanwhile, we collected 400 phantom eye images to construct a novel shadow dataset incorporating realistic light intensity

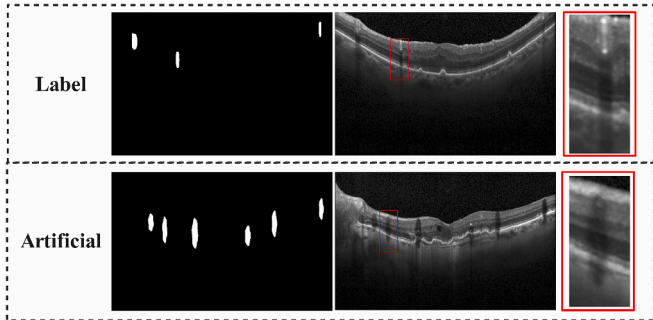


Fig. 5. Comparison of artificially generated shadow regions with actual shadow regions. The first row of images in the figure shows the actual shadow regions in B-Scan images, and the second row of images shows the artificial shadow regions.

Table 1

Hyperparameter settings for model training.

Model Name	Hyperparameter	Value
OSSNet	Epoch	200
	Batch_size	16
	Init_lr	$1e-4$
	Min_lr	$Init_lr * 0.01$
	Optimizer	Adam
	Momentum	0.9
	Weight_decay	0.01
	Lr_decay_type	cos
	Epoch	400
	Batch_size	16
OSRNet	Init_lr	$2e-4$
	Weight_decay	0.02
	Optimizer	Adamw

attenuation along with corresponding ground-truth labels. Finally, we evaluated and analyzed the shadow recovery performance of the model across three datasets, the details of which are presented in Table 2.

2.7. Compared method

In the task of B-Scan shadow detection, there is no research about establishing the model for shadow segmentation. Therefore, this research selects the most common and latest models in the medical field for comparison: Resnet-50, U-Net [35] (Baseline model), DeeplabV3+ [36], U-Net++ [37], LightReSeg [38], DA_Transunet [39], RelayNet [40], EMVNet [41], OSMGUnet [42], MK_Unet [43].

For the task of shadow recovery in B-Scan images, the publicly available method currently is the adaptive compensation algorithm based on pixel statistics. Therefore, this paper selects methods based on advanced generative models and pixel statistics for comparison: Compensation [44], OCT-GAN [16], MISF [45], Shadowformer [22] (Baseline model).

3. Results and discussion

This section presents the experimental results of the two-stage model for segmentation and recovery tasks. We further elaborates on the factors that affect model performance, as well as a general approach for shadow recovery tasks in B-Scan images.

3.1. Shadow area segmentation

This research trains a segmentation model based on the backbone feature extraction network ResNet50 and compares it with commonly used models in the medical field. And then, we reveal the factors that influence the segmentation accuracy of ocular B-Scan images. The performance parameters of the various models are compared in Table 3.

In Table 3, the model with the best performance is indicated in bold font. mIoU (Mean Intersection over Union) can comprehensively consider the pixel-level prediction accuracy of the model. It provides a fair evaluation of the model's performance in handling targets of different categories and sizes. Its calculation formula is as follows:

$$mIoU = \frac{TP}{TP + FP + FN} \quad (14)$$

Table 2

Details of the datasets.

Name	Number
OCT2017	1440
EYE FOUNDS	1113
OCT_DL	2064
ISTD	1870
Phantom	400

Table 3

Comparison of shadow segmentation models.

Model	mIoU↑	MP↑	MR↑	mDice↑
ResNet-50	78.04	86.67	85.49	85.42
U-Net	77.91	86.54	85.18	85.85
DeepLabV3+	77.03	87.13	83.32	85.12
U-Net++	78.16	86.00	86.11	86.05
DA_Transunet	77.75	86.23	85.22	85.72
RelayNet	78.12	86.31	85.73	86.02
EMVNet	67.42	68.87	91.80	75.95
OSMGUNet	77.69	86.03	85.31	85.67
LightReSeg	77.76	85.28	86.19	85.73
MK_Unet	77.69	85.68	85.64	85.66
OSSNet	79.07	87.56	86.23	86.69

where *TP* represents True Positives, *FP* represents False Positives, and *FN* represents False Negatives. The model is evaluated by calculating the comprehensive recognition ability of positive and negative samples. MP (Mean Precision) reflects the model's ability to recognize positive samples. MR (Mean Recall) reflects the model's coverage of positive samples. mDice (Mean Dice coefficient) measures the overlap between the predicted results and the ground truth labels.

According to the data in Table 3, the proposed method achieves the highest mIoU of 79.07% and the highest mDice of 86.69%. Compared with the baseline ResNet-50 model, which yields an mIoU of 78.04%, an MP of 86.67%, an MR of 85.49%, and an mDice of 85.42%, the proposed framework exhibits a consistent performance improvement across these evaluated metrics. Furthermore, the integration of the multi-scale

training strategy allows the network to learn feature representations across varying receptive fields. This explicit scale-invariant learning alleviates the performance degradation caused by the substantial size variations of retinal vascular shadows, thereby adapting to both large localized trunk shadows and fine-grained peripheral shadow zones. This comparison indicates that the model has a higher recognition accuracy for positive samples. Through comparisons between the other models, it is observed that the model possesses reliable shadow tracking capabilities. This model demonstrates accurate recognition in shadow segmentation tasks for different types of ocular B-Scan images (as shown in Fig. 6).

By comparing the segmentation results depicted in Fig. 6, it is evident that the method proposed achieves a high accuracy in delineating shadow areas. This observation aligns with the performance indicators presented in Table 3. During the segmentation process, our proposed method demonstrates improved accuracy in classifying pixels located at edges and exhibits enhanced discriminative capability when distinguishing structural areas that resemble shadows. However, a qualitative analysis reveals that all models encounter errors in misidentification and non-identification during the segmentation task. The causes of these errors can be attributed to the following three factors: Small proportion of shadow areas: For B-Scan images, the pixel occupied by shadow areas constitutes approximately 1% of the total image. When averaged across individual shadow areas, the pixel proportion becomes even smaller. This presents a challenge for small object recognition, as models can readily fall into local minima during the recognition process. Notably, when focal loss and Dice loss strategies

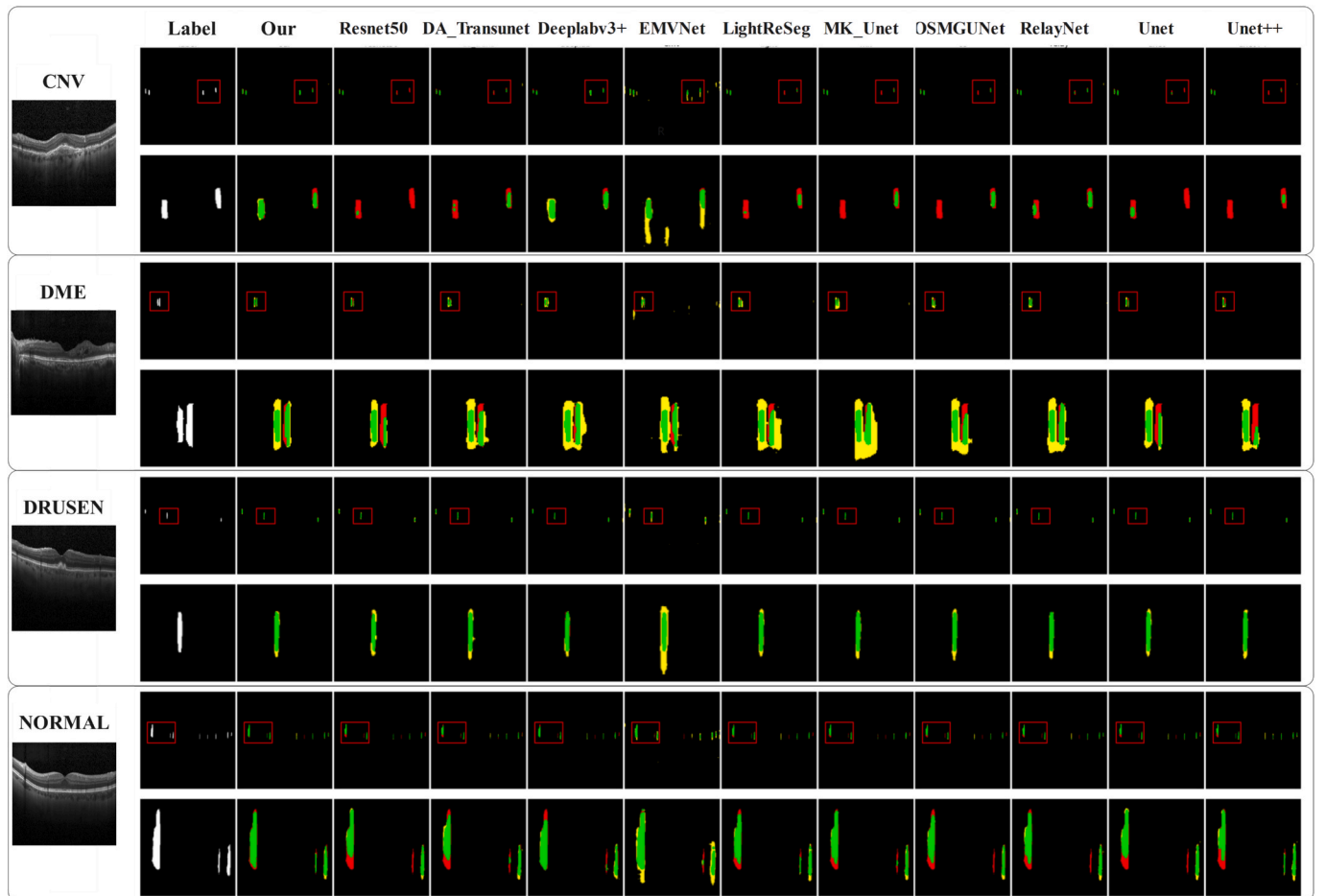


Fig. 6. OSSNet segments shadow areas of different B-Scan images. The green regions in the figure represent the intersection of the label areas and the model's predicted results. The red areas indicate the label areas not successfully predicted by the model. The yellow regions represent pixels that are not present in the label areas but are predicted by the model.

were not incorporated into our training process, the model resorted to predicting the entire image as background to maintain accuracy. It leads to erroneous convergence and model failure due to the inability of the gradient descent strategy to constrain the model effectively. Confusion between retinal structure and shadows: Retinal B-Scan images are grayscale representations that solely depict the reflection intensity of tissue layers after Fourier transform of interference spectra. Consequently, areas with weak reflection intensity, such as retinal fluid accumulation, exhibit features akin to vessel shadows, thereby inducing segmentation errors in the model. Light intensity attenuation: The principle of OCT imaging involves the reconstruction of images through mathematical operations based on the reflection of incident light. However, the intensity of the incident light diminishes due to absorption by tissues. This results in the obscuring of features at the boundaries of shadow regions, rendering them difficult to segment accurately.

In order to address the aforementioned issues, we propose the strategies of Focal Loss and Dice Loss to mitigate the challenges posed by the small pixel proportion of shadow areas. By incorporating auxiliary branch, we achieve skip connections for full-size feature maps, as illustrated in Fig. 3. This can take the useful information from different scale feature maps to the detection head. Finally, we enhance the segmentation performance by establishing three-level auxiliary detection heads. These auxiliary detection heads leverage combined multi-scale feature maps to enable gradient mining across different scales, thereby improving the distinction between tissue structures and shadow areas and enhancing the ability to extract boundary features.

Lastly, through the analysis of segmentation results from 1440 ocular B-Scan images, this research demonstrates that the auxiliary training strategy constructed based on auxiliary detection heads can enhance the efficiency of gradient backpropagation and improve the robustness of the model.

3.2. Shadow area recovery

In this research, we train a recovery model based on the Encode-Decode architecture and compare it with existing methods. This comparison aims to reveal the core challenges in the task of B-Scan image recovery. The performance parameters are compared in Table 4.

In Table 4, MBD_{area} (Mean Brightness Difference within the Minimum Bounding Box) represents the average brightness deviation calculated only within the shadow recovery area (as shown in Fig. 7), which is a direct expression of the brightness improvement in the area. $SSIM_{total}$ (Structure Similarity Index Measure for the Total Image) indicates the structural similarity between the output image of the model and the original image, reflecting the model's ability to preserve the original image's structure and brightness invariance. $SSIM_{area}$ represents the Sobel edge structural similarity calculated exclusively within the localized minimum bounding boxes. Crucially, the spatial position and boundaries of each bounding box are dynamically determined by the shadow masks predicted in Stage 1, which strictly constrains the evaluation to the affected regions and maximizes the suppression of background pixel dilution. Sobel edges are extracted because they are invariant to localized image brightness variations induced by illumination compensation, thereby capturing the intrinsic structural edge features. Consequently, this target-specific metric eliminates non-shadow background interference and provides the most direct assessment of the

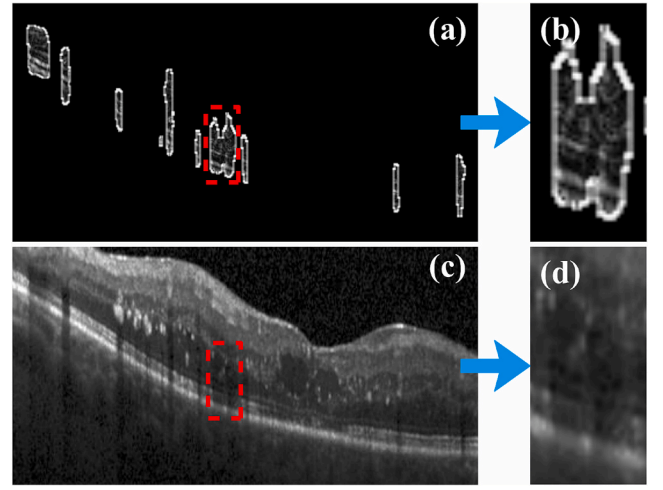


Fig. 7. Minimum Bounding Box of the Shadow Area. (a) Shows the Sobel edge image of the shadow area. The Sobel edge structural similarity within the red rectangular box in (b) is calculated as $SSIM_{area}$. (c) is the original image of B-Scan, and the $RMSE_{area}$ of pixel brightness is obtained by comparing it with the bounding box area (d) is the original image of (b).

morphological fidelity and continuity of anatomical layer structures between the recovered regions and the original tissue matrix. $RMSE_{area}$ represents the brightness error between each pixel in the recovery shadow area within the minimum bounding box and the corresponding pixel in the original image, directly reflecting the model's shadow recovery capability. In Table 4, “—” indicates that the value is unavailable or meaningless. Among the aforementioned metrics, SSIM is the parameter that comprehensively assesses the model's recovery performance, and its calculation formula is as follows:

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (15)$$

where μ_x and μ_y denote the mean values of images x and y , which reflect the brightness information of the images. σ_x^2 and σ_y^2 represent the variances of images x and y , which indicate the contrast information. σ_{xy} denotes the covariance between images x and y , which reflects the degree of structural similarity between the two images.

Based on the quantitative analysis presented in Table 4, several key insights can be observed regarding the trade-off between targeted enhancement and structural distortion. Our method achieves competitive values on both structural and brightness restoration metrics, validating its design. A further comparison with traditional compensation methods shows that by leveraging the segmentation mask from our model, we can precisely localize shadow regions. This strategy effectively minimizes the unintended alterations in non-shadow areas, thereby preserving a high degree of structural similarity. This localized advantage is particularly evident when compared against GAN-based generative approaches, such as OCTGAN. Because GAN-based methods typically apply global reconstructions, they tend to yield a lower level of structural similarity, indicating that they frequently introduce structurally inconsistent features in non-shadow regions and fail to preserve critical anatomical edges. Moreover, in comparisons with other mask-based methods such as MISF and ShadowFormer, our approach exhibits more localized and structurally constrained capabilities for shadow restoration. While ShadowFormer leverages the global context of a Transformer mechanism to maintain a relatively high structural similarity (an $SSIM_{total}$ of 0.875), its performance remains constrained; without multi-scale loss constraints and effective domain adaptation strategies, Transformer-based restorations are highly susceptible to the interference of surrounding high-intensity vascular pixels and inherent

Table 4
Comparison of Shadow Recovery Models.

Model	$MBD_{area} \uparrow$	$SSIM_{total} \uparrow$	$SSIM_{area} \uparrow$	$RMSE_{area} \uparrow$
Compensation	—	0.620	—	—
OCT-GAN	—	0.470	—	—
MISF	—	0.852	0.821	—
Shadowformer	—	0.875	0.877	—
OSRNet	21.13	0.999	0.953	27.06

speckle noise, which degrades their restoration fidelity. In contrast, our framework maintains strict structural integrity across both the original image and the restored regions, effectively suppressing ungrounded feature hallucination. This is rigorously evidenced by the baseline $SSIM_{total}$ of 0.953.

As evidenced by the comparative results in Fig. 8, the proposed method excels in preserving the global structural fidelity of the image, thereby achieving a $SSIM_{total}$ of 0.99. Furthermore, within the restored shadow regions, our approach effectively maintains the natural intensity contrast between retinal layers, moving beyond a simple uniform increase in pixel values. Notably, it demonstrates superior capability in handling structural edges, successfully restoring the shaded areas while preserving edge sharpness. This results in significantly improved visibility of the underlying layer structures beneath the blood vessels. This finding underscores the critical importance of precise vessel region localization for successful shadow restoration in retinal images. To further address the critical distinction between physical enhancement and generative hallucination—a paramount concern in medical imaging safety—we conduct a granular comparative analysis against existing GAN and Transformer architectures. Specifically, while OCTGAN can effectively recover pixel intensity within vessel shadows, its visualized Sobel edge profiles exhibit a pronounced discrepancy compared to the authentic ground-truth structures, indicating a tendency toward generating visually plausible yet anatomically incorrect hallucinations. Meanwhile, the Transformer-based ShadowFormer tends to introduce chromatic-like boundary artifacts at shadow edges, and as highlighted by the heatmap in the red dashed box, its internal layer attention is overly concentrated on isolated high-brightness edges. In contrast, by introducing a multi-scale auxiliary gradient framework, our method adaptively balances the contributions of high-contrast boundaries and their surrounding contextual pixels, effectively eliminating non-existent artifacts and structural hallucinations to deliver superior, anatomically faithful restoration.

3.3. Effectiveness verification of phantom eye

To validate the efficacy of the proposed method in preserving structural consistency within shadow regions, a labeled paired dataset was constructed under measured optical attenuation using a phantom eye. To mitigate the confounding effects of optical dispersion and optical path length on the structural integrity of the pairs, a 5um thick attenuation film was employed to locally attenuate the B-scan measurement beam, thereby simulating the localized light intensity attenuation induced by retinal blood vessels. On this dataset, a comprehensive quantitative and qualitative evaluation of existing methods was conducted. The quantitative comparison results are summarized in Table 5, where MBD_{gt} , $RMSE_{gt}$, and $SSIM_{total_gt}$ denote the MBD, RMSE, and SSIM calculated over the entire image relative to the ground truth (GT), while $SSIM_{area_gt}$ represents the SSIM specifically computed within the shadowed area.

As indicated by the data in Table 5, due to the temporal interval between the acquisition of the ground truth and the shadowed images, the variations in speckle distribution (speckle decorrelation) inevitably affected the calculation of structural similarity, leading to an overall decrease in the metric values. Nevertheless, the quantitative results in Table 5 demonstrate that our method achieves the state-of-the-art performance among all competing methods. It effectively recovers shadowed regions while superiorly preserving the structural consistency of non-shadowed areas; furthermore, the structures within the restored

Table 5

Comparison of Shadow Recovery Models on Phantom Eye.

Model	$MBD_{gt} \downarrow$	$SSIM_{total_gt} \uparrow$	$SSIM_{area_gt} \uparrow$	$RMSE_{gt} \downarrow$
Compensation	16.55	0.611	0.348	40.07
OCT-GAN	13.20	0.499	0.318	43.58
MISF	10.08	0.817	0.476	37.74
Shadowformer	9.830	0.821	0.694	31.13
OSRNet	8.596	0.932	0.818	28.69

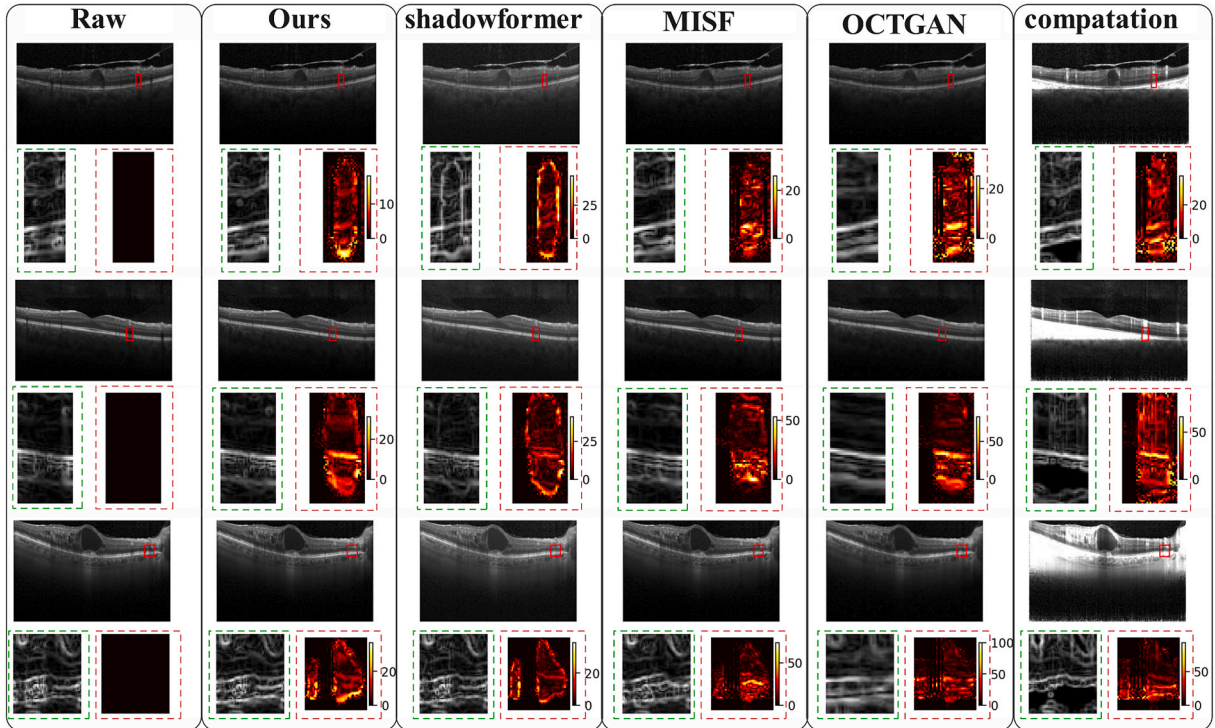


Fig. 8. Comparison of the shadow restoration results. The green and red dashed boxes highlight the Sobel edge map of the shadow region and the difference map between the restored and original shadow areas, respectively. In the difference map, higher intensity signifies a larger error.

regions closely approximate the ground-truth images. Consequently, our method yields the optimal $SSIM_{total_gt}$ of 0.932 and $SSIM_{area_gt}$ of 0.818. This outcome aligns consistently with the performance trends observed in real human eye shadow recovery, further substantiating the efficacy of the proposed method.

From a qualitative perspective, as illustrated in Fig. 9, the images restored by the proposed method are visually closer to the ground truth. Furthermore, the Sobel edge maps reveal that our method successfully preserves the edge sharpness and structural integrity of the scattering particles inside the phantom eye during the restoration process. In contrast, GAN-based methods exhibit poor preservation of edge structures within the shadow regions, introducing structural artifacts (false information). Both Shadowformer and MISF suffer from speckle-induced edge artifacts, as indicated by the green arrows. Although the Compensation method enhances the brightness in shadowed areas, it amplifies the contrast of high-intensity pixels and compresses the dynamic range of the image, ultimately leading to localized pixel over-exposure. Therefore, by constructing paired shadow and shadow-free images featuring realistic light intensity attenuation, we further validate the efficacy of the proposed method from both quantitative and qualitative perspectives.

3.4. Ablation study

To validate the individual contributions of the transfer learning strategy and the auxiliary training branch to the shadow restoration results, we conducted ablation studies by incrementally adding each component to the baseline model. The experimental results are presented in Table 6, where “Baseline” denotes the model trained solely on synthetic shadow data, “T” indicates the model incorporating the transfer learning strategy, and “Aux.” refers to the model enhanced with the auxiliary training head.

A comparison of the data in Table 6 indicates that training exclusively on a dataset with artificial shadows is insufficient to perfectly simulate the pixel distribution of vascular shadows, consequently leading to low global and regional structural similarity. After transfer training, the model gains a better understanding of the pixel distribution in shadow regions, thus achieving higher structural similarity. Finally, with the addition of the auxiliary training head, gradients are propagated more effectively during the training process, which further improves the accuracy of shadow restoration. The model utilizing transfer learning and the auxiliary training head showed the best performance on all three datasets. It achieved the best restoration of pixel values in the retinal shadow areas, with the greatest improvement in terms of both MBD and RMSE.

A qualitative analysis from the comparative results in Fig. 10 shows that the original baseline method processes the overall image pixels, which increases the pixel values in noisy areas and leads to a reduced signal-to-noise ratio. It also lacks processing of the contrast relationships between layers in the restored structural areas. After introducing transfer training, the model's performance improved significantly. It could capture the characteristics of shadow areas and effectively restore

Table 6
Ablation study results.

Dataset	Model	MBD _{area} ↑	SSIM _{total} ↑	SSIM _{area} ↑	RMSE _{area} ↑
OCT2017	Baseline	—	0.875	0.877	—
	Baseline + T	19.93	0.996	0.947	25.52
	Baseline + T	21.13	0.999	0.953	27.06
	+ Aux				
EYE	Baseline	—	0.837	0.920	—
	Baseline + T	34.16	0.995	0.941	37.89
	Baseline + T	36.73	0.997	0.944	40.03
	+ Aux				
OCT_DL	Baseline	—	0.954	0.961	—
	Baseline + T	26.43	0.991	0.974	33.37
	Baseline + T	28.74	0.997	0.982	36.15
	+ Aux				

them when processing global pixels. However, due to the single detection head, the model gradients were not fully utilized, resulting in weaker capabilities for restoring inter-layer structures and minute intra-layer structures. When the auxiliary training head was added, the restoration of these structures improved. This enhancement allows the model to restore obvious structural locations while simultaneously maintaining an increase in contrast within the layers, thus strengthening the comparative ability of the results.

3.5. Recovery area segmentation

To verify the impact of shadow recovery on the accuracy of retinal layer segmentation measurements, this research validates the experimental results using a graph-theory-based segmentation method [46] commonly employed in the medical field. By comparing the RPE-Choroid layer segmentation fitting curves before and after recovery in Fig. 11, we observe that the segmentation algorithm is prone to deviations from the layer boundaries due to the influence of shadows when the shadow area spans a large area. When the brightness in the shadow area is very low, the image segmentation effect is also significantly affected. Notably, the impact is more pronounced when the shadow area is located at the edge of the image on both sides. These issues are caused by the segmentation algorithm, which identifies pixels belonging to the same boundary edge and detects grayscale changes between adjacent pixels in the image to achieve segmentation. This confirms that the recovered shadow area contains more usable information, enabling more precise measurement indicators.

4. Conclusion

In summary, through a comparison of various methods, it can be concluded that traditional approaches lack robustness in handling shadow regions, failing to maintain the consistency of the overall structure and content of the image. This limitation affects the accuracy of subsequent B-Scan image detection metrics. On the other hand, methods based on deep learning neural networks exhibit stronger robustness and generalization capabilities, minimizing the impact on the

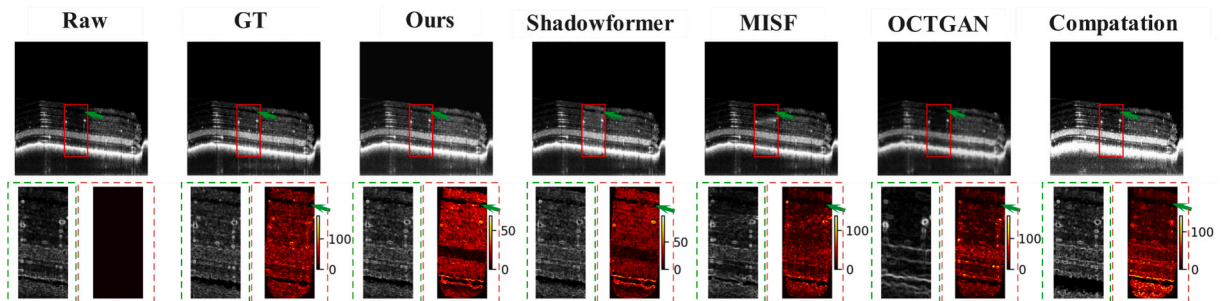


Fig. 9. Comparison of shadow recovery performance on phantom eye images. The green and red dashed boxes share the same meaning as in Fig. 8.

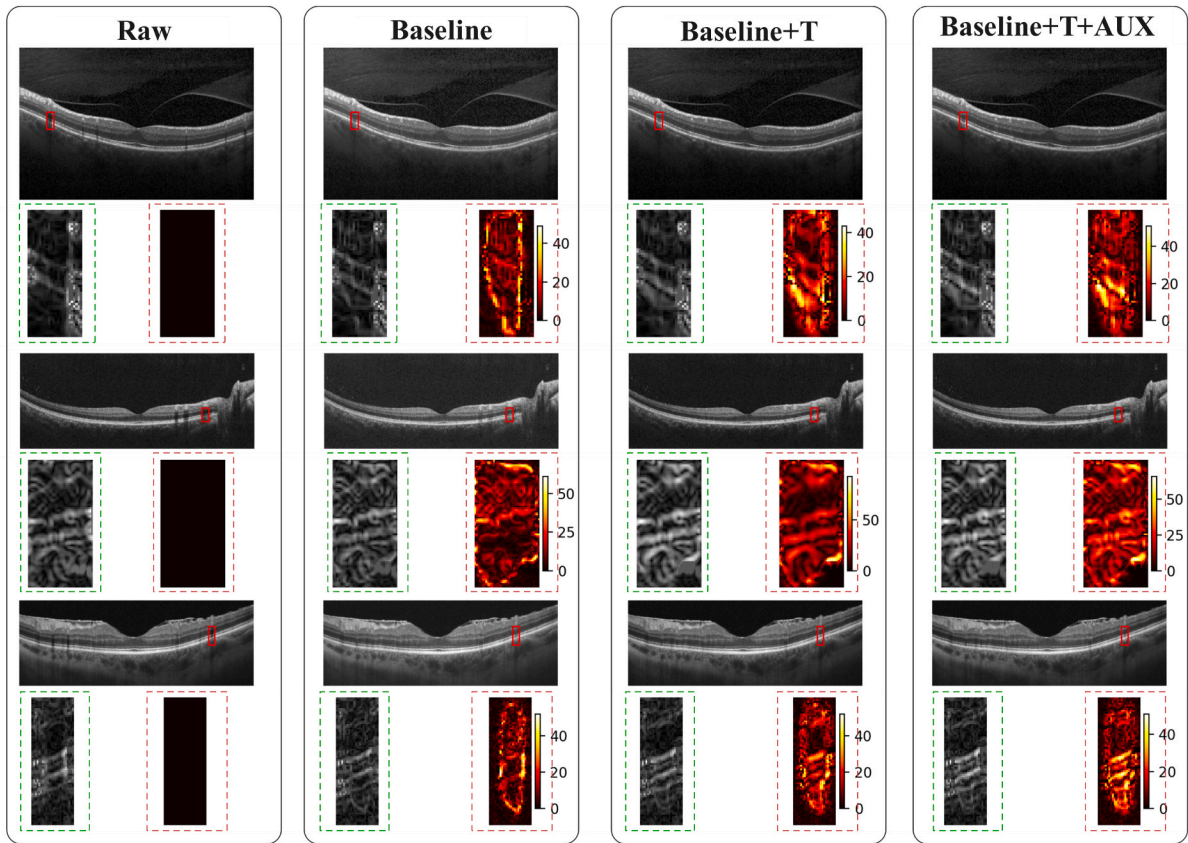


Fig. 10. Comparison of different parts for shadow restoration. The green and red dashed boxes share the same meaning as in Fig. 8.

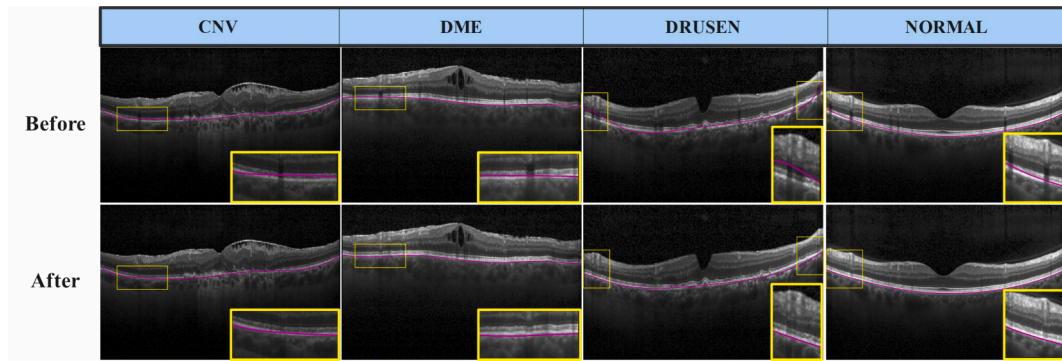


Fig. 11. Comparison of RPE-choroid layer segmentation before and after recovery.

original image to the greatest extent. However, to date, methods for shadow area recovery in OCT B-Scan images remain in the exploratory stage. Previous studies have generally employed generative techniques to directly redraw the shadowed areas, just like GAN. While this approach makes the image appear more natural, it results in the loss of structural pixel information behind the vascular shadows, rendering it ineffective for the recovery of subtle lesions that may exist within the shadowed areas.

Therefore, this paper proposes a two-stage model comprising segmentation and recovery processes to accurately locate shadow areas. Building upon this, the model leverages the global receptive field capabilities of the transformer mechanism to enable information exchange between shadow pixels and surrounding pixels. The segmentation phase enhances a binarized mask, which is then used for precise recovery of the shadow regions, ensuring minimal alteration to the original image structure while enhancing the pixel brightness of the shadowed areas.

Notably, the auxiliary training strategy proposed in this study improves the model's accuracy by providing multi-head gradient information during the training phase, enabling highly competitive segmentation accuracy and structural similarity across the evaluations in Tables 3 and 4. During the inference phase, a pruning strategy is employed to eliminate the auxiliary branches, maintaining the same number of parameters as the baseline model while improving inference speed. Our work provides a novel approach for future research on B-Scan shadow recovery, focusing on recovering the image by leveraging the pixel structure of the shadowed areas rather than simply redrawing them.

In future work, we will explore models with higher accuracy and methods to simulate vascular shadows better, aiming to achieve improved restoration results. Admittedly, our current shadow simulation approach has certain limitations in capturing the full complexity of true OCT attenuation. To address this, we will incorporate the attenuation design of the ocular and optical systems and simulate the

absorption effects of blood vessels, thereby further exploring shadow simulation methods that better align with fundus vasculature. This will enhance the quality of B-Scan images and provide greater convenience for clinical applications.

CRedit authorship contribution statement

Yukuan Wang: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Funding acquisition, Formal analysis, Data curation. **Wei Yi:** Visualization, Validation. **Hua Wei:** Writing – review & editing, Visualization. **Yi Wan:** Writing – review & editing, Visualization. **Zhengkai Yao:** Writing – review & editing, Visualization. **Xiaoqi Lu:** Writing – review & editing, Visualization. **Peng Sheng:** Writing – review & editing, Visualization. **Xianchun Jiang:** Writing – review & editing, Visualization. **Weiye Song:** Visualization, Validation, Formal analysis.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yukuan Wang reports financial support was provided by the National Natural Science Foundation of China (62205181) and the Natural Science Foundation of Shandong Province (ZR2022QF017). If there are other authors, they 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

Data will be made available on request.

References

- [1] D. Huang, E.A. Swanson, C.P. Lin, J.S. Schuman, W.G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C.A. Puliafito, J.G. Fujimoto, Opt. Coherence Tomogr. 254 (1991) 1178–1181.
- [2] F. Rossant, I. Bloch, I. Ghorbel, M.J.P.R. Pâques, Parallel double Snakes. Application to the segmentation of retinal layers in 2D-OCT for pathological subjects, Pattern Recogn. 48 (2015) 3857–3870.
- [3] A. Geetha, M. Carmel Sobia, D. Santhi, A. Ahilan, DEEP GD: deep learning based snapshot ensemble CNN with EfficientNet for glaucoma detection, Biomed. Signal Process. Control 100 (2025) 106989.
- [4] R. Wang, G. Zhang, X. Xi, L. Xu, X. Nie, J. Nie, X. Meng, Y. Zhang, X. Chen, Y. Yin, Discriminative atoms embedding relation dual network for classification of choroidal neovascularization in OCT images, Pattern Recogn. 156 (2024) 110757.
- [5] X. Li, K. Huang, Y. Zhang, S. Yuan, S. Niu, Q. Chen, MT-DENet: Prediction of post-therapy OCT images in diabetic macular edema by multi-temporal disease evolution network, Biomed. Signal Process. Control 112 (2026) 108721.
- [6] Ş. Kaya, M.K. Kaya, Assessment of RNFL and macular changes in the eye related to multiple substance use using OCT, Psychiatr. Res. Neuroimag. 345 (2024) 111889.
- [7] C.K. Leung, N. Choi, R.N. Weinreb, S. Liu, C. Ye, L. Liu, G.W. Lai, J. Lau, D.S. Lam, Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: pattern of RNFL defects in glaucoma, Ophthalmology 117 (2010) 2337–2344.
- [8] M.J. Girard, N.G. Strouthidis, C.R. Ethier, J.M. Mari, Shadow removal and contrast enhancement in optical coherence tomography images of the human optic nerve head, Invest. Ophthalmol. Vis. Sci. 52 (2011) 7738–7748.
- [9] C.d. Vente, B.v. Ginneken, C.B. Hoyng, C.C.W. Klaver, C.I.J.M.i.a. Sanchez, Uncertainty-aware multiple-instance learning for reliable classification: Application to optical coherence tomography, 97 (2023) 103259.
- [10] T.J.H. Krause, T.R. Allen, J.M. Fraser, Self-witnessing coherent imaging for artifact removal and noise filtering, Opt. Lasers Eng. 151 (2022) 106936.
- [11] T. Fabritius, S. Makita, Y. Hong, R. Myllylä, Y. Yasuno, Automated retinal shadow compensation of optical coherence tomography images, J. Biomed. Opt. 14 (2009) 010503.
- [12] D.M.M. George, H. Nandakumar, S. Srivastava, Enhancement of dynamic range and complete elimination of self-interference artifacts in a spectral domain OCT: potential for high performance at much lowered-cost, Results Opt. 5 (2021) 100144.
- [13] X. Chen, S. Zheng, L. Peng, Q. Zhong, L. He, A novel method based on shifted rank-1 reconstruction for removing EMG artifacts in ECG signals, Biomed. Signal Process. Control 85 (2023) 104967.
- [14] H. Hallez, M. De Vos, B. Vanrumste, P. Van Hese, S. Assecondi, K. Van Laere, P. Dupont, W. Van Paesschen, S. Van Huffel, I. Lemahieu, Removing muscle and eye artifacts using blind source separation techniques in ictal EEG source imaging, Clin. Neurophysiol. 120 (2009) 1262–1272.
- [15] Q. Jiao, Z. Xu, S. Wu, H.-S. Wong, DA-GAN: dual-attention generative adversarial networks for real-world exquisite makeup transfer, Pattern Recogn. 158 (2025) 111049.
- [16] H. Cheong, S. Krishna Devalla, T. Chuangsuwanich, T.A. Tun, X. Wang, T. Aung, L. Schmetterer, M.L. Buist, C. Boote, A.H. Thiéry, M.J.A. Girard, OCT-GAN: single step shadow and noise removal from optical coherence tomography images of the human optic nerve head, Biomed. Opt. Exp. 12 (2021) 1482–1498.
- [17] M.H. Akpinar, A. Sengur, O. Faust, L. Tong, F. Molinari, U.R. Acharya, Artificial intelligence in retinal screening using OCT images: a review of the last decade (2013–2023), Comput. Methods Programs Biomed. 254 (2024) 108253.
- [18] A. Camino, Y. Jia, J. Yu, J. Wang, L. Liu, D. Huang, Automated detection of shadow artifacts in optical coherence tomography angiography, Biomed. Opt. Exp. 10 (2019) 1514–1531.
- [19] H. Cheong, S.K. Devalla, T.H. Pham, L. Zhang, T.A. Tun, X. Wang, S. Perera, L. Schmetterer, T. Aung, C. Boote, A. Thiéry, M.J.A. Girard, DeshadowGAN: a deep learning approach to remove shadows from optical coherence tomography images, Transl. Vis. Sci. Technol. 9 (2020) 23.
- [20] V. Nguyen, T.F.Y. Vicente, M. Zhao, M. Hoai, D. Samaras, Shadow detection with conditional generative adversarial networks, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 4520–4528.
- [21] J. Wang, X. Li, J. Yang, Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 1788–1797.
- [22] L. Guo, S. Huang, D. Liu, H. Cheng, B.J.A. Wen, ShadowFormer: Global Context Helps Image Shadow Removal, abs/2302.01650 (2023).
- [23] D.M. Vigneault, W. Xie, C.Y. Ho, D.A. Bluemke, J.A. Noble, Ω-net (Omega-net): fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks, Med. Image Anal. 48 (2018) 95–106.
- [24] X. Quan, G. Hou, W. Yin, H.J.I.F. Zhang, A multi-modal and multi-stage fusion enhancement network for segmentation based on OCT and OCTA images, 113 (2024) 102594.
- [25] D. Oh, J. Moon, K. Park, W. Kim, S. Yoo, H. Lee, J.J.E.S.A. Yoo, GCN-assisted attention-guided UNet for automated retinal OCT segmentation 249 (2024) 123620.
- [26] X. He, Y. Wang, F. Poiesi, W. Song, Q. Xu, Z. Feng, Y.J.S.E.J. Wan, Exploiting Multi-Granularity Visual Features for Retinal Layer Segmentation in Healthy Human Eyes, (2023).
- [27] M. Luo, Z. Xu, Z. Ye, Z. Liang, H. Xiao, Y. Li, Z. Li, Y. Zhu, Y. He, Y. Zhuo, Deep learning for anterior segment OCT angiography automated denoising and vascular quantitative measurement, Biomed. Signal Process. Control 83 (2023) 104660.
- [28] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J. Liang, UNet++: redesigning skip connections to exploit multiscale features in image segmentation, IEEE Trans. Med. Imaging 39 (2020) 1856–1867.
- [29] A. Cazañas-Gordón, L.A.D.S. Cruz, Multiscale Attention Gated Network (MAGNet) for retinal layer and macular cystoid edema segmentation, IEEE Access 10 (2022) 85905–85917.
- [30] D. Ma, D. Lu, S. Chen, M. Heisler, S. Dabiri, S. Lee, H. Lee, G.W. Ding, M.V. Sarunic, M.F. Beg, LF-UNet - a novel anatomical-aware dual-branch cascaded deep neural network for segmentation of retinal layers and fluid from optical coherence tomography images, Comput. Med. Imag. Graph.: Off. J. Comput. Med. Imag. Soc. 94 (2021) 101988.
- [31] C.-Y. Wang, I.-H. Yeh, H.-Y. Mark Liao, YOLOv9: Learning What You Want to Learn using Programmable Gradient Information, Springer Nature, Switzerland, Cham, 2025, pp. 1–21.
- [32] R. Ray, S.S. Stinnett, G.J. Jaffe, Evaluation of image artifact produced by optical coherence tomography of retinal pathology, Am. J. Ophthalmol. 139 (2005) 18–29.
- [33] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N.J.A. Houlsby, An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, abs/2010.11929 (2020).
- [34] Q. Yan, Y. Ma, W. Wu, L. Mou, W. Huang, J. Cheng, Y.J.C.i.b. Zhao, medicine, Choroidal Layer Analysis in OCT images via Ambiguous Boundary-aware Attention 175 (2024) 108386.
- [35] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.
- [36] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-Decoder with Atrous Separable Convolution for Semantic image Segmentation, Springer International Publishing, Cham, 2018, pp. 833–851.
- [37] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, UNet++: A Nested U-Net Architecture for Medical image Segmentation, Springer International Publishing, Cham, 2018, pp. 3–11.
- [38] X. He, W. Song, Y. Wang, F. Poiesi, J. Yi, M. Desai, Q. Xu, K. Yang, Y.J.I.t.o.i. Wan, measurement, Lightweight retinal layer segmentation with global reasoning 73 (2024) 1–14.
- [39] G. Sun, Y. Pan, W. Kong, Z. Xu, J. Ma, T. Racharak, L.-M. Nguyen, J. Xin, DA-TransUNet: integrating spatial and channel dual attention with transformer U-net for medical image segmentation 12 - 2024 (2024).
- [40] A.G. Roy, S. Conjeti, S.P.K. Karri, D. Sheet, A. Katouzian, C. Wachinger, N. Navab, ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks, Biomed. Opt. Exp. 8 (2017) 3627–3642.

- [41] X. He, Y. Wang, F. Poiesi, W. Song, Q. Xu, Z. Feng, Y. Wan, Exploiting multi-granularity visual features for retinal layer segmentation in human eyes, *Front. Bioeng. Biotechnol.* 11 (2023) 1191803.
- [42] J. Li, P. Jin, J. Zhu, H. Zou, X. Xu, M. Tang, M. Zhou, Y. Gan, J. He, Y. Ling, Y. Su, Multi-scale GCN-assisted two-stage network for joint segmentation of retinal layers and discs in peripapillary OCT images, *Biomed. Opt. Exp.* 12 (2021) 2204–2220.
- [43] M.M. Rahman, R. Marculescu, MK-UNet: multi-kernel lightweight CNN for medical image segmentation, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 1042–1051.
- [44] J.M. Mari, N.G. Strouthidis, S.C. Park, M.J.A. Girard, Enhancement of lamina cribrosa visibility in optical coherence tomography images using adaptive compensation, *Invest. Ophthalmol. Vis. Sci.* 54 (2013) 2238–2247.
- [45] X. Li, Q. Guo, D. Lin, P. Li, W. Feng, S. Wang, MISF: multi-level interactive siamese filtering for high-fidelity image inpainting, in: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 1859–1868.
- [46] K. Hu, B. Shen, Y. Zhang, C. Cao, F. Xiao, X.J.N. Gao, Automatic segmentation of retinal layer boundaries in OCT images using multiscale convolutional neural network and graph search, 365 (2019) 302–313.