Zero-Shot Enhancement with Cross-Modal Applicability for Low-Light Vis-μOCT Images

Shujiang Chen[†], Yanshuo Li[†], Hua Wei[†], Fuwang Wu[†] and Weiye Song[*]
[*]Department of Mechanical Engineering
Shandong University, No.17923, Jingshi Road, Lixia District, Jinan 250061, China.
Email: songweiye@sdu.edu.cn.
[†]Department of Mechanical Engineering
Shandong University, No.17923, Jingshi Road, Lixia District, Jinan 250061, China.

Abstract—**Objective:Optical coherence tomography (OCT) is a rapid and non-destructive imaging technique, but image brightness decreases when imaging deep tissues or under low power and short exposure due to insufficient backscattered light. This issue is more pronounced in visible-light micro-OCT (vis-μOCT), where shorter wavelengths increase scattering and limit penetration, restricting its application. Method:In this paper, we propose Dif-NIR, a novel framework for enhancing low-light OCT images. The framework begins with a preliminary denoising stage. Image enhancement is then performed using a neural implicit representation (NIR) network, in which pixel values are incorporated as auxiliary input to mitigate the oversmoothing effect of fully connected layers. To enable unsupervised learning, custom-designed loss functions is employed. The proposed method is validated through qualitative and quantitative comparisons on a self-collected *en face* image dataset. To further assess its generalizability, we also performed experiments on B-scan images and retinal images acquired from other OCT devices. Result: On the *en face* image dataset, Dif-NIR outperforms existing methods in terms of visual quality, SNR (58.99 dB), CNR (49.56 dB), and NIQE (9.0553). It also effectively generalizes to OCT B-scan images and retinal images acquired by other devices. Conclusion: The proposed network effectively mitigates unpredictable brightness degradation, producing clearer and better-illuminated images while exhibiting strong generalization capability. Significance: The network effectively reveals deep-layer information in OCT images and can be applied to expand its usage scenarios to cost-effective and high-speed imaging settings.**

Index Terms—**Vis-μOCT, Zero-shot learning, Image enhancement, Neural implicit representation**

I. Introduction

OPTICAL coherence tomography (OCT) [1] is a fast, three-dimensional, non-invasive optical imaging modality that provides high resolution in both axial and lateral directions. It has been widely applied in various clinical and research fields, including cardiology [2], otology [3], dermatology [4], cytology [5], dentistry [6], urology [7], and ophthalmology [8]–[10]. The conventional OCT system typically achieves axial and lateral resolutions on the order of 3-5 μm and 5-8 μm, respectively. With the advent of supercontinuum sources(SC), which provide a smooth, continuous ultra-broad spectrum, researchers have demonstrated that by combining visible and near-infrared wavelengths, a Micro-OCT (μOCT) system can be developed that achieves a five-fold improvement in resolution across all spatial directions compared to traditional OCT [11]. The μOCT system built using the visible light spectrum is referred to as visible-light μOCT (vis-μOCT).

En face images provide unique advantages in scenarios that demand detailed lateral information or multi-layered structural analysis. The resolution of *en face* images is predominantly determined by the numerical aperture (NA) of the focusing objective in the sample arm of the OCT system. In general, a larger NA results in higher lateral resolution and the ability to distinguish smaller details, but it also reduces the depth of focus.

Although visible-light illumination sources hold promise for achieving high resolution, according to Beer-Lambert law, the intensity of the light beam in the sample arm diminishes as it penetrates deeper into the tissue, resulting in reduced light reaching deeper layers and weaker backscattered signals [12]. Furthermore, scenarios with lower hardware costs are more prone to exhibit conditions such as low input power, low quantum efficiency, and low exposure time. As a result, *en face* images captured at different depths exhibit variable illumination conditions. In addition, the excessive relative intensity noise (RIN) of SC introduces significantly higher noise levels, further complicating the imaging process.

Consequently, *en face* vis-μOCT images are frequently affected by several quality issues, including low illumination and high noise levels. To mitigate these degradations, researchers have investigated a range of potential solutions. Among hardware advancements, the most commonly employed technique is compounding [13]. Another approach involves analyzing the slope of the A-line profiles within homogeneous tissue to calculate the attenuation coefficient and compensate for intensity loss [14]. These methods are typically complex to implement and may require the acquisition of additional optical components, increasing their overall cost. Therefore, low-cost deep learning methods are gradually attracting our attention.

In order to achieve image enhancement under low light, we propose a zero-shot enhancement framework based on the Retinex theory [15] in this paper. Our key contributions are summarized as follows.

- We analyze the limitations of existing algorithms in addressing the specific challenges of *en face* vis-μOCT images and propose the Dif-NIR capable of simultane-

ously mitigating various forms of image degradation. By seamlessly integrating denoising and enhancement, this network progressively improves image quality.
- We further investigate an image enhancement module based on neural implicit representations (NIR), wherein grayscale values are incorporated as additional information for the network. Two parallel branches are designed to explore the optimal weighting of these grayscale values. To facilitate zero-shot training, we introduce multiple loss functions that effectively guide the image restoration process.
- We compile a dataset of *en face* vis-μOCT images from orange, cucumber, and chicken lung using our system. The proposed method is validated on this self-collected dataset, demonstrating superior performance compared to traditional methods and other deep learning approaches, as evidenced by both visual quality and quantitative metrics.

II. Related work

A. Traditional Methods

Traditional methods, such as grayscale histogram equalization [16], rely heavily on carefully designed optimization rules. Abdullah *et al.* [17] proposed a dynamic histogram equalization (DHE) technique, which partitions the histogram and applies equalization to each region independently. However, this method does not account for the actual brightness information, which can lead to overexposure or insufficient enhancement [18]. To overcome this problem, Chao *et al.* [19] introduced the histogram equalization preserving brightness with maximum entropy (BPHEME), which enhances the image while maintaining the original brightness. Reza *et al.* [20] proposed a system-level implementation of contrast-limited adaptive histogram equalization (CLAHE), which enhances image contrast while preserving local features. However, these methods are based on well-established prior assumptions, which limit their applicability in real-world scenarios where the lighting conditions may vary significantly.

B. Deep Learning-Based Methods

Recent advancements in low-light image enhancement using deep learning [21]–[26] have outperformed traditional methods. However, these methods heavily depend on high-precision paired datasets, limiting their applicability in medical imaging, thereby making unsupervised image enhancement networks more advantageous. Unsupervised image enhancement networks can be broadly categorized into two types: one that applies dynamic pixel adjustments inspired by techniques like gamma transformation, with works such as ZeroDce and its lightweight variant ZeroDce++ [27]. The other type is based on the Retinex theory. Anqi *et al.* [28] extended the Retinex theory by incorporating noise components, decomposing the image into illumination, reflectance, and noise components, which improved restoration quality to some extent. Yifan *et al.* proposed an unsupervised generative adversarial network, EnlightenGAN [29], which leverages information from the input itself to guide unpaired training, achieving excellent visual results. Risheng *et al.* [30] introduced a novel Retinex-inspired method called Unrolling with Architecture Search (RAUS), which develops a collaborative dual-layer search strategy, achieving enhanced results while also saving memory. Long *et al.* proposed the SCI [31] framework, which estimates low-illumination images from the input and incorporates a self-calibration module to reduce inference costs. Fu *et al.* [32] took an alternative approach by feeding two low-light images with different illumination conditions into the network (PairLIE) to train the model's ability of accurately decompose them, subsequently recovering the image by correcting the illumination map. Jiang *et al.* extended the PairLIE approach and proposed a two-stage training method, LightenDiffusion [33], achieving state-of-the-art performance.

While unsupervised deep learning methods demonstrate inherent suitability for medical imaging scenarios where ground-truth annotations are scarce, their direct deployment on OCT images risks amplifying noise during processing. Moreover, most existing networks require retraining when switching datasets, which significantly compromises their generalization capability. Therefore, we attempt to integrate a denoising component into the network architecture, suppress noise simultaneously throughout the enhancement process, and redesign a zero-shot network based on NIR.

III. Method

A. Overview

The general architecture of the proposed framework is illustrated in Fig. 1. Given an original *en face* vis-μOCT image $x_0 \in \mathbb{R}^{H\times W\times 1}$, the process begins with a Preliminary Denoising Module (PDM) that utilizes the denoising diffusion probability model (DDPM) to iteratively remove noise. The image is then resized and passed through the Image Enhancement Module (IEM), in the end, subsequently restored to its original size through a fast guide filter [34].

B. Preliminary Denoising Module

In the denoising phase, the PDM follows the standard diffusion model [35], where forward diffusion is employed to train the network, and reverse sampling is subsequently applied to generate an initial image with effectively removed noise. To improve denoising performance, the model was retrained using OCT images. It is worth noting that the PDM serves as a preprocessing step for OCT image enhancement and can be replaced by other denoising networks in principle. Additionally, if the noise level in the images is low, removing this component can still achieve good enhancement results.

C. Image Enhancement Module

The Retinex theory serves as a fundamental framework in the field of low-light image enhancement, establishing a mathematical relationship between a low-light image X and its corresponding ideal image under normal illumination R. By introducing an illumination map L that represents the true local lighting conditions, the theory models the mapping between the X and R. In low-light image enhancement research,

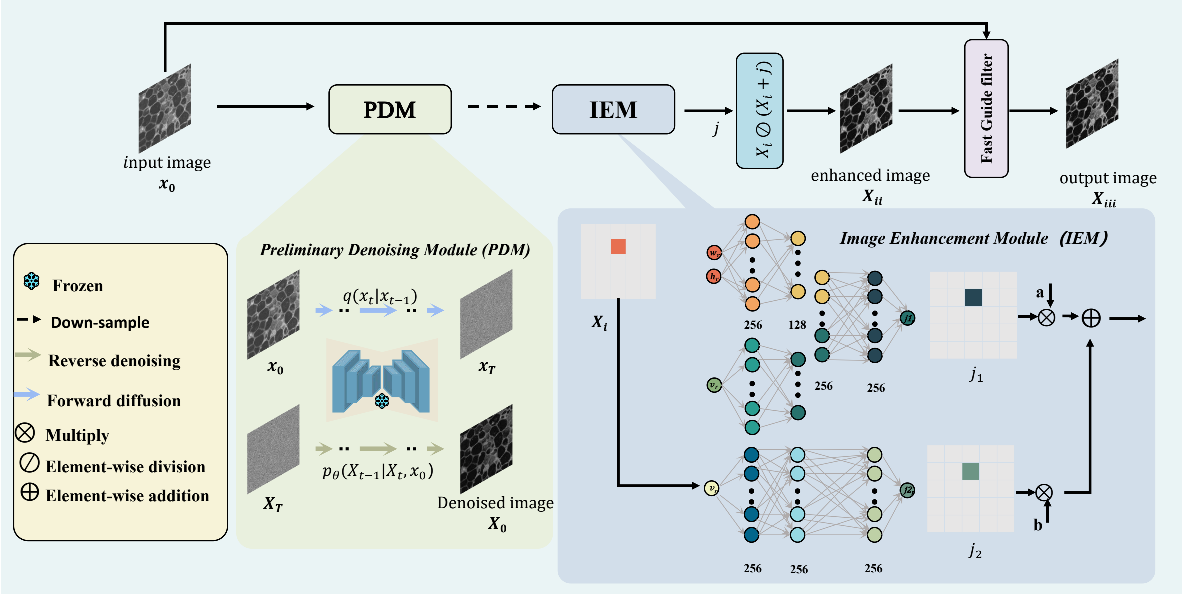

Fig. 1. The overall architecture of the Dif-NIR. Preliminary denoising is performed using a pre-trained PDM. Subsequently, during the zero-shot training phase, the image is first downsampled to a smaller size, followed by pixel-wise processing through a constructed MLP layer with a dual-branch structure. Finally, the image is restored to its original resolution using a fast guide filter, guided by the original image.

accurately estimating the illumination map L is critical for preserving image details and avoiding processing artifacts. The mathematical relationship among X, R, and L is formulated as follows:

$$X = R \odot L \tag{1}$$

where $\odot$ denotes the operation of the Hadamard product.

To achieve accurate estimation of the illumination map l, this study introduces a parameterized mapping function $F(x)$, as shown in Equation 2. Based on the estimated illumination map l and the observed low-light image x, the corresponding image r under ideal lighting conditions can be further derived. Additionally, to reduce computational complexity and enhance the stability and robustness of exposure control while avoiding overexposure, a function $f_\iota(\cdot)$ is employed to learn the residual between l and X_i. This design is inspired by the widely accepted assumption that the illumination map and low-light images exhibit structural similarity, with a near-linear relationship in most regions.

$$F(x): \begin{cases} l = x + f_\iota(\cdot) \\ r = x \oslash l \end{cases} \tag{2}$$

In constructing the model for $f_\iota(\cdot)$, a parameterized Multi-layer Perceptron (MLP) is employed to map the input image x to a residual image, enabling precise modeling of the differences between the illumination map and the low-light image. Based on the framework of NIR, the MLP establishes a continuous mapping between image intensity values and spatial coordinates, providing theoretical support for the model design. Unlike conventional RGB image mappings that involve handling three separate channels as defined in Equation 3, OCT grayscale images contain only a single channel. Therefore, the mapping is directly learned from the two-dimensional spatial coordinates to the corresponding grayscale value at the target location.

$$f : (x, y) \longrightarrow (r, g, b) \tag{3}$$

Furthermore, to address the tendency of fully connected networks to produce overly smooth global features, the grayscale value at each pixel is also included in the input. Moreover, inspired by the residual connection structure, we employ a dual-branch structure and further investigate the weight adaptation for pixel value in-formation. To overcome the representational limitations of the ReLU activation function, which lacks second-order derivatives, SIREN [36] layers are used to construct the MLP. As illustrated in Fig. 1, the architecture consists of a main branch that processes the concatenated pixel coordinates and grayscale values, and a secondary branch that uses only the grayscale value to provide fine-grained corrections. The main branch reduces its feature dimensions by half before being connected to the output layer, while the correction branch maintains a fixed width of 256 channels across all hidden layers. The residual maps generated by the two branches are then fused with the appropriate weights to produce the final residual map. The residual map estimation process can be represented by Equations 4 and 5.

$$f_l(\cdot): \begin{cases} ([h_r, w_r], v_r) \rightarrow j_{1r} \\ (v_r) \rightarrow j_{2r} \end{cases} \tag{4}$$

$$j = a \cdot j_1 + b \cdot j_2 \tag{5}$$

Where $[h_r, w_r]$ denotes the coordinates of the pixel at position r in X_i, and v_r represents the grayscale value of the pixel at this location. j_1 and j_2 are the residual images estimated by the two branches, with a and b being the fusion weights for the two images. Finally, j is the resulting residual image obtained after the fusion process.

IV. Network Training

A. Diffusion Model

Specifically, PDM optimizes the parameters θ of the noise estimation network by training the network's ability to estimate noise, ensuring that the noise ϵ_t estimated by the network closely approximates the noise added to the image at the current time step. The loss function of the diffusion model is as follows:

$$L_{diff} = \parallel \epsilon_t - \epsilon_\theta(X_t, x_0, t) \parallel_2 \tag{6}$$

B. Zero-shot Training

The fidelity loss L_{spa} is designed to maintain the consistency of the pixel level between the estimated illumination map l and the image X_i, which is expressed as:

$$L_{spa} = \parallel X_i - l \parallel_2 \tag{7}$$

The smoothness loss L_{tv} is designed to ensure the smoothness of the estimated illumination map. This is based on the common assumption that the illumination of an image should be locally or even globally smooth, without large gradients. The illumination smoothness loss is defined as:

$$L_{tv} = (\parallel \nabla_i l \parallel_2 + \parallel \nabla_j l \parallel_2)^2 \tag{8}$$

Where ∇_i and ∇_j represent the vertical and horizontal gradient operations, respectively.

L_{exp} is designed to adjust the overall brightness level of the l, thereby indirectly influencing the overall brightness level of the restored image.

$$L_{exp} = \frac{1}{N} \sum_{k=1}^{N} \parallel \sqrt{T_k} - 0.6 \parallel_2 \tag{9}$$

Where N represents the number of subdivisions of l into N small regions, and T_k denotes the average gray value of each region. The value 0.6 is a hyperparameter we set manually, where different values result in different brightness levels.

To maintain the fidelity of different regions in the image and prevent excessively high brightness values, we employ L_g to constrain the brightness values of each pixel in X_{ii} and M represents the total number of pixels in the image.

$$L_g = \frac{1}{M} \sum_{1}^{M} \mid X_{ii} \mid \tag{10}$$

Although most of the noise is filtered out during PDM, the unavoidable amplification of noise during IEM remains a persistent challenge. Therefore, the learned perceptual image patch similarity (LPIPS) is used to minimize the perceptual differences between the output image X_{iii} and X_0. The vgg network is selected as a measure of perceived similarity.

$$L_p = vgg(X_0, X_{iii}) \tag{11}$$

Finally, the losses are assigned different weight values and then summed. The total loss of the zero-shot enhancement network is expressed as:

$$L_{totle} = \alpha L_{spa} + \beta L_{tv} + \gamma L_{exp} + \delta L_g + \eta L_p \tag{12}$$

V. Experiments

A. Dataset

The vis-μOCT system developed in our laboratory is used for data acquisition. The system utilizes a light source with a wavelength range of 500-750 nm and incorporates a custom-designed spectrometer, achieving an axial resolution of 1.3 μm and a lateral resolution of 1.5 μm. The schematic and corresponding photograph of the device are presented in Fig. 2. Imaging was performed on various biological tissues, including cucumber, orange pulp, and chicken lung. Due to the shallow depth of focus of the system, only *en face* images near the focal plane were selected to construct the dataset.

B. Implementation Details

The DDPM is trained using the U-Net architecture as the noise estimation network. The images were randomly cropped to a size of 64×64, and the Adam optimizer was used for training with parameters $\beta_1 = 0.9, \beta_2 = 0.999$, and a learning rate set to 2×10^{-5}. The forward diffusion time step T was set to 1000, and the sampling step S was set to 20. After freezing the DDPM, the learning rate for the Adam optimizer in IEM was set to 0.001, with training conducted over 300 epochs. After the 100th epoch, the learning rate was halved every 10 epochs. The weight coefficients for the various loss terms were set as follows:$\alpha = 1, \beta = 20, \gamma = 8, \delta = 5$ and $\eta = 8$, while the fusion weights for the residual images are set as $a = 0.9$ and $b = 0.1$. To improve efficiency, the images were downsampled to a resolution of 256×256 before entering the enhancement network. It is worth noting that in IEM, the network does not require extensive training datasets but instead selectively focuses on the current image, enhancing its generalization capability, which demonstrates the zero-shot nature of our framework.

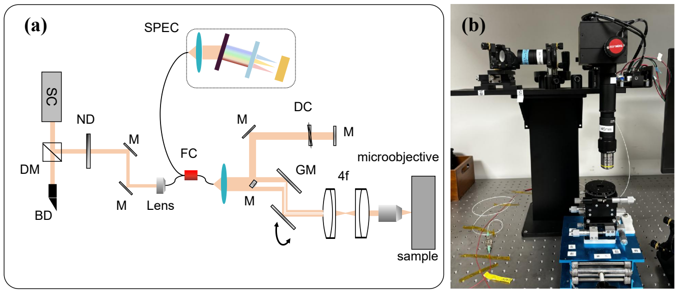

Fig. 2. (a) the schematic of the vis-μOCT. (b) the photograph of the system.SC: supercontinuum;DM: dichroic mirror;ND: neutral density filter;BD: beam dump;SPEC: spectrometer;FC: fiber coupler;M: mirror;GM: galvanometer mirror;DC: glass plate.

C. Comparision With Existing Methods

1) Comparison Methods: To validate the superiority and advancement of the proposed method on *en face* vis-μOCT images, we perform a comparative analysis using traditional methods and recent state-of-the-art unsupervised methods on our dataset. Since ground truth data is unavailable, supervised methods are excluded from the comparison. Traditional methods include HE [16], CLAHE [20], and methods based on entropy curves and homomorphic filtering [37]. Unsupervised methods include ZeroDce [27], ZeroDce++ [27], RRDNet [28], and SCI [31].

2) Quantitative Comparison: No-reference evaluation metrics, including the Average Gradient (AG), Contrast-to-Noise Ratio (CNR) and Natural Image Quality Evaluator (NIQE) [38], are used to assess the contrast and naturalness of the enhanced images. The Signal-to-Noise Ratio (SNR) evaluates the noise reduction capability of the network, while the LPIPS [39] measures the similarity between the enhanced and original images. To more accurately differentiate between signal and noise regions, we measured the SNR using epoxy resin mixed with polystyrene microspheres as the sample material, designating the microspheres as the signal region and the remaining area as the background. Moreover, to prevent AG from failing to reflect true image quality due to large gradient differences between background and signal regions, we calculate AG using only the signal region for a fairer comparison. This approach may yield higher AG values because the denominator in the formula, which represents the total number of pixels, is reduced by excluding background areas. As a result, the AG value may increase even though boundary gradients are removed. The calculation methods for AG, SNR and CNR are shown in Equation 13, Equation 14 and Equation 15. $\nabla_i x$ and $\nabla_j x$ represent the gradients of the image in the row and column directions, respectively. $mean_{signal}$ represents the mean of the signal region, std_{noise} represents the standard deviation of the noise, and ϵ represents a very small constant. Table I summarizes the quantitative results for these metrics. The data clearly show that our method outperforms existing approaches in terms of SNR, CNR and NIQE, and ranks among the top three in AG.

$$AG = (\sum_{i=1}^{N}\sum_{j=1}^{M}\sqrt{(\nabla_i x)^2 + (\nabla_j x)^2})/(M * N) \tag{13}$$

$$SNR = 10 * log(mean_{signal}/(std_{noise} + \epsilon)) \tag{14}$$

$$CNR = 10 * log((mean_1 - mean_2)/(std_{noise} + \epsilon)) \tag{15}$$

TABLE I
QUANTITATIVE COMPARISONS (SNR, NIQE, AG, LPIPS,CNR) OF DIFFERENT METHODS ON THE *en face* μOCT DATASETS. THE TOP THREE RESULTS ARE HIGHLIGHTED IN RED.

Method	SNR↑	NIQE↓	AG↑	LPIPS↓	CNR↑
Original	**30.78**	**10.4673**	132.00	-	16.90
He [16]	12.49	10.9743	**151.97**	0.2210	-62.52
Clahe [20]	28.26	10.9733	142.78	**0.0368**	**14.96**
Entropy [37]	25.96	**10.3081**	143.42	0.1723	-17.34
ZeroDce [27]	24.76	10.6654	138.54	0.0893	9.71
ZeroDce++ [27]	28.22	10.6227	135.38	**0.0447**	10.59
RRDNet [28]	27.10	10.5193	140.58	**0.0288**	13.58
SCI [31]	27.87	11.4727	**166.67**	0.0986	9.23
Base [35]	**55.20**	11.1965	138.32	0.1577	**47.99**
Dif-NIR	**58.99**	**9.0553**	**147.50**	**0.1382**	**49.56**

3) Qualitative Comparison: As shown in Fig. 3 and 4, traditional methods tend to result in over-enhancement, with inadequate control over contrast. Although deep learning-based approaches are more effective in adjusting overall brightness, they often introduce noticeable noise. In Fig. 5, we present the enhanced images of polystyrene microspheres embedded in mixing epoxy glue, together with the gradient contrast within the microspheres and the noise regions, from which the above conclusions can be drawn more easily. In comparison, the method proposed in this study demonstrates significant advantages in noise suppression, preservation of image gradient fidelity, and improvement of image illumination.

VI. DISCUSSION

A. Ablation Study

1) The Role of Pixel Value: Fig. 6 illustrates the results of experiments comparing the use of only image coordinate information in the INR with those that also incorporate the residual error of the grayscale value at the current coordinate. Subsequently, we introduced a second branch that independently inputs pixel values, and we evaluated performance metrics under various weight configurations for the two branches, as detailed in Table II. The results demonstrate that incorporating pixel value information significantly enhances the restoration

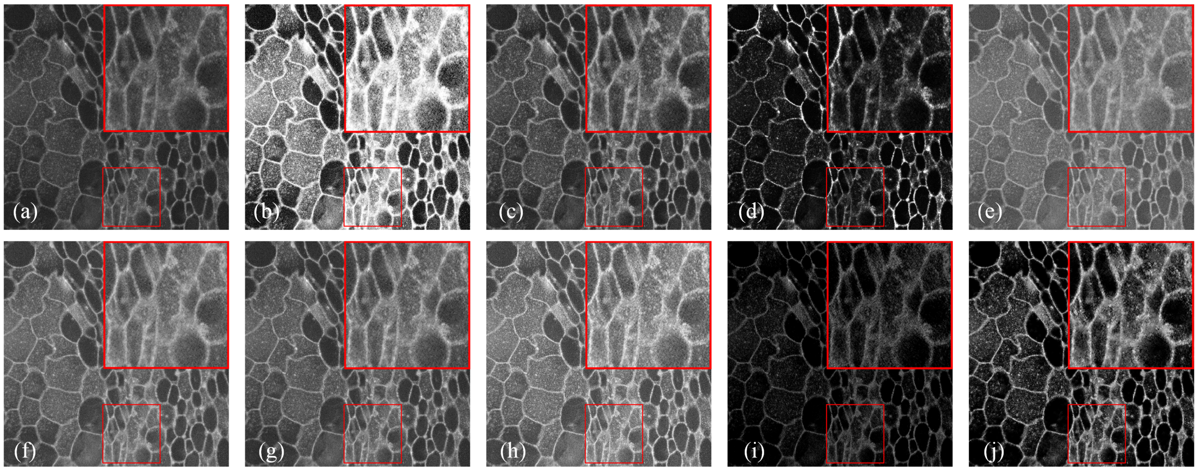

Fig. 3. Visual comparisons of different methods on orange pulp images. The grayscale range of all images is linearly scaled to 0-255. (a)Input; (b)He; (c)Clahe; (d)entropy; (e)ZeroDce; (f)ZeroDce++; (g)RRDNet; (h)SCI; (i)Base; (j)Dif-NIR

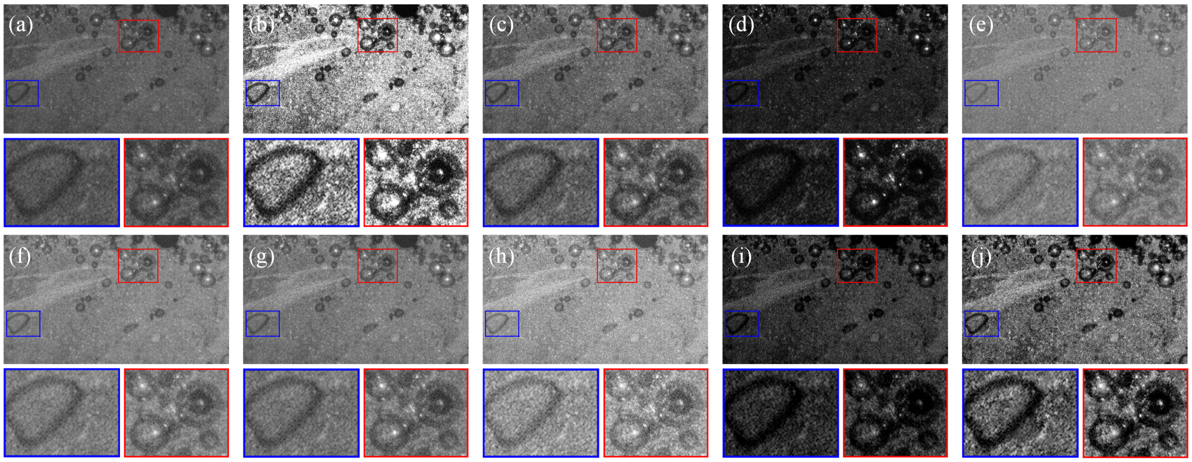

Fig. 4. Visual comparisons of different methods on chicken lungs images. The grayscale range of all images is linearly scaled to 0-255. (a)Input; (b)He; (c)Clahe; (d)entropy; (e)ZeroDce; (f)ZeroDce++; (g)RRDNet; (h)SCI; (i)Base; (j)Dif-NIR

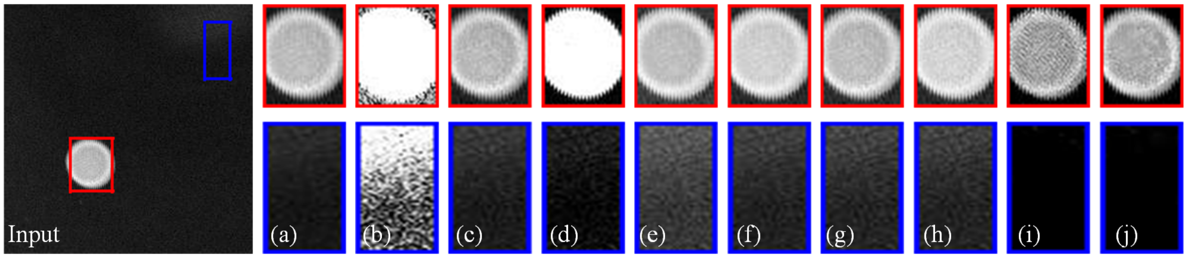

Fig. 5. Visual comparisons of different methods on polystyrene microsphere images. As can be seen from the figure, Dif-NIR clearly demonstrates superior performance in SNR and fidelity, effectively suppressing noise enhancement.The grayscale range of all images is linearly scaled to 0-255. (a)Input; (b)He; (c)Clahe; (d)entropy; (e)ZeroDce; (f)ZeroDce++; (g)RRDNet; (h)SCI; (i)Base; (j)Dif-NIR

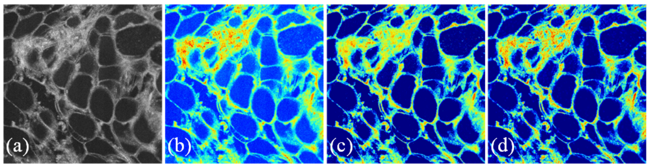

Fig. 6. The role of pixel values in the network. Absence of pixel values as input leads to a smoother enhanced image with loss of texture details, whereas providing grayscale values significantly improves texture restoration. (a)Original image; (b) Grayscale colormap of the original image; (c) Grayscale colormap of the enhanced image without pixel value input; (d) Grayscale colormap of the enhanced image with pixel value input

of the image's original gradient details. Moreover, when the weights are appropriately adjusted, the second branch further improves the image quality, enabling the preservation of even more fine-grained details.

2) Contribution of Each Loss.: Table III summarizes the impact of each loss function by removing different loss terms. Quantitative results show that progressively refining the loss design leads to optimal LPIPS, NIQE, and CNR. Introducing L_{tv} significantly improves AG, indicating that constraining illumination smoothness effectively preserves the inherent contrast of image regions. Adding L_g regularization slightly reduces SNR and AG, as the exposure constraint helps maintain a brightness distribution closer to human visual perception. Finally, L_p loss further boosts SNR and CNR by introducing a noise suppression mechanism. Although L_{exp}'s functionality remains challenging to evaluate through standard metrics, we have additionally presented its performance variations under different hyperparameters in Fig. 7. This confirms that multi-loss synergy uniquely enables balanced enhancement across detail preservation, noise suppression, and exposure control.

TABLE II
THE PERFORMANCE METRICS UNDER DIFFERENT WEIGHT ASSIGNMENTS FOR THE TWO BRANCHES WERE EVALUATED. THE BEST RESULTS ARE MARKED IN BOLD.

Weight	SNR↑	NIQE↓	AG↑	LPIPS↓	CNR↑
$a = 1, b = 0$	56.51	9.0633	147.34	0.1390	43.67
$a = 0.9, b = 0.1$	**58.99**	**9.0553**	**147.50**	**0.1382**	**49.56**
$a = 0.7, b = 0.3$	58.30	9.0342	148.67	0.1391	49.44
$a = 0.5, b = 0.5$	58.19	9.0037	147.80	0.1389	49.20
$a = 0.3, b = 0.7$	57.63	8.949	145.40	0.1395	46.88
$a = 0.1, b = 0.9$	55.93	8.966	139.75	0.1431	42.35

TABLE III
CONTRIBUTIONS OF EACH LOSS.

	L_{spa}	$L_{spa}+$ L_{tv}	$L_{spa} +$ $L_{tv} +$ L_{exp}	$L_{spa} + L_{tv} +$ $L_{exp} + L_g$	$L_{spa} + L_{tv} +$ $L_{exp} + L_g +$ L_p
SNR ↑	53.03	44.99	54.75	54.04	**58.99**
AG↑	107.84	151.21	152.97	146.22	**147.50**
NIQE↓	8.6624	9.177	9.1948	9.2213	**9.0553**
LPIPS↓	0.2701	0.1687	0.1769	0.1485	**0.1382**
CNR↑	31.47	33.10	40.66	44.69	**49.56**

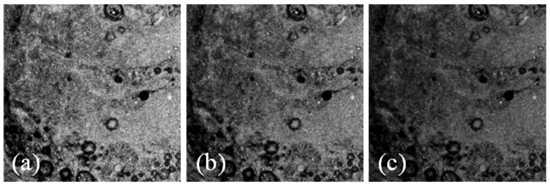

Fig. 7. Overview of the effect of the hyperparameter in the exposure loss term on the brightness of the output image. The brightness hyperparameters used in (a), (b), and (c) are set to 0.1, 0.5, and 0.9, respectively.

B. Migration Verification

In this section, the capability of the proposed network is further validated through the acquisition of B-scan images of transparent adhesive tape, which is a material with a highly reflective surface. As shown in Fig. 8, images captured under normal illumination and reduced exposure time are presented in Fig. 8a and Fig. 8b, respectively. Comparable brightness levels are observed in the surface layer of both images. However, a gradual decline in clarity and a loss of fine details with increasing imaging depth are observed in Fig. 8b, which cannot be resolved through threshold selection. In Fig. 8c and Fig. 8d, brightness in the dark regions of low-light images is effectively enhanced by the network, revealing additional structural details. Furthermore, by modifying the loss function L_{exp} within the network, precise control over brightness levels is achieved.

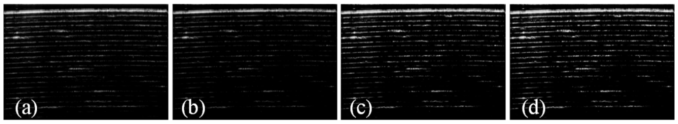

Fig. 8. Effect of Dif-NIR on the B-scan images of transparent adhesive tape. (a) and (b) represent images acquired under normal illumination and reduced exposure time, respectively, while (c) and (d) show the results of network enhancement with different brightness settings.

Furthermore, the proposed network was validated in both invisible low-illumination chicken lung images and retinal images captured by other OCT devices. As shown in Fig. 9, an extremely low signal image (Fig. 9a) was acquired by reducing the exposure time, where almost no tissue structures were visible. In contrast, the restored result (Fig. 9b) demonstrates adequate illumination and clear structural details. For the retinal image in Fig. 9c, the choroid layer exhibits weak signals due to optical signal attenuation, while the restored version (Fig. 9d) shows enhanced visibility of the choroid. Furthermore, brightness enhancement appears to improve layer delineation in retinal images. Table IV summarizes the quantitative metrics of both the original and enhanced images on the retinal dataset, further demonstrating the superiority of our method.

TABLE IV
QUANTITATIVE ASSESSMENT ON THE RETINAL DATASET.

	SNR↑	AG↑	NIQE↓	CNR↑	LPIPS↓
Input	35.41	98.55	5.5227	61.02	-
Output	37.75	106.45	5.0026	64.70	0.0447

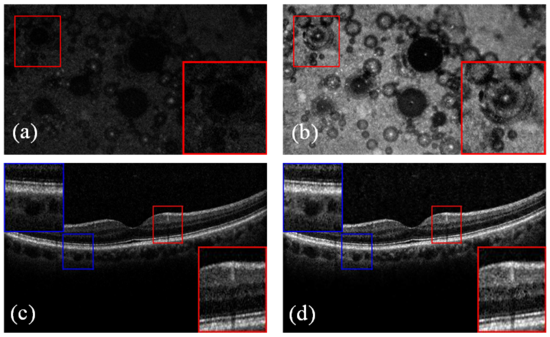

Fig. 9. Network validation in invisible low-light image of chicken lungs and retinal cross-sectional imaging. (a) Chicken lung image captured under extremely low-light conditions; (b) Enhanced chicken lung image; (c) Retinal image captured under low-light conditions; (d) Enhanced retinal image.

VII. CONCLUSION

To address the challenges of unpredictable brightness degradation in *en face* vis-μOCT images, we propose Dif-NIR, a zero-shot enhancement framework integrating denoising and enhancement. The network processes input images sequentially through a frozen PDM and an IEM. The IEM, guided by carefully designed loss functions, enables zero-shot training. We perform quantitative and qualitative comparisons with existing methods on a self-constructed *en face* vis-μOCT dataset. The results demonstrate that Dif-NIR outperforms current approaches in both metrics and visual quality. Furthermore, we evaluated the generalizability of the proposed network by applying it to B-scan images and retinal images acquired from other OCT devices. The results indicate that our method remains effective on both of them. In particular, since the network does not rely on dataset-specific training or weight tuning, it exhibits strong generalization ability and holds promise for broader applicability. In summary, the proposed framework improves both *en face* and B-scan images under low-light conditions and demonstrates robust generalization across different OCT modalities and imaging systems. This work offers a deep learning-based solution to enable the application of OCT systems in low-light conditions, particularly in settings involving low-cost, low-efficiency detectors or high-speed imaging.

REFERENCES

[1] D. Huang et al., "Optical coherence tomography," *science*, vol. 254, no. 5035, pp. 1178–1181, 1991.
[2] X. Yao et al., "Myocardial imaging using ultrahigh-resolution spectral domain optical coherence tomography," *Journal of biomedical optics*, vol. 21, no. 6, pp. 061 006–061 006, 2016.
[3] J. S. Iyer et al., "Micro-optical coherence tomography of the mammalian cochlea," *Scientific Reports*, vol. 6, no. 1, p. 33288, 2016.
[4] O. Babalola et al., "Optical coherence tomography (oct) of collagen in normal skin and skin fibrosis," *Archives of dermatological research*, vol. 306, pp. 1–9, 2014.
[5] X. Yu et al., "Toward high-speed imaging of cellular structures in rat colon using micro-optical coherence tomography," *IEEE Photonics Journal*, vol. 8, no. 4, pp. 1–8, 2016.
[6] Y.-S. Hsieh et al., "Dental optical coherence tomography," *Sensors*, vol. 13, no. 7, pp. 8928–8949, 2013.
[7] M. T. Bus et al., "Volumetric in vivo visualization of upper urinary tract tumors using optical coherence tomography: a pilot study," *The Journal of urology*, vol. 190, no. 6, pp. 2236–2242, 2013.

[8] W. Song et al., "Wavelength-dependent optical properties of melanosomes in retinal pigmented epithelium and their changes with melanin bleaching: a numerical study," *Biomedical optics express*, vol. 8, no. 9, pp. 3966–3980, 2017.

[9] W. Song et al., "Multimodal volumetric retinal imaging by oblique scanning laser ophthalmoscopy (oslo) and optical coherence tomography (oct)," *Journal of Visualized Experiments: Jove*, no. 138, p. 57814, 2018.

[10] W. Song et al., "Fiber-based visible and near infrared optical coherence tomography (vnoct) enables quantitative elastic light scattering spectroscopy in human retina," *Biomedical Optics Express*, vol. 9, no. 7, pp. 3464–3480, 2018.

[11] M. Kashiwagi et al., "Feasibility of the assessment of cholesterol crystals in human macrophages using micro optical coherence tomography," *PloS one*, vol. 9, no. 7, p. e102669, 2014.

[12] J. M. Schmitt, "Optical coherence tomography (oct): a review," *IEEE Journal of selected topics in quantum electronics*, vol. 5, no. 4, pp. 1205–1215, 1999.

[13] S. Adabi et al., "Optical coherence tomography technology and quality improvement methods for optical coherence tomography images of skin: a short review," *Biomedical engineering and computational biology*, vol. 8, p. 1179597217713475, 2017.

[14] A. Hojjatoleslami and M. R. Avanaki, "Oct skin image enhancement through attenuation compensation," *Applied optics*, vol. 51, no. 21, pp. 4927–4935, 2012.

[15] E. H. Land and J. J. McCann, "Lightness and retinex theory," *Josa*, vol. 61, no. 1, pp. 1–11, 1971.

[16] A. C. Bovik, *Handbook of image and video processing*. Academic press, 2010.

[17] M. Abdullah-Al-Wadud et al., "A dynamic histogram equalization for image contrast enhancement," *IEEE transactions on consumer electronics*, vol. 53, no. 2, pp. 593–600, 2007.

[18] X. Guo et al., "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on image processing*, vol. 26, no. 2, pp. 982–993, 2016.

[19] C. Wang and Z. Ye, "Brightness preserving histogram equalization with maximum entropy: a variational perspective," *IEEE Transactions on Consumer Electronics*, vol. 51, no. 4, pp. 1326–1334, 2005.

[20] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement," *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 38, pp. 35–44, 2004.

[21] S. W. Zamir et al., "Learning enriched features for real image restoration and enhancement," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*. Springer, 2020, pp. 492–511.

[22] X. Xu et al., "Snr-aware low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 17 714–17 724.

[23] Y. Wang et al., "Low-light image enhancement with illumination-aware gamma correction and complete image modelling network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 13 128–13 137.

[24] Y. Wu et al., "Learning semantic-aware knowledge guidance for low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 1662–1671.

[25] X. Xu et al., "Low-light image enhancement via structure modeling and guidance," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 9893–9903.

[26] H. Jiang et al., "Revisiting coarse-to-fine strategy for low-light image enhancement with deep decomposition guided training," *Computer Vision and Image Understanding*, vol. 241, p. 103952, 2024.

[27] C. Li et al., "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 8, pp. 4225–4238, 2021.

[28] A. Zhu et al., "Zero-shot restoration of underexposed images via robust retinex decomposition," in *2020 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2020, pp. 1–6.

[29] Y. Jiang et al., "Enlightengan: Deep light enhancement without paired supervision," *IEEE transactions on image processing*, vol. 30, pp. 2340–2349, 2021.

[30] R. Liu et al., "Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 10 561–10 570.

[31] L. Ma et al., "Toward fast, flexible, and robust low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5637–5646.

[32] Z. Fu et al., "Learning a simple low-light image enhancer from paired low-light instances," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 22 252–22 261.

[33] H. Jiang et al., "Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models," in *European Conference on Computer Vision*. Springer, 2025, pp. 161–179.

[34] K. He and J. Sun, "Fast guided filter," *arXiv preprint arXiv:1505.00996*, 2015.

[35] J. Ho et al., "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[36] V. Sitzmann et al., "Implicit neural representations with periodic activation functions," *Advances in neural information processing systems*, vol. 33, pp. 7462–7473, 2020.

[37] P. S. Yadav et al., "A new approach of contrast enhancement for medical images based on entropy curve," *Biomedical Signal Processing and Control*, vol. 88, p. 105625, 2024.

[38] N. Hautiere et al., "Blind contrast enhancement assessment by gradient ratioing at visible edges," *Image Analysis and Stereology*, vol. 27, no. 2, pp. 87–95, 2008.

[39] R. Zhang et al., "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.