frontiers | Frontiers in Neurorobotics

TYPE Original Research
PUBLISHED 02 October 2025
DOI 10.3389/fnbot.2025.1676787

Check for updates

OPEN ACCESS

EDITED BY
Mei Liu,
Multi-scale Medical Robotics Center Limited, China

REVIEWED BY
Mingyu Liu,
Technical University of Munich, Germany
Minqi Li,
Xi'an Polytechnic University, China

*CORRESPONDENCE
Zhixiong Huang
✉ hzxcyanwind@mail.dlut.edu.cn

RECEIVED 31 July 2025
ACCEPTED 12 September 2025
PUBLISHED 02 October 2025

CITATION
Fu W, Wang X, Yang C, Zhang L, Feng L and Huang Z (2025) DWMamba: a structure-aware adaptive state space network for image quality improvement. *Front. Neurorobot.* 19:1676787. doi: 10.3389/fnbot.2025.1676787

COPYRIGHT
© 2025 Fu, Wang, Yang, Zhang, Feng and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

DWMamba: a structure-aware adaptive state space network for image quality improvement

Wenjun Fu[1], Xiaobin Wang[2], Chuncai Yang[3], Liang Zhang[1], Lin Feng[4] and Zhixiong Huang[4]*

[1]Beijing China Coal Mine Engineering Co., Ltd., Beijing, China, [2]Gansu Coal First Engineering Co. Ltd., Baiyin, Gansu, China, [3]Wanyi First Mine, China Energy Baotou Energy Co. Ltd., Baotou, Inner Mongolia, China, [4]School of Information and Communication Engineering, Dalian Minzu University, Dalian, Liaoning, China

Overcoming visual degradation in challenging imaging scenarios is essential for accurate scene understanding. Although deep learning methods have integrated various perceptual capabilities and achieved remarkable progress, their high computational cost limits practical deployment under resource-constrained conditions. Moreover, when confronted with diverse degradation types, existing methods often fail to effectively model the inconsistent attenuation across color channels and spatial regions. To tackle these challenges, we propose **DWMamba**, a degradation-aware and weight-efficient Mamba network for image quality enhancement. Specifically, DWMamba introduces an Adaptive State Space Module (ASSM) that employs a dual-stream channel monitoring mechanism and a soft fusion strategy to capture global dependencies. With linear computational complexity, ASSM strengthens the models ability to address non-uniform degradations. In addition, by leveraging explicit edge priors and region partitioning as guidance, we design a Structure-guided Residual Fusion (SGRF) module to selectively fuse shallow and deep features, thereby restoring degraded details and enhancing low-light textures. Extensive experiments demonstrate that the proposed network delivers superior qualitative and quantitative performance, with strong generalization to diverse extreme lighting conditions. The code is available at https://github.com/WindySprint/DWMamba.

KEYWORDS

image quality improvement, multi-scenario enhancement, vision mamba, state space model, structural cue

1 Introduction

Images captured in real-world environments are often affected by various degradation factors, such as reduced visibility, blurred textures, and color distortion (Zhuang et al., 2022), which substantially hinder subsequent visual perception tasks. Developing a robust and efficient image enhancement method is therefore crucial, with two key requirements: (1) real-time performance for deployment on resource-constrained devices (Zhang et al., 2024c), and (2) strong adaptability to diverse types of degradation. Among these challenging scenarios, underwater environments represent a particularly complex case. Due to light absorption and scattering, underwater images frequently suffer from uneven brightness, amplified noise, and severe color deviation. Designing a robust enhancement method for underwater scenes is not only vital for underwater computer vision applications but also serves as a meaningful benchmark for general image

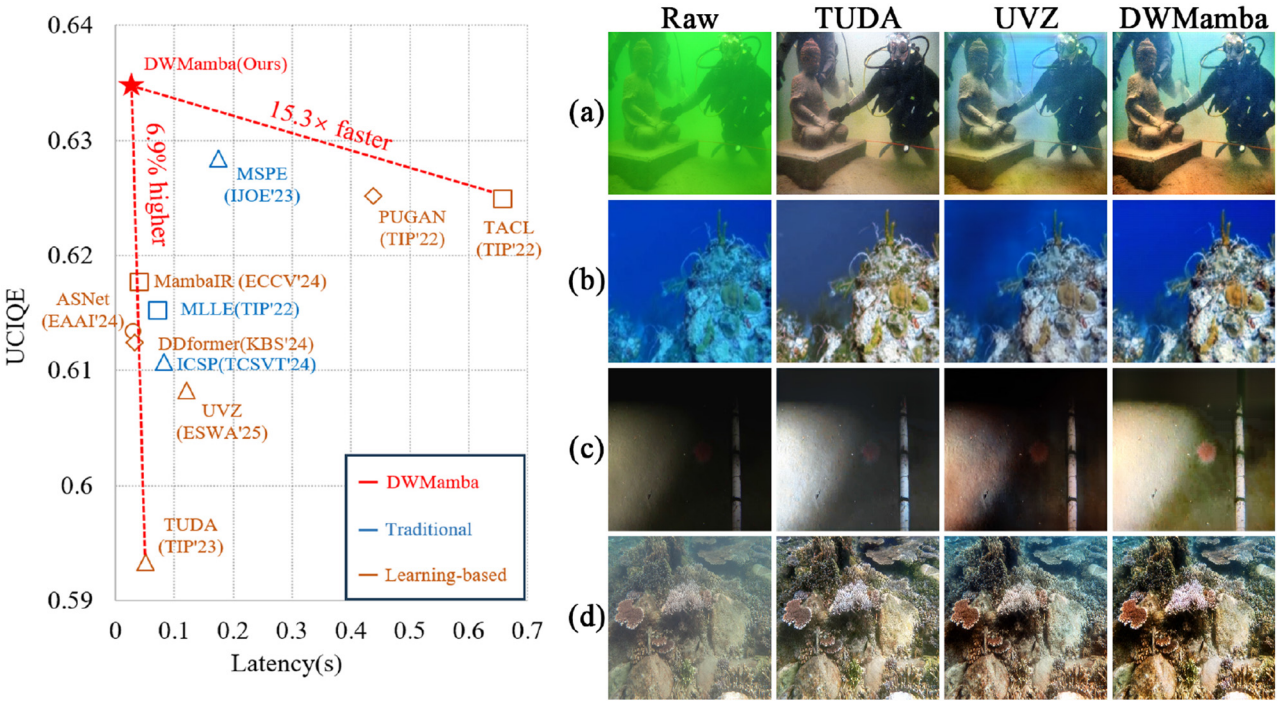

FIGURE 1
Metrics and generalizability comparison of advanced methods. The left side shows UCIQE and single-image latency. The right side shows underwater images including color deviation, low-light, overexposure, and sandy scenarios from top to bottom, along with enhanced results from different methods.

enhancement. Figure 1 presents a comparison of parameters and enhancement results for several state-of-the-art methods. While some methods achieve satisfactory underwater visibility enhancement, they often fail when confronted with other severe degradations (e.g., uneven brightness or strong visual interference). In contrast, DWMamba delivers both high efficiency and robust enhancement across a wide range of challenging scenarios.

Traditional methods (Berman et al., 2020; Zhou et al., 2023c; Kang et al., 2023) aim to improve image quality by reversing the degradation process or adjusting pixel values. However, as they rely on pre-defined priors or handcrafted region partitioning, their effectiveness is often limited under variations in lighting, turbidity, and shooting conditions (Zhou et al., 2024b). In recent years, deep learning methods have received considerable attention owing to advances in computing hardware and model optimization techniques (Liu et al., 2021; Jin et al., 2022; Liu M. et al., 2024; Fan et al., 2025; Huang H. et al., 2025). Convolutional neural network (CNN)-based approaches (Fu et al., 2022; Jiang N. et al., 2022) have demonstrated strong performance in local feature extraction and achieved remarkable visibility enhancement. Nevertheless, limited by the convolutional kernel size, CNNs struggle to capture long-range dependencies, making it difficult to adequately address locally degraded regions (e.g., overexposed areas or low-light regions) in uneven degradation scenarios. To overcome this issue, Transformer-based methods (Peng et al., 2023; Khan et al., 2024) leverage global receptive fields and dynamic weights, enabling more comprehensive color correction and detail enhancement. However, their computational complexity grows quadratically with image resolution, which severely restricts their practicality in resource-constrained environments. Although some approaches attempt to improve efficiency by partitioning input regions (Chen et al., 2021; Huang et al., 2022), these strategies inevitably reduce the receptive field and impair high-level semantic understanding.

Mamba (Gu and Dao, 2023), a recently proposed selectively structured state space model, demonstrates strong capability in modeling long-range dependencies and efficiently processing long natural language sequences due to its linear computational complexity. However, when applied to 2D image sequences, Mambas limited receptive field strategy fails to capture relationships between unscanned regions. To better adapt to the visual domain, VMamba (Yue et al., 2024) introduces a four-way scanning mechanism to establish a more comprehensive receptive field. Inspired by this property, we extend VMamba to the task of image quality improvement (IQI), aiming to achieve efficient and robust visual enhancement. In practice, directly applying VMamba suffers from several limitations: (1) The absence of explicit structural cues prevents the model from adequately focusing on degradation details, particularly in low-light conditions, which limits the enhancement of degraded textures. (2) Degraded images often exhibit significant inter-channel differences, yet the model lacks a mechanism to monitor global channel dependencies during channel mapping, leading to insufficient correction of color deviations. (3) During four-way scanning, uneven regional degradation causes feature inconsistencies across different directions, and directly aggregating these results reduces the models attention to critical features.

Therefore, how to further strengthen the model's perception and compensation ability for uneven degradation while maintaining efficient computation is an issue of concern.

In this work, we propose a degradation-aware and weight-efficient Mamba network for image quality improvement, called DWMamba. The framework primarily adopts an adaptive state space module (ASSM) to effectively capture global feature dependencies. Compared to VMamba, our method introduces a lightweight channel monitoring mechanism before the scanning mechanism, and implements a fusion learning strategy based on directional properties, which proves to be beneficial for the completeness of global dependencies. Furthermore, we designed a structure-guided residual fusion module (SGRF), which employs explicit edge priors and region partitioning as cues to guide the targeted fusion of shallow and deep features. Extensive comparative experiments demonstrate that DWMamba achieves competitive enhancement results with ideal computational resource. Moreover, we extended the model to different lighting environments (e.g., low-light, overexposure, haze, and sandy scenarios) without additional training, the impressive visual improvements further validate the robustness of DWMamba. The main contributions of this paper are summarized as follows:

- We present DWMamba, a novel Mamba-based network that integrates an adaptive state space module and a structure-guided residual fusion module (SGRF), offering new insights for IQI model design.
- We design an Adaptive State Space Module (ASSM) that incorporates a dual-stream channel monitoring mechanism and a soft fusion strategy, enabling finer-grained dependency modeling for accurate degradation restoration.
- We conduct extensive experiments showing that DWMamba achieves state-of-the-art performance on multiple benchmark datasets with lower computational cost, and demonstrates strong generalization across diverse visual degradation scenarios.

2 Related work

2.1 Traditional methods

Traditional methods can be broadly categorized into two types: restoration-based and enhancement-based methods. Restoration-based methods restore the input image by simulating degradation imaging and reversing this process. Dark channel prior (DCP) (He et al., 2011) provided insights into the image pixel distribution, the works (Peng et al., 2018; Liang et al., 2021) successfully extended DCP to various degraded scenarios. To address differences in channel attenuation, the works (Berman et al., 2020; Zhou et al., 2024d) introduces multiple priors to obtain higher quality images. Enhancement-based methods improve image quality through rational region partitioning and pixel adjustment. By focusing on color loss and pixel balancing, the works (Ancuti et al., 2019; Zhou et al., 2023a) effectively addressed the problems of color distortion and backscattering. Utilizing various image decomposition and fusion strategies, the works (Kang et al., 2023; An et al., 2024; Mishra et al., 2024) achieved multi-level enhancement of degraded images.

However, due to reliance on pre-set prior knowledge or single statistical feature, traditional methods may fail to generate the desired results in out-of-range scenarios.

2.2 Learning-based methods

With powerful feature learning capabilities and reasonable data matching, deep learning methods demonstrate good visual enhancement effects. The work (Liu M. et al., 2024) integrates two transformers that focused on multi-scale and global features, enhancing the network's response to attenuated color channels and degraded spatial regions. Introducing the color compensation mechanism, the works (Liu C. et al., 2024; Wang Y. et al., 2024) provide reasonable enhancement for the severely degraded channels, resulting in more natural images. The works (Li et al., 2021; Jiang Z. et al., 2022; Zhou et al., 2024c; Zhang et al., 2024a) adopt the multi-branch architecture to enrich feature representation, enabling the network to flexibly address the complex degradation characteristics. The works (Zhang et al., 2024d; Zhao et al., 2024) introduce the frequency domain to refine the image details, further extending the representation abilities of models.

To reduce the requirement of training data, unsupervised techniques or cross-domain knowledge transfer are applied to IQI tasks. The work (Jiang Q. et al., 2022) employs transfer learning for translating underwater images to the air domain, and then uses shared dehazing weights to further enhance the images. The work (Nathan et al., 2025) combines the in-air natural outdoor dataset with the imaging model to train the diffusion model, and obtains restored images through multiple rounds of denoising. By using target detection or evaluation metrics as training rewards, the works (Liu et al., 2022; Wang H. et al., 2024) generate enhanced images conducive to related tasks. The work (Fu et al., 2022) reconstructs raw image by estimating the relevant elements in the imaging process while utilizing the homology constraint to supervise image quality.

Deep learning methods have made significant progress in IQI tasks, but high parameters limit their practicality in resource-limited environments. More critically, the generalization of these methods is greatly challenged by variations in lighting conditions.

2.3 State space models

State Space Models (SSMs), as a classical sequence modeling structure in control theory, recently received extensive attention in the field of deep learning. Structured SSM (S4) (Gu et al., 2021) is a pioneering work in deep state space modeling, significantly improved modeling efficiency by improving the structure of the state matrix. Building on S4, Mamba (Gu and Dao, 2023) introduces a selectivity mechanism to break the constraints of constant transition, allowing SSM to pass or forget correlations between elements along the sequence. To introduce Mamba's properties into the visual domain, VMamba (Yue et al., 2024)

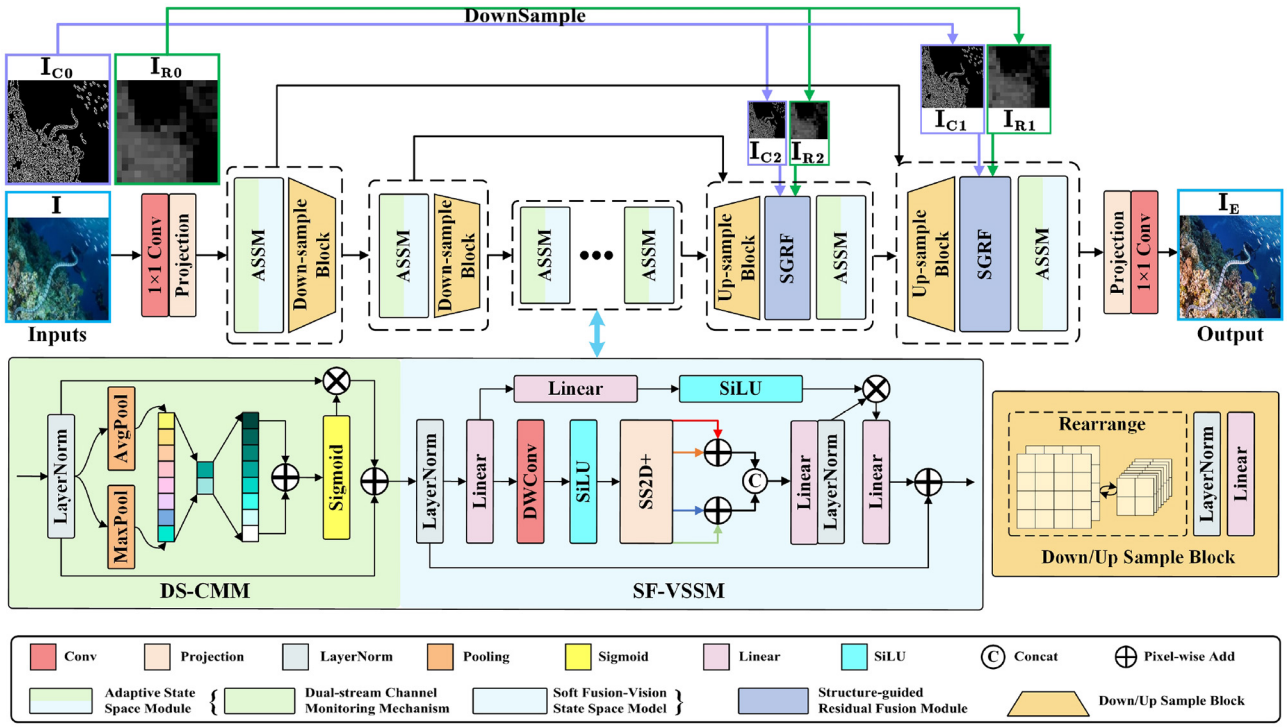

FIGURE 2
Network architecture of our proposed DWMamba. The **left** side below figure shows the detailed structure of the Adaptive State Space Module (ASSM), including the Dual-stream Channel Monitoring Module (DS-CMM) and the Soft Fusion Visual State Space Module (SF-VSSM); the **right** side shows the sample block. The details of the Structure-guided Residual Fusion Module (SGRF) are shown in Figure 5.

proposes a cross-scan module to apply the four-way scanning mechanism, which fully expands the restricted receptive field. Inspired by the outstanding performance of VMamba, various low-level vision tasks [e.g., super-resolution Guo et al., 2025; Shi et al., 2025, low-light enhancement Bai et al., 2024; Liu et al., 2025, image dehazing Zhou et al., 2024a, and image denoising Liu Y. et al., 2024] quickly saw a boom in applications. Meanwhile, the work (Chen and Ge, 2024) first introduces the VMamba in the underwater image enhancement field and achieves high accuracy with a small number of parameters. In addition, the works (Guan et al., 2024; Dong et al., 2024) introduce the channel interaction mechanisms to capture the dependencies, while the work (Zhang et al., 2024b) introduces a physical imaging model to constrain the enhancement process, these methods maintain the linear remote modeling capability and effectively cope with wavelength-dependent channel attenuation.

However, existing VMamba-based IQI methods lack explicit cues to guide the targeted enhancement of local degradation details. Additionally, scanning features exhibit significant directional characteristics, and directly adding them results in insufficient attention to important features.

3 Methodology

3.1 Overall architecture

As shown in Figure 2, DWMamba adopts a multi-scale UNet architecture and contains a three-stage feature extraction process. For the degraded image $\boldsymbol{I} \in \mathbb{R}^{H\times W\times 3}$, we first utilize 11 convolution and channel projection to obtain the feature embedding $\boldsymbol{E}_0 \in \mathbb{R}^{H\times W\times C_0}$, where H and W represent the height and width, respectively, and C_0 represents the initial number of expanded channels. Subsequently, the features are passed through two layers of symmetric encoder-decoder, each containing ASSM and corresponding sampling operation. In the lightweight network, we configure the number of ASSM as [1, 1, 2, 1, 1] to ensure effective feature extraction and low computational complexity. The sampling process is conducted by Rearrange operation, LN layer, and Linear layer, with the scale and channel of features undergoing a two-fold opposite change each time. Unlike the encoding process, we choose to up-sample the results of skip connections in the decoder. With the structure-explicit modeling $\boldsymbol{I}_{C0} \in \mathbb{R}^{H\times W\times 1}$ and the region modeling $\boldsymbol{I}_{R0} \in \mathbb{R}^{H\times W\times 1}$ extracted from $\boldsymbol{I}$, we construct a SGRF to guide the network in performing targeted enhancement of degraded textures and low-light regions. After two stages of feature reconstruction, the high-quality enhanced image $\boldsymbol{I}_E \in \mathbb{R}^{H\times W\times 3}$ is refined by 1×1 convolution.

3.2 Adaptive state space module

During multi-dimensional mapping in deep networks, the degradation differences between feature channels are further amplified. However, due to the absence of an effective mechanism to capture interactions between channels, VMamba has limitations in correcting color deviations, which diminishes its practical effectiveness. Considering the above, we design the ASSM as shown

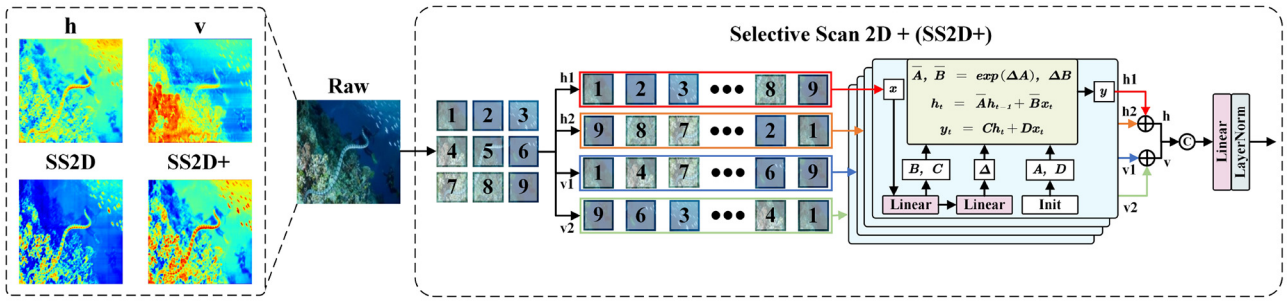

FIGURE 3
Specific flow of SS2D+. Unlike the original SS2D, we employ an improved strategy SS2D+ at the end, which discriminatively adds features from different directions and performs fusion learning. The heatmaps of horizontal adding (h) and vertical adding (v) results, SS2D, and SS2D+ are shown on the left.

in Figure 2 to assist in feature attention and region partitioning in the network. For the input features $\boldsymbol{f} \in \mathbb{R}^{H\times W\times C}$, we first normalize them through a LayerNorm (LN) layer, and then use DS-CMM to capture channel dependencies. Unlike traditional channel attention, it adopts a dual-branch structure and captures channel interactions through different pooling mechanisms and multiple residual connections, which is computed as follows:

$$\boldsymbol{f}_{CI} = \boldsymbol{f} + \boldsymbol{f} \times \left(CLC\left(\boldsymbol{f}_A\right) + CLC\left(\boldsymbol{f}_M\right)\right)_{Sig}, \tag{1}$$

where $\boldsymbol{f}_A$ and $\boldsymbol{f}_M$ represent the features after average pooling and maximum pooling, respectively, and $(\cdot)_{Sig}$ is the sigmoid function. $CLC(\cdot)$ is a feature extraction module containing a LeakyReLU function placed between two convolutional layers. Next, the features $\boldsymbol{f}_{CI}$ labeling the channel importance are fed into the vision state space module. The shallow features $\boldsymbol{f}_S$ are obtained by LN and linear layers (Lin), while the deeper features $\boldsymbol{f}'_S$ are extracted by depth-wise convolution (DWC) and SiLU function:

$$\boldsymbol{f}_S = \mathrm{Lin}\left(\mathrm{LN}\left(\boldsymbol{f}_{CI}\right)\right), \boldsymbol{f}'_S = \left(\mathrm{DWC}\left(\boldsymbol{f}_S\right)\right)_{SiLU}. \tag{2}$$

Through the four-way scanning mechanism, VMamba overcomes the modeling limitations of SSM in 2D images and enhances the perception of non-causal features. The focus of this mechanism is to extend multiple scanning directions to capture dependencies. However, in practical underwater applications, we have observed notable directional properties in the modeling of horizontal and vertical directions. For instance, in the heatmaps of Figure 3, the horizontal scanning result highlights attention between horizontally adjacent texture features, while the vertical scanning result distinctly delineates the vertical division of image regions. Due to the differences between these two results, directly adding them not only interferes with the unique feature contributions, but also compromises the integrity of important feature modeling. To this end, we have introduced a softer fusion strategy SF to enhance the aggregation of features from different directions.

Following the settings in the four-way scanning mechanism, $\boldsymbol{f}'_S$ is serialized as a patch sequence and captures the interactions between sequence elements from four directions. To capture more accurate spatial remote dependencies, we employ an adding-fusion learning process SS2D+. Specifically, SS2D+ performs a discriminative addition on scanning results in the same direction, then superimposes the addition results from different directions at the channel level through concatenation. Further feature fusion is achieved through linear and LN layers, allowing the network to integrate the modeling properties of both directions, thereby highlighting the attention to important features:

$$\boldsymbol{f}_{SS2D+} = LN\left(\mathrm{Lin}\left(\left(\boldsymbol{f}_{h1+h2}\right), \left(\boldsymbol{f}_{v1+v2}\right)\right)_{CAT}\right), \tag{3}$$

where $\left(\boldsymbol{f}_{h1+h2}\right)$ and $\left(\boldsymbol{f}_{v1+v2}\right)$ represent the adding results in horizontal and vertical directions, respectively. Finally, the remote dependencies are labeled on the shallow feature $\boldsymbol{f}_S$, and connected with $\boldsymbol{f}_{CI}$ in a residual manner to obtain the final modeling result:

$$\boldsymbol{f}_{ASSM} = \boldsymbol{f}_{CI} + \mathrm{Lin}\left(\boldsymbol{f}_{SS2D+} \otimes \mathrm{Lin}\left(\boldsymbol{f}_S\right)_{SiLU}\right). \tag{4}$$

3.3 Structure-guided residual fusion module

As shown in Figure 4, explicit modeling of structural cues brings significant benefits in degraded scenarios, mainly reflected in two aspects: Firstly, edge modeling can directly highlight key textures in blurred details, which is highly effective in improving image clarity and contrast. Secondly, due to uneven brightness distribution, high-quality region modeling tends to yield richer feature information. By leveraging this property to distinguish regions with varying brightness, the model can more thoroughly understand the imaging structure and depth information. Due to the differences between encoded and decoded features, the fused features require a guiding mechanism to focus on degradation details and illumination loss. Based on structural cues, this mechanism can promote the model to enhance texture details and increase sensitivity to illumination changes, thereby further improving image enhancement performance.

To this end, we introduce a Structure-guided Residual Fusion Module (SGRF) during the decoding stage, the structure is shown in Figure 5. Firstly, SGRF generates an initial edge modeling $\boldsymbol{I}_C \in \mathbb{R}^{H\times W\times 1}$ by the Canny operator (Liu et al., 2025) to guide the edge texture enhancement of the up-sampled fusion result. The process

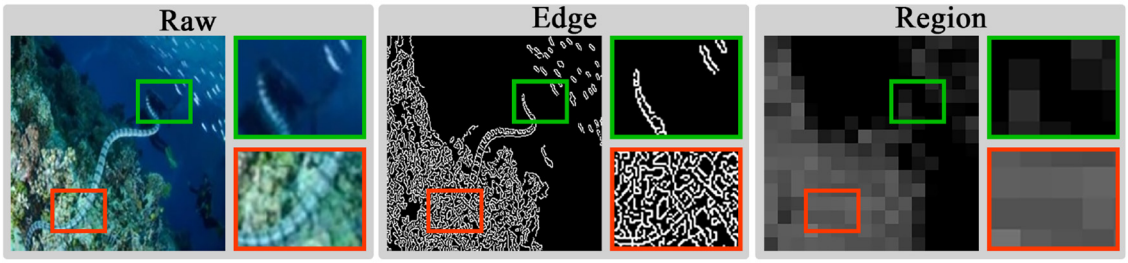

FIGURE 4
Modeling examples of different image regions. High-quality regions are framed in red and low-quality regions are framed in green.

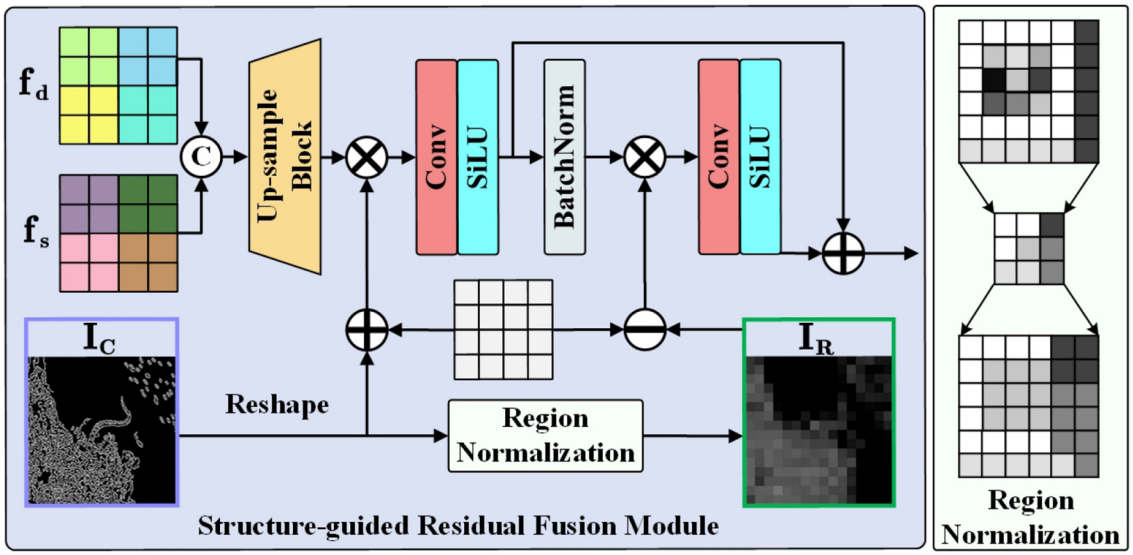

FIGURE 5
Structure of the structure-guided residual fusion module. With the region normalization shown on the right, we transform edge modeling I_C into region modeling I_R with rich feature distinctions.

is represented as follows:

$$f_{\text{edge}} = CS\left((\boldsymbol{\rho} + \boldsymbol{I}_C) * up\left(f_s, f_d\right)_{CAT}\right), \tag{5}$$

where f_s and f_d represent shallow encoded features and deep decoded features, and $up(\cdot)$ represents the up-sampling process. In addition, since black pixels will distort the original features, we set a constant matrix $\boldsymbol{\rho}$ to emphasize the edge information. All pixels of $\boldsymbol{\rho}$ are assigned a value of 1. After adding it to $\boldsymbol{I}_C$, the original black pixel regions (i.e., pixels with a value of 0) exert no influence on the original features, while the texture distribution regions enhance the original features. Then, SGRF learns the edge enhancement feature f_{edge} through convolutional layers and SiLU function (CS).

To fully utilize the structural modeling properties, we construct a clear illumination region division $\boldsymbol{I}_R \in \mathbb{R}^{H\times W\times 1}$ by region normalization to guide the visibility improvement in low-light regions. The procedure first divides $\boldsymbol{I}_C$ into multiple regions $r_1, r_2, ..., r_i, r_n$ of size $R\times R$, and then calculates the mean μ to replace the pixel value p of the region. The computation for the i-region is as follows:

$$\mu_i = \frac{1}{R^2}\sum_{p_{ij}r_i} p_{ij}. \tag{6}$$

In region modeling $\boldsymbol{I}_R$, regions with rich edge detail exhibit higher pixel values, which also implies low pixel values in low-light regions. We use the constant matrix $\boldsymbol{\rho}$ to invert the pixel distribution of the region. When $\boldsymbol{\rho}$ is subtracted from $\boldsymbol{I}_R$, pixel values in low-light areas are amplified while those in bright regions are suppressed, thereby guiding the network to focus on regions with light illumination loss. The relevant process is as follows:

$$f_{SGRF} = f_{\text{edge}} + CS\left((\boldsymbol{\rho} - \boldsymbol{I}_R) * BN\left(f_{\text{edge}}\right)\right), \tag{7}$$

where $BN(\cdot)$ represents the BatchNorm layer. Finally, SGRF combines the two enhancement features by residual connection, enabling the network to enhance degraded details and low-light regions in a more targeted manner.

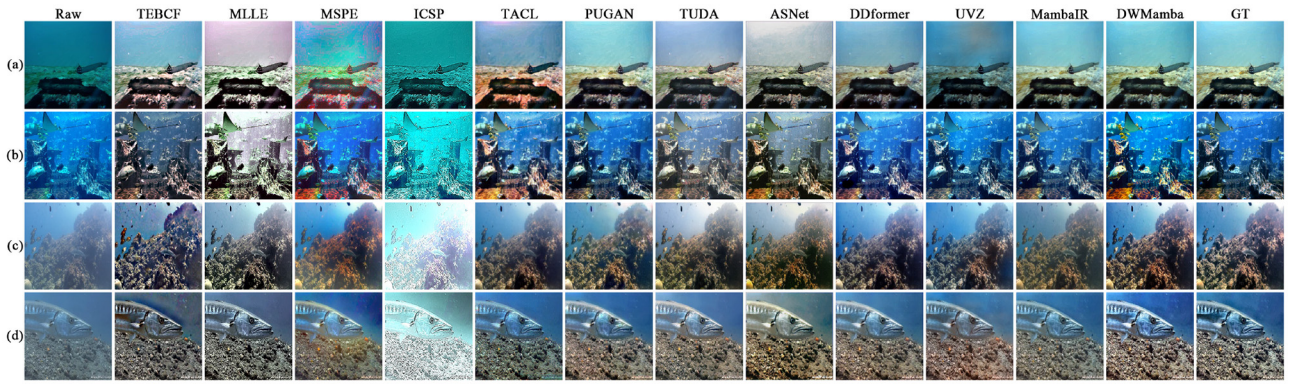

FIGURE 6
Comparison of enhancement results for all methods on the UIEB dataset. Each row represents a different scene, with variations in color and clarity reflecting distinct processing methods.

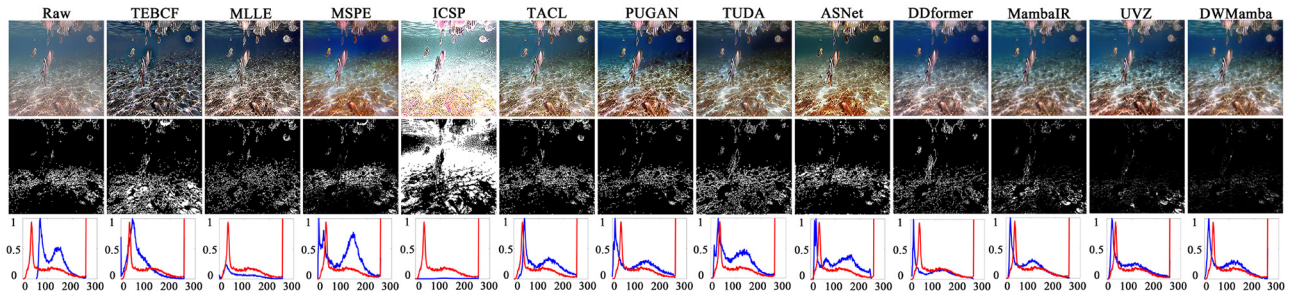

FIGURE 7
Comparison of residuals and red channel distribution between all enhanced images and GT. The first row shows the original images and the enhanced images from different methods, while the second row shows the corresponding residual images. The third row displays the red channel pixel distribution, with the blue and red lines representing the enhanced image and GT, respectively.

3.4 Loss functions

(1) Charbonnier Loss: As a variant of the L_1 loss, the charbonnier loss varies more gently when the gradient is large, which helps to measure high-frequency details (e.g., edges, textures, etc.) between images. In addition, this loss is more robust to outliers, ensuring training stability.

$$L_C = \sqrt{\|\boldsymbol{I}_E - \boldsymbol{I}_{GT}\|^2 + \varepsilon^2}, \tag{8}$$

where $\boldsymbol{I}_E$ and $\boldsymbol{I}_{GT}$ represent enhanced and ground truth (GT) images, and ε represents a constant used to stabilize the loss, which is set to 10^{-3}.

(2) Structural Similarity Index Measure (SSIM) Loss: Based on the measures of luminance, contrast and structure, the SSIM loss evaluates the image similarity from multiple perspectives, benefiting the network in generating visually enhanced images with improved brightness and structure.

$$L_S = 1 - \left(l\left(\boldsymbol{I}_E, \boldsymbol{I}_{GT}\right) * c\left(\boldsymbol{I}_E, \boldsymbol{I}_{GT}\right) * s\left(\boldsymbol{I}_E, \boldsymbol{I}_{GT}\right)\right), \tag{9}$$

where $l(\cdot,\cdot)$ is luminance similarity, $c(\cdot,\cdot)$ is contrast similarity, and $s(\cdot,\cdot)$ is structural similarity.

To train the proposed DWMamba, we adopt a combination of the two loss functions, aiming to encourage the network to generate enhanced results that are more similar to GT images in terms of detail and brightness. The total loss function is calculated as follows:

$$L_{\text{total}} = L_C + \lambda L_S, \tag{10}$$

where λ represents the balance weight, which we set to 0.5 by default.

4 Experiments

4.1 Experiment setup

(1) Implementation details: We implemented DWMamba with the PyTorch 2.1.1 framework on a machine with an Intel Core i7-12700KF CPU, two NVIDIA GeForce RTX 4090 GPUs, and 64 GB of memory. The model was trained by the ADAM optimizer, with the learning rate set to 0.0001, the epochs to 100, and the batch size to 8.

During testing, DWMamba was compared with eleven advanced methods, including four traditional methods: TEBCF (Yuan et al., 2021), MLLE (Zhang et al., 2022),

TABLE 1 Quantitative comparison for all methods on the UIEB dataset (optimal, sub-optimal).

	TEBCF	MLLE	MSPE	ICSP	TACL	PUGAN	TUDA	ASNet	DD former	UVZ	Mamba IR	DW Mamba
PSNR↑	20.48	17.25	21.92	12.61	22.79	25.67	22.82	21.68	23.53	23.61	22.52	26.37
SSIM↑	0.883	0.739	0.890	0.585	0.825	0.895	0.927	0.876	0.889	0.935	0.786	0.953
UCIQE↑	0.630	0.616	0.630	0.603	0.629	0.629	0.594	0.608	0.614	0.621	0.608	0.633
ENTROPY↑	7.496	7.547	7.405	6.945	7.557	7.572	7.594	7.599	7.491	7.511	7.425	7.591
CEIQ↑	3.433	3.469	3.369	3.159	3.457	3.474	3.479	3.482	3.412	3.441	3.384	3.488
Params (M)	-	-	-	-	11.37	95.65	31.36	0.755	7.580	5.387	0.626	0.430
FLOPs (G)	-	-	-	-	56.86	72.05	174.4	48.97	17.87	499.5	10.21	5.018
Times (s)	1.479	0.072	0.176	0.083	0.658	0.438	0.051	0.030	0.033	0.122	0.037	0.028

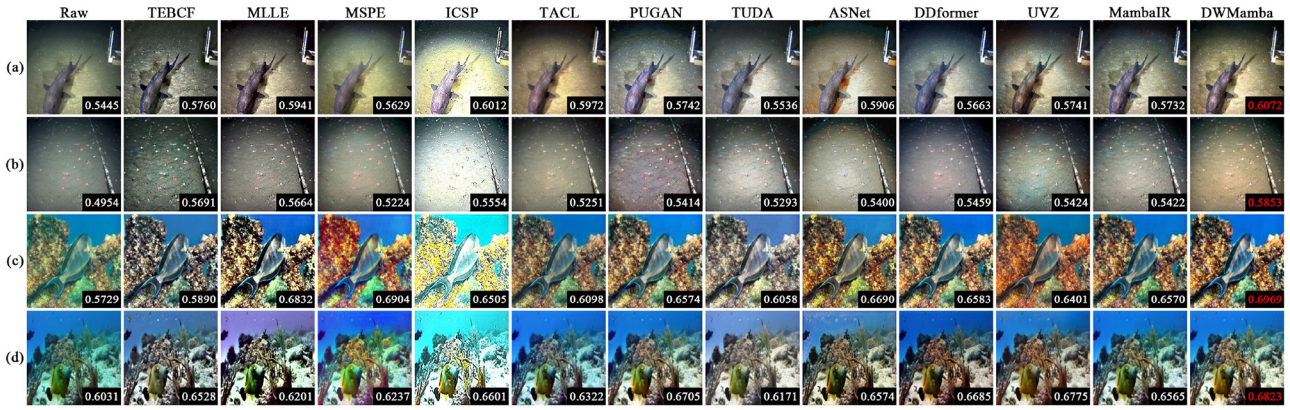

FIGURE 8
Comparison of enhancement results and UCIQE for all methods on the OceanDark and UFO datasets, with red values indicating optimal metrics. Each row (AD) depicts a different underwater scene.

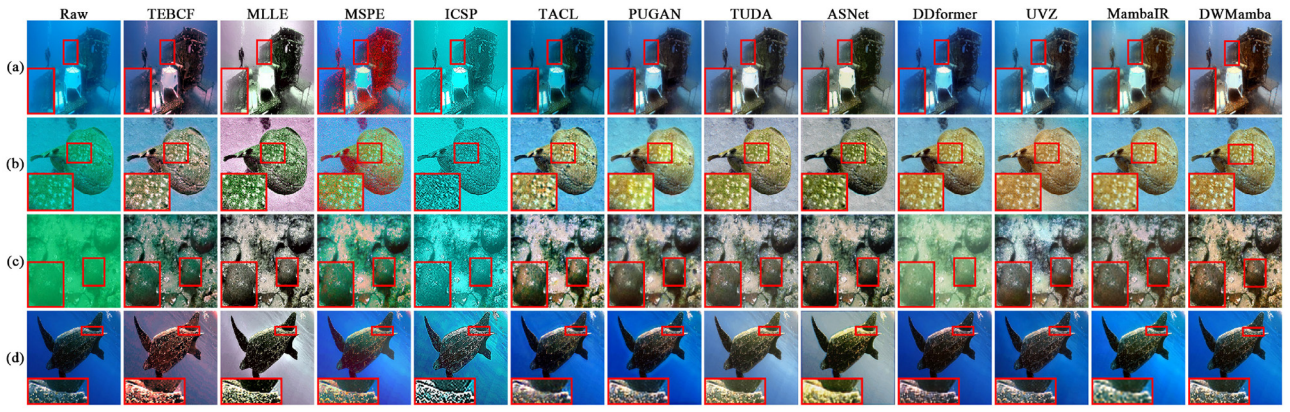

FIGURE 9
Detail enlargement comparison for all methods on the EUVP and LNRUD Datasets. Rows labeled (AD) show images of different underwater objects.

MSPE (Zhou et al., 2023b), and ICSP (Hou et al., 2024), six deep learning methods: TACL (Liu et al., 2022), PUGAN (Cong et al., 2023), TUDA (Wang et al., 2023), ASNet (Park and Eom, 2024), DDformer (Gao et al., 2024), and UVZ (Huang Z. et al., 2025), one mamba-based method: MambaIR (Guo et al., 2025). These methods were based on CNN, unsupervised techniques, Transformer, Mamba, and hybrid frameworks, respectively, and were experimented using the

TABLE 2 Quantitative comparison for all methods on the four non-reference datasets (**optimal**, **sub-optimal**).

Dataset	Metric	TEBCF	MLLE	MSPE	ICSP	TACL	PUGAN	TUDA	ASNet	DDformer	UVZ	MambaIR	DWMamba
OceanDark	UCIQE↑	0.584	0.583	0.568	0.596	0.591	0.580	0.556	0.573	0.572	0.588	0.574	0.598
	ENTROPY↑	7.377	7.456	7.621	6.929	7.709	7.559	7.581	7.528	7.570	7.619	7.683	7.735
	CEIQ↑	3.354	3.369	3.526	3.183	3.556	3.459	3.473	3.455	3.437	3.489	3.557	3.577
UFO	UCIQE↑	0.628	0.626	0.632	0.627	0.624	0.627	0.591	0.617	0.617	0.610	0.622	0.640
	ENTROPY↑	7.450	7.533	7.438	7.037	7.453	7.509	7.584	7.510	7.394	7.288	7.462	7.541
	CEIQ↑	3.397	3.425	3.375	3.252	3.378	3.420	3.465	3.424	3.337	3.279	3.398	3.445
EUVP	UCIQE↑	0.616	0.617	0.621	0.631	0.621	0.614	0.585	0.593	0.601	0.612	0.605	0.621
	ENTROPY↑	7.619	7.329	7.465	6.679	7.646	7.615	7.654	7.588	7.533	7.554	7.539	7.662
	CEIQ↑	3.502	3.330	3.401	3.159	3.510	3.489	3.516	3.474	3.428	3.449	3.450	3.530
LNRUD	UCIQE↑	0.626	0.615	0.628	0.610	0.624	0.625	0.593	0.613	0.612	0.608	0.618	0.634
	ENTROPY↑	7.504	7.552	7.463	7.102	7.533	7.551	7.611	7.557	7.454	7.393	7.489	7.580
	CEIQ↑	3.432	3.447	3.393	3.289	3.433	3.451	3.484	3.457	3.380	3.347	3.417	3.472
NPE	ENTROPY↑	7.286	6.706	7.399	6.172	7.265	7.461	7.309	7.247	7.152	7.348	7.542	7.546
	CEIQ↑	3.279	2.904	3.374	3.068	3.256	3.376	3.309	3.263	3.165	3.314	3.475	3.476

10.3389/fnbot.2025.1676787

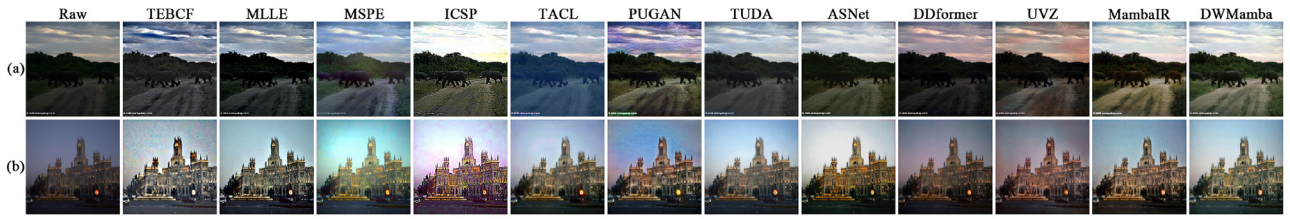

FIGURE 10
Comparison of enhancement results for all methods on the NPE dataset. Row (A, B) display different low-light outdoor scenes.

TABLE 3 Component settings and metrics comparison for all ablation configurations (optimal, sub-optimal).

Configuration	#1	#2	#3	#4	#5	#6	#7	#8
Base	✓	✓	✓	✓	✓	✓	✓	✓
DS-CMM		✓			✓	✓		✓
SF			✓		✓		✓	✓
SGRF				✓		✓	✓	✓
PSNR↑	24.88	25.70	25.90	26.17	26.22	26.36	26.11	26.37
SSIM↑	0.941	0.945	0.946	0.948	0.943	0.949	0.949	0.953
UCIQE↑	0.621	0.624	0.625	0.621	0.622	0.627	0.628	0.633
ENTROPY↑	7.571	7.547	7.570	7.535	7.545	7.556	7.585	7.591
CEIQ↑	3.479	3.465	3.475	3.454	3.456	3.468	3.486	3.488

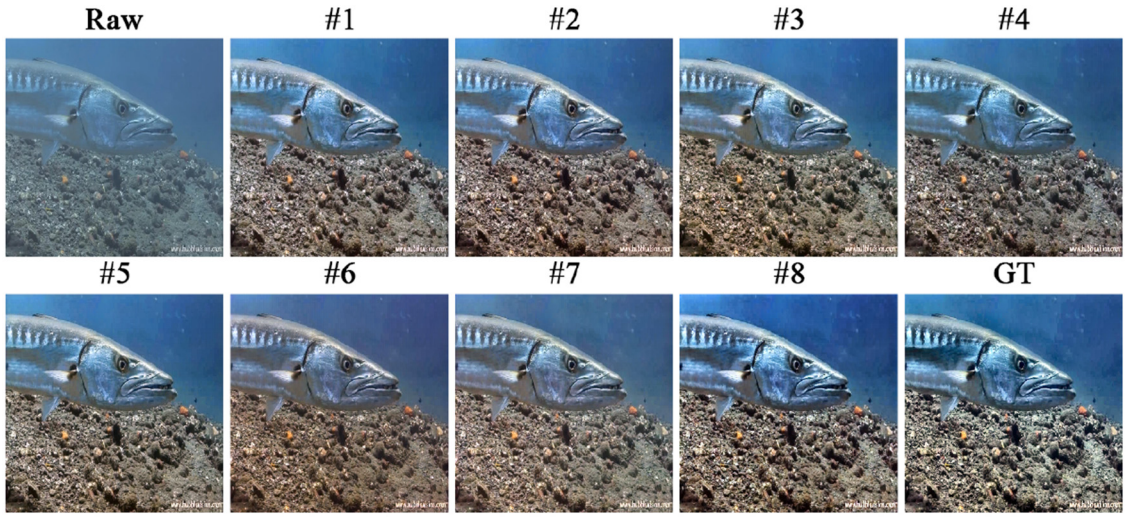

FIGURE 11
Enhanced results of all ablation configurations.

training models provided in the original paper. As MambaIR is not specifically designed for underwater image enhancement, we retrained it under same experimental configurations for fair comparison.

(2) Datasets and metrics: We randomly selected 810 pairs of real images from the UIEB (Li et al., 2019) dataset to train DWMamba, using the remaining 80 pairs as a reference test dataset. To verify the applicability, we also conducted extensive experiments on five non-reference datasets, namely OceanDark (Marques and Albu, 2020), UFO (Islam et al., 2020a), EUVP (Islam et al., 2020b), LNRUD (Ye et al., 2022), and NPE (Wang et al., 2013). The image sizes of all datasets were resized to 256 × 256 to facilitate the experiments.

For the referenced datasets, we used two full-reference metrics such as PSNR and SSIM to measure the similarity between the enhancement results and the GT images. Additionally, three non-reference metrics such as UCIQE (Yang and Sowmya, 2015), ENTROPY, and CEIQ (Fang

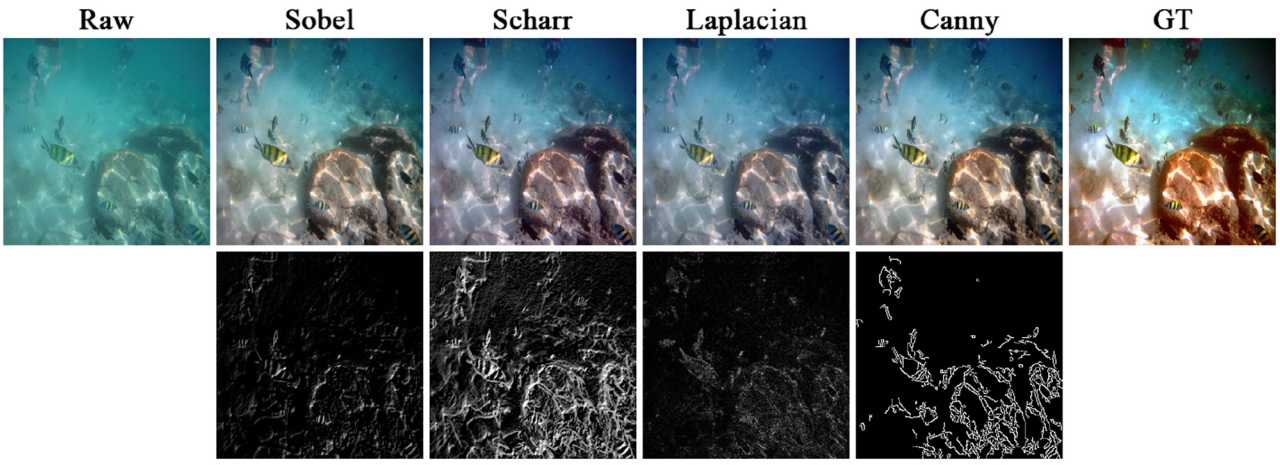

FIGURE 12
Comparison of results using different operators. The top shows the raw image and the enhanced results using different operators, and the bottom shows the corresponding edge detection results.

et al., 2014) were used to evaluate the quality of all enhancement results.

4.2 Comparison on the reference dataset

Figure 6 shows the enhancement examples of all methods on the UIEB dataset. The results from ICSP and ASNet exhibited varying degrees of overexposure, leading to oversaturated image colors. On the contrary, TEBCF and MLLE darkened the image brightness and masked the original image details, TEBCF and UVZ introduced artifacts in the enhancement process. The results from MSPE and TACL showed over-enhancement of the red channel, which disturbed the background color. The results of TUDA and MambaIR presented a relatively blurred scene, thereby reducing the color attractiveness of its results. PUGAN and DDformer effectively enhanced the image visibility, but the image details still need to be further enhanced. Our DWMamba demonstrated excellent performance in terms of color, clarity, and detail, particularly in the hazy scene (d), where the fish texture and scene colors were significantly enhanced.

In Figure 7, we further compared the similarity between the different methods and GT using residual figures and red channel distribution figures. The white regions in the residual figures directly reflects the differences between the enhanced images and GT, while the distribution figures highlight the most damaged channels. Compared with the residual figures of the raw images, the white regions from TEBCF, ICSP, and ASNet were more extensive, indicating a more significant difference from GT. In contrast, UVZ and DWMamba had the smallest white regions, and the red channel distributions more closely matched those of GT, proving that the enhancement results were highly consistent with GT images.

Table 1 provides a performance metrics comparison for all methods on the UIEB dataset. Our method stands out among all methods for its competitive performance, lower complexity, and higher operational efficiency. Notably, DWMamba achieves optimal values in most of the metrics and is slightly behind TUDA and ASNet in Entropy, demonstrating its well-balanced trade-off between efficiency and effectiveness.

TABLE 4 Metrics comparison of networks using different operators (**optimal**, **sub-optimal**).

	PSNR↑	SSIM↑	UCIQE↑	ENTROPY↑	CEIQ↑
Sobel	25.92	0.948	0.630	7.609	3.501
Scharr	26.29	0.950	0.628	7.583	3.482
Laplacian	26.27	0.949	0.627	7.532	3.450
Canny	26.37	0.968	0.633	7.591	3.488

4.3 Comparison on the non-reference datasets

In this section, we present the performance of all methods on the four non-reference datasets. Figure 8 visualizes the enhancement effects and UCIQE metrics of the different methods on the OceanDark and UFO datasets. For dark scenes in OceanDark, most methods overexposed or darkened the brightness and failed to reasonably enhance the scene visibility. In contrast, TACL, ASNet, and our DWMamba effectively enhanced the brightness and contrast of the scene. However, TACL resulted in textures blurred by hazy lighting, while ASNet excessively enhanced the red channel in shadow regions. The UFO dataset highlights the color deviation and hierarchical scenes. TEBCF, MLLE, and ASNet rendered distant scenes less colorful, with even MLLE producing an illogical purple background. MSPE, UVZ, and MambaIR significantly enhanced color saturation, but there were still noticeable color deviations. PUGAN and DWMamba not only effectively enhanced color attractiveness while retaining

TABLE 5 Metrics comparison of different baseline model and model size (optimal, sub-optimal).

	PSNR↑	SSIM↑	UCIQE↑	ENTROPY↑	CEIQ↑	Parameter (M)	FLOPs (G)
-r/w conv-S	22.22	0.916	0.583	7.358	3.330	0.324	4.400
-r/w vit-S	21.69	0.904	0.575	7.289	3.281	0.627	7.386
-r/w swin vit-S	22.59	0.921	0.592	7.352	3.337	0.694	7.876
DWMamba-S	26.37	0.953	0.633	7.591	3.488	0.404	5.018
-r/w conv-L	21.86	0.919	0.614	7.543	3.451	3.826	32.41
-r/w vit-L	22.30	0.910	0.581	7.338	3.312	8.139	62.45
-r/w swin vit-L	24.98	0.944	0.619	7.523	3.445	9.547	71.25
DWMamba-L	26.80	0.953	0.620	7.549	3.465	5.715	43.32

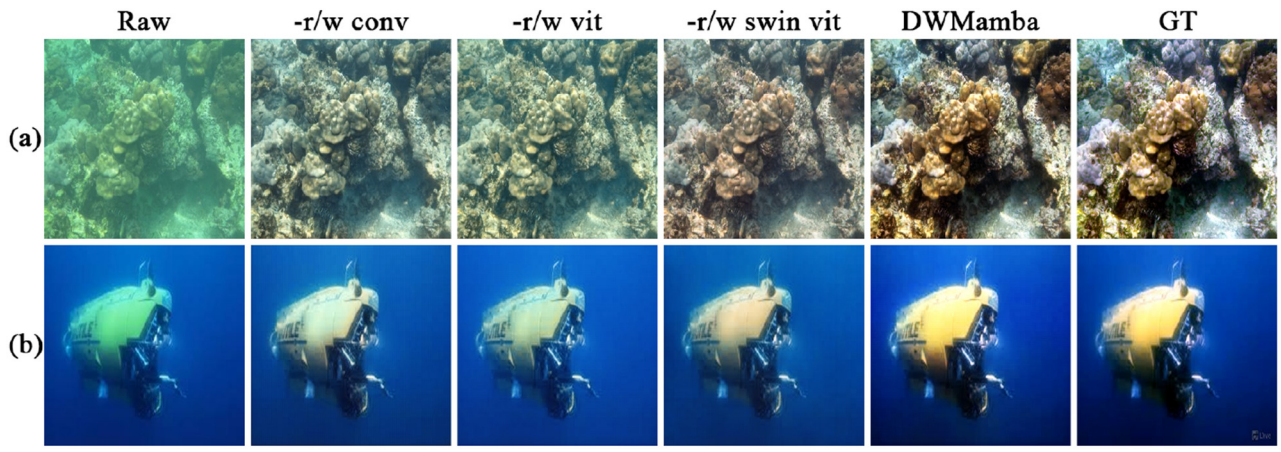

FIGURE 13
Comparison of results using baseline model. Row (A, B) display classic blue and green deviation scenes.

well-defined scene boundaries, and DWMamba achieved the best UCIQE with the most vibrant colors.

Figure 9 enlarges the enhanced details for the EUVP and LNRUD datasets. ASNet introduced unnatural yellow colors, significantly deviating from the true color distribution. The results of ICSP generally suffered from over-rendering of cyan, while MSPE and TEBCF excessively enhanced the red channel, and these over-saturated colors severely interfered with textural appearance. On the other hand, MLLE reduced the overall brightness of the image, making the enlarged details less clear. Although TACL and PUGAN considerably improved the visual effect of the underwater scene, the enlarged details were still affected by color deviation. The enhanced results of MambaIR demonstrated excellent color reproduction, but introduced other colors and blurred details. Compared to the above methods, TUDA and DWMamba significantly improved scene visibility, with the enlarged texture presenting clearer details. As shown in Table 2, our DWMamba achieved both optimal and sub-optimal values on the four unreferenced datasets, further proving its excellent enhancement effect and wide applicability. In addition, TUDA achieves outstanding performance in ENTROPY and CEIQ through excellent brightness enhancement and detail restoration.

In addition, Figure 10 presents the enhancement results of all methods on the low-light NPE dataset. It can be observed that most methods fail to effectively improve the overall brightness, resulting in unclear and dim outputs. The results of MSPE and ICSP suffer from oversaturation and overexposure, which obscure fine details. PUGAN, DDformer, and UVZ excessively amplify the red channel, leading to noticeable color distortions. In contrast, MambaIR and DWMamba consistently enhances both brightness and contrast in low-light scenes while preserving natural color tones without introducing artifacts. For quantitative evaluation, since UCIQE is specifically designed for underwater scenarios, only ENTROPY and CEIQ were adopted for the NPE dataset. The results demonstrate that DWMamba achieves the best performance on both metrics.

TABLE 6 Comparison of computational complexity at different image sizes.

Size	64	128	256	512	1024
FLOPs (G)	0.3136	1.2546	5.0182	20.073	80.2916
Time (s)	0.3830	0.4115	0.4116	0.5110	1.1534

4.4 Ablation study

In Table 3, we provided a detailed configuration and analysis of the key components of DWMamba, using the UNet architecture and the original VMamba as a baseline (#1). Figure 11 illustrates that the network with complete components produces more prominent texture details and more natural background color,

FIGURE 14
Generalizability of DWMamba across diverse poor-lighting scenarios. **(a–d)** Are low-light images from the NPE (Wang et al., 2013) dataset, **(e–h)** are overexposure images from the MIT (Bychkovsky et al., 2011) dataset, **(i–l)** are haze images from the O-HAZE (Ancuti et al., 2018) dataset, and **(m–p)** are sandy images from the WEAPD (Xiao et al., 2021) dataset. The results clearly demonstrate that DWMamba possesses domain-agnostic properties.

effectively avoiding color deviation and the introduction of other colors. In contrast, in the absence of SF, i.e. using the original SS2D, that is, when using the original SS2D, the background exhibits excessive enhancement of the red channel, while the overall brightness of the image decreases. Without DS-CMM or SGRF, the fish texture and the background exhibit an uncorrected blue deviation, highlighting deficiencies in region division. When lacking both DS-CMM and SF, i.e. using the original VSSM, the enhanced results showed blurred textures and dull colors. This highlights the importance of more comprehensive feature dependency modeling. As shown in Table 3, the complete network achieves the most outstanding metric performance, indicating that all three components are beneficial for the performance improvement of the baseline model. Notably, the absence of SGRF results in the most significant performance degradation, demonstrating its effectiveness in focusing on degradation details and illumination loss.

To verify the impact of different operators (Su et al., 2021) on the network performance, we adopted Sobel, Scharr, Laplacian, and Canny operators to construct the edge modeling of the raw image, and employed region normalization for subsequent region modeling. Figure 12 and Table 4 compare the enhancement results and metrics using different operators, respectively. Specifically, the network with the Sobel operator effectively improves the visual quality, but fails to completely eliminate the color deviation. The network using Scharr and Laplacian operators generated hazier enhancement results, accompanied by significant red over-enhancement. The network using the Canny operator generates superior results in terms of color and clarity, achieving optimal or sub-optimal evaluations across all metrics. Further analyzing the edge detection results, we found that the edge information from Sobel and Laplacian operators were relatively weak and cannot clearly reflect the rich details of the actual scene. Scharr and Canny operators generated more prominent edge information, but Scharr introduced substantial scene noise, which was detrimental to subsequent region division. Therefore, we prioritized the Canny operator for edge modeling.

Focusing on baseline model and model size, we conducted an ablation study as shown in Table 5. First, we fixed the remaining module configurations and replaced the core module ASSM with a single convolutional layer (-r/w conv), Vision Transformer (-r/w vit), and Swin Transformer (-r/w swin vit), where -r/w conv can also be regarded as the network without ASSM. Furthermore, we design small-scale configurations S: (layers: [1, 1, 2, 1, 1], dimensions: [24, 48, 96, 48, 24]) and large-scale configurations L: (layers. [2, 2, 9, 2, 2], dimensions: [48, 96, 192, 96, 48]). Among these, DWMamba-S is the main experimental architecture in this paper. Based on the relatively good performance of -r/w swin vit, DWMamba-S achieved a significant reduction of 41.8% in parameters and 36.2% in FLOPs, and increased PSNR by 16.7%. Similarly, DWMamba-L reduced parameters by 40.1%, FLOPs by 39.1%, and increased PSNR by 7.2%. The results in Figure 13 further demonstrate the visual advantages of DWMamba over other baseline models.

nOTABLY, although DWMamba exhibits significant advantages in all metrics, the marginal gains in performance diminish when scaled to large-scale model configurations, especially in the non-reference quality assessment metrics. In this regard, we believe that although more complex model

architectures enhance the model's ability to approximate GT images, the information loss of deeper architectures and the inherent limitations of GT data constrain the performance improvement. Therefore, in future research, striking a balance between model complexity and feature extraction capability will be the essential direction to further improve the performance of DWMamba and its similar models.

To assess the computational complexity of DWMamba, we compared the model's FLOPs and inference time across image inputs of varying sizes. As shown in Table 6, when the width and height of the input image doubled, the FLOPs increased nearly fourfold, aligning with the linear distribution of pixel count growth. This conclusively demonstrates that DWMamba inherits the linear complexity characteristic of the State Space Model, without introducing quadratic overhead.

4.5 Generalizability test

The complex and variable lighting conditions pose a challenge to IQI algorithm's generalization performance. Specifically, this challenge is reflected in multiple aspects: insufficient brightness due to light attenuation, overexposure from artificial light sources, image hazing due to transmission media, and visual interference from suspended particles, all of which significantly impact image quality. To this end, we conducted generalizability tests on four corresponding land datasets, and applied the DWMamba model without parameter tuning to these scenarios. As shown in Figure 14, for the low-light and overexposed scenarios, DWMamba demonstrates excellent adaptive brightness adjustment, making the dull colors and textures more attractive. For the hazy and sandy scenarios with more complex degradation, DWMamba does not completely eliminate the adverse effects of these extreme environments, but still shows significant visual improvement and plays a positive role in enhancing image visibility and scene understanding.

4.6 Limitation

While DWMamba achieves strong results across various real-world scenarios, it struggles with severe degradation. When images lose most of their color and detail, the enhancement task shifts toward generation, and our method fails to produce satisfactory visibility. Future research will focus on expanding multi-source information fusion for underwater scenes and advancing its real-world application.

5 Conclusion

In this paper, we present DWMamba, a novel adaptive Mamba network for image quality improvement. Specifically, we introduce a dual-stream channel monitoring mechanism and a soft fusion mechanism in the vision state space model, significantly enhancing its ability to capture global dependencies. In addition, we establish explicit structural and regional modeling to facilitate the targeted fusion of shallow and deep features. Extensive experiments across various datasets demonstrate that DWMamba exhibits not only excellent generalization but also significant quality improvements under diverse extreme lighting conditions. Notably, this multi-scenario applicability comes without high computational costs, highlighting its potential for practical applications.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

WF: Conceptualization, Writing – original draft, Writing – review & editing. XW: Investigation, Writing – original draft. CY: Conceptualization, Data curation, Writing – review & editing. LZ: Formal analysis, Visualization, Writing – review & editing. LF: Funding acquisition, Supervision, Writing – review & editing. ZH: Methodology, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the National Natural Science Foundation of China (61972064), the Science and Technology Program Joint Project of Liaoning Province (2024JH2/102600090), and the LiaoNing Revitalization Talents Program (XLYC1806006).

Conflict of interest

WF and LZ were employed by Beijing China Coal Mine Engineering Co. Ltd. XW was employed by Gansu Coal First Engineering Co. Ltd. CY was employed by Wanyi First Mine, China Energy Baotou Energy Co. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

An, S., Xu, L., Senior Member, I., Deng, Z., and Zhang, H. (2024). HFM: a hybrid fusion method for underwater image enhancement. *Eng. Appl. Artif. Intell.* 127:107219. doi: 10.1016/j.engappai.2023.107219

Ancuti, C. O., Ancuti, C., De Vleeschouwer, C., and Sbert, M. (2019). Color channel compensation (3c): A fundamental pre-processing step for image enhancement. *IEEE Trans. Image Proc.* 29, 2653–2665. doi: 10.1109/TIP.2019.2951304

Ancuti, C. O., Ancuti, C., Timofte, R., and De Vleeschouwer, C. (2018). "O-haze: a dehazing benchmark with real hazy and haze-free outdoor images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages* 754–762. doi: 10.1109/CVPRW.2018.00119

Bai, J., Yin, Y., and He, Q. (2024). Retinexmamba: retinex-based mamba for low-light image enhancement. *arXiv preprint arXiv:2405.03349*. doi: 10.1007/978-981-96-6596-9_30

Berman, D., Levy, D., Avidan, S., and Treibitz, T. (2020). Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 2822–2837. doi: 10.1109/TPAMI.2020.2977624

Bychkovsky, V., Paris, S., Chan, E., and Durand, F. (2011). "Learning photographic global tonal adjustment with a database of input/output image pairs," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE), 97–104. doi: 10.1109/CVPR.2011.5995332

Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., et al. (2021). "Pre-trained image processing transformer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12299–12310. doi: 10.1109/CVPR46437.2021.01212

Chen, Z., and Ge, Y. (2024). Mambauie&sr: unraveling the ocean's secrets with only 2.8 flops. *arXiv preprint arXiv:2404.13884*.

Cong, R., Yang, W., Zhang, W., Li, C., Guo, C.-L., Huang, Q., et al. (2023). Pugan: Physical model-guided underwater image enhancement using gan with dual-discriminators. *IEEE Trans. Image Proc.* 32, 4472–4485. doi: 10.1109/TIP.2023.3286263

Dong, C., Zhao, C., Cai, W., and Yang, B. (2024). O-mamba: O-shape state-space model for underwater image enhancement. *arXiv preprint arXiv:2408.12816*.

Fan, J., Jin, L., Li, P., Liu, J., Wu, Z.-G., and Chen, W. (2025). Coevolutionary neural dynamics considering multiple strategies for nonconvex optimization. *Tsinghua Sci. Technol.* 2025:9010120. doi: 10.26599/TST.2025.9010120

Fang, Y., Ma, K., Wang, Z., Lin, W., Fang, Z., and Zhai, G. (2014). No-reference quality assessment of contrast-distorted images based on natural scene statistics. *IEEE Signal Process. Lett.* 22, 838–842. doi: 10.1109/LSP.2014.2372333

Fu, Z., Lin, H., Yang, Y., Chai, S., Sun, L., Huang, Y., et al. (2022). "Unsupervised underwater image restoration: from a homology perspective," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 643–651. doi: 10.1609/aaai.v36i1.19944

Gao, Z., Yang, J., Jiang, F., Jiao, X., Dashtipour, K., Gogate, M., et al. (2024). Ddformer: Dimension decomposition transformer with semi-supervised learning for underwater image enhancement. *Knowl. Based Syst.* 297:111977. doi: 10.1016/j.knosys.2024.111977

Gu, A., and Dao, T. (2023). Mamba: linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.

Gu, A., Goel, K., and Ré, C. (2021). Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*.

Guan, M., Xu, H., Jiang, G., Yu, M., Chen, Y., Luo, T., et al. (2024). Watermamba: visual state space model for underwater image enhancement. *arXiv preprint arXiv:2405.08419*.

Guo, H., Li, J., Dai, T., Ouyang, Z., Ren, X., and Xia, S.-T. (2025). "Mambair: a simple baseline for image restoration with state-space model," in *European Conference on Computer Vision* (Springer), 222–241. doi: 10.1007/978-3-031-72649-1_13

He, K., Sun, J., and Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 2341–2353. doi: 10.1109/TPAMI.2010.168

Hou, G., Li, N., Zhuang, P., Li, K., Sun, H., and Li, C. (2024). Non-uniform illumination underwater image restoration via illumination channel sparsity prior. *IEEE Trans. Circ. Syst. Video Technol.* 34, 799–814. doi: 10.1109/TCSVT.2023.3290363

Huang, H., Jin, L., and Zeng, Z. (2025). A momentum recurrent neural network for sparse motion planning of redundant manipulators with majorization-minimization. *IEEE Trans. Ind. Electr.* 2025, 1–10. doi: 10.1109/TIE.2025.3566731

Huang, Z., Li, J., Hua, Z., and Fan, L. (2022). Underwater image enhancement via adaptive group attention-based multiscale cascade transformer. *IEEE Trans. Instrum. Meas.* 71, 1–18. doi: 10.1109/TIM.2022.3189630

Huang, Z., Wang, X., Xu, C., Li, J., and Feng, L. (2025). Underwater variable zoom: depth-guided perception network for underwater image enhancement. *Expert Syst. Appl.* 259:125350. doi: 10.1016/j.eswa.2024.125350

Islam, M. J., Luo, P., and Sattar, J. (2020a). Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. *arXiv preprint arXiv:2002.01155*.

Islam, M. J., Xia, Y., and Sattar, J. (2020b). Fast underwater image enhancement for improved visual perception. *IEEE Robot. Autom. Lett.* 5, 3227–3234. doi: 10.1109/LRA.2020.2974710

Jiang, N., Chen, W., Lin, Y., Zhao, T., and Lin, C.-W. (2022). Underwater image enhancement with lightweight cascaded network. *IEEE Trans. Multim.* 24, 4301–4313. doi: 10.1109/TMM.2021.3115442

Jiang, Q., Zhang, Y., Bao, F., Zhao, X., Zhang, C., and Liu, P. (2022). Two-step domain adaptation for underwater image enhancement. *Pattern Recognit.* 122:108324. doi: 10.1016/j.patcog.2021.108324

Jiang, Z., Li, Z., Yang, S., Fan, X., and Liu, R. (2022). Target oriented perceptual adversarial fusion network for underwater image enhancement. *IEEE Trans. Circ. Syst. Video Technol.* 32, 6584–6598. doi: 10.1109/TCSVT.2022.3174817

Jin, L., Wei, L., and Li, S. (2022). Gradient-based differential neural-solution to time-dependent nonlinear optimization. *IEEE Trans. Automat. Contr.* 68, 620–627. doi: 10.1109/TAC.2022.3144135

Kang, Y., Jiang, Q., Li, C., Ren, W., Liu, H., and Wang, P. (2023). A perception-aware decomposition and fusion framework for underwater image enhancement. *IEEE Trans. Circ. Syst. Video Technol.* 33, 988–1002. doi: 10.1109/TCSVT.2022.3208100

Khan, R., Mishra, P., Mehta, N., Phutke, S. S., Vipparthi, S. K., Nandi, S., et al. (2024). "Spectroformer: multi-domain query cascaded transformer network for underwater image enhancement," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1454–1463. doi: 10.1109/WACV57701.2024.00148

Li, C., Anwar, S., Hou, J., Cong, R., Guo, C., and Ren, W. (2021). Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Trans. Image Proc.* 30, 4985–5000. doi: 10.1109/TIP.2021.3076367

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., et al. (2019). An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Proc.* 29, 4376–4389. doi: 10.1109/TIP.2019.2955241

Liang, Z., Ding, X., Wang, Y., Yan, X., and Fu, X. (2021). Gudcp: generalization of underwater dark channel prior for underwater image restoration. *IEEE Trans. Circ. Syst. Video Technol.* 32, 4879–4884. doi: 10.1109/TCSVT.2021.3114230

Liu, C., Shu, X., Xu, D., and Shi, J. (2024a). Gccf: a lightweight and scalable network for underwater image enhancement. *Eng. Appl. Artif. Intell.* 128:107462. doi: 10.1016/j.engappai.2023.107462

Liu, M., Chen, L., Du, X., Jin, L., and Shang, M. (2021). Activated gradients for deep neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* 34, 2156–2168. doi: 10.1109/TNNLS.2021.3106044

Liu, M., Cui, Y., Ren, W., Zhou, J., and Knoll, A. C. (2025). Liednet: a lightweight network for low-light enhancement and deblurring. *IEEE Trans. Circ. Syst. Video Technol.* 35, 6602–6615. doi: 10.1109/TCSVT.2025.3541429

Liu, M., Li, Y., Chen, Y., Qi, Y., and Jin, L. (2024). A distributed competitive and collaborative coordination for multirobot systems. *IEEE Trans. Mobile Comput.* 23, 11436–11448. doi: 10.1109/TMC.2024.3397242

Liu, R., Jiang, Z., Yang, S., and Fan, X. (2022). Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Trans. Image Proc.* 31, 4922–4936. doi: 10.1109/TIP.2022.3190209

Liu, Y., Xiao, J., Guo, Y., Jiang, P., Yang, H., and Wang, F. (2024). Hsidmamba: exploring bidirectional state-space models for hyperspectral denoising. *arXiv preprint arXiv:2404.09697*.

Marques, T. P., and Albu, A. B. (2020). "L2uwe: a framework for the efficient enhancement of low-light underwater images using local contrast and multi-scale fusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 538–539. doi: 10.1109/CVPRW50498.2020.00277

Mishra, A. K., Kumar, M., and Choudhry, M. S. (2024). Fusion of multiscale gradient domain enhancement and gamma correction for underwater image/video enhancement and restoration. *Opt. Lasers Eng.* 178:108154. doi: 10.1016/j.optlaseng.2024.108154

Nathan, O. B., Levy, D., Treibitz, T., and Rosenbaum, D. (2025). "Osmosis: RGBD diffusion prior for underwater image restoration," in *European Conference on Computer Vision* (Springer), 302–319. doi: 10.1007/978-3-031-73033-7_17

Park, C. W., and Eom, I. K. (2024). Underwater image enhancement using adaptive standardization and normalization networks. *Eng. Appl. Artif. Intell.* 127:107445. doi: 10.1016/j.engappai.2023.107445

Peng, L., Zhu, C., and Bian, L. (2023). U-shape transformer for underwater image enhancement. *IEEE Trans. Image Proc.* 32, 3066–3079. doi: 10.1109/TIP.2023.3276332

Peng, Y.-T., Cao, K., and Cosman, P. C. (2018). Generalization of the dark channel prior for single image restoration. *IEEE Trans. Image Proc.* 27, 2856–2868. doi: 10.1109/TIP.2018.2813092

Shi, Y., Xia, B., Jin, X., Wang, X., Zhao, T., Xia, X., et al. (2025). Vmambair: visual state space model for image restoration. *IEEE Trans. Circ. Syst. Video Technol.* 35, 5560–5574. doi: 10.1109/TCSVT.2025.3530090

Su, Z., Liu, W., Yu, Z., Hu, D., Liao, Q., Tian, Q., et al. (2021). "Pixel difference networks for efficient edge detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5117–5127. doi: 10.1109/ICCV48922.2021.00507

Wang, H., Sun, S., Chang, L., Li, H., Zhang, W., Frery, A. C., et al. (2024). Inspiration: a reinforcement learning-based human visual perception-driven image enhancement paradigm for underwater scenes. *Eng. Appl. Artif. Intell.* 133:108411. doi: 10.1016/j.engappai.2024.108411

Wang, S., Zheng, J., Hu, H.-M., and Li, B. (2013). Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Proc.* 22, 3538–3548. doi: 10.1109/TIP.2013.2261309

Wang, Y., Hu, S., Yin, S., Deng, Z., and Yang, Y.-H. (2024). A multi-level wavelet-based underwater image enhancement network with color compensation prior. *Expert Syst. Appl.* 242:122710. doi: 10.1016/j.eswa.2023.122710

Wang, Z., Shen, L., Xu, M., Yu, M., Wang, K., and Lin, Y. (2023). Domain adaptation for underwater image enhancement. *IEEE Trans. Image Proc.* 32, 1442–1457. doi: 10.1109/TIP.2023.3244647

Xiao, H., Zhang, F., Shen, Z., Wu, K., and Zhang, J. (2021). Classification of weather phenomenon from images by using deep convolutional neural network. *Earth Space Sci.* 8:e2020EA001604. doi: 10.1029/2020EA001604

Yang, M., and Sowmya, A. (2015). An underwater color image quality evaluation metric. *IEEE Trans. Image Proc.* 24, 6062–6071. doi: 10.1109/TIP.2015.2491020

Ye, T., Chen, S., Liu, Y., Ye, Y., Chen, E., and Li, Y. (2022). "Underwater light field retention: neural rendering for underwater imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 488–497. doi: 10.1109/CVPRW56347.2022.00064

Yuan, J., Cai, Z., and Cao, W. (2021). TEBCF: real-world underwater image texture enhancement model based on blurriness and color fusion. *IEEE Trans. Geosci. Rem. Sens.* 60, 1–15. doi: 10.1109/TGRS.2021.3110575

Yue, L., Yunjie, T., Yuzhong, Z., Hongtian, Y., Lingxi, X., Yaowei, W., et al. (2024). Vmamba: visual state space model. *arXiv preprint arXiv:2401.10166*.

Zhang, D., Wu, C., Zhou, J., Zhang, W., Lin, Z., Polat, K., et al. (2024a). Robust underwater image enhancement with cascaded multi-level sub-networks and triple attention mechanism. *Neural Netw.* 169, 685–697. doi: 10.1016/j.neunet.2023.11.008

Zhang, S., Duan, Y., Li, D., and Zhao, R. (2024b). Mamba-uie: Enhancing underwater images with physical model constraint. *arXiv preprint arXiv:2407.19248*.

Zhang, S., Zhao, S., An, D., Li, D., and Zhao, R. (2024c). Liteenhancenet: A lightweight network for real-time single underwater image enhancement. *Expert Syst. Appl.* 240:122546. doi: 10.1016/j.eswa.2023.122546

Zhang, W., Zhou, L., Zhuang, P., Li, G., Pan, X., Zhao, W., et al. (2024d). Underwater image enhancement via weighted wavelet visual perception fusion. *IEEE Trans. Circ. Syst. Video Technol.* 34, 2469–2483. doi: 10.1109/TCSVT.2023.3299314

Zhang, W., Zhuang, P., Sun, H.-H., Li, G., Kwong, S., and Li, C. (2022). Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Trans. Image Proc.* 31, 3997–4010. doi: 10.1109/TIP.2022.3177129

Zhao, C., Cai, W., Dong, C., and Hu, C. (2024). "Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8281–8291. doi: 10.1109/CVPR52733.2024.00791

Zhou, H., Wu, X., Chen, H., Chen, X., and He, X. (2024a). Rsdehamba: Lightweight vision mamba for remote sensing satellite image dehazing. *arXiv preprint arXiv:2405.10030*.

Zhou, J., Gai, Q., Zhang, D., Lam, K.-M., Zhang, W., and Fu, X. (2024b). Iacc: cross-illumination awareness and color correction for underwater images under mixed natural and artificial lighting. *IEEE Trans. Geosci. Rem. Sens.* 62, 1–15. doi: 10.1109/TGRS.2023.3346384

Zhou, J., Liu, Q., Jiang, Q., Ren, W., Lam, K.-M., and Zhang, W. (2023a). Underwater camera: Improving visual perception via adaptive dark pixel prior and color correction. *Int. J. Comput. Vision* 2023, 1–19. doi: 10.1007/s11263-023-01853-3

Zhou, J., Pang, L., Zhang, D., and Zhang, W. (2023b). Underwater image enhancement method via multi-interval subhistogram perspective equalization. *IEEE J. Oceanic Eng.* 48, 474–488. doi: 10.1109/JOE.2022.3223733

Zhou, J., Sun, J., Li, C., Jiang, Q., Zhou, M., Lam, K.-M., et al. (2024c). Hclr-net: Hybrid contrastive learning regularization with locally randomized perturbation for underwater image enhancement. *Int. J. Comput. Vision* 132, 4132–4156. doi: 10.1007/s11263-024-01987-y

Zhou, J., Wang, S., Lin, Z., Jiang, Q., and Sohel, F. (2024d). A pixel distribution remapping and multi-prior retinex variational model for underwater image enhancement. *IEEE Trans. Multim.* 26, 7838–7849. doi: 10.1109/TMM.2024.3372400

Zhou, J., Wang, Y., Li, C., and Zhang, W. (2023c). Multicolor light attenuation modeling for underwater image restoration. *IEEE J. Oceanic Eng.* 48, 1322–1337. doi: 10.1109/JOE.2023.3275615

Zhuang, P., Wu, J., Porikli, F., and Li, C. (2022). Underwater image enhancement with hyper-laplacian reflectance priors. *IEEE Trans. Image Proc.* 31, 5442–5455. doi: 10.1109/TIP.2022.3196546