



ARTICLE

Spectral-Semantic Decoupled Rectification Network for Non-Uniform Underwater Image Restoration

Jinshuo Ma, Yang Li*, Can Guo, Wen Gao and Ruiming Zhang

School of Mathematical Sciences, Dalian Minzu University, Dalian, China

*Corresponding Author: Yang Li. Email: liyang@dlmu.edu.cn

Received: 22 May 2026; Accepted: 02 July 2026; Published: 13 August 2026

ABSTRACT: Underwater image restoration is severely hindered by a tightly coupled degradation process: wavelength-dependent spectral distortion combined with non-uniform, multi-scale spatial scattering. Standard Convolutional Neural Networks (CNNs) and rigid physical priors frequently fail in these dynamic environments, limited by restricted receptive fields, overlooked inter-channel spectral correlations, and severe over-enhancement in photon-starved regions. To break this bottleneck, we propose the Phased Feature Rectification Network (PFR-Net), a decoupled architecture that transforms the ill-posed restoration task into a sequential global spectral calibration and deep semantic refinement paradigm. In the first phase, an efficient Multi-Layer Perceptron (MLP)-based Color Mapping (MLP-CM) module acts as a front-end calibrator to rectify global color casts with minimal parameter overhead. Following this, semantic structures are systematically reconstructed through Dilated Multi-Scale Residual Blocks (DMSRB), which are designed to prevent the truncation of delicate structural gradients and dynamically stabilize content-adaptive feature propagation. To bridge the gap between spatial processing and spectral consistency, we introduce Cross-Covariance Global Attention (XCA), capturing long-range inter-channel dependencies with a linear spatial complexity of $\mathcal{O}(NC^2)$. Finally, we formulate an illumination-aware Adaptive Dark Channel Prior (DCP) loss that dynamically modulates penalty constraints based on global scene radiance, robustly suppressing backscatter while shielding dark regions from visual artifacts. Extensive zero-shot cross-dataset evaluations validate the effectiveness of our paradigm; PFR-Net achieves competitive performance, including a PSNR of 27.625 and an SSIM of 0.885 on the rigorous LSUI400 benchmark. By synergizing physically grounded modulations with decoupled feature learning, PFR-Net provides a robust and high-fidelity solution for underwater visual perception.

KEYWORDS: Spectral-semantic decoupling; adaptive dark channel prior; cross-covariance global attention; non-uniform degradation; underwater image restoration

1 Introduction

High-fidelity underwater vision is a cornerstone for advancing marine engineering and deep-sea exploration [1,2]. Recent advances—from physical model-guided enhancement [3] and diffusion-based restoration [4] to comprehensive deep learning surveys [5] and studies on the benefits of Underwater Image Enhancement (UIE) for downstream perception [6]—underscore the growing demand for robust visual perception in aquatic environments. However, light propagation underwater is governed by severe wavelength-dependent attenuation and multi-path scattering from suspended particulates [7], inducing pervasive spectral shifts and low-contrast distortions that degrade downstream perception algorithms [6,8]. Traditional restoration strategies rely on hand-crafted physical priors such as the Dark Channel Prior

(DCP) [9] to invert this degradation, yet their rigid assumptions often produce severe artifacts and poor generalization across diverse water types [1].

Deep Convolutional Neural Networks (CNNs) have emerged as powerful alternatives, yet standard architectures face intrinsic limitations: their localized receptive fields often fail to capture the expansive structural degradation in underwater scenes [1,10]. Furthermore, a more profound challenge is “spectral discontinuity”—where different wavelengths attenuate at substantially different rates. Most existing spatial-centric networks, including recent Mamba-based architectures [11,12], focus primarily on spatial dependencies, while only a few works [13] explicitly model inter-channel spectral correlations. Consequently, these methods struggle to simultaneously correct color casts and recover fine structural details, particularly in scenes where spectral distortion and spatial scattering are deeply coupled. To address these limitations, we prioritize visual restoration capability, accepting higher computational cost to establish a high-capacity baseline for offline oceanographic analysis.

We propose PFR-Net, a Phased Feature Rectification Network that tackles underwater degradation through a “divide-and-conquer” strategy. Recognizing that restoring color and semantic structure within an entangled feature space is suboptimal, we decouple the restoration into two phases. Phase one performs pixel-wise spectral realignment via an efficient Multi-Layer Perceptron (MLP)-based Color Mapping (MLP-CM) module [14]. Phase two performs semantic refinement using a Nonlinear Activation Free Gate (NAFG) mechanism [15] to dynamically gate structural details. Following both phases, a Cross-Covariance Global Attention (XCA) module [16] captures global channel dependencies on the fully rectified features at linear spatial complexity $\mathcal{O}(NC^2)$.

To broaden the receptive field without excessive parameter proliferation, we propose the Dilated Multi-Scale Residual Block (DMSRB) with multi-branch dilated convolutions and Mish activation [17] for smooth, non-monotonic gradient flow. To overcome the rigidity of physical priors, we introduce an illumination-aware Adaptive DCP loss that dynamically modulates penalty weights based on global scene radiance—ensuring rigorous backscatter suppression in well-lit regions while safeguarding dark regions from artifacts.

The primary contributions of this work are:

- PFR-Net, a decoupled two-phase architecture that separates spectral calibration from semantic refinement for underwater image restoration.
- The Dilated Multi-Scale Residual Block (DMSRB) with Mish activation, expanding the receptive field to capture multi-scale structural semantics.
- Cross-Covariance Global Attention (XCA) integrated into the phased pipeline, modeling global channel dependencies at linear spatial complexity.
- An illumination-aware Adaptive DCP loss that dynamically modulates physical constraints to balance dehazing and structural preservation.

2 Related Work

2.1 Traditional Underwater Image Enhancement

Early methods relied on idealized optical assumptions and deterministic physical models. The Dark Channel Prior (DCP) [9], while seminal for terrestrial haze, yields suboptimal results when directly applied underwater. Domain-specific adaptations such as the Underwater Dark Channel Prior (UDCP) [18] omit the rapidly attenuating red channel for improved transmission estimation; related formulations were explored in maximum attenuation identification [19]. Statistics-based methods including Histogram Equalization and Retinex-based fusion [20] enhance contrast but, operating independently of optical physics, frequently

amplify noise and introduce color deviations. Consequently, traditional techniques lack the robustness to handle diverse real-world underwater environments [1].

2.2 Deep Learning for UIE

Data-driven UIE methods can be categorized along two axes: architectural paradigm and degradation modeling strategy. Architecturally, early CNN-based methods (WaterNet [21], FUnIE-GAN [2]) established end-to-end baselines but were limited by local receptive fields. Multi-scale architectures (Ucolor [22], RAUNE-Net [23]), lightweight frameworks [10], U-Net variants [24], and hybrid CNN-Transformer designs [25] partially alleviated this limitation, while physical model-guided methods [26] complement data-driven learning. More recently, state-space models (MambaIR [27], PixMamba [11], UWMambaNet [12]) and diffusion-based approaches (e.g., UDBE [28], DACA-Net [29]) have emerged; however, Mamba methods remain spatial-centric and diffusion models incur substantial latency. Regarding degradation strategy, most methods adopt monolithic processing—addressing color and structure in an entangled feature space—overlooking that spectral absorption and spatial scattering are orthogonal mechanisms. PFR-Net bridges these axes via phased decoupling: spectral calibration (Phase 1) followed by semantic refinement (Phase 2), leveraging channel-wise attention (XCA) and multi-scale dilated convolutions (DMSRB). Mamba- and diffusion-based methods are excluded from our quantitative comparison; a systematic benchmarking study incorporating these emerging paradigms is identified as an important direction for future work.

3 Proposed Method

Before presenting the architectural details, we clarify the provenance of individual components. Several building blocks are adapted from existing studies: the MLP-CM module draws on [14], XCA on [16], NAFG on [15], Mish on [17], CBAM on [30], and DCP on [9]. Our original contributions are: (i) the two-phase decoupled architecture separating spectral calibration from semantic refinement; (ii) the DMSRB design with multi-branch dilated configuration and Mish-driven gradient preservation; (iii) the integration of channel-wise XCA into a phased pipeline for global spectral dependencies at linear complexity; and (iv) the illumination-aware Adaptive DCP loss, a formulation not previously explored in this domain.

3.1 Overall Architecture

PFR-Net employs a symmetric encoder-decoder architecture (Fig. 1). The encoder begins with the Initial Spatial Perception Module (ISPM): three 3×3 convolutions with reflection padding (effective 7×7 receptive field), Instance Normalization, and ReLU. The Attentional Feature Compression Module (AFCM) then downsamples by a factor of 2 via parallel Max and Average pooling, fuses the concatenated features with a 1×1 convolution and LeakyReLU, applies CBAM [30] for adaptive spatial/channel weighting, and further compresses by $2\times$ via strided convolution, yielding feature maps of $\frac{H}{4} \times \frac{W}{4} \times C_3$ for the bottleneck.

The bottleneck consists of the Phased Feature Rectification and Learning Module (PFR-LM) with 30 cascaded DMSRB blocks performing hierarchical color-semantic rectification. The decoder upsamples bottleneck features by $2\times$, then the Hierarchical Reconstruction Module (HRM) restores full resolution via bilinear interpolation and 3×3 convolutions with Instance Normalization (chosen to mitigate checkerboard artifacts). Finally, the Refined Mapping and Smoothing Module (RMSM) projects to RGB space via a 3×3 convolution with Tanh activation.

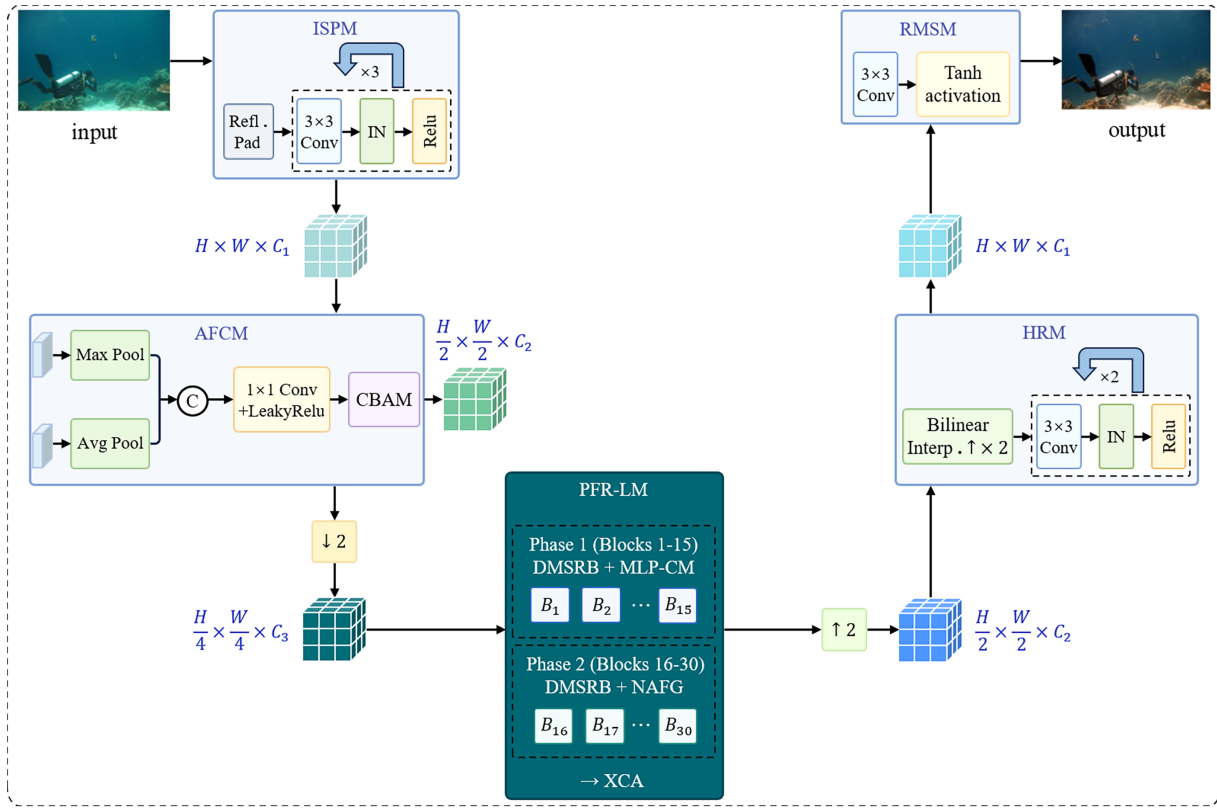


Figure 1: Overall architecture of the proposed PFR-Net. The network takes degraded underwater images as input and sequentially extracts and refines features through multiple specialized modules (ISPM, AFCM, PFR-LM, HRM, and RMSM) to output high-quality enhanced images.

3.2 Phased Feature Rectification and Learning Module (PFR-LM)

Underwater degradation is governed by two orthogonal physical processes: wavelength-dependent absorption induces global spectral shifts, while particle scattering degrades local structures. Processing both simultaneously in an entangled feature space leads to suboptimal convergence. To decouple this, the PFR-LM partitions 30 consecutive DMSRB blocks into two phases (Fig. 2).

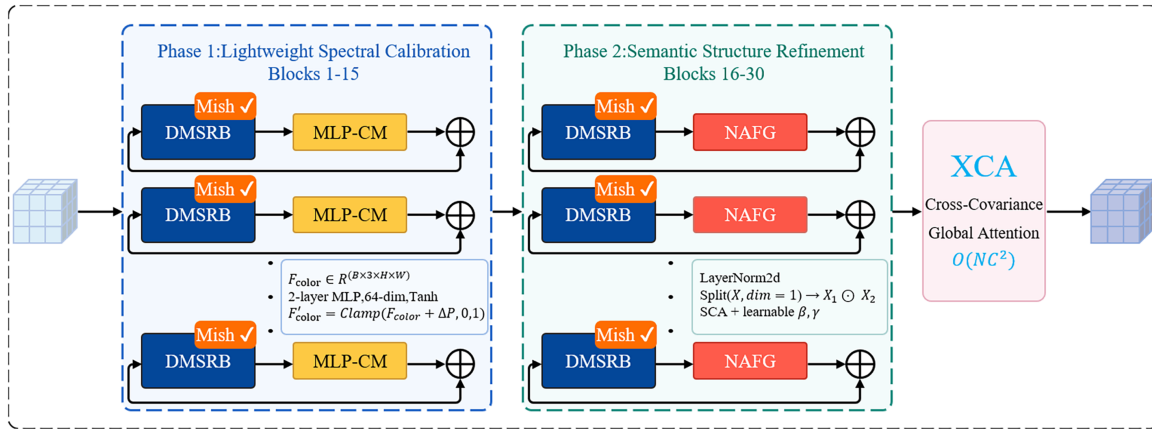


Figure 2: Architecture of the proposed PFR-LM module, illustrating the two-phase refinement process.

Phase 1: Lightweight Spectral Calibration (Blocks 1–15). Applying dense spatial convolutions across all channels solely for color correction is computationally redundant. Instead, each of the 15 Phase-1 DMSRB blocks integrates an independent MLP-based Color Mapping (MLP-CM) module [14] for pixel-wise spectral adjustment. Within each block, the three leading feature channels $F_{color} \in \mathbb{R}^{B \times 3 \times H \times W}$ —structurally designated for colorimetric information—are normalized to $[0,1]$ and processed by a compact MLP ($3 \rightarrow 64 \rightarrow 64 \rightarrow 64 \rightarrow 3$, with ReLU after the first layer and Tanh thereafter) to learn a color residual ΔP :

$$F'_{color} = \text{Clamp}(F_{color} + \Delta P, 0, 1). \quad (1)$$

The calibrated F'_{color} is rescaled to $[-1, 1]$ and concatenated with the remaining semantic channels. Each MLP-CM instance contains $3 \times 64 + 64 + 2 \times (64 \times 64 + 64) + 64 \times 3 + 3 = 8771$ parameters; across 15 blocks the total Phase-1 overhead is approximately $15 \times 8.8K \approx 132K$.

Phase 2: Semantic Structure Refinement (Blocks 16–30). With spectral bias removed, blocks 16–30 focus on structural recovery. All 30 DMSRBs employ Mish activation to prevent gradient truncation; Phase 2 additionally augments each block with the Nonlinear Activation Free Gate (NAFG) [15] for content-adaptive gating. NAFG uses an improved LayerNorm2d and a SimpleGate mechanism that splits the tensor X channel-wise and multiplies the halves:

$$X_1, X_2 = \text{Split}(X, \text{dim} = 1), \quad (2)$$

$$\text{SimpleGate}(X) = X_1 \odot X_2. \quad (3)$$

Coupled with Simplified Channel Attention (SCA) pooling and learnable residual scaling (β, γ), this parameter-free gating selectively amplifies high-frequency components for detail-rich reconstruction.

3.3 Dilated Multi-Scale Residual Block (DMSRB)

Underwater degradation spans heterogeneous spatial scales—from fine texture loss to broad illumination shifts. To capture this hierarchy, we deploy 30 identical DMSRB blocks as the backbone feature extractors. As shown in Fig. 3, each DMSRB employs three parallel convolutional branches with dilation rates $d = 1, 2, 4$, targeting fine local textures (3×3 receptive field), intermediate semantics (5×5), and broad contextual information (9×9), respectively.

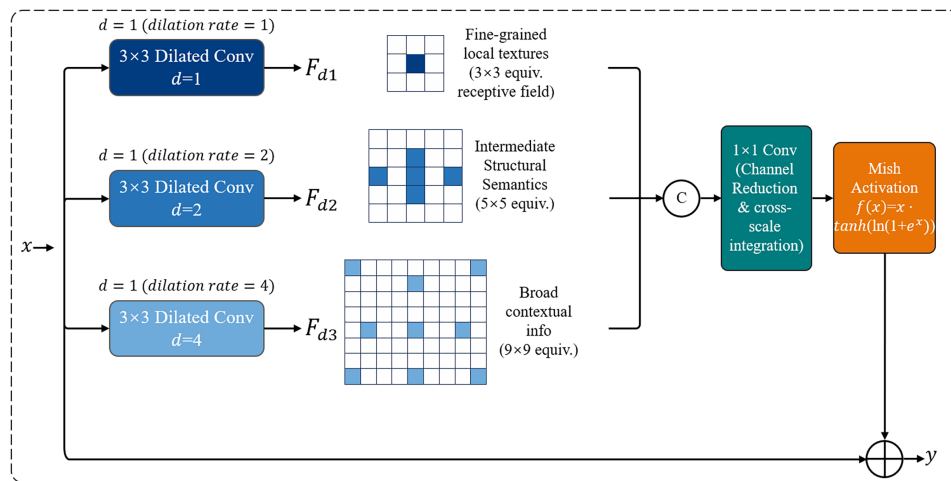


Figure 3: Schematic diagram of the Dilated Multi-Scale Residual Block (DMSRB). It employs three parallel branches with varying dilation rates ($d = 1, 2, 4$) to synergistically capture localized details and broad contextual features.

The multi-scale representations are fused via:

$$y = \mathcal{F}_{1 \times 1}(\text{Concat}[F_{d1}, F_{d2}, F_{d3}]) + x, \quad (4)$$

where $\mathcal{F}_{1 \times 1}$ is a 1×1 convolution for channel reduction and cross-scale integration.

A critical bottleneck in deep restoration is the irreversible loss of subtle gradients caused by standard activations. We therefore replace ReLU with Mish [17] across all DMSRB internal convolutions:

$$f(x) = x \cdot \tanh(\ln(1 + e^x)). \quad (5)$$

As shown in Fig. 4, ReLU's hard zero threshold truncates negative features, causing zero gradients that obliterate faint underwater structures. Mish's C^∞ smooth, non-monotonic profile permits small negative activations to propagate, preventing gradient death and preserving delicate textures critical for underwater restoration.

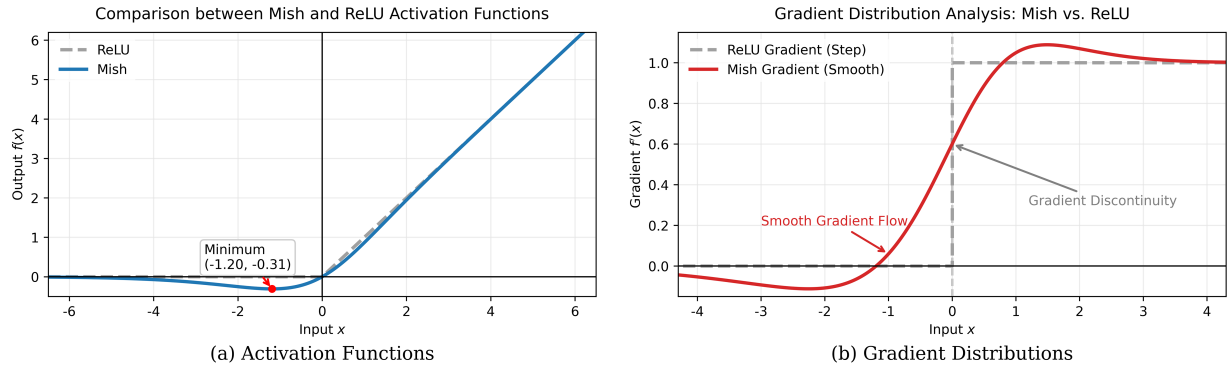


Figure 4: Mathematical comparison between Mish and standard ReLU. (a) Activation function; (b) Gradient distributions. Unlike ReLU's abrupt gradient truncation, Mish maintains C^∞ smoothness. In underwater restoration, this preserves subtle textural gradients—fine sand ripples, distant object silhouettes, low-contrast biological structures—that ReLU's hard zero threshold would obliterate.

3.4 Cross-Covariance Global Attention (XCA)

While DMSRB captures local multi-scale semantics, underwater degradation also exhibits global spectral shifts requiring long-range modeling. Standard spatial self-attention costs $\mathcal{O}(N^2C)$ ($N = H \times W$), prohibitive for high-resolution imagery. XCA [16] addresses this by computing cross-covariance along the channel dimension (Fig. 5). Given projected Query (Q), Key (K), and Value (V) matrices, we apply L_2 normalization along the channel dimension:

$$\hat{Q} = \frac{Q}{\|Q\|_{2,C}}, \quad \hat{K} = \frac{K}{\|K\|_{2,C}}, \quad (6)$$

stabilizing feature magnitudes against illumination-induced amplitude variations. The cross-covariance matrix $A \in \mathbb{R}^{C \times C}$ is obtained via:

$$A = \text{Softmax}\left(\frac{\hat{Q}^T \hat{K}}{\tau}\right), \quad (7)$$

with a learnable temperature τ adaptively controlling attention sharpness. The output is reconstructed as:

$$\text{Output} = VA^T, \quad (8)$$

reducing complexity to $\mathcal{O}(NC^2)$. We use num_heads = 8 with per-head τ initialized to 1.0. XCA is positioned after the PFR-LM bottleneck, operating on fully rectified features before the decoder.

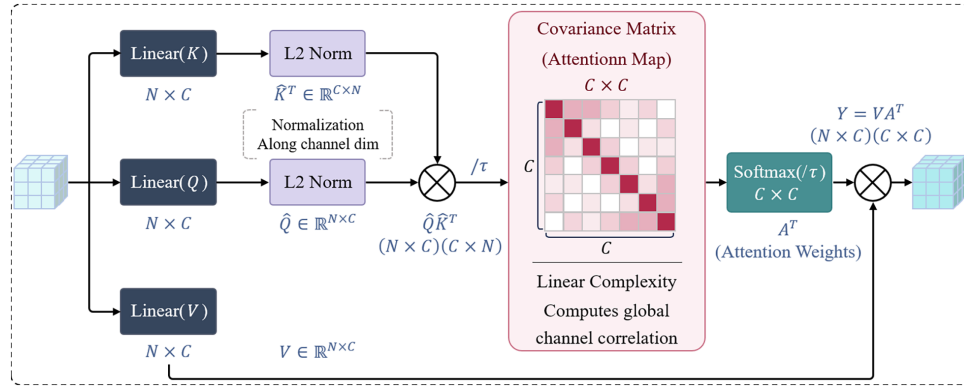


Figure 5: Schematic illustration of the cross-covariance global attention (XCA) mechanism. By computing cross-covariance along the channel dimension, the module models global spectral correlations at linear spatial complexity.

3.5 Loss Function with Adaptive Physical Constraints

We formulate a multi-task objective combining pixel-level fidelity, structural preservation, and physics-based regularization. The L_1 loss provides a baseline for color and illumination reconstruction:

$$\mathcal{L}_1 = \frac{1}{N} \sum_{i=1}^N \|I_{en}(i) - I_{gt}(i)\|_1. \quad (9)$$

To preserve high-frequency textures degraded by scattering, we incorporate SSIM:

$$\mathcal{L}_{ssim} = 1 - \text{SSIM}(I_{en}, I_{gt}). \quad (10)$$

The core innovation is the illumination-aware Adaptive DCP loss. Standard DCP assumes near-zero intensity in at least one color channel; enforcing this on underwater images suppresses weak signals in photon-starved regions, causing artifacts. Our Adaptive DCP modulates the penalty weight $\alpha_{adaptive}$ via the global average brightness B_{global} of the input:

$$\alpha_{adaptive} = \begin{cases} 0.05, & \text{if } B_{global} < 0.35 \\ 1.0, & \text{if } 0.35 \leq B_{global} < 0.5 \\ 1.5, & \text{if } B_{global} \geq 0.5 \end{cases}. \quad (11)$$

As shown in Fig. 6, well-lit regions receive aggressive dehazing ($\alpha = 1.5$) while photon-starved regions are protected ($\alpha = 0.05$). The thresholds (0.35, 0.5) were selected from the brightness distribution of LSUI3879, where $B_{global} < 0.35$ corresponds to low-light scenes, 0.35–0.5 to moderate lighting, and ≥ 0.5 to well-lit conditions with strong backscatter. Applied uniformly across all datasets without tuning, the consistent gains on all four benchmarks (Table 1)—including OceanEx where $\alpha = 0.05$ is automatically applied—indicate reasonable stratification. We do not claim optimality; per-dataset calibration is left for future work. Since $\alpha_{adaptive}$ is a scalar constant computed before the forward pass, it does not participate in gradient computation ($\nabla \mathcal{L}_{dcp} = \alpha_{adaptive} \cdot \nabla \mathcal{L}_{dcp}^{base}$), so discrete threshold transitions do not affect backpropagation. Mini-batch training across diverse brightness regimes partially smooths the optimization trajectory.

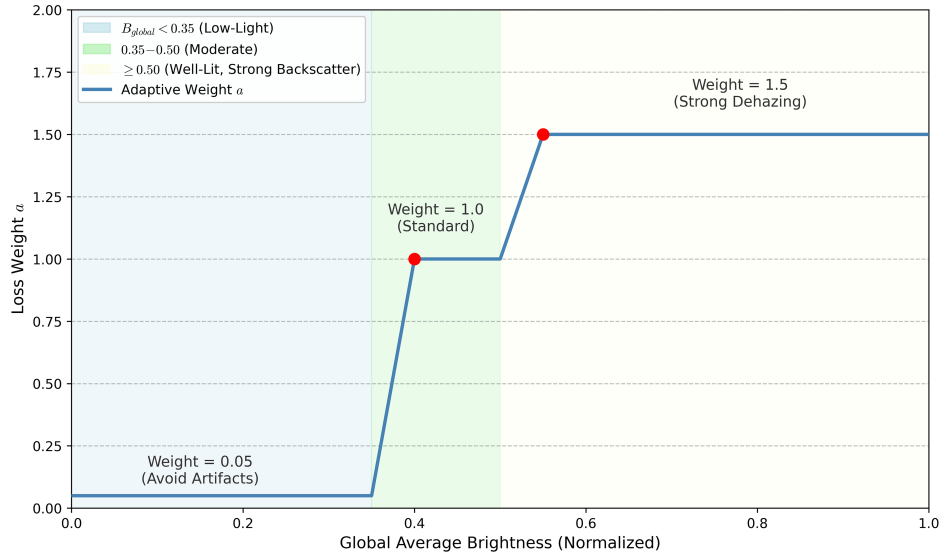


Figure 6: Adaptive dark channel weighting guided by global brightness, stratifying scenes into low-light, medium, and well-lit zones with dynamic loss weights.

Table 1: Quantitative comparison on four full-reference datasets. Best, second, and third results are highlighted in **red**, **blue**, and **green**, respectively. All highlighted results are also bolded to ensure visibility in black-and-white prints.

Methods	LSUI400		EUVP_515		UIEB100		OceanEx	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
UT-UIE	24.351	0.829	25.214	0.813	20.916	0.764	21.272	0.822
SyreaNet	18.051	0.766	17.721	0.743	16.501	0.836	20.243	0.865
ERD	25.964	0.874	24.365	0.838	22.343	0.873	21.655	0.892
FUnIE-GAN	23.272	0.818	24.077	0.794	19.614	0.813	20.448	0.855
MAC-Lookup	23.153	0.844	23.643	0.833	20.212	0.846	22.302	0.892
WaterNet	26.688	0.874	25.285	0.833	22.279	0.868	22.132	0.887
SGUIE-Net	19.911	0.819	19.187	0.766	21.178	0.872	18.677	0.834
SINET-UIE	20.695	0.838	23.911	0.844	17.995	0.833	20.987	0.827
RAUNE-Net	26.812	0.876	26.331	0.845	22.751	0.879	22.728	0.876
GHSUIR	25.996	0.859	25.879	0.834	21.635	0.871	22.002	0.853
Ours	27.625	0.885	27.596	0.855	22.864	0.874	22.053	0.876

The weighted Adaptive DCP loss is:

$$\mathcal{L}_{dcp} = \alpha_{adaptive} \cdot \frac{1}{HW} \sum_{x,y} \min_{c \in \{R,G,B\}} \left(\min_{(x',y') \in \Omega(x,y)} I_{en}^c(x', y') \right), \quad (12)$$

where $\Omega(x, y)$ is a 15×15 patch following the original DCP [9]. The nested minima are implemented differentiably via $\min(\cdot) = -\max(-\cdot)$ using max-pooling with stride 1 and appropriate padding. The total loss is:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_{ssim} + \lambda_3 \mathcal{L}_{dcp}, \quad (13)$$

with $\lambda_1 = 1.0, \lambda_2 = 1.0, \lambda_3 = 0.03$.

4 Experiments

4.1 Datasets and Implementation Details

PFR-Net is trained on LSUI3879 [31] (3879 paired images with predefined splits) and tested zero-shot on four full-reference benchmarks: LSUI400 [31] (400 pairs, no overlap with the training set), UIEB100 [21], EUVP_515 [2], and OceanEx—a 41-pair extreme low-light dataset distributed with RAUNE-Net [23]. No fine-tuning is permitted on any test set. For no-reference validation, we use U45 [32], RUIE_Color90 [8], and UPoor200 [21]. Training images are resized to 256×256 via bicubic interpolation, normalized to $[-1, 1]$, with random horizontal flipping ($p = 0.5$) as the only augmentation. Test images are evaluated at original resolution; any size mismatch is resolved by resizing the ground truth via bicubic interpolation. The network is implemented in PyTorch 1.13.1 on an NVIDIA RTX 4090 D (24 GB), trained for 100 epochs with batch size 4. We use Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$, no weight decay) with initial learning rate 2×10^{-4} , decayed by 0.1 at epochs 50, 75, and 100. No fixed random seed was used. The MLP-CM is initialized with near-zero weights for identity mapping at $t = 0$. PSNR and SSIM are computed via Kornia (PSNR: max = 1.0, SSIM: window = 11), averaged over all test images. All ten competing methods (WaterNet [21], FUnIE-GAN [2], SyreaNet [33], SGUIE-Net [34], UT-UIE [31], MAC-Lookup [35], SINET-UIE [36], RAUNE-Net [23], GHSUIR [37], and ERD [38]) are retrained by us on the identical LSUI3879 training set under a unified protocol.

4.2 Quantitative Evaluation

As shown in Table 1, PFR-Net achieves the best PSNR and SSIM on LSUI400 (27.625/0.885) and EUVP_515 (27.596/0.855). On UIEB100, it attains the highest PSNR (22.864) but ranks second in SSIM (0.874 vs. RAUNE-Net's 0.879). On OceanEx—the most challenging benchmark with extreme low-light—PFR-Net ranks fourth in PSNR (22.053) and third in SSIM (0.876, tied with RAUNE-Net), attributable to photon starvation, where Phase 1 spectral calibration may over-correct the limited color information available. No statistical significance testing was performed; rankings are thus indicative rather than definitive. These results demonstrate that the decoupled phased paradigm provides a robust foundation, with clear opportunities for improvement in extreme low-light scenarios.

4.3 Qualitative Evaluation

Perceptual superiority is best viewed through the lens of extreme degradation. The RUIE_Color90 results (Fig. 7) show that while some competitive algorithms introduce artifacts or leave residual color masks, PFR-Net executes precise spectral realignment, restoring natural textures. Fig. 8 targets the backscatter in U45, where PFR-Net enhances visibility of diving equipment with high global consistency. A more challenging test is UPoor200 (Fig. 9), where PFR-Net, aided by XCA and Adaptive DCP, neutralizes the color cast and amplifies the contrast of structures in heavily degraded zones.

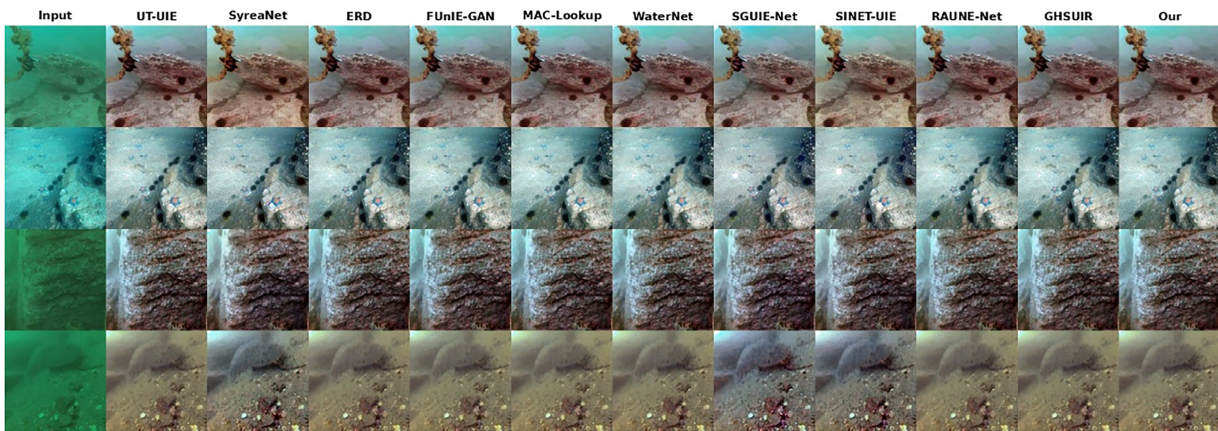


Figure 7: Visual comparison on RUIE_Color90 [8]. Our method successfully mitigates severe greenish casts.

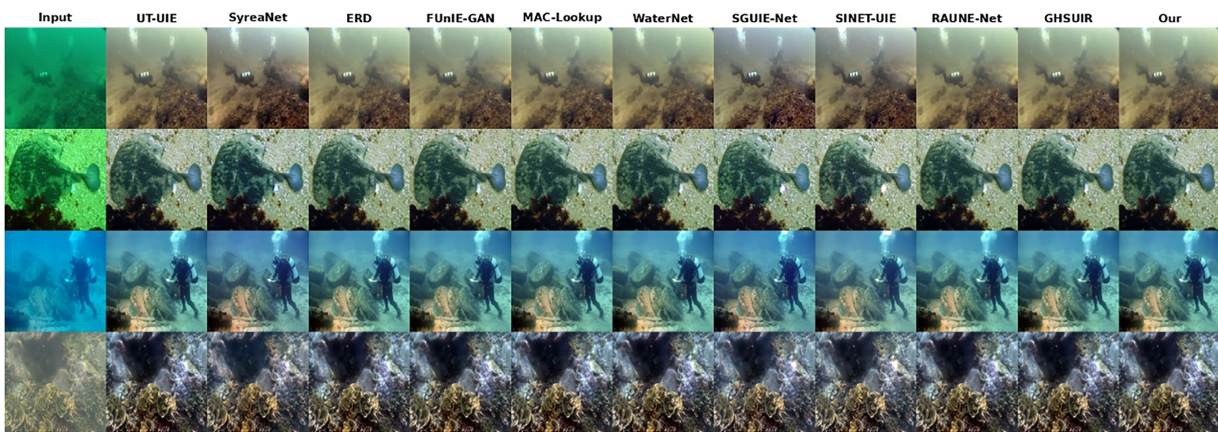


Figure 8: Qualitative results on U45 [32]. We achieve superior clarity in high-haze scenarios.

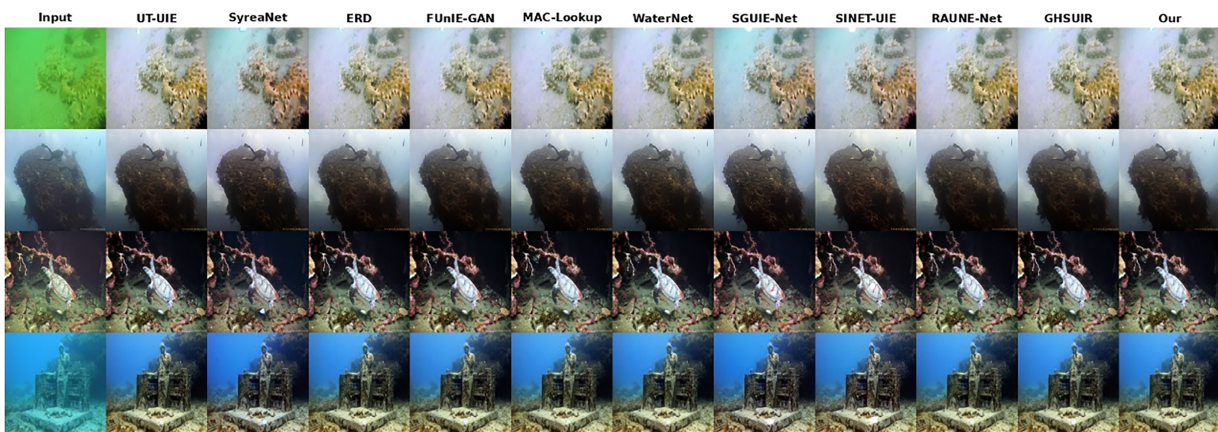


Figure 9: Visual comparison on the heavily degraded UPoor200 dataset [21].

4.4 Ablation Study

We conduct a comprehensive ablation study to evaluate the contribution of each architectural component and validate the central design choices. All variants are trained under the identical protocol described in Section 4.1. Table 2 presents results on LSUI400, organized into two groups: component removal analysis (upper panel) and architectural design validation (lower panel).

Table 2: Quantitative ablation study on the LSUI400 dataset. In the upper panel, each row removes one component from the full model. In the lower panel, architectural design variants validate the two-phase decoupling, module allocation strategy, and activation function choice. Best results are **bolded**.

Variant	Components				Metrics		
	DMSRB	CM	NAFG	XCA	DCP	PSNR	SSIM
<i>Component Removal</i>							
Fixed DCP ($\alpha = 1.0$)	✓	✓	✓	✓	Fixed	26.814	0.871
w/o XCA	✓	✓	✓	×	Adaptive	27.125	0.875
w/o NAFG	✓	✓	×	✓	Adaptive	27.051	0.872
w/o CM	✓	×	✓	✓	Adaptive	26.432	0.865
ResNet Backbone	×	✓	✓	✓	Adaptive	26.115	0.860
<i>Architectural Design Validation</i>							
All 30 DMSRB, no phase split	✓	×	×	✓	Adaptive	26.252	0.862
All 30 DMSRB + CM (all blocks)	✓	✓*	×	✓	Adaptive	27.481	0.882
All 30 DMSRB + NAFG (all blocks)	✓	×	✓*	✓	Adaptive	27.055	0.884
Full Model (ReLU)	✓	✓	✓	✓	Adaptive	27.513	0.882
Full Model (Mish)	✓	✓	✓	✓	Adaptive	27.625	0.885

Note: ✓: component present; ×: component removed or replaced. *Applied to all 30 blocks instead of the phased arrangement (15 blocks each). “fixed” denotes a constant DCP weight $\alpha = 1.0$; “adaptive” denotes the illumination-aware $\alpha_{adaptive}$ from Eq. (11).

Component Removal. Bypassing MLP-CM (w/o CM) causes the largest drop (PSNR: −1.193, SSIM: −0.020), confirming the critical role of spectral calibration. Replacing DMSRB with a standard ResNet backbone yields the worst result (PSNR: 26.115), and removing XCA (−0.500) or NAFG (−0.574) produces moderate degradation. Replacing Adaptive DCP with Fixed DCP ($\alpha = 1.0$) reduces PSNR by 0.811 and SSIM by 0.014.

Architectural Design Validation. A non-phased baseline (30 identical DMSRB blocks without MLP-CM, NAFG, or phase split) achieves only 26.252 dB—a 1.373 dB degradation—directly validating the two-phase decoupling. Applying MLP-CM uniformly to all 30 blocks yields 27.481 dB (0.144 dB below full model, with ~264K extra parameters); uniform NAFG yields 27.055 dB (−0.570). The phased allocation—MLP-CM in Phase 1, NAFG in Phase 2—outperforms both, confirming the “divide-and-conquer” principle. Replacing Mish with ReLU (Full Model, ReLU) reduces PSNR by 0.112 dB (27.513 vs. 27.625); Mish consistently outperforms ReLU across all benchmarks (Table 3).

Cross-Dataset Ablation. Table 3 evaluates key design choices on all four benchmarks. The non-phased baseline consistently underperforms (average PSNR gap: 1.20 dB). Adaptive DCP outperforms Fixed DCP on every dataset, with the largest margin on OceanEx (+0.828), where adaptive protection is most critical. Mish surpasses ReLU with average gains of 0.116 dB PSNR and 0.003 SSIM.

Table 3: Cross-dataset ablation of key design choices. PSNR/SSIM are reported for each dataset. “Non-phased” denotes 30 identical DMSRB blocks without MLP-CM, NAFG, or phase partitioning. “Fixed DCP” uses $\alpha = 1.0$; “Adaptive DCP” uses Eq. (11).

Variant	LSUI400		EUVP_515		UIEB100		OceanEx	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Non-phased (all 30 DMSRB)	26.252	0.862	26.102	0.835	21.450	0.862	20.802	0.865
Fixed DCP ($\alpha = 1.0$)	26.814	0.871	26.791	0.855	22.171	0.861	21.225	0.862
Full Model (ReLU)	27.513	0.882	27.487	0.852	22.724	0.871	21.932	0.874
Full Model (Mish + Adaptive DCP)	27.625	0.885	27.596	0.855	22.864	0.874	22.053	0.876

4.5 Complexity and Performance Trade-Off

Table 4 compares parameter counts, FLOPs, MACs, inference time, FPS, and GPU memory on 256×256 inputs (NVIDIA RTX 4090 D, 24 GB).

Table 4: Computational complexity and runtime performance (evaluated on 256×256 input, NVIDIA RTX 4090 D). FLOPs and MACs are measured using the thop library; inference time and FPS are averaged over 1000 forward passes; GPU memory is measured via PyTorch native utilities.

Method	Params (M)	FLOPs (G)	MACs (G)	Time (ms)	FPS	Mem (GB)
WaterNet	1.09	142.85	71.42	2.70	370.6	0.01
FUnIE-GAN	7.02	20.48	10.24	0.77	1303.6	0.04
RAUNE-Net	36.73	335.64	167.85	12.23	81.8	2.03
Ours (Backbone)	60.59	265.68	132.84	19.08	52.4	4.37
Ours (Full)	67.94	291.79	145.90	29.34	34.1	6.67
Ours (Lightweight)	14.88	77.42	38.71	9.48	105.5	1.76

Design Rationale and Computational Cost. The full PFR-Net (67.94M parameters, 291.79 G FLOPs) is substantially heavier than FUnIE-GAN (7.02M) and WaterNet (1.09M), primarily due to the 30-block DMSRB backbone with multi-branch dilated convolutions. The phased rectification suite (15 MLP-CM instances, 15 NAFG modules, and XCA) adds approximately 7.35M parameters. At 256×256 , the full model achieves 34.1 FPS (29.34 ms) with 6.67 GB GPU memory, unsuitable for real-time AUV deployment. We position it as a high-capacity offline model where fidelity supersedes speed.

Lightweight Variant. We additionally configure a lightweight variant with 4 DMSRB blocks (first two with MLP-CM, last two with NAFG), preserving the two-phase decoupling. This variant has 14.88M parameters and 77.42 G FLOPs, achieving 105.5 FPS (9.48 ms) with 1.76 GB memory. As shown in Table 5, it achieves competitive results (LSUI400: 27.02 dB/0.883 SSIM), surpassing all non-PFR-Net competitors on LSUI400 and UIEB100 at over $3\times$ the throughput of the full model. On OceanEx, it slightly exceeds the full model (22.210 vs. 22.053 dB), suggesting reduced depth mitigates over-correction in extreme low-light. Further compression via knowledge distillation is left for future work.

Table 5: Quantitative results of the lightweight PFR-Net variant (4 DMSRB blocks with two-phase decoupling) compared to the full model on all four full-reference benchmarks.

Variant	LSUI400		EUVP_515		UIEB100		OceanEx	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Ours (Full)	27.625	0.885	27.596	0.855	22.864	0.874	22.053	0.876
Ours (Lightweight)	27.020	0.883	26.170	0.854	22.990	0.883	22.210	0.878

5 Discussion and Limitation

While PFR-Net achieves competitive restoration by harmonizing physical priors with deep feature learning, it encounters the fundamental limits of passive optical imaging. In extreme photon-starved environments, the underlying signals are not merely degraded but absent; no deterministic mapping can reconstruct information the sensor never captured. Under such constraints, forcing structural recovery amplifies sensor noise.

To transcend these limits, we will explore physics-guided Diffusion Models [28,29], whose stochastic generative processes can synthesize plausible high-frequency details where physical signals are lost. For deployment on power-constrained AUVs, we have introduced a lightweight variant (4 DMSRB blocks, 14.88M parameters, 105.5 FPS) that preserves the two-phase decoupling principle while achieving real-time inference. Further compression through Knowledge Distillation (KD) and structural pruning for edge deployment on embedded platforms is left for future work. Furthermore, the dynamic turbidity of estuarine environments motivates investigating Unsupervised Domain Adaptation (UDA) to enable robust cross-domain generalization without costly paired data.

Finally, the current evaluation relies on PSNR and SSIM, which do not fully capture perceptual naturalness. We plan to incorporate LPIPS, UIQM, and UCIQE in future evaluations to provide a more comprehensive assessment across diverse water types and illumination conditions.

6 Conclusion

This paper presented PFR-Net, a Phased Feature Rectification Network that resolves the entangled nature of underwater image degradation through a decoupled two-phase strategy. Preemptive spectral calibration via lightweight MLP-CM modules eliminates global color casts, while DMSRB blocks with Mish activations and NAFG gating recover multi-scale structural details. XCA captures global channel dependencies at linear spatial complexity, and the illumination-aware Adaptive DCP loss dynamically balances dehazing against structural preservation. Extensive evaluations demonstrate that PFR-Net establishes a competitive baseline for marine optical restoration. Future work will focus on knowledge distillation for real-time AUV deployment.

Acknowledgement: None.

Funding Statement: This research was partially supported by the National Natural Science Foundation of China [Grant No. 12271076].

Author Contributions: Jinshuo Ma: Conceptualization, Methodology, Software, Formal analysis, Validation, Writing—original draft. Yang Li: Writing—review & editing, Supervision, Project administration, Funding acquisition. Can Guo: Data curation, Formal analysis. Wen Gao: Validation, Visualization. Ruiming Zhang: Validation, Visualization. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The datasets used in this study are publicly available from their respective sources as cited. The trained models and evaluation results are available from the corresponding author upon reasonable request. The source code is part of an ongoing funded project and is not currently publicly released; it may be shared for academic collaboration purposes upon reasonable request to the corresponding author.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Anwar S, Li C. Diving deeper into underwater image enhancement: a survey. *Signal Process Image Commun.* 2020;89(6):115978. doi:10.1016/j.image.2020.115978.
2. Islam MJ, Xia Y, Sattar J. Fast underwater image enhancement for improved visual perception. *IEEE Robot Autom Lett.* 2020;5(2):3227–34. doi:10.1109/lra.2020.2974710.
3. Cong R, Yang W, Zhang W, Li C, Guo CL, Huang Q, et al. PUGAN: physical model-guided underwater image enhancement using GAN with dual-discriminators. *IEEE Trans Image Process.* 2023;32:4472–85. doi:10.1109/tip.2023.3286263.
4. Shi X, Wang YG. CPDM: content-preserving diffusion model for underwater image enhancement. *Sci Rep.* 2024;14(1):31309. doi:10.1038/s41598-024-82803-y.
5. Cong X, Zhao Y, Gui J, Hou J, Tao D. A comprehensive survey on underwater image enhancement based on deep learning. *IEEE Trans Emerg Top Comput Intell.* 2026:1–19. doi:10.1109/tetci.2026.3668671.
6. Wang Y, Guo J, He W, Gao H, Yue H, Zhang Z, et al. Is underwater image enhancement all object detectors need? *IEEE J Ocean Eng.* 2024;49(2):606–21. doi:10.1109/joe.2023.3302888.
7. Jaffe JS. Computer modeling and the design of optimal underwater imaging systems. *IEEE J Ocean Eng.* 1990;15(2):101–11. doi:10.1109/48.50695.
8. Liu R, Fan X, Zhu M, Hou M, Luo Z. Real-world underwater enhancement: challenges, benchmarks, and solutions under natural light. *IEEE Trans Circuits Syst Video Technol.* 2020;30(12):4861–75.
9. He K, Sun J, Tang X. Single image haze removal using dark channel prior. *IEEE Trans Pattern Anal Mach Intell.* 2010;33(12):2341–53. doi:10.1109/tpami.2010.168.
10. Liu T, Zhu K, Wang X, Song W, Wang H. Lightweight underwater image adaptive enhancement based on zero-reference parameter estimation network. *Front Mar Sci.* 2024;11:1378817. doi:10.3389/fmars.2024.1378817.
11. Lin WT, Lin YX, Chen JW, Hua KL. PixMamba: leveraging state space models in a dual-level architecture for underwater image enhancement. In: *Proceedings of the Asian Conference on Computer Vision (ACCV)*; 2024 Dec 8–12; Hanoi, Vietnam. p. 3622–37.
12. Zhang Y, Yu X, Cai Z. UWMambaNet: dual-branch underwater image reconstruction based on W-shaped mamba. *Mathematics.* 2025;13(13):2153. doi:10.3390/math13132153.
13. Sun T, Zhang Y, Hu J, Cui H, Yu T. A multi-domain enhanced network for underwater image enhancement. *Information.* 2025;16(8):627. doi:10.3390/info16080627.
14. Zeng H, Cai J, Li L, Cao Z, Zhang L. Learning image-adaptive 3D lookup tables for high performance photo enhancement in real-time. *IEEE Trans Pattern Anal Mach Intell.* 2022;44(4):2058–73. doi:10.1109/tpami.2020.3026740.
15. Chen L, Chu X, Zhang X, Sun J. Simple baselines for image restoration. In: *Proceedings of the Computer Vision—ECCV 2022*; 2022 Oct 23–27; Tel Aviv, Israel. Cham, Switzerland: Springer Nature; 2022. p. 17–33.
16. Ali A, Touvron H, Caron M, Bojanowski P, Douze M, Joulin A, et al. XcIT: cross-covariance image transformers. In: *Advances in neural information processing systems*. Vol. 34. Red Hook, NY, USA: Curran Associates, Inc.; 2021. p. 20014–27.
17. Misra D. Mish: a self regularized non-monotonic activation function. In: *Proceedings of the British Machine Vision Conference (BMVC)*; 2020 Sep 7–10; Virtual.

18. Drews P Jr, do Nascimento E, Moraes F, Botelho S, Campos M. Transmission estimation in underwater single images. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops; 2013 Dec 2–8; Sydney, NSW, Australia. p. 825–30.
19. Peng YT, Cao K, Cosman PC. Generalization of the dark channel prior for single image restoration. *IEEE Trans Image Process.* 2018;27(6):2856–68. doi:10.1109/tip.2018.2813092.
20. Ancuti CO, Ancuti C, De Vleeschouwer C, Bekaert P. Color balance and fusion for underwater image enhancement. *IEEE Trans Image Process.* 2017;27(1):379–93. doi:10.1109/tip.2017.2759252.
21. Li C, Guo C, Ren W, Cong R, Hou J, Kwong S, et al. An underwater image enhancement benchmark dataset and beyond. *IEEE Trans Image Process.* 2019;29:4376–89. doi:10.1109/TIP.2019.2955241.
22. Li C, Anwar S, Hou J, Cong R, Guo C, Ren W. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Trans Image Process.* 2021;30:4985–5000. doi:10.1109/tip.2021.3076367.
23. Peng W, Zhou C, Hu R, Cao J, Liu Y. RAUNE-Net: a residual and attention-driven underwater image enhancement method. In: Digital multimedia communications. Singapore: Springer Nature; 2024. p. 15–27.
24. Zhu S, Geng Z, Xie Y, Zhang Z, Yan H, Zhou X, et al. New underwater image enhancement algorithm based on improved U-net. *Water.* 2025;17(6):808. doi:10.3390/w17060808.
25. Yang C, Shao L, Deng Y, Wang J, Zhai H. SwinCNet leveraging Swin Transformer V2 and CNN for precise color correction and detail enhancement in underwater image restoration. *Front Mar Sci.* 2025;12:1523729. doi:10.3389/fmars.2025.1523729.
26. Du D, Si L, Xu F, Niu J, Sun F. A physical model-guided framework for underwater image enhancement and depth estimation. *IEEE Trans Circuits Syst Video Technol.* 2026;36(7):10244–59. doi:10.1109/tcsvt.2026.3674091.
27. Guo H, Li J, Dai T, Ouyang Z, Ren X, Xia ST. MambaIR: a simple baseline for image restoration with state-space model. In: Proceedings of the Computer Vision—ECCV 2024; 2024 Sep 29–Oct 4; Milan, Italy. Cham, Switzerland: Springer Nature; 2025. p. 222–41.
28. Schein TT, De Almeida GP, Brião SL, De Bem RA, De Oliveira FG, Drews PLJ Jr. UDBE: unsupervised diffusion-based brightness enhancement in underwater images. In: Proceedings of the 2024 International Conference on Machine Learning and Applications (ICMLA); 2024 Dec 18–20; Miami, FL, USA. New York, NY, USA: IEEE; 2024. p. 664–70.
29. Huang C, Cao J, Ma J, Yu K, Li C, Yang H, et al. DACA-Net: a degradation-aware conditional diffusion network for underwater image enhancement. In: Proceedings of the 33rd ACM International Conference on Multimedia; 2025 Oct 27–31; Dublin, Ireland. p. 4962–71.
30. Ye R, Qian Y, Huang X. RT-CBAM: refined transformer combined with convolutional block attention module for underwater image restoration. *Sensors.* 2024;24(18):5893. doi:10.3390/s24185893.
31. Peng L, Zhu C, Bian L. U-shape transformer for underwater image enhancement. *IEEE Trans Image Process.* 2023;32(2):3066–79. doi:10.1109/tip.2023.3276332.
32. Li H, Li J, Wang W. A fusion adversarial underwater image enhancement network with a public test dataset. *arXiv:1906.06819.* 2019.
33. Wen J, Cui J, Zhao Z, Yan R, Gao Z, Dou L, et al. SyreaNet: a physically guided underwater image enhancement framework integrating synthetic and real images. In: Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA); 2023 May 29–Jun 3; London, UK. New York, NY, USA: IEEE; 2023. p. 5177–83.
34. Qi Q, Li K, Zheng H, Gao X, Hou G, Sun K. SGUIE-net: semantic attention guided underwater image enhancement with multi-scale perception. *IEEE Trans Image Process.* 2022;31:6816–30. doi:10.1109/tip.2022.3216208.
35. Yi F, Zheng Z, Liang Z, Dong Y, Fang X, Wu W, et al. MAC-lookup: multi-axis conditional lookup model for underwater image enhancement. In: Proceedings of the 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC); 2025 Oct 5–8; Vienna, Austria. New York, NY, USA: IEEE; 2025. p. 1556–61.
36. Panda G, Kundu S, Bhattacharya S, Routray A. SINET: sparsity-driven interpretable neural network for underwater image enhancement. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2025 Apr 6–11; Hyderabad, India. New York, NY, USA: IEEE; 2025. p. 1–5.

37. Guo X, Chen X, Wang S, Pun CM. Underwater image restoration through a prior guided hybrid sense approach and extensive benchmark analysis. *IEEE Trans Circuits Syst Video Technol.* 2025;35(5):4784–800. doi:10.1109/tcsvt.2025.3525593.
38. Cao J, Peng W, Liu Y, Dong J, Le Callet P, Kwong S. ERD: encoder-residual-decoder neural network for underwater image enhancement. *IEEE Trans Circuits Syst Video Technol.* 2025;35(9):8958–72. doi:10.1109/tcsvt.2025.3556203.