PGMamba: A Physical Model-Guided Global Mamba for Underwater Image Enhancement

Zijun Tan[1], Chuan Fu[1], Tan Guo[2], Zhixiong Nan[1], Pengzhan Zhou[1], Xinggan Peng[3], Fulin Luo[1]*

[1] College of Computer Science, Chongqing University, China
[2] School of Communication and Information Engineering, Chongqing University of Post and Telecommunications, China
[3] CMCU Engineering Co.,Ltd., China
20241401031G@stu.cqu.edu.cn, fuchuan@cqu.edu.cn, guot@cqupt.edu.cn, nanzx@cqu.edu.cn, pzzhou@cqu.edu.cn, pxg@cmcu.cn, luoflyn@cqu.edu.cn

Abstract

Underwater image enhancement (UIE) aims to address image degradation caused by water absorption and scattering effects. Despite significant progress in deep learning-based UIE methods, existing approaches still face key challenges due to the neglect of physical imaging principle. Moreover, while current Mamba models achieve global modeling via multi-directional scanning, their local sequential strategy lacks sufficient global context. To this end, we propose a novel Physical Model-Guided Global Mamba (PGMamba) that combines the efficient sequential modeling capability of Mamba with underwater imaging physical model. Specifically, we first design a Spatial-Aware Global Mamba (SAGMamba) that achieves efficient long-range dependency modeling through a spatial-aware ranking strategy with global context information. Second, we develop a Physical Model-Guided Feed-Forward Network (PMGFFN) that explicitly incorporates underwater optical imaging principles into the network architecture. Extensive experimental results and comprehensive ablation studies demonstrate the outstanding performance and importance of our proposed method.

Code — https://github.com/ZijunTan/PMGMamba

Introduction

Underwater imaging technology (Akkaynak and Treibitz 2019) serves as a vital tool for humanity to understand, exploit, and protect the marine environment, playing a significant role in fields such as seabed resource exploration (Cai et al. 2020) and marine biological research (Islam, Xia, and Sattar 2020). However, in underwater environments, light absorption, scattering, and suspended matter cause significant image degradation and color distortion; therefore, the advancement of underwater image enhancement (UIE) techniques is essential for reliable underwater visual perception.

Early research on UIE methods primarily focused on traditional approaches. Traditional UIE methods based on physical models incorporate prior knowledge of physical imaging models, typically derived from optical imaging principles or mathematical modeling, to reconstruct clear images from degraded inputs through inverse derivation (Ancuti et al. 2012; Drews et al. 2013; Peng, Cao, and Cosman 2018; Akkaynak et al. 2017; Zhang et al. 2022). However, these methods require simultaneous estimation of multiple parameters, resulting in poor model generalizability.

In recent years, the rapid advancement of deep learning has shifted the paradigm toward data-driven approaches. With the rise of deep learning, convolutional neural networks (CNNs) have emerged as a dominant solution due to their powerful feature extraction capabilities. Nevertheless, their limitation of local receptive fields restricts their performance (Li et al. 2019). The introduction of Transformers (Dosovitskiy et al. 2020) marked a significant breakthrough, as their self-attention mechanism enables global dependency modeling, overcoming the locality constraints of CNNs and

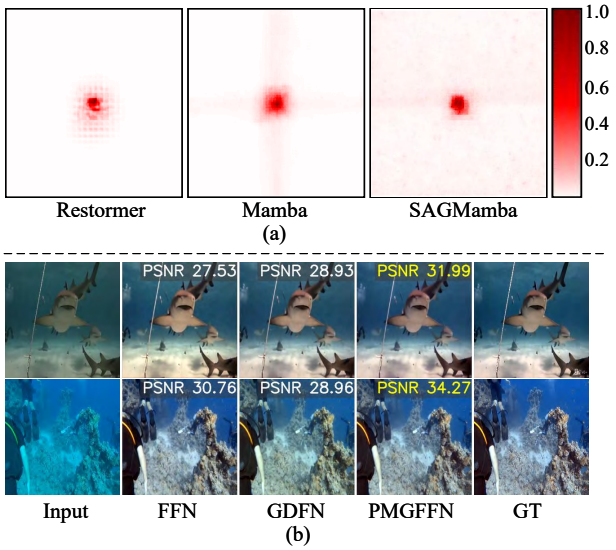


Figure 1: (a) The Effective Receptive Field (ERF) visualization for the Restormer (Zamir et al. 2022), Mamba in VMamba (Liu et al. 2024) and our SAGMamba. A broader distribution of bright areas signifies a larger ERF. (b) The visual result of FFN (Dosovitskiy et al. 2020), GDFN (Zamir et al. 2022) and our PMGFFN.

*Corresponding author

elevating UIE performance to new heights. Despite the impressive performance of Transformer-based methods, their high computational complexity in self-attention operations makes them prohibitively expensive for all image restoration tasks. Although some studies have attempted to reduce computational overhead by shifting self-attention from spatial to channel dimension (Zamir et al. 2022), the resulting performance remains suboptimal. Recently, state space model (SSM) (Liu et al. 2024) has emerged as a promising alternative to UIE, offering a novel research direction. By leveraging the unique scanning mechanism, Mamba (Gu and Dao 2023) achieves global feature learning while maintaining favorable computational efficiency.

While recent works have attempted to combine physical constraints with deep learning approaches (Cong et al. 2023; Li et al. 2021; Zhang et al. 2023, 2025), they suffer from oversimplified architectures or complex cascaded designs for physical parameter estimation, which hinders their scalability. Furthermore, despite employing multi-directional scanning for global modeling, Mamba's local sequential strategy still lacks sufficient global context integration. To address this limitation, we propose a Physical Model-Guided Global Mamba (PGMamba) for UIE. Departing from existing Mamba architectures (Liu et al. 2024; Lianghui et al. 2024; Huang et al. 2024; Xiao et al. 2024), our designed Spatial-Aware Global Mamba (SAG-Mamba) replaces the multi-directional scanning method with a spatial-aware ranking strategy to improve efficiency, while augmenting the SSM with global enhancement components to better capture long-range dependencies, as illustrated in Fig. 1(a). Furthermore, we design a Physical Model-Guided Feed-Forward Network (PMGFFN) that hierarchically integrates physical priors with deep features by embedding the underwater imaging model within the network architecture, thus achieving better results than previous Feed-Forward Networks (FFNs), as shown in Fig. 1(b). The main contributions of this work are summarized as follows:

- We propose a Spatial-Aware Global Mamba (SAG-Mamba), which integrates globally enhanced SSM to complement local sequential scanning with global context and improve global feature perception.
- We design a Physical Model-Guided Feed-Forward Network (PMGFFN) that effectively integrates physical priors into data-driven learning, enhancing image restoration performance under physical constraints.
- By integrating these two components, we introduce the Physical Model-Guided Global Mamba (PGMamba). Extensive experiments demonstrate that our method achieves state-of-the-art performance.

Related Works

Physical Model-Based Methods

Traditional physical model-based UIE methods treat enhancement as an inverse problem, where image reconstruction is guided by priors derived from underwater imaging physics. Common priors include the dark channel prior (Peng and Cosman 2017), attenuation curve prior (Wang, Liu, and Chau 2017), blur prior (Chiang and Chen 2011) and minimum information prior (Li et al. 2016). For instance, Drews et al. (Drews et al. 2013) estimated light transmission based on the statistical characteristics of natural scenes. Building on this, Li et al. (Li et al. 2017) improved background light estimation by incorporating underwater optical cues. Meanwhile, Peng et al. (Peng, Cao, and Cosman 2018) addressed color casts and contrast issues via depth-aware ambient light estimation and adaptive correction. Akkaynak and Treibitz (Akkaynak and Treibitz 2019) revised the atmospheric model for RGBD-based enhancement.

Learning-Based Methods

CNN-Based Methods. CNNs have significantly advanced UIE by learning direct mappings from degraded to clear images. Li et al. (Li et al. 2019) introduced a multi-branch CNN that processes raw images along with their white-balanced, histogram-equalized, and gamma-corrected counterparts to improve enhancement robustness. Extending this work, Li et al. (Li et al. 2021) incorporated transmission priors to improve color accuracy, while Zhang et al. (Zhang et al. 2024) proposed a multi-scale refinement framework for better detail restoration. Generative adversarial networks (Goodfellow et al. 2020) have also been widely adopted for UIE. For instance, Cong et al. (Cong et al. 2023) proposed PUGAN, which employs dual subnetworks to improve perceptual realism, and Quan et al. (Quan et al. 2024) introduced AMSIM, which enhances visual quality through frequency-aware decomposition. Moreover, Zhao et al. (Zhao et al. 2024b) emphasized structural preservation by integrating spatial-frequency fusion mechanisms.

Transformer-Based Methods. Transformers have garnered increasing attention due to their capability in modeling long-range dependencies via self-attention mechanisms. Peng et al. (Peng, Zhu, and Bian 2023) introduced a U-shaped Transformer that incorporates both channel and spatial attention to enhance global contextual understanding. To capture richer representations, Khan et al. (Khan et al. 2024) proposed SpectroFormer, which simultaneously utilizes frequency and spatial domain features. To address uneven lighting conditions in underwater scenes, Shang et al. (Shang et al. 2025) introduced a luminance-guided strategy that enhances illumination balance. In addition, Zhuang et al. (Zhuang et al. 2024) utilized deformable attention to reduce computational cost while maintaining performance. Furthermore, Mishra et al. (Mishra et al. 2025) incorporated wavelet-based feature decomposition in USWFormer to better capture multi-scale frequency characteristics.

State Space Model

As emerging architectures following CNN and Transformer, SSMs have attracted widespread attention in computer vision due to their powerful long-range dependency modeling capabilities and linear computational complexity. VMamba (Liu et al. 2024) and Vim (Lianghui et al. 2024) first introduced the Mamba (Gu and Dao 2023) to vision tasks, achieving global visual perception through bidirectional cross-scanning. LocalMamba (Huang et al. 2024) further

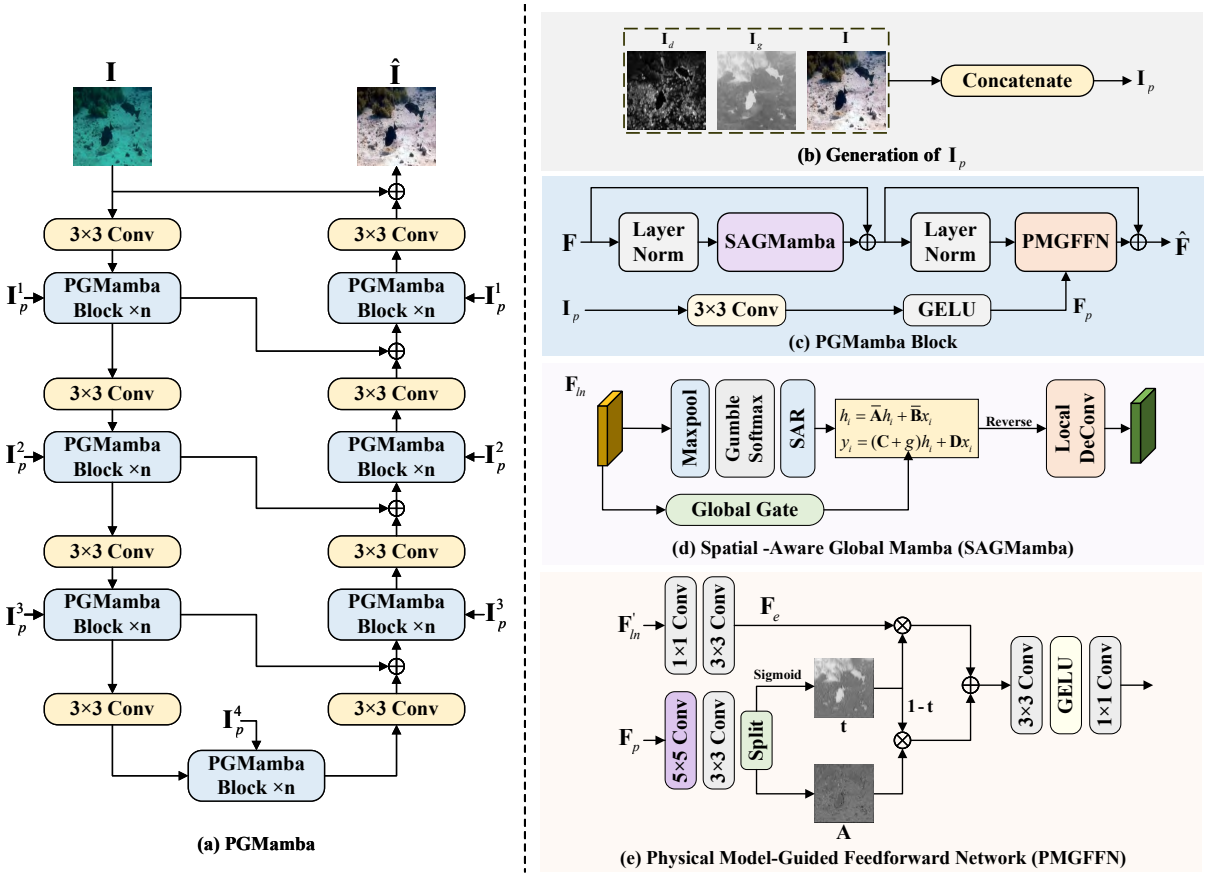


Figure 2: The architecture of the proposed PGMamba. (a) PGMamba mainly consists of seven stages, each of which contains n PGMamba blocks. (b) The $\mathbf{I}_p$ is formed by concatenating the degraded image $\mathbf{I}$ with its corresponding depth map $\mathbf{I}_d$ and gradient map $\mathbf{I}_g$. (c) PGMamba mainly consists of a SAGMamba and a PMGFFN. (d) SAGMamba achieves accurate contextual understanding through SAR strategy and a global gate. (e) PMGFFN achieves image enhancement under physical constraints.

optimized local feature extraction via window-based local scanning strategies, enhancing fine-grained feature capture. SpatialMamba (Xiao et al. 2024) redesigned the scanning paradigm, aggregating global information in a single scan while maintaining satisfactory performance. Recently, SSMs have shown great potential in image restoration tasks (Guo et al. 2024, 2025). However, existing Mamba methods haven't considered UIE's unique requirements, urgently necessitating novel Mamba architectures specifically designed for underwater vision characteristics.

Method

Overall Architecture

The architecture of PGMamba is illustrated in Fig. 2(a). The model follows a canonical U-shaped network design. Given a degraded underwater image $\mathbf{I}$ as input, the framework first extracts low-level features through a 3×3 convolutional layer. The proposed architecture employs a 4-level symmetric encoder-decoder structure. Each encoder-decoder level integrates n PGMamba blocks. Every PGMamba block takes image features and the physical priors $\mathbf{I}_p$ as input. To facilitate feature restoration, skip connections are established between corresponding encoder and decoder stages. The enhanced output is obtained through residual learning, where the final 3×3 convolutional features are added to the original input image $\mathbf{I}$, yielding clear image $\hat{\mathbf{I}}$.

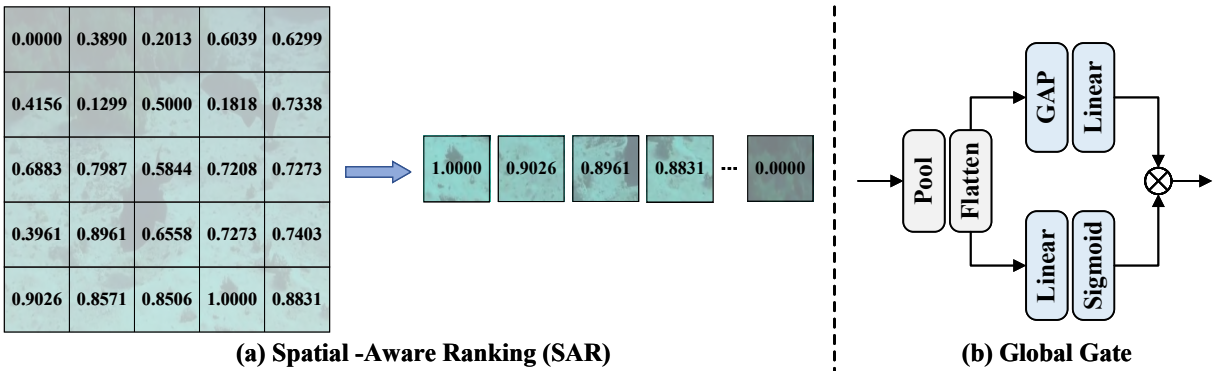
PGMamba Block

As the core operation for feature extraction and aggregation, our PGMamba block adopts a Transformer-like architecture. However, the substantial computational demands of self-attention and cross-attention mechanisms significantly impair the efficiency of Transformer-based approaches, while the complexity of underwater environments makes standard FFN unsuitable for direct application in UIE. To address these limitations, we eliminate both self-attention computations and conventional FFN, replacing them with SAGMamba and PMGFFN, as shown in Fig. 2(c). Given the input feature $\mathbf{F}$ and physical prior $\mathbf{I}_p$, the feature flow of our PGMamba block can be described as:

$$\mathbf{F}^{'} = \mathbf{F} + \mathrm{SAGMamba}(\mathbf{F}_{ln}) \tag{1}$$

$$\mathbf{F}_p = \mathrm{GELU}(\mathrm{Conv}(\mathbf{I}_p)) \tag{2}$$

$$\hat{\mathbf{F}} = \mathbf{F}^{'} + \mathrm{PMGFFN}(\mathbf{F}^{'}_{ln}, \mathbf{F}_p) \tag{3}$$



(a) Spatial -Aware Ranking (SAR) **(b) Global Gate**

Figure 3: The SAR and Global Gate.

where $\mathbf{F}_{ln}$ is the layer-normalized form of $\mathbf{F}$.

Spatial-Aware Global Mamba

Existing SSM-based methods primarily originate from the Mamba architecture. Formally, Mamba employs discrete state space equations to model interactions between tokens, with the mathematical representation given by:

$$\begin{cases} h_t = \overline{\mathbf{A}}_t h_{t-1} + \overline{\mathbf{B}}_t x_t \\ y_t = \mathbf{C}_t h_t + \mathbf{D} x_t \end{cases} \tag{4}$$

where $\overline{\mathbf{A}} = \exp(\Delta\mathbf{A})$ represents the state transition matrix, $\overline{\mathbf{B}} = (\Delta\mathbf{A})^{-1}(\exp(\Delta\mathbf{A}) - \mathbf{I})\Delta\mathbf{B}$ denotes the input matrix, $\mathbf{C}$ is the output matrix, and $\mathbf{D}$ stands for the skip-connection matrix. In Mamba, the parameters achieve context-aware and adaptive processing through a selection function. This is implemented by modifying parameters Δ, $\mathbf{B}$ and $\mathbf{C}$ into simple functions of the input sequence x_t, thereby obtaining input-dependent parameters $\Delta = f_\Delta(x_t)$, $\mathbf{B} = f_B(x_t)$ and $\mathbf{C} = f_C(x_t)$.

Existing Mamba-based methods typically use unidirectional or multi-directional scanning strategies to flatten 2D images into 1D sequences for sequential modeling. However, in unidirectional scanning, the i^{th} pixel can only access the previous $i-1$ pixels, failing to capture global information, while multidirectional scanning increases computational complexity despite expanding the receptive field. From Eq. (4), we observe that the stability of the state equation depends only on $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$, independent of $\mathbf{C}$, where $\mathbf{C}$ only regulates local dynamics. Directly modifying $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ may disrupt the inherent stability of the state evolution. Therefore, we propose injecting global information $\mathbf{g}$ into the output matrix $\mathbf{C}$, incorporating the gradient information of global context into local dynamics to enrich feature representation. This design maintains the stability of the state equation while enabling global perception within a unidirectional scanning framework. Thus, the modified state space equation can be rewritten as:

$$\begin{cases} h_t = \overline{\mathbf{A}}_t h_{t-1} + \overline{\mathbf{B}}_t x_t \\ y_t = (\mathbf{C}_t + \mathbf{g}) h_t + \mathbf{D} x_t \end{cases} \tag{5}$$

where $\mathbf{g}$ contains global contextual information from input x and is time-step invariant.

Building upon these findings, we propose the SAGMamba, as illustrated in Fig. 2(d). It is well-established that image data typically contains substantial redundant features. While several studies have explored sparse learning to address this challenge (Chen et al. 2023; Zhou et al. 2024b; Lou, Fu, and Yu 2025), our work adopts a single-sparsity strategy with maxpooling as the sparsification operation. Considering that feature importance ranking enables more effective modeling of long-range dependencies in Mamba (Guo et al. 2025; Dong et al. 2024), we employ a Spatial-Aware Ranking (SAR) strategy to process the sparsified features, as illustrated in Fig. 3(a).

To achieve effective integration of global contextual information with SSM, we design a simple global gate, as illustrated in Fig. 3(b). Specifically, the input features undergo pooling and flattening before being processed through two parallel branches: one branch employs global average pooling followed by linear transformation, while the other utilizes linear layers with Sigmoid activation. The outputs of these pathways are multiplied to implement the gating operation, with the resultant global information $\mathbf{g}$ being incorporated into our enhanced SSM.

The improved SSM simultaneously processes both the spatially-ranked features and global gated features, enabling the joint capture of rich contextual information and spatial awareness. Finally, local enhancement is achieved through feature sequence restoration and transposed convolution operations, yielding the optimized output of SAGMamba.

Physical Model-Guided Feed-Forward Network

Conventional FFNs often fail to account for the unique characteristics of underwater images, leading to limited enhancement accuracy. To address this, we combine an underwater physical model (Akkaynak and Treibitz 2018) as the guiding principle with FFN. The model is formally expressed as:

$$\mathbf{I}(\mathbf{x}) = \mathbf{J}(\mathbf{x})\mathbf{T}(\mathbf{x}) + \mathbf{A}(\mathbf{1} - \mathbf{T}(\mathbf{x})) \tag{6}$$

where $\mathbf{I}$ denotes the observed degraded underwater image, $\mathbf{J}$ represents the target enhanced image, $\mathbf{A}$ is the global background light intensity, and $\mathbf{T}$ is the medium transmission map. Following the Beer-Lambert law (Swinehart 1962), the transmission map can be further defined as $\mathbf{T}(\mathbf{x}) = e^{-\beta d(x)}$.

Here, β characterizes the water attenuation coefficient, and d is the scene depth map. To achieve underwater image enhancement, we reformulate the imaging model as:

$$\mathbf{J}(\mathbf{x}) = \frac{1}{\mathbf{T}(\mathbf{x})}\mathbf{I}(\mathbf{x}) + \mathbf{A}(\mathbf{1} - \frac{1}{\mathbf{T}(\mathbf{x})}) \tag{7}$$

Previous approaches typically require complex networks to jointly estimate multiple physical parameters. Recognizing that gradient maps effectively capture local blurring caused by light scattering, while depth maps encode critical information about water attenuation and background light distribution, we design a simple physical parameter estimation method. Specifically, we derive the depth map $\mathbf{I}_d$ and gradient map $\mathbf{I}_g$ from the input image $\mathbf{I}$ based on DCPSIR (Peng, Cao, and Cosman 2018) and concatenate them to form a physical prior $\mathbf{I}_p$, as shown in Fig. 2(b).

As illustrated in Fig. 2(d), the proposed PMGFFN employs a dual-branch architecture, the first branch processes the input feature $\mathbf{F}_{ln}^{'}$ through convolutional operations to obtain $\mathbf{F}_e$, while the second branch is the prior branch which enhances local features $\mathbf{F}_p$ of $\mathbf{I}_p$ via a large kernel (5×5) convolution, followed by a 3×3 convolution with Sigmoid function to decompose the transmission inverse $\mathbf{t}$ and background light $\mathbf{A}$, where $\mathbf{t}$ is the reciprocal of $\mathbf{T}$. Strictly adhering to the physical model, the network combines the outputs as $\mathbf{F}_e \times \mathbf{t} + \mathbf{A} \times (\mathbf{1} - \mathbf{t})$, followed by a 3×3 convolution, GELU activation, and 1×1 convolution to produce the final enhanced result under the guidance of the physical model.

Loss Function

The loss function in this work combines L1 loss and structural similarity loss with weighted summation, expressed as:

$$\mathcal{L}(\mathbf{I}, \hat{\mathbf{I}}) = \mathcal{L}_1(x, y) + \omega \mathcal{L}_{ssim}(\mathbf{I}, \hat{\mathbf{I}}) \tag{8}$$

where ω represents the weighting coefficient, $\mathcal{L}_{ssim}$ denotes the structural similarity loss.

Experiments

Experiment Settings

Datasets. We conduct experiments on two widely-used UIE datasets: UIEB (Li et al. 2019) and LSUI (Peng, Zhu, and Bian 2023). The UIEB dataset consists of a full-reference subset with 890 image pairs, from which we randomly select 800 pairs for training and 90 pairs for testing, and a non-reference subset containing 60 challenging degraded images (denoted as C60). The LSUI dataset comprises 4279 pairs of degraded and clear images, with 3879 pairs randomly selected for training and 400 pairs for testing. In all experiments, we uniformly resize all images to 256×256.

Implementation Details. Our method is evaluated on an NVIDIA GeForce RTX 4090 GPU. The number of PGMamba n in each level is set to 4. For optimization, we employ the Adam optimizer with momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate follows a cosine annealing schedule with an initial value of 0.0004. During training, we set the batch size to 8 and train the model for 200 epochs. Random horizontal flipping is applied as a data augmentation strategy to enhance generalization. The factor ω in the loss function is set to 0.1.

Evaluation Metrics. For full-reference evaluation, we employ two main metrics: PSNR and SSIM (Wang et al. 2004), where higher values indicate better image quality. In addition, we employ LPIPS (Zhang et al. 2018) and FID (Heusel et al. 2017) metrics for full-reference image evaluation. LPIPS is a deep neural network-based image quality metric that assesses perceptual similarity between image pairs. FID measures the distance between two image distributions. For non-reference assessment, we utilize UIQM (Panetta, Gao, and Agaian 2015) and UCIQE (Yang and Sowmya 2015) to evaluate the enhancement performance.

Comparison methods. We conduct comprehensive comparisons between PGMamba and eight state-of-the-art (SOTA) UIE methods: WaterNet (Li et al. 2019), SUWNet (Naik, Swarnakar, and Mittal 2021), PUIE (Fu et al. 2022), UShape (Peng, Zhu, and Bian 2023), NU2Net (Guo et al. 2023), DM-Water (Tang, Kawasaki, and Iwaguchi 2023), WFDiff (Zhao et al. 2024a), and SS-UIE (Peng and Bian 2025). For fair comparison, all methods are evaluated using their officially released source codes while strictly adhering to the experimental settings specified in their publications.

Full-Reference Evaluation. As demonstrated in Table 1, the proposed PGMamba achieves best performance across nearly all full-reference evaluation metrics. Notably, it exhibits significant advantages in PSNR and SSIM. To provide a comprehensive assessment, Fig. 4 presents visual comparisons between PGMamba and other methods. Four representative samples randomly selected from the UIEB and LSUI show that our enhanced results demonstrate superior color fidelity, detail restoration, and visual naturalness, closely matching the reference images. These results provide compelling evidence for PGMamba's superior generalization in practical applications.

Non-Reference Evaluation. For non-reference assessment, the quantitative results and visual comparisons are presented in Table 1 and Fig. 5, respectively. Quantitative analysis reveals our method achieves the highest scores on both UIQM and UCIQE, demonstrating PGMamba's exceptional generalization performance. Visual comparisons in Fig. 5 clearly show that existing methods, lacking physical model guidance, produce enhanced images with insufficient global contrast and lost local details. In contrast, PGMamba effectively incorporates physical priors to improve detail restoration while maintaining natural image quality.

Efficiency Analysis. As shown in Table 1, PGMamba achieves a favorable balance between model parameters (Params) and floating-point operations per second (Flops). With a moderate parameter size, it attains substantially lower FLOPs than most competing methods. Notably, its computational cost approaches that of the most lightweight architectures while delivering superior performance. These results demonstrate that PGMamba effectively optimizes the trade-off between model capacity and computational efficiency.

Ablation Experiment

Contribution Analysis of PGMamba. The ablation studies presented in Table 2 systematically validate the importance

	Method	WaterNet TIP'2019	SUWNet AAAI'2021	PUIE ECCV'2022	UShape TIP'2023	NU2Net AAAI'2023	DMWater MM'2023	WFDiff CVPR'2024	SSUIE AAAI'2025	Ours
UIEB	FID↓	40.43	58.68	**35.65**	73.08	51.38	42.49	48.60	62.98	44.30
	LPIPS↓	0.1111	0.1804	0.0995	0.1670	0.0873	0.0754	0.1126	0.1500	**0.0740**
	PSNR↑	21.66	18.79	21.10	19.98	23.28	22.99	21.57	19.48	**24.34**
	SSIM↑	0.8846	0.7363	0.8838	0.7651	0.9146	0.9165	0.8736	0.8379	**0.9235**
LSUI	FID↓	27.37	43.64	27.03	35.05	29.91	22.40	25.16	31.85	**21.17**
	LPIPS↓	0.1344	0.2020	0.1437	0.1116	0.1123	0.1089	0.0917	0.1180	**0.0769**
	PSNR↑	23.60	25.27	24.16	24.41	26.10	27.79	26.44	25.45	**28.66**
	SSIM↑	0.8880	0.8203	0.8934	0.8511	0.9139	0.9217	0.9067	0.8740	**0.9314**
C60	UIQM↑	2.779	2.393	2.568	2.807	2.860	2.874	2.828	2.764	**3.2037**
	UCIQE↑	0.5871	0.5234	0.5745	0.5474	0.5825	0.5868	0.5648	0.5443	**0.5894**
Params(M)↓		1.09	**0.22**	1.40	65.60	3.15	10.71	43.72	46.31	4.61
Flops(G)↓		13.7	21.6	30.0	66.2	**10.5**	67.1	94.1	53.9	15.7

Table 1: Quantitative comparison of different UIE methods on UIEB, LSUI and C60 datasets. The best results are highlighted in bold and the second best results are underlined.

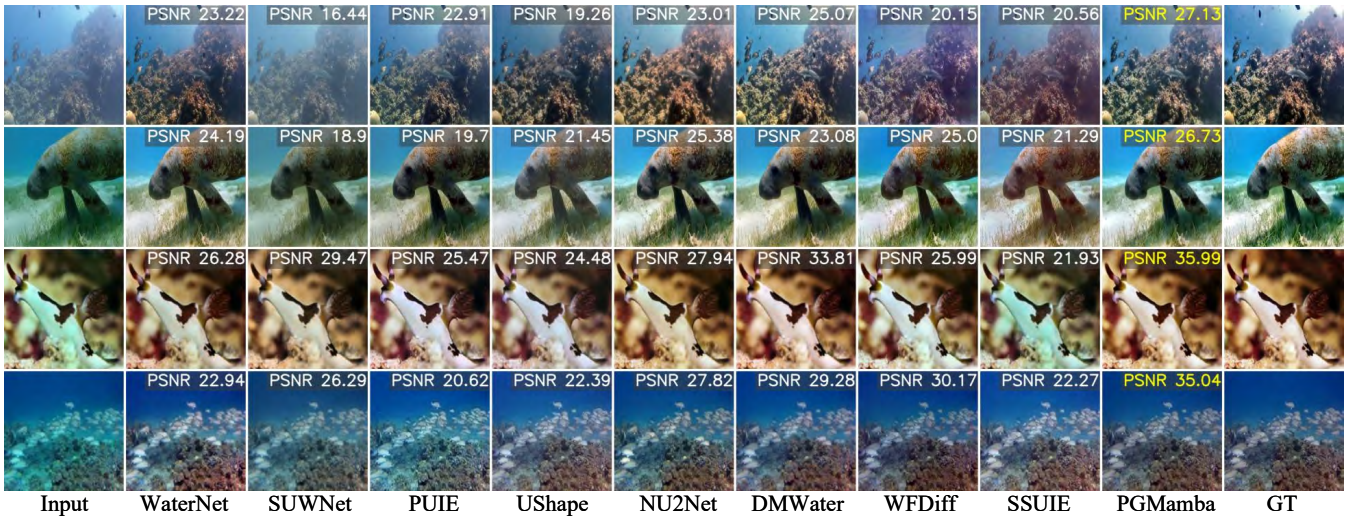


Figure 4: Visual comparison with other SOTA methods on UIEB and LSUI. The highest PSNR scores are marked in yellow.

Method	PSNR	SSIM	Params(M)	Flops(G)
w/o g	23.75	0.9169	4.57	**15.6**
w/o maxpool	22.33	0.8923	4.60	24.9
Mamba	22.73	0.9115	5.55	39.5
self-attention	23.05	0.9142	**4.14**	20.3
SAGMamba	**24.34**	**0.9235**	4.61	15.7

Table 2: Ablation study of SAGMamba on UIEB.

of PGMamba. The performance degradation observed upon removing the global gate confirms the critical role of global information in feature enhancement. Furthermore, the significant accuracy drop when eliminating the maxpool operation demonstrates the substantial impact of redundant features on model performance. Notably, comparative experiments show that neither self-attention (Zamir et al. 2022) nor Mamba block (Dong et al. 2024) can match the performance of PGMamba, highlighting its superior design.

Contribution Analysis of global gate in PGMamba. As shown in Table 3, adding output $\mathbf{g}$ of global gate into $\overline{\mathbf{A}}$ in state equation leads to performance degradation compared to removing $\mathbf{g}$ entirely, because state space is affected. While adding $\mathbf{g}$ to $\overline{\mathbf{B}}$ or $\mathbf{D}$ yields slight improvement, it still underperforms relative to inserting $\mathbf{g}$ into $\mathbf{C}$. Moreover, replacing $\mathbf{g}$ with model weights as global information (denoted as $\mathbf{p}$) (Guo et al. 2025) results in inferior performance, further validating the effectiveness of our proposed design.

Contribution Analysis of PMGFFN. For PMGFFN, the ablation results in Table 4 provide compelling evidence of its effectiveness. Complete removal of this module leads to measurable performance deterioration, while specifically disabling its prior branch (PB) also results in degraded output quality, verifying the essential role of physical model constraints. The inferior performance observed when substituting PMGFFN with either Restormer's GDFN (Zamir et al. 2022) or ViT's standard FFN (Dosovitskiy et al. 2020) fur-

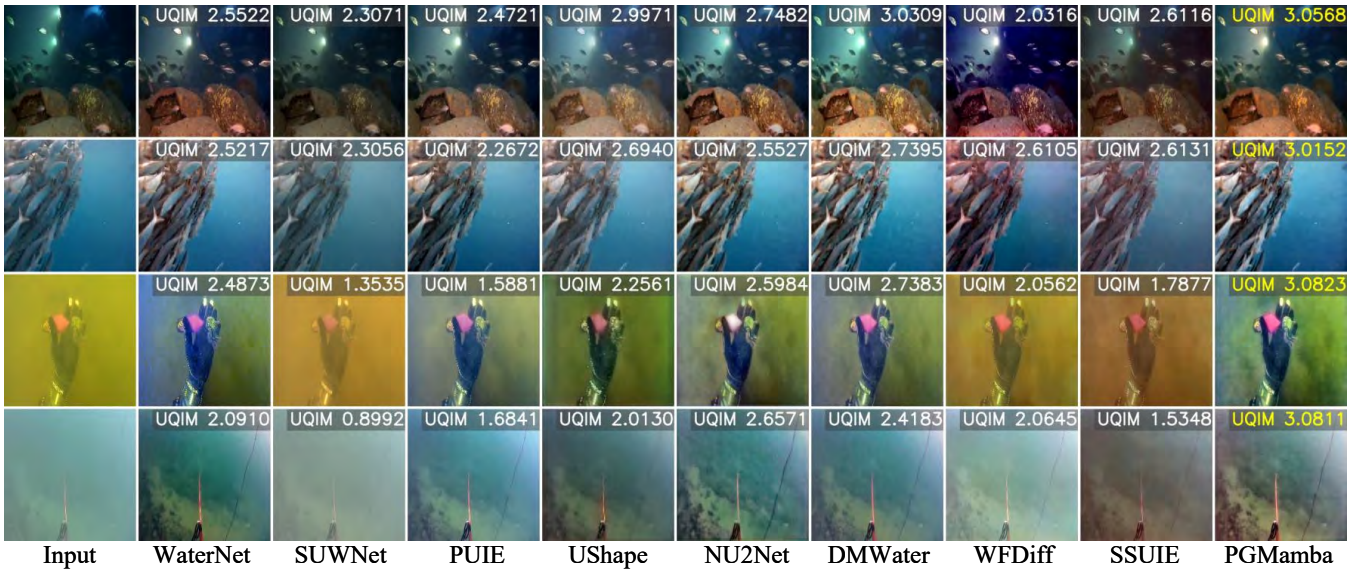


Figure 5: Visual comparison with other SOTA methods on C60. The highest UIQM scores are marked in yellow.

Method	PSNR	SSIM
w/o g	23.75	0.9169
$\overline{\mathbf{A}}+\mathbf{g}$	22.70	0.8957
$\overline{\mathbf{B}}+\mathbf{g}$	23.95	0.9198
$\mathbf{D}+\mathbf{g}$	23.85	0.9129
$\mathbf{C}+\mathbf{p}$	23.31	0.9064
$\mathbf{C}+\mathbf{g}$	**24.34**	**0.9235**

Table 3: Ablation study of global gate on UIEB.

Method	PSNR	SSIM	Params(M)	Flops(G)
w/o PMGFFN	23.82	0.9126	**2.95**	**5.9**
w/o PB	23.89	0.9170	4.35	12.5
FFN	23.59	0.9019	4.26	11.7
GDFN	23.66	0.9182	5.65	18.4
PMGFFN	**24.34**	**0.9235**	4.61	15.7

Table 4: Ablation study of PMGFFN on UIEB.

ther substantiates the unique advantages of our PMGFFN. These comprehensive experiments collectively demonstrate the value of our proposed architectural innovations.

Conclusion

In this work, we propose PMGMamba for UIE. The method incorporates global information into Mamba to enhance the performance. Furthermore, we incorporate physical model into FFN, achieving synergistic integration of physical constraints with data-driven learning. Experimental results demonstrate that PGMamba outperforms current methods. While PGMamba achieves promising results, the method remains dependent on physical priors. Nevertheless, under extreme conditions, such priors may degrade the model's performance. Future research could explore architectures that minimize the environmental influence on prior information, thereby mitigating the degradation in model accuracy.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62371076 and Grant 62201109; in part by the New Chongqing Youth Innovative Talents Project under Grant CSTB2024NSCQ-QCXMX0071, and in part by the Natural Science Foundation of Chongqing under Grant CSTB2024NSCQ-MSX0393.

References

Akkaynak, D.; and Treibitz, T. 2018. A revised underwater image formation model. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6723–6732.

Akkaynak, D.; and Treibitz, T. 2019. Sea-thru: A method for removing water from underwater images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1682–1691.

Akkaynak, D.; Treibitz, T.; Shlesinger, T.; Loya, Y.; Tamir, R.; and Iluz, D. 2017. What is the space of attenuation coefficients in underwater computer vision? In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4931–4940.

Alsakar, Y. M.; Sakr, N. A.; El-Sappagh, S.; Abuhmed, T.; and Elmogy, M. 2025. Underwater image restoration and enhancement: a comprehensive review of recent trends, challenges, and applications. *The Visual Computer*, 41(6): 3735–3783.

Ancuti, C.; Ancuti, C. O.; Haber, T.; and Bekaert, P. 2012. Enhancing underwater images and videos by fusion. In *2012 IEEE conference on computer vision and pattern recognition*, 81–88. IEEE.

Cai, M.; Wang, Y.; Wang, S.; Wang, R.; Ren, Y.; and Tan, M. 2020. Grasping marine products with hybrid-driven underwater vehicle-manipulator system. *IEEE transactions on automation science and engineering*, 17(3): 1443–1454.

Chen, X.; Li, H.; Li, M.; and Pan, J. 2023. Learning a sparse transformer network for effective image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5896–5905.

Chiang, J. Y.; and Chen, Y.-C. 2011. Underwater image enhancement by wavelength compensation and dehazing. *IEEE transactions on image processing*, 21(4): 1756–1769.

Cong, R.; Yang, W.; Zhang, W.; Li, C.; Guo, C.-L.; Huang, Q.; and Kwong, S. 2023. Pugan: Physical model-guided underwater image enhancement using gan with dual-discriminators. *IEEE Transactions on Image Processing*, 32: 4472–4485.

Dong, W.; Zhou, H.; Zhang, Y.; Liu, X.; and Chen, J. 2024. Ecmamba: Consolidating selective state space model with retinex guidance for efficient multiple exposure correction. *Advances in Neural Information Processing Systems*, 37: 53438–53457.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Drews, P.; Nascimento, E.; Moraes, F.; Botelho, S.; and Campos, M. 2013. Transmission estimation in underwater single images. In *Proceedings of the IEEE international conference on computer vision workshops*, 825–830.

Fu, Z.; Wang, W.; Huang, Y.; Ding, X.; and Ma, K.-K. 2022. Uncertainty inspired underwater image enhancement. In *European conference on computer vision*, 465–482. Springer.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2020. Generative adversarial networks. *Communications of the ACM*, 63(11): 139–144.

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

Guo, C.; Wu, R.; Jin, X.; Han, L.; Zhang, W.; Chai, Z.; and Li, C. 2023. Underwater ranker: Learn which is better and how to be better. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 702–709.

Guo, H.; Guo, Y.; Zha, Y.; Zhang, Y.; Li, W.; Dai, T.; Xia, S.-T.; and Li, Y. 2025. Mambairv2: Attentive state space restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 28124–28133.

Guo, H.; Li, J.; Dai, T.; Ouyang, Z.; Ren, X.; and Xia, S.-T. 2024. Mambair: A simple baseline for image restoration with state-space model. In *European conference on computer vision*, 222–241. Springer.

Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; and Hochreiter, S. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.

Huang, T.; Pei, X.; You, S.; Wang, F.; Qian, C.; and Xu, C. 2024. Localmamba: Visual state space model with windowed selective scan. In *European Conference on Computer Vision*, 12–22. Springer.

Islam, M. J.; Xia, Y.; and Sattar, J. 2020. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2): 3227–3234.

Kar, A.; Dhara, S. K.; Sen, D.; and Biswas, P. K. 2021. Zero-shot single image restoration through controlled perturbation of koschmieder's model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16205–16215.

Khan, R.; Mishra, P.; Mehta, N.; Phutke, S. S.; Vipparthi, S. K.; Nandi, S.; and Murala, S. 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.

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 Transactions on Image Processing*, 30: 4985–5000.

Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; and Tao, D. 2019. An underwater image enhancement benchmark dataset and beyond. *IEEE transactions on image processing*, 29: 4376–4389.

Li, C.; Guo, J.; Chen, S.; Tang, Y.; Pang, Y.; and Wang, J. 2016. Underwater image restoration based on minimum information loss principle and optical properties of underwater imaging. In *2016 IEEE International Conference on Image Processing (ICIP)*, 1993–1997. IEEE.

Li, C.; Guo, J.; Guo, C.; Cong, R.; and Gong, J. 2017. A hybrid method for underwater image correction. *Pattern Recognition Letters*, 94: 62–67.

Li, Y.; Mi, Z.; Wang, Y.; Jiang, S.; and Fu, X. 2024. TAFormer: A Transmission-Aware Transformer for Underwater Image Enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*.

Lianghui, Z.; Bencheng, L.; Qian, Z.; Xinlong, W.; Wenyu, L.; and Xinggang, W. 2024. Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model. In *International Conference on Machine Learning (ICML)*.

Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; Jiao, J.; and Liu, Y. 2024. Vmamba: Visual state space model. *Advances in neural information processing systems*, 37: 103031–103063.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022.

Lou, M.; Fu, Y.; and Yu, Y. 2025. Sparx: A sparse cross-layer connection mechanism for hierarchical vision mamba and transformer networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 19104–19114.

Mishra, P.; Mehta, N.; Vipparthi, S. K.; and Murala, S. 2025. Uswformer: Efficient sparse wavelet transformer for underwater image enhancement. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 3372–3382. IEEE.

Naik, A.; Swarnakar, A.; and Mittal, K. 2021. Shallow-uwnet: Compressed model for underwater image enhancement (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 15853–15854.

Panetta, K.; Gao, C.; and Agaian, S. 2015. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3): 541–551.

Peng, L.; and Bian, L. 2025. Adaptive Dual-domain Learning for Underwater Image Enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 6461–6469.

Peng, L.; Zhu, C.; and Bian, L. 2023. U-shape transformer for underwater image enhancement. *IEEE Transactions on Image Processing*, 32: 3066–3079.

Peng, Y.-T.; Cao, K.; and Cosman, P. C. 2018. Generalization of the dark channel prior for single image restoration. *IEEE Transactions on Image Processing*, 27(6): 2856–2868.

Peng, Y.-T.; and Cosman, P. C. 2017. Underwater image restoration based on image blurriness and light absorption. *IEEE transactions on image processing*, 26(4): 1579–1594.

Qi, B.; Gao, J.; Li, D.; Zhang, K.; Liu, J.; Wu, L.; and Zhou, B. 2024. S4++: Elevating Long Sequence Modeling with State Memory Reply.

Quan, Y.; Tan, X.; Huang, Y.; Xu, Y.; and Ji, H. 2024. Enhancing Underwater Images via Asymmetric Multi-Scale Invertible Networks. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 6182–6191.

Shang, J.; Li, Y.; Xing, H.; and Yuan, J. 2025. LGT: Luminance-guided transformer-based multi-feature fusion network for underwater image enhancement. *Information Fusion*, 102977.

Swinehart, D. F. 1962. The beer-lambert law. *Journal of chemical education*, 39(7): 333.

Tang, Y.; Kawasaki, H.; and Iwaguchi, T. 2023. Underwater image enhancement by transformer-based diffusion model with non-uniform sampling for skip strategy. In *Proceedings of the 31st ACM international conference on multimedia*, 5419–5427.

Wang, Y.; Liu, H.; and Chau, L.-P. 2017. Single underwater image restoration using adaptive attenuation-curve prior. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 65(3): 992–1002.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.

Xiao, C.; Li, M.; Zhang, Z.; Meng, D.; and Zhang, L. 2024. Spatial-Mamba: Effective Visual State Space Models via Structure-Aware State Fusion. *arXiv preprint arXiv:2410.15091*.

Yang, M.; and Sowmya, A. 2015. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12): 6062–6071.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5728–5739.

Zhang, D.; Zhou, J.; Guo, C.; Zhang, W.; and Li, C. 2024. Synergistic multiscale detail refinement via intrinsic supervision for underwater image enhancement. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 7033–7041.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.

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 Transactions on Image Processing*, 31: 3997–4010.

Zhang, Z.; Jiang, Z.; Liu, J.; Fan, X.; and Liu, R. 2023. Waterflow: Heuristic normalizing flow for underwater image enhancement and beyond. In *Proceedings of the 31st ACM International conference on multimedia*, 7314–7323.

Zhang, Z.; Jiang, Z.; Ma, L.; Liu, J.; Fan, X.; and Liu, R. 2025. HUPE: Heuristic Underwater Perceptual Enhancement with Semantic Collaborative Learning. *International Journal of Computer Vision*, 1–19.

Zhao, C.; Cai, W.; Dong, C.; and Hu, C. 2024a. 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.

Zhao, C.; Cai, W.; Dong, C.; and Zeng, Z. 2024b. Toward sufficient spatial-frequency interaction for gradient-aware underwater image enhancement. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3220–3224. IEEE.

Zhou, J.; He, Z.; Lam, K.-M.; Wang, Y.; Zhang, W.; Guo, C.; and Li, C. 2024a. AMSP-UOD: When vortex convolution and stochastic perturbation meet underwater object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 7659–7667.

Zhou, S.; Chen, D.; Pan, J.; Shi, J.; and Yang, J. 2024b. Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2952–2963.

Zhuang, J.; Zheng, Y.; Guo, B.; and Yan, Y. 2024. Globally Deformable Information Selection Transformer for Underwater Image Enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*.

Zuo, S.; Liu, X.; Jiao, J.; Charles, D.; Manavoglu, E.; Zhao, T.; and Gao, J. 2022. Efficient long sequence modeling via state space augmented transformer. *arXiv preprint arXiv:2212.08136*.