


Article

FACMamba: Frequency-Aware Coupled State Space Modeling for Underwater Image Enhancement

Li Wang ¹ , Keyong Shen ^{2,*}, Haiyang Sun ¹, Xiaoling Cheng ², Jun Zhu ¹ and Bixuan Wang ³

¹ School of Computer and Software, Nanjing University of Industry Technology, Nanjing 210023, China; li1019wang@niit.edu.cn (L.W.); sunhy@niit.edu.cn (H.S.); junzhu@niit.edu.cn (J.Z.)

² School of Computer Information and Engineering, Nanchang Institute of Technology, Nanchang 330044, China

³ School of Electronic Science and Engineering, Southeast University, Nanjing 211189, China; dlutwbx2021@gmail.com

* Correspondence: sky@nut.edu.cn

Abstract

Recent advances in underwater image enhancement (UIE) have achieved notable progress using deep learning techniques; however, existing methods often struggle with limited receptive fields, inadequate frequency modeling, and poor structural perception, leading to sub-optimal visual quality and weak generalization in complex underwater environments. To tackle these issues, we propose FACMamba, a Mamba-based framework augmented with frequency-aware mechanisms, enabling efficient modeling of long-range spatial relations for underwater image restoration. Specifically, FACMamba incorporates three key components: a Multi-Directional Vision State-Space Module (MVSM) to model directional spatial context via the proposed 8-direction selective scan block (SS8D), a Frequency-Aware Guidance Module (FAGM) for learning informative frequency representations with low overhead, and a Structure-Aware Fusion Module (SAFM) to preserve fine-grained structural cues through adaptive multi-scale integration. Recognizing the importance of spatial-frequency interaction, our model fuses these representations via lightweight architecture to enhance both texture and color fidelity. Experiments on standard UIE benchmarks demonstrate that FACMamba achieves a favorable balance between enhancement quality and computational efficiency, outperforming many existing UIE methods.

Keywords: underwater image enhancement; frequency domain; mamba; perceptual fusion



Academic Editor: Weicheng Cui

Received: 4 November 2025

Revised: 21 November 2025

Accepted: 24 November 2025

Published: 27 November 2025

Citation: Wang, L.; Shen, K.; Sun, H.; Cheng, X.; Zhu, J.; Wang, B.

FACMamba: Frequency-Aware Coupled State Space Modeling for Underwater Image Enhancement. *J. Mar. Sci. Eng.* **2025**, *13*, 2258. <https://doi.org/10.3390/jmse13122258>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Underwater environments exhibit complex optical properties, including light scattering, absorption, and reflection, which collectively degrade the quality of captured images. Consequently, acquired underwater images typically suffer from degradation characteristics such as low visibility, color distortion, and reduced contrast. These low-quality images not only impair human visual perception but also reduce the performance and efficiency of underwater robotic systems. For autonomous underwater vehicles (AUVs) and remotely operated vehicles (ROVs) in particular, improved visibility can significantly enhance navigation stability, obstacle avoidance capabilities, and mission planning capabilities. In marine ecological monitoring, clearer images facilitate accurate identification of marine species, assessment of coral reef health, and long-term environmental surveillance. These applications rely on high-level vision tasks performed on underwater images and videos. Severe visual degradation in underwater scenes, however, significantly undermines

the performance and reliability of these tasks. To improve the usability and accuracy of underwater visual information, it is crucial to acquire high-quality images. To this end, the core objective of Underwater Image Enhancement (UIE) is to boost visual quality and information expressiveness by suppressing light scattering effects, correcting color distortion, and restoring image details.

Addressing the underwater image enhancement problem, several recent studies have explored deep learning-based solutions. Among them, Li et al. [1] presented the Ucolor model, which employs a medium transmission-guided decoder to effectively address color bias and low contrast issues in underwater scenes. Peng et al. [2] reported a U-shape Transformer network where the attention mechanism enhances sensitivity to color channels and spatially degraded regions, thereby effectively eliminating color artifacts and casts. Han et al. [3] integrated an encoder–decoder architecture into a generative adversarial network to correct underwater images, enabling the model to simultaneously address pixel-level degradations while preserving both global appearance and local structural details. Tao et al. [4] enhanced CNN architectures with a dual channel-spatial attention mechanism, allowing the model to emphasize salient regions in underwater scenes adaptively. Chen et al. [5] developed a cross-scale Transformer network capable of preserving fine details and correcting color distortions by leveraging global context modeling. Fan et al. [6] devised an efficient multi-scale joint prior network that enhances color estimation accuracy under various underwater conditions by integrating background priors and multi-scale spatial priors. Despite their effectiveness, these methods still encounter limitations in computational efficiency and real-time applicability.

Frequency-domain analysis has emerged as a promising direction in underwater image research. Zhang et al. [7] pioneered the use of wavelet transform techniques in underwater image restoration, leveraging frequency-domain decomposition to achieve finer detail reconstruction. Liu et al. [8] introduced a frequency-domain attention mechanism capable of effectively differentiating useful signals from noise, thereby substantially enhancing reconstruction quality. Pramanick et al. [9] identified frequency-domain spatial regularities in underwater images, which serve as a foundational insight for designing more effective super-resolution techniques. Although the application of frequency-domain approaches in underwater image restoration is still relatively limited, their demonstrated success in related areas underscores their considerable potential for advancing underwater image enhancement.

The State Space Model (SSM) architecture, recently developed, offers notable advantages for image restoration by efficiently capturing long-range dependencies while maintaining linear computational costs [10]. Guo et al. [11] initiated the application of Mamba to natural image super-resolution and reported enhanced performance relative to Transformer-driven techniques, highlighting its potential for high-fidelity restoration tasks. Guan et al. [12] introduced a spatial-channel omnidirectional scan mechanism that captures bidirectional dependencies along both spatial and channel dimensions in four orientations, addressing pixel-channel coupling challenges. Lin et al. [13] utilized SSM to efficiently capture global dependencies, resulting in refined and high-quality enhancements of underwater images. Dong et al. [14] proposed an O-shaped dual-branch architecture that independently encodes spatial and cross-channel cues, exploiting globally receptive properties of state-space models for improved UIE. Collectively, these studies highlight a shift toward lightweight yet powerful architectures that improve restoration quality while maintaining computational efficiency.

Building on these advancements, we propose FACMamba, a novel multi-domain feature integration network that synergistically combines frequency-domain analysis with state-space modeling to enhance the efficiency and effectiveness of UIE tasks. To enable

effective feature interaction across scales, we introduce the Bidirectional Awareness Integration Module (BAIM) between encoder and decoder stages. The BAIM framework integrates three key components: a Multi-Directional Vision State-Space Module (MVSM) for capturing long-range dependencies, a Frequency-Aware Guidance Module (FAGM) to enhance high-frequency information, and a Structure-Aware Fusion Module (SAFM) for adaptive fusion of global and local features. Extensive evaluations on multiple underwater benchmarks confirm that FACMamba delivers better performance in visual quality and quantitative indices, while maintaining lower model complexity and computational overhead compared to existing approaches.

To summarize, this work makes the following three major contributions:

- To address the inefficiency of global modeling and detail enhancement in UIE tasks, a novel state-space-based U-Net architecture named FACMamba is proposed. FACMamba integrates a lightweight yet expressive Bidirectional Awareness Integration Module (BAIM) between the encoder and decoder stages to facilitate efficient bidirectional information flow across scales. Such a design allows the network to effectively integrate frequency-domain information with long-range spatial modeling, all within a framework of linear computational complexity.
- To cope with the inherent challenges of underwater image degradation, the proposed BAIM incorporates three submodules: the Multi-Directional Vision State-Space Module (MVSM) for long-range spatial interaction, the Frequency-Aware Guidance Module (FAGM) for high-frequency detail restoration, and the Structure-Aware Fusion Module (SAFM) for adaptive integration of global and local features. This comprehensive integration enhances the model's perceptual quality and robustness in diverse underwater scenarios.
- Comprehensive evaluations on multiple UIE benchmarks show that FACMamba delivers performance comparable to state-of-the-art methods while maintaining markedly lower computational complexity.

2. Related Work

2.1. Underwater Image Enhancement

Deep learning-based approaches [7,15,16] have risen as a prevailing direction in UIE task. Unlike traditional methods based on physical models [17–19], these approaches leverage end-to-end learning to automatically extract effective feature representations from underwater images, eliminating the need for hand-crafted features. Li et al. [15] put forward a CNN structure that eliminates the need for estimating underwater imaging model parameters, enabling direct image restoration through a fully end-to-end learning process. Building upon this, Li et al. [1] described Ucolor, a model that utilizes medium transmission-guided multi-color space embedding to mitigate color distortion and enhance contrast in underwater imagery. Incorporating physical priors into deep learning, Cong et al. [20] proposed a generative adversarial network (GAN) framework guided by a physical imaging model. The architecture features a Parameter Estimation network for recovering model parameters, a Two-Stream Enhancement module leveraging degradation quantization, and a dual-discriminator design aimed at improving image realism and perceptual quality. To further integrate semantic understanding into enhancement tasks, Qi et al. [21] proposed SGUIE-Net, a network that integrates a semantic region enhancement module to learn and leverage semantic-level features as high-level guidance for improving underwater image quality. Moving toward hybrid architectures, Peng et al. [2] designed a U-shaped Transformer architecture that integrates a channel-wise multi-scale feature fusion transformer module and a spatial-wise global feature modeling transformer module. This design reinforces the model's concern for severely degraded color and spatial areas. In

parallel, An et al. [22] presented a hybrid model named UWMamba, which couples the local feature extraction capabilities of convolution with the global modeling strength of SSM. To address the issue of spatial and chromatic inhomogeneous degradation in underwater images, they also designed the Mamba Attention Fusion Module for effective feature fusion and enhancement. In summary, UIE research is shifting toward hybrid designs that integrate global modeling, semantic guidance, and domain knowledge. However, effectively capturing long-range dependencies and handling diverse underwater conditions remains a challenge.

2.2. Fourier Transform

Fast Fourier Transform (FFT) offers a robust mathematical foundation for representing global image statistics in the frequency domain, inherently capturing long-range spatial dependencies [23]. Leveraging this capability, recent advances in computer vision have explored FFT-based methods suitable for numerous tasks, such as image restoration [24], semantic segmentation [25], and image classification [26]. In the context of underwater image enhancement, Cheng et al. [27] proposed FDCE-Net, which integrates frequency-domain learning with dual-color encoding to effectively mitigate color distortion and blurring, while maintaining computational efficiency. Walia et al. [28] employed frequency-domain priors to guide spatial reconstruction, thereby improving image clarity and structural detail in underwater scenes. Zhao et al. [29] put forward a wavelet-based Fourier interaction network that jointly incorporates multi-scale frequency diffusion to enhance texture fidelity and structural consistency. Similarly, Zhang et al. [7] proposed a frequency-spatial fusion network that combines FFT-extracted frequency features with spatial features, effectively enhancing underwater images while maintaining detail fidelity and suppressing artifacts. Zhu et al. [30] proposed a Fourier transform-guided dual-channel diffusion network for UIE task, where frequency-domain structural information is leveraged to improve the quality of the diffusion model's input, thereby enhancing key attributes such as brightness and contrast. Observing Figure 1 shows that the phase component dominates the contour and detail information of the image, while the amplitude component affects the brightness and overall structure of the image. Overall, frequency-domain methods offer strong potential for underwater enhancement by modeling global structures and fine details. However, challenges remain in fully exploiting phase–amplitude relationships and integrating them with spatial-domain cues in a unified framework.

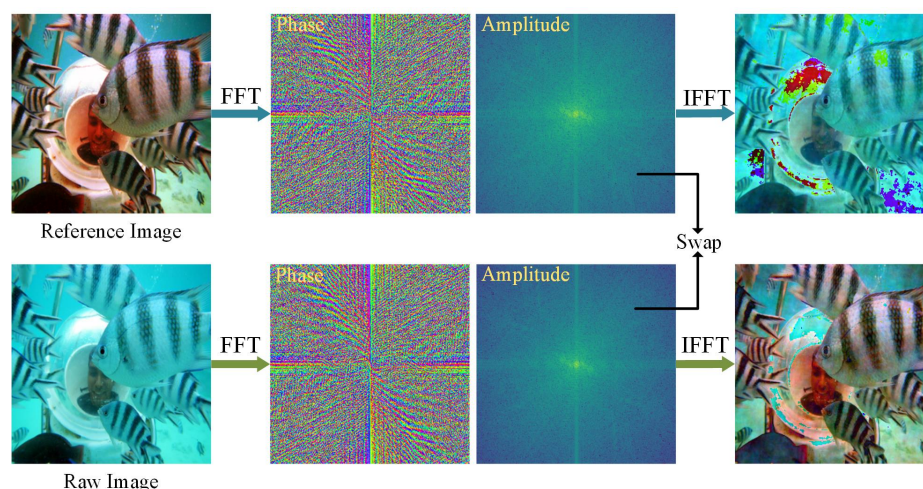


Figure 1. Observation of spectral exchange by FFT. The degradation occurs primarily in the amplitude component, and FFT can decouple the image's degradation information to a certain extent.

3. Proposed Method

3.1. Architecture

Our proposed FACMamba adopts a U-Net style design, as depicted in Figure 2, where a Bidirectional Awareness Integration Module (BAIM) is embedded between the encoder and decoder to promote efficient cross-scale information flow and exploit spatial-frequency interactions for superior enhancement quality.

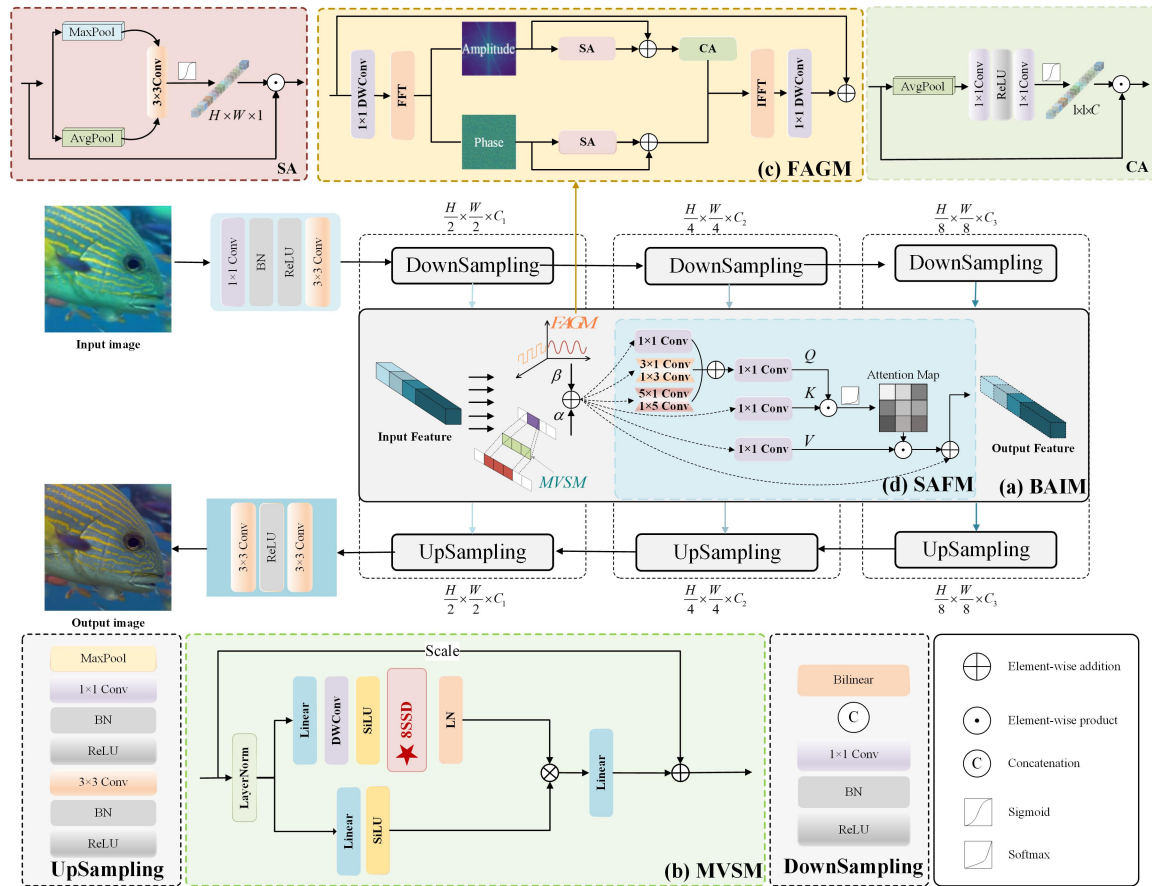


Figure 2. Overview of our FACMamba framework. (a) BAIM is embedded between the encoder and decoder to maintain seamless feature transmission and enhance cross-scale information flow. (b) The MVSM exploits directional-aware SS8D operations (red star marked) to model long-range spatial dependencies. (c) The FAGM performs frequency-domain enhancement by separately refining the amplitude and phase components via spatial and channel attention mechanisms. (d) The SAFM adaptively integrates multi-scale structural information through asymmetric convolutions and attention-based fusion.

3.2. Multi-Directional Vision State-Space Module

Multi-Directional Vision State-Space Module (MVSM) is designed to utilize the state space equations to capture long-distance dependencies, as illustrated in Figure 3.

The input features $F \in \mathbb{R}^{H \times W \times C}$ is initially divided into two parallel processing branches [31]. In the first pathway, the feature dimension is expanded to λC through a linear projection, followed by depth-wise convolution and activation using the SiLU function. The resulting representations are further enhanced via the 8-direction selective scan block (SS8D), which captures spatial dependencies from diverse directional perspectives, and are subsequently normalized by a LayerNorm operation. In the second pathway, the features are also transformed by a linear projection and SiLU activation to generate a modulation signal. The outcomes of two branches are then combined through an element-wise product and projected back to C channels via another linear transformation. Finally, a

learnable scaling parameter s regulates the flow of information, producing the final output $F_{MVSM} \in \mathbb{R}^{H \times W \times C}$:

$$\begin{aligned}
 \dot{F} &= LN(F) \\
 F' &= LN(SS8D(SiLU(DWConv(Linear(\dot{F})))))) \\
 F'' &= SiLU(Linear(\dot{F})) \\
 F_{MVSM} &= Linear(F' \odot F'') + s \cdot F
 \end{aligned} \tag{1}$$

where $DWConv(\cdot)$ is depth-wise convolution. \odot is the Hadamard product.

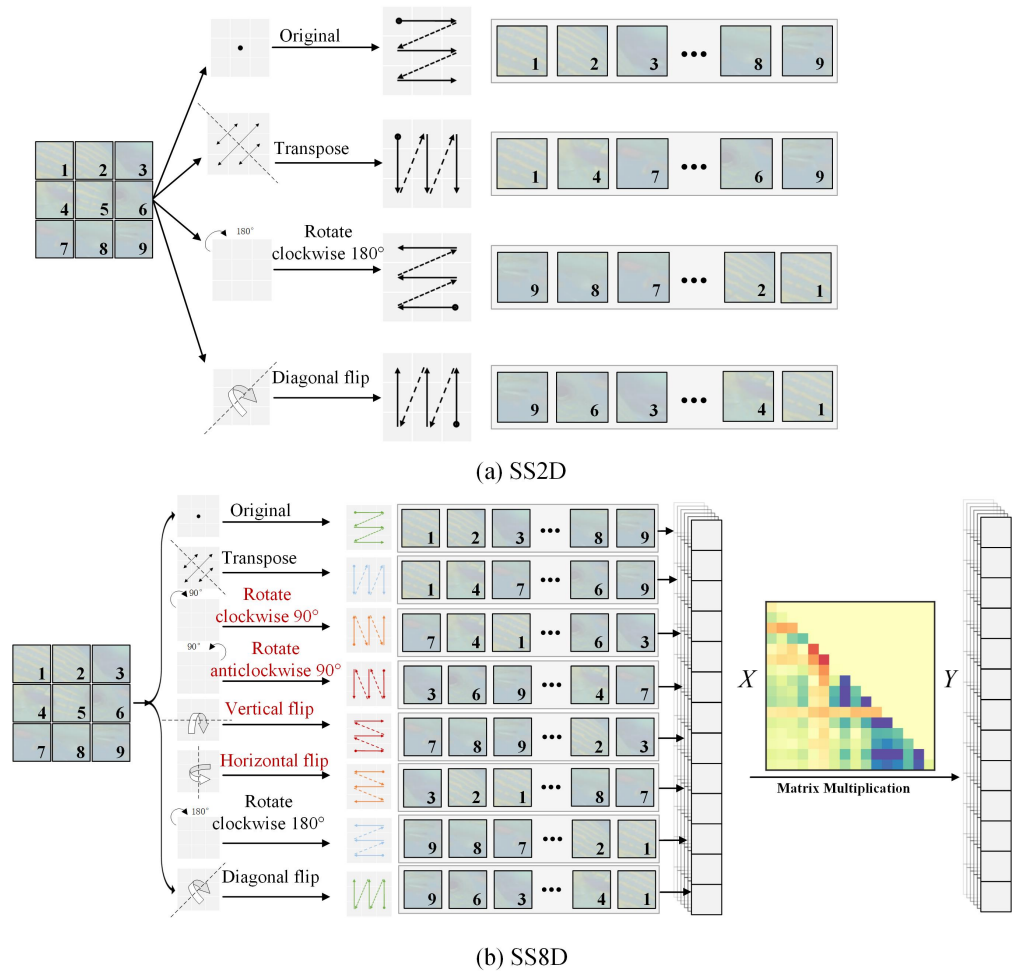


Figure 3. (a) Original SS2D using four basic scanning directions. (b) Proposed SS8D incorporating eight directional transformations for enhanced spatial dependency modeling and richer contextual feature extraction.

Because the refractive index of water in underwater environments differs from that of air, light refracts at the water-air interface, causing geometrical distortions in the image, particularly at the edges, which may lead to stretching or compression. The presence of suspended particles leads to wavelength-specific scattering and absorption, which degrades image quality through color distortion and blurring. To counteract these effects, we devise an SS8D to extract multi-directional features that capture geometric distortion patterns and compensate for color deviations, as illustrated in Figure 3b.

As shown in Figure 3a, the original 2D Selective Scan Module (SS2D) contains horizontal and vertical feature extraction, which cannot adequately capture diagonal or rotation-related feature information, and has limited ability to model complex geometries and texture changes in the underwater scene. Comparatively, our proposed 8-direction

selective scan block adds sub-diagonal directions as well as rotational (90° clockwise, 90° counterclockwise, 180°) features to provide richer orientation-awareness capabilities, especially when dealing with diagonally distributed targets (e.g., irregular targets such as corals and fishes), and the 8-direction extracted features are more robust. Plus, the fusion of 8-direction features expands the sensory field of the network, which helps to recover the global brightness, contrast, and color consistency of the image, thus showing more robustness and superiority in dealing with underwater color fading, scattering effects, and dynamic scenes.

3.3. Frequency-Aware Guidance Module

The high-frequency details of an image often undergo degradation, with the magnitude information specifically impacting the contrast and sharpness of images. Consequently, specialized processing of the magnitude information can effectively restore image details. Recognizing the advantages of frequency-based methods in addressing image degradation, we propose a Frequency-Aware Guidance Module (FAGM).

The input feature $F \in \mathbb{R}^{H \times W \times C}$ is initially passed through a convolutional layer, after which it is converted into the frequency domain via the Fourier transform, represented as $fre = FFT(Conv(F))$. The amplitude and phase components are obtained separately:

$$A = abs(fre), P = angle(fre) \tag{2}$$

Subsequently, we exploit Spatial Attention (SA) to enhance critical regions within the amplitude and phase information domains, thereby improving the representation ability of target features. Given that image degradation primarily affects the amplitude information, following SA operation in the amplitude branch, channel attention (CA) is then applied to emphasize critical channel features and suppress redundant ones. After the above operation, two refined components \bar{A} and \bar{P} are recovered. The process can be represented as:

$$\begin{aligned} \bar{A} &= H_{CA}(H_{SA}(A) + A) \\ \bar{P} &= H_{SA}(P) + P \end{aligned} \tag{3}$$

where $H_{SA}(\cdot)$ and $H_{CA}(\cdot)$ correspond to SA and CA operation. To enhance feature extraction efficiency while reducing computational complexity, the structures of SA and CA are illustrated in Figure 2.

The enhanced frequency-domain representations are obtained by combining two refined components:

$$\begin{cases} real = \bar{A} \times \cos(\bar{P}) \\ imag = \bar{A} \times \sin(\bar{P}) \end{cases} \tag{4}$$

$$fre_{ref} = complex(real, imag)$$

Finally, the output features F_{FAGM} are generated utilizing the Fourier inverse transform (IFFT).

$$F_{FAGM} = Conv\left(abs\left(IFTT\left(fre_{ref}\right)\right) + F\right) \tag{5}$$

In the literature [32], the authors process frequency domain information using only two 1×1 convolutional layers, resulting in low computational overhead. However, convolution operations struggle to differentiate the importance of various channels or regions in the frequency domain, particularly for highly degraded parts of the image or specific channel features, making targeted enhancement challenging. Contrastively, the proposed FAGM enhances the structure by incorporating an attention mechanism to more precisely emphasize important features while preserving the frequency domain structure.

Additionally, the attention mechanism adds adaptive learning capabilities, enabling the model to adapt more flexibly to images with varying levels of degradation.

3.4. Structure-Aware Fusion Module

The above analysis demonstrates that the MVSM module primarily emphasizes global feature extraction, enhancing the contrast and color reproduction of underwater images. In contrast, the FAGM module concentrates on enhancing high-frequency details and preserving structural integrity through frequency domain processing. To more effectively co-optimize global and local features, we propose SAFM to enhance the comprehensive perception of image degradation.

As depicted in Figure 2d, for a given input $F_f = \alpha \cdot F_{MVSM} + \beta \cdot F_{FAGM}$, where α and β are learnable parameters. SAFM directs the fused features F_f to three branches, processing them in a heterogeneous manner. The first branch combines structural information across different scales using parallel asymmetric convolutions and generates the query key Q via a 1×1 convolution, while the remaining two branches obtain the key K and value V using other 1×1 convolutions. After that, Softmax is used to compute the attention map between Q and K . Finally, it is weighted and summed with V to obtain the structure-aware fusion features F_{SAFM} .

$$\begin{aligned} Q &= Conv_{1 \times 1} \left(H_{Conv} \left(F_f \right) \right) \\ K &= Conv_{1 \times 1} \left(F_f \right) \\ V &= Conv_{1 \times 1} \left(F_f \right) \\ F_{SAFM} &= softmax(QK^T)V \end{aligned} \tag{6}$$

where $Conv(\cdot)$ represents the convolution operation, with the subscript indicating the size of the corresponding convolution kernel. $H_{Conv}(\cdot)$ contains parallel asymmetric convolution operations. $softmax(\cdot)$ is the operation of Softmax function.

3.5. Loss Function

To refine the perceptual appeal and authenticity of the generated outputs, we employ both L_2 loss and SSIM loss. L_2 loss more effectively maintains the overall brightness consistency of the image during the optimization process. However, it tends to over-smooth details during image restoration. To compensate for this, we incorporate SSIM loss to preserve the image's detail and texture information. Let X be the distorted underwater, Y be the ground truth, and \hat{Y} be the predicted image from FACMamba. L_2 loss and SSIM loss are calculated separately as follows:

$$\begin{aligned} L_2 &= E_{X,Y} [\|Y - \hat{Y}\|_2] \\ L_{SSIM} &= E_{X,Y} [1 - SSIM(Y - \hat{Y})] \end{aligned} \tag{7}$$

To optimize the model, the total loss is formulated as:

$$L = \gamma_1 L_2 + \gamma_2 L_{SSIM} \tag{8}$$

where γ_1 and γ_2 are set to 1 and 0.4, respectively, for balancing the loss term.

4. Experiments

4.1. Datasets

The proposed method is trained on the UIEB dataset [15], which contains 950 underwater images divided into three subsets: 800 for training (U800), 90 for validation (T90), and 60 for challenge evaluation (C60). Each image in U800 and T90 has a paired high-quality reference, while C60 consists only of raw degraded inputs. The dataset encompasses a

wide variety of underwater scenes, including marine life and divers, where image quality suffers from severe degradation.

To evaluate generalization and color correction performance, the method is further tested on four benchmark datasets: UCCS [33], EUVP [34], UFO-120 [35], and U45 [36]. UCCS includes 300 images equally distributed across three color variations: blue, green, and blue-green. EUVP provides a large collection of underwater images from diverse conditions; following prior works [6,37], 130 validation images of underwater scenes are selected from EUVP dataset for testing. UFO-120 comprises 120 images captured under varied underwater scenarios, representing different degradation types. U45 contains 45 degraded images exhibiting color bias, low contrast, and hazy appearance, with no corresponding ground truths.

4.2. Implementation Details

The AdamW [38] optimizer is exploited to minimize the loss function, with parameters set to $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The learning rate is adjusted using a 20-epoch warm-up and subsequently decayed using a cosine annealing scheduler [39]. Batch size is set to 2, and the network has been trained for 1000 epochs. All training samples are uniformly resized to 256×256 pixels and augmented through random flipping and rotation. Experiments on FACMamba are conducted using the PyTorch (<https://pytorch.org/>) framework with an NVIDIA GeForce RTX 4090 GPU.

To remain consistent with mainstream algorithms, we evaluate the proposed network using MSE, PSNR [40], SSIM [41], UIQM [42], UCIQE, and NIQE [43]. The source code for our FACMamba can be fetched at: <https://github.com/wwaanngllii/FACMamba> (accessed on 17 October 2025).

4.3. Model Analysis

To verify the contributions of SS8D in MVSM and FAGM, we conduct ablation studies by substituting them with widely used counterparts. The quantitative results are reported in Table 1, and the corresponding visual and statistical comparisons are illustrated in Figure 4.

Table 1. Analysis of four networks on T90 dataset. \uparrow denotes higher is better.

Model	PSNR \uparrow	SSIM \uparrow	UIQM \uparrow	UCIQE \uparrow	Params	FLOPs
FACMamba-SSM	23.45	0.918	2.977	0.550	1.92 M	25.91 G
FACMamba-FRB	23.51	0.915	3.004	0.569	8.78 M	31.48 G
FACMamba-FSM	23.49	0.916	3.001	0.566	3.62 M	25.90 G
FACMamba	23.53	0.920	3.011	0.572	3.03 M	25.91 G

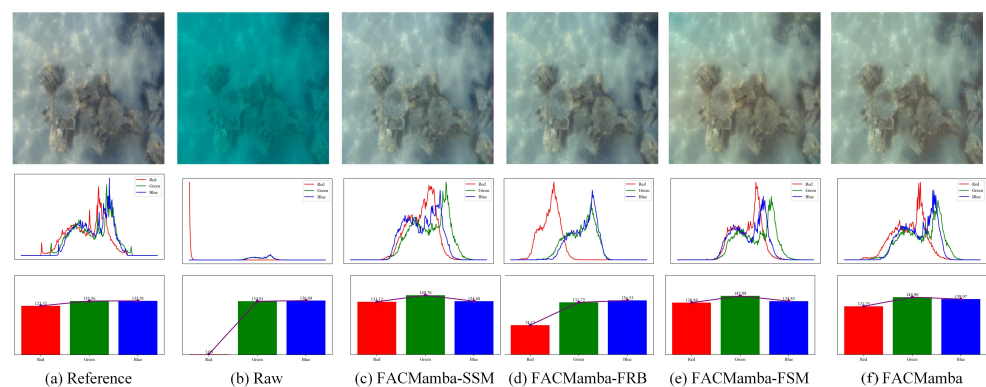


Figure 4. Visual and statistical analysis of different variants. First Row: Enhanced Images; Second Row: RGB Pixel Brightness Curves; Third Row: RGB Average Brightness Histograms. In comparison to different variants, FACMamba restores more natural colors and balanced distributions.

Analysis of MVSM. To evaluate the effectiveness of SS8D, we replace it with the standard SS2D, yielding the variant FACMamba-SSM. Compared to FACMamba-SSM, the full FACMamba achieves a 0.08 dB gain in PSNR (23.53 dB vs. 23.45 dB), a 0.002 increase in SSIM (0.920 vs. 0.918), and higher perceptual scores in UIQM (3.011 vs. 2.977) and UCIQE (0.572 vs. 0.550). Although FACMamba contains 1.11M more parameters than FACMamba-SSM, this slight increase represents a reasonable trade-off for the observed accuracy gains. More importantly, the visual improvements are considerably more noticeable than the numerical differences suggest. As illustrated in Figure 4, it can be found that the image produced by FACMamba-SSM remains visually dull and lacks contrast, with an RGB histogram exhibiting low red-channel intensity and compressed color variation. In contrast, FACMamba generates a more vivid image with stronger edge definition, a more balanced RGB curve, and average brightness closer to the reference. This confirms that SS8D captures richer global context and long-range dependencies, thereby offering advantages in color correction and visual clarity.

Analysis of FAGM. To assess the effectiveness of FAGM, we replace it with other frequency-domain blocks, including the frequency residual block (FRB [32]) and frequency selection module (FSM [23]), as depicted in Figure 5, yielding the variants FACMamba-FRB and FACMamba-FSM. Table 1 shows that while these variants achieve comparable PSNR and SSIM, their perceptual quality scores (UIQM and UCIQE) are consistently lower than those of FACMamba. In particular, FACMamba records the best UIQM (3.011) and UCIQE (0.572), demonstrating superior color enhancement and visual fidelity. Figure 6 displays a comparison of average feature maps produced by different frequency-domain blocks after the final upsampling stage. The visual results corroborate the quantitative findings: FRB and FSM variants capture basic structural cues but exhibit blurred textures and weaker edge responses, whereas FACMamba with FAGM produces sharper boundaries, richer textures, and stronger activations in target regions. Similarly, Figure 4 reveals that FACMamba with FAGM restores a more natural color distribution in the RGB histogram, achieving a visual effect closer to the reference image.

In summary, both SS8D and FAGM make significant contributions to FACMamba. They not only improve image enhancement results but also achieve a superior equilibrium between quality and efficiency compared with traditional alternatives.

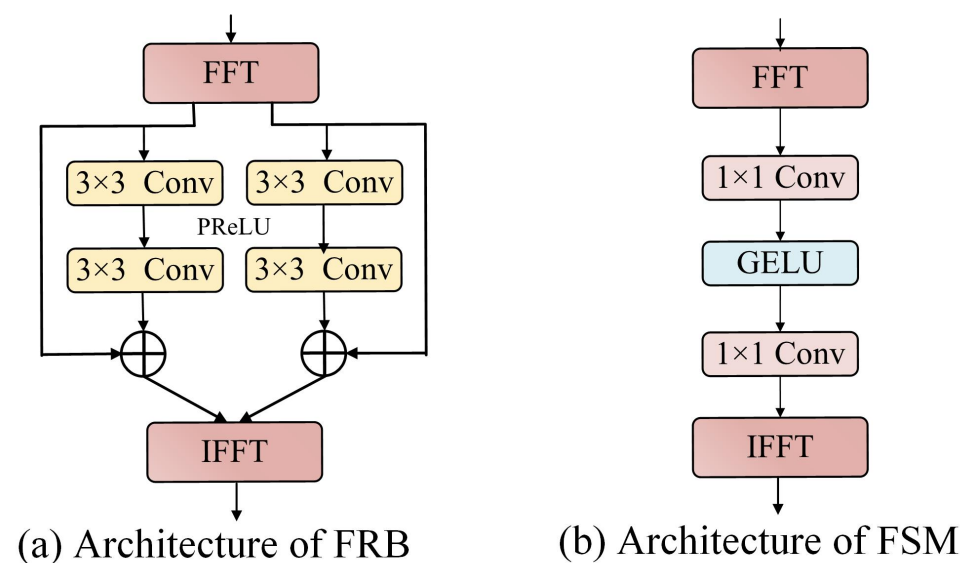


Figure 5. Architectures of two frequency-domain blocks. (a) The FRB is proposed by the literature [32] (b) The FSM is proposed by the literature [23].

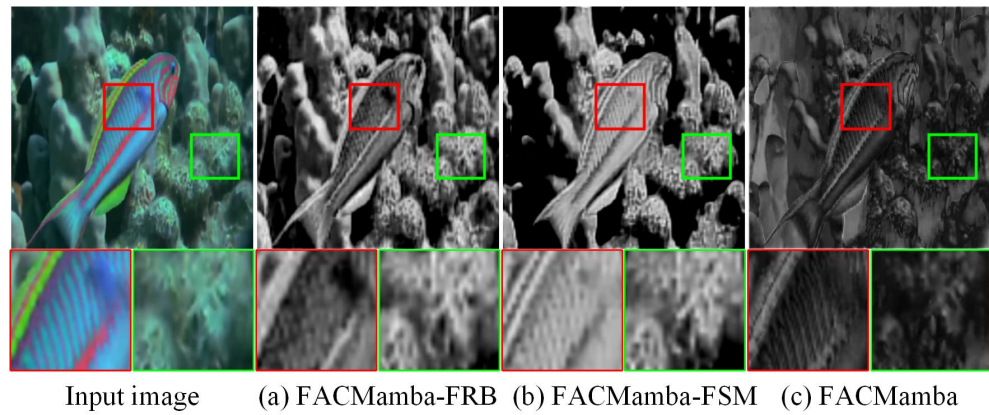


Figure 6. Visualization of average feature maps with grayscale colormap. One can see that FACMamba preserves sharper edges and richer textures than FRB and FSM variants.

4.4. Ablation Study

Table 2 shows the results of the ablation experiments on the T90 dataset. These experiments sequentially removed the three core modules of the FACMamba network architecture (MVSM, FAGM, and SAFM) to validate each module’s contribution to image enhancement performance. Additionally, we give a visual heat map of the second BAIM after removing the relevant components in Figure 7.

Table 2. Ablation study for investigating the components of FACMamba on T90 dataset. \uparrow denotes higher is better.

Model	PSNR \uparrow	SSIM \uparrow	UIQM \uparrow	UCIQE \uparrow	Params	FLOPs
FACMamba w/o MVSM	22.94	0.899	2.942	0.542	0.56 M	24.24 G
FACMamba w/o FAGM	23.44	0.915	2.988	0.567	3.01 M	25.90 G
FACMamba w/o SAFM	23.32	0.903	2.983	0.554	2.63 M	4.44 G
FACMamba	23.53	0.920	3.011	0.572	3.03 M	25.91 G

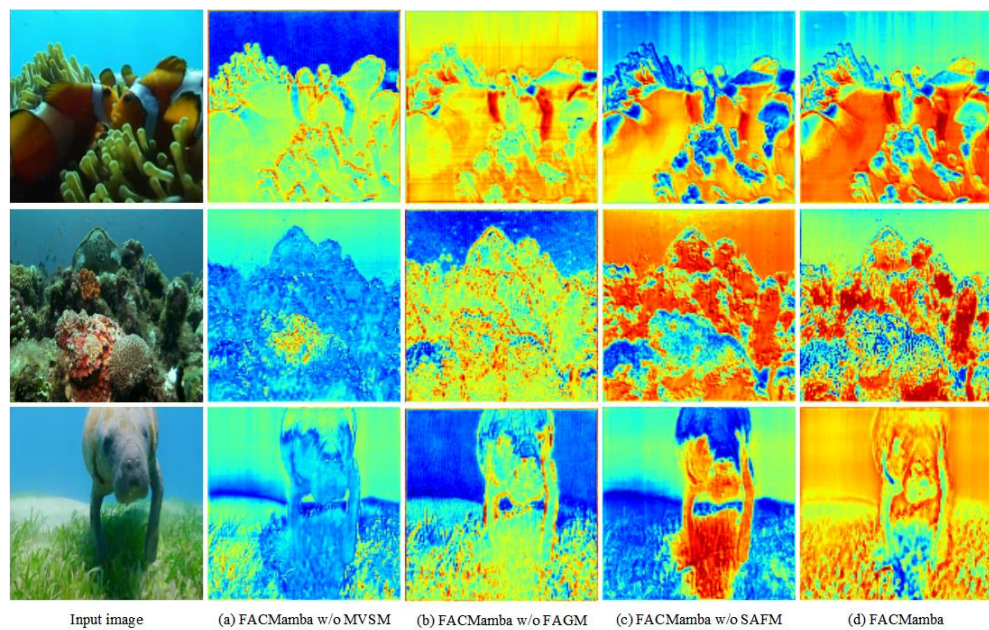


Figure 7. Average feature maps visualization from the second BAIM. It can be found that FACMamba produces clearer feature maps with sharper edges, richer textures, and more active responses in target regions than its ablated variants.

Ablation of MVSM. As shown in Table 2, removing the MVSM module, specifically its SS8D, leads to the most significant performance degradation, with PSNR dropping from 23.53 dB to 22.94 dB and SSIM from 0.920 to 0.899. UIQM and UCIQE also decrease by 0.069 and 0.030 respectively, underscoring the pivotal role of SS8D in capturing multi-directional spatial dependencies. Although this removal reduces the parameter count to 0.56 M and FLOPs to 24.24 G, the substantial loss in perceptual and structural quality confirms the necessity of MVSM.

Ablation of FAGM. We can see from Table 2 that compared to FACMamba w/o FAGM, FACMamba supported by FAGM achieves significant performance gains with only a marginal increase in parameters of approximately 0.02 M. Figure 7 further manifests that the introduction of FAGM yields a more positive response to the details within the target region. This is attributed to the ability of FAGM to selectively enhance frequency components of information through its attention mechanism, thereby effectively improving image quality in damaged areas.

Ablation of SAFM. Similarly, eliminating SAFM causes a noticeable drop in SSIM from 0.920 to 0.903, underscoring its importance in structural consistency preservation. Despite having fewer parameters (2.63 M) and the lowest computational cost (4.44 G FLOPs), its removal leads to considerable performance degradation, indicating its key role in multi-scale feature fusion. The reduction in complexity after removing SAFM can be attributed to its convolution-intensive design, where multiple convolutional operations are employed to capture spatial dependencies across scales. While these convolutions inevitably increase parameters and FLOPs, they are essential for enhancing structural representation.

As depicted in Figure 7, the complete FACMamba model (rightmost column) exhibits more uniform, high-contrast feature activations with clear structure boundaries across all three examples, indicating effective representation of both global and local image features. In contrast, removing key modules such as MVSM, FAGM, or SAFM leads to visibly degraded feature responses.

Overall, the ablation study validates that the integration of MVSM, FAGM, and SAFM significantly improves both accuracy and computational efficiency in FACMamba.

4.5. Comparison with SOTA Methods

Quantitative Comparisons: We conduct a quantitative evaluation by comparing our proposed method with a range of advanced UIE approaches. Table 3 presents a comprehensive quantitative comparison of competing image enhancement methods on the C60 and UCCS datasets, evaluated using UIQM, UCIQE, parameter count, and FLOPs. One can see that FACMamba realizes competitive performance on C60 and leading performance on UCCS. This is attributable to our approach utilizing MVSM, a technique that aggregates features from eight directions, significantly enhancing global context modeling and color correction capabilities. Consequently, it proves particularly advantageous for the UCCS dataset with severe color deviations. Additionally, FACMamba displays a lower model capacity (3.03 M/25.91 G), significantly outperforming models like X-CAUNet (31.78 M/261.48 G) and Restormer (26.10 M/140.99 G), demonstrating a favorable balance between performance and efficiency. Overall trends show that Transformer-based and Mamba-based architectures outperform traditional CNN methods such as Ucolor and PUIE-Net in both image quality and computational cost.

Table 3. Quantitative comparisons are presented with respect to the C60 and UCCS datasets, model parameters, and FLOPs. Bold indicates the best performance. ↑ denotes higher is better. ↓ denotes lower is better.

Method	C60		UCCS		Params ↓	FLOPs ↓
	UIQM ↑	UCIQE ↑	UIQM ↑	UCIQE ↑		
Ucolor [1]	2.482	0.553	3.019	0.550	157.4 M	34.68 G
PUIE-Net [44]	2.521	0.558	3.003	0.536	1.41 M	30.09 G
URSCT [45]	2.642	0.543	2.947	0.544	11.41 M	18.11 G
Restormer [46]	2.688	0.572	2.981	0.542	26.10 M	140.99 G
PUGAN [20]	2.652	0.566	2.977	0.536	95.66 M	72.05 G
MFEF [47]	2.652	0.566	2.977	0.556	61.86 M	26.52 G
Semi-UIR [48]	2.667	0.574	3.079	0.554	1.65 M	36.44 G
UIE-Convformer [49]	2.684	0.572	2.946	0.555	25.9 M	36.9 G
X-CAUNET [50]	2.683	0.564	2.922	0.541	31.78 M	261.48 G
WaterMamba [12]	2.853	0.582	3.057	0.550	3.69 M	7.53 G
PixMamba [50]	2.868	0.586	3.053	0.561	8.68 M	7.60 G
FACMamba (Ours)	2.690	0.570	3.074	0.597	3.03 M	25.91 G

A comparative analysis of UIE approaches on the T90 dataset is provided in Table 4, based on five evaluation criteria: PSNR, SSIM, MSE, UIQM, and UCIQE. The table indicates that our proposed FACMamba achieves the best overall performance, attaining the highest SSIM (0.938), lowest MSE (0.060), and leading UIQM (3.074), while also maintaining a competitive PSNR (23.727 dB) and UCIQE (0.602). In contrast, traditional methods like Shallow-uwnet and Ucolor show consistently lower scores across all metrics, suggesting weaker capability in preserving image quality and color information. WaterMamba outperforms FACMamba by 0.998 dB in PSNR and PixMamba improves by 0.015 in UCIQE, but these gains come at the cost of a increase increased parameter budgets (3.69 M and 8.68 M vs. 3.03 M). This highlights that FACMamba offers a more favorable balance between enhancement quality and computational efficiency.

Table 4. Quantitative comparisons on T90 dataset. Bold indicates the best performance. ↑ denotes higher is better. ↓ denotes lower is better.

Method	T90				
	PSNR ↑	SSIM ↑	MSE ↓	UIQM ↑	UCIQE ↑
Ucolor [1]	21.093	0.872	0.096	3.049	0.555
Shallow-uwnet [51]	18.278	0.855	0.131	2.942	0.544
UIEC ² -Net [52]	22.958	0.907	0.078	2.999	0.599
PUIE-Net [44]	21.382	0.882	0.093	3.021	0.566
NU2Net [53]	23.061	0.923	0.086	2.936	0.587
WaterMamba [12]	24.715	0.931	-	-	-
PixMamba [50]	23.587	0.921	0.061	3.048	0.617
FACMamba (Ours)	23.727	0.938	0.060	3.074	0.602

Table 5 reports the quantitative results on the EUVP, UFO-120, and U45 datasets using UIQM, UCIQE, and NIQE. Across all datasets, our method consistently yields the highest UIQM scores, demonstrating its advantage in perceptual quality. In terms of UCIQE and NIQE, FACMamba also demonstrates competitive performance, reflecting its strong capability in color contrast enhancement and naturalness preservation under diverse underwater conditions.

Qualitative Comparison: As depicted in Figure 8, traditional methods like Ucolor and PUGAN show partial improvements but often suffer from over-saturation or unnatural

tones. MFEF and X-CAUNet restore more details but leave some haze or color imbalance. WaterMamba and PixMamba generate clearer, more natural-looking results, but subtle color inconsistencies remain. Our proposed FACMamba, however, delivers the most visually appealing outputs, with enhanced sharpness, well-balanced colors, and preserved fine textures in both datasets.

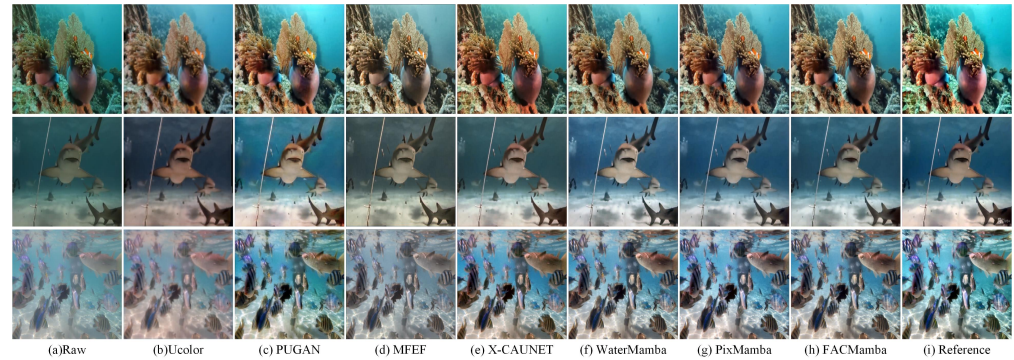


Figure 8. Qualitative comparison of advanced approaches on T90 dataset.

Table 5. Quantitative comparisons on EUVP, UFO-120, and U45 datasets. Bold indicates the best performance. \uparrow denotes higher is better. \downarrow denotes lower is better.

Method	EUVP			UFO-120			U45		
	UIQM \uparrow	UCIQE \uparrow	NIQE \downarrow	UIQM \uparrow	UCIQE \uparrow	NIQE \downarrow	UIQM \uparrow	UCIQE \uparrow	NIQE \downarrow
FUnIE-GAN [34]	2.763	0.588	4.323	2.755	0.601	4.376	4.981	0.534	4.981
Water-Net [15]	3.079	0.587	4.639	3.023	0.592	4.423	3.125	0.567	6.048
PUIE-Net [44]	3.039	0.588	4.523	3.015	0.601	4.367	3.189	0.566	4.107
MLLE [54]	2.229	0.611	4.636	2.309	0.621	4.628	2.484	0.597	6.352
UNTV [55]	2.497	0.619	5.307	2.523	0.621	5.174	1.766	0.627	8.939
U-shape [2]	2.947	0.574	4.478	2.898	0.575	4.387	3.104	0.553	4.569
ASNet [56]	3.064	0.616	4.579	3.015	0.618	4.473	3.217	0.603	6.165
LiteEnhanceNet [57]	2.957	0.620	4.593	2.954	0.622	4.436	3.171	0.580	4.064
MESA-WNet [37]	3.057	0.622	4.228	3.032	0.623	4.027	3.238	0.604	4.048
FACMamba (Ours)	3.090	0.607	4.387	3.057	0.612	4.365	3.260	0.593	4.004

Similarly, FACMamba stands out by producing results most visually aligned with the reference in Figure 9. It achieves a natural color balance, sharper details, and enhanced contrast across all scenes. Compared to its Mamba-family counterparts, FACMamba consistently renders more accurate marine hues (e.g., fish scales, coral textures) and clearer object boundaries, suggesting superior handling of both global and local features.

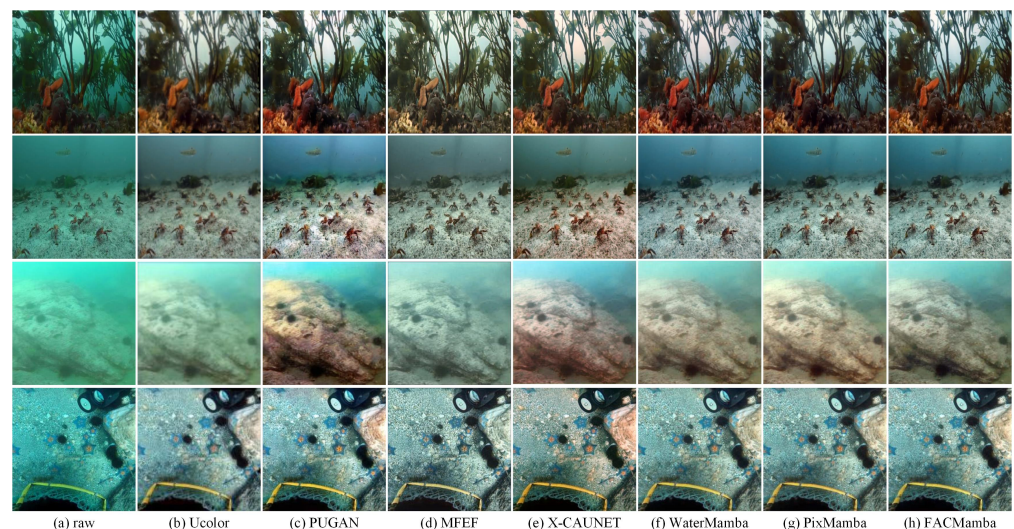


Figure 9. Qualitative comparison of advanced approaches on C60 and UCCS datasets. The first and second rows are for C60 and the third and fourth rows are for UCCS dataset.

In line with these observations, Figure 10 further demonstrates the superiority of FAC-Mamba over CNN-based approaches such as Water-Net, PUIE-Net, and LiteEnhanceNet, which often suffer from residual haze, color distortion, or loss of detail. While MESA-UNet improves tonal consistency, its outputs may appear overly smoothed. By contrast, FACMamba consistently produces visually natural underwater scenes with sharper edges, balanced colors, and well-preserved textures.

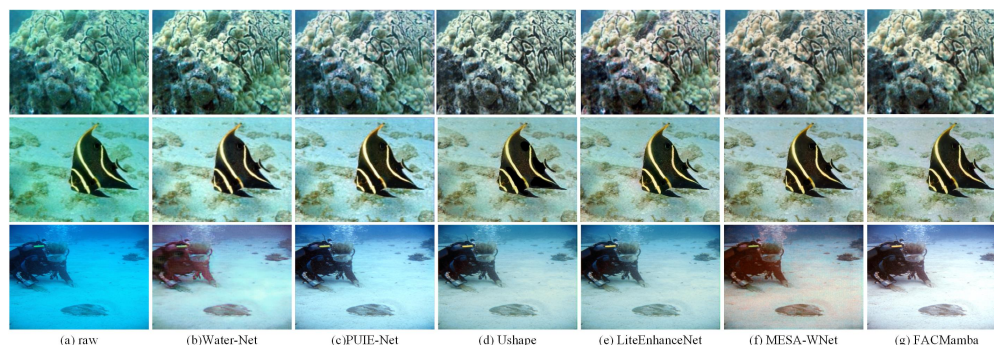


Figure 10. Qualitative comparison of advanced approaches on EUVP, UFO-120, and U45 datasets. The first row is for EUVP dataset, the second row is for UFO-120 dataset, and the third row is for U45 dataset.

4.6. Model Efficiency

Table 6 summarizes the computational efficiency of advanced UIE methods on the U45 dataset (256×256 input), reporting parameter size, FLOPs, and inference time. Among CNN-based methods, LiteEnhanceNet is the most lightweight (0.01 M parameters, 0.64 G FLOPs) and fastest (0.7608 s), while PUIE-Net, though relatively low in complexity (1.41 M, 30.09 G), exhibits the slowest inference (4.3551 s). U-shape and MESA-UNet are parameter-intensive, requiring 31.59 M and 26.87 M, respectively. For Mamba-based methods, FACMamba offers a favorable trade-off, with only 3.03 M parameters and 25.91 G FLOPs, while maintaining competitive inference time (2.2674 s).

Table 6. Computational efficiency of advanced UIE approaches on U45 dataset.

Index	CNN-Based				Mamba-Based			
	Water-Net	PUIE-Net	U-Shape	LiteEnhanceNet	MESA-WNet	WaterMamba	PixMamba	FACMamba (Ours)
Params (M)	24.81	1.41	31.59	0.01	26.87	3.69	8.68	3.03
FLOPs (G)	193.7	30.09	26.10	0.64	26.38	7.53	7.60	25.91
Inference Time (s)	0.82245	4.3551	1.5238	0.7608	2.9241	2.2382	2.7109	2.2674

4.7. Applications

To further validate the scalability of the proposed method, Figure 11 illustrates the comparative results of Harris corner detection [58] and Canny edge detection [59] applied to images enhanced by different approaches. The first row displays the enhanced images, the second row overlays detected Harris corners (in red), and the third row shows corresponding edge maps.

As illustrated in Figure 11g, the proposed FACMamba method yields images with the most clearly defined and accurately positioned corners and edges, closely resembling the reference shown in Figure 11a. In contrast, the raw input image (Figure 11b) exhibits low contrast and weak structural information. FACMamba effectively enhances structural clarity, rendering edges and corners more pronounced. While alternative methods such as MFEF (Figure 11c), PUGAN (Figure 11d), and WaterMamba (Figure 11e) offer partial improvements, they often result in blurred or inconsistent edge representations. Although PixMamba (Figure 11f) performs comparatively well, FACMamba still surpasses it in both corner localization density and edge definition.

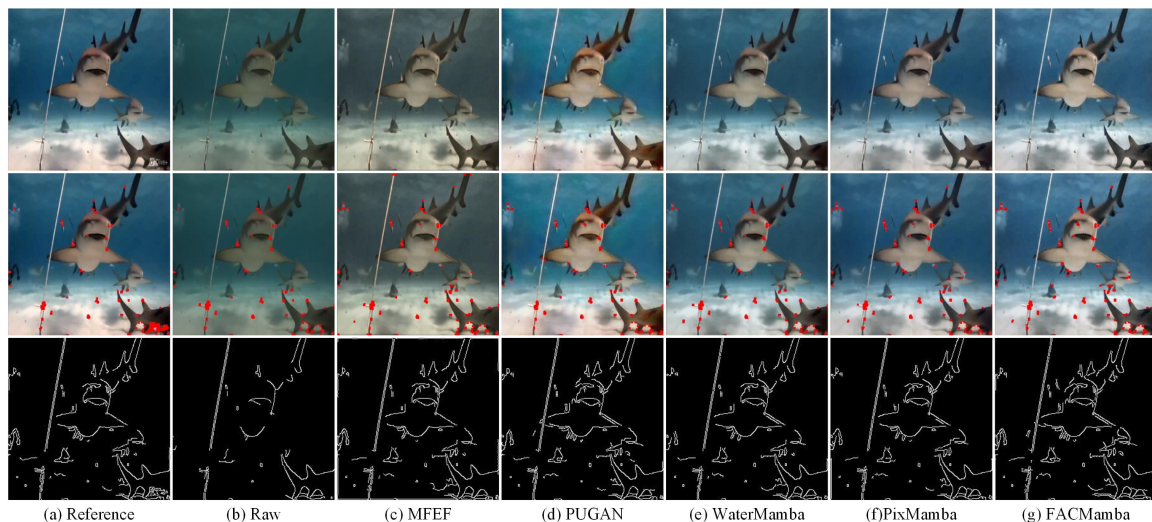


Figure 11. Comparative analysis of corner and edge detection results across enhancement methods. FACMamba delivers the most accurate and clearly defined corners and edges.

These findings confirm that, beyond visual enhancement, FACMamba effectively supports tasks like feature correspondence and object recognition by improving local textures, geometric structures, and salient edge features.

5. Conclusions

We propose FACMamba, an underwater image enhancement framework that fuses vision state-space dynamics with frequency-aware and structure-preserving guidance to effectively handle visual distortions, reduced contrast, and color shifts inherent in underwater scenes. FACMamba effectively captures long-range spatial dependencies through the Multi-Directional Vision State-Space Module (MVSM), which leverages a directional-aware SS8D mechanism to enhance contextual feature modeling. To further improve perceptual quality, we introduce the Frequency-Aware Guidance Module (FAGM), enabling adaptive frequency-domain enhancement with low computational overhead. Additionally, the Structure-Aware Fusion Module (SAFM) is employed to reinforce multi-scale structural consistency via lightweight fusion strategies. Extensive experiments on benchmark datasets have shown that FACMamba is comparable to state-of-the-art methods in terms of both objective and perceptual metrics, while maintaining a favorable balance between accuracy and efficiency. Although FACMamba exhibits favorable performance, certain limitations remain. FACMamba relies on a degenerate prior specific to underwater environments and may demonstrate reduced robustness under extremely turbid or highly dynamic illumination conditions. Future work will focus on developing adaptive frequency modeling to extend the proposed method to a broader range of low-visibility imaging applications.

Author Contributions: Conceptualization, L.W. and K.S.; Methodology, L.W., K.S. and X.C.; Software, L.W.; Validation, K.S., H.S., J.Z. and B.W.; Formal analysis, H.S. and B.W.; Investigation, J.Z. and B.W.; Resources, X.C.; Data curation, H.S. and X.C.; Writing—original draft, L.W.; Writing—review & editing, K.S.; Funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Start-up Fund for New Talented Researchers of Nanjing Vocational University of Industry Technology (Grant No. YK24-05-02), National Natural Science Foundation of China (Grant No. 62563030 and No. 42507422), Natural Science Foundation of Jiangsu Province (Grant No. BK20241070), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant No. 25KJD520006).

Data Availability Statement: Data available in a publicly accessible repository The original data presented in this study are openly available in multiple public underwater image datasets, including the UIEB dataset [15] the UCCS dataset [33], the EUVP dataset [34], the UFO 120 dataset [35], and the U45 dataset [36]. All datasets are publicly accessible through their respective repositories or DOI links provided in the cited references.

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

References

1. 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. [[CrossRef](#)]
2. Peng, L.; Zhu, C.; Bian, L. U-Shape Transformer for Underwater Image Enhancement. *IEEE Trans. Image Process.* **2023**, *32*, 3066–3079. [[CrossRef](#)]
3. Han, G.; Wang, M.; Zhu, H.; Lin, C. UIEGAN: Adversarial Learning-Based Photorealistic Image Enhancement for Intelligent Underwater Environment Perception. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5611514. [[CrossRef](#)]
4. Tao, Y.; Tang, J.; Zhao, X.; Zhou, C.; Wang, C.; Zhao, Z. Multi-scale network with attention mechanism for underwater image enhancement. *Neurocomputing* **2024**, *595*, 127926. [[CrossRef](#)]
5. Chen, W.; Lei, Y.; Luo, S.; Zhou, Z.; Li, M.; Pun, C.M. Uwformer: Underwater image enhancement via a semi-supervised multi-scale transformer. In Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN), Yokohama, Japan, 30 June–5 July 2024; pp. 1–8.
6. Fan, J.; Xu, J.; Zhou, J.; Meng, D.; Lin, Y. See Through Water: Heuristic Modeling Toward Color Correction for Underwater Image Enhancement. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 4039–4054. [[CrossRef](#)]
7. Zhang, W.; Li, X.; Huang, Y.; Xu, S.; Tang, J.; Hu, H. Underwater image enhancement via frequency and spatial domains fusion. *Opt. Lasers Eng.* **2025**, *186*, 108826. [[CrossRef](#)]
8. Liu, X.; Gu, Z.; Ding, H.; Zhang, M.; Wang, L. Underwater image super-resolution using frequency-domain enhanced attention network. *IEEE Access* **2024**, *12*, 6136–6147. [[CrossRef](#)]
9. Pramanick, A.; Megha, D.; Sur, A. Attention-Based Spatial-Frequency Information Network for Underwater Single Image Super-Resolution. In Proceedings of the ICASSP 2024—2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Republic of Korea, 14–19 April 2024; pp. 3560–3564. [[CrossRef](#)]
10. Gu, A.; Dao, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. In Proceedings of the First Conference on Language Modeling, Pennsylvania, PA, USA, 7–10 October 2024.
11. Guo, H.; Li, J.; Dai, T.; Ouyang, Z.; Ren, X.; Xia, S.T. MambaIR: A Simple Baseline for Image Restoration with State-Space Model. *arXiv* **2024**, arXiv:2402.15648.
12. Guan, M.; Xu, H.; Jiang, G.; Yu, M.; Chen, Y.; Luo, T.; Song, Y. WaterMamba: Visual State Space Model for Underwater Image Enhancement. *arXiv* **2024**, arXiv:2405.08419. [[CrossRef](#)]
13. Lin, W.T.; Lin, Y.X.; Chen, J.W.; Hua, K.L. PixMamba: Leveraging State Space Models in a Dual-Level Architecture for Underwater Image Enhancement. *arXiv* **2024**, arXiv:2406.08444.
14. Dong, C.; Zhao, C.; Cai, W.; Yang, B. O-Mamba: O-shape State-Space Model for Underwater Image Enhancement. *arXiv* **2024**, arXiv:2408.12816.
15. Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; Tao, D. An Underwater Image Enhancement Benchmark Dataset and Beyond. *IEEE Trans. Image Process.* **2019**, *29*, 4376–4389. [[CrossRef](#)]
16. Guan, M.; Xu, H.; Jiang, G.; Yu, M.; Chen, Y.; Luo, T.; Zhang, X. DiffWater: Underwater Image Enhancement Based on Conditional Denoising Diffusion Probabilistic Model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 2319–2335. [[CrossRef](#)]
17. Li, C.; Guo, J.; Chen, S.; Tang, Y.; Pang, Y.; Wang, J. Underwater image restoration based on minimum information loss principle and optical properties of underwater imaging. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 1993–1997. [[CrossRef](#)]
18. Drews, P.L.; Nascimento, E.R.; Botelho, S.S.; Montenegro Campos, M.F. Underwater Depth Estimation and Image Restoration Based on Single Images. *IEEE Comput. Graph. Appl.* **2016**, *36*, 24–35. [[CrossRef](#)]
19. Wang, Y.; Liu, H.; Chau, L.P. Single Underwater Image Restoration Using Adaptive Attenuation-Curve Prior. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2018**, *65*, 992–1002. [[CrossRef](#)]
20. Cong, R.; Yang, W.; Zhang, W.; Li, C.; Guo, C.L.; Huang, Q.; Kwong, S. PUGAN: Physical Model-Guided Underwater Image Enhancement Using GAN with Dual-Discriminators. *IEEE Trans. Image Process.* **2023**, *32*, 4472–4485. [[CrossRef](#)] [[PubMed](#)]
21. 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*, 6816–6830. [[CrossRef](#)] [[PubMed](#)]

22. An, G.; He, A.; Wang, Y.; Guo, J. UWMamba: UnderWater Image Enhancement with State Space Model. *IEEE Signal Process. Lett.* **2024**, *31*, 2725–2729. [[CrossRef](#)]
23. Xiao, Y.; Yuan, Q.; Jiang, K.; Chen, Y.; Zhang, Q.; Lin, C.W. Frequency-Assisted Mamba for Remote Sensing Image Super-Resolution. *arXiv* **2024**, arXiv:2405.04964. [[CrossRef](#)]
24. Zhang, X.; Su, Q.; Yuan, Z.; Liu, D. An efficient blind color image watermarking algorithm in spatial domain combining discrete Fourier transform. *Optik* **2020**, *219*, 165272. [[CrossRef](#)]
25. Yang, Y.; Soatto, S. FDA: Fourier Domain Adaptation for Semantic Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 4085–4095.
26. Xu, Q.; Zhang, R.; Zhang, Y.; Wang, Y.; Tian, Q. A Fourier-Based Framework for Domain Generalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 14383–14392.
27. Cheng, Z.; Fan, G.; Zhou, J.; Gan, M.; Chen, C.L.P. FDCE-Net: Underwater Image Enhancement with Embedding Frequency and Dual Color Encoder. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 1728–1744. [[CrossRef](#)]
28. Walia, J.S.; Venkatraman, S.; LK, P. FUSION: Frequency-guided Underwater Spatial Image reconstruction. In Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville TN, USA, 11–15 June 2025.
29. Zhao, C.; Cai, W.; Dong, C.; Hu, C. Wavelet-based Fourier Information Interaction with Frequency Diffusion Adjustment for Underwater Image Restoration. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 8281–8291. [[CrossRef](#)]
30. Zhu, Z.; Li, X.; Ma, Q.; Zhai, J.; Hu, H. FNet: Fourier transform guided dual-channel underwater image enhancement diffusion network. *Sci. China Technol. Sci.* **2025**, *68*, 1100403. [[CrossRef](#)]
31. Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; Liu, Y. VMamba: Visual State Space Model. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 103031–103063.
32. Wang, C.; Jiang, J.; Zhong, Z.; Liu, X. Spatial-Frequency Mutual Learning for Face Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 22356–22366.
33. 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*, 4861–4875. [[CrossRef](#)]
34. Islam, M.J.; Xia, Y.; Sattar, J. Fast Underwater Image Enhancement for Improved Visual Perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [[CrossRef](#)]
35. Islam, M.J.; Luo, P.; Sattar, J. Simultaneous Enhancement and Super-Resolution of Underwater Imagery for Improved Visual Perception. *arXiv* **2020**, arXiv:2002.01155. [[CrossRef](#)]
36. Li, H.; Li, J.; Wang, W. A Fusion Adversarial Underwater Image Enhancement Network with a Public Test Dataset. *arXiv* **2019**, arXiv:1906.06819. [[CrossRef](#)]
37. Chang, B.; Yuan, G.; Li, J. Mamba-enhanced spectral-attentive wavelet network for underwater image restoration. *Eng. Appl. Artif. Intell.* **2025**, *143*, 109999. [[CrossRef](#)]
38. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. *arXiv* **2019**, arXiv:1711.05101. [[CrossRef](#)]
39. Loshchilov, I.; Hutter, F. SGDR: Stochastic Gradient Descent with Warm Restarts. *arXiv* **2017**, arXiv:1608.03983. [[CrossRef](#)]
40. Korhonen, J.; You, J. Peak signal-to-noise ratio revisited: Is simple beautiful? In Proceedings of the 2012 Fourth International Workshop on Quality of Multimedia Experience, Melbourne, Australia, 5–7 July 2012; pp. 37–38. [[CrossRef](#)]
41. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
42. Panetta, K.; Gao, C.; Agaian, S. Human-Visual-System-Inspired Underwater Image Quality Measures. *IEEE J. Ocean. Eng.* **2016**, *41*, 541–551. [[CrossRef](#)]
43. Liu, L.; Liu, B.; Huang, H.; Bovik, A.C. No-reference image quality assessment based on spatial and spectral entropies. *Signal Proc. Image Commun.* **2014**, *29*, 856–863. [[CrossRef](#)]
44. Fu, Z.; Wang, W.; Huang, Y.; Ding, X.; Ma, K.K. Uncertainty Inspired Underwater Image Enhancement. In *Computer Vision—ECCV 2022*; Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T., Eds.; Springer: Cham, Switzerland, 2022; pp. 465–482.
45. Ren, T.; Xu, H.; Jiang, G.; Yu, M.; Zhang, X.; Wang, B.; Luo, T. Reinforced Swin-ConvS Transformer for Simultaneous Underwater Sensing Scene Image Enhancement and Super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4209616. [[CrossRef](#)]
46. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H. Restormer: Efficient Transformer for High-Resolution Image Restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 5728–5739.
47. Zhou, J.; Sun, J.; Zhang, W.; Lin, Z. Multi-view underwater image enhancement method via embedded fusion mechanism. *Eng. Appl. Artif. Intell.* **2023**, *121*, 105946. [[CrossRef](#)]

48. Huang, S.; Wang, K.; Liu, H.; Chen, J.; Li, Y. Contrastive Semi-Supervised Learning for Underwater Image Restoration via Reliable Bank. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 18145–18155.
49. Wang, B.; Xu, H.; Jiang, G.; Yu, M.; Ren, T.; Luo, T.; Zhu, Z. UIE-Convformer: Underwater Image Enhancement Based on Convolution and Feature Fusion Transformer. *IEEE Trans. Emerg. Top. Comput. Intell.* **2024**, *8*, 1952–1968. [[CrossRef](#)]
50. Pramanick, A.; Sarma, S.; Sur, A. X-CAUNET: Cross-Color Channel Attention with Underwater Image-Enhancing Transformer. In Proceedings of the ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Republic of Korea, 14–19 April 2024; pp. 3550–3554. [[CrossRef](#)]
51. Naik, A.; Swarnakar, A.; Mittal, K. Shallow-UWnet: Compressed Model for Underwater Image Enhancement (Student Abstract). *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 15853–15854. [[CrossRef](#)]
52. Wang, Y.; Guo, J.; Gao, H.; Yue, H. UIEC²-Net: CNN-based underwater image enhancement using two color space. *Signal Proc. Image Commun.* **2021**, *96*, 116250. [[CrossRef](#)]
53. Guo, C.; Wu, R.; Jin, X.; Han, L.; Zhang, W.; Chai, Z.; Li, C. Underwater Ranker: Learn Which Is Better and How to Be Better. *Proc. AAAI Conf. Artif. Intell.* **2023**, *37*, 702–709. [[CrossRef](#)]
54. Zhang, W.; Zhuang, P.; Sun, H.H.; Li, G.; Kwong, S.; Li, C. Underwater Image Enhancement via Minimal Color Loss and Locally Adaptive Contrast Enhancement. *IEEE Trans. Image Process.* **2022**, *31*, 3997–4010. [[CrossRef](#)]
55. Xie, J.; Hou, G.; Wang, G.; Pan, Z. A Variational Framework for Underwater Image Dehazing and Deblurring. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 3514–3526. [[CrossRef](#)]
56. Park, C.W.; Eom, I.K. Underwater image enhancement using adaptive standardization and normalization networks. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107445. [[CrossRef](#)]
57. Zhang, S.; Zhao, S.; An, D.; Li, D.; Zhao, R. LiteEnhanceNet: A lightweight network for real-time single underwater image enhancement. *Expert Syst. Appl.* **2024**, *240*, 122546. [[CrossRef](#)]
58. Harris, C.; Stephens, M. A Combined Corner and Edge Detector. In Proceedings of the Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988; pp. 23.1–23.6. [[CrossRef](#)]
59. Canny, J. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *PAMI-8*, 679–698. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.