FUSION: Frequency-guided Underwater Spatial Image recOnstructioN

Jaskaran Singh Walia* Shravan Venkatraman* Pavithra L K
Vellore Institute of Technology, Chennai, India
corresponding email: pavithra.lk@vit.ac.in
* Equal contribution

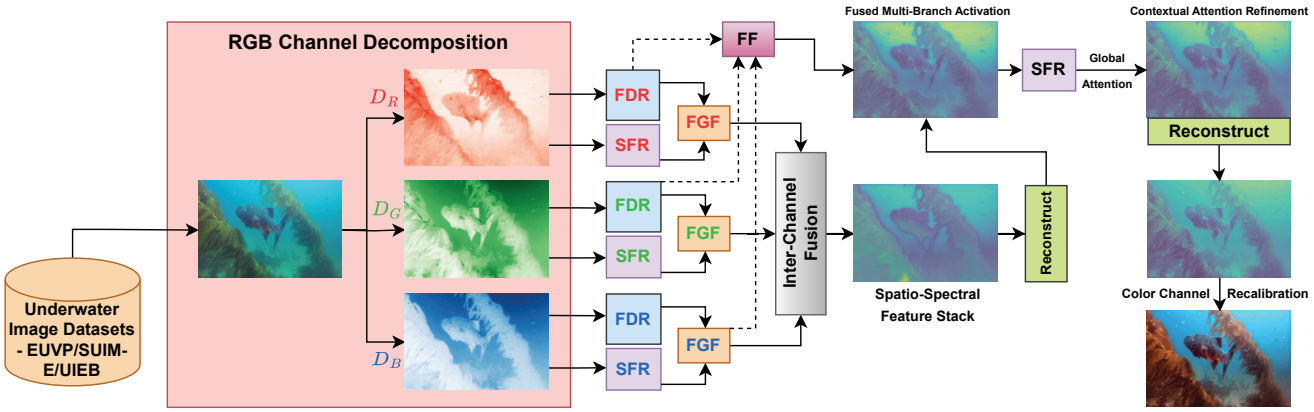

Figure 1. An overview of the proposed FUSION pipeline, illustrating the dual-domain (spatial and frequency) processing, contextual attention refinement, and final channel calibration for UIE.

Abstract

Underwater images suffer from severe degradations, including color distortions, reduced visibility, and loss of structural details due to wavelength-dependent attenuation and scattering. Existing enhancement methods primarily focus on spatial-domain processing, neglecting the frequency domain's potential to capture global color distributions and long-range dependencies. To address these limitations, we propose FUSION, a dual-domain deep learning framework that jointly leverages spatial and frequency domain information. FUSION independently processes each RGB channel through multi-scale convolutional kernels and adaptive attention mechanisms in the spatial domain, while simultaneously extracting global structural information via FFT-based frequency attention. A Frequency Guided Fusion module integrates complementary features from both domains, followed by inter-channel fusion and adaptive channel recalibration to ensure balanced color distributions. Extensive experiments on benchmark datasets (UIEB, EUVP, SUIM-E) demonstrate that FUSION achieves state-of-the-art performance, consistently outperforming existing methods in reconstruction fidelity (highest PSNR of 23.717 dB and SSIM of 0.883 on UIEB), perceptual quality (lowest LPIPS of 0.112 on UIEB), and visual enhancement metrics (best UIQM of 3.414 on UIEB), while requiring significantly fewer parameters (0.28M) and lower computational complexity, demonstrating its suitability for real-time underwater imaging applications.

1. Introduction

Underwater imaging plays a crucial role in fields like marine biology, underwater archaeology, and autonomous underwater vehicle (AUV) navigation. However, it faces challenges such as light absorption and scattering, leading to low contrast, color casts (bluish and greenish hues), and blurriness, which hinder high-level vision tasks like object detection and segmentation [3, 9, 16]. Traditional underwater image enhancement (UIE) methods, such as histogram equalization and dehazing algorithms, struggle with the complex degradations in underwater environments [17]. Advanced cameras also fail to address non-uniform light attenuation, where shorter wavelengths like blue and green

penetrate deeper underwater, distorting color balance and reducing task performance [25].

Deep learning-based techniques have recently shown promise in low-level vision tasks. State-of-the-art underwater image restoration (UIR) methods often use identical receptive field sizes for R-G-B channels, ignoring wavelength-dependent degradation patterns. Sharma et al. [18] demonstrated that varying receptive field sizes (e.g., R(3×3), G(5×5), B(7×7)) improves UIR by capturing local and global features. Encoder-decoder networks commonly used in UIR capture broad contexts but lose spatial details during downsampling [19, 36]. High-resolution networks avoid downsampling but struggle with encoding global context needed for coherent enhancement. Most UIR methods focus solely on spatial-domain processing, overlooking long-range dependencies and global color distributions essential for effective UIE.

To address these limitations, we propose FUSION: Frequency-guided Underwater Spatial Image recONstructioN—a dual-domain framework tailored for underwater image enhancement. FUSION integrates spatial and frequency domain processing through four key modules: the Multi-Scale Spatial Module processes RGB channels using dedicated kernel sizes to handle wavelength-dependent attenuation; the Frequency Extraction Module refines magnitude information to capture global structural cues; the Frequency-Guided Fusion (FGF) Module combines spatial and frequency features for balanced local detail and global color consistency; and the Inter-Channel Fusion and Channel Calibration Module uses global attention and adaptive scaling to produce enhanced images with balanced color distribution.

Our dual-path architecture (Figure 1) processes RGB channels (D_R, D_G, D_B) independently. Spatial features are refined using multi-scale convolutions and attention mechanisms, while frequency features are extracted using Fourier analysis to capture global information. Spatial-frequency features are fused via FGF blocks for each channel before inter-channel fusion integrates RGB dependencies. A decoder stage with deconvolutional layers, attention mechanisms, residual connections, and adaptive recalibration balances RGB channels to produce enhanced images with improved visibility, color accuracy, and detail preservation.

To summarize, our contributions are as follows:

- **Dual-Domain Enhancement:** We introduce a parallel frequency pathway that captures long-range dependencies and global color distributions, complementing traditional spatial processing.
- **Dedicated Frequency Attention Module:** By preserving original phase while applying adaptive attention to the magnitude spectrum, our method captures global structural information critical for handling complex underwater degradations.
- **Inter-Channel Calibration for Color Correction:** A global recalibration stage, which employs learnable scaling factors to balance color intensities adaptively.

2. Related Works

2.1. Underwater Image Enhancement

Traditional methods for UIE have relied on image processing techniques such as histogram equalization, white balance adjustment, and dehazing algorithms based on physical models of light propagation underwater [4, 7, 12]. While these methods can enhance contrast and correct color casts to some extent, they generally lack adaptability to varying underwater conditions and often fail to restore fine details and textures. Li et al. proposed a dehazing and color correction method using convolutional neural networks (CNNs) that leverage the statistical properties of underwater images [12]. FUnIE-GAN and Water-Net have shown promising results by learning mappings from degraded images to their enhanced counterparts using generative adversarial networks (GANs) [22, 37].

There exists minimal literature on frequency-based methods for image enhancement, with no prior application to underwater imaging. Kersting et al. used a GAN-based approach to enhance ultra-fast PSMA-PET scans via synthetic reconstruction, showing improved detection in prostate cancer staging [11]. Liu et al. applied frequency decomposition in PID^{2}Net for underwater descattering and denoising, though not explicitly using frequency-domain learning [23]. Li et al. fused polarization cues with wavelet-based subband processing to improve defect visibility on reflective surfaces [21]. Agaian et al. proposed transform-based enhancement using orthogonal bases like Fourier and Hadamard with quantifiable performance metrics [1]. Wang proposed a parallel frequency-domain low-light framework that decouples contrast and structure restoration [32], while Wang et al. designed FourLLIE, leveraging Fourier amplitude mappings and SNR-guided fusion for efficient low-light enhancement [31].

2.2. Attention Mechanisms in Image Enhancement

Attention mechanisms are integrated into deep learning models to improve feature representation by focusing on the most informative parts of the input. In the context of image enhancement, attention modules can help models learn where to emphasize or suppress features, leading to better restoration of degraded images.

Chen et al. introduced an attention-based UIE method that employs a multi-scale attention mechanism to adaptively enhance features at different resolutions [22, 37]. Similarly, Li et al. utilized channel attention in their network to weigh the importance of different feature maps, improving the overall enhancement quality. While these

methods have shown effectiveness, they often increase the model's complexity and computational load [12, 33].

2.3. Shortcomings Addressed

The primary limitation in current underwater image restoration and enhancement approaches is that they focus predominantly on spatial-domain processing, overlooking the frequency domain's ability to capture global color distributions and long-range dependencies. This omission often results in residual color imbalances and artifacts, especially under severe wavelength-dependent attenuation [33]. Additionally, certain models that are able to capture these domains (like Fine-tuned GANs) require very heavy computational power, which makes them not viable for deployment and scalability scenarios [12]. Our proposed FUSION addresses these issues through a dual-domain design that efficiently processes each color channel in both spatial and frequency domains while having a quick inference time and low-memory compute. By incorporating multi-stage and multi-domain attention mechanisms with channel-wise recalibration, FUSION also preserves fine details, reduces artifacts, and balances color distributions.

3. Proposed Method: *FUSION*

The proposed architecture enhances underwater images through a dual-path framework that integrates spatial domain processing and frequency domain processing. The input image $D^{h\times w\times 3}$ is split into three independent channels, D_R, D_G, and D_B, which are processed separately in both domains to extract complementary features.

In the spatial domain, each channel undergoes multi-scale convolution with kernel sizes 3×3, 5×5, and 7×7 to capture features at varying receptive fields. This ensures that color-specific distortions are addressed independently, preventing the propagation of noisy features while preserving crucial wavelength-driven contextual information, as suggested in [29]. These features are refined using a Channel and Spatial Attention Module (CBAM) and residual connections to preserve information.

In parallel, frequency domain features are extracted by transforming each channel into the frequency domain using a 2D Fast Fourier Transform (FFT). The magnitude of the frequency representation is processed using 1×1 convolutional layers and refined with a Frequency Attention mechanism. The inverse FFT (IFFT) reconstructs these features back into the spatial domain.

The outputs from the spatial ($f^3_{R/G/B,\dots}$) and frequency ($freq_{R/G/B,\dots}$) domains are fused using FGF blocks. Finally, the fused features are passed through a decoder with a global attention (CBAM) and channel recalibration to adaptively balance RGB channels, producing the enhanced underwater image $E^{h\times w\times 3}$.

3.1. Spatial Domain Processing

The spatial domain processing path extracts features from each channel of the input image $D^{h\times w\times 3}$ by leveraging multi-scale feature extraction, attention mechanisms, and residual refinement. Each channel, D_R, D_G, and D_B, is processed independently to capture spatial patterns at multiple scales $\{s_1, s_2, s_3\}$.

Initially, feature maps $f_i^1 = \Phi_i(D_i)$ are extracted from each channel $i \in \{R, G, B\}$ using convolutional operations with varying receptive fields. Specifically, f_R^1 represents the features extracted from the red channel using kernel size 3×3, while f_G^1 and f_B^1 are obtained with 5×5 and 7×7 kernels, respectively. This multi-scale extraction $\{f_R^1, f_G^1, f_B^1\}$ enables the network to capture hierarchical features across the feature dimension with varying spatial contexts.

To enhance these features, a two-stage attention mechanism $\mathcal{A} = \mathcal{A}_c \circ \mathcal{A}_s$ is applied independently to each channel. In the first stage, channel attention $\mathcal{A}_c$ aggregates global information by computing scaling weights W_{channel} based on pooled statistics of the feature map:

$$W_{\text{channel}} = \sigma\big(\mathbf{W}_2 \cdot \phi(\mathbf{W}_1 \cdot g(f_i^1))\big) \tag{1}$$

Here $g(f_i^1)$ represents global average pooling, $\phi(\cdot)$ implements ReLU activation, and $\mathbf{W}_1$, $\mathbf{W}_2$ are learnable weight matrices with reduction ratio r. The feature map is then scaled element-wise as $f_{\text{channel-att}} = W_{\text{channel}} \odot f_i^1$.

In the second stage, spatial attention $\mathcal{A}_s$ refines these channel-weighted features by focusing on spatially significant regions through attention mapping. This is achieved by computing spatial attention weights:

$$W_{\text{spatial}} = \sigma\big(h(f_{\text{channel-att}})\big) \tag{2}$$

$$h(f_{\text{channel-att}}) = \psi\Big([\mathcal{P}_{avg}(f_{\text{channel-att}}); \mathcal{P}_{max}(f_{\text{channel-att}})]\Big) \tag{3}$$

The function h aggregates information across channels via concatenated max and average pooling operations, followed by a spatial transformation. The final attention-refined feature map is given by $f_{\text{spatial-att}} = W_{\text{spatial}} \odot f_{\text{channel-att}}$.

After applying both attention mechanisms, the refined feature maps for each channel are denoted as $f_i^2 = \mathcal{A}(f_i^1) = \mathcal{A}_s(\mathcal{A}_c(f_i^1))$ for $i \in \{R, G, B\}$. To preserve original spatial information and improve gradient flow during training, residual connections are added:

$$f_i^3 = f_i^2 + f_i^1 \quad \forall i \in \{R, G, B\} \tag{4}$$

These skip connections ensure that low-level features are preserved throughout the network while allowing the learning of residual mappings. The outputs, f_R^3, f_G^3, f_B^3, represent the final spatial representations for each channel after

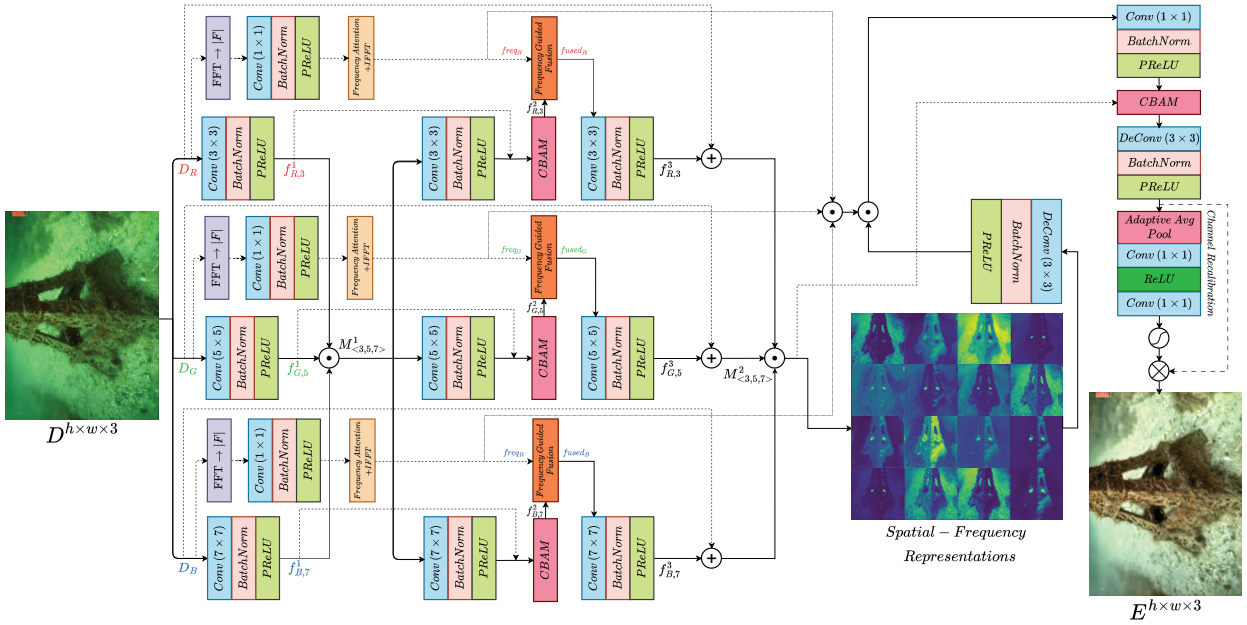

Figure 2. Overview of our proposed FUSION architecture for UIE. The model takes a degraded underwater image as input and restores it with enhanced visual quality.

multi-scale feature extraction, attention-based refinement, and residual enhancement.

By processing each color channel independently, we address the unique degradation patterns in underwater images where different wavelengths of light are attenuated at rates dependent on depth and water properties. The multi-scale feature extraction with varying kernel sizes is specifically designed to capture the diverse spatial characteristics present in underwater scenes, from fine-grained textures to broader structural elements.

3.2. Frequency Domain Processing

The frequency domain processing path complements the spatial domain by extracting and refining frequency features from each channel of the input image $D^{h\times w\times 3}$. This path leverages Fourier transforms, magnitude extraction, frequency attention, and inverse reconstruction to capture global contextual information that is often inaccessible in the spatial domain.

Each channel, D_i for $i \in \{R, G, B\}$, is independently transformed into the frequency domain using a 2D Fast Fourier Transform (FFT). For a given channel, the FFT produces a complex-valued representation $F_i = \mathcal{F}(D_i)$ containing both real and imaginary components. The magnitude of this representation is extracted as:

$$|F_i| = \sqrt{\text{Re}(F_i)^2 + \text{Im}(F_i)^2} \tag{5}$$

This magnitude $|F_i|$ captures global structural information about the input channel, where $\text{Re}(F_i)$ and $\text{Im}(F_i)$ denote

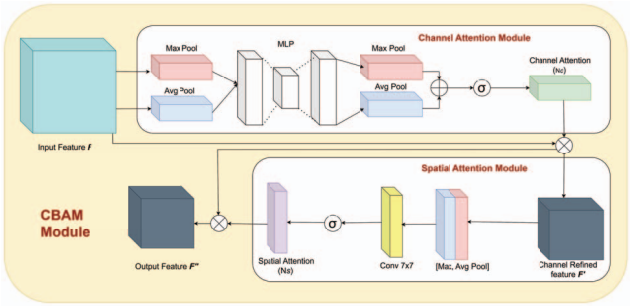

Figure 3. Architecture of the CBAM block [34]

the real and imaginary parts of the frequency representation. To refine these magnitude features, we apply a series of transformations in the frequency domain. The magnitude map $|F_i|$ undergoes linear transformations with learnable weight matrices W_1 and W_2 to reduce dimensionality and enhance discriminative features:

$$\hat{F}_i = W_2 \cdot \phi(W_1 \cdot |F_i|) \tag{6}$$

These transformations incorporate PReLU activation function $\phi(\cdot)$ and are followed by normalization to stabilize feature distributions across varying underwater conditions. Since underwater images suffer from wavelength-dependent attenuation that manifests differently in the frequency spectrum, these transformations help isolate discriminative frequency features that carry reliable information about the scene. A Frequency Attention Module fur-

ther enhances these features by computing attention weights W_{freq} for each channel:

$$W_{\text{freq}} = \sigma(W_4 \cdot \phi(W_3 \cdot g(|F_i|))) \tag{7}$$

Here $g(|F_i|)$ represents global average pooling, W_3 and W_4 are learnable weights, and $\sigma(\cdot)$ denotes sigmoid activation. The refined magnitude map $|F_i|_{\text{refined}} = W_{\text{freq}} \odot |F_i|$ adaptively amplifies important frequency components while suppressing less informative ones. This attention mechanism is particularly crucial for underwater imagery where certain frequency bands may be more degraded than others depending on water properties and depth. The refined magnitude is then recombined with the original phase information $\Theta_i = \text{Phase}(F_i)$ to reconstruct a complex-valued frequency representation:

$$F_i' = |F_i|_{\text{refined}} \cdot e^{j \cdot \Theta_i} \tag{8}$$

This phase preservation is essential as it maintains structural coherence while allowing magnitude enhancement. The exponential phase term can be expressed as $e^{j \cdot \Theta_i} = \cos(\Theta_i) + j \cdot \sin(\Theta_i)$ where $\Theta_i = \arctan\left(\frac{\text{Im}(F_i)}{\text{Re}(F_i)}\right)$. Finally, an inverse FFT (IFFT) transforms the refined frequency representation back into the spatial domain:

$$f_{\text{freq},i} = \mathcal{F}^{-1}(F_i') \tag{9}$$

The resulting frequency-derived feature maps, $f_{\text{freq},i}$ for $i \in \{R, G, B\}$, capture global contextual information that complements the localized details extracted in the spatial domain. These frequency features effectively represent long-range dependencies between pixels and global color distributions, which are particularly valuable for underwater image enhancement where visibility degradation affects the entire image non-uniformly.

By processing frequency information independently for each color channel, our approach addresses the channel-specific degradation patterns common in underwater environments, where red wavelengths attenuate more rapidly than green and blue wavelengths with increasing depth according to $I(\lambda, d) = I_0(\lambda)e^{-\beta(\lambda)d}$ [35].

3.3. Frequency Guided Fusion

We integrate spatial and frequency features through our FGF blocks, which operate independently for each channel (Red, Green, Blue). These blocks combine complementary information from spatial domain ($f_{\text{spatial},i}$) and frequency domain ($f_{\text{freq},i}$) to produce fused features $f_{\text{fused},i}$ for each channel $i \in \{R, G, B\}$.

For each color channel, we first concatenate the spatial feature map $f_{\text{spatial},i}$ and the frequency feature map $f_{\text{freq},i}$ along the channel dimension:

$$f_{\text{concat},i} = \mathcal{C}(f_{\text{spatial},i}, f_{\text{freq},i}) \tag{10}$$

This creates a unified representation containing both local spatial details and global frequency characteristics crucial for underwater image enhancement. We then transform the concatenated feature map through a convolution operation:

$$f_{\text{fused},i} = W_i * f_{\text{concat},i} \tag{11}$$

with learnable weights W_i to reduce dimensionality while integrating the two complementary modalities. This ensures that we preserve discriminative features from both domains while managing computational complexity.

The outputs of our FGF blocks, $f_{\text{fused},i}$ for $i \in \{R, G, B\}$, represent channel-specific fused representations that combine both fine-grained spatial details and comprehensive frequency information, capturing both local textures and global color distributions.

3.4. Inter-Channel Fusion and Channel Calibration

In the final stage of our architecture, we refine the fused feature representations from each RGB channel to produce the enhanced underwater image E. To ensure consistency in feature representation while mitigating underwater distortions, we integrate residual enhancements, spatial-frequency fusion, and adaptive recalibration.

First, we reinforce each fused feature map by adding back the corresponding input channel, ensuring that the residual information is preserved without disrupting learned features:

$$f_{\text{residual},i} = f_{\text{fused},i} + f_{\text{input},i}, \quad i \in \{R, G, B\} \tag{12}$$

We concatenate these residual-enhanced features to form a unified representation f_{concat}, enabling our model to leverage inter-channel dependencies effectively. To increase feature expressivity and capture richer spatial characteristics, we project this representation into a higher-dimensional feature space using transformation $\mathcal{T}_d$, yielding:

$$f_d = \phi(\mathcal{T}_d(f_{\text{concat}})) \tag{13}$$

where ϕ denotes a non-linear activation function. Parallel to this, we extract frequency domain features $f_{\text{freq},i}$ for each RGB channel to capture structural variations that may be less evident in the spatial domain. These features are concatenated as f_{freq}, providing complementary information for the fusion process. To effectively integrate spatial and frequency domain representations, we apply a learned transformation $\mathcal{T}_f$:

$$f_{\text{fusion}} = \phi(\mathcal{T}_f(f_d, f_{\text{freq}})) \tag{14}$$

This allows us to capture localized textures and global structures simultaneously, ensuring effective feature aggregation. Since different regions of the image may require varying levels of enhancement, we refine the fused features

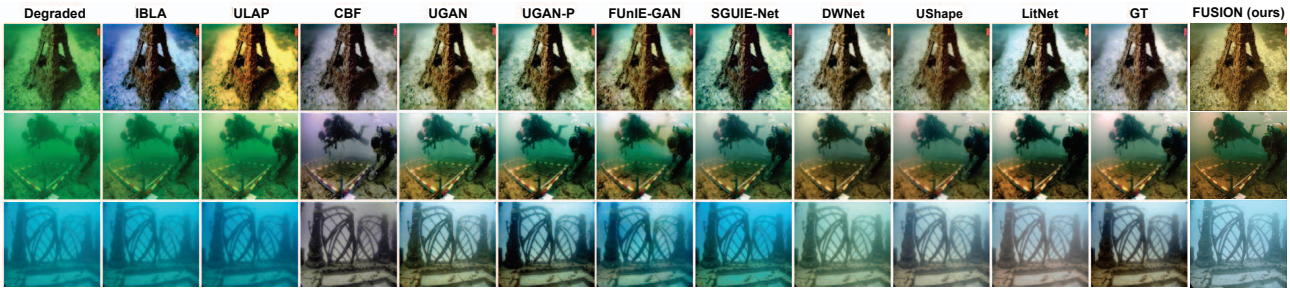

Figure 4. Visual comparisons on the UIEB dataset.

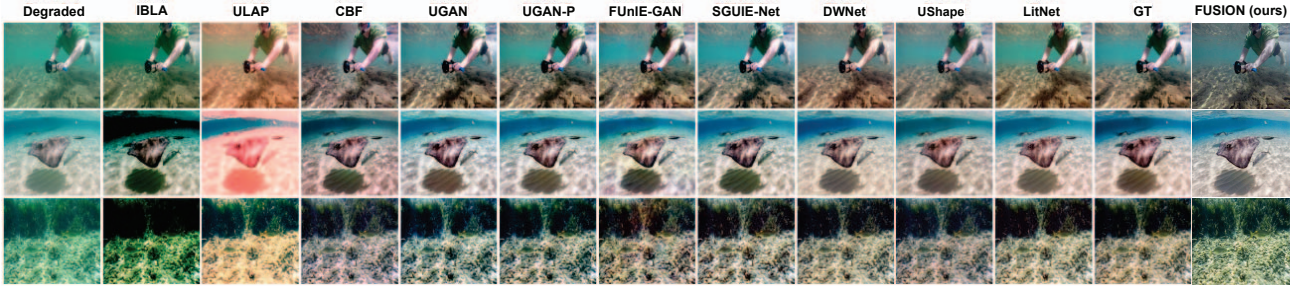

Figure 5. Visual comparisons on the EUVP dataset.

using a CBAM-based global attention mechanism $\mathcal{A}$ that selectively emphasizes important regions:

$$f_{\text{attn}} = \mathcal{A}(f_{\text{fusion}}, f_{\text{concat}}) \tag{15}$$

The attention-refined representation undergoes transformation through $\mathcal{T}_e$, reconstructing a coherent spatial representation $E = \phi(\mathcal{T}_e(f_{\text{attn}}))$. However, this yields pre-channel-calibrated reconstructions, which need further color distribution balancing. To address this and mitigate unwanted shifts, we implement an adaptive recalibration mechanism that generates per-channel scaling factors:

$$W_{\text{calibration}} = \sigma(W_2 \cdot \phi(W_1 \cdot g(E))) \tag{16}$$

where $g(E)$ extracts global descriptors summarizing the image's color characteristics, and σ normalizes the scaling factors to maintain RGB channel balance. The final enhanced image is obtained through element-wise calibration:

$$E_{\text{final}} = E \odot W_{\text{calibration}} \tag{17}$$

This adaptive weighting scheme ensures a visually coherent and perceptually enhanced underwater image by dynamically adjusting color balance and preserving structural details, mitigating common artifacts found in traditional enhancement techniques.

4. Results

Table 1. Evaluation on UIEB test set with the best-published works for UIE. First, second, and third best performances are represented in red, blue, and green colors, respectively. ↓ indicates lower is better.

Method	PSNR	SSIM	LPIPS↓	UIQM	UISM	BRISQUE↓
UDCP [5]	13.026	0.545	0.283	1.922	7.424	24.133
GBdehaze [13]	15.378	0.671	0.309	2.520	7.444	23.929
IBLA [26]	19.316	0.690	0.233	2.108	7.427	23.710
ULAP [30]	19.863	0.724	0.256	2.328	7.362	25.113
CBF [2]	20.771	0.836	0.189	3.318	7.380	20.534
UGAN [6]	23.322	0.815	0.199	3.432	7.241	27.011
UGAN-P [6]	23.550	0.814	0.192	3.396	7.262	25.382
FUnIE-GAN [10]	21.043	0.785	0.173	3.250	7.202	24.522
SGUIE-Net [28]	23.496	0.853	0.136	3.004	7.362	24.607
DWNet [29]	23.165	0.843	0.162	2.897	7.089	24.863
Ushape [24]	21.084	0.744	0.220	3.161	7.183	24.128
Lit-Net [27]	23.603	0.863	0.130	3.145	7.396	23.038
FUSION (Ours)	23.717	0.883	0.112	3.414	7.429	23.193

Experimental Settings We first evaluate the performance of our proposed FUSION framework on three widely used underwater image datasets: UIEB [14], EUVP [10], and SUIM-E [28]. All images across these datasets are resized to a uniform resolution of 256×256 prior to training and evaluation. For training, we utilize the EUVP dataset, which contains 11,435 paired underwater images, while its test set consists of 515 image pairs of the same resolution. The UIEB dataset comprises 890 paired images, from which 800 are randomly selected for training, and the remaining 90 images are used for testing (following [14]). The SUIM-E dataset includes 1,635 images, with 1,525 used for training and 110 for evaluation (following [27]).

Table 2. Ablation hardware comparisons with respect to average performance across datasets ($\overline{\text{Metric}}$ denotes the average of that metric across the 3 datasets used).

Configuration	Freq. Attn	Freq. Branch	Freq. Fusion	Chan. Calib	Local Attn	Global Attn	Inference Time (ms)	GFLOPs	$\overline{\text{UISM}}$	$\overline{\text{LPIPS}}$	$\overline{\text{BRISQUE}}$
Full Model (FUSION)	✓	✓	✓	✓	✓	✓	128.68	36.73	7.385	0.135	23.797
No Frequency Attention	✗	✓	✓	✓	✓	✓	128.53	36.71	6.395	0.207	27.643
No Frequency Branch	✓	✗	✓	✓	✓	✓	88.89	36.55	5.996	0.255	29.553
No Frequency Guided Fusion	✓	✓	✗	✓	✓	✓	90.29	36.71	6.192	0.226	28.370
No Channel Calibration	✓	✓	✓	✗	✓	✓	128.70	36.73	6.164	0.230	28.517
No Local Attention	✓	✓	✓	✓	✗	✓	75.87	36.69	6.453	0.210	27.627
No Global Attention	✓	✓	✓	✓	✓	✗	110.69	36.72	6.561	0.200	27.167
Spatial Only	✗	✗	✗	✓	✓	✓	89.01	36.55	5.908	0.250	29.320
Minimal Model	✗	✗	✗	✗	✗	✗	18.49	36.49	5.704	0.276	30.607

Table 3. Evaluation on EUVP dataset with the best-published works for UIE. First, second, and third best performances are represented in red, blue, and green colors, respectively. ↓ indicates lower is better.

Method	MSE↓	PSNR	SSIM	UIQM	LPIPS↓	UISM	BRISQUE↓
UGAN [6]	0.355	26.551	0.807	2.896	0.220	6.833	35.859
UGAN-P [6]	0.347	26.549	0.805	2.931	0.223	6.816	35.099
FUnIE-GAN [10]	0.386	26.220	0.792	2.971	0.212	6.892	30.912
FUnIE-GAN-UP [10]	0.600	25.224	0.788	2.935	0.246	6.853	34.070
Deep SESR [8]	0.325	27.081	0.803	3.099	0.206	7.051	35.179
DWNet [29]	0.276	28.654	0.835	3.042	0.173	7.051	30.856
Ushape [24]	0.370	26.822	0.811	3.052	0.187	6.843	35.648
Lit-Net [27]	0.225	29.477	0.851	3.027	0.169	7.011	32.109
FUSION (Ours)	0.208	28.671	0.862	3.220	0.174	7.048	29.547

To comprehensively assess the visual quality and perceptual fidelity of enhanced images, we compare our method against a range of state-of-the-art (SOTA) underwater image enhancement (UIE) approaches using both full-reference and no-reference metrics. These include Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS), along with perceptual quality measures such as the Underwater Image Quality Measure (UIQM), Underwater Image Sharpness Measure (UISM), and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE). Tables 1 and 3 present a detailed comparison of the quantitative results on the UIEB and EUVP datasets, respectively.

4.1. Comparison with State-of-the-Art

We present a quantitative and quantitative evaluation demonstrating that FUSION consistently outperforms competing methods across all evaluated metrics, achieving state-of-the-art results. In particular, on the UIEB test set (Table 1), FUSION achieves a PSNR of 23.717 dB and an SSIM of 0.883, indicating a very high reconstruction fidelity and structural similarity. It also records the lowest LPIPS score (0.112), reflecting superior perceptual quality and detail preservation. We observe similar trends on the EUVP dataset (Table 3), where FUSION attains a PSNR of 28.671 dB and the highest SSIM value of 0.862, alongside a low LPIPS score (0.174). Figure 6 depicts a bubble chart illustrating the trade-off between average PSNR and GFLOPs for various models, further validating the balance between efficiency and effectiveness of our approach.

Table 4. Comparison with the model parameters and GFLOPs of SOTA models at an input size of 256×256. Lower is better. First best is in red, second best in blue.

Method	Parameters (M)	FLOPs (G)
WaterNet [15]	24.8	193.7
UGAN [6]	57.17	18.3
FUnIE-GAN [10]	7.71	10.7
Ucolor [20]	157.4	443.9
SGUIE-Net [28]	18.55	123.5
DWNet [29]	0.48	18.2
Ushape [24]	65.6	66.2
LitNet [27]	0.54	17.8
Ours	0.28	36.73

We evaluate the visual quality of our FUSION framework through qualitative comparisons. Figures 4 and 5 show enhancement results for the UIEB and EUVP datasets alongside outputs from state-of-the-art methods. FUSION recovers finer structural details and preserves subtle textures, restoring balanced color distributions and improving contrast to mitigate underwater distortions like color casts and low visibility. UIEB and EUVP results (Figure 4) enhance natural hues and recover important scene details better than competing methods.

In addition to quantitative performance, we also assess the efficiency of our approach. Table 4 summarizes the model parameters and GFLOPs for our method compared to other leading UIE models at an input size of 256×256. Notably, FUSION achieves superior enhancement results with a significantly lower number of parameters (0.28M) and competitive GFLOPs (36.73), justifying its potential for deployment in real-time and resource-constrained settings.

4.2. Ablation Study

Quantitative Analysis. From the ablation studies across UIEB and EUVP, it is evident that each architectural component contributes meaningfully to overall performance. Removing frequency attention, branch, or guided fusion consistently leads to notable degradation in perceptual quality (higher LPIPS, lower UIQM and UISM), affirming the

Table 5. Ablation performance on UIEB.

Configuration	Freq. Attn	Freq. Branch	Freq. Fusion	Chan. Calib	Local Attn	Global Attn	UIQM	UISM	LPIPS	BRISQUE
Full Model (FUSION)	✓	✓	✓	✓	✓	✓	3.414	7.429	0.112	23.19
No frequency attention	✗	✓	✓	✓	✓	✓	2.978	7.235	0.153	24.81
No Frequency Branch	✓	✗	✓	✓	✓	✓	2.903	6.606	0.231	27.25
No Frequency Guided Fusion	✓	✓	✗	✓	✓	✓	2.961	6.821	0.202	26.34
No Channel Calibration	✓	✓	✓	✗	✓	✓	2.827	6.751	0.214	26.68
No Local Attention	✓	✓	✓	✓	✗	✓	3.005	7.102	0.169	25.22
No Global Attention	✓	✓	✓	✓	✓	✗	3.000	7.268	0.148	24.37
Spatial Only	✗	✗	✗	✓	✓	✓	2.896	6.660	0.225	26.91
Minimal Model	✗	✗	✗	✗	✗	✗	2.720	6.410	0.258	28.43

Table 6. Ablation Study Results on the EUVP Dataset

Configuration	Freq. Attn	Freq. Branch	Freq. Fusion	Chan. Calib	Local Attn	Global Attn	UIQM	UISM	LPIPS	BRISQUE
Full Model (FUSION)	✓	✓	✓	✓	✓	✓	3.220	7.048	0.174	29.547
No Frequency Attention	✗	✓	✓	✓	✓	✓	2.839	6.118	0.227	34.21
No Frequency Branch	✓	✗	✓	✓	✓	✓	2.674	5.709	0.249	35.68
No Frequency Guided Fusion	✓	✓	✗	✓	✓	✓	2.665	5.744	0.247	35.53
No Channel Calibration	✓	✓	✓	✗	✓	✓	2.640	5.646	0.252	35.89
No Local Attention	✓	✓	✓	✓	✗	✓	2.538	6.222	0.232	34.51
No Global Attention	✓	✓	✓	✓	✓	✗	2.640	6.392	0.224	33.92
Spatial Only	✗	✗	✗	✓	✓	✓	2.373	5.557	0.261	36.43
Minimal Model	✗	✗	✗	✗	✗	✗	2.106	5.553	0.278	37.21

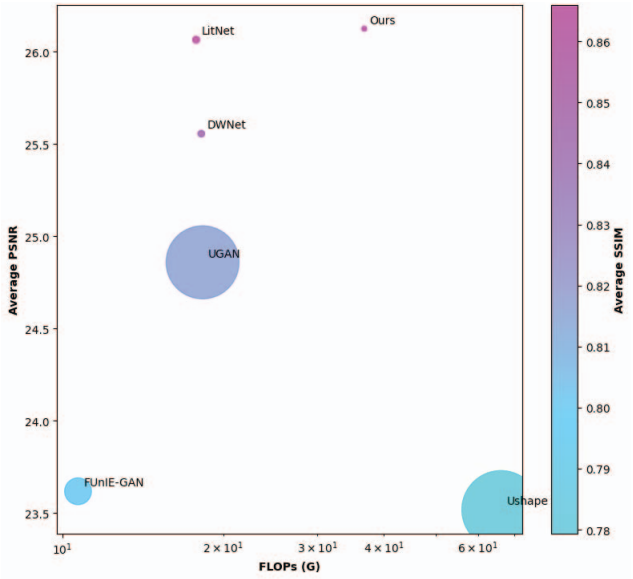

Figure 6. Bubble chart comparing the trade-off between average PSNR and GFLOPs for various UIE models

critical role of frequency-aware design in FUSION. Similarly, channel calibration and attention blocks - both local and global - also drive significant gains, especially in structural sharpness and perceptual realism. Interestingly, global attention appears to be particularly vital in retaining fine-grained global coherence, while local attention improves texture fidelity. Models stripped of frequency modules or reduced to spatial-only designs suffer from reduced enhancement quality, confirming the synergy between spectral and spatial representations in underwater image enhancement.

Hardware Efficiency. Beyond accuracy, FUSION maintains competitive inference efficiency, showcasing a balanced trade-off between performance and resource footprint. The full model runs at 128.68 ms with just 36.73 GFLOPs, which is notably efficient given its multi-branch design. Ablating the frequency branch or removing attention mechanisms reduces inference time - e.g., down to 75.87 ms without local attention - but at the cost of performance. While the minimal model is fastest at 18.49 ms, it offers the weakest performance, backing the need for our architectural complexity to achieve enhancement fidelity. Overall, FUSION demonstrates that strategic architectural additions, particularly those exploiting frequency and attention cues, yield meaningful gains without sacrificing deployability in real-time or resource-limited scenarios.

5. Conclusion

We propose FUSION (Frequency-guided Underwater Spatial Image recOnstructioN), a novel dual-domain framework that combines multi-scale spatial feature extraction with FFT-based frequency processing for underwater image enhancement. Leveraging adaptive attention, FUSION effectively addresses complex degradations in underwater scenes. Extensive evaluations on UIEB, EUVP, and SUIM-E show superior performance across PSNR, SSIM, LPIPS, UIQM, UISM, and BRISQUE metrics. FUSION also offers a strong balance between quality and efficiency, making it ideal for real-time use on AUVs.

References

[1] S.S. Agaian, K. Panetta, and A.M. Grigoryan. Transform-based image enhancement algorithms with performance measure. *IEEE Transactions on Image Processing*, 10(3): 367–382, 2001. 2

[2] Codruta O Ancuti, Cosmin Ancuti, Christophe De Vleeschouwer, and Philippe Bekaert. Color balance and fusion for underwater image enhancement. *IEEE Transactions on image processing*, 27(1):379–393, 2017. 6, 3

[3] Saeed Anwar, Chongyi Li, and Fatih Porikli. Deep underwater image enhancement. *arXiv preprint arXiv:1807.03528*, 2018. 1

[4] Xiaofeng Cong, Yu Zhao, Jie Gui, Junming Hou, and Dacheng Tao. A comprehensive survey on underwater image enhancement based on deep learning. *arXiv preprint arXiv:2405.19684*, 2024. 2

[5] Paul Drews, Erickson Nascimento, Filipe Moraes, Silvia Botelho, and Mario Campos. Transmission estimation in underwater single images. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 825–830, 2013. 6, 3

[6] Cameron Fabbri, Md Jahidul Islam, and Junaed Sattar. Enhancing underwater imagery using generative adversarial networks. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7159–7165. IEEE, 2018. 6, 7, 3

[7] Y. R. Gogireddy and J. R. Gogireddy. Advanced underwater image quality enhancement via hybrid super-resolution convolutional neural networks and multi-scale retinex-based defogging techniques. *arXiv preprint arXiv:2410.14285*, 2024. 2

[8] Md Jahidul Islam, Peigen Luo, and Junaed Sattar. Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. *arXiv preprint arXiv:2002.01155*, 2020. 7

[9] Md Jahidul Islam, Yue Xia, and Junaed Sattar. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2):3227–3234, 2020. 1

[10] Md Jahidul Islam, Youya Xia, and Junaed Sattar. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2):3227–3234, 2020. 6, 7, 3

[11] D. Kersting, K. Borys, A. Küper, et al. Staging of prostate cancer with ultra-fast psma-pet scans enhanced by ai. *European Journal of Nuclear Medicine and Molecular Imaging*, 52:1658–1670, 2025. 2

[12] Naresh Kumar, Juveria Manzar, Shivani, and Shubham Garg. Underwater image enhancement using deep learning. *Multimedia Tools and Applications*, 82:46789–46809, 2023. 2, 3

[13] Chongyi Li, Jichang Quo, Yanwei Pang, Shanji Chen, and Jian Wang. Single underwater image restoration by blue-green channels dehazing and red channel correction. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1731–1735. IEEE, 2016. 6, 3

[14] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29:4376–4389, 2019. 6

[15] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond, 2019. 7

[16] Chongyi Li, Chunle Guo, Wenhan Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond, 2020. 1

[17] Chongyi Li, Chunle Guo, Wenhan Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29:4376–4389, 2020. 1

[18] Chongyi Li, Chunle Guo, Wenhan Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29:4376–4389, 2020. 2

[19] Chongyi Li, Chunle Guo, Wenhan Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29:4376–4389, 2020. 2

[20] Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Transactions on Image Processing*, 30: 4985–5000, 2021. 7

[21] Y. Li, Y. Wang, H. Yu, R. Xin, J. Chu, and R. Zhang. Image enhancement method for paint defects based on polarization fusion technique. *Journal of Modern Optics*, 71(10–12): 364–374, 2024. 2

[22] B. Liu, X. Ning, S. Ma, and Y. Yang. Multi-scale dense spatially-adaptive residual distillation network for lightweight underwater image super-resolution. *Frontiers in Marine Science*, 10:1328436, 2023. 2

[23] Hedong Liu, Wenjie Zhang, Yilin Han, Xiaobo Li, Tiegen Liu, Jingsheng Zhai, Yefei Mao, Lin Xiao, and Haofeng Hu. Pid2net: A neural network for joint underwater polarimetric images descattering and denoising. *IEEE Sensors Journal*, 24(17):27803–27814, 2024. 2

[24] Lintao Peng, Chunli Zhu, and Liheng Bian. U-shape transformer for underwater image enhancement. *IEEE Transactions on Image Processing*, 2023. 6, 7, 3

[25] Yan-Tsung Peng and Pamela C. Cosman. Underwater image restoration based on image blurriness and light absorption. *IEEE Transactions on Image Processing*, 26(4):1579–1594, 2017. 2

[26] Yan-Tsung Peng and Pamela C Cosman. Underwater image restoration based on image blurriness and light absorption. *IEEE transactions on image processing*, 26(4):1579–1594, 2017. 6, 3

[27] Alik Pramanick, Arijit Sur, and V. Vijaya Saradhi. Harnessing multi-resolution and multi-scale attention for underwater image restoration, 2024. 6, 7, 3

[28] Qi Qi, Kunqian Li, Haiyong Zheng, Xiang Gao, Guojia Hou, and Kun Sun. Sguie-net: Semantic attention guided underwater image enhancement with multi-scale perception. *IEEE Transactions on Image Processing*, 31:6816–6830, 2022. 6, 7, 3

[29] Prasen Sharma, Ira Bisht, and Arijit Sur. Wavelength-based attributed deep neural network for underwater image restoration. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(1):1–23, 2023. 3, 6, 7

[30] Wei Song, Yan Wang, Dongmei Huang, and Dian Tjondronegoro. A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration. In *Advances in Multimedia Information Processing–PCM 2018: 19th Pacific-Rim Conference on Multimedia, Hefei, China, September 21-22, 2018, Proceedings, Part I 19*, pages 678–688. Springer, 2018. 6, 3

[31] Chenxi Wang, Hongjun Wu, and Zhi Jin. Fourllie: Boosting low-light image enhancement by fourier frequency information. In *Proceedings of the 31st ACM International Conference on Multimedia*, page 7459–7469, New York, NY, USA, 2023. Association for Computing Machinery. 2

[32] H. Wang. Frequency-based unsupervised low-light image enhancement framework. In *MultiMedia Modeling. MMM 2025*, pages Springer, Singapore. Springer, 2025. 2

[33] H. Wang, Z. Li, Y. Zhang, Y. Wang, and J. Li. Learning hybrid dynamic transformers for underwater image super-resolution. *Frontiers in Marine Science*, 11:1389553, 2024. 3

[34] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 3–19, 2018. 4

[35] Wending Xiang, Ping Yang, Shuai Wang, Bing Xu, and Hui Liu. Underwater image enhancement based on red channel weighted compensation and gamma correction model. *Opto-Electronic Advances*, 1:18002401–18002409, 2018. 5

[36] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 2

[37] Y. Zhang, J. Li, H. Wang, Y. Zhang, and Y. Wang. Daegan: Underwater image super-resolution based on symmetric dual attention and edge enhancement. *Symmetry*, 16(5):588, 2024. 2