

# Underwater Image Enhancement Via Dual-Path Frequency-Domain Fusion Network

Yupeng Ma, Qiuling Yang, Rongxin Zhu *Member, IEEE*, Yifei Li, Magd M. Abdel Wahab *Fellow, IEEE*,

**Abstract**—Underwater images typically suffer from color distortion and reduced clarity due to wavelength-dependent light absorption and scattering. Many existing learning-based methods learn a direct pixel-domain mapping from degraded images to enhanced results, while seldom exploiting explicit frequency-domain priors that are important for high-quality restoration. To address this, we propose DPF-Net, a Dual-Path Frequency-domain Fusion Network. Specifically, we first decompose the input into low- and high-frequency components via the discrete wavelet transform (DWT). We then design a low-frequency branch integrating Low-Frequency Guided Attention (LFGA) and the Color Feature Extraction Module (CFEM), and a high-frequency branch integrating High-Frequency Aware Modulation (HFAM) and the Detail Feature Extraction Module (DFEM), to capture frequency-specific characteristics. During upsampling, we introduce a WaveUp module consisting of the Dual-path Up-Interaction Module (DUIM) and a Multi-kernel Convolution-based Fusion Network (MCFN) to enable effective cross-frequency feature interaction and fusion, producing high-quality enhanced images. Extensive experiments on benchmark datasets demonstrate that DPF-Net achieves superior qualitative and quantitative performance compared with state-of-the-art methods, and real-world tests further verify its practical potential. Code is available at <https://github.com/YuMa-star-lab/DPF-Net>.

**Index Terms**—Underwater image enhancement, Discrete wavelet transform, Fast fourier transform

## I. INTRODUCTION

UNDERWATER imaging plays a critical role in underwater vision tasks. However, light absorption, scattering, and interference from suspended particles often cause image degradations, including color cast, insufficient illumination, haze effects, and noise [1]. Since degraded underwater images can significantly impact the efficiency of underwater vision tasks, there is a strong need to develop specialized methods for improving underwater image quality.

This work was partially supported by the National Natural Science Foundation of China under Grant 62362026; in part by the Key Project of Hainan Province under Grant ZDYF2023GXJS158; in part by the specific research fund of The Innovation Platform for Academicians of Hainan Province under Grant YSPTZX202314; in part by the Hainan Provincial Natural Science Foundation of China under Grants 626MS0092 and 626QN0556.

Yupeng Ma, Qiuling Yang, and Rongxin Zhu are with the School of Computer Science and Technology, Hainan University, Haikou 570228, China (e-mails: phpl@hainanu.edu.cn; qlyang@hainanu.edu.cn; Rongxin.zhu@hainanu.edu.cn). Corresponding author: Qiuling Yang.

Yifei Li is with the School of Engineering, Huzhou University, Huzhou 313000, China (e-mail: 03354@zjhu.edu.cn).

Magd M. Abdel Wahab is with Soete Laboratory, Dept. of Electrical Energy, Metals, Mechanical Constructions, and Systems, Faculty of Engineering and Architecture, Ghent University, Gent, Belgium; and Parallel and Distributed Systems Laboratory, Jožef Stefan Institute, Jamova cesta 39, Ljubljana SI-1000, Slovenia (e-mail: magd.abdelwahab@ugent.be).

Copyright © 2026 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

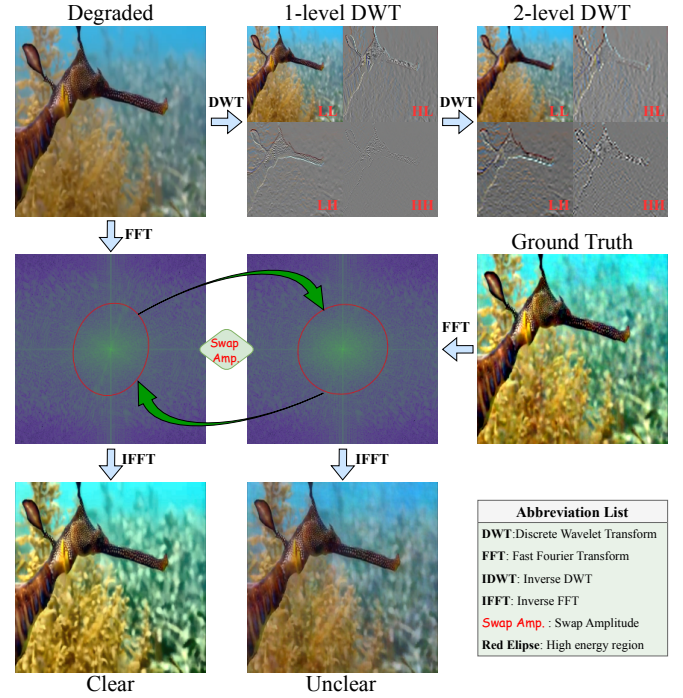


Fig. 1: Motivation. First row: 1-level and 2-level DWT decomposition results of degraded images; second and third row: results of amplitude spectrum swapping between degraded and ground-truth images.

To date, a variety of UIE approaches have been proposed, which can generally be categorized into two groups: physics-based and deep learning-based methods [2]. Physics-based methods mainly estimate and correct degradation parameters, such as scattering components, transmission, and background light, by constructing underwater imaging models [3], [4], [5], [6], [7], [8], [9], [10]. These approaches can mitigate the impact of suspended particles and scattering to some extent, thereby enhancing image contrast and clarity. However, they still suffer from several limitations, including susceptibility to inaccurate parameter estimation, limited generalization of model assumptions, and relatively high computational complexity.

Several representative deep learning UIE methods have been proposed. In [11], Jiang et al. proposed the lightweight CNN, utilizing underwater scene priors and a two-stage structure with multi-branch color enhancement and pixel attention modules for UIE. In [12], Li et al. introduced a fusion architecture to learn three confidence maps, which serves as

a baseline to evaluate UIE models. Generative Adversarial Networks (GANs) exhibit robust image generation capabilities. In [13], Fabbri et al. leveraged GANs to learn matching pairs, addressing the challenge of limited real data, and effectively enhanced underwater images. In addition, in [14], [15], [16], [17], and [18], Zhang et al. validated the effectiveness of multi-scale methods in applying to similar low-quality image scenarios. Some methods that integrate physical knowledge with CNNs have been also proposed. In [19], Guo et al. leveraged the gray world assumption theory to propose a color balance prior-guided hybrid sensing framework for underwater image restoration. In [20] and [21], Zhang et al. respectively introduced innovative methods combining wavelet transform fusion and wavelet-based adaptive contrast strategies, aiming to address the issues of color distortion, contrast degradation, and detail loss in underwater images.

However, most existing methods primarily rely on implicit pixel-level mappings and seldom explicitly model wavelet subbands, thus underutilizing frequency-domain priors that are essential for high-quality restoration. To address this, motivated by this frequency-aware perspective, we analyze degraded underwater images in the frequency domain. Specifically, as illustrated in Fig. 1 (first row), the low-frequency subbands of the DWT capture global image structures, such as color and illumination distortions, which are strongly related to the overall degradation. This makes DWT's low-frequency components well-suited for modeling global distortions caused by underwater light scattering and absorption. Meanwhile, the high-frequency subbands retain fine details and textures crucial for restoring sharpness and clarity, and are more sensitive to noise. These components help recover local structural details lost in the pixel domain due to noise. Moreover, the 2-level DWT decomposition exhibits more refined features than 1-level DWT or raw images. Fig. 1 (second and third rows) further shows that swapping the amplitude spectra of degraded and ground truth images in high-energy regions results in the ground truth image becoming unclear, while the degraded image appears clear. This demonstrates that the amplitude and phase spectra, particularly in high-energy regions, provide strong frequency-domain priors, which help restore key image structures by preserving important global features and suppressing noise in both low and high-frequency components.

Therefore, we propose a Dual-Path Frequency-domain Fusion Network for UIE, termed DPF-Net. In this method, the degraded image is first decomposed into low- and high-frequency features using DWT, providing learning frequency priors. Then, in the low-frequency path, we propose a joint module comprising Low-Frequency Guided Attention (LFGA) and Color Feature Extraction Module (CFEM), which leverages the Fourier transform to explicitly model the amplitude-phase spectrum of the low-frequency components, preserving global color/illumination consistency and reducing information loss. In the high-frequency path, our proposed joint module integrating High-Frequency Aware Modulation (HFAM) and Detail Feature Extraction Module (DFEM) employs a dynamic reweighting mechanism to adaptively emphasize informative textures while suppressing noise. Finally, to better achieve interaction and fusion of low- and high-frequency, we propose

the WaveUp module. This module enables cross-frequency interaction through the Dual-path Up-Interaction Module (DUIM), which integrates PixelShuffle and IDWT. It also ensures effective global fusion via the Multi-kernel Convolution-based Fusion Network (MCFN), ultimately restoring high-quality and clear images.

The main contributions of the paper are summarized as follows:

- Starting from the frequency-domain perspective of degraded images, leveraging the frequency prior information provided by wavelet and fourier transform, we propose a dual-path frequency-domain fusion network, termed DPF-Net, specifically designed to address the UIE task. Additionally, we also discuss the implementation and application of the lightweight DPF-Net-Lite method.
- In the low-frequency path, to mitigate the loss of color and illumination information, we newly develop a joint module comprising LFGA and CFEM. Specifically, LFGA generates guided frequency priors and content-adaptive attention maps to preserve global color/illumination consistency, while CFEM is designed to capture local structural and color cues.
- In the high-frequency path, to strengthen the network's capability in representing fine-grained details while effectively suppressing redundant noise, we propose a joint module integrating HFAM and DFEM. Specifically, HFAM employs an adaptive residual decomposition mechanism to modulate high-frequency features, while DFEM is designed to learn multi-scale structural detail information.
- To achieve interaction and fusion between low- and high-frequency paths, we propose the WaveUp module. This module consists of a DUIM, which integrates PixelShuffle and IDWT, and a MCFN that enables global features fusion.
- Extensive experiments and analyses demonstrate that the proposed DPF-Net not only significantly improves underwater image quality and consistently outperforms state-of-the-art (SOTA) UIE methods, but also achieves a favorable balance between computational efficiency and restoration performance, exhibiting strong potential for practical applications.

## II. RELATED WORKS

### A. Conventional UIE Methods

Conventional UIE methods are generally categorized into supplementary information-based, non-IFM (heuristic), and physical model-based approaches. Supplementary methods [22], [23] exploit extra cues (e.g., depth/stereo/polarization or multi-view frames) to improve visibility, but often require specialized acquisition and are scenario-dependent. Non-IFM methods enhance images by directly manipulating pixel statistics, such as HE [7], CLAHE [5], UCM [10], ICM [8], and MIP [6], which are efficient yet may amplify noise or cause over-enhancement under severe scattering. Physical model-based methods incorporate the underwater image formation model to compensate for absorption and scattering (e.g.,



IBLA [3], DCP [4], UDCP [9]), but remain sensitive to unknown, spatially varying water properties, limiting robustness in real scenes.

### B. Deep Learning-Based UIE Methods

Deep learning-based UIE methods [11]–[13], [24]–[29] learn the mapping from degraded underwater observations to visually enhanced outputs using data-driven models, and can be broadly categorized into supervised/paired training, unsupervised adaptation, and GAN-based frameworks. WaterNet [12] is a representative supervised approach that fuses multiple enhancement cues/branches to cope with diverse underwater degradations, leading to improved color rendition and visibility. For efficient deployment, FIVE A+ [11] presents an ultra-lightweight network with only 9K parameters for real-time UIE on resource-constrained devices. Unsupervised UIE has also gained attention in image restoration [30]; for example, USUIR [24] enhances underwater images by enforcing cross-domain consistency without paired labels, while UGAN [13] leverages GAN training with synthetically degraded pairs to improve color fidelity and contrast in real scenes. Despite promising progress, most existing deep UIE models primarily operate in the pixel/spatial domain, and may be less robust to complex degradations where frequency-domain priors (e.g., global appearance consistency and detail-noise separation) are critical.

Beyond UIE-specific networks, multi-scale decomposition has been widely explored in general image restoration and video processing [14]–[18], [28], [31]–[34]. UIEC<sup>2</sup>-Net [33] exploits dual color spaces to jointly correct color and preserve details, while CCMSRNet [34] incorporates multiscale Retinex-inspired modeling for illumination and color correction. Representative examples from generic restoration include ESTINet [14], which couples deep residual learning with recurrent modeling for efficient deraining, and DDMSNet [15], which integrates semantic and depth priors in a multi-stage pipeline for challenging desnowing. Transformer-based restorers such as GridFormer [17] and MB-TaylorFormer V2 [18] further improve multi-scale representation learning, where the former combines attention with residual dense features and the latter reduces complexity via Taylor expansion and multi-branch multi-scale patch embedding. Although these designs highlight the effectiveness of multi-scale modeling and global-local interaction for degradation removal, their direct applicability to UIE still warrants careful study due to wavelength-dependent attenuation and scattering, which entangle global color cast with local detail degradation.

### C. Hybrid UIE Strategies

Hybrid UIE strategies [19]–[21], [35]–[37] incorporate explicit priors into learning- or optimization-based pipelines to improve interpretability and generalization with less reliance on large-scale real underwater data. UWCNN [36] synthesizes training pairs using physically motivated degradation and scene priors to enhance color and contrast. GUPDM [37] embeds IFM-derived constraints into a dynamic architecture to

adapt to diverse water types and depths. ICSP-Retinex [35] formulates UIE via illumination-constrained, structure-preserved variational optimization, while GuidedHySensUIR [19] combines deep restoration with prior guidance for robust enhancement. However, most physics-guided hybrids mainly regularize UIE in the spatial domain or impose coarse global constraints, which may be insufficient for preserving fine-grained global color/illumination consistency and structural fidelity under complex degradations. From a frequency-aware perspective, WWPF [20] and WFAC [21] are wavelet-only pipelines relying on multi-scale subband fusion and hand-crafted contrast/color adjustments. Although wavelets offer spatial locality for edges and textures, they often leave (i) low-frequency global appearance (Fourier amplitude–phase statistics) and (ii) high-frequency noise-detail trade-offs without explicit, content-adaptive modeling. This limitation motivates our dual-path design that couples DWT-based subband decoupling with learnable Fourier amplitude–phase refinement.

## III. THE PROPOSED METHOD

As shown in Fig. 2, the proposed DPF-Net is designed to generate clear and high-quality underwater images. As observed, DPF-Net adopts an encoder-decoder design like the U-Net architecture. The encoder comprises dual-path WaveDown based on low- and high-frequency subbands, where DWT decouples global distortions (low-frequency) and local details (high-frequency) of the input features. Since pure wavelet subbands do not explicitly model amplitude and phase, which are crucial for global consistency, we introduce a learnable FFT-based APRM in the low-frequency path to refine amplitude (color/contrast) and phase (structural layout), enhancing global color/illumination consistency while preserving structural stability. In the high-frequency path, we apply adaptive re-weighting mechanism and feature enhancement to suppress noise and emphasize important edges and textures. The decoder incorporates the dual-path upsampling WaveUp module, which facilitates global interaction and fusion of low- and high-frequency features to generate high-quality images. To refine features and stabilize the global representation, we employ only the CFEM module within low-frequency path between the encoder and decoder. Additionally, the lightweight DPF-Net version, DPF-Net-Lite, includes only one WaveDown and WaveUp.

To achieve enhanced color restoration and detail improvement in images, we designed a joint loss function that comprehensively considers pixel consistency, structural preservation, and frequency-domain constraints.

### A. Dual-Path Encoder with Local Enhancement

Given an input degraded image  $I \in \mathbb{R}^{H \times W \times 3}$ , we first apply pointwise convolution (PW Conv)  $\tilde{f}_{pw}$  to extract shallow features  $F_0 \in \mathbb{R}^{H \times W \times C}$ , where  $C = 32$ . Subsequently, we employ the DWT based on the Haar wavelet to decompose the input. The Haar wavelet consists of a low-pass filter  $L$  and a high-pass filter  $H$ , as follows:

$$L = \frac{1}{\sqrt{2}}[1, 1]^T, \quad H = \frac{1}{\sqrt{2}}[1, -1]^T. \quad (1)$$

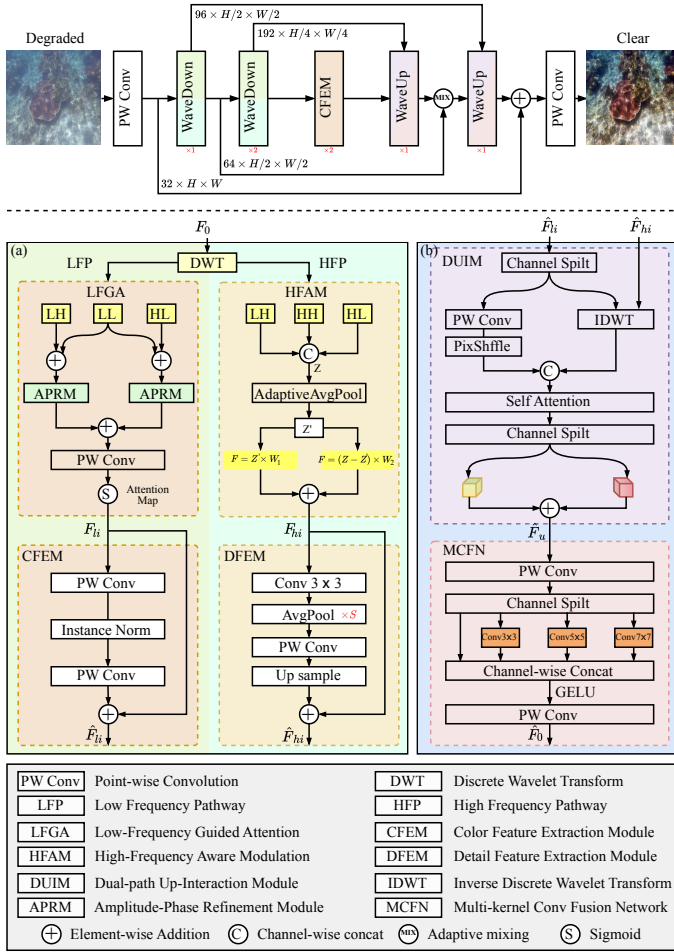


Fig. 2: The architecture of the proposed DPF-Net model. It consists of dual-path downsampling module, WaveDown as shown in (a). and dual-path upsampling module, WaveUp as shown in (b).

This yields four subbands, which can be expressed as:

$$I_{LL}, \{I_{LH}, I_{HL}, I_{HH}\} = \text{DWT}(I), \quad (2)$$

where  $I_{LL}, \{I_{LH}, I_{HL}, I_{HH}\} \in \mathbb{R}^{H/2 \times W/2 \times c}$  represent the low-frequency feature of the input image, and the high-frequency features in the vertical, horizontal, and diagonal directions, respectively. More specifically, the low-frequency component encapsulates the content and color information of the input image, while the other three high-frequency coefficients capture detailed information regarding global structures and textures [38]. These subbands are downsampled to half the resolution of the input. Next, we design low- and high-frequency path to capture features of the low- and high-frequency features, respectively.

**Low Frequency Path:** As illustrated in Fig. 2(a), to improve the loss of detail in the low-frequency features, we consider the subbands  $\{I_{LH}, I_{LL}, I_{HL}\}$  as the input to the low-frequency path. Within this path, we propose a joint module, composed of the LFGA and CFEM, to adaptively enhance the low-frequency features. LFGA leverages attention mechanisms to highlight important regions in the low-frequency features, mitigating detail loss, while CFEM focuses on capturing

local structural and color features, further strengthening the low-frequency representation. Specially, in LFGA module, to alleviate low-frequency deficiency in fine details, we inject the features  $\{I_{LH}, I_{HL}\}$  into the features  $\{I_{LL}\}$ :

$$F_1 = I_{LH} + I_{LL}, \quad F_2 = I_{HL} + I_{LL}. \quad (3)$$

To fully exploit frequency-domain priors for enhancing low-frequency features, we then design an Amplitude-Phase Refinement Module (APRM) to perform amplitude-phase domain refinement on  $F_i$  ( $i = 1, 2$ ), as shown in Fig. 3. The APRM module aims to refine the amplitude and phase information independently, enhancing the global structure and color consistency. Each  $F_i$  is first mapped to the complex frequency domain via the FFT [39], from which the amplitude and phase are extracted as:

$$\text{mag}_i, \text{pha}_i = \mathcal{F}(F_i), \quad (4)$$

where  $\text{mag}_i = |\mathcal{F}(F_i)|$  and  $\text{pha}_i = \angle \mathcal{F}(F_i)$ . Subsequently, as shown in Fig. 3, a lightweight residual bottleneck (LRB) is employed to refine these features independently:

$$\begin{aligned} \text{mag}'_i &= \tilde{f}_{\text{pw}}(\text{Concat}(\text{ReLU}(\text{BN}(\tilde{f}_{\text{pw}}(\text{mag}_i))), \text{mag}_i)), \\ \text{pha}'_i &= \tilde{f}_{\text{pw}}(\text{Concat}(\text{ReLU}(\text{BN}(\tilde{f}_{\text{pw}}(\text{pha}_i))), \text{pha}_i)), \end{aligned} \quad (5)$$

where  $\tilde{f}_{\text{pw}}(\cdot)$  denotes a PW Conv, and BN and Concat represent batch normalization and channel-wise concat, respectively. Leveraging Euler's formula, the refined amplitude and phase are then utilized to reconstruct the real and imaginary parts of the spectrum:

$$\begin{aligned} \text{real}_i &= \text{mag}'_i \odot \cos(\text{pha}'_i), \\ \text{imag}_i &= \text{mag}'_i \odot \sin(\text{pha}'_i), \end{aligned} \quad (6)$$

where  $\odot$  indicates element-wise multiplication. These parts are combined to form the complex spectrum  $\text{real}_i + j\text{imag}_i$ , which is mapped back to the spatial domain through the IFFT. To maintain feature stability, a residual connection is incorporated, followed by a PW Conv for effective integration. This procedure is formalized as

$$F_{di} = \tilde{f}_{\text{pw}}(\mathcal{F}^{-1}(\text{real}_i + j\text{imag}_i) + F_i). \quad (7)$$

Due to the focus of APRM on learning amplitude and phase, the full model incorporating APRM significantly enhances the stability of color and structural representation of degraded images. Therefore, we use PSNR, which focuses on overall quality, SSIM, which emphasizes structural information, and LPIPS, which prioritizes texture and detail preservation, for evaluation. As evidenced by the results in Fig. 4, the PSNR

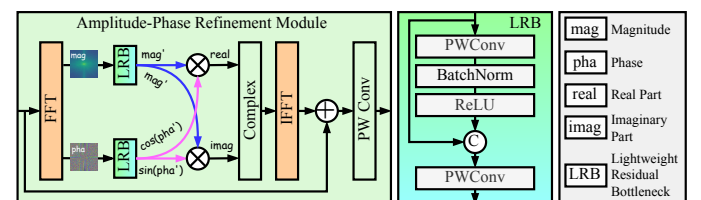


Fig. 3: The architecture of the APRM.

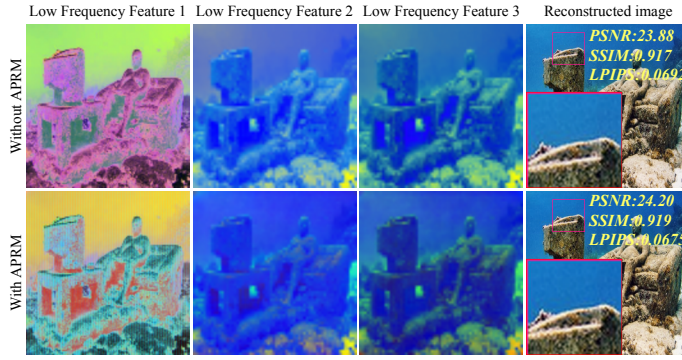


Fig. 4: Low-frequency path feature output results on T-90 dataset. The first row represents the scenario without APRM, while the second row depicts the full model within APRM.

value increases from 23.88 to 24.20, the SSIM value improves from 0.917 to 0.919, and LPIPS decreases from 0.0692 to 0.0675. These enhancements collectively make a substantial contribution to the quality of the final output.

To ensure the model effectively captures and emphasizes the salient low-frequency information for subsequent processing, the refined features  $F_{di}(i = 1, 2)$  are aggregated by PW Conv  $\tilde{f}_{pw}$  and sigmoid activation function. The outputs are  $F_{li}$ :

$$F_{li} = \text{Sigmoid}(\tilde{f}_{pw}(F_{d1} + F_{d2})). \quad (8)$$

Finally, as shown in Fig. 2(a), within the low-frequency path, we design the CFEM to capture the output features of the LFGE. Specifically, given the input  $F_{li}$ , a PW Conv  $\tilde{f}_{pw}$  followed by instance normalization IN is applied to extract and normalize local representations. A residual connection then integrates the processed features with the original input, thereby mitigating over-smoothing. This process is formalized as:

$$\hat{F}_{li} = \tilde{f}_{pw}(\text{IN}(\tilde{f}_{pw}(F_{li}))) + F_{li}. \quad (9)$$

where  $\hat{F}_{li}$  represents the output of the low-frequency path.

**High Frequency Path:** To enhance the capability of high-frequency features in modeling multi-directional textures and edges while effectively suppressing noise, as illustrated in Fig. 2(a), we propose a joint module comprising the HFAM and DFEM as the high-frequency path. In this path, HFAM helps adaptively re-weight high-frequency features to emphasize important edges and textures, while suppressing noise. Meanwhile, DFEM further refines these features by extracting fine-grained edge and texture information, allowing for enhanced feature representation. Specifically, within HFAM, we first concatenate the high-frequency features  $\{HL, HH, LH\}$  along the channel dimension  $z = \text{Concat}(LH, HH, HL)$ . Subsequently, adaptive average pooling is applied to derive the compressed feature  $z'$ , with the primary objective of reducing computational burden. Then, to achieve fine-grained modeling of multi-directional textures and edges while suppressing noise, we propose a residual decomposition mechanism to dynamically re-weight the high-frequency features, with the output feature  $F_{hi}$  formulated as:

$$F_{hi} = z' \times W_1 + (z - z') \times W_2, \quad (10)$$

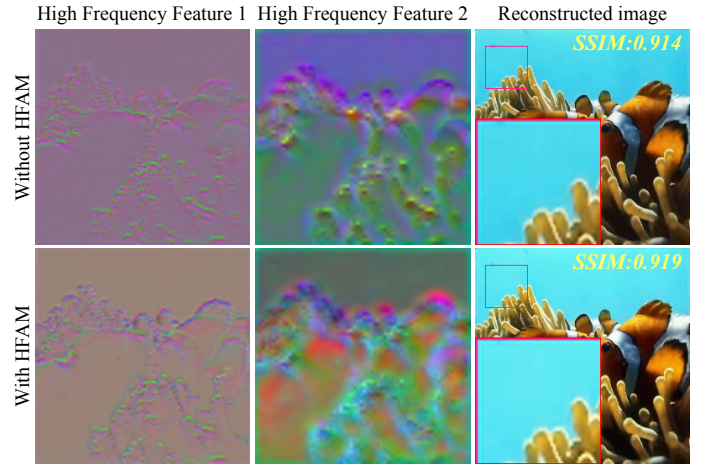


Fig. 5: High-frequency path feature output results on T-90 dataset. The first row represents the scenario without HFAM, while the second row depicts the full model within HFAM.

where  $W_1$  and  $W_2$  are learnable weight parameters that adaptively adjust the contributions of the compressed and residual features, respectively. The full model incorporating HFAM significantly enhances the representation of fine textures and edges in the high-frequency features compared to the model without HFAM. Therefore, we use the SSIM metric, which focuses on structural similarity, for evaluation. As shown in Fig. 5, the SSIM value improves from 0.914 to 0.919, and this enhancement contributes to clearer and more natural reconstructed outputs.

Following the modulation in HFAM, the DFEM is designed to further extract fine-grained edge and texture information from the high-frequency features. Given  $F_{hi}$  as input, a  $3 \times 3$  convolution is first applied to capture local structures, followed by average pooling over multiple scales  $S = \{16 \times 16, 32 \times 32, 64 \times 64\}$  to obtain multi-scale contextual information. The resulting features are then refined via PW Conv  $\tilde{f}_{pw}$ , upsampled to the original resolution, and integrated with the input through a residual connection. This process is formalized as

$$\hat{F}_{hi} = \text{Up}(\tilde{f}_{pw}(\text{AvgPool}_S(\text{Conv3}(F_{hi})))) + F_{hi}, \quad (11)$$

where Up denotes the upsampling operation, and  $\hat{F}_{hi}$  represents the output of the high-frequency path.

## B. Dual-Path Decoder with Global Fusion

To fully achieve interaction and fusion of low- and high-frequency features, as shown in Fig. 2(b), we propose a dual upsampling module, termed WaveUp. The module is composed of a DUIM, built upon PixelShuffle and IDWT to interact low- and high-frequency features, and a MCFN to achieve global feature integration. Specifically, in DUIM, the input low-frequency feature  $\hat{F}_{li}$  is first split along the channel dimension into low- and high-frequency features:

$$F_{ul}, F_{uh} = \text{ChannelSplit}(\hat{F}_{li}), \quad (12)$$

where  $F_{ul}$  and  $F_{uh}$  denote the low- and high-frequency feature, respectively. For low-frequency features  $F_{ul}$ , we apply



a PW Conv  $\tilde{f}_{pw}$  for channel expansion, followed by a PixelShuffle layer to achieve upsampling. Due to the invertibility of wavelet transforms [40], we employ IDWT for upsampling the high-frequency features  $F_{uh}$ . This process can be formulated as:

$$\begin{aligned}\hat{F}_{ul} &= \text{PixelShuffle}(\tilde{f}_{pw}(F_{ul})), \\ \hat{F}_{uh} &= \text{IDWT}(\text{Concat}(F_{uh}, F_{hi})),\end{aligned}\quad (13)$$

where  $\hat{F}_{ul}$  and  $\hat{F}_{uh}$  denote the upsampled outputs of the low- and high-frequency branches, respectively, with IDWT implemented using the Haar wavelet basis. Next, to get joint representation and interaction between low- and high-frequency features, the upsampled outputs are concatenated along the channel dimension to form  $F_u$ :

$$F_u = \text{Concat}(\hat{F}_{ul}, \hat{F}_{uh}). \quad (14)$$

This concatenated feature is then processed by a self-attention module  $\text{SA}(\cdot)$  to capture global dependencies and enable cross-frequency interactions:

$$F_{sa} = \text{SA}(F_u) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (15)$$

where  $Q$ ,  $K$ , and  $V$  are the query, key, and value matrices derived from  $F_u$  through linear transformations. To further promote cross-channel interaction between the low- and high-frequency features,  $F_{sa}$  is split along the channel dimension into two equal parts, which are subsequently combined via element-wise addition to yield the preliminarily fused feature:

$$\begin{aligned}F_{u1}, F_{u2} &= \text{ChannelSplit}(F_{sa}), \\ \hat{F}_u &= F_{u1} + F_{u2},\end{aligned}\quad (16)$$

where  $\hat{F}_u$  represents the output of the upsampling stage after initial fusion.

To further fusion the features  $\hat{F}_u$  and enhance multi-scale representations for global integration, we introduce the MCFN module [41]. First, this module applies a  $\tilde{f}_{pw}$  to linearly map and expand the channel dimension of  $\hat{F}_u$ , thereby aggregating low-level features while preparing for subsequent multi-branch processing. The resulting feature is then evenly split along the channel dimension into four branches:

$$F_{mi} = \text{ChannelSplit}(\tilde{f}_{pw}(\hat{F}_u)), \quad i = 0, 1, 2, 3. \quad (17)$$

The branches  $i = 1, 2, 3$  undergo convolutional operations with varying kernel sizes  $\tilde{f}_K$ , where  $K = \{3 \times 3, 5 \times 5, 7 \times 7\}$ , to capture contextual information across diverse receptive fields, while the branch  $i = 0$  remains unaltered to retain the original feature details. The outputs from all branches are subsequently concatenated along the channel dimension to form a comprehensive multi-scale representation:

$$F_c = \text{Concat}(\tilde{f}_K(F_{mi})), \quad i = 0, 1, 2, 3. \quad (18)$$

To incorporate non-linearity and achieve effective channel-wise fusion of the multi-scale features,  $F_c$  is processed through a GELU activation function, followed by a PW Conv  $\tilde{f}_{pw}$ , yielding the final fused output:

$$\hat{F}_0 = \tilde{f}_{pw}(\text{GELU}(F_c)). \quad (19)$$

As mentioned above, the full model incorporating MCFN outperforms the scenario without MCFN in feature fusion. Therefore, as shown in Fig. 6, we used the more comprehensive LPIPS metric for evaluation. The full model incorporating MCFN achieved a lower value of 0.0692, resulting in better visual quality.

In the overall decoder architecture, the WaveUp module is applied sequentially at multiple scales, with an adaptive mixing operation inserted between them. This mix operation dynamically weighs and combines low-frequency features based on learned parameters, enabling more effective fusion across resolutions. To stabilize training and preserve fine-grained details from the input, a residual connection is integrated, mitigating potential vanishing gradient issues while ensuring the retention of low-level information. The resulting features are then processed through a final PW Conv to generate the clear image.

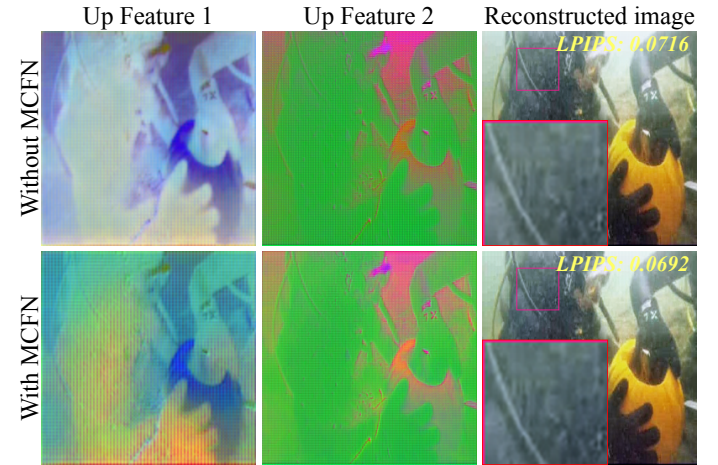


Fig. 6: Upsampling output features and results on T-90 dataset. The first row represents the scenario without MCFN, while the second row depicts the full model within MCFN.

### C. Loss Function

To achieve both color restoration and detail enhancement in underwater images, we design a joint loss function that comprehensively considers pixel consistency, structural preservation, and frequency-domain constraints. The overall loss is defined as:

$$\mathcal{L}_{total} = w_1 \mathcal{L}_c + w_2 \mathcal{L}_s + w_3 \mathcal{L}_f, \quad (20)$$

where  $w_1, w_2, w_3$  denote the weighting coefficients of the three loss terms. These weights are determined through random search method [42] on the validation set and set to  $w_1 = 1$ ,  $w_2 = 0.2$ , and  $w_3 = 0.4$ , which achieves a good balance between spatial consistency and frequency fidelity.

The Charbonnier loss is employed to measure the pixel-level discrepancy between the restored image  $\hat{I}$  and the ground truth  $G$ , as illustrated in Fig. 7. The formulation is:

$$\mathcal{L}_c = \frac{1}{N} \sum_{i=1}^N \sqrt{(\hat{I}_i - G_i)^2 + \epsilon^2}, \quad (21)$$



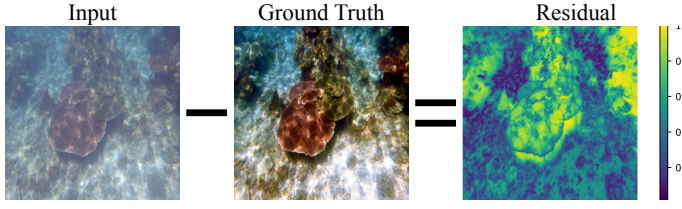


Fig. 7: Illustration of the Charbonnier loss. It computes the discrepancy between the input image pixels and the ground-truth pixels, measuring the pixel-wise difference between the restored image  $\hat{I}$  and the ground truth  $G$ .

where  $\epsilon = 10^{-6}$  is introduced to avoid numerical instability. As a smooth approximation of the  $L_1$  loss, the Charbonnier loss provides more stable gradients during training, effectively capturing pixel-wise consistency.

To preserve edges and texture structures, we introduce a structural similarity (SSIM)-based loss:

$$\mathcal{L}_s(\hat{I}, G) = 1 - \frac{(2\mu_{\hat{I}}\mu_G + C_1)(2\sigma_{\hat{I}G} + C_2)}{(\mu_{\hat{I}}^2 + \mu_G^2 + C_1)(\sigma_{\hat{I}}^2 + \sigma_G^2 + C_2)}, \quad (22)$$

where  $\mu$  and  $\sigma$  denote the mean and variance,  $\sigma_{\hat{I}G}$  is the covariance, and  $C_1, C_2$  are constants for stabilization. This term ensures consistency in luminance, contrast, and structural information during reconstruction.

Furthermore, as discussed in the previous research motivation and illustrated in Fig. 8, we first employ the FFT to convert both the input image  $I$  and the ground-truth image  $G$  from the spatial domain to the frequency domain. We then extract their corresponding amplitude and phase features. Specifically, we combine the amplitude of the input image with the phase of the ground truth, and conversely, the amplitude of the ground truth with the phase of the input image. Finally, by applying the IFFT, both combinations are transformed back into the spatial domain to reconstruct images. As shown in Fig. 8, the reconstructed images exhibit significant variations in tone and brightness compared with the original images, while the content and structure remain unchanged. This observation suggests that the brightness and structural characteristics of underwater images can be decomposed in the frequency domain to a certain extent.

Motivated by the above analysis, and to better guide the network in learning the brightness and structure information from ground truth images, we design a frequency-domain loss function that enforces consistency between the generated and ground-truth images in the frequency domain. The proposed frequency-domain loss is mathematically defined as:

$$\mathcal{L}_f = \|\mathcal{F}(\hat{I}) - \mathcal{F}(G)\|_1, \quad (23)$$

where  $\mathcal{F}(\cdot)$  denotes the Fourier transform and  $\|\cdot\|_1$  is the  $L_1$ -norm.

#### IV. EXPERIMENTS

In this section, we outline the experimental setup, present a frequency-domain analysis of DPF-Net, compare it with state-of-the-art methods, and conduct ablation studies to verify the

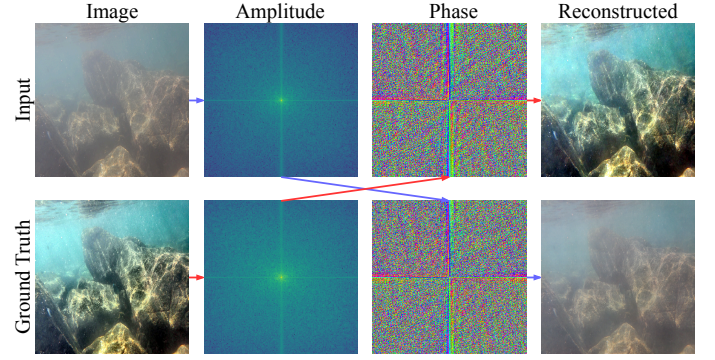


Fig. 8: Illustration of the frequency-domain loss. By exchanging amplitude and phase between the input and the ground truth image, the analysis demonstrates that color and brightness are mainly represented by amplitude, while content and structure are primarily associated with phase.

contribution of each component. We further evaluate DPF-Net on downstream applications, followed by discussions, limitations, and future directions.

##### A. Basic experimental details

1) *Datasets*: As shown in Tab. I, the datasets used in this study include both paired and unpaired datasets. For paired training datasets, we employed UIEB-T [43] and LSUI-T [44]. The unpaired test datasets consist of U45 [45], Challenge-60 [43], LOL-low [46], RRSID-thin-fog [47], Dense-Haze [48], and UIQS-D [49]. Among these, LOL-low [46], RRSID-thin-fog [47], and Dense-Haze [48] are used to evaluate the generalization of our proposed method on similar low-level vision tasks, while UIQS-D [49] is used to assess the performance of our method on keypoint matching tasks. The UIEB dataset consists of 950 real-world underwater images, of which 890 are paired images. We use 800 paired images for training (UIEB-T) and the remaining images for testing (T-90). The remaining 60 images constitute the Challenge-60 subset. The LSUI dataset is a large-scale underwater image dataset containing 4279 image pairs. We use 3825 image pairs for training (LSUI-T), and the remaining 427 for testing (LSUI-TE). The U45 dataset includes 45 real underwater images, presenting challenges such as color casts, low contrast, and haze effects. Further details of the remaining datasets are provided in Tab. I.

2) *Implementation Details*: All experiments were implemented in PyTorch 2.4 with CUDA 12.1 acceleration and executed on a workstation equipped with one NVIDIA RTX 4090 GPUs. The network parameters were optimized using AdamW, with momentum terms set to  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ . The initial learning rate was configured as  $2 \times 10^{-4}$  and dynamically adjusted using the cosine annealing scheduler [51], which gradually decayed the learning rate to  $1 \times 10^{-6}$  during training. For fair and consistent evaluation, all input images were resized to  $256 \times 256$  pixels, and the mini-batch size was fixed at 8. The network was trained for 200 epochs on the UIEB-T dataset and 400 epochs on the LSUI-T

TABLE I: Summary of Datasets Used in This Paper

| Setting | Train/Test | Datasets                | Type     | Role in this Paper                                                                     |
|---------|------------|-------------------------|----------|----------------------------------------------------------------------------------------|
| Set 1   | Training   | UIEB-T [43]             | Paired   | Small-scale real underwater scene training                                             |
|         |            | T-90 [43]               | Unpaired | Small-scale real underwater scene testing                                              |
|         | Testing    | U45 [45]                | Unpaired | Generalization evaluation under underwater color shift, low contrast, and haze effects |
|         |            | Challenge-60 [43]       | Unpaired | Generalization evaluation of challenging underwater scenes                             |
| Set 3   | Testing    | LOL-test-low [46]       | Unpaired | Cross-task generalization evaluation of low-light scenes                               |
|         |            | RRSHID-thin-fog [47]    | Unpaired | Cross-task generalization evaluation of remote sensing dehazing scenes                 |
|         |            | Dense-Haze [48]         | Unpaired | Cross-task generalization evaluation of indoor denoising scenes                        |
|         |            | UIQS-D [49]             | Unpaired | Keypoint feature matching performance evaluation                                       |
| Set 3   | Training   | LSUI-T [44]             | Paired   | Large-scale real underwater scene training                                             |
|         | Testing    | LSUI-TE [44]            | Paired   | Large-scale real underwater scene evaluation                                           |
| Set 4   | Training   | LOL-485 [46]            | Paired   | Model training in real low-light scenes                                                |
|         | Testing    | LOL-15 [46]             | Paired   | Evaluation in low-light scenes                                                         |
|         | Training   | Haze1k-moderate-T [50]  | Paired   | Model training in remote sensing dehazing scenes                                       |
|         | Testing    | Haze1k-moderate-TE [50] | Paired   | Testing on remote sensing dehazing scenes                                              |

dataset, ensuring convergence on both synthetic and real-world underwater image distributions.

3) *Baseline*: We compared ours with twenty-two methods. The traditional methods considered include IBLA [3], DCP [4], CLAHE [5], MIP [6], ICM [8], UDCP [9], UCM [10], and HE [7]. The deep learning-based approaches comprise WaterNet [12], UIEC<sup>2</sup>-Net [33], SCNet [29], GUPDM [37], Shallow-UWnet [27], UWCNN [36], CCMSRNet [34], USUIR [24], SFGNet [28], UGAN [13], FIVE A+ [11], PUIE-Net [26], GuidedHySensUIR [19], and MB-TaylorFormer v2 [18]. In addition, for comparison in low-level vision tasks, we selected three representative methods: LIME [52], URetinex [53], and BMFH-Net [54].

4) *Evaluation Metrics*: For quantitative evaluation, we adopted two complementary assessment paradigms: full-reference and non-reference evaluation. The full-reference evaluation was carried out on the T-90 and LSUI-TE datasets using widely recognized image quality metrics, including PSNR [55], SSIM [55], LPIPS [56], UCIQE [57], and UIQM [58]. The non-reference evaluation was performed on the Challenge-60 and U45 datasets, focusing on UCIQE and UIQM to assess perceptual quality in the absence of ground-truth images. Furthermore, to evaluate the frequency-domain functionality of the proposed dual-path architecture, we selected the frequency-domain energy ratio ( $R_E$ ) [59] and phase consistency metrics ( $C_\phi$ ) [60] for quantitative analysis. In addition, to evaluate the computational cost of the model, we use Multiply-Accumulate Operations (MACs), the number of parameters (Param.), and the end-to-end time per image (E2E Time) as evaluation metrics.

## B. Dual-Path Feature Maps and Frequency Domains Analysis

1) *Qualitative Visualizations Analysis*: As shown in Fig. 9, the qualitative results reveal the distinct roles of the dual-path design. The low-frequency path (Low-Frequency Features 1–3) generates smooth, stable color feature maps, with magnitude

spectra energy concentrated in the central region and high-frequency components suppressed, confirming its focus on global information recovery, such as luminance and chroma, which helps correct overall color tone and illumination deviations. The high-frequency path (High-Frequency Features 1–2) captures prominent texture, edge, and detail information (e.g., fish contours or coral textures), exhibiting enhanced peripheral magnitude energy and directional phase patterns (e.g., grid or line patterns), indicating its responsibility for local structural restoration, thereby enhancing image sharpness and clarity. During the upsampling stage (Features 1–2), low- and high-frequency information is progressively integrated, leading to magnitude spectra that incorporate both central and peripheral energy components, and phase spectra that progressively align from simple patterns to approach the Ground Truth, confirming that the cross-frequency fusion stage achieves amplitude-phase coordination and ensures the Reconstructed approximates the ideal image in terms of color correction and detail enhancement. Through cross-dataset comparison, this design demonstrates strong generalization, with the low-frequency path effectively addressing color deviations and the high-frequency path tackling texture challenges, overall proving the model's robustness and innovative potential.

2) *Quantitative Metrics Analysis*: As shown in Fig. 10–11, quantitative results corroborate these observations. The energy ratio  $R_E$  increases along the low-frequency branch (from 295 at the input to 12194 at Low-Frequency Feature 3), while the high-frequency branch peaks at High-Frequency Feature 2 (30699). The final reconstructed output achieves  $R_E = 159$ , closely matching the ground truth value of 162, validating balanced spectral energy recovery. For phase consistency, intermediate features often exhibit near-zero or negative  $C_\phi$ , indicating incomplete phase alignment, whereas the final output reaches  $C_\phi = 0.872$ , a significant improvement over the input, demonstrating that phase alignment is primarily achieved during the fusion stage. The LSUI-TE dataset exhibits similar trends, with both  $R_E$  and  $C_\phi$  approaching



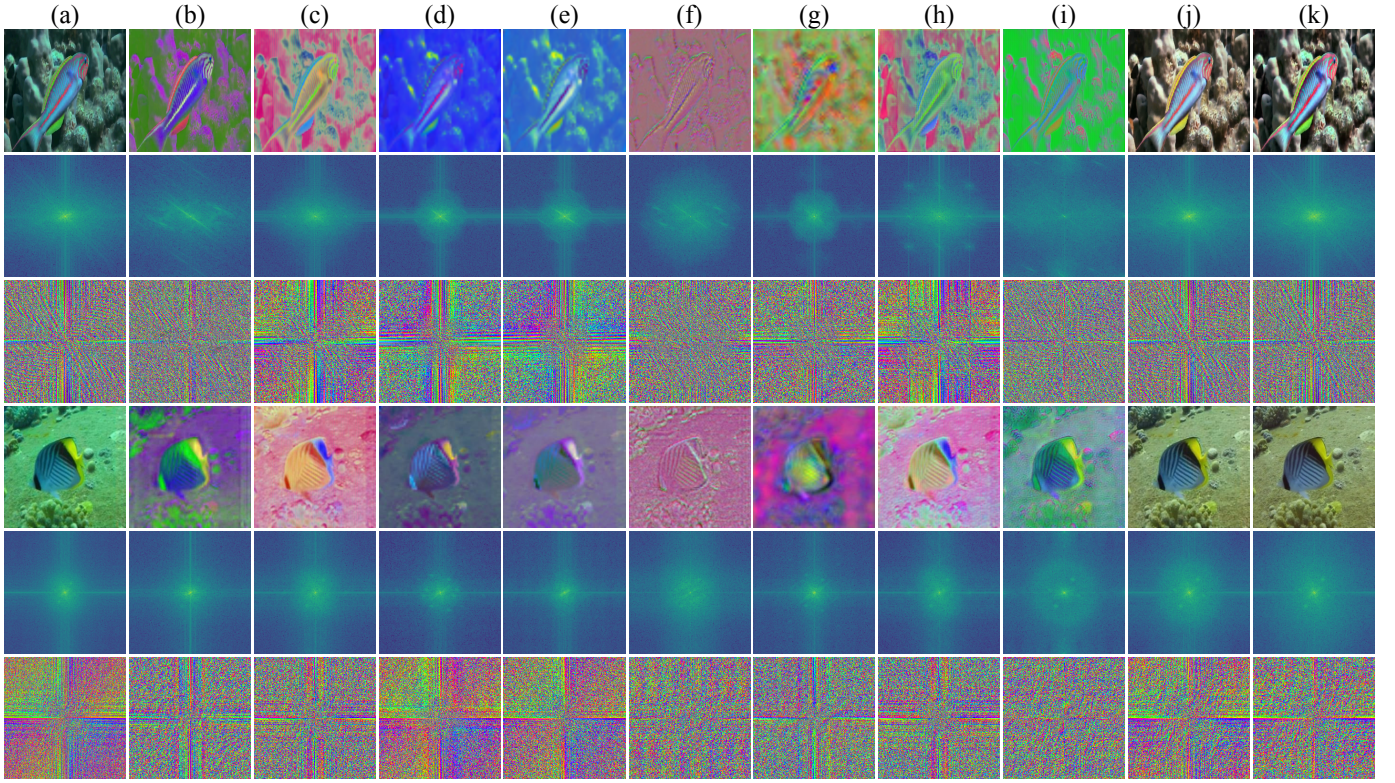


Fig. 9: Visualization of dual-path feature maps along with their corresponding magnitude and phase spectra. Represented from left to right: (a) Input, (b) Initial Feature, (c) Low-Frequency Feature 1, (d) Low-Frequency Feature 2, (e) Low-Frequency Feature 3, (f) High-Frequency Feature 1, (g) High-Frequency Feature 2, (h) Upsampling Feature 1, (i) Upsampling Feature 2, (j) Reconstructed(Our), (k) GroundTruth. The first row presents an example from T-90 dataset, while the second and third rows display its respective magnitude and phase spectra. The fourth row shows an example from LSUI-TE dataset, with the fifth and sixth rows illustrating its corresponding magnitude and phase spectra.

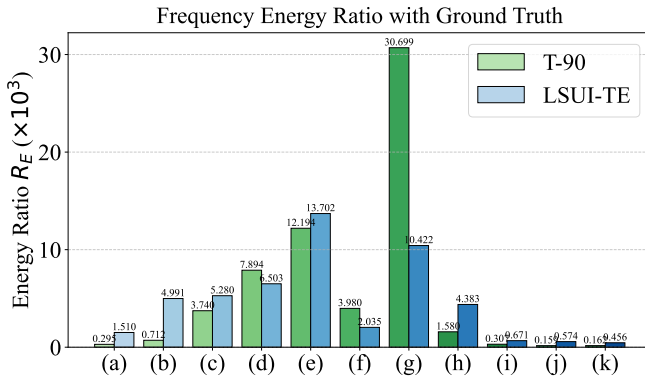


Fig. 10: Frequency energy ratio of the dual-path model on T-90 and LSUI-TE dataset. (a)-(k) represent the same as in Figure 9.

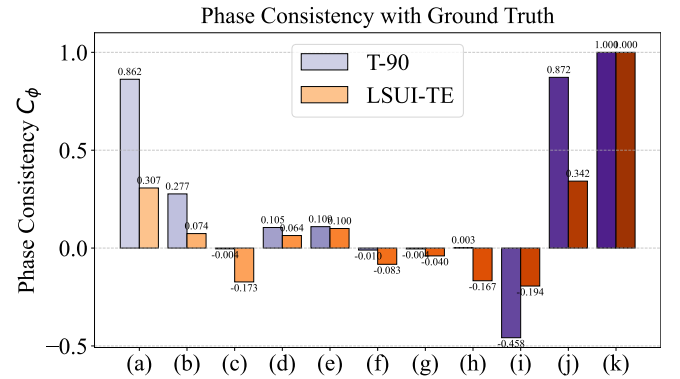


Fig. 11: Phase consistency with respect to the ground truth on T-90 and LSUI-TE dataset. (a)-(k) represent the same as in Figure 9.

ground truth values, further confirming the efficacy of the dual-path design and cross-frequency fusion.

### C. Comparison with SOTA UIE Methods

1) *Quantitative Evaluation*: This section presents a quantitative comparison and analysis of twenty-two methods on both reference and no-reference datasets.

Tab. II presents the full-reference evaluation results of twenty-two UIE methods on the T-90 and LSUI-TE datasets. Our DPF-Net consistently achieves top performance across key metrics, including PSNR, SSIM, LPIPS, and UIQM. Specifically, on T-90 dataset, DPF-Net achieves the highest PSNR of 24.20, outperforming MB-TaylorFormer v2 by 1.31%. In SSIM, our method achieves 0.919, surpassing GuidedHybSensUIR by 1.10%. For LPIPS, DPF-Net



TABLE II: Quantitative Results On the T-90 and LSUI-TE Datasets. Best results in red, second-best in blue, and the third-best are purple.

| Methods                     | T-90 [43] |       |        |        |        | LSUI-TE [44] |       |        |        |        |
|-----------------------------|-----------|-------|--------|--------|--------|--------------|-------|--------|--------|--------|
|                             | PSNR↑     | SSIM↑ | LPIPS↓ | UCIQE↑ | UIQM↑  | PSNR↑        | SSIM↑ | LPIPS↓ | UCIQE↑ | UIQM↑  |
| DCP [4]                     | 14.60     | 0.693 | 0.1915 | 0.4536 | 2.2569 | 15.71        | 0.685 | 0.2696 | 0.4488 | 2.1064 |
| CLAHE [5]                   | 19.22     | 0.865 | 0.1177 | 0.4184 | 3.2264 | 17.46        | 0.802 | 0.2376 | 0.4179 | 3.1649 |
| MIP [6]                     | 14.67     | 0.669 | 0.2110 | 0.4524 | 2.3057 | 15.65        | 0.663 | 0.2622 | 0.4532 | 2.2627 |
| ICM [8]                     | 17.56     | 0.807 | 0.1494 | 0.4148 | 2.7355 | 20.24        | 0.846 | 0.1781 | 0.4252 | 2.9204 |
| UDCP [9]                    | 15.55     | 0.720 | 0.1788 | 0.4394 | 2.3868 | 17.10        | 0.739 | 0.2473 | 0.4288 | 2.3600 |
| IBLA [3]                    | 15.92     | 0.659 | 0.2024 | 0.4734 | 2.0642 | 17.96        | 0.740 | 0.2326 | 0.4568 | 2.4456 |
| UCM [10]                    | 17.22     | 0.822 | 0.1740 | 0.4827 | 2.9744 | 19.93        | 0.846 | 0.2047 | 0.4640 | 3.0790 |
| HE [7]                      | 17.12     | 0.812 | 0.2090 | 0.5001 | 3.2291 | 14.85        | 0.701 | 0.3264 | 0.4993 | 3.3199 |
| WaterNet [12]               | 21.96     | 0.845 | 0.1495 | 0.4516 | 3.6751 | 20.58        | 0.813 | 0.1813 | 0.3883 | 3.0310 |
| UIEC <sup>2</sup> -Net [33] | 20.11     | 0.892 | 0.1114 | 0.4417 | 3.6636 | 23.48        | 0.903 | 0.0705 | 0.3977 | 3.6883 |
| SCNet [29]                  | 22.19     | 0.895 | 0.0933 | 0.4060 | 3.2674 | 25.27        | 0.883 | 0.0726 | 0.4082 | 3.1135 |
| GUPDM [37]                  | 21.69     | 0.853 | 0.1309 | 0.3656 | 3.1967 | 24.81        | 0.868 | 0.1221 | 0.3947 | 3.1987 |
| Shallow-UWnet [27]          | 19.21     | 0.845 | 0.1456 | 0.3925 | 3.7856 | 21.59        | 0.845 | 0.1099 | 0.3738 | 3.7372 |
| UWCNN [36]                  | 18.86     | 0.859 | 0.1222 | 0.3871 | 3.8554 | 22.89        | 0.886 | 0.1187 | 0.3937 | 3.8471 |
| CCMSRNet [34]               | 22.31     | 0.880 | 0.1054 | 0.4109 | 3.7147 | 27.42        | 0.904 | 0.0861 | 0.4190 | 3.6604 |
| USUIR [24]                  | 18.69     | 0.776 | 0.1755 | 0.3704 | 3.6190 | 18.34        | 0.778 | 0.2501 | 0.3964 | 3.4792 |
| SFGNet [28]                 | 21.49     | 0.896 | 0.1347 | 0.4428 | 3.0335 | 25.13        | 0.874 | 0.1278 | 0.4203 | 3.1059 |
| UGAN [13]                   | 22.60     | 0.847 | 0.1396 | 0.4495 | 3.4664 | 26.24        | 0.880 | 0.1444 | 0.4480 | 3.0886 |
| FIVE A+ [11]                | 22.71     | 0.904 | 0.0861 | 0.4514 | 3.9447 | 25.55        | 0.903 | 0.1039 | 0.4288 | 3.7766 |
| PUIE-Net [26]               | 19.48     | 0.822 | 0.1332 | 0.3389 | 3.7955 | 26.99        | 0.899 | 0.0575 | 0.4190 | 3.8694 |
| GuidedHybSensUIR [19]       | 23.36     | 0.909 | 0.0722 | 0.4393 | 3.8085 | 28.10        | 0.917 | 0.0455 | 0.4243 | 3.7116 |
| MB-TaylorFormer v2 [18]     | 23.89     | 0.908 | 0.0754 | 0.4706 | 4.1671 | 28.32        | 0.925 | 0.0605 | 0.4325 | 3.8359 |
| Ours (DPF-Net)-Lite         | 23.78     | 0.910 | 0.0684 | 0.4289 | 3.9448 | 27.86        | 0.921 | 0.0791 | 0.4214 | 3.6383 |
| Ours                        | 24.20     | 0.919 | 0.0675 | 0.4340 | 3.9908 | 29.54        | 0.927 | 0.0731 | 0.4264 | 3.8683 |

achieves 0.0675, demonstrating a 6.51% reduction compared to GuidedHybSensUIR, indicating better perceptual quality with lower perceptual distortion. Regarding UIQM, DPF-Net ranks second with a score of 3.8683, closely following MB-TaylorFormer v2, which achieves 4.1671. On the LSUI-TE dataset, DPF-Net achieves the highest PSNR of 29.54 and SSIM of 0.927, and ranks second in UIQM of 3.8683, outperforming MB-TaylorFormer v2 by 4.31% in PSNR and 0.22% in SSIM. Compared to the best-performing method in LPIPS and UIQM, our approach also exhibits minor gaps.

To evaluate robustness, we assessed the performance on the non-reference dataset, as shown in Tab. III. Specifically, all methods were trained on the UIEB-T dataset and then directly tested on the non-reference dataset. On the U45 dataset, DPF-Net achieves the highest UIQM score of 3.9284, and on the Challenge-60 dataset, it achieves the second-highest score of 3.8795. These results demonstrate its effectiveness in improving overall image quality, particularly in color restoration, clarity, and contrast enhancement.

Notably, to enhance the real-time practicality of DPF-Net, we propose a lightweight version, DPF-Net-Lite. We also evaluate the performance of DPF-Net-Lite across all datasets. DPF-Net-Lite achieves near-optimal results, with PSNR, SSIM, and UIQM scores slightly lower than the full DPF-Net but still outperforming many other methods. In particular, on the T-90 dataset, DPF-Net-Lite achieves a PSNR of 23.78, SSIM of 0.910, and LPIPS of 0.0684, which are very close to the full version, confirming the effectiveness of the lightweight design.

Additionally, although HE and UCM achieved the highest and second-highest scores in the UCIQE metric on all datasets, respectively, they exhibited poor visual results, as shown in Fig. 12 and 14. This discrepancy highlights a key limitation of

TABLE III: Quantitative Results On The U45 and Challenge-60 Datasets. The best results on all datasets are shown in red, and the second-best are blue.

| Methods                     | U45 [45] |        | Challenge-60 [43] |        |
|-----------------------------|----------|--------|-------------------|--------|
|                             | UCIQE ↑  | UIQM ↑ | UCIQE ↑           | UIQM ↑ |
| DCP [4]                     | 0.4622   | 2.258  | 0.4367            | 1.4776 |
| CLAHE [5]                   | 0.4203   | 3.0829 | 0.3920            | 2.5819 |
| MIP [6]                     | 0.4377   | 2.1890 | 0.4031            | 1.6259 |
| ICM [8]                     | 0.4459   | 2.8486 | 0.3917            | 2.2142 |
| UDCP [9]                    | 0.4291   | 2.5016 | 0.4007            | 1.8867 |
| IBLA [3]                    | 0.4816   | 1.7088 | 0.4517            | 1.7793 |
| UCM [10]                    | 0.4973   | 2.9000 | 0.4815            | 2.2189 |
| HE [7]                      | 0.4994   | 3.0475 | 0.4934            | 2.3089 |
| WaterNet [12]               | 0.4297   | 3.6633 | 0.3722            | 3.8638 |
| UIEC <sup>2</sup> -Net [33] | 0.4427   | 3.6442 | 0.4244            | 3.7985 |
| SCNet [29]                  | 0.3712   | 2.8560 | 0.3677            | 3.8531 |
| GUPDM [37]                  | 0.2584   | 3.6766 | 0.2547            | 3.8003 |
| Shallow-UWnet [27]          | 0.3662   | 3.7063 | 0.3503            | 3.7394 |
| UWCNN [36]                  | 0.3615   | 3.7234 | 0.3525            | 3.8178 |
| CCMSRNet [34]               | 0.3362   | 3.3475 | 0.3848            | 3.7504 |
| USUIR [24]                  | 0.3608   | 3.5556 | 0.3698            | 3.6426 |
| SFGNet [28]                 | 0.4554   | 3.5307 | 0.4168            | 1.0787 |
| UGAN [13]                   | 0.4423   | 3.5060 | 0.4320            | 3.3858 |
| FIVE A+ [11]                | 0.4437   | 3.8989 | 0.4369            | 4.1746 |
| PUIE-Net [26]               | 0.3241   | 3.7595 | 0.3117            | 3.8220 |
| GuidedHybSensUIR [19]       | 0.4311   | 3.7706 | 0.4107            | 3.8748 |
| MB-TaylorFormer v2 [18]     | 0.4745   | 2.7584 | 0.2962            | 3.1804 |
| Ours (DPF-Net)-Lite         | 0.4243   | 3.7635 | 0.3871            | 3.7648 |
| Ours                        | 0.4202   | 3.9284 | 0.4317            | 3.8795 |

UCIQE, which, as noted in [61] and [62], relies on simplified human visual system features and linear models, failing to capture perceptual nonlinearity. Consequently, UCIQE struggles to accurately assess the perceptual quality of underwater images. Our observations further corroborate this limitation,



Fig. 12: Visualization examples on T-90 dataset, where the first two rows are from T-90 dataset and the latter two rows are from LSUI-TE dataset. (a) Raw, (b) DCP, (c) CLAHE, (d) MIP, (e) ICM, (f) UDCP, (g) IBLA, (h) UCM, (i) HE, (j) WaterNet, (k) UIEC<sup>2</sup>-Net, (l) SCNet, (m) GUPDM, (n) Shallow-UWnet, (o) UWCNN, (p) CCMsRNet, (q) USUIR, (r) SFGNet, (s) UGAN, (t) FIVE A+, (u) PUIE-Net, (v) GuidedHybSensUIR, (w) MB-TaylorFormer v2, (x) Ours(DPF-Net)-Lite, (y) Ours, (z) Ground Truth.

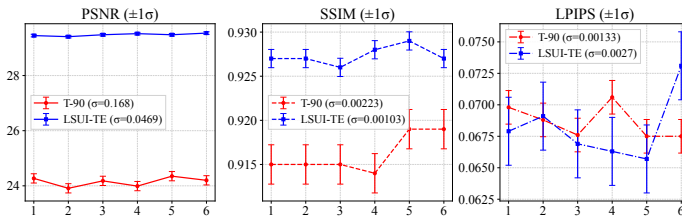


Fig. 13: Robustness to random seeds on T-90 and LSUI-TE.

emphasizing that UCIQE's reliance on these simplified metrics does not align well with the complex perceptual characteristics of enhanced underwater images. In contrast, our method, DPF-Net, prioritizes structural preservation and perceptual realism, achieving superior visual results despite the lower UCIQE score.

Overall, these results demonstrate that our method consistently enhances underwater images with superior structural preservation and perceptual quality across diverse datasets.

2) *Random-seed robustness*: To quantify the sensitivity of our results to random initialization and training stochasticity,

we conduct experiments on T-90 and LSUI-TE while keeping all training settings fixed. We vary only the random seed and repeat the full training pipeline six times. We report mean  $\pm$  sample standard deviation (delta degrees of freedom,  $ddof = 1$ ) across the six independent runs, and Fig.13 visualizes the per-run results with  $\pm 1\sigma$  error bars. Overall, the variance is small: on T-90, PSNR is  $24.15 \pm 0.17$ , SSIM is  $0.916 \pm 0.002$ , and LPIPS is  $0.0686 \pm 0.0013$ ; on LSUI-TE, PSNR is  $29.48 \pm 0.05$ , SSIM is  $0.927 \pm 0.0010$ , and LPIPS is  $0.0682 \pm 0.0027$ . These results indicate that our method is stable across different random seeds, and the reported improvements are not attributable to favorable initialization or training noise.

3) *Qualitative Evaluation*: This part presents a comprehensive visual comparison across both reference and non-reference datasets with twenty-two methods.

Fig. 12 shows four common underwater scenarios: low-light deep-sea conditions, detail blur caused by suspended particles, deep-sea green color cast, and shallow water scenes. Specifically, traditional methods often introduce color distortions and suffer from issues in color fidelity such as IBLA, UCM and HE. Deep learning-based methods generally outperform





Fig. 14: Non-reference image visualization, with the top two rows showing samples from U45 and the bottom two rows from Challenge-60. (a) Raw, (b) DCP, (c) CLAHE, (d) MIP, (e) ICM, (f) UDCP, (g) IBLA, (h) UCM, (i) HE, (j) WaterNet, (k) UIEC<sup>2</sup>-Net, (l) SCNet, (m) GUPDM, (n) Shallow-UWnet, (o) UWCNN, (p) CCMSRNet, (q) USUIR, (r) SFGNet, (s) UGAN, (t) FIVE A+, (u) PUIE-Net, (v) GuidedHybSensUIR, (w) MB-TaylorFormer v2, (x) Ours(DPF-Net)-Lite, (y) Ours, (z) No Reference

traditional techniques but still have limitations. For instance, WaterNet performs poorly in deep-sea green color cast correction, while UWCNN and Shallow-UWnet struggle with the removal of suspended particles. USUIR, FIVE A+, and PUIE-Net occasionally introduce color distortions in detail blur scenarios. Except for UIEC<sup>2</sup>-Net, CCMSRNet, PUIE-Net, and GuidedHybSensUIR, the remaining methods fail to enhance images under low-light conditions in deep-sea scenarios. Among these, UIEC<sup>2</sup>-Net and GuidedHybSensUIR introduce color distortions, while PUIE-Net suffers from overexposure issues. In contrast, although our DPF-Net-Lite performs moderately in low-light deep-sea environments, our DPF-Net method effectively mitigates the aforementioned problems, producing images with overall quality closest to the reference images, while minimizing color distortions and preserving fine details during low-light enhancement.

As shown in Fig. 14, we selected four representative common underwater image scenarios, including green color cast, detail blur, blue color cast, and cyan color cast. Fig. 14 presents the visualization results of the generalization tests for various methods. Specifically, many traditional methods and deep learning approaches perform poorly on non-

reference datasets due to the diverse and complex distortions in the dataset. For instance, IBLA, UCM, and HE often introduce color distortions. UIEC<sup>2</sup>-Net, GUPDM, USUIR, and PUIE-Net struggle with detail blur removal. Shallow-UWnet, SFGNet, CCMSRNet, and USUIR exhibit poor performance on green color cast. SFGNet, MB-TaylorFormer, and Shallow-UWnet perform moderately on blue color cast. UIEC<sup>2</sup>-Net and Shallow-UWnet also struggle with cyan color cast. Although FIVE A+ and GuidedHybSensUIR show better performance in addressing green, blue, and cyan color casts, their performance in detail blur removal is mediocre. Our DPF-Net-Lite achieves good results in blue and cyan color cast removal, but its performance in green color cast and detail blur removal is average. In contrast, our DPF-Net demonstrates the best overall performance across all visual results.

#### D. Ablation Study

To comprehensively validate the effectiveness of each component in the proposed DPF-Net, all ablation experiments were conducted using the T-90 dataset for testing.

1) *Effects of the Dual-Path Design:* As shown in Tab. IV, we evaluated the effectiveness of the dual-path design



by comparing it with the full DPF-Net model. Removing the Low-frequency Path resulted in a decrease of 0.92 in PSNR, a decrease of 0.011 in SSIM, and an increase of 0.0061 in LPIPS. Similarly, removing the High-frequency Path led to a decrease of 0.66 in PSNR, a decrease of 0.018 in SSIM, and an increase of 0.0096 in LPIPS. The Dual-Path Fusion plays a crucial role in the integration of high-frequency and low-frequency features. Removing the Dual-Path Fusion module resulted in a significant deterioration across all metrics. Specifically, PSNR decreased by 1.75, SSIM decreased by 0.032, and LPIPS increased by 0.0217. Clearly, the clarity progressively improves as follows: Fig. 15 (b), Fig. 15 (a), and Fig. 15 (c), compared to Fig. 15 (o).

2) *Effects of DWT and FFT*: To assess the multi-scale decomposition of DWT, we replace DWT/IDWT with convolutional down/up-sampling. As shown in Tab. IV and Fig. 15(d), this replacement causes a clear performance drop, confirming that subband decoupling is crucial for separating global structures and local details. We then remove the FFT/IFFT-based refinement in the low-frequency path. The metric drop is relatively limited, suggesting that DWT captures most of the gains; nevertheless, amplitude-phase refinement provides a complementary constraint that stabilizes global appearance (e.g., overall tone and illumination), as suggested by Fig. 15(e). Overall, DWT enables localized multi-scale detail restoration, while FFT-based amplitude-phase modeling encourages low-frequency global color/illumination consistency; their combination yields a better balance than using wavelet decomposition alone.

3) *Effects of Different Components*: Since the LFGA and CFEM modules together form the Low-frequency Path, the HFAM and DFEM modules form the High-frequency Path, and the DUIM and MCFN modules constitute the Dual-Path Fusion module, removing these submodules individually has a minimal effect on DPF-Net, as shown in Tab. IV. The removal of the MCFN module has the least impact, as the DUIM module can effectively implement the initial interaction between high-frequency and low-frequency features. In contrast, the most significant impact occurs when the DUIM module is removed, as its removal prevents effective interaction and supplementation of local features between the low-frequency and high-frequency paths, thereby affecting the overall restoration performance. Accordingly, Fig. 15(k) is closest to Fig. 15(o).

4) *Effects of Each Loss Term*: As shown in Tab. IV, removing any loss term leads to a decrease in performance for most metrics, indicating the overall effectiveness of these loss terms. This aligns with the specific aspects each loss term

TABLE IV: Ablation Study on the T-90 Dataset. The Low-frequency Path includes the LFGA and CFEM modules, the High-frequency Path includes the HFAM and DFEM modules, and the Path-Fusion contains the DUIM and MCFN modules.

| Methods                  | Metrics         |                 |                    |
|--------------------------|-----------------|-----------------|--------------------|
|                          | PSNR $\uparrow$ | SSIM $\uparrow$ | LPIPS $\downarrow$ |
| -w/o Low-frequency Path  | 23.28           | 0.908           | 0.0736             |
| -w/o High-frequency Path | 23.54           | 0.901           | 0.0771             |
| -w/o Dath-Path Fusion    | 22.45           | 0.887           | 0.0892             |
| -w/o DWT/IDWT            | 23.41           | 0.906           | 0.0728             |
| -w/o FFT/IFFT (APRM)     | 23.88           | 0.917           | 0.0692             |
| -w/o LFGA                | 23.89           | 0.914           | 0.0691             |
| -w/o CFEM                | 23.66           | 0.910           | 0.0771             |
| -w/o HFAM                | 23.94           | 0.914           | 0.0698             |
| -w/o DFEM                | 24.04           | 0.911           | 0.0707             |
| -w/o DUIM                | 23.54           | 0.908           | 0.0734             |
| -w/o MCFN                | 24.09           | 0.912           | 0.0716             |
| -w/o $\mathcal{L}_c$     | 23.73           | 0.917           | 0.0695             |
| -w/o $\mathcal{L}_s$     | 23.94           | 0.905           | 0.0708             |
| -w/o $\mathcal{L}_f$     | 23.77           | 0.918           | 0.0713             |
| Full Model (DPF-Net)     | <b>24.20</b>    | <b>0.919</b>    | <b>0.0675</b>      |

targets. Removing the pixel domain constraint loss  $\mathcal{L}_c$  results in the most significant decrease in PSNR, while removing the structural similarity loss  $\mathcal{L}_s$  leads to the most noticeable drop in SSIM. Removing the frequency domain constraint loss  $\mathcal{L}_f$  leads to the most significant increase in LPIPS. The visual perceptual clarity of three loss term decreases in the following order: Fig. 15(l), Fig. 15(m), and Fig. 15(n).

### E. Applications

In this section, we analyze the application potential of the DPF-Net method in downstream tasks, focusing on keypoint matching, keypoint detection, edge detection, and object detection tasks.

By analyzing previous comparative experimental results, we selected the following well-performing image enhancement methods as baseline methods for comparison: PUIE-Net, MB-TaylorFormer v2, UCM, FIVE A+, HE, and GuidedHybSenSUIR. We then used the SIFT [63] operator to compare the number of valid keypoint matches in images processed by our method and these baselines on the paired UIQS-D dataset. The qualitative comparison results, as shown in Fig. 16, clearly

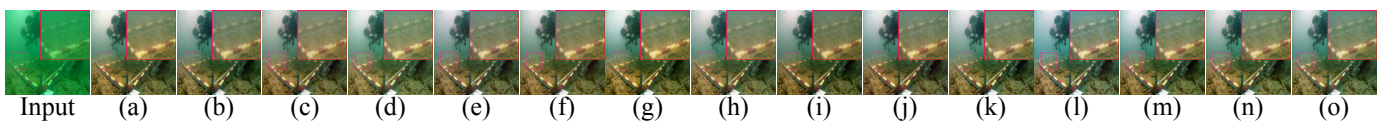


Fig. 15: Visualization of ablation study results on T-90 dataset. (a)-w/o low-frequency path (LFGA-CFEM), (b) -w/o the high-frequency path (HFAM-DFEM), (c) -w/o the dual-path fusion (DUIM-MCFN), (d) -w/o FFT/IFFT(APRM), (e) -w/o DWT/IDWT, (f) -w/o LFGA, (g) -w/o CFEM, (h) -w/o HFAM, (i) -w/o DFEM, (j) -w/o DUIM, (k) -w/o MCFN, (l) -w/o  $\mathcal{L}_c$ , (m) -w/o  $\mathcal{L}_s$ , (n) -w/o  $\mathcal{L}_f$ , (o) Ours (DPF-Net)

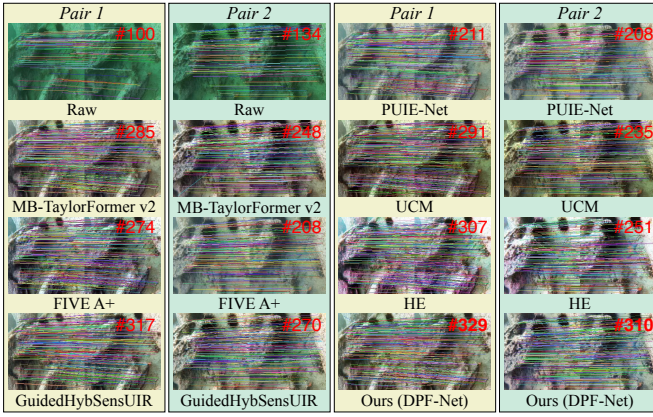


Fig. 16: Visualization results of SOTA methods for SIFT keypoint matching on Pair 1 and Pair 2 from the UIQS-D dataset.

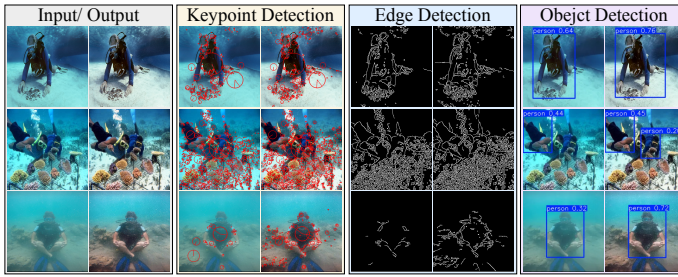


Fig. 17: Results of downstream task application tests for the DPF-Net method. From Left to Right: Distorted and Enhanced Images, Keypoint Detection, Edge Detection, and Object Detection Results on T-90 Dataset.

demonstrate that our method has the most keypoint matches in both Pair 1 and Pair 2. Fig. 16 also show that our method achieves the highest number of matched keypoints in both Pair 1 and Pair 2 images.

As shown in Fig. 17, we evaluated the enhanced images on three representative downstream tasks using the T-90 dataset: SIFT-based keypoint detection, Canny-based edge detection [64], and YOLO-based object recognition [65]. In the keypoint detection task, images enhanced by our method enable the detection of more keypoints. Similarly, in the edge detection task, our results reveal more edges. Furthermore, on the YOLO platform, our enhanced images show significantly higher human recognition accuracy compared to degraded images. Evidently, images enhanced by the proposed DPF-Net preserve features more effectively and are more readily detected in downstream applications, highlighting their strong potential for practical deployment.

## F. Discussion

This section discusses the impact of loss-function weight combinations on model enhancement performance, the trade-off between computational efficiency and performance, the generalization of DPF-Net in low-level vision tasks, and its scalability to low-level tasks, as well as the efficiency-performance trade-offs of its lightweight variant.

1) *Weight Combinations Vs. Model Performance:* In this study, the weight combination  $c1 : (w_1 = 1.0, w_2 = 0.2, w_3 = 0.4)$  was identified as the optimal configuration through random search on the validation set. To further investigate the impact of these weights on model performance, we designed three additional weight combinations to assess their effects on model enhancement. These combinations include  $c2 : (w_1 = 0.8, w_2 = 0.4, w_3 = 0.4)$ ,  $c3 : (w_1 = 1.0, w_2 = 0.1, w_3 = 0.5)$ , and  $c4 : (w_1 = 1.2, w_2 = 0.2, w_3 = 0.3)$ . Fig. 18 illustrates the influence of these weight combinations on the PSNR and SSIM scores, along with their corresponding visual results. From the experiments, we observe the following: In  $c3$ , the relatively low  $w_2$  value caused a significant loss of structural information during the restoration process, resulting in the lowest PSNR and SSIM scores. The generated images also exhibited noticeable detail blur. In  $c2$  and  $c4$ , the higher values of  $w_2 = 0.4$  and  $w_2 = 0.2$  respectively resulted in smaller structural losses, both yielding an SSIM score of 0.912. However, due to the values of  $w_1 = 0.8$  in  $c2$  and  $w_1 = 1.2$  in  $c4$ , the PSNR of  $c4$  was higher than that of  $c2$ . Correspondingly, the images generated by  $c4$  appeared sharper.

In conclusion, while the configuration  $c1$  achieved the best PSNR and SSIM results, further experimental analysis validated the significant impact of different weight combinations on model performance. Based on these results, we conclude that the proper configuration of weights is crucial for enhancing the model's effectiveness.

2) *Performance Vs. Computational Efficiency:* As summarized in Tab. V, DPF-Net achieves a favorable trade-off between restoration quality and computational efficiency. In terms of MACs and E2E. Times, DPF-Net outperforms GuidedHybSensUIR, PUIE-Net, and several other methods in efficiency, while surpassing them in visual quality. Although HE and CLAHE offer faster E2E speeds, their visual results are relatively poor. While FIVE A+, UWCNN, and USUIR achieve lower parameter counts, their significantly reduced capacities may limit restoration performance. Compared to recent methods such as MB-TaylorFormer, DPF-Net provides better performance in terms of MACs. Furthermore, DPF-Net-Lite demonstrates a highly competitive balance between restoration quality and computational efficiency, outperforming all methods in this regard.

As shown in Tab. VI, the full DPF-Net model achieves the

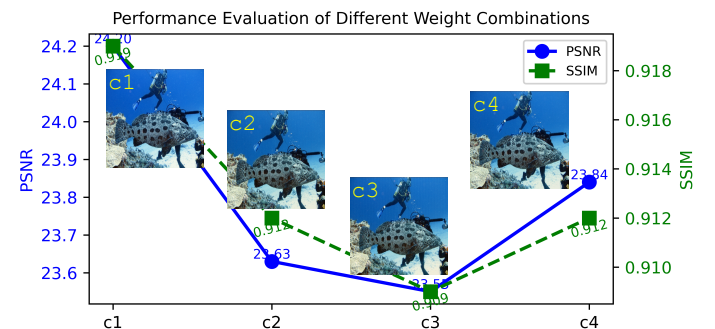


Fig. 18: Qualitative and quantitative evaluation results of different weight combinations.



TABLE V: Computational Cost Comparison For All Methods. The best results are highlighted in red, and the second-best are blue, and the third-best are purple, - Indicates no corresponding MACs and Params.

| Methods                     | Computational Cost |             |               |
|-----------------------------|--------------------|-------------|---------------|
|                             | MACs(G) ↓          | Param.(M) ↓ | E2E.Time(s) ↓ |
| DCP [4]                     | -                  | -           | 3.3555        |
| CLAHE [5]                   | -                  | -           | 0.0040        |
| MIP [6]                     | -                  | -           | 2.6000        |
| ICM [8]                     | -                  | -           | 0.1333        |
| UDCP [9]                    | -                  | -           | 2.6888        |
| IBLA [3]                    | -                  | -           | 6.9555        |
| UCM [10]                    | -                  | -           | 1.5555        |
| HE [7]                      | -                  | -           | 0.0035        |
| WaterNet [12]               | 71.42              | 1.09        | 0.0111        |
| UIEC <sup>2</sup> -Net [33] | 26.06              | 0.53        | 0.0888        |
| SCNet [29]                  | 23.56              | 0.82        | 0.0222        |
| GUPDM [37]                  | 95.80              | 1.49        | 0.4888        |
| Shallow-UWnet [27]          | 21.63              | 0.22        | 0.0651        |
| UWCNN [36]                  | 2.61               | 0.04        | 0.0666        |
| CCMSRNet [34]               | 43.60              | 21.13       | 0.0800        |
| USUIR [24]                  | 14.81              | 0.23        | 0.0528        |
| SFGNet [28]                 | 81.58              | 1.30        | 0.0228        |
| UGAN [13]                   | 19.82              | 57.17       | 0.1072        |
| FIVE A+ [11]                | 0.58               | 0.01        | 0.0666        |
| PUIE-Net [26]               | 30.09              | 1.40        | 0.1111        |
| GuidedHybSensUIR [19]       | 10.05              | 1.15        | 0.1111        |
| MB-TaylorFormer v2 [18]     | 31.36              | 2.62        | 0.0718        |
| Ours (DPF-Net)-Lite         | 6.43               | 0.31        | 0.0446        |
| Ours                        | 14.69              | 2.27        | 0.0746        |

best restoration quality, albeit at the cost of higher computational load and longer E2E. Time. In contrast, DPF-Net-Lite exhibits outstanding computational efficiency while maintaining strong competitive performance in terms of visual quality, as illustrated in Fig. 19. Moreover, analyzing the variants of DPF-Net reveals that the removal of key components, such as LFGA, CFEM, and HFAM, reduces complexity and E2E. Time but leads to performance degradation. The visual results of these variants, as shown in Fig. 19, further confirm the importance of these components in maintaining high-quality restoration.

Overall, DPF-Net strikes an optimal balance between complexity and restoration quality, achieving low MACs while delivering superior restoration performance. For resource-constrained application scenarios, DPF-Net-Lite provides an efficient alternative, ensuring that good restoration quality is maintained while meeting the computational efficiency

requirements.

3) *Generalization capability of Our method*: To evaluate the generalization ability of our method and the transferability of the model, we conducted generalization tests using the pre-trained UIEB-T model on the low-light dataset LOL-test-low, the remote sensing image dataset RRSHID-thin-fog, and the natural image dataset Dense-Haze. As shown in Fig. 20, it can be observed that, compared to the GuidedHybSensUIR and MB-TaylorFormer v2 methods, our approach demonstrates stronger performance in low-light enhancement and remote sensing image dehazing. However, our method produces relatively poorer results in indoor image denoising. These results indicate that our method exhibits superior performance in low-light enhancement and remote sensing image dehazing tasks.

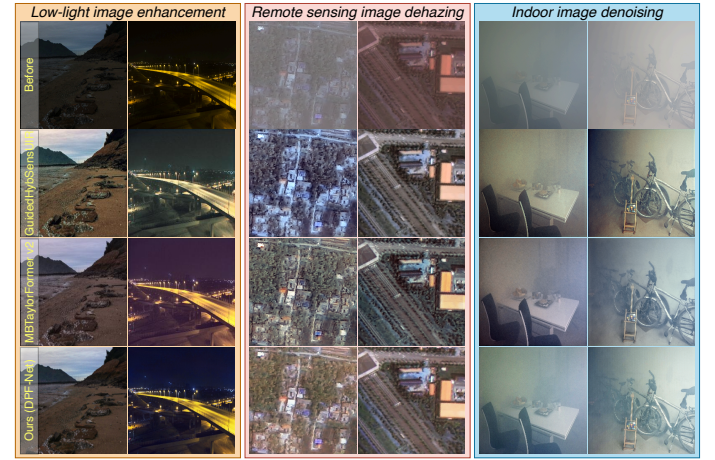


Fig. 20: The visual results of our method on low-light, remote sensing, and haze images are presented. The first row shows the original images, while the second row displays the results enhanced by our method.

4) *Other Low-Level Vision Tasks*: To further validate the performance of our method in low-light image enhancement and remote sensing image dehazing tasks, we trained on the low-light image dataset LOL-485 and tested on LOL-15. Additionally, we selected two representative methods, LIME and URetinex, for both quantitative and qualitative comparisons with our method, as shown in Fig. 21. Similarly, we conducted training on the remote sensing image dehazing dataset Hazelk-moderate-T and tested on Hazelk-moderate-TE. We also performed both qualitative and quantitative comparisons with the DCP and BMFH-Net methods, as depicted in Fig. 21. Clearly, our method generates higher-quality images in both low-quality image enhancement and remote sensing image dehazing tasks. This further corroborates the potential

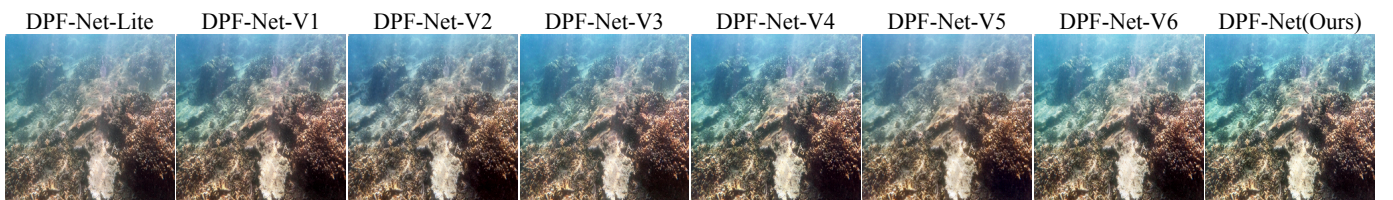


Fig. 19: The visual results of Performance Efficiency.

TABLE VI: Comparison of Performance, Model Complexity, and Efficiency Across DPF-Net Components

| Methods               | MACs (G) ↓ | Param. (M) ↓ | E2E.Time (s) ↓ | UCIQE ↑ | UIQM↑  | LPIPS↓ |
|-----------------------|------------|--------------|----------------|---------|--------|--------|
| DPF-Net-Lite          | 6.43       | 0.31         | 0.0446         | 0.4289  | 3.9437 | 0.0685 |
| DPF-Net-V1 (w/o LFGA) | 14.17      | 2.21         | 0.0588         | 0.4345  | 3.9481 | 0.0691 |
| DPF-Net-V2 (w/o CFEM) | 11.96      | 1.71         | 0.0650         | 0.4267  | 3.9097 | 0.0771 |
| DPF-Net-V3 (w/o HFAM) | 14.69      | 2.27         | 0.0610         | 0.4202  | 3.8142 | 0.0698 |
| DPF-Net-V4 (w/o DFEM) | 6.46       | 0.76         | 0.0597         | 0.4262  | 3.8899 | 0.0707 |
| DPF-Net-V5 (w/o DUIM) | 12.37      | 2.18         | 0.0716         | 0.4350  | 4.0100 | 0.0734 |
| DPF-Net-V6 (w/o MCFN) | 14.03      | 2.24         | 0.0598         | 0.4331  | 3.9103 | 0.0716 |
| DPF-Net               | 14.69      | 2.27         | 0.0746         | 0.4340  | 3.9908 | 0.0675 |

TABLE VII: Computational complexity, memory usage, inference latency, and restoration performance under different input resolutions.

| Methods      | Resolution | MACs (G)↓ | Param. Mem. (M)↓ | Act. Mem. (M)↓ | Peak GPU Mem (M)↓ | ΔPeak GPU Mem (M)↓ | Inf. Time (s)↓ | PSNR↑ | SSIM↑  | LPIPS↓ |
|--------------|------------|-----------|------------------|----------------|-------------------|--------------------|----------------|-------|--------|--------|
| UWCNN [36]   | 128×128    | 0.65      | 0.15             | 18.19          | 47.23             | 37.27              | 0.0008         | 18.76 | 0.844  | 0.1769 |
|              | 256×256    | 2.61      | 0.15             | 72.75          | 176.34            | 150.56             | 0.0011         | 18.82 | 0.856  | 0.1840 |
|              | 512×512    | 10.44     | 0.15             | 291.0          | 631.98            | 599.02             | 0.0031         | 18.83 | 0.8611 | 0.1964 |
| FIVE A+ [11] | 128×128    | 0.14      | 0.03             | 35.25          | 29.69             | 13.03              | 0.0024         | 20.23 | 0.867  | 0.0969 |
|              | 256×256    | 0.58      | 0.03             | 141.00         | 85.73             | 53.06              | 0.0027         | 21.55 | 0.890  | 0.0909 |
|              | 512×512    | 2.34      | 0.03             | 564.00         | 249.10            | 209.12             | 0.0058         | 22.12 | 0.895  | 0.0933 |
| DPF-Net-Lite | 128×128    | 1.61      | 7.93             | 100.69         | 71.22             | 45.5               | 0.0041         | 20.92 | 0.879  | 0.0839 |
|              | 256×256    | 6.43      | 7.93             | 402.13         | 208.28            | 182.0              | 0.0048         | 21.80 | 0.894  | 0.0812 |
|              | 512×512    | 25.71     | 7.93             | 1607.25        | 756.53            | 728.0              | 0.0111         | 23.03 | 0.906  | 0.0799 |

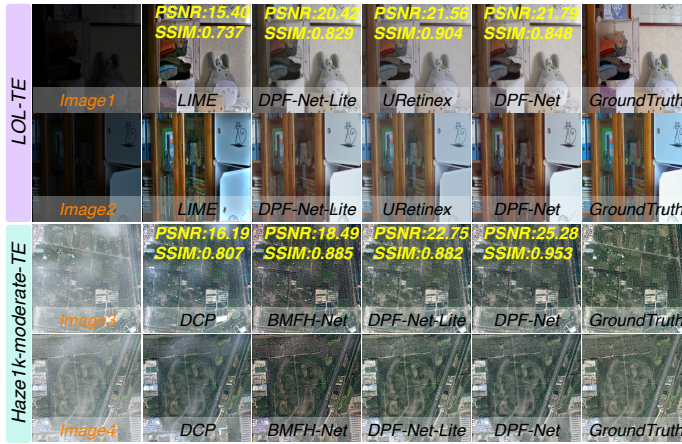


Fig. 21: The qualitative and quantitative results on the LOL-TE and Haze1k-moderate-TE datasets.

applicability of our model in other low-quality image enhancement tasks.

5) *Efficiency-Performance Trade-offs Under Multi-Resolution Inputs*: Tab.VII provides a multi-resolution profiling of compute, memory footprint, and latency for DPF-Net-Lite. As resolution increases from  $128 \times 128$  to  $512 \times 512$ , DPF-Net-Lite scales from 1.61G to 25.71G MACs and from 0.0041s to 0.0111s inference time, consistent with the quadratic growth in pixel count. Meanwhile, memory usage becomes increasingly dominated by activations: at  $512 \times 512$ , DPF-Net-Lite reaches 1607.25M activation memory and 756.53M peak GPU memory, indicating that runtime footprint is mainly driven by intermediate features rather than parameters at high resolutions. Compared with lightweight baselines, DPF-Net-Lite is not the fastest/smallest,

but it delivers a clearer quality advantage at higher resolution (e.g.,  $512 \times 512$ : 23.03/0.906/0.0799), illustrating a favorable efficiency-performance trade-off toward restoration fidelity. Although we do not conduct on-device deployment experiments, the presented profiling offers a concrete reference for evaluating resolution-dependent efficiency in practical settings.

### G. Limitation and Future Work

Although the proposed method consistently outperforms SOTA approaches in UIE, several limitations remain. First, as shown in Fig. 22, existing datasets mainly cover common degradations (e.g., color shift and low contrast) but provide limited samples of extreme underwater conditions; therefore, DPF-Net is constrained in learning robust representations for rare yet critical scenarios such as extreme blurriness and severe color shifts, making accurate detail recovery and quality restoration challenging. Second, while DPF-Net-Lite is introduced to reduce computational resources for resource-constrained settings, its simplified architecture may weaken feature extraction for complex distortions (e.g., greenish color shifts and fine-detail restoration), leading to suboptimal performance in these aspects. Third, as observed from the multi-resolution analysis in Discussion subsection five, multi-resolution training often fails to match the optimal performance achieved under a single fixed resolution, suggesting that jointly learning across resolutions can introduce optimization conflicts due to distribution discrepancies and misaligned objectives among different scales, which may hinder the model from fully learning key capabilities such as detail restoration and color correction.

To address these issues, future work will explore several directions. First, we aim to adopt image generation tech-



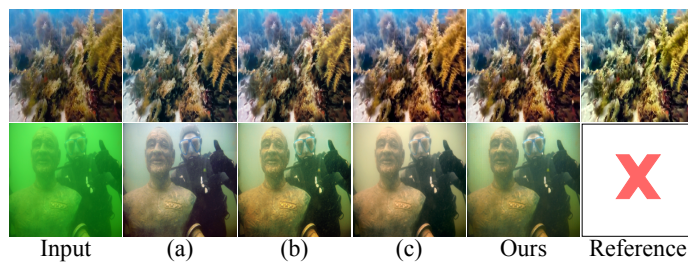


Fig. 22: Demonstration of failure cases for the proposed DPF-Net and SOTA methods is derived from the T-90 and U45 datasets. Specifically, (a) represents FIVE A+, (b) represents GuidedHybSensUIR, and (c) represents MB-TaylorFormer v2.

niques to augment the training dataset by synthesizing extreme underwater environments, thereby helping the network learn richer features for extreme scenarios. Second, we plan to introduce environmental priors for underwater images, such as depth, temperature, and turbidity, which can provide additional context for the model and improve its performance under complex underwater conditions. Finally, to address the performance drop in cross-resolution learning, we will explore improved cross-resolution training strategies to reduce multi-scale optimization conflicts and close the gap between multi- and single-resolution training.

## V. CONCLUSION

In this paper, we systematically review the current challenges in UIE and propose the DPF-Net method. To effectively exploit the frequency-domain priors of degraded images, this method ingeniously designs a dual-path downsampling and upsampling architecture based on DWT. The dual-path structure focuses on enhancing local features and performing global feature interaction fusion in degraded images, thereby enhancing underwater degraded images in both the frequency and pixel domains. Compared to SOTA UIE methods, DPF-Net demonstrates superior performance in extensive qualitative and quantitative evaluations. Additionally, it shows significant application potential in cross-task learning, such as low-light enhancement and remote sensing image dehazing. Furthermore, to meet real-time requirements, we propose the DPF-Net-Lite model, which generates high-quality images with reduced computational resources.

## REFERENCES

- [1] R. Cong, Y. Zhang, L. Fang, J. Li, Y. Zhao, and S. Kwong, "Rrnet: Relational reasoning network with parallel multiscale attention for salient object detection in optical remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.
- [2] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2019.
- [3] Y. -T. Peng and P. C. Cosman, "Underwater Image Restoration Based on Image Blurriness and Light Absorption," in *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1579–1594, April 2017.
- [4] Y. Xu, X. Guo, H. Wang, F. Zhao and L. Peng, "Single image haze removal using light and dark channel prior," 2016 IEEE/CIC International Conference on Communications in China (ICCC), Chengdu, China, 2016, pp. 1–6.

- [5] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, M. P. Braeuning, and S. M. Pizer, "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *Journal of Digital Imaging*, vol. 11, pp. 193–200, 1998.
- [6] L. B. Dorini and N. J. Leite, "Multiscale Methods for Image Processing: The Wavelet and the Scale-Space Approaches," 2009 Tutorials of the XXII Brazilian Symposium on Computer Graphics and Image Processing, Rio de Janeiro, Brazil, 2009, pp. 31–44.
- [7] Yeong-Taeg Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," in *IEEE Transactions on Consumer Electronics*, vol. 43, no. 1, pp. 1–8, Feb. 1997.
- [8] S. Gao and D. -y. Bi, "Equivalence Relation Between Intersecting Cortical Model and Binary Mathematic Morphology in Image Processing," 2009 International Conference on Artificial Intelligence and Computational Intelligence, Shanghai, China, 2009, pp. 490–494.
- [9] P. Drews Jr, E. do Nascimento, F. Moraes, S. Botelho and M. Campos, "Transmission Estimation in Underwater Single Images," 2013 IEEE International Conference on Computer Vision Workshops, Sydney, NSW, Australia, 2013, pp. 825–830.
- [10] K. Iqbal, M. Odetayo, A. James, Rosalina Abdul Salam and Abdullah Zawawi Hj Talib, "Enhancing the low quality images using Unsupervised Colour Correction Method," 2010 IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 2010, pp. 1703–1709.
- [11] J. Jiang, T. Ye, S. Chen, E. Chen, Y. Liu, S. Jun, J. Bai, and W. Chai, "Five A+ Network: You Only Need 9K Parameters for Underwater Image Enhancement," in 34th British Machine Vision Conference 2023, BMVC 2023, Aberdeen, UK, November 20–24, 2023, BMVA, 2023.
- [12] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2019.
- [13] C. Fabbri, M. J. Islam and J. Sattar, "Enhancing Underwater Imagery Using Generative Adversarial Networks," 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 2018, pp. 7159–7165.
- [14] K. Zhang, D. Li, W. Luo, W. Ren and W. Liu, "Enhanced Spatio-Temporal Interaction Learning for Video Deraining: Faster and Better," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 1287–1293, 1 Jan. 2023.
- [15] K. Zhang, R. Li, Y. Yu, W. Luo and C. Li, "Deep Dense Multi-Scale Network for Snow Removal Using Semantic and Depth Priors," in *IEEE Transactions on Image Processing*, vol. 30, pp. 7419–7431, 2021.
- [16] K. Zhang, W. Luo, Y. Zhong, L. Ma, W. Liu and H. Li, "Adversarial Spatio-Temporal Learning for Video Deblurring," in *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 291–301, Jan. 2019.
- [17] T. Wang, K. Zhang, Z. Shao, W. Luo, B. Stenger, T. Lu, T.-K. Kim, W. Liu, and H. Li, "GridFormer: Residual Dense Transformer with Grid Structure for Image Restoration in Adverse Weather Conditions," in *Springer International Journal of Computer Vision*, vol. 132, pp. 4541–4563, Oct. 2024.
- [18] Z. Jin, Y. Qiu, K. Zhang, H. Li and W. Luo, "MB-TaylorFormer V2: Improved Multi-Branch Linear Transformer Expanded by Taylor Formula for Image Restoration," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 7, pp. 5990–6005, July 2025.
- [19] X. Guo, X. Chen, S. Wang and C. -M. Pun, "Underwater Image Restoration Through a Prior Guided Hybrid Sense Approach and Extensive Benchmark Analysis," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 5, pp. 4784–4800, May 2025.
- [20] W. Zhang et al., "Underwater Image Enhancement via Weighted Wavelet Visual Perception Fusion," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 4, pp. 2469–2483, April 2024.
- [21] W. Zhang, Q. Liu, H. Lu, J. Wang and J. Liang, "Underwater Image Enhancement via Wavelet Decomposition Fusion of Advantage Contrast," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 8, pp. 7807–7820, Aug. 2025.
- [22] Y. -Z. Hsieh and M. -C. Chang, "Underwater Image Enhancement and Attenuation Restoration Based on Depth and Backscatter Estimation," in *IEEE Transactions on Computational Imaging*, vol. 11, pp. 321–332, 2025.
- [23] Z. Jingchun, S. Jiaming, Z. Weishi, and L. Zifan, "Multi-view underwater image enhancement method via embedded fusion mechanism," *Engineering Applications of Artificial Intelligence*, vol. 121, pp. 105946, 2023.
- [24] Z. Fu, "Unsupervised Underwater Image Restoration: From a Homology Perspective", *AAAI*, vol. 36, no. 1, pp. 643–651, Jun. 2022.
- [25] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in 2017 IEEE

- International Conference on Computer Vision (ICCV). Venice, Italy: IEEE Computer Society, 2017, pp. 2242–2251.
- [26] Z. Fu, W. Wang, Y. Huang, X. Ding, and K.-K. Ma, "Uncertainty inspired underwater image enhancement," in European Conference on Computer Vision (ECCV), 2022, pp. 465–482.
- [27] A. Naik, A. Swarnakar, and K. Mittal, "Shallow-UWNet: Compressed model for underwater image enhancement (student abstract)," in AAAI Conference on Artificial Intelligence, vol. 35, no. 18, 2021, pp. 15853–15854.
- [28] C. Zhao, W. Cai, C. Dong, and Z. Zeng, "Toward sufficient spatial-frequency interaction for gradient-aware underwater image enhancement," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024, pp. 3220–3224.
- [29] K. Han, R. S. Rezende, B. Ham, K.-Y. K. Wong, M. Cho, C. Schmid, and J. Ponce, "Scnet: Learning semantic correspondence," in IEEE International Conference on Computer Vision 2017, ICCV 2017, pp. 1831–1840.
- [30] J. Tan et al., "Unsupervised Lifelong Person Re-identification via Affinity Harmonization," in ACM Transactions on Multimedia Computing, Communications and Applications, vol. 21, no. 2, pp. 123–134, 2025.
- [31] H. Wang, Y. Chen, M. Yao, W. Liu, J. Peng and X. Fu, "Tensor Completion Framework by Graph Refinement for Incomplete Multi-View Clustering," in IEEE Transactions on Multimedia, vol. 27, pp. 9385–9398, 2025.
- [32] T. Xiao et al., "Disentangled Pose and Appearance Guidance for Multi-Pose Generation," 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2025, pp. 5646–5655.
- [33] Y. Wang, J. Guo, H. Gao, and H. Yue, "UIEC<sup>2</sup>-Net: CNN-based underwater image enhancement using two color space," Signal Processing: Image Communication, vol. 96, pp. 116250, 2021.
- [34] H. Qi, H. Zhou, J. Dong and X. Dong, "Deep Color-Corrected Multiscale Retinex Network for Underwater Image Enhancement," in IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1–13, 2024.
- [35] S. Liu et al., "Turbid Underwater Image Enhancement With Illumination-Constrained and Structure-Preserved Retinex Model," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 35, no. 11, pp. 10844–10861, Nov. 2025.
- [36] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," Pattern Recognition, vol. 98, 2020.
- [37] P. Mu, H. Xu, Z. Liu, Z. Wang, S. Chan and C. Bai, "A generalized physical-knowledge-guided dynamic model for underwater image enhancement," Proc. of the 31st ACM International Conference on Multimedia (ACM MM), pp. 7111–7120, 2023.
- [38] R. Priyadharsini, T. Sree Sharmila, and V. Rajendran, "A wavelet transform based contrast enhancement method for underwater acoustic images," Multidimensional Systems and Signal Processing, vol. 29, pp. 1845–1859, 2018.
- [39] E. O. Brigham and R. Morrow, "The fast fourier transform," IEEE Spectrum, vol. 4, no. 12, pp. 63–70, 1967.
- [40] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp. 674–693, July 1989.
- [41] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2818–2826.
- [42] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," The Journal of Machine Learning Research, vol. 13, no. 1, pp. 281–305, 2012.
- [43] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," Pattern Recognition, vol. 98, 2020.
- [44] L. Peng, C. Zhu, and L. Bian, "U-shape transformer for underwater image enhancement," IEEE Transactions on Image Processing, vol. 32, pp. 3066–3079, 2023.
- [45] H. Li, J. Li, and W. Wang, "A fusion adversarial underwater image enhancement network with a public test dataset," arXiv, 2019.
- [46] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," arXiv preprint arXiv:1808.04560, 2018.
- [47] Z. H. Zhu, W. Lu, S. B. Chen, C. H. Q. Ding, J. Tang, and B. Luo, "Real-World Remote Sensing Image Dehazing: Benchmark and Baseline," IEEE Transactions on Geoscience and Remote Sensing, pp. 1–1, 2025.
- [48] C. O. Ancuti, C. Ancuti, M. Sbert and R. Timofte, "Dense-Haze: A Benchmark for Image Dehazing with Dense-Haze and Haze-Free Images," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 1014–1018.
- [49] R. Liu, X. Fan, M. Zhu, M. Hou and Z. Luo, "Real-World Underwater Enhancement: Challenges, Benchmarks, and Solutions Under Natural Light," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 12, pp. 4861–4875, Dec. 2020.
- [50] T. Song, S. Fan, P. Li, J. Jin, G. Jin and L. Fan, "Learning an Effective Transformer for Remote Sensing Satellite Image Dehazing," in IEEE Geoscience and Remote Sensing Letters, vol. 20, pp. 1–5, 2023.
- [51] Z. Liu, "Super convergence cosine annealing with warm-up learning rate," in Proc. CAIBDA 2022, 2nd Int. Conf. Artif. Intell., Big Data and Algorithms, VDE, pp. 1–7, 2022.
- [52] X. Guo, Y. Li and H. Ling, "LIME: Low-Light Image Enhancement via Illumination Map Estimation," in IEEE Transactions on Image Processing, vol. 26, no. 2, pp. 982–993, Feb. 2017.
- [53] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang, "URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2022, pp. 5901–5910.
- [54] H. Sun et al., "Bidirectional-Modulation Frequency-Heterogeneous Network for Remote Sensing Image Dehazing," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 35, no. 11, pp. 10649–10664, Nov. 2025.
- [55] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [56] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 586–595.
- [57] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," IEEE Transactions on Image Processing, vol. 24, no. 12, pp. 6062–6071, 2015.
- [58] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," IEEE Journal of Oceanic Engineering, vol. 41, no. 3, pp. 541–551, 2015.
- [59] R. C. Gonzalez and R. E. Woods, "Digital Image Processing", 3rd ed. Upper Saddle River, NJ, USA: Pearson Prentice Hall, 2008.
- [60] P. Kovesi, "Image features from phase congruency," Videre: Journal of Computer Vision Research, vol. 1, no. 3, pp. 1–26, 1997.
- [61] C. Li et al., "An underwater image enhancement benchmark dataset and beyond," IEEE Trans. Image Process., vol. 29, pp. 4376–4389, 2019.
- [62] P. Guo, H. Liu, D. Zeng, T. Xiang, L. Li, and K. Gu, "An underwater image quality assessment metric," IEEE Trans. Multimedia, vol. 25, pp. 5093–5106, 2022.
- [63] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, 2004.
- [64] J. Canny, "A computational approach to edge detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, 1986.
- [65] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," arXiv preprint arXiv:2107.08430, 2021.



**Yupeng Ma** received the B.S. degree in Computer Science from Tianjin Normal University in 2021 and the M.S. degree in Computer Science from Tianjin Polytechnic University in 2024. He is currently pursuing the Ph.D. degree in Computer Science at Hainan University.

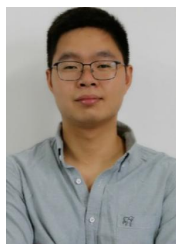
His research interests include underwater image enhancement, underwater image transmission, and image coding.





**Qiuling Yang** received the Ph.D. degree from Tianjin University, Tianjin, China, in 2016. She is currently a professor and doctoral supervisor in Hainan University, Haikou, China.

Her research interests include protocols design and security of the underwater acoustics sensor networks.



**Rongxin Zhu** (Member, IEEE) received his M.E. degree in software engineering from Nanjing University, Nanjing in 2017 and the Ph.D. degree in Hainan University, Haikou, China in 2024.

He is currently a professor and doctoral supervisor in Hainan University, Haikou, China. His research interests include the security of underwater sensor networks and intelligent algorithms.



**YiFei Li** received the Ph.D. degree in Engineering Mechanics from Hohai University, Nanjing, China, in March 2024, and the Ph.D. degree in Mechanical and Electronic Engineering from Ghent University, Ghent, Belgium, in June 2024.

He is currently an Associate Professor and Master's Supervisor with the School of Engineering, Huzhou University. His research interests include structural health monitoring based on applied soft computing methods.



**Magd Abdel Wahab** (Fellow, IEEE) received the B.Sc. and M.Sc. degrees from Cairo University, Cairo, Egypt, and the Ph.D. degree from KU Leuven, Leuven, Belgium, as well as the D.Sc. degree from the University of Surrey, Surrey, U.K. He is currently a Full Professor and Chair of Applied Mechanics, and the Head of the Applied Mechanics Research Group at Laboratorium Soete, Faculty of Engineering and Architecture, Ghent University, Belgium. His research interests include applied mechanics, fatigue analysis, and structural dynamics.