

WCI-Net: A Wavelet-guided Cross-domain Interaction Network for Underwater Image Enhancement

Yupeng Ma, and Qiuling Yang

Abstract—Underwater images suffer from color distortion and detail loss due to wavelength-dependent absorption and scattering. Existing deep learning-based enhancement methods primarily operate in the pixel domain or rely on a single frequency-domain transform, limiting their ability to jointly preserve fine-grained details and model global dependencies. To address this gap, we propose WCI-Net, a wavelet-guided cross-domain interaction network for underwater image enhancement. WCI-Net decomposes features into multi-scale wavelet subbands to capture local structural details, while a cross-domain interaction module integrates wavelet-domain representations with Fourier-domain global refinement. A dual-domain loss further regularizes learning in both spatial and frequency domains. Extensive experiments on real-world benchmarks demonstrate that WCI-Net achieves state-of-the-art enhancement performance with low computational cost.

Index Terms—Discrete wavelet transform, Fast Fourier transform, Underwater image enhancement, Underwater Image Restoration

I. INTRODUCTION

Underwater image enhancement (UIE) plays a crucial role in computer vision. Influenced by multiple factors such as water depth, light attenuation, and suspended particles, underwater images often suffer from color distortion, reduced contrast, and loss of details [1]. In particular, due to the absorption and scattering of red light in water, the red component decays rapidly, causing the overall image to appear bluish or greenish. Therefore, UIE is of great importance for improving image quality and advancing practical applications such as underwater target recognition [2], object detection [3], and autonomous underwater robotics [4].

To alleviate these challenges, various image enhancement techniques have been developed. Traditional UIE methods mainly improve image quality by adjusting the pixel values of a given image or by relying on assumptions in physical models [5]–[12]. Although these methods are effective to some extent, they often struggle to address color distortions and low contrast in complex underwater environments. In recent years, deep

learning has advanced UIE through its strong nonlinear mapping capabilities. In these approaches, deep neural networks are typically trained on paired datasets consisting of degraded underwater images and corresponding ground-truth references, thereby enabling effective contrast enhancement and color correction [13]–[25]. However, traditional CNNs are constrained by their limited receptive field, which impedes their ability to model long-range dependencies in images. Recent studies have shown that frequency-domain feature extraction methods offer new insights for addressing this limitation [26]–[33]. Specifically, some methods utilize the Fast Fourier Transform (FFT) [30], [31] to explore the relationship between frequency features and image degradation (e.g., haze). However, the global nature of the FFT representation makes it relatively insensitive to local detail variations. Meanwhile, another line of research leverages wavelet transforms to exploit their multi-scale analysis strengths [29], [32]. These approaches use wavelet transforms to decompose images into high- and low-frequency components, then apply convolutional or attention mechanisms on the resulting sub-bands for enhancement [28], [29], achieving good results in tasks such as dehazing. Although these methods demonstrate the effectiveness of different transform domains individually, existing works often employ only a single transform in isolation and thus fail to fully explore or integrate the complementary information from different frequency-domain representations. Therefore, a key open question is how to systematically combine the multi-scale local characterization capability of wavelet transforms with the global representation capability of Fourier transforms in UIE. Achieving effective interaction and synergy between these two forms of representation remains a problem worthy of in-depth investigation.

Accordingly, we propose WCI-Net, a wavelet-guided cross-domain interaction network with a hierarchical encoder-decoder architecture. In the encoder, the network employs DWT to decompose the input and generate multi-scale sub-bands, while in the decoder it performs upsampling via inverse DWT (IDWT), and introduces a novel cross-domain interaction module (CIM) at each stage. This module fuses wavelet sub-bands with Fourier-domain refinements through dense convolutional layers, enabling the joint modeling of spatial details, frequency context, and global dependencies. Furthermore, to simultaneously regularize learning in both the frequency and spatial domains, we adopt a dual-domain loss allowing WCI-Net to achieve stable convergence as well as

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 YSPITZX202314; in part by the Hainan Provincial Natural Science Foundation of China under Grants 626MS0092 and 626QN0556.

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

superior enhancement performance. The main contributions are as follows:

- We propose WCI-Net, a wavelet-guided cross-domain interaction network tailored for UIE, which adopts a hierarchical encoder-decoder architecture with DWT/IDWT-based multi-scale representation learning.
- We propose a cross-domain interaction module CIM, which consists of DWT-based local feature extraction and FFT-based global feature extraction to capture fine-grained features and global contextual information.
- We introduce a dual-domain constrained loss function to regulate both the spatial and frequency domains between the input and output images.
- Extensive experiments demonstrating state-of-the-art results on benchmark UIE datasets, with ablation studies validating each component.

II. PROPOSED METHODOLOGY

A. Fundamentals

We select the orthogonal Haar wavelet, defined by its low-pass and high-pass analysis filters:

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

The four subbands are labeled as $\mathcal{N} = \{LL, HL, LH, HH\}$, with L and H applied in vertical and horizontal directions. Consequently, HL highlights vertical edges with horizontal smoothing, LH emphasizes horizontal variations, and HH detects diagonal features. For a given input tensor $F \in \mathbb{R}^{H \times W \times C}$, DWT operates channel-by-channel via separable 1D convolutions and subsequent downsampling:

$$I_{pq} = \downarrow_2 ((F *_{\nu} p) *_{\mu} q), \quad p, q \in \{L, H\}, \quad (1)$$

where $*_{\nu}$ and $*_{\mu}$ represent convolutions along height and width, and $\downarrow_2$ indicates $2 \times$ spatial subsampling. This produces subbands $I_{pq} \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times C}$. In underwater scenes, global color casts and haze-like degradations mainly reside in the low-frequency LL component, whereas edges and fine textures of marine organisms or man-made targets are predominantly captured by the HL , LH , and HH subbands.

Primarily, the LL subband maintains broad color and layout information, while $\{HL, LH, HH\}$ isolate directional textures such as vertical, horizontal, and oblique details [34]. Thanks to the Haar wavelet's orthogonality, perfect signal recovery is achieved through the IDWT from these subbands:

$$F' = \sum_{p,q \in \{L,H\}} (\uparrow_2 I_{pq} *_{\nu} \tilde{p} *_{\mu} \tilde{q}) = \text{IDWT}(\{I_{pq}\}; L, H), \quad (2)$$

with $\uparrow_2$ denoting $2 \times$ zero-insertion upsampling, and synthesis filters $\tilde{L}, \tilde{H}$ matching L and H (accounting for convolution reversal). Appropriate boundary handling and even dimensions ensure exact reconstruction, $F' = F$. Compared with standard strided convolutions or pooling, such a DWT/IDWT pyramid changes the spatial resolution while preserving invertibility, which is desirable for enhancement tasks where both faithful reconstruction and multi-scale analysis are important.

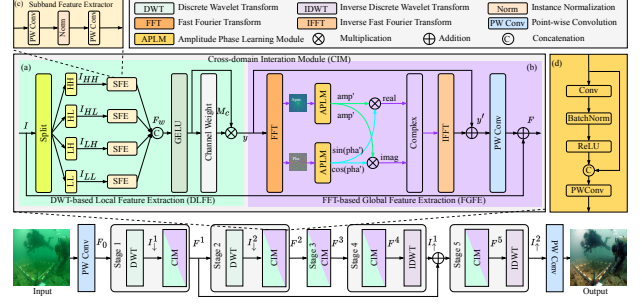


Fig. 1. Overview of the proposed WCI-Net. The method primarily consists of three modules: DWT, CIM, and IDWT, where CIM includes DWT-based Local Feature Extraction for capturing intricate local and finer-grained spatial features, and FFT-based Global Feature Extraction for analyzing amplitude and phase spectra to provide a holistic view of global frequency content.

B. Proposed WCI-Net Method

1) *Overall Architecture*: Our goal is to maximize the expressive potential of neural networks for UIE. As shown in Fig. 1, given an degraded image $x \in \mathbb{R}^{H \times W \times 3}$ as input, the proposed WCI-Net first extracts low-level features $F_0 \in \mathbb{R}^{H \times W \times C}$ by using a Point-wise Convolution (PW Conv), where $H \times W$ represents the spatial dimensions, and C denotes the number of channels, with $C = 32$. Then, the feature F_0 passes through a five-stage hierarchical encoder-decoder architecture. In the encoder, DWT performs decoupling and generates guide sub-bands, allowing the network to explicitly separate low-frequency color/illumination components from high-frequency structural details that are easily degraded in underwater environments. Correspondingly, the decoder employs IDWT for multi-scale upsampling, progressively restoring the spatial resolution. At each stage's CIM, the output features are $F^1 \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times 4C}$, $F^2 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times 16C}$, $F^3 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times 16C}$, $F^4 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times 16C}$, and $F^5 \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times 4C}$. Due to the ability of multi-level wavelet transform to obtain richer fine-grained details, we denote the feature maps obtained after the i -th DWT-based downsampling as $I_{\downarrow}^i$ ($i = 1, 2$), and those produced by the i -th IDWT-based upsampling as $I_{\uparrow}^i$ ($i = 1, 2$). To preserve structural and textural information and enhance restoration performance, the feature F^1 is concatenated with the corresponding upsampled feature $I_{\uparrow}^1$ via a skip connection, which mitigates the loss of fine details caused by repeated transforms and downsampling. Finally, PW Conv is applied to the upsampled feature $I_{\uparrow}^2$ to generate the restored, clear underwater image.

2) *Cross-domain Interaction Module*: To capture finer-grained spatial features and global frequency content, we develop the Cross-domain Interaction Module (CIM), as illustrated in Fig. 1. The CIM includes DWT-based Local Feature Extraction (DLFE) and FFT-based Global Feature Extraction (FGFE).

DLFE: Given an input feature map I , DLFE focuses on local, multi-directional textures by first applying DWT to split I into four sub-band components $I_{LL}, I_{LH}, I_{HL}, I_{HH}$.

To further refine them individually, we introduce a Subband Feature Extractor (SFE), as shown in Fig. 1(c). This process can be formalized as:

$$F_j = \tilde{f}_{pw}(\text{IN}(\tilde{f}_{pw}(I_j))), \quad j \in \mathcal{N}. \quad (3)$$

where IN denotes Instance Normalization, $\tilde{f}_{pw}$ denotes PW Conv. The refined sub-band features are concatenated to form a unified wavelet-domain representation:

$$F_w = \text{Concat}(F_{LL}, F_{LH}, F_{HL}, F_{HH}). \quad (4)$$

In parallel, to emphasize salient channels, we first apply GELU activation to F_w . Let $G = \text{GELU}(F_w)$. We then compute an attention map M_c from the globally aggregated G . Specifically, global average pooling yields a channel descriptor $z = \text{GAP}(G)$, which is fed into a lightweight two-layer MLP to produce M_c :

$$M_c = \sigma \left(\tilde{f}_{pw2} \left(\text{ReLU} \left(\tilde{f}_{pw1}(z) \right) \right) \right), \quad (5)$$

where $\tilde{f}_{pw1}$ reduces channels from $4C$ to $2C$, $\tilde{f}_{pw2}$ expands back to $4C$, and σ denotes the sigmoid activation. The channel-aware response is

$$y = G \odot M_c, \quad (6)$$

where $\odot$ denotes element-wise multiplication with channel-wise broadcasting, and y is the output of DLFE. This wavelet-guided branch therefore acts as a local feature extractor that selectively emphasizes informative directional details while suppressing less relevant responses.

FGFE: Given an input feature y from DLFE, to achieve global contextual modeling and handle large-scale color casts and veiling effects that extend beyond local receptive fields, FGFE applies the FFT to y :

$$\text{amp}, \text{pha} = \mathcal{F}(y), \quad (7)$$

where $\mathcal{F}$ is FFT, amp and pha denote the magnitude and phase, respectively.

To precisely capture and refine the magnitude and phase features independently, we develop a Amplitude Phase Learning Module (APLM), as shown in Fig. 1(d). APLM consists of a Conv-BN-ReLU unit, followed by channel-wise concatenation with the original input and a final $\tilde{f}_{pw}$. The refinement for magnitude and phase components is formally expressed as:

$$\begin{aligned} \text{amp}' &= \tilde{f}_{pw}(\text{Concat}(\text{ReLU}(\text{BN}(\text{Conv}(\text{amp}))), \text{amp})), \\ \text{pha}' &= \tilde{f}_{pw}(\text{Concat}(\text{ReLU}(\text{BN}(\text{Conv}(\text{pha}))), \text{pha})). \end{aligned} \quad (8)$$

By treating amplitude and phase explicitly, the network can globally adjust the energy distribution related to contrast and color while preserving the structural layout encoded in the phase. Leveraging Euler's formula, the refined magnitude and phase are then utilized to reconstruct the real and imaginary parts of the spectrum:

$$\begin{aligned} \text{real} &= \text{amp}' \odot \cos(\text{pha}'), \\ \text{imag} &= \text{amp}' \odot \sin(\text{pha}'), \end{aligned} \quad (9)$$

where $\odot$ indicates element-wise multiplication.

These parts are combined to form the complex spectrum $\text{real} + j\text{imag}$, which is mapped back to the spatial domain through the inverse FFT (IFFT). To maintain feature stability, a residual connection is incorporated. This procedure is formalized as

$$y' = \mathcal{F}^{-1}(\text{real} + j\text{imag}) + y, \quad (10)$$

where $y' \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times 4C}$, $\mathcal{F}^{-1}$ denotes IFFT. Finally, a PW Conv fuses the complementary wavelet- and Fourier-domain features, and a residual addition with I produces the FGFE output:

$$F = I + \tilde{f}_{pw}(y'). \quad (11)$$

C. Loss Function

To jointly constrain reconstruction quality in the spatial and frequency domains, we adopt the following dual-domain loss:

$$L = \frac{1}{N} \sum_{i=1}^N \sqrt{(\hat{x}_i - G_i)^2 + \epsilon^2} + \lambda \|\mathcal{F}(\hat{x}) - \mathcal{F}(G)\|_1, \quad (12)$$

where the first term is the Charbonnier loss (a differentiable variant of the ℓ_1 norm), and the second term penalizes the ℓ_1 discrepancy between the Fourier transforms of the prediction and the ground truth. The Charbonnier term mainly encourages spatial-domain fidelity and edge preservation, whereas the Fourier discrepancy term regularizes the global and mid-frequency components, effectively reducing color shifts and ringing artifacts that are typical in underwater enhancement. Here, $\hat{x}$ and G denote the reconstructed and ground-truth images, respectively; N is the number of pixels; $\epsilon = 10^{-6}$ is a small constant for numerical stability; and $\mathcal{F}(\cdot)$ is the FFT. The balancing coefficient is set to $\lambda = 0.1$.

III. EXPERIMENTS

A. Experimental Details

1) *Experimental Setup:* WCI-Net was implemented in PyTorch 2.4 with CUDA 12.1 and trained on two NVIDIA RTX 4090 GPUs using AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\text{lr} = 2 \times 10^{-4}$) with Cosine Annealing [35]. Training ran for 200 and 400 epochs on UIEBD and LSUI, respectively, with images resized to 256×256 and batch size 8.

2) *Datasets:* We use UIEBD [13] and LSUI [36] for training and evaluation. UIEBD contains 800 training pairs and 90 test pairs (T-90). LSUI contains 3852 training pairs and 427 test pairs (LSUI-TE).

3) *Evaluation Metrics:* We use five evaluation metrics: PSNR [37], SSIM [37], LPIPS [38], UIQM [39], and UCIQE [40] to quantitatively assess the enhancement performance of all methods.

B. Qualitative and Quantitative Assessment

1) *Analysis of Intermediate Features, Magnitude and Phase Spectra:* As shown in Fig. 2, spatial-domain feature maps progress from coarse to fine representations. At initialization, the network smooths dominant color shifts while preserving gross structural cues. During downsampling (Stages 1–2),

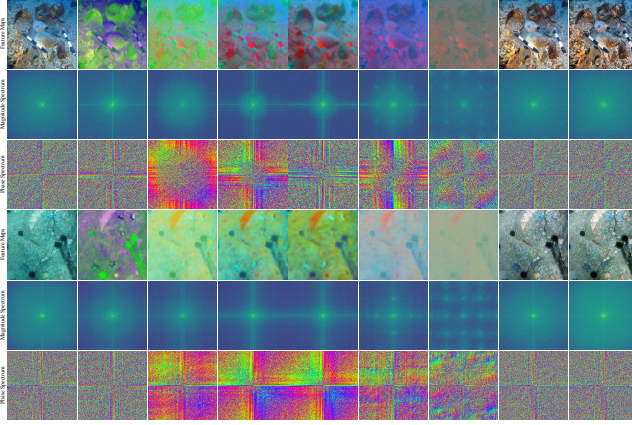


Fig. 2. Visualization of intermediate feature maps and their corresponding magnitude and phase spectra produced by WCI-Net on T-90 and LSUI-TE datasets. Each row block corresponds to one dataset: rows 1–3 shows the T-90 example and rows 4–6 shows the LSUI-TE example.

low-frequency components dominate, with magnitude spectra energy concentrating near the center and phase spectra exhibiting disordered patterns. In the latent stage, global structural priors are emphasized, with subtle re-emergence of medium-frequency content and increasingly structured phase patterns. During upsampling (Stages 4–5), magnitude spectra progressively regain high-frequency energy and phase maps become more coherent, reflecting the reconstruction of edges and textures. In the final output, both spectra closely approximate the ground truth, confirming that WCI-Net effectively restores corrected color and fine structural details through coordinated magnitude and phase recovery.

2) *Comparative Analysis with State-of-the-Art UIE Methods*: Table I quantitatively compares WCI-Net with representative traditional and deep learning methods on T-90 and LSUI-TE datasets. Compared with GuidedHybSensUIR, WCI-Net achieves relative improvements of 4.32%, 0.88%, and 5.26% in PSNR, SSIM, and LPIPS on T-90, and gains of 5.87% and 1.42% in PSNR and SSIM on LSUI-TE. Although UCIQE and UIQM do not always rank WCI-Net first, possibly due to their reliance on simplified human visual system characteristics [41], the results remain close to the overall averages, confirming that WCI-Net meets basic visual quality expectations. Qualitatively, Figs. 3 and 4 demonstrate that WCI-Net’s RGB channel distributions and pixel frequency curves are most consistent with the Ground Truth, confirming more balanced restoration and natural color reproduction across multi-spectral channels.

3) *Comparison of Computational Cost*: Table II presents the computational cost comparison in terms of MACs, parameter counts, and End to End (E2E) time. Traditional methods such as HE and CLAHE exhibit the lowest latency but lack generalization in complex underwater conditions. Among deep learning methods, UWCNN [14] achieves the lowest MACs (2.61G) with high efficiency, while FIVE_APLUS [16] attains the smallest model size (0.01M) at the cost of higher

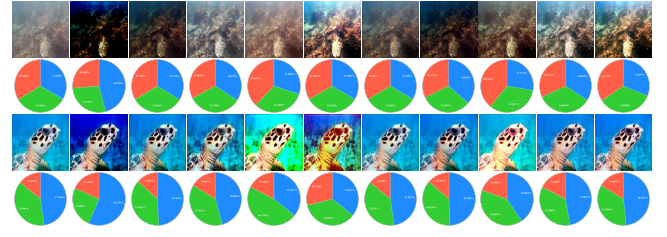


Fig. 3. Visualization of enhancement results from traditional methods and WCI-Net with RGB channel proportion statistics. Rows 1–2: T-90 example with pie charts. Rows 3–4: LSUI-TE example with pie charts.

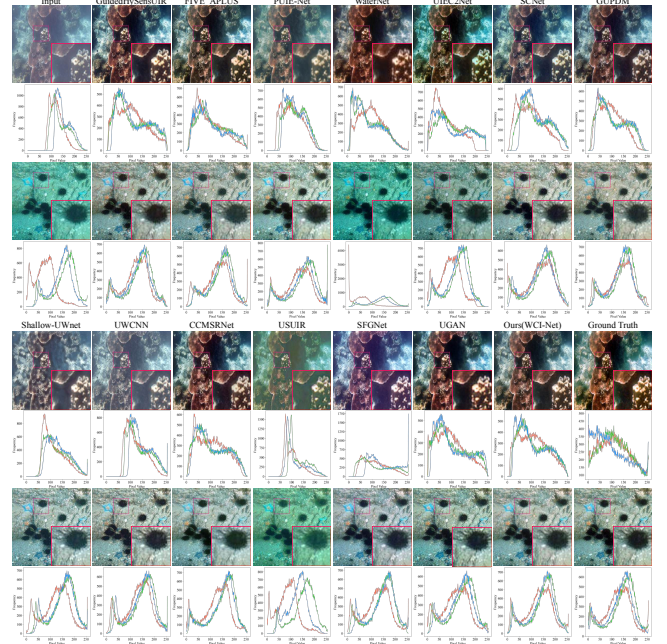


Fig. 4. Visualization of enhancement results from deep learning methods and WCI-Net with pixel frequency line charts. Rows 1–2: T-90 example. Rows 3–4: LSUI-TE example.

computation. The proposed WCI-Net achieves a balanced trade-off with 17.56G MACs, 5.65M parameters, and 0.0444s inference time, delivering superior enhancement quality while maintaining reasonable computational cost, demonstrating its practical potential for real-time underwater imaging.

C. Ablation Study

To validate the effectiveness of each component in WCI-Net, we conduct an ablation study on the T-90 dataset, with results shown in Table III and Fig. 5. Removing SFE significantly degrades performance, while ablating Channel Weight results in PSNR/SSIM/LPIPS of 23.77/0.908/0.0689. Excluding DLFE causes the most significant drop, confirming its importance in local refinement. For FGFE, removing APLM reduces performance, and excluding FGFE entirely yields 23.49/0.908/0.0721. Removing $\mathcal{L}_c$ and $\mathcal{L}_f$ decreases PSNR to 22.76 and 23.44, respectively. The full model outperforms

TABLE I
QUANTIATIVE RESULTS ON THE T-90 AND LSUI-TE DATASETS. BEST RESULTS IN RED AND SECOND-BEST IN BLUE.

Method	T-90					LSUI-TE				
	PSNR↑	SSIM↑	LPIPS↓	UCIQE↑	UIQM↑	PSNR↑	SSIM↑	LPIPS↓	UCIQE↑	UIQM↑
IBLA [5]	15.92	0.659	0.2024	0.4734	2.0642	17.96	0.740	0.2326	0.4568	2.4456
DCP [6]	14.60	0.693	0.1915	0.4536	2.2569	15.71	0.685	0.2696	0.4488	2.1064
CLAE [7]	19.22	0.865	0.1177	0.4184	3.2264	17.46	0.802	0.2376	0.4179	3.1649
MIP [8]	14.67	0.669	0.2110	0.4524	2.3057	15.65	0.663	0.2622	0.4532	2.2627
HE [9]	17.12	0.812	0.2090	0.5001	3.2291	14.85	0.701	0.3264	0.4993	3.3199
ICM [10]	17.56	0.807	0.1494	0.4148	2.7355	20.24	0.846	0.1781	0.4252	2.9204
UDCP [11]	15.55	0.720	0.1788	0.4394	2.3868	17.10	0.739	0.2473	0.4288	2.3600
UCM [12]	17.22	0.822	0.1740	0.4827	2.9744	19.93	0.846	0.2047	0.4640	3.0790
FIVE_APLUS [16]	22.71	0.904	0.0861	0.4514	3.9447	25.55	0.903	0.1039	0.4288	3.7766
PUIE-Net [17]	19.48	0.822	0.1332	0.3389	3.7955	26.99	0.899	0.0575	0.4190	3.8694
WaterNet [13]	21.96	0.845	0.1495	0.4516	3.6751	20.58	0.813	0.1813	0.3883	3.0310
UIEC2Net [15]	20.11	0.892	0.1114	0.4417	3.6636	23.48	0.903	0.0705	0.3977	3.6883
SCNet [23]	22.19	0.895	0.0933	0.4060	3.2674	25.27	0.883	0.0726	0.4082	3.1135
GUPDM [18]	21.69	0.853	0.1309	0.3656	3.1967	24.81	0.868	0.1221	0.3947	3.1987
Shallow-UWnet [22]	19.21	0.845	0.1456	0.3925	3.7856	21.59	0.845	0.1099	0.3738	3.7372
UWCNN [14]	18.86	0.859	0.1222	0.3871	3.8554	22.89	0.886	0.1187	0.3937	3.8471
CCMSRNet [20]	22.31	0.880	0.1054	0.4109	3.7147	27.42	0.904	0.0861	0.4190	3.6604
USUIR [19]	18.69	0.776	0.1755	0.3704	3.6190	18.34	0.778	0.2501	0.3964	3.4792
SFGNet [21]	21.49	0.896	0.1347	0.4428	3.0335	25.13	0.874	0.1278	0.4203	3.1059
UGAN [24]	22.60	0.847	0.1396	0.4495	3.4664	26.24	0.880	0.1444	0.4480	3.0886
GuidedHybSensUIR [25]	23.36	0.909	0.0722	0.4393	3.8085	28.10	0.917	0.0455	0.4243	3.7116
Ours (WCI-Net)	24.37	0.917	0.0684	0.4459	3.7375	29.75	0.930	0.0622	0.4252	3.6775

TABLE II
COMPUTATIONAL COST COMPARISON FOR ALL METHODS WITH PUBLICATION YEAR AND VENUE. THE BEST RESULTS ARE IN RED, THE SECOND-BEST IN BLUE, AND THE THIRD-BEST IN PURPLE. - INDICATES MISSING MACs OR PARAMS.

Method	Year	Journal/Conf.	MACs(G) ↓	Params(M) ↓	E2E.Time(s) ↓
IBLA [5]	2017	TIP	-	-	6.9555
DCP [6]	2010	TPAMI	-	-	3.3555
CLAE [7]	1998	JDI	-	-	0.0040
MIP [8]	2009	SIBGRAP	-	-	2.6000
HE [9]	1997	TCE	-	-	0.0035
ICM [10]	2007	IAENG IJCS	-	-	0.1333
UDCP [11]	2013	ICCVW	-	-	2.6888
UCM [12]	2010	SMC	-	-	1.5555
FIVE_APLUS [16]	2023	BMVC	0.58	0.01	0.0666
PUIE-Net [17]	2022	ECCV	30.09	1.40	0.1111
WaterNet [13]	2019	TIP	71.42	1.09	0.0111
UIEC2Net [15]	2021	SPIC	26.06	0.53	0.0888
SCNet [23]	2022	ICASSP	23.56	0.82	0.0222
GUPDM [18]	2023	ACM MM	95.80	1.49	0.4888
Shallow-UWnet [22]	2021	AAAI	21.63	0.22	0.0651
UWCNN [14]	2020	PR	2.61	0.04	0.0666
CCMSRNet [20]	2023	TGRS	43.60	21.13	0.0800
USUIR [19]	2022	AAAI	14.81	0.23	0.0528
SFGNet [21]	2023	TIP	81.58	1.30	0.0228
UGAN [24]	2018	ICRA	19.82	57.17	0.1072
GuidedHybSensUIR [25]	2025	TCSVT	10.05	1.15	0.1111
Ours (WCI-Net)	-	-	17.55	5.65	0.0444

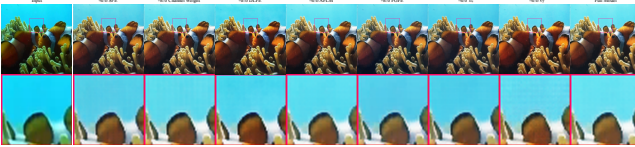


Fig. 5. Visual comparison of ablation study variants on the T-90 dataset.

all variants, confirming the synergistic contributions of each component.

TABLE III
ABLATION STUDY OF INDIVIDUAL COMPONENTS AND LOSS FUNCTIONS ON THE T-90 DATASET

Method	Metric		
	PSNR ↑	SSIM ↑	LPIPS ↓
-w/o SFE	22.69	0.891	0.0819
-w/o Channel Weight	23.77	0.908	0.0689
-w/o DLFE	21.61	0.872	0.0959
-w/o APLM	24.06	0.909	0.0682
-w/o FGFE	23.49	0.908	0.0721
-w/o $\mathcal{L}_c$	22.76	0.902	0.0697
-w/o $\mathcal{L}_f$	23.44	0.881	0.0808
Full Model	24.37	0.917	0.0684

IV. CONCLUSION

In this paper, we proposed WCI-Net for underwater image enhancement. By leveraging DWT for local feature capture and FFT for global property modeling, we designed the CIM module to enable cross-domain interaction between wavelet and Fourier representations. Extensive experiments demonstrate that WCI-Net effectively corrects color distortions and restores fine-grained details with acceptable computational cost, providing a powerful and generalizable solution for real-world marine vision applications.

REFERENCES

- [1] O. Almutiry, K. Iqbal, S. Hussain, A. Mahmood, and H. Dhahri, "Underwater images contrast enhancement and its challenges: a survey,"

- Multimedia Tools and Applications, vol. 83, no. 5, pp. 15125–15150, 2024.
- [2] F. Zhao, J. Li, J. Zhao, and J. Feng, “Weakly supervised phrase localization with multi-scale anchored transformer network,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5696–5705, 2018.
- [3] J. Zhao, Y. Cheng, Y. Cheng, Y. Yang, F. Zhao, J. Li, H. Liu, S. Yan, and J. Feng, “Look across elapse: Disentangled representation learning and photorealistic cross-age face synthesis for age-invariant face recognition,” Proceedings of the AAAI conference on artificial intelligence, pp. 9251–9258, 2019.
- [4] F. Zhang, S. You, Y. Li, and Y. Fu, “Atlantis: Enabling underwater depth estimation with stable diffusion,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11852–11861, 2024.
- [5] Y. Peng and P. C. Cosman, “Underwater image restoration based on image blurriness and light absorption,” IEEE Transactions on Image Processing, vol. 26, no. 4, pp. 1579–1594, 2017.
- [6] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 12, pp. 2341–2353, 2010.
- [7] 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.
- [8] 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, pp. 31–44, 2009.
- [9] Y.-T. Kim, “Contrast enhancement using brightness preserving bi-histogram equalization,” IEEE Transactions on Consumer Electronics, vol. 43, no. 1, pp. 1–8, 1997.
- [10] 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, vol. 3, pp. 490–494, 2009.
- [11] D. Huang, Y. Wang, W. Song, J. Sequeira, and S. Mavromatis, “Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition,” MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5–7, 2018, Proceedings, Part I 24, pp. 453–465, 2018.
- [12] K. Iqbal, M. Odetayo, A. James, R. A. Salam, and A. Z. H. Talib, “Enhancing the low quality images using unsupervised colour correction method,” 2010 IEEE International Conference on Systems, Man and Cybernetics, pp. 1703–1709, 2010.
- [13] 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.
- [14] C. Li, S. Anwar, and F. Porikli, “Underwater scene prior inspired deep underwater image and video enhancement,” Pattern Recognition, vol. 98, p. 107038, 2020.
- [15] Y. Wang, J. Guo, H. Gao, and H. Yue, “UIEC 2-Net: CNN-based underwater image enhancement using two color space,” Signal Processing: Image Communication, vol. 96, p. 116250, 2021.
- [16] J. Jiang, T. Ye, J. Bai, S. Chen, W. Chai, S. Jun, Y. Liu, and E. Chen, “Five A+ Network: You Only Need 9K Parameters for Underwater Image Enhancement,” arXiv preprint arXiv:2305.08824, 2023.
- [17] Z. Fu, W. Wang, Y. Huang, X. Ding, and K.-K. Ma, “Uncertainty inspired underwater image enhancement,” *European Conference on Computer Vision*, pp. 465–482, 2022.
- [18] P. Mu, H. Xu, Z. Liu, Z. Wang, S. Chan, and C. Bai, “A generalized physical-knowledge-guided dynamic model for underwater image enhancement,” Proceedings of the 31st ACM International Conference on Multimedia, pp. 7111–7120, 2023.
- [19] Z. Fu, H. Lin, Y. Yang, S. Chai, L. Sun, Y. Huang, X. Ding, “Unsupervised underwater image restoration: From a homology perspective,” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, 2022.
- [20] H. Qi, H. Zhou, J. Dong, and X. Dong, “Deep color-corrected multiscale retinex network for underwater image enhancement,” IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1–13, 2023.
- [21] C. Zhao, W. Cai, C. Dong, and Z. Zeng, “Toward Sufficient Spatial-Frequency Interaction for Gradient-aware Underwater Image Enhancement,” arXiv preprint arXiv:2309.04089, 2023.
- [22] A. Naik, A. Swarnakar, and K. Mittal, “Shallow-uwnet: Compressed model for underwater image enhancement (student abstract),” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 18, pp. 15853–15854, 2021.
- [23] J.-J. Liu, Q. Hou, M.-M. Cheng, C. Wang, and J. Feng, “Improving Convolutional Networks with Self-Calibrated Convolutions,” IEEE CVPR, 2020.
- [24] C. Fabbri, M. J. Islam, and J. Sattar, “Enhancing underwater imagery using generative adversarial networks,” 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 7159–7165, 2018.
- [25] X. Guo, X. Chen, S. Wang, and C.-M. Pun, “Underwater Image Restoration Through a Prior Guided Hybrid Sense Approach and Extensive Benchmark Analysis,” IEEE Transactions on Circuits and Systems for Video Technology, 2025.
- [26] Y. Cui, W. Ren, X. Cao, and A. Knoll, “Image restoration via frequency selection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no. 2, pp. 1093–1108, 2023.
- [27] H. Shen, C. Wang, L. Deng, L. He, X. Lu, M. Shao, and D. Meng, “FDDN: frequency-guided network for single image dehazing,” Neural Computing and Applications, vol. 35, no. 25, pp. 18309–18324, 2023.
- [28] W. Yi, L. Dong, M. Liu, M. Hui, L. Kong, and Y. Zhao, “Frequency-guidance collaborative triple-branch network for single image dehazing,” Displays, vol. 80, p. 102577, 2023.
- [29] Y. Wen, T. Gao, Z. Li, J. Zhang, K. Zhang, and T. Chen, “All-in-one Weather-degraded Image Restoration via Adaptive Degradation-aware Self-prompting Model,” IEEE Transactions on Multimedia, 2025.
- [30] H. Shen, Z.-Q. Zhao, Y. Zhang, and Z. Zhang, “Mutual information-driven triple interaction network for efficient image dehazing,” *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 7–16, 2023.
- [31] H. Yu, N. Zheng, M. Zhou, J. Huang, Z. Xiao, and F. Zhao, “Frequency and spatial dual guidance for image dehazing,” European Conference on Computer Vision, pp. 181–198, 2022.
- [32] X. Song, D. Zhou, W. Li, H. Ding, Y. Dai, L. Zhang, “WSAMF-Net: Wavelet spatial attention-based MultiStream feedback network for single image dehazing,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 33, no. 2, pp. 575–588, 2022.
- [33] Z. Li, Z.-S. Kuang, Z.-L. Zhu, H.-P. Wang, and X.-L. Shao, “Wavelet-based texture reformation network for image super-resolution,” IEEE Transactions on Image Processing, vol. 31, pp. 2647–2660, 2022.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] 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.
- [41] P. Guo, H. Liu, D. Zeng, T. Xiang, L. Li and K. Gu, “An Underwater Image Quality Assessment Metric,” in IEEE Transactions on Multimedia, vol. 25, pp. 5093–5106, 2023.
- [42] J. Canny, “A computational approach to edge detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, 1986.