DRM-Net: Explicit Residual Modelling with Subaquatic Multi-Scale Context Fusion for Underwater Image Enhancement

Chang Huang[1,2], **Zhexin Zhou**[2], **Jun Ma**[1], **Jiatong Shen**[1], **Peixuan Xiong**[1], **Huayong Yang**[2], **Kaishun Wu**[1] *

[1]The Hong Kong University of Science and Technology (Guangzhou)
[2]Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou)
chuang932@connect.hkust-gz.edu.cn, 2023022473@m.scnu.edu.cn, eejma@hkust-gz.edu.cn,
jshen854@connect.hkust-gz.edu.cn, pxiong843@connect.hkust-gz.edu.cn, yanghy@gmlab.ac.cn, wuks@hkust-gz.edu.cn

Abstract

Clear and high-quality underwater images are essential for marine applications, including autonomous navigation, ecological monitoring, and infrastructure inspection. However, underwater images typically suffer from severe colour distortion, low contrast, and diminished structural visibility due to wavelength-dependent attenuation, scattering, and uneven illumination conditions. Recent deep learning-based underwater image enhancement (UIE) methods primarily adopt end-to-end frameworks, directly regressing enhanced images from degraded inputs. While these approaches have achieved significant progress, they often lack explicit modeling of the degradation process, leading to limited interpretability and suboptimal recovery of fine-grained details. To address these limitations, we propose DRM-Net, an explicit residual learning framework for UIE. Rather than estimating the enhanced image directly, DRM-Net first predicts a pixel-wise Degradation Residual Map (DRM) in the perceptually uniform CIELab colour space. This map explicitly quantifies local colour, contrast, and structural degradations, thereby enabling the network to precisely reconstruct missing visual information. Furthermore, we design a lightweight Subaquatic Multi-Scale Context Fusion module, which utilizes parallel atrous convolutions with softmax-weighted feature aggregation, significantly enhancing robustness against spatially heterogeneous scattering. Trained jointly with pixel-wise DRM and VGG-based perceptual losses, DRM-Net achieves superior colour fidelity, perceptual realism, and structural detail recovery. Comprehensive experiments conducted on multiple benchmarks demonstrate that our proposed approach attains competitive quantitative results and superior qualitative visual performance compared to state-of-the-art UIE methods, while maintaining low computational overhead, making it particularly suitable for resource-constrained underwater robotic systems.

Introduction

The increasing demand for marine resource exploitation and ocean environmental protection has driven significant advancements in efficient underwater monitoring and perception technologies. Compared to active sensing modalities such as sonar and lidar, optical cameras provide a cost-effective, high-resolution alternative that offers rich visual semantic information crucial for underwater exploration (Afzal et al. 2022; Schettini and Corchs 2010; Raveendran, Patil, and Birajdar 2021). Nevertheless, underwater imagery suffers substantially from colour distortion, reduced contrast, and degraded structural details due to the absorption and scattering of light as well as suspended particulates in water, severely affecting the effectiveness of visual interpretation and subsequent intelligent analysis tasks (Islam, Xia, and Sattar 2020). In response, underwater image enhancement (UIE) has emerged as a critical research area within computer vision. Existing methods roughly fall into two categories: traditional models (He, Sun, and Tang 2010; Akkaynak and Treibitz 2018; Schechner and Karpel 2005) and machine learning-based techniques (Jian et al. 2021; Cong et al. 2024). Traditional physics-based approaches, grounded on optical imaging theory, typically estimate backscatter and transmission maps to compensate for degradation. Although effective under specific assumptions, these methods often rely on consistent lighting conditions and homogeneous water properties, and thus struggle to handle spatially varying degradations (Chiang and Chen 2011).

Machine learning methods have recently demonstrated superior performance in the field of underwater image enhancement. This data-driven technique is mainly categorised into supervised and unsupervised learning. Supervised learning approaches such as WaterGAN (Li et al. 2017), UWCNN (Li, Anwar, and Porikli 2020), and UIE-Net (Wang et al. 2017) have significantly improved enhancement quality by learning direct mappings from degraded images to clear counterparts. However, these methods rely heavily on paired training data, limiting their generalisation capabilities. When it comes to the unsupervised learning methods, for instance, URD-UIE proposed by Zhu et al. (Zhu et al. 2023) avoids paired-data dependency but often introduces artifacts like unrealistic colours and structural distortions. Recent approaches incorporating Transformer architectures and diffusion models have further boosted performance but at the expense of increased computational complexity and reduced interpretability (Guan et al. 2023; Pramanick, Sarma, and Sur 2024; Guo et al. 2023; Tang, Kawasaki, and Iwaguchi 2023). Critically reviewing the literature reveals that existing methods predominantly focus on directly predicting enhanced images end-to-end without

*Corresponding author

explicitly modelling the image degradation process. This monolithic approach inevitably blends multiple sub-tasks, such as colour correction, contrast enhancement, and structural detail recovery, diluting gradients, complicating optimisation, and impairing generalisation. Moreover, the neglect of explicit degradation modelling reduces interpretability, making it challenging to diagnose and improve performance systematically.

To address these challenges, this paper proposes a novel underwater image enhancement method built upon the concept of a Degradation Residual Map (DRM). Specifically, Degradation Residual Map quantifies pixel-wise differences between degraded and clear reference images in terms of colour, contrast, and structural detail, explicitly modelling the visual information lost due to underwater conditions. We further design a specialised CNN network named DRM-Net, trained end-to-end, to accurately predict the Degradation Residual Map. Enhanced images are then reconstructed by superimposing the predicted Degradation Residual Map onto the degraded input, explicitly compensating for the underwater degradation. Experimental results on public benchmark datasets demonstrate that our proposed method achieves state-of-the-art enhancement performance and exhibits visual consistency and robustness in real-world underwater scenarios. Furthermore, the explicit introduction of DRM significantly enhances the interpretability of the enhancement process, opening a new avenue for future research in the area of underwater image enhancement. Overall, our main contributions can be summarised as follows:

- **Degradation Residual Map:** We reformulate underwater enhancement as explicit residual prediction in the CIELab domain, where DRM encodes pixel-wise discrepancies between raw and latent clean images.
- **Subaquatic Multi-Scale Context Fusion Module:** We design a Subaquatic Multi-Scale Context Fusion module that adaptively fuses near- and far-field features via soft-weighted atrous branches, thereby boosting robustness to heterogeneous underwater scattering.
- **Dual-loss training and extensive validation:** We adopt a dual-loss strategy, combining pixel-level DRM L1 loss with a VGG-based perceptual loss. Additionally, comprehensive experiments and ablation settings demonstrate superior performance over state-of-the-art DL-based baselines.

Related Work

Underwater image enhancement methods are generally divided into model-based, physics-based and deep learning-based approaches. Model-based methods model underwater optical degradation or utilise signal priors, but their performance is limited by simplistic and idealised assumptions like uniform lighting and homogeneous water, which hinder generalisation in complex conditions such as strong scattering or uneven illumination (He, Sun, and Tang 2010; Akkaynak and Treibitz 2018; Schechner and Karpel 2005; Jian et al. 2021; Cong et al. 2024; Chiang and Chen 2011). With the advance of deep learning, CNNs have significantly improved enhancement performance (Wang et al. 2017). However, these black-box models often lack physical interpretability and may introduce artifacts that violate optical principles, such as unrealistic colour mapping or texture distortion (Yu, Qu, and Hong 2018). To address this, physics-guided deep learning has emerged as a promising direction to improve both interpretability and robustness. Saini et al. (Saini, Singh, and Anand 2025) proposed a Minimum Distortion Dehazing module that embeds the underwater optical model into a lightweight network, including background light and transmittance. MonoUIR (Tong, Zhang, and Zhang 2022) introduced a physically accurate imaging framework. It first generates depth maps via a monocular depth estimation network, then applies nonlinear least squares fitting to estimate the backscattering component. The result is passed into a dual-branch network, where the backbone predicts the transmittance map and an auxiliary branch corrects red-channel bias, thereby enhancing the final image. Chai et al. (Chai et al. 2022) proposed a four-branch network that jointly estimates scene radiance, transmittance, and background light. By integrating self-supervised loss and colour constancy constraints, the method achieves synergistic optimisation between the physical degradation model and the network. ADCNet (Yang, Li, and Li 2023) adopts a reflection-aware framework to separate background light and transmittance, producing a de-scattered representation. Both the original and de-scattered images are then fed into a multi-scale feature fusion module to enhance texture details. In Prompt-UIE (Zhang et al. 2025), the Reverse Transmission Map guides the loss function to focus on non-uniformly degraded regions. Wu et al. (Wu et al. 2022) incorporates attention mechanisms into the co-feature fusion and Specific Feature Extraction modules to enforce consistency constraints on depth and chromatic aberration features. ITW-DehazeFormer (Wang et al. 2025) designs a soft reconstruction formula based on learnable parameters—transmittance and background light derived from the physical model. The third approach focuses on learning the degradation paradigm of underwater images through a data-driven approach and thus enhancement. HIFI-Net (Zhuang et al. 2025) employs Haar wavelet transform to decompose the image into one low-frequency and three high-frequency bands, capturing structural and detailed features. Focusing on edges, WaterSNet (Chen et al. 2022) introduces a multi-scale feature extraction and cross-domain attention module for adversarial frequency-domain learning. Wavelet-based Dual-Stream (Ma and Oh 2022) processes RGB, HSV, and Lab spaces in parallel, applying low-frequency features for colour correction and high-frequency residuals for edge enhancement, reinforced by adversarial learning. Built on Discrete Wavelet Transform, the Attention-Based Spatial-Frequency Network (Pramanick, Megha, and Sur 2024) extracts four subbands to suppress redundancy. Efficient-USR (Pramanick, Bheda, and Sur 2025) combines depthwise separable convolutions with FFT to extract spatial textures and frequency cues.

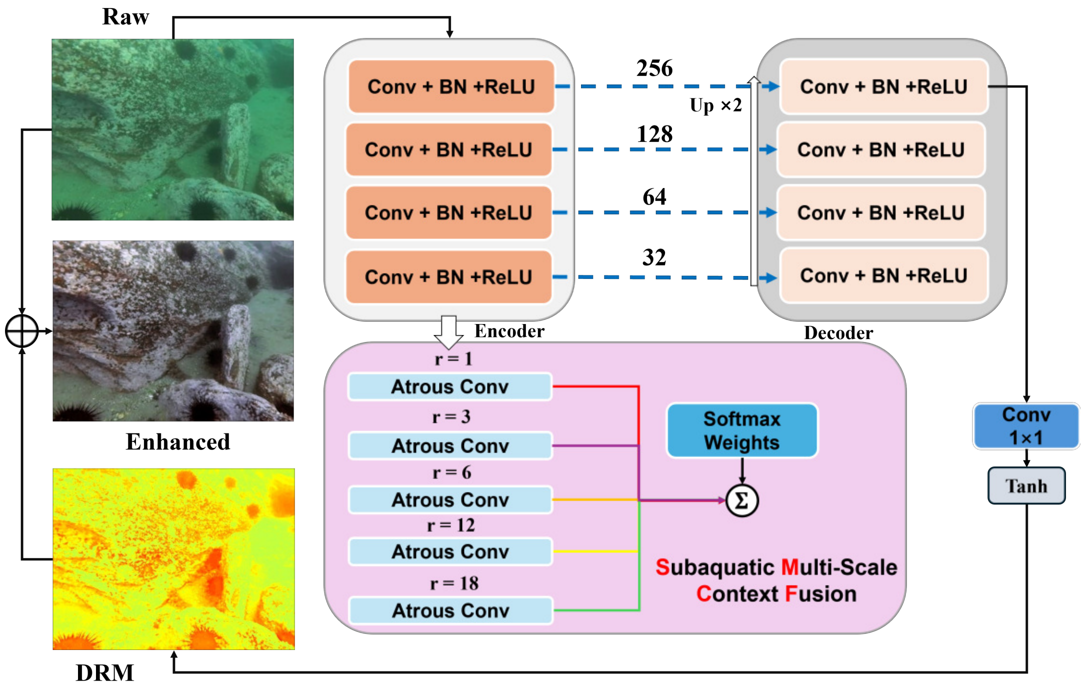

Figure 1: Overview of the proposed DRM-Net framework. Given a raw underwater image, the encoder-decoder architecture explicitly predicts a Degradation Residual Map in the CIELab domain. At the bottleneck, the SMCF module adaptively integrates multi-scale contextual information using parallel atrous convolutions with learnable softmax-normalized weights. The final enhanced image is obtained by adding the predicted DRM back to the original degraded input.

Methodology

Overview

This section presents an overview of the proposed DRM-Net framework, as illustrated in Figure 1. The Figure depicts the overall pipeline of DRM-Net, whose motivation is to enhance underwater images by explicitly learning a Degradation Residual Map. The network is organised in three consecutive stages: an encoder, the Subaquatic Multi-Scale Context Fusion module, and a decoder. The degraded input image is first processed by a four-stage convolutional encoder that progressively downsamples spatial resolution while distilling increasingly abstract features. The deepest feature tensor is then fed into the SMCF module, which deploys parallel atrous-convolution branches with distinct dilation rates to perceive multi-scale "water-layer" contexts. A soft-max–normalised set of learnable weights adaptively fuses these branches, yielding a context-enriched representation that is highly responsive to non-uniform scattering and wavelength-dependent colour shifts. Subsequently, a decoder with three transposed-convolution stages upsamples the fused features. Skip connections concatenate encoder activations of matching resolutions, facilitating the recovery of fine spatial details. The enhanced image is then reconstructed, thereby explicitly compensating for the lost visual information. The entire network is trained end-to-end under the joint supervision of pixel-wise DRM loss and perceptual loss, achieving coherent restoration of both local textures and global semantics across diverse underwater conditions.

Generation of the Degradation Residual Map

To explicitly quantify differences between underwater degraded images and clear reference images, we introduce the concept of a Degradation Residual Map (DRM). Specifically, given a degraded underwater image $\mathrm{I}_{\mathrm{raw}}$ and its corresponding clear reference image $\mathrm{I}_{\mathrm{ref}}$, DRM is calculated to capture pixel-wise discrepancies in terms of colour, contrast, and structural details. In practical implementation, we first convert both input images from the standard RGB colour space into the CIELab colour space, which is known for its perceptual uniformity and closer alignment with the human visual system (Luo, Cui, and Li 2006; Reinhard et al. 2008). This conversion effectively highlights the perceptual differences in luminance and colour. Formally, the conversion is described as follows:

$$I_{\mathrm{Lab}} = \mathrm{RGB2Lab}(I_{\mathrm{RGB}}) \tag{1}$$

Subsequently, the DRM is computed by subtracting the degraded image from the clear reference image in the Lab domain, resulting in a preliminary residual map DRM' :

$$\mathrm{DRM}' = \frac{I_{\mathrm{ref}}^{\mathrm{Lab}} - I_{\mathrm{raw}}^{\mathrm{Lab}}}{255} \tag{2}$$

To ensure numerical stability and prevent extreme outlier values during model training, we apply a clipping operation

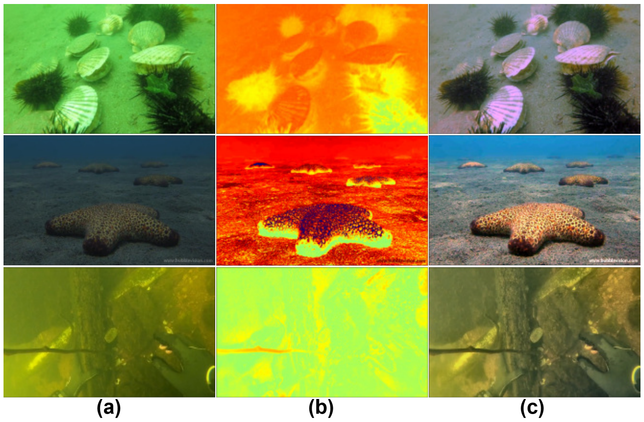

Figure 2: Examples of Degradation Residual Map visualisation. (a) Original raw images; (b) Generated DRMs, intuitively representing the degradation levels: regions depicted in red denote areas of significant degradation typically corresponding to object edges and detailed structures, while regions in yellow-green indicate areas with relatively minor degradation, usually uniform or background regions; (c) Corresponding reference images.

to constrain the DRM values within the interval $[-1, 1]$:

$$\mathrm{DRM} = \mathrm{clip}(\mathrm{DRM}', -1, 1) \tag{3}$$

where $\mathrm{clip}(\mathrm{x}, \mathrm{a}, \mathrm{b})$ confines values x within the specified range $[\mathrm{a}, \mathrm{b}]$. During training, this computed DRM serves as a supervisory signal, enabling the network to explicitly learn and reconstruct visual information lost due to underwater degradation. Moreover, for intuitive interpretation and visualization, we further define the norm of DRM as:

$$\mathrm{DRM}_{\text{vis}}(x, y) = \|\mathrm{DRM}(x, y)\|_2 \tag{4}$$

This norm effectively illustrates spatial variations of degradation intensities across different image regions. A pseudocolor mapping scheme (e.g., Jet colormap (Borland and Ii 2007)) is then applied to facilitate visual analysis.

As illustrated in Figure 2, column (a) shows original degraded underwater images exhibiting typical issues such as color distortion, low contrast, and loss of structural details. Column (b) visualises the proposed DRM, clearly depicting pixel-wise degradation intensity between degraded and reference images. Specifically, regions highlighted in red correspond to substantial degradation, prominently located at object contours, detailed textures, and significant colour distortions, whereas regions colored yellow-green reflect areas of comparatively minor degradation, typically background or homogeneous areas. This visualisation demonstrates how DRM can guide the restoration process by highlighting the spatial distribution of degradation severity. Column (c) shows corresponding clear reference images (ground truth), which serve as training targets and evaluation benchmarks for the enhancement method proposed in this work.

Subaquatic Multi-Scale Context Fusion Module

To address the spatially heterogeneous nature of underwater degradation, we design the Subaquatic Multi-Scale Context Fusion (SMCF) Module. Inspired by atrous spatial pyramid pooling (Chen et al. 2017), SMCF is re-engineered to cope with underwater-specific challenges such as non-uniform illumination, scattering, and wavelength-dependent colour cast. Given an input feature map $\mathrm{F} \in \mathcal{R}^{\mathrm{C}\times\mathrm{H}\times\mathrm{W}}$, SMCF builds K parallel water-layer perception branches $\phi_{\mathrm{k}}(\cdot)$, each employing an atrous convolution with dilation rate $\mathrm{r_k}$ to perceive contextual cues ranging from near-field to far-field depths. A learnable scale-attention mechanism adaptively fuses these branches:

$$\mathbf{F}_{\text{out}} = \psi\left(\sum_{k=1}^{K} w_k\, \phi_k(F)\right), \qquad w_k = \frac{\exp(\alpha_k)}{\sum_{j=1}^{K}\exp(\alpha_j)} \tag{5}$$

where α_{k} are trainable scalars and $\psi(\cdot)$ denotes a 1×1 projection convolution for recombination and channel compression. Unlike conventional ASPP, which merely concatenates multi-scale features, SMCF employs Softmax–normalised learnable weights to dynamically emphasize the most discriminative "water-layer" scales.

Loss Function

To effectively train our proposed DRM-based enhancement network, we design a composite loss function integrating pixel-wise loss and perceptual loss. Formally, the overall loss function is defined as:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{DRM}}\mathcal{L}_{\text{DRM}} + \lambda_{\text{PERC}}\mathcal{L}_{\text{PERC}} \tag{6}$$

where λ_{DRM} and $\lambda_{\mathrm{PERC}}$are hyperparameters balancing the contributions of respective loss terms during training.

Pixel-wise DRM Loss: This loss term supervises the discrepancy between predicted DRM ($\mathrm{DRM_{pred}}$) and ground-truth DRM ($\mathrm{DRM_{gt}}$), computed using the L1 loss to ensure training stability:

$$\mathcal{L}_{\text{DRM}} = \frac{1}{N}\sum_{i=1}^{N}\left\|\mathrm{DRM}_{\text{pred}}^{(i)} - \mathrm{DRM}_{\text{gt}}^{(i)}\right\|_1 \tag{7}$$

where N denotes the total number of pixels. The L1 loss effectively mitigates the influence of outliers during DRM prediction.

Perceptual Loss: To further improve the perceptual quality of enhanced images, we introduce a perceptual loss (Johnson, Alahi, and Fei-Fei 2016) based on a pretrained VGG network (Simonyan and Zisserman 2014). Perceptual loss quantifies differences in deep feature representations between the enhanced image ($\mathrm{I_{enh}}$) and the ground-truth clear image ($\mathrm{I_{gt}}$). Specifically, images are first normalised. Then, deep features $\mathrm{f}(\cdot)$ are extracted from the normalised images using the VGG model, and the perceptual loss is calculated as the L1 difference between these features:

$$\mathcal{L}_{\text{PERC}} = \|f(I_{\text{enh}}) - f(I_{\text{gt}})\|_1 \tag{8}$$

This perceptual loss better preserves structural details significant to human visual perception, resulting in more visually natural and appealing enhanced images.

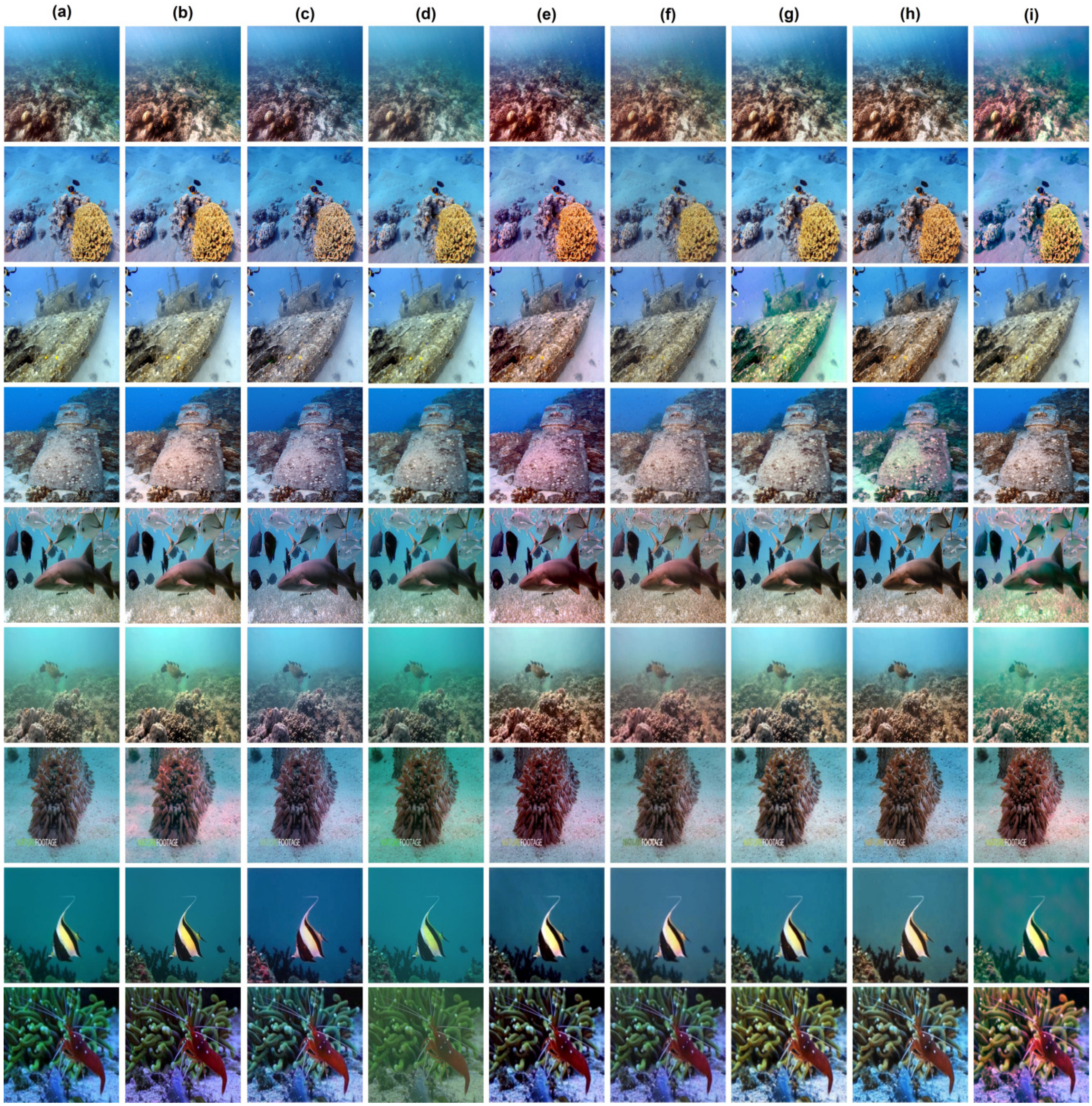

Figure 3: Visual comparison with different methods on real-world datasets. The underwater images and their corresponding enhanced images from the following methods are presented. Besides, the raw and reference images are displayed in the first two columns. (a) Raw (Input). (b) Reference (Ground Truth). (c) WaterNet. (d) DM_Underwater. (e) SFG-Net. (f) U-Shape. (g) URanker. (h) UIR-PolyKernerl. (i) DRM-Net (Ours). Compared to other approaches, our proposed method performs better in terms of colour restoration, detail retention and noise removal, achieving a better visual quality.

Method	UIEB			LSUI			Overhead		
	PSNR	**SSIM**	**MSE**	**PSNR**	**SSIM**	**MSE**	**GFLOPs**	**Par. (M)**	**Inference (ms)**
WaterNet	19.98	0.820	1280.61	20.04	0.804	720.64	8.28	**1.09**	30.96
UIR-PolyKernel	20.60	0.875	759.17	19.87	0.796	739.97	13.70	1.84	15.06
URanker	23.16	0.896	415.57	20.24	0.807	747.74	10.48	3.15	4.38
SFG-Net	17.36	0.774	1761.53	21.69	0.792	476.00	81.57	1.30	44.75
DM_Underwater	25.47	0.905	297.25	20.41	0.837	775.22	133.79	18.34	431.74
U-Shape	21.05	0.772	754.29	23.86	0.845	371.29	**2.98**	22.82	46.20
DRM-Net (Ours)	**25.59**	**0.912**	**259.79**	**24.14**	**0.878**	**306.85**	6.77	1.47	**1.02**

Table 1: Quantitative comparison with state-of-the-art methods on UIEB and LSUI benchmarks in terms of image quality and computational overhead. Bold values are best, while underlined values are second best. Image quality metrics include PSNR, SSIM and MSE. The Overhead metric contains FLOPs (G), Parameters (M), Inference Time (ms).

Component		Metrics		
$\mathcal{L}_{\text{DRM}}$	$\mathcal{L}_{\text{PERC}}$	PSNR	SSIM	MSE
✓	✓	25.59	0.912	259.79
✓	-	22.52	0.871	487.28
-	✓	21.63	0.789	720.15
-	-	21.74	0.850	503.59

Table 2: Ablation study of individual loss components in our hybrid loss design on UIEB dataset. The last row adopts the MSE loss as a baseline substitute when all components are removed. The full combination of all losses yields the best performance, confirming their complementary effects.

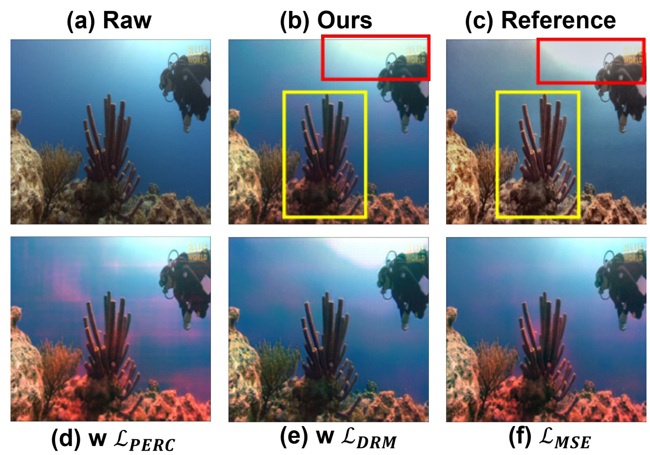

Figure 4: Visualisation for ablation study of loss components on UIEB dataset. The results demonstrate the significance of the proposed loss components.

Experiments

Implementation Details

We implement the proposed DRM-Net with PyTorch using a commodity workstation with a single GTX-4080 GPU. In our experiments, we set the maximum number of training epochs to 400 and the batch size to 32. We use the Adam optimiser and set the initial learning rate to 0.001.

Datasets

We conduct experiments using two widely recognised publicly available underwater image enhancement datasets: UIEB (Li et al. 2019) and LSUI (Peng, Zhu, and Bian 2023). The complementary characteristics of these datasets, in terms of scale, scenario diversity, and degradation complexity, allow for a comprehensive evaluation of the effectiveness and generalisation of our proposed method. Additionally, both the UIEB and LSUI datasets are randomly partitioned into training and testing sets. For standardised processing and improved training efficiency, all images from both UIEB and LSUI datasets are uniformly resized to a fixed dimension (256×256) before inputting into the network, and the pixel value will be normalised to $[0, 1]$.

Evaluation Metrics

To comprehensively and objectively evaluate the proposed method's performance, we employ widely used metrics for image enhancement: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Mean Squared Error (MSE).

Results and Analysis

Benchmarks We compare our proposed method with a variety of representative DL-based underwater image enhancement approaches, including WaterNet (Li et al. 2019), UIR-PolyKernel (Guo et al. 2025), URanker (Guo et al. 2023), SFG-Net (Zhao et al. 2024), DM_Underwater (Tang, Kawasaki, and Iwaguchi 2023) and U-Shape (Peng, Zhu, and Bian 2023). All compared methods are reproduced using the same implementations on the UIEB and LSUI datasets to ensure fair comparison.

Quantitative comparison: Table 1 presents the quantitative evaluation results of DRM-Net and several state-of-the-art underwater image enhancement methods on the UIEB and LSUI datasets. Evaluation metrics are all full-reference measures (PSNR, SSIM and MSE).

As summarised in Table 1, DRM-Net presents a comprehensive quantitative evaluation of DRM-Net against recent state-of-the-art underwater enhancement approaches on UIEB and LSUI datasets. DRM-Net consistently achieves top-ranking restoration performance, yielding improvements of 0.12 dB and 0.007 in PSNR and

Module	Metrics			Overhead		
SMCF	PSNR	SSIM	MSE	GFLOPs	Parameters(M)	Inference(ms)
✓	25.59	0.912	259.79	6.77	1.47	1.02
-	23.85	0.876	326.61	6.23	0.95	0.78

Table 3: Ablation study on the effectiveness of SMCF modules of DRM-Net on UIEB dataset. We evaluate the impact of this component on the quality of the enhancement and the overhead problem it poses.

SSIM, respectively, over the best-performing competitor (DM_Underwater) on the UIEB benchmark, accompanied by a 12% reduction in MSE. On the challenging LSUI dataset, DRM-Net surpasses the previous leading method (U-Shape) by 0.28 dB PSNR, 0.033 SSIM, and 17% MSE reduction. Notably, DRM-Net achieves this performance at remarkably low computational cost, requiring only 6.77 GFLOPs and 1.47 million parameters, while operating at an inference speed of merely 1.02 ms per image, significantly outperforming most existing approaches. These results demonstrate that DRM-Net provides an excellent balance between visual restoration quality and computational efficiency, making it particularly suitable for real-time deployment in underwater robotic and monitoring systems with limited computing resources.

Qualitative comparison: Figure 4 provides a side-by-side visual comparison of nine real-world underwater scenes. Visually, our method restores the missing red–orange spectrum, preserves coral and fish textures, and suppresses background haze far better than WaterNet, U-Shape and other baselines, which still exhibit colour casts or oversharpened artefacts. The combination of explicit residual learning and the SMCF module therefore delivers the highest fidelity in colour restoration, detail retention and noise suppression, confirming the practical superiority of DRM-Net.

Ablation Study

Loss Components: Table 2 reveals that employing the dual-loss configuration—pixel-wise DRM loss $\mathcal{L}$DRM combined with perceptual loss $\mathcal{L}_{\text{PERC}}$—yields the best restoration results, achieving a PSNR of 25.59 dB, an SSIM of 0.912, and the lowest MSE of 259.79. Compared with using either loss alone, the dual supervision improves PSNR and SSIM by roughly 3 dB and 0.04, respectively, while reducing the pixel error by 40–60%. This demonstrates the complementary roles of the two terms: $\mathcal{L}_{\text{DRM}}$ enforces pixel-level fidelity, whereas $\mathcal{L}_{\text{PERC}}$ promotes perceptual realism. Relying on a single loss biases the network toward only one of these objectives, and removing both losses leads to the poorest quality, confirming that each component provides indispensable guidance during training. Figure 4 visualises the ablation study of different loss functions on the UIEB dataset. Compared with using individual loss components, our dual-loss strategy combining DRM loss and perceptual loss clearly achieves results closest to the reference image in terms of colour fidelity, detail preservation, and structural clarity. Specifically, employing only perceptual loss tends to produce visually appealing but overly saturated results, introducing unwanted colour artifacts. Conversely, using only the DRM loss recovers accurate colour distributions but results in blurry textures and reduced perceptual quality. Relying solely on MSE loss severely degrades colour accuracy, introducing evident reddish colour bias and compromising structural details. The visual results thus confirm that our combined loss function effectively leverages the complementary strengths of both DRM and perceptual losses, ensuring superior visual realism and accurate restoration simultaneously.

Module: As summarised in Table 3, integrating the SMCF module yields a notable gain of +1.74 dB in PSNR and +0.036 in SSIM, while reducing MSE by roughly 20%. These improvements come at a modest cost of only 0.54 GFLOPs, 0.52 million additional parameters, and 0.24 ms extra inference time. The results demonstrate that SMCF offers a favourable trade-off between restoration quality and computational overhead, confirming its effectiveness as a pivotal component of our framework.

Discussion

From the Figure 4, our DRM-Net framework successfully recovers fine-grained structural textures and largely restores accurate colour distributions, particularly evident in the improved reconstruction of the challenging highlight regions near the top-right light source. However, certain limitations remain: for instance, the primary subjects within the enhanced images still exhibit slightly darker tonalities, suggesting room for improvement in adaptive brightness adjustment. Future work could focus on incorporating region-aware or adaptive tone mapping techniques to further enhance brightness consistency across various image regions

Conclusion

This work introduced DRM-Net, an explicit residual-learning framework for UIE, which explicitly predicts a pixel-wise DRM in the CIELab space, guiding the network to accurately restore lost colour, contrast, and structural details. We further designed a lightweight SMCF module to effectively address spatially heterogeneous scattering through adaptive multi-scale feature fusion. Joint supervision using pixel-level DRM and perceptual losses enabled the framework to achieve robust visual restoration with minimal computational overhead. Experiments on multiple benchmarks confirm that our approach attains competitive quantitative results and enhanced qualitative performance relative to SOTA methods. Future research will explore adaptive brightness correction strategies, real-time video applications, and hybrid enhancement schemes incorporating physical priors.

Acknowledgements

This work is supported by Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things (No.2023B1212010007), China NSFC Grant award number (No.62472366), 111 Center (No.D25008), the Project of DEGP (No.2024GCZX003, 2023KCXTD042), Shenzhen Science and Technology Foundation (ZDSYS20190902092853047) and The Hong Kong Branch of the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou).

References

Afzal, S. S.; Akbar, W.; Rodriguez, O.; Doumet, M.; Ha, U.; Ghaffarivardavagh, R.; and Adib, F. 2022. Battery-free wireless imaging of underwater environments. *Nature communications*, 13(1): 5546.

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

Borland, D.; and Ii, R. M. T. 2007. Rainbow color map (still) considered harmful. *IEEE computer graphics and applications*, 27(2): 14–17.

Chai, S.; Fu, Z.; Huang, Y.; Tu, X.; and Ding, X. 2022. Unsupervised and untrained underwater image restoration based on physical image formation model. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2774–2778. IEEE.

Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; and Yuille, A. L. 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4): 834–848.

Chen, R.; Fu, Z.; Huang, Y.; Cheng, E.; and Ding, X. 2022. A robust object segmentation network for underwater scenes. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2629–2633. IEEE.

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

Cong, X.; Zhao, Y.; Gui, J.; Hou, J.; and Tao, D. 2024. A comprehensive survey on underwater image enhancement based on deep learning. *arXiv preprint arXiv:2405.19684*.

Guan, M.; Xu, H.; Jiang, G.; Yu, M.; Chen, Y.; Luo, T.; and Zhang, X. 2023. DiffWater: Underwater image enhancement based on conditional denoising diffusion probabilistic model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17: 2319–2335.

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

Guo, X.; Dong, Y.; Chen, X.; Chen, W.; Li, Z.; Zheng, F.; and Pun, C.-M. 2025. Underwater image restoration via polymorphic large kernel cnns. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

He, K.; Sun, J.; and Tang, X. 2010. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12): 2341–2353.

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

Jian, M.; Liu, X.; Luo, H.; Lu, X.; Yu, H.; and Dong, J. 2021. Underwater image processing and analysis: A review. *Signal Processing: Image Communication*, 91: 116088.

Johnson, J.; Alahi, A.; and Fei-Fei, L. 2016. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, 694–711. Springer.

Li, C.; Anwar, S.; and Porikli, F. 2020. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern recognition*, 98: 107038.

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

Li, J.; Skinner, K. A.; Eustice, R. M.; and Johnson-Roberson, M. 2017. WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robotics and Automation letters*, 3(1): 387–394.

Luo, M. R.; Cui, G.; and Li, C. 2006. Uniform colour spaces based on CIECAM02 colour appearance model. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, 31(4): 320–330.

Ma, Z.; and Oh, C. 2022. A wavelet-based dual-stream network for underwater image enhancement. In *ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2769–2773. IEEE.

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

Pramanick, A.; Bheda, U.; and Sur, A. 2025. Efficient-USR: Prompt Guided Dual-Domain Feature Information for Efficient Underwater Image Super-Resolution. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Pramanick, A.; Megha, D.; and Sur, A. 2024. Attention-based spatial-frequency information network for underwater single image super-resolution. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3560–3564. IEEE.

Pramanick, A.; Sarma, S.; and Sur, A. 2024. X-caunet: Cross-color channel attention with underwater image-enhancing transformer. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3550–3554. IEEE.

Raveendran, S.; Patil, M. D.; and Birajdar, G. K. 2021. Underwater image enhancement: a comprehensive review, recent trends, challenges and applications. *Artificial Intelligence Review*, 54(7): 5413–5467.

Reinhard, E.; Khan, E. A.; Akyuz, A. O.; and Johnson, G. 2008. *Color imaging: fundamentals and applications*. CRC Press.

Saini, P.; Singh, N.; and Anand, A. S. 2025. AUIED3K: A New Andaman Underwater Image Enhancement Dataset for Deep Learning-Driven Image Enhancement with Minimum Loss Dehazing. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Schechner, Y. Y.; and Karpel, N. 2005. Recovery of underwater visibility and structure by polarization analysis. *IEEE Journal of oceanic engineering*, 30(3): 570–587.

Schettini, R.; and Corchs, S. 2010. Underwater image processing: state of the art of restoration and image enhancement methods. *EURASIP journal on advances in signal processing*, 2010(1): 746052.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

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

Tong, J.; Zhang, T.; and Zhang, L. 2022. Towards underwater image restoration: A physical-accurate pipeline and a large scale full-reference benchmark. In *2022 IEEE International Conference on Multimedia and Expo (ICME)*, 01–06. IEEE.

Wang, Q.; Zhong, X.; Lu, D.; and Tian, Y. 2025. ITW-DehazeFormer: Imaging through Turbid Water Using Improved DehazeFormer. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Wang, Y.; Zhang, J.; Cao, Y.; and Wang, Z. 2017. A deep CNN method for underwater image enhancement. In *2017 IEEE international conference on image processing (ICIP)*, 1382–1386. IEEE.

Wu, B.; Ye, X.; Xue, F.; and Xu, R. 2022. Learning Data Hallucination and Reciprocal Guidance for Underwater Depth Estimation and Color Correction. In *2022 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6. IEEE.

Yang, J.; Li, C.; and Li, X. 2023. Underwater image restoration with light-aware progressive network. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Yu, X.; Qu, Y.; and Hong, M. 2018. Underwater-GAN: Underwater image restoration via conditional generative adversarial network. In *International Conference on Pattern Recognition*, 66–75. Springer.

Zhang, Y.; Luo, L.; Mu, P.; and Bai, C. 2025. Prompt-UIE: A Unified Prompt-Driven Framework for Underwater Image Enhancement. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

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

Zhu, P.; Liu, Y.; Wen, Y.; Xu, M.; Fu, X.; and Liu, S. 2023. Unsupervised underwater image enhancement via content-style representation disentanglement. *Engineering Applications of Artificial Intelligence*, 126: 106866.

Zhuang, J.; Zhou, J.; Zheng, Y.; Chang, Y.; and Mazhar, S. 2025. Multi-domain fusion network for underwater image enhancement. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.