Conditional Prompt Learning via Degradation Perception for Underwater Image Enhancement

Mingze Yao, Zhiying Jiang*, Xianping Fu*, Huibing Wang

College of Information Science and Technology, Dalian Maritime University, Dalian, China

{ymz0284, zyjiang, fxp, huibing.wang}@dlmu.edu.cn

Abstract

Underwater Image Enhancement (UIE) focuses on improving visual quality from various underwater scenes. Existing methods simplistically treat various degradations as homogeneous, disregarding their intrinsic connections and causing models to blindly learn, resulting in conflicting optimization goals and visual distortions. To address above limitations, we propose a Conditional Prompt Learning via Degradation Perception (CPLDP) model, which employs conditional prompt as degradation perception priors and guides underwater image enhancement. Specifically, we show that the natural language prompts not only promote distinguishing different degraded images, but also aid in exploring more details with semantic information. Therefore, our method generates five key degradation prompts (green/blue/green-blue color casts, uneven illumination and haze) with conditional prompt learning. Subsequently, considering the intrinsic relationships among different degradations, we employ degradation perceptions as priors and fine-tune the learning strategy to enhance underwater images. During training, an adaptive loss function with multi-degradations is designed, allowing it to effectively handle the task conflicts among multiple underwater degradations. Additionally, we conduct a human visual-based underwater dataset with various degradation types by subjective statistics. Extensive experiments on both full-reference and non-reference datasets demonstrate that our CPLDP can achieve better visual results and outperforms state-of-the-art UIE methods across various degradation scenarios.

Code — https://github.com/yg76cv/CPLDP.

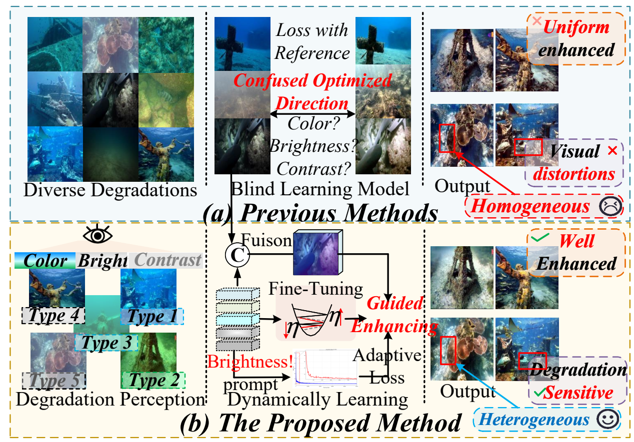


Figure 1: Comparison between previous deep learning-based UIE methods and our CPLDP. (a) Previous methods treated diverse degradations as a homogeneous phenomenon, while fail in ineffective optimization by indiscriminate degradation treatment with uniform results as visual distortions. (b) The proposed method leverages human visual-based degradation prompts and proposes dynamic learning model with prompt guidance, enabling to achieve sensitive degradation perception and heterogeneous enhancement in complex underwater scenarios.

Introduction

Underwater imaging plays a critical role in a wide range of marine applications, including ecological monitoring, underwater exploration, and aquaculture (Jiang et al. 2022b; Wang et al. 2024). However, underwater images often suffer from complex visual degradations such as color cast, uneven illumination, and haze, due to light absorption and scattering by water particles.hese complex degradations in underwater images hinder human perception of the underwater environment and pose substantial challenges for marine resource development., which motivates the need for robust underwater image enhancement (UIE) techniques (Han et al. 2018; Zhang et al. 2023b; Wang et al. 2025).

Existing UIE approaches can be broadly classified as traditional physics-based methods and deep learning-based methods. Traditional methods rely on assumptions about light propagation or scene priors to estimate and correct degradation (Ancuti et al. 2012; Li et al. 2016), but they often lack robustness and generalization in diverse underwater environments. In contrast, deep learning-based methods, utilizing CNNs and Transformers structures (Yao et al. 2025b; Liu et al. 2022; Jiang et al. 2024), has shown impressive performance improvements in UIE filed. Besides, Generative Adversarial Networks (GANs) and diffusion models (Hambarde, Murala, and Dhall 2021; Zhao et al. 2024) are also employed with considering the reliance on large-scale paired

*Corresponding Author.

datasets. However, they almost train indiscriminately on the whole underwater datasets with mixed-degradation strategy (Hou et al. 2023; Chen et al. 2024), while neglecting complex degradations interplay and fundamentally incompatible enhancement goals between different degradations types. As shown in Figure 1 (a), with a fundamental assumption that diverse degradations are regarded as a homogeneous phenomenon, while fail in confused optimized direction, resulting in suboptimal learning objectives and visual artifacts.

This "One-for-All" paradigm fundamentally conflicts with incompatible enhancement goals across distortions, inevitably compromising generalization. Therefore, it is worth considering whether it is reasonable to conduct image enhancement model with undifferentiated underwater degradations. As illustrated in Figure 1 (b), with the subjective statistics from human visual, we conduct an underwater dataset with various degradation perceptions. Meanwhile, our proposed method introduce corresponding prompts to dynamically guide enhancement and adaptively adjust loss function, enabling to effectively enhance image quality with fewer distortions and present various natural results.

To resolve this goal conflict rooted in various underwater degradations, we propose a Conditional Prompt Learning via Degradation Perception (CPLDP) model, designed to sufficiently perceive underwater image degradations and effectively enhance the image quality. By conducting subjective analysis on the underwater dataset, our model first generate conditional prompt formed by fusing the degradation-specific and positive reference prompts, guides the enhancement process more effectively. And then, we design a prompt-guided image enhancement network with conditional prompt as priors to jointly explore the enhance directions during training and preserve semantic details. In addition, to resolve conflicts from diverse degradations, the conditional prompts are enabling dynamic adaptation of learning strategy while flexibly adjust the loss function based on image content. Finally, we conduct a human visual-based underwater dataset by subjective statistics, which provide corresponding degradation types for each image.

Overall, our contribution can be summarized as follows:

- We develop the Conditional Prompt Learning via Degradation Perception (CPLDP) model, specifically designed by perceiving complex degradation patterns and natural reference as conditional prompts. which is able to address the challenge of diverse underwater degradations and effectively resolving conflicting optimization problems and better enhancing underwater images.
- In order to achieve conditional prompt guidance, we leverage a dynamic learning network with prompt fusion and fine-tuning to guide enhancement process. The degradation-specific prompts allow the network to adaptively focus on degradation-aware regions and flexibly fine-tuning learn enhancement strategy, resulting in visually natural and context-consistent results.
- A human visual-based underwater dataset is constructed for learning condition prompt, incorporating natural statistical subjective analysis. Besides, an adaptive adjustment loss is designed to prevent distortions caused by conflicting degradations, leading to robust optimization.

Related Work

Underwater Image Enhancement

Existing UIE methods are primarily categorized as physics-driven traditional and data-driven deep learning-based approaches. Traditional methods primarily focused on the mechanisms of imaging, enhancing underwater images through pixel adjustment or color parameter modification(Chiang and Chen 2011), modeling light scattering and medium transmission using physics-based mathematical formulations (Wang, Liu, and Chau 2017). Nevertheless, they mostly require pre-calibrated the underwater imaging models as priors, resulting in low robustness for dynamic underwater environments. Data-driven deep learning-based methods have gradually become the mainstream solution for UIE tasks (Panetta, Gao, and Agaian 2015; Tang, Kawasaki, and Iwaguchi 2023; Jiang et al. 2025). This is because their excellent feature extraction and nonlinear mapping capabilities, enabling them to effectively enhance image quality (Fu et al. 2022a; Peng and Bian 2025; Zhang et al. 2025b). The aforementioned progress notwithstanding, they typically treat heterogeneous underwater degradations indiscriminately (Zhang et al. 2023a; Shang et al. 2025). The simple assumption causes models to focus on the dominant degradation types within the datasets, such as color cast or low contrast, while overlooking the interplay between different degradations and conflicts among competing optimization objectives. Consequently, their performance remains limited in complex underwater scenarios characterized by multiple simultaneous degradations.

Prompt Learning in Vision

Prompt learning has gained widespread adoption due to its superior performance in vision-related tasks. By generating task-specific prompts as prior knowledge, these methods can guide models toward progressively refined fine-tuning, enabling them to better understand downstream tasks (Wang et al. 2023; Yao et al. 2025a). In computer vision, prompt learning is typically categorized into two types: Hard prompt (Wen et al. 2023; Conde, Geigle, and Timofte 2024) derived from textual descriptions of images, Soft prompt (Liang et al. 2023; Zhang et al. 2025a) learned from various inputs as optimizable vectors. Recently, considering the limitations of providing accurate textual descriptions for each image, PromptIR (Potlapalli et al. 2023) encodes degradation-specific information as soft prompt and integrates with image features for guiding image restoration. However, implicitly learning prompt information as dynamic feature level weights fails to fully capture the complex interdependencies among multiple degradations and conflicting optimization objectives across degradations. To overcome this, this paper proposes a conditional prompt learning with explicit degradation perception from constructed underwater dataset.

Method

As depicted in Figure 2, our proposed method contains two core components: conditional prompt learning and

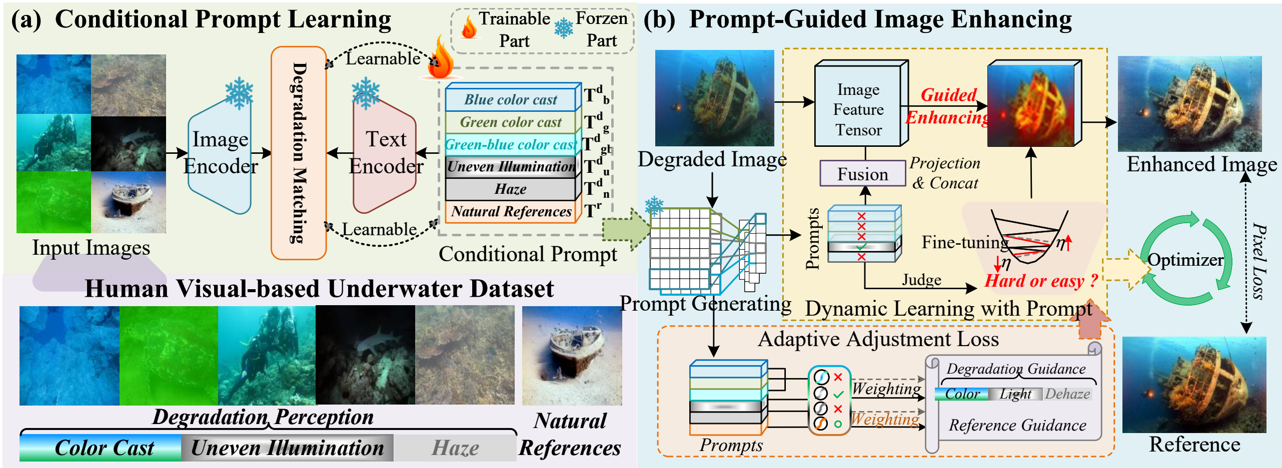


Figure 2: Overview of the proposed CPLDP. (a) Conditional Prompt Learning: A human visual-based underwater dataset is first constructed, where images are categorized by perceived degradation types. Prompt learning is then employed to extract reliable degradation-specific and reference prompts as guidance priors. (b) Prompt-Guided Image enhancing.: Leveraging the learned prompts, a dynamic learning network flexibly enhances degraded images with fusing prompt-image features, and an adaptive adjustment loss is introduced to enhance various degraded images, resulting in fewer distortions and more natural results.

prompt-guided image enhancing. Specifically, the conditional prompt learning module is designed to extract reliable degradation-specific and positive reference prompts as informative priors for guiding enhancement. Once trained, this frozen module generates degradation-specific prompts for each input image, which are subsequently fused with intermediate image features through a learnable projection and concat layer. This feature-prompt interaction forms a joint representation that enables the network to dynamically focus on degradation-relevant regions. Moreover, based on the inferred degradation types from prompts, our network conditionally fine-tunes its enhancement strategy to mitigate optimization conflicts among various degradations. Finally, an adaptive adjustment loss guided by the conditional prompt further improves enhancement performance with fewer distortions and natural results.

Besides, before training, we conduct a human visual-based underwater dataset with degradation perception for data preparation. Underwater images are divided into three main degradation types as color cast, uneven illumination and haze. The color cast is further classified into blue/green/green-blue color cast.

Conditional Prompt Learning

In conditional prompt learning, various degraded images and natural references are provided to learn degradation-specific and positive reference prompts respectively, so as to distinguish between underwater images with different degradation types and high quality references from semantic perspective.

Degradation Prompts Learning. Given the degraded underwater images $\mathbf{I}_i^d \in \mathbb{R}^{H\times W\times 3}$, we randomly initialize a degradation-specific prompt embeddings as $\mathbf{T}_i^d \in \mathbb{R}^{1\times N\times 512}$. N represents the length of embedded tokens in each gradation prompts and $i \in \{b, g, gb, u, n\}$ denotes the degradation types. Different from hard prompt with language descriptions, we utilize the frozen image and text encoders ($\Phi_{\text{image}}, \Phi_{\text{text}}$) to extract soft prompt from input degraded images and initial prompts for matching degradation correspondences. In order to enable learnable degradation-specific prompts to capture sufficient information from input images, we employ the cross entropy loss $\mathcal{L}_{dp}$ to classify different degradations with degraded underwater images. The $\mathcal{L}_{dp}$ can be expressed as:

$$\mathcal{L}_{dp} = -\frac{1}{N_p}\sum_{i\in G}\sum_{j\in G}\mathbf{Y}_{ij}\log(\frac{\exp(\mathcal{S}(i,j))}{\sum\limits_{k\in G}\exp(\mathcal{S}(i,k))}), \quad (1)$$

where N_p is the prompt batch size and G is batch set. $\mathbf{Y}_{ij}$ is the degradation label of degraded image. The similarity function $S(\cdot)$ can be described as:

$$\mathcal{S}(i,j) = \frac{\mathbf{X}_i^d \cdot \mathbf{Y}_j^{d^T}}{\|\mathbf{X}_i^d\| \cdot \|\mathbf{Y}_j^{d^T}\|}, \quad (2)$$

where image features are extracted as $\mathbf{X}_i^d = \Phi_{\text{image}}(\mathbf{I}_i^d)$ and prompt vectors are obtained as $\mathbf{Y}_j^d = \Phi_{\text{text}}(\mathbf{T}_j^d)$.

Positive Prompts Learning. Inspired by contrastive learning, we utilize natural references to learn positive prompts for guiding the optimization direction of subsequent enhancement network. Similar to the degradation-specific prompts learning, given the natural reference image $\mathbf{I}^r$, the corresponding prompt embedding is also initialized as $\mathbf{T}^r$. Meanwhile, the loss function of positive prompts learning $\mathcal{L}_{rp}$ is utilized to measures the similarity between reference image and the positive prompt.

During training prompt learning, the parameters in image and text encoders are frozen and only learnable prompt is trainable. Therefore, through contrastive learning on diverse degraded underwater images and natural references, the degradation-specific and positive reference prompts can be efficiently optimized, enabling the model to achieve an optimal bias-variance balance.

Prompt-Guided Image Enhancing

The prompt-guided image enhancing network, as depicted in Figure 2 (b), which utilizes degraded images as input and generates corresponding prompt for guiding enhancement. To better explore the reliable information from degradation perceptions and prompt the enhancement network effectively, we design a dynamic learning network as backbone with the gradation-specific prompts. Meanwhile, an adaptive adjustment loss by degradation perceptions is introduced to train enhancement network. From an input underwater image tensor $\mathbf{X} \in \mathbb{R}^{H\times W\times 3}$ and a conditional prompt $\mathbf{Y} \in \{\mathbf{Y}^d \in \mathbb{R}^{G\times 512}, \mathbf{Y}^r \in \mathbb{R}^{1\times 512}\}$, we prompt the enhancement process with learning rate η as:

$$\mathcal{P}(\mathbf{X},\mathbf{Y}) = F(\mathbf{X};\mathbf{Y}^d : \eta) + \mathcal{L}_{ie}(\mathbf{X},\mathbf{Y}^r), \tag{3}$$

where $\mathcal{P}(\cdot)$ represents the dynamic learning network with the degradation-specific prompts $\mathbf{Y}_d$ and an adaptive adjustment loss as $\mathcal{L}_{ie}$. Prompt fusion is described as $F(\mathbf{X};\mathbf{Y}^d)$.

Dynamic Learning with Conditional Prompt. In order to leverage conditional prompt effectively, we input underwater images to prompt generating module and deign a dynamic learning network. After obtaining responding degradation-specific prompts, we first fuse textual representations and intermediate image features with a projection and concat layer. The degradation-specific prompts can significantly provide reliable information, which lacking in image features. Subsequently, to leverage these prompts as enhancement priors, we attempt to calculate the similarity between learnable prompts and image features like Eq. 2. Therefore, we can obtain the degradation type of input images $\widehat{k}_i$, which can be described as:

$$\widehat{k}_i = \arg\max \mathcal{S}(\mathbf{X}_i, \mathbf{Y}_i^d). \tag{4}$$

With obtaining degradation types, we intend to fine-tune the training process for mitigating optimized conflicts between different degradations. Drawing on multi-task learning (Loshchilov and Hutter 2017; Khodamoradi et al. 2021), hard tasks (with small gradients and slow convergence) receive higher weights, boosting their effective learning rates, while easier tasks receive lower weights. We regard UIE task as multiple enhancement tasks and design a dynamic fine-tune strategy by degradation perceptions, which adjust learning rate based on difficulty of enhance degradation types and improve the learning and generalization abilities of model.

Specifically, we consider the one with highest quantity as the main degradation type in one batch, and calculate the corresponding performance scores to fine-tune learning strategy. And then, for the multi-task optimization of enhancing multiple degradations, we increase low-scoring types (indicating poor enhancement quality), while decreasing it for high-scoring ones. Accordingly, we introduce a scaling factor based on enhancement results during training, which can be expressed as:

$$\eta_t = \eta_{base} \times \min(m_0, \max(m_1, a - \frac{\text{avgscore}_i}{\beta})), \tag{5}$$

where $\mathrm{avgscore_i}$ is the average PSNR performance score of different types and β is selected from training performance as 30. The scaling parameters m_0, m_1 are set as $\{2.0, 0.5\}$ due to the cyclical learning rates inspiration(Shen et al. 2024) and $a = m_0 - m_1$. This strategy focuses the model more on degradation types that are harder to enhance and balance optimization during training like:

$$\theta_{t+1} = \theta_t - \eta_t \nabla_\theta (\sum\nolimits_{i=1}^{M} \mathcal{L}_i(\theta)), \tag{6}$$

where θ_t is network parameters for the t-th iteration and $\mathcal{L}_i$ represents various loss functions. Notably, the enhancement network is trainable and degradation perception module is frozen during training.

Adaptive Adjustment Loss. Enhancing image details in severely various degraded images is challenging due to the visual distortion in different regions caused by complex degradations. Existing UIE methods mostly regard these regions as a uniform or utilize implicit attention mechanism to enhance these details, which ignores the intrinsic connections between different degradations.

Therefore, our enhancement network is uniquely designed with conditional prompts as priors, and designs an adaptive adjustment loss. Specifically, with the conditional prompts, we can obtain the degradation types $\widehat{k}_i$ from Eq. 4. Based on this guidance, we propose an adaptive loss framework that dynamically tailors enhancement through specialized components: $\mathcal{L}_{A1}$ corrects color casts by aligning channel-wise means to a neutral target; $\mathcal{L}_{A2}$ mitigates low-light conditions via minimum-illumination constraints and exposure preservation; $\mathcal{L}_{A3}$ enhances contrast and reduces haze by maximizing Laplacian edge energy to amplify high-frequency structural details. They are defined as:

$$\mathcal{L}_{A1} = \sum_{i=1}^{B} \sqrt{(\mathbf{C}_r^i - 0.5)^4 + (\mathbf{C}_g^i - 0.5)^4 + (\mathbf{C}_b^i - 0.5)^4}, \tag{7}$$

$$\mathcal{L}_{A2} = \sum_{i=1}^{B} [\mathrm{ReLu}(0.5 - \mu_{\mathrm{g}}^{\mathrm{i}}) + 0.5 \times |\mu_g^i - \mu_{og}^i|], \tag{8}$$

$$\mathcal{L}_{A3} = -\frac{1}{H \times W} \sum_{h=1}^{H} \sum_{w=1}^{W} |\nabla X_g(h,w)|, \tag{9}$$

where $\mathbf{C}_{r,g,b}^i$ mean the average of three color channels and μ_g represents the global average grayscale of image. B is the image batch size and ∇X_g means feature tensor. Meanwhile, the adaptive adjustment weights can be expressed as:

$$\omega_t^{\mathcal{L}} = \min(m_0, \max(m_2, \overline{\mathcal{L}}_t)), \tag{10}$$

where $\overline{\mathcal{L}}_t$ denotes the mean MSE variation in t type of main degradations. Similar to Eq. 5, m_2 is scaling parameter as lower limit to ensures updates and set as 0.7. Hence, we utilize the MSE loss history to update the adaptive weight of different degradations, which aims to balance the learning progress of different degradation types and prevent the model from optimizing simple types excessively and ignoring difficult types. This tripartite loss function automatically adjusts loss weights of each image, enabling targeted optimization for various degradations while maintaining natural visual properties.

Dataset	Metric	Water-Net	FUnIE	UWCNN	MLLE	PUIE	URanker	Ushape	CCLNet	SMDR-IS	CPLDP
UIEB	PSNR ↑	20.07	18.20	15.47	17.99	21.71	22.05	21.11	19.78	22.45	**22.60**
	SSIM ↑	0.809	0.849	0.792	0.799	0.854	0.868	0.818	0.848	**0.914**	0.891
	UCIQE ↑	0.567	0.551	0.487	0.603	0.575	0.589	0.551	0.540	0.598	**0.641**
	UIQM ↑	3.459	4.041	3.359	4.070	3.389	3.397	4.138	3.021	3.280	**4.443**
	UIConM ↑	0.708	0.784	0.632	0.715	0.678	0.703	0.747	0.566	0.693	**0.809**
	UISM ↑	5.316	7.133	4.978	5.991	5.850	5.385	7.217	4.989	5.528	**7.256**
	UICM ↑	-22.75	-26.55	-23.08	-19.05	-27.03	-24.98	-23.53	-28.62	-29.42	**-20.98**
	CCF ↑	18.13	20.36	12.14	27.43	21.70	20.03	20.55	24.75	25.16	**28.33**
U45	UCIQE ↑	0.568	0.535	0.485	0.598	0.566	0.597	0.553	0.633	0.588	**0.636**
	UIQM ↑	4.361	4.169	4.282	4.600	4.297	4.487	4.261	4.612	4.219	**4.732**
	UIConM ↑	0.813	0.808	0.736	0.713	0.768	0.809	0.772	0.820	0.794	**0.879**
	UISM ↑	7.193	7.135	7.276	**7.472**	7.205	7.184	7.120	7.266	7.263	7.300
	UICM ↑	-23.71	-29.28	-17.63	-16.56	-20.48	-18.68	-21.29	-16.53	-27.05	**-15.97**
	CCF ↑	23.25	19.44	15.15	50.03	25.51	26.05	21.39	24.73	27.37	**51.44**
UCCS	UCIQE ↑	0.499	0.503	0.462	0.554	0.526	0.553	0.539	0.564	0.503	**0.568**
	UIQM ↑	3.501	3.722	2.330	4.068	4.009	3.520	3.534	3.445	2.677	**4.183**
	UIConM ↑	0.738	0.738	0.646	0.747	0.686	0.730	0.717	0.730	0.658	**0.776**
	UISM ↑	3.929	4.618	3.338	5.053	4.731	4.039	**5.580**	4.326	3.694	4.995
	UICM ↑	-17.40	-22.51	-34.27	**-12.78**	-18.66	-15.04	-21.62	-24.38	-27.10	-13.45
	CCF ↑	18.25	17.68	13.73	**31.37**	19.64	20.14	20.30	20.68	17.52	30.93

Table 1: Quantitative comparison of different UIE methods and our proposed CPLDP on the UIEB val, U45, UCCS datasets and best results are highlighted in bold.

Loss Functions

In the conditional prompt learning, we employ $\mathcal{L}_{dp}$ and $\mathcal{L}_{rp}$ to learn the degradation-specific prompts and positive reference prompt pairs, which are expressed in Eq. 1. Subsequently, in prompt-guided image enhancing, we introduce a positive prompt loss $\mathcal{L}_{re}$ with enhanced image for further guiding the enhancement process. This loss function $\mathcal{L}_{re}$ is formulated as follows:

$$\mathcal{L}_{re} = 1 - \frac{\Phi_{\text{image}}(\mathbf{X}^E) \cdot \Phi_{\text{text}}(\mathbf{T}^r)}{\|\Phi_{\text{image}}(\mathbf{X}^E)\| \cdot \|\Phi_{\text{text}}(\mathbf{T}^r)\|}. \quad (11)$$

Besides, we also utilize the pixel loss between enhanced images $\mathbf{X}^E$ and references $\mathbf{X}^R$ to encourages enhanced results to be similar to references in terms of content and structure, which can be expressed as:

$$\mathcal{L}_e = \tfrac{1}{B}\|\mathbf{X}^E - \mathbf{X}^R\|_2. \quad (12)$$

Overall, in order to supervise the learning process of the enhancement network, we organize the above loss functions as training objectives in prompt-guided image enhancing to guide the optimization, which are calculated as follows:

$$\mathcal{L}_{total} = \mathcal{L}_e + \lambda\mathcal{L}_{re} + \sum_{t=1}^{3}\omega_t^{\mathcal{L}}\mathcal{L}_{At}, \quad (13)$$

where the trade-off weight λ is introduced for balancing the total training loss and is set to 0.9 empirically.

Experiments

Experiment Setup

Datasets. Experiments for evaluation are trained on the constructed human visual-based underwater dataset. Specifically, we invite a group of experienced volunteers to gather representative 800 samples from related underwater datasets (Li et al. 2019) and divide into five types of degradation (blue/green/green-blue color casts, uneven illumination, and haze). For testing, 90 examples with references are regarded as test set to ensure consistency with existing methods. Besides, to further evaluate the effectiveness, two extra underwater dataset U45 (Li, Li, and Wang 2019) and UCSS (Liu et al. 2020) are utilized as non-reference benchmarks.

Implementation Details. Our method is implemented based on Pytorch framework and all experiments are conducted on a NVIDIA RTX 3090Ti GPU. Meanwhile, the optimizer of CPLDP is selected as Adam with $\beta_1 = 0.9, \beta_2 = 0.99$ and each image is randomly cropped to 256×256. Besides, we need to emphasize that a two-stage training strategy is designed for our prompts guided underwater image enhancement. In the conditional prompt learning, we initialize all prompts with a batch size of 16 and CLIP backbone (Radford et al. 2021).The prompts learning network is trained by the constructed dataset with learning rate as $5 \cdot 10^{-6}$. In enhancing stage, the parameters in prompts learning network are frozen, and the L_{re} is introduced for guiding image enhancement at a batch size of 16. The initialization of learning rate in enhancement network is set to $2 \cdot 10^{-5}$. The scaling parameters in enhancement network m_0, m_1, m_2 are set as $\{2.0, 0.5, 0.7\}$

Evaluation Metrics. To evaluate and compare the performance of UIE methods, both full-reference and non-reference evaluation metrics are adopted in experiments. Specifically, PSNR (Wang et al. 2004) and SSIM (Korhonen and You 2012) are employed to evaluate the full-reference image enhancement results, with higher values suggesting higher similarity to the reference image. For non-reference experiments, we introduce UCIQE, UIQM, UISM, UICM, UIConM (Jiang et al. 2022a) and CCF (Wang et al. 2018)

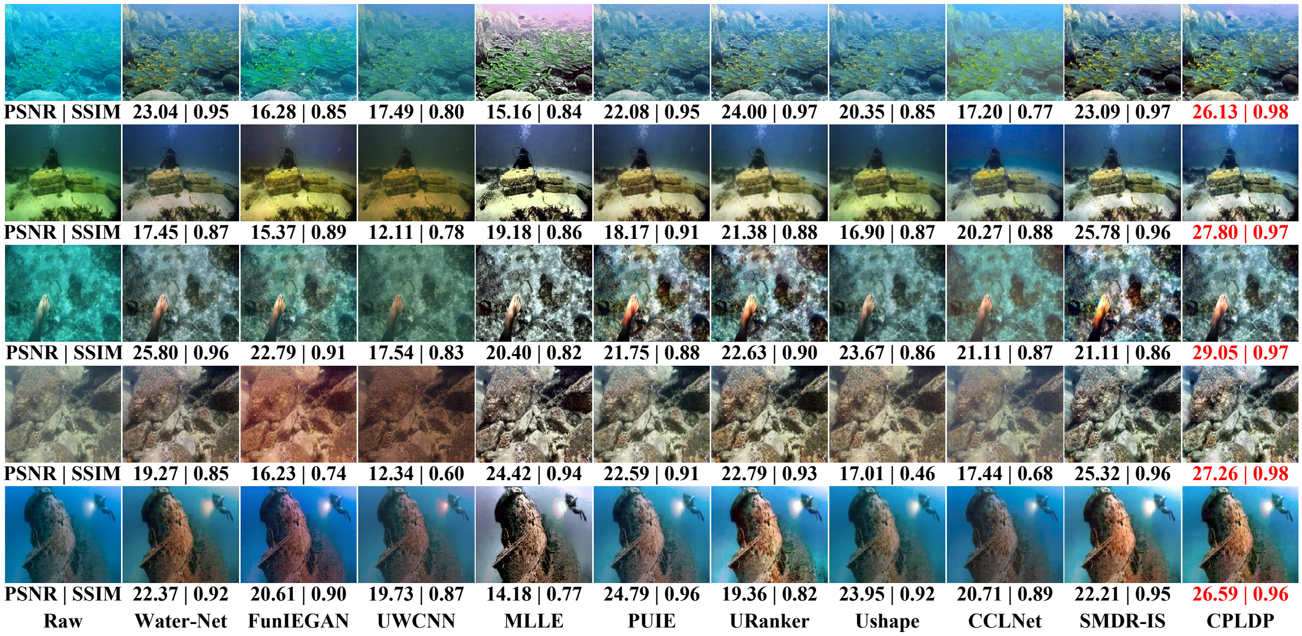


Figure 3: Qualitative comparison of enhancement results sampled from each degradation. The highest PSNR and SSIM scores are marked in red. It can be observed our CPLDP can effectively enhance various degraded underwater images.

to evaluate our method, which are specifically designed for underwater image enhancement.

Experiment Results

In this section, we conducted comparison experiments with state-of-the-art UIE methods to validate the effectiveness of our CPLDP, including Water-Net (Li et al. 2019), FUnIE (Islam, Xia, and Sattar 2020), UWCNN (Li, Anwar, and Porikli 2020), MLLE (Zhang et al. 2022), PUIE (Fu et al. 2022b), URanker (Guo et al. 2023), Ushape (Peng, Zhu, and Bian 2023), CCLNet (Liu et al. 2024) and SMDR-IS (Zhang et al. 2024). For a fair comparison, we employ the original source codes released by the respective authors and strictly maintain the settings in papers.

Full-Reference Evaluation. To comprehensively evaluate our proposed CPLDP framework, we conducted qualitative and quantitative comparisons against nine state-of-the-art UIE methods using full-reference underwater datasets. As presented in Table 1, CPLDP achieves the highest PSNR values and secures second-best SSIM performance among all competitors. The quantitative superiority is reinforced by visual evidence in Figure 3, where our method consistently demonstrates the intrinsic color correction and reduces uneven illumination and haze interferences.

For a detailed analysis, while established methods like MLLE, PUIE, and Ushape deliver competent enhancement under specific degradation profiles, they frequently introduce visible distortions and exhibit limited generalization across diverse underwater conditions. Meanwhile, SMDR-IS generates results closer to reference images but suffers from systematic over-enhancement artifacts such as localized over-exposure. In contrast, CPLDP achieves significantly more balanced enhancement, exhibiting minimal color casts and artifacts while demonstrating robustness across varied scenarios. These performances advantage stems fundamentally from our method's explicit modeling of inherent degradation relationships through conditional prompt learning. By dynamically adapting to degradation interdependencies, our method achieves context-aware enhancement that maintains natural appearance with positive prompt while suppressing distortions.

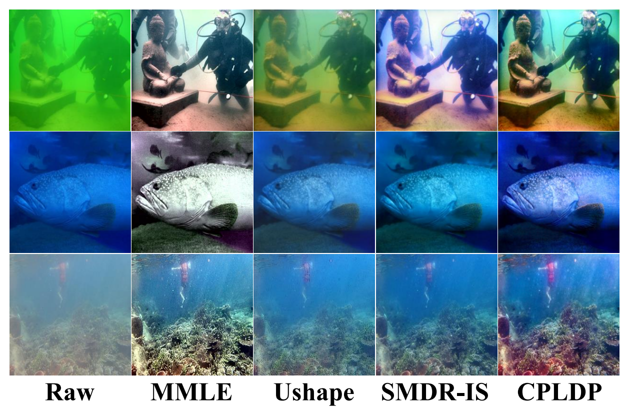


Figure 4: Visual results of state-of-the-art methods and CPLDP on U45 datasets. Compared with them, our method exhibits fewer visual distortions, and enhances details better.

Non-Reference Evaluation. The U45 and UCCS datasets are utilized for conducting non-reference comparative experiments. As evidenced by both quantitative metrics in Table 1 and visual results in Figure 4, our proposed CPLDP

Components			Metrics			
$\mathbf{Y}^F$	$\mathbf{Y}^d$	$\mathbf{Y}^r$	PSNR↑	SSIM↑	UCIEQ↑	UIQM↑
			19.045	0.832	0.557	4.029
✓			18.339	0.807	0.542	3.386
	✓		19.582	0.826	0.585	3.461
		✓	21.107	0.853	0.602	4.137
	✓	✓	**22.597**	**0.891**	**0.641**	**4.443**

Table 2: Ablation study of conditional prompt learning.

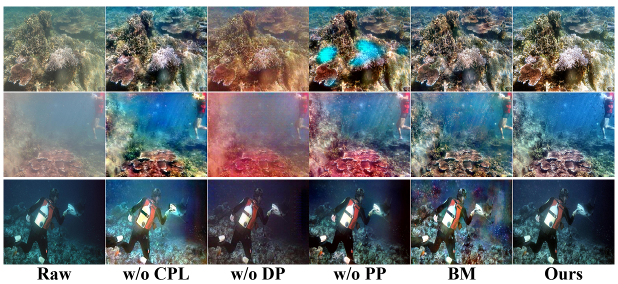


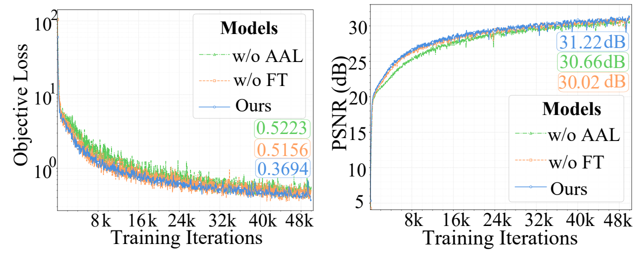
Figure 5: Visual results of different ablation studies settings on the UIEB dataset.

demonstrates superior underwater image enhancement performance across nearly all non-reference metrics. Visual results reveals critical limitations in comparative methods like homogeneous training strategy MMLE, tending to produce over-enhanced results and lost details and other approaches frequently introduce unnatural distortions and artifacts. In stark contrast, our method consistently delivers accurate enhancement across diverse degradation scenarios, and natural color enhancement with preserved structural integrity. This is because the degradation perceptions can dynamically reinforce to focus on degraded regions though conditional prompts. Combined with our adaptive loss function that prioritizes solve the conflicting optimization goals, the CPLDP framework significantly elevates UIE quality by focusing capacity where enhancement is most critically needed.

Ablation Study

To validate the contribution of each component in our CPLDP network, ablation studies are conducted with systematically evaluating the impact of key designs: conditional prompt learning, dynamic learning and adaptive loss functions. Therefore, we set a baseline model with fixed text prompts like "This is a low-quality underwater image" and " This is a high-quality underwater image". Besides, BM also removes dynamic learning and adaptive loss.

Effectiveness of Conditional Prompts. To investigate the effect about our proposed conditional prompt learning, we conduct a clear model without any prompts. As provided in Table. 2, $\mathbf{Y}^F, \mathbf{Y}^d, \mathbf{Y}^r$ represent fixed prompt, degradation-specific prompt and positive reference prompts, respectively. It is obvious that both the model without prompts and basic model with $\mathbf{Y}^F$ achieve lower PSNR and SSIM, as their lack of effective prompt guidance fails to properly direct the enhancement process. Besides, the visual results are provided to directly distinguish the contribution of these components. The enhanced results with diminished distortions confirm effective enhancements based on conditional prompts in Figure 5. Unlike single-prompt model with $\mathbf{Y}^d$: DP and $\mathbf{Y}^r$: PP, conditional prompt learning model integrates degradation-specific and positive reference prompts, simultaneously guiding image enhancement while mitigating interference between distinct degradations.



(a) Objective Loss (b) PSNR Performance

Figure 6: The ablation study of Dynamic Learning & Adaptive Adjustment Loss with objective loss and PSNR performance curves during training.

Effectiveness of Dynamic Learning & Adaptive Adjustment Loss. During the training process of prompt-guided image enhancing network, we design a dynamic learning network to effectively guide training strategy and an adaptive adjustment loss to solve conflicts between different degradations optimization. Therefore, to verify the effectiveness of these components, ablation models as w/o FT and w/o AAL are conducted, represent without dynamic learning and without adaptive adjustment loss. As shown in Figure 6, we have conducted comparative experiments with the optimized loss and PSNR metric, which can directly observe the effect of our proposed two components. The results clearly demonstrate that CPLDP achieves superior performance and faster convergence when equipped with the dynamic learning strategy and adaptive adjustment loss.

Conclusion

In this work, we propose a novel underwater image enhancement method with conditional prompt learning as degradation perception, which can fully explore intrinsic connections between different degradations and provide guidance with positive reference prompt. To mitigate interference from optimizations of various degradations, we introduced a dynamic learning network, which is meticulously designed to regulate the network optimization process and integrated prompt features in propagation. Furthermore, our adaptive adjustment loss function can flexibly guide the enhancement network to process different degradations. Quantitative and qualitative experimental results show that underwater images with diverse degradations can be effectively enhanced by manipulating the conditional prompt as degradation perception.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China Grants 62176037,62576067, National Key Research and Development Program of China Grant 2022YFE0132600,2024YFB4710800, Liaoning Provincial Natural Science Foundation Grant 2025-YQ-01 and 2024-MS-012, Dalian Science and Technology Talent Innovation Support Plan Grant 2024RY010.

References

Ancuti, C.; Ancuti, C. O.; Haber, T.; and Bekaert, P. 2012. Enhancing underwater images and videos by fusion. In *2012 IEEE conference on computer vision and pattern recognition*, 81–88. IEEE.

Chen, H.; Chen, X.; Lu, J.; and Li, Y. 2024. Rethinking multi-scale representations in deep deraining transformer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 1046–1053.

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.

Conde, M. V.; Geigle, G.; and Timofte, R. 2024. Instructir: High-quality image restoration following human instructions. In *European Conference on Computer Vision*, 1–21. Springer.

Fu, Z.; Lin, H.; Yang, Y.; Chai, S.; Sun, L.; Huang, Y.; and Ding, X. 2022a. Unsupervised underwater image restoration: From a homology perspective. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 643–651.

Fu, Z.; Wang, W.; Huang, Y.; Ding, X.; and Ma, K.-K. 2022b. Uncertainty inspired underwater image enhancement. In *European conference on computer vision*, 465–482. Springer.

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.

Hambarde, P.; Murala, S.; and Dhall, A. 2021. UW-GAN: Single-image depth estimation and image enhancement for underwater images. *IEEE Transactions on Instrumentation and Measurement*, 70: 1–12.

Han, M.; Lyu, Z.; Qiu, T.; and Xu, M. 2018. A review on intelligence dehazing and color restoration for underwater images. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 50(5): 1820–1832.

Hou, G.; Li, N.; Zhuang, P.; Li, K.; Sun, H.; and Li, C. 2023. Non-uniform illumination underwater image restoration via illumination channel sparsity prior. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(2): 799–814.

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.

Jiang, Q.; Gu, Y.; Li, C.; Cong, R.; and Shao, F. 2022a. Underwater image enhancement quality evaluation: Benchmark dataset and objective metric. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(9): 5959–5974.

Jiang, Y.; Wang, H.; Peng, J.; Fu, X.; and Wang, Y. 2024. Scene-adaptive person search via bilateral modulations. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 938–946.

Jiang, Z.; Li, Z.; Yang, S.; Fan, X.; and Liu, R. 2022b. Target oriented perceptual adversarial fusion network for underwater image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(10): 6584–6598.

Jiang, Z.; Liu, R.; Yang, S.; Zhang, Z.; and Fan, X. 2025. DRNet: Learning a dynamic recursion network for chaotic rain streak removal. *Pattern Recognition*, 158: 111004.

Khodamoradi, A.; Denolf, K.; Vissers, K.; and Kastner, R. C. 2021. Aslr: An adaptive scheduler for learning rate. In *2021 International Joint Conference on Neural Networks (IJCNN)*, 1–8. IEEE.

Korhonen, J.; and You, J. 2012. Peak signal-to-noise ratio revisited: Is simple beautiful? In *2012 Fourth international workshop on quality of multimedia experience*, 37–38. IEEE.

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, C.-Y.; Guo, J.-C.; Cong, R.-M.; Pang, Y.-W.; and Wang, B. 2016. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *IEEE Transactions on Image Processing*, 25(12): 5664–5677.

Li, H.; Li, J.; and Wang, W. 2019. A fusion adversarial underwater image enhancement network with a public test dataset. *arXiv preprint arXiv:1906.06819*.

Liang, Z.; Li, C.; Zhou, S.; Feng, R.; and Loy, C. C. 2023. Iterative prompt learning for unsupervised backlit image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8094–8103.

Liu, R.; Fan, X.; Zhu, M.; Hou, M.; and Luo, Z. 2020. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE transactions on circuits and systems for video technology*, 30(12): 4861–4875.

Liu, R.; Jiang, Z.; Yang, S.; and Fan, X. 2022. Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Transactions on Image Processing*, 31: 4922–4936.

Liu, Y.; Jiang, Q.; Wang, X.; Luo, T.; and Zhou, J. 2024. Underwater image enhancement with cascaded contrastive learning. *IEEE Transactions on Multimedia*.

Loshchilov, I.; and Hutter, F. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Panetta, K.; Gao, C.; and Agaian, S. 2015. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3): 541–551.

Peng, L.; and Bian, L. 2025. Adaptive Dual-domain Learning for Underwater Image Enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 6461–6469.

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

Potlapalli, V.; Zamir, S. W.; Khan, S. H.; and Shahbaz Khan, F. 2023. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 36: 71275–71293.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763. PmLR.

Shang, J.; Li, Y.; Xing, H.; and Yuan, J. 2025. LGT: Luminance-guided transformer-based multi-feature fusion network for underwater image enhancement. *Information Fusion*, 118: 102977.

Shen, L.; Sun, Y.; Yu, Z.; Ding, L.; Tian, X.; and Tao, D. 2024. On efficient training of large-scale deep learning models. *ACM Computing Surveys*, 57(3): 1–36.

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.

Wang, H.; Chen, Y.; Yao, M.; Liu, W.; Peng, J.; and Fu, X. 2025. Tensor Completion Framework by Graph Refinement for Incomplete Multi-view Clustering. *IEEE Transactions on Multimedia*.

Wang, H.; Yao, M.; Chen, Y.; Xu, Y.; Liu, H.; Jia, W.; Fu, X.; and Wang, Y. 2024. Manifold-based incomplete multi-view clustering via bi-consistency guidance. *IEEE Transactions on Multimedia*, 26: 10001–10014.

Wang, H.; Yao, M.; Jiang, G.; Mi, Z.; and Fu, X. 2023. Graph-collaborated auto-encoder hashing for multiview binary clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 35(7): 10121–10133.

Wang, Y.; Li, N.; Li, Z.; Gu, Z.; Zheng, H.; Zheng, B.; and Sun, M. 2018. An imaging-inspired no-reference underwater color image quality assessment metric. *Computers & Electrical Engineering*, 70: 904–913.

Wang, Y.; Liu, H.; and Chau, L.-P. 2017. Single underwater image restoration using adaptive attenuation-curve prior. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 65(3): 992–1002.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.

Wen, Y.; Jain, N.; Kirchenbauer, J.; Goldblum, M.; Geiping, J.; and Goldstein, T. 2023. Hard prompts made easy: Gradient-based discrete optimization for prompt tuning and discovery. *Advances in Neural Information Processing Systems*, 36: 51008–51025.

Yao, M.; Wang, H.; Chen, Y.; and Fu, X. 2025a. Between/Within View Information Completing for Tensorial Incomplete Multi-View Clustering. *IEEE Transactions on Multimedia*, 27: 1538–1550.

Yao, M.; Wang, H.; Li, Y.; Liu, W.; and Fu, X. 2025b. Detail-focused and polarization-guided multi-modality fusion for underwater image clarity enhancing. *Engineering Applications of Artificial Intelligence*, 159: 111677.

Zhang, D.; Zhou, J.; Guo, C.; Zhang, W.; and Li, C. 2024. Synergistic multiscale detail refinement via intrinsic supervision for underwater image enhancement. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 7033–7041.

Zhang, D.; Zhou, J.; Zhang, W.; Lin, Z.; Yao, J.; Polat, K.; Alenezi, F.; and Alhudhaif, A. 2023a. ReX-Net: A reflectance-guided underwater image enhancement network for extreme scenarios. *Expert Systems with Applications*, 231: 120842.

Zhang, W.; Zhuang, P.; Sun, H.-H.; Li, G.; Kwong, S.; and Li, C. 2022. Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Transactions on Image Processing*, 31: 3997–4010.

Zhang, X.; Ma, J.; Wang, G.; Zhang, Q.; Zhang, H.; and Zhang, L. 2025a. Perceive-ir: Learning to perceive degradation better for all-in-one image restoration. *IEEE Transactions on Image Processing*.

Zhang, Z.; Jiang, Z.; Liu, J.; Fan, X.; and Liu, R. 2023b. Waterflow: Heuristic normalizing flow for underwater image enhancement and beyond. In *Proceedings of the 31st ACM International conference on multimedia*, 7314–7323.

Zhang, Z.; Jiang, Z.; Ma, L.; Liu, J.; Fan, X.; and Liu, R. 2025b. HUPE: Heuristic underwater perceptual enhancement with semantic collaborative Learning. *International Journal of Computer Vision*, 1–19.

Zhao, C.; Cai, W.; Dong, C.; and Hu, C. 2024. Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8281–8291.