








DiffWater: Underwater Image Enhancement Based on Conditional Denoising Diffusion Probabilistic Model

Meisheng Guan , Haiyong Xu , *Member, IEEE*, Gangyi Jiang , *Senior Member, IEEE*, Mei Yu , Yeyao Chen , *Member, IEEE*, Ting Luo , *Member, IEEE*, and Xuebo Zhang , *Senior Member, IEEE*

Abstract—Underwater imaging is often affected by light attenuation and scattering in water, leading to degraded visual quality, such as color distortion, reduced contrast, and noise. Existing underwater image enhancement (UIE) methods often lack generalization capabilities, making them unable to adapt to various underwater images captured in different aquatic environments and lighting conditions. To address these challenges, a UIE method based on the conditional denoising diffusion probabilistic model (DDPM) is proposed (DiffWater), which leverages the advantages of DDPM and trains a stable and well-converged model capable of generating high-quality and diverse samples. Considering the multiple distortion issues in underwater imaging, unconditional DDPM may not achieve satisfactory enhancement and restoration results. Therefore, DiffWater utilizes the degraded underwater image with added color compensation as a conditional guide, through which the DiffWater achieves high-quality restoration of degraded underwater images. Particularly, the proposed DiffWater introduces a color compensation method that performs channelwise color compensation in the RGB color space, tailored to different water conditions and lighting scenarios, and utilizes this condition to guide the denoising process. In the experimental section, the proposed DiffWater method is tested on four real underwater image datasets and compared against existing methods. Experimental results demonstrate that DiffWater outperforms existing comparison methods in terms of enhancement quality and effectiveness, exhibiting stronger generalization capabilities and robustness.

Index Terms—Color compensation, conditional denoising diffusion probabilistic model (DDPM), underwater image enhancement (UIE).

Manuscript received 11 October 2023; revised 30 November 2023 and 9 December 2023; accepted 12 December 2023. Date of publication 19 December 2023; date of current version 4 January 2024. This work was supported in part by the Natural Science Foundation of China under Grant 62171243, Grant 61971247, Grant 62071266, Grant 61931022, and Grant 61671412, in part by the Zhejiang Natural Science Foundation of China under Grant LY19F020009, Grant LY21F010003, Grant LY19F010002, Grant LQ20F010002, Grant LY21F010014, Grant LY22F020020, Grant LQ24F010002, and Grant Z24F010019, and in part by the Natural Science Foundation of Ningbo under Grant 2022JCGY010527, Grant 2022J066, and Grant 202003N4088. (*Corresponding author: Haiyong Xu.*)

Meisheng Guan and Haiyong Xu are with the School of Mathematics and Statistics, Ningbo University, Ningbo 315211, China (e-mail: meisheng-guan@gmail.com; xuhaiyong@nbu.edu.cn).

Gangyi Jiang, Mei Yu, Yeyao Chen, and Ting Luo are with the Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China (e-mail: jianggangyi@nbu.edu.cn; yumei@nbu.edu.cn; cheneyao@nbu.edu.cn; luoting@nbu.edu.cn).

Xuebo Zhang is with the Northwest Normal University, Lanzhou 730071, China (e-mail: xuebo_zhang@sina.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3344453

I. INTRODUCTION

WITH the advancement of technology, underwater imaging has been widely used in marine detection, underwater robotics, underwater archaeology, and other fields [1], [2], [3]. However, due to the complexity and variability of the underwater environment, underwater imaging is often affected by factors such as light attenuation and scattering, resulting in severe degradation of underwater image quality, color distortion, low contrast, etc., [4], as shown in Fig. 1 (*Top*). This directly affects subsequent underwater image analysis and understanding tasks, such as underwater target detection, recognition, and tracking [5], [6], [57]. Therefore, how to effectively recover and enhance underwater images to improve their contrast, sharpness, and realism has become an important and difficult problem in underwater imaging technologies.

To improve the quality of underwater images, traditional methods [7], [16], [17], [18], [19], [20], [51], [58] and deep-learning-based methods [8], [10], [11], [12], [13], [23], [24], [25], [31], [32], [33], [55], [56], [59] are proposed for underwater image enhancement (UIE). Traditional UIE methods rely on a priori knowledge or assumptions, design rules, or models to process underwater images, such as histogram equalization [7] and white balance [9]. Although these methods are simple to implement, the effect is limited and cannot adapt to different water qualities and lighting conditions.

Recently, deep-learning-based UIE methods have made tremendous progress in computer vision, providing new insights and tools for UIE. By leveraging a large amount of data to learn the features and patterns of underwater images, deep-learning-based UIE methods achieve automated, intelligent, and end-to-end enhancement of underwater images. Generally, deep-learning-based UIE methods are mainly classified into two categories: 1) convolutional neural networks (CNN)-based [10], [11] and 2) generative adversarial networks (GAN)-based [12]. The CNN-based UIE methods, which train deep CNNs with a large amount of data to learn the mapping relationship from the degraded underwater image to the clear water image or the enhanced underwater image [13], can adaptively deal with underwater images under different scenes and achieve better performance. The GAN-based UIE can realize the conversion between the degraded underwater image domain and the distortion-free image domain and has achieved great success [14]. However, the GAN-based UIE method often trains unstably and suffers from mode collapse, and the generated samples exist in mode collapse

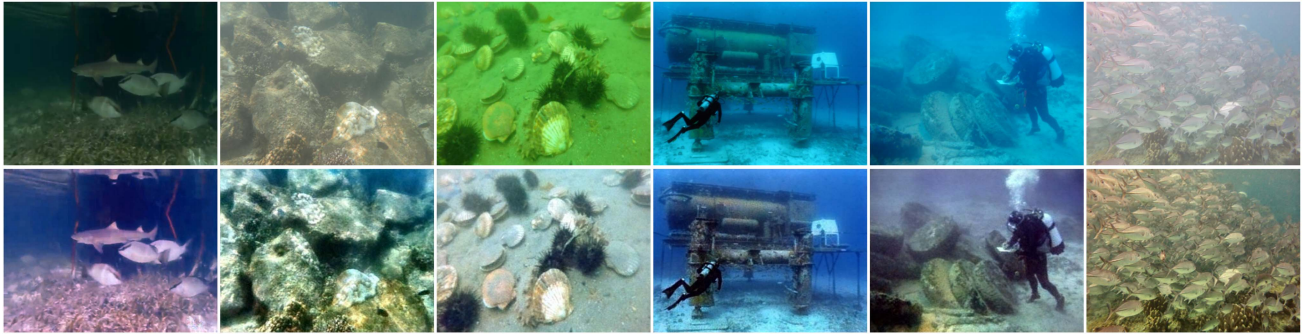


Fig. 1. Some results of the proposed DiffWater in different underwater visual scenes. (Top) Raw underwater images with serious color deviation, low contrast, blur, and green. (Bottom) Enhanced images correspond to the proposed DiffWater.

or lack of diversity. Therefore, in order to further improve the quality and effect of UIE, it is necessary to explore more stable and diversified methods.

With the introduction of the denoising diffusion probabilistic model (DDPM) [15] in the field of image generation, the diffusion model has begun to attract the wide attention of researchers due to its relatively stable training process and good convergence performance. Generally, the DDPM contains two processes: 1) the forward diffusion process and 2) the reverse diffusion process. The forward diffusion process refers to the gradual addition of Gaussian noise to the real image so that the image distribution gradually becomes flat and isotropic. The reverse diffusion process refers to the gradual removal of noise from the noisy image to recover the real image. More recently, the DDPM has achieved significant results in the field of image generation, and its generated samples have high quality and high diversity. However, unconditional DDPM has certain limitations, because it lacks flexibility and can only perform unconditional generation, which may lead to the generation of some images that do not match the data distribution. Since the degree of degradation of underwater images is higher, unconditional DDPM may not be able to generate high-quality images. Therefore, it is necessary to consider using effective conditions to guide the enhancement and recovery of underwater images, thereby improving the quality and relevance of the generated images.

In this article, a novel generative method based on conditional DDPM for UIE, named DiffWater, is proposed to improve the quality of underwater images. The proposed DiffWater method formulates the UIE problem as a conditional diffusion process, leveraging the relative stability and strong convergence properties of DDPM to generate high-quality and diverse samples. By gradually adding noise to diffuse a random noisy image and obtaining the enhanced underwater image with denoising processes, the proposed DiffWater considers the complexity and uncertainty of the underwater environment, the differences of underwater target features, as well as noise and interference. Using the underwater image obtained by color compensation of the degraded underwater image as the condition, it realizes the conversion from the degraded underwater image to the enhanced underwater image. The effectiveness and superiority of the DiffWater method are verified through experiments by comparing and analyzing with other methods. Results show that

the proposed method can effectively improve image sharpness and contrast while retaining detailed information, leading to improved visual quality of underwater images.

The main contributions of the proposed DiffWater are summarized as follows.

- 1) Considering the diversity and complexity of real underwater environments and the issues of poor image quality and color bias resulting from directly using simple priors in DDPM, an optimized DiffWater method was proposed. In the proposed DiffWater method, optimized conditional mechanisms are utilized to extract more information from conditional images through a denoising process. This allows for enhanced underwater images with higher quality, clarity, truthfulness, and naturalness to be achieved.
- 2) To address the issues of poor image quality and color bias in underwater images, a color channel compensation (3C) method was introduced. In the DiffWater method, compensated underwater images with color channels are used as conditional guidance to direct the diffusion denoising process, to improve the color appearance of image enhancement.
- 3) The proposed DiffWater method is tested on four real underwater image datasets and compared and analyzed with existing comparison methods. Experimental results show that the proposed DiffWater method outperforms existing comparison methods in terms of enhancement quality and effects and exhibits better generalizability and robustness.

The rest of the article is organized as follows: Section II reviews the related work. In Section III, the proposed DiffWater is described in detail. In Section IV, extensive experiments are performed to validate the effectiveness and performance of the proposed DiffWater method. Finally, Section V concludes this article.

II. RELATED WORK

A. Traditional UIE Method

Traditional UIE methods rely on prior knowledge or assumptions and design rules or models to process underwater images. Image degradation priors are required to be utilized by UIE methods based on physical models to perform the inverse process of image degradation. Peng et al. [61] proposed

an expanded and modified underwater scene depth estimation method based on the dark channel prior. This method can decompose underwater images into direct light and scattering light components, enabling the restoration of true colors and details by eliminating or suppressing the scattered light component. Li et al. [62] proposed an underwater image restoration algorithm based on dehazing the blue-green channel and correcting the red channel. It utilizes the expansion and modification of the dark channel prior to recover the blue-green channel, then applies the gray-world assumption to correct the red channel and, finally, uses an adaptive exposure map to balance brightness and contrast. Berman et al. [63] proposed an underwater image color recovery method based on an underwater imaging model with parameters estimated from haze-lines. This method applies horizontal or vertical lines in underwater images as haze-lines, calculates the attenuation ratios of different water areas, and estimates the attenuation ratios of the blue-red and blue-green color channels based on the pixel values on the haze-lines. This simplifies the underwater image problem to a single-image haze removal problem, thus achieving color calibration and contrast enhancement of underwater images. Ancuti et al. [16] proposed a color balance and fusion-based method, which combines white balance and the Laplacian pyramid to generate two enhanced results and fuse them through weight mapping. Jin et al. [17] proposed an adaptive histogram transformation method that adjusts the shape of the transformation function using the local mean gray-level and local gray-level variance to enhance image details and contrast. However, these methods did not consider the negative effects of noise and artifacts in underwater images, which can easily lead to overenhancement or loss of details, and cannot eliminate color distortion in underwater images. To address these issues, Li et al. [4] proposed the minimum information loss principle and histogram distribution prior [13], which estimates the transmission map of the underwater scene by minimizing the information loss of the enhanced image and applies a color correction term based on the prior distribution of natural image histograms to restore the clear image. However, this method cannot completely solve the problems of contrast enhancement, noise suppression, and edge preservation. In order to further improve the quality of underwater images, Li et al. [18] proposed multialgorithm fusion in both RGB and HSV color spaces. Drews et al. [19] proposed a UIE method based on underwater dark channel prior (UDCP), which uses statistical priors of outdoor natural scene images and mainly considers blue and green color channels as the sources of underwater visual information. Berman et al. [20] proposed a single-image color restoration method based on the fog line for underwater images, which can estimate the light attenuation ratio based on different water types, thereby simplifying the problem of single-image dehazing. However, these methods also have some issues, such as not considering the scattering effect in underwater images, noise, and detail protection in the image, which may lead to color distortion and reduced contrast. Therefore, UIE remains a challenging research area that requires further exploration and improvement.

B. Deep-Learning-Based UIE Method

Convolutional Neural Networks: CNN-based UIE methods [21], [22] extract multilevel high-level image features through convolutional and pooling layers, which can automatically learn effective features of underwater images. For example, Li et al. [23] proposed a lightweight CNN (UWCNN) with an enhancement branch designed for each type of water, trained with the corresponding data to adapt to different underwater scenes. Li et al. [25] proposed a CNN-based UIE method (Ucolor) that can adaptively integrate and highlight the most discriminative features from multiple color spaces. Additionally, Sun et al. [26] proposed a novel CNN-based UIE using convolutional layers for noise filtering and deconvolutional layers for detail recovery and image optimization. Naik et al. [33] further proposed a shallow neural network architecture called Shallow-wnet for UIE. This network has fewer parameters than existing models while maintaining strong performance. Continuous development and improvements of CNN-based UIE methods will provide more potential and possibilities for UIE.

Generative Adversarial Networks: GAN [12], [27], [28], [55], [56] is a deep-learning-based generative method. Recently, GAN has been widely applied to improve the visual quality and usability of underwater images. Among them, Li et al. [24] proposed a UIE network (Water-Net) as a baseline method, revealing the performance and limitations of existing UIE methods, and providing guidance and inspiration for future research. Zhu et al. [29] proposed a novel UIE method by using CycleGAN for unsupervised image-to-image conversion, which can effectively solve the problems of low contrast, color distortion, and noise existing in underwater images. Li et al. [30] proposed WaterGAN, which utilizes unlabeled underwater video sequences to learn underwater imaging models and generates realistic synthetic underwater images. Fabbri et al. [14] proposed UGAN for improving the quality of underwater images, which can generate realistic underwater images from aerial images and depth map pairs without any paired underwater data, can achieve real-time enhancement, and implicitly learns coarse depth estimation of underwater scenes. Liu et al. [31] proposed a conditional GAN (MLFcGAN), which improves the color and contrast of underwater images through multilevel feature fusion. The multilevel feature fusion method enhances local features into global features, thereby enhancing the learning ability and performance of the network. Islam et al. [32] proposed a method based on conditional GAN (FUnIEGAN) for real-time UIE and designed an objective function that comprehensively considers global content, color, local texture, and style information to guide adversarial training, which can learn to improve the visual quality of underwater images from paired or unpaired data. However, these GAN-based UIE methods training processes are often unstable and difficult to converge, and the generated results often have diversity and uncertainty, which may have certain biases or differences from the ideal clear images, and it is difficult to ensure consistency and accuracy of color and structure.

C. Diffusion Model

With the rapid development of the diffusion model [15], an increasing number of researchers have begun to explore various image processing problems in the fields of image generation [35], [37], [38] and enhancement [39], [40], [41] using diffusion models. Sohl-Dickstein et al. [34] initially proposed the diffusion model, which is a generative method inspired by nonequilibrium thermodynamics, that destroys data structures by gradually adding noise through a Markov chain and then recovers data details by gradually removing noise. Although the diffusion model has a good theoretical basis, its training is unstable, and the sampling speed is slow, so it has not attracted widespread attention.

To solve these problems, DDPM [15] is a simplified diffusion model that uses variational inference for modeling and samples data through reparameterization techniques. In fact, DDPM is a deep generative method that learns the data distribution through the forward diffusion process and reverse diffusion process. The forward diffusion process refers to gradually adding Gaussian noise to the image, making it become more and more blurry and random until it approaches an isotropic Gaussian distribution. The reverse diffusion process is to reconstruct the original data from a random Gaussian distribution, that is, to gradually remove noise. This process requires learning a neural network method to approximate the conditional probability distribution at each step, that is, given the current state of the data, predicting the data of the previous step, DDPM demonstrates stable training performance and achieves good results in image generation. Subsequently, Nicholas et al. [35] proposed improved DDPMs (IDDPM), which improved DDPM to achieve competitive log likelihoods while maintaining high sample quality. Saharia et al. [36] proposed two diffusion-model-based image synthesis methods: Image superresolution via iterative refinement (SR3) and cascaded diffusion models for high-fidelity image generation (CDM) [37]. SR3 is a superresolution diffusion model that builds a corresponding high-resolution image from low-resolution images as an input condition, from pure noise. CDM is a conditional diffusion model that uses category labels as input conditions and generates images of the corresponding category from pure noise.

Currently, diffusion models are applied to image reconstruction tasks [38], [39], [40]. Kawar et al. [41] proposed denoising diffusion restoration models (DDRM) for image recovery, a variational inference framework that uses a pretrained denoising diffusion generative model as a prior for natural images and combines with the linear measurement method to obtain an approximate posterior distribution, from which images are efficiently generated. It can be applied to tasks such as image superresolution, deblurring, restoration, and colorization, generating more realistic and diverse images with broad application prospects. Lu et al. [60] first proposed a UIE method based on DDPM (UW-DDPM), which employs two U-Net networks for image denoising and image distribution transformation, effectively improving the quality of underwater images. However, the UW-DDPM method did not encompass a sufficient amount of datasets during the validation process, thereby failing to

provide a more comprehensive explanation of its generalizability. Furthermore, Lu et al. [64] proposed an accelerated and fused method for DDPM, which improves the speed of the inference process by modifying the initial sampling distribution and reducing the number of iterations in the denoising stage. Additionally, in the diffusion stage, the degraded image and the reference image are fused, enhancing the quality of the enhanced image and avoiding the issues of poor image quality and color deviation resulting from the direct utilization of conditional DDPM. Therefore, inspired by diffusion models, we propose DiffWater to improve the quality of underwater images.

III. UIE BASED ON CONDITIONAL DDPM

In this article, we propose a novel UIE method by physical priors as conditions to guide the denoising process in the diffusion model, with the aim of improving image quality for the characteristics of underwater imaging. In this section, we first introduce the underwater imaging model (see Section A) and then describe the proposed overall DiffWater framework (see Section B). Then, we introduce in detail the key processes of DiffWater, including the forward diffusion process (see Section C), the reverse diffusion process (see Section D) and the loss function (see Section E).

A. Characteristics of Underwater Imaging

Underwater imaging remains a challenging research area due to the effects of absorption, scattering, and reflection of optical or acoustic signals in underwater environments. These phenomena severely deteriorate the image quality, resulting in color cast, blurriness, low contrast, and other issues that hinder the analysis and understanding of underwater images by computer vision systems. Additionally, factors such as water depth, quality, and temperature further complexify underwater imaging by significantly impacting its effectiveness. To obtain clearer underwater images, it is necessary to mathematically model the image degradation process and reverse it by estimating model parameters. The mathematical model of underwater imaging is expressed as [42]

$$I_\lambda(x) = B_\lambda(x)e^{(-\eta \cdot d(x))} + J_\lambda(x)(1 - e^{(-\eta \cdot d(x))}) \quad (1)$$

where $\lambda \in \{R, G, B\}$, $I_\lambda(x)$ is the observed underwater image, $B_\lambda(x)$ is the clean image, $J_\lambda(x)$ is the background radiance, η is the attenuation coefficient, and $d(x)$ is the distance between the camera and the target.

During underwater image acquisition, the observed image $I_\lambda(x)$ is the result of water absorption and scattering while the true clean image is represented as $B_\lambda(x)$. Suspended particles scatter light in both forward and backward directions, reducing image clarity and visibility, while also differently attenuating wavelengths in water, causing color distortion and low contrast. These effects can be described by background brightness $J_\lambda(x)$, attenuation coefficient η , and distance $d(x)$ between camera and target, an important factor. Considering the underwater imaging model, to recover true color and details in underwater images, we introduce a conventional method—3-channel compensation

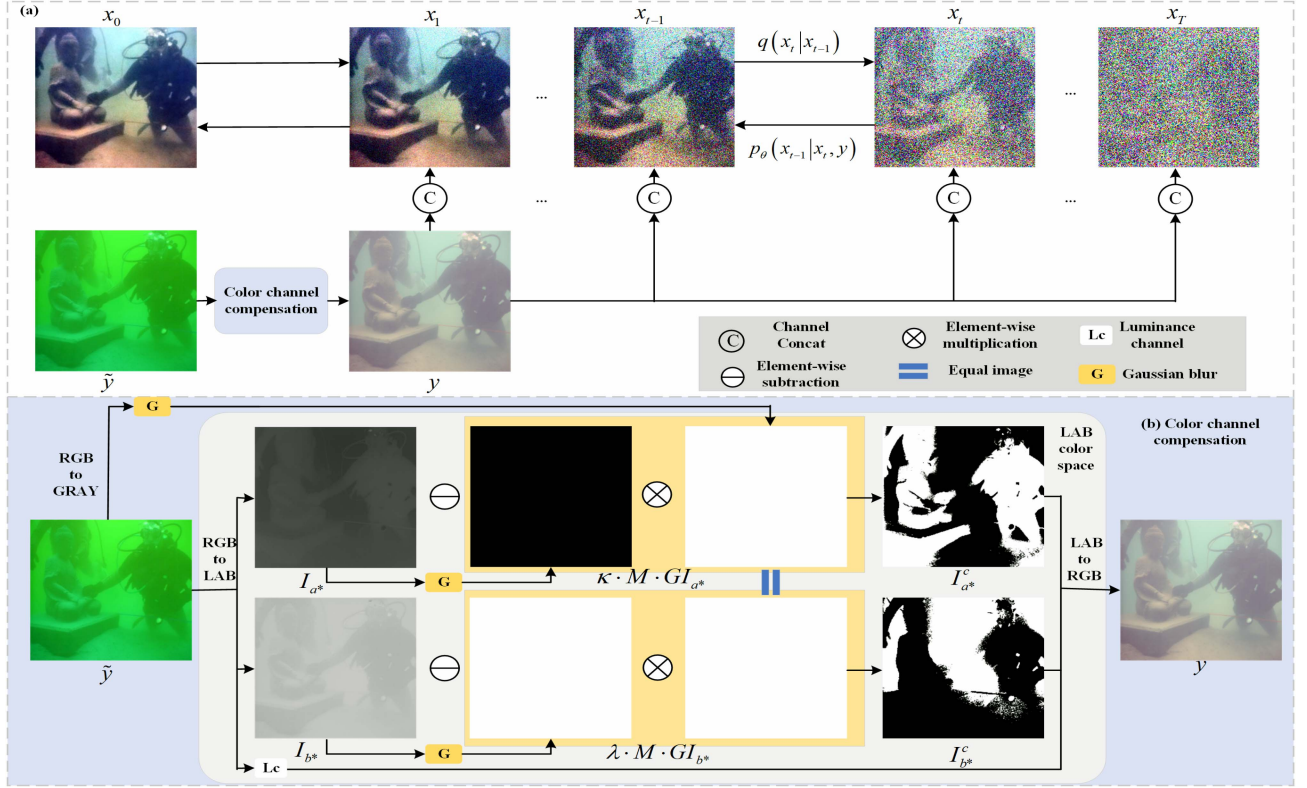


Fig. 2. Architecture of the proposed DiffWater method. (a) Forward diffusion and inverse diffusion processes. (b) 3C. The degraded underwater image \tilde{y} obtains the conditional image y through 3C [43], where in the forward diffusion process q (from left to right), Gaussian noise is gradually added to the underwater reference image. The reverse inference process p (from right to left) uses the conditional image y and the noise image concatenated in the channel dimension as conditions to iteratively denoise the target image.

(3C) [43], which reconstructs lost channels utilizing complementary color channels

$$I_{a^*}^c(x) = I_{a^*}(x) - \kappa \cdot M(x) \cdot GI_{a^*}(x) \quad (2)$$

$$I_{b^*}^c(x) = I_{b^*}(x) - \lambda \cdot M(x) \cdot GI_{b^*}(x) \quad (3)$$

where $I_{a^*}^c$ and $I_{b^*}^c$ are the compensated chromatic channels, I_{a^*} and I_{b^*} are the original chromatic channels of the underwater image, κ and λ are two parameters used to adjust the degree of compensation for the two opposing chromatic channels, and G is the Gaussian blur. According to the experimental results of 3C [43], a value of approximately 0.7 is generally used to achieve better results. M is a mask used to avoid excessive brightness changes at the location of the light source, where its value is zero for pixels with average brightness greater than 0.85 and one otherwise. To avoid artifacts, the mask M is smoothed with a Gaussian filter. Specifically, as shown in Fig. 2(b), the degraded underwater RGB image is initially transformed into the Lab color space, yielding the L , a , and b channels. The a and b channels of the Lab space, represented by I_{a^*} and I_{b^*} in (2), respectively, are then subjected to the following processing steps. First, the degraded underwater RGB image is converted into a grayscale image and undergoes mask processing to generate the mask $M(x)$. Subsequently, for the I_{a^*} channel, it is compensated by subtracting the elementwise multiplication of the result obtained by applying Gaussian blur and the mask

$M(x)$, aiming to restore the detailed information within the I_{a^*} channel. A similar processing approach is applied to the I_{b^*} channel. It is compensated by subtracting the elementwise multiplication of the result obtained by applying Gaussian blur and the mask $M(x)$ to restore the detailed information within the I_{b^*} channel. Finally, the processed Lab image is converted back to the RGB color space, resulting in the color-compensated image. We aimed to identify the optimal denoising conditions for underwater images by attempting various physical priors, including utilizing UDCP [19] as a guiding criterion (see Section IV-F).

B. Architecture of the Proposed DiffWater Method

Distinct from the reverse diffusion process of DDPM, the goal of the UIE task is to iteratively refine the mapping from an underwater degraded image to an underwater reference image. This aims to approximate the distribution $p(x_0|y)$, where y represents the color-compensated underwater degraded image and x_0 represents the underwater reference image.

DiffWater [see Fig. 2(a)] consists of forward and reverse diffusion processes. In the forward diffusion, Gaussian noise is gradually introduced into x_0 via a Markov chain $(q(x_t|x_{t-1}))$ resulting in x_T . In the reverse diffusion, x_T is iteratively refined through T steps, introducing y at each step based on $p_\theta(x_{t-1}|x_t, y)$ to reconstruct $x_0 \sim p(x_0|y)$.

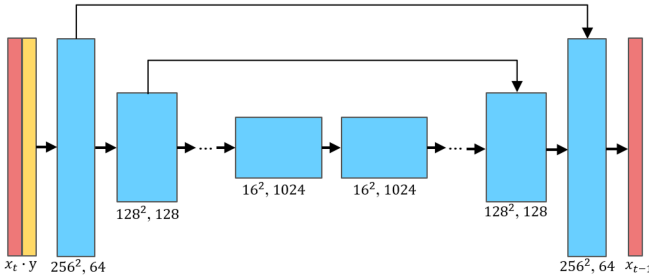


Fig. 3. Description of the U-Net architecture with skip connections. 3C [43] is applied to the degraded underwater image to obtain the input image y , which is then concatenated with the noisy underwater reference image x_t .

To improve the clarity and quality of underwater images, the DiffWater architecture adopts the SR3 network [36] as the denoising model. As shown in Fig. 3, it uses a U-Net structure containing 3 ResNet blocks. First, the channel number is set to 64 for the first layer, and the depth multipliers are $\{1, 2, 4, 8, 16\}$. Second, the degraded image is color compensated and then concatenated with the input noisy image to guide the denoising process.

In the following section, we first outline the forward diffusion procedure and then discuss how to train the denoising model f_θ and employ it for inference.

C. DiffWater: Diffusion Process

First, a forward Markov chain q is defined [15], which, through T iterations, gradually introduces Gaussian noise into the underwater reference image x_0 according to the variance sequence $\beta_1, \beta_2, \dots, \beta_T$, until it eventually becomes a completely random noisy image x_T

$$q(\mathbf{x}_{1:T} | \mathbf{x}_0) = \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1}) \quad (4)$$

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t | \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}) \quad (5)$$

where t denotes the diffusion step and the scalar parameter $\beta_t \in (0, 1)$ is a hyperparameter that determines the variance of the Gaussian noise introduced at each iteration. The parameter β_t is utilized to quantify the intensity of noise at each step. During the forward diffusion process, the noise gradually increases with the increment of step T . On the other hand, the inverse diffusion process starts from fully Gaussian noise and functions as a denoising process. β_t serves as a crucial parameter determining the strength of noise at each step. The model incorporates embeddings of the input time step t , enabling parameter sharing across time. As the time step t undergoes linear transformation, the noise intensity β_t linearly adjusts to accommodate different time steps.

Letting $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t := \prod_{s=1}^t \alpha_s$, x_t can be sampled in closed form at any time step t

$$q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I}). \quad (6)$$

To simplify notation, we let $\gamma_t = \bar{\alpha}_t$, expressing (6) as

$$q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\gamma_t} \mathbf{x}_0, (1 - \gamma_t) \mathbf{I}) \quad (7)$$

where β_t decreases with increasing t to ensure that the variance of the random variable remains bounded as $t \rightarrow \infty$. After reparameterization, we obtain the noisy image x_t

$$\mathbf{x}_t = \sqrt{\gamma_t} \mathbf{x}_0 + \sqrt{1 - \gamma_t} \boldsymbol{\epsilon} \quad (8)$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$ follows the standard Gaussian distribution. Next, we will discuss how to learn a neural network to reverse this forward diffusion process.

D. DiffWater: Inverse Diffusion by Adding Iterative Refinement of Conditions

In the proposed DiffWater, the inference is defined as a reverse Markov process that is the opposite of the forward diffusion process. Inference starts from a Gaussian noise x_T

$$p_\theta(\mathbf{x}_{0:T} | \mathbf{y}) = p(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{y}) \quad (9)$$

$$p(\mathbf{x}_T) = \mathcal{N}(\mathbf{x}_T | \mathbf{0}, \mathbf{I}) \quad (10)$$

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{y}) = \mathcal{N}(\mathbf{x}_{t-1} | \mu_\theta(\mathbf{y}, \mathbf{x}_t, t), \sigma_t^2 \mathbf{I}) \quad (11)$$

where y represents the input degraded image after color compensation, and θ represents the model parameters.

Furthermore, the inference process is defined according to the isotropic Gaussian conditional [36], which starts from Gaussian noise x_T and reverses the forward diffusion process to finally obtain the target image x_0 . This inference process contains T iterative refinement steps, each step uses the conditional distribution $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{y})$, which is learned by the neural network model f_θ . If the variance of the noise in the forward diffusion process is set small enough (i.e., $\alpha_{1:T} \approx 1$), the optimal reverse process $p(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{y})$ will approximate a Gaussian distribution [34]. Therefore, choosing a Gaussian conditional distribution in the inference process (11) can provide a reasonable fit. At the same time, $1 - \gamma_T$ is large enough to ensure that x_T conforms to the prior $p(\mathbf{x}_T) = (\mathbf{x}_T | 0, \mathbf{I})$, thus making the sampling process start from completely random Gaussian noise.

According to Ho et al.'s equations [15] for $\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0) := \frac{\sqrt{\bar{\alpha}_{t-1}} \beta_t}{1 - \bar{\alpha}_t} \mathbf{x}_0 + \frac{\sqrt{\bar{\alpha}_t} (1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} \mathbf{x}_t$ and $\tilde{\beta}_t := \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_t$, the reverse process can be parameterized as follows:

$$\mu_\theta(\mathbf{x}, \mathbf{y}_t, \gamma_t) = \frac{\sqrt{\gamma_{t-1}} (1 - \alpha_t)}{1 - \gamma_t} \mathbf{x}_0 + \frac{\sqrt{\bar{\alpha}_t} (1 - \gamma_{t-1})}{1 - \gamma_t} \mathbf{x}_t \quad (12)$$

$$\sigma_t^2 \mathbf{I} = \frac{(1 - \gamma_{t-1})(1 - \alpha_t)}{1 - \gamma_t}. \quad (13)$$

The key is to train a denoising score model f_θ that can estimate the noise level β_t given a noisy image x_t and diffusion timestep t . During training, we use the denoising model f_θ to estimate ϵ . Therefore, given a noisy image x_t , by replacing ϵ with $f_\theta(y, x_t, t)$ in (8) and rearranging terms, we obtain an expression approximating the target image x_0

$$\hat{\mathbf{x}}_0 = \frac{1}{\sqrt{\gamma_t}} \left(\mathbf{x}_t - \sqrt{1 - \gamma_t} f_\theta(\mathbf{y}, \mathbf{x}_t, t) \right). \quad (14)$$

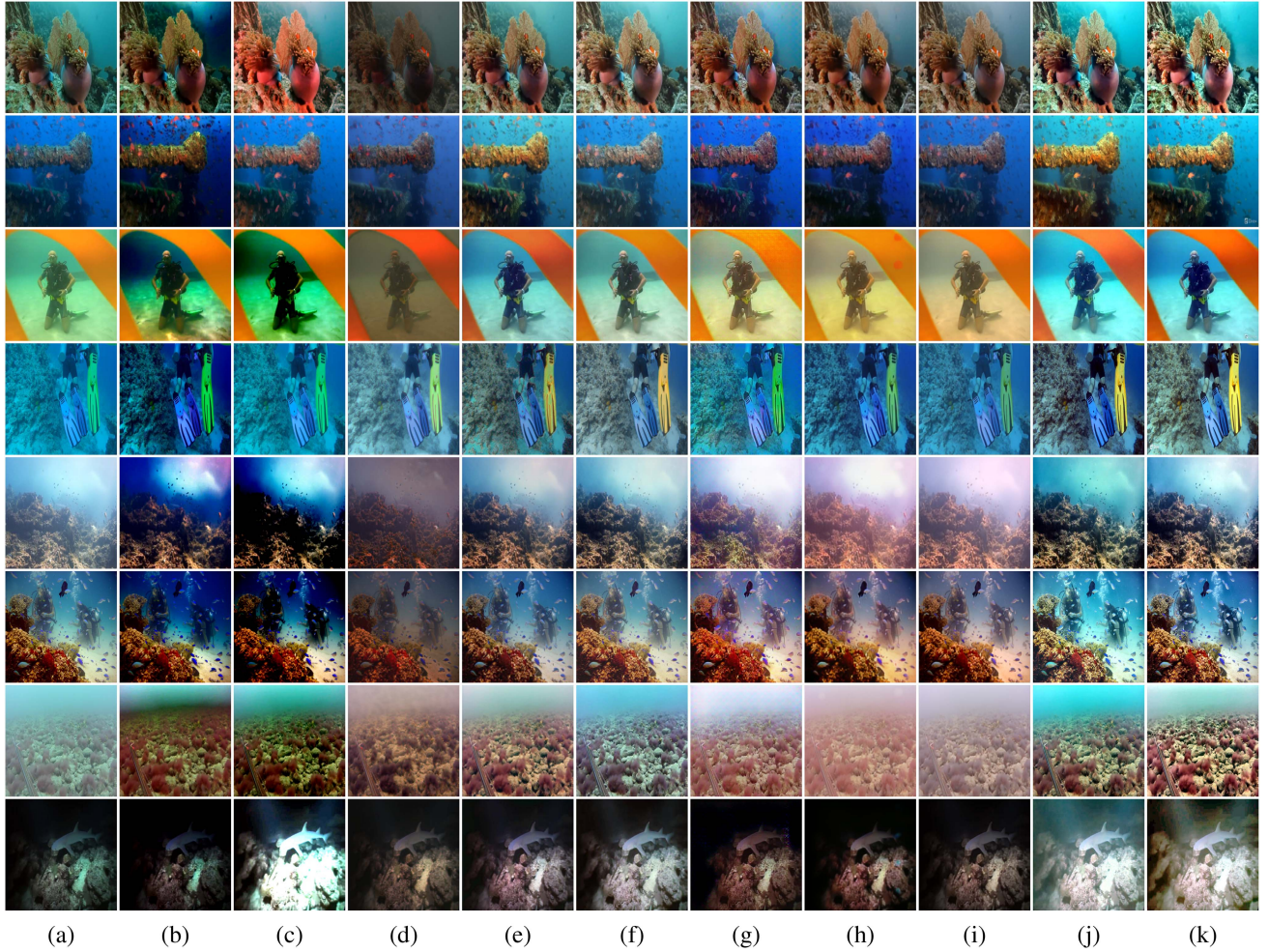


Fig. 4. Visual effects of eight different enhancement methods and our method on the TEST-U90 [24] dataset. (a) Raw. (b) UDCP [19]. (c) UIBLA [51]. (d) UWCNN-typeII [23]. (e) WaterNet [24]. (f) Ucolor [25]. (g) FUnIEGAN [32]. (h) MLFcGAN [31]. (i) Shallow-uwnet [33]. (j) DiffWater. (k) Reference.

By solving (12) and (14), the mean of $p_\theta(x_{t-1}|x_t, y)$ in (11) can be parameterized as

$$\mu_\theta(\mathbf{y}, \mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \gamma_t}} f_\theta(\mathbf{y}, \mathbf{x}_t, t) \right) \quad (15)$$

where α_t and γ_t denote the variance parameters for injecting noise in the forward diffusion process, respectively.

After this parameterization, each refinement step of our method samples from

$$\mathbf{x}_{t-1} \leftarrow \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \gamma_t}} f_\theta(\mathbf{y}, \mathbf{x}_t, t) \right) + \frac{1 - \gamma_{t-1}}{1 - \gamma_t} \beta_t \epsilon_t \quad (16)$$

where $\epsilon_t \sim \mathcal{N}(0, I)$. This is similar to a Langevin dynamics [44] step where f_θ provides an estimate of the data log density gradient.

In the T-step inverse diffusion process, the color channel compensated image y is obtained through 3C operation and is applied at each step. When only the degraded image \tilde{y} is used as the input condition, more information can be provided to the neural network. The degraded image \tilde{y} reflects the structural

and contour features of the original image x_0 , aiding the neural network in accurately predicting the noise ϵ and faithfully restoring the original image x_0 . However, due to severe degradation in underwater images, we obtain a preliminarily optimized underwater degraded image y through color compensation as the input condition, effectively providing prior knowledge to the neural network. In this manner, the neural network can more effectively utilize the variations in noise variance β_t , assisting in accurate noise prediction and distortion reduction, thereby enhancing the effectiveness and quality of image restoration. Algorithm 2 gives the pseudocode for the denoising process of the proposed DiffWater.

In the inverse diffusion process of the DiffWater method, a neural network is employed to predict the noise at each step. To alleviate the computational burden, a relatively simple U-Net model is commonly used. The input to this neural network consists of the current image x_t and the time step t , whereas the output is the noise estimation ϵ . During the training process, the neural network is utilized only once to predict the noise. However, in the reverse diffusion process, T iterations are required, with T typically set to around 1000. In each

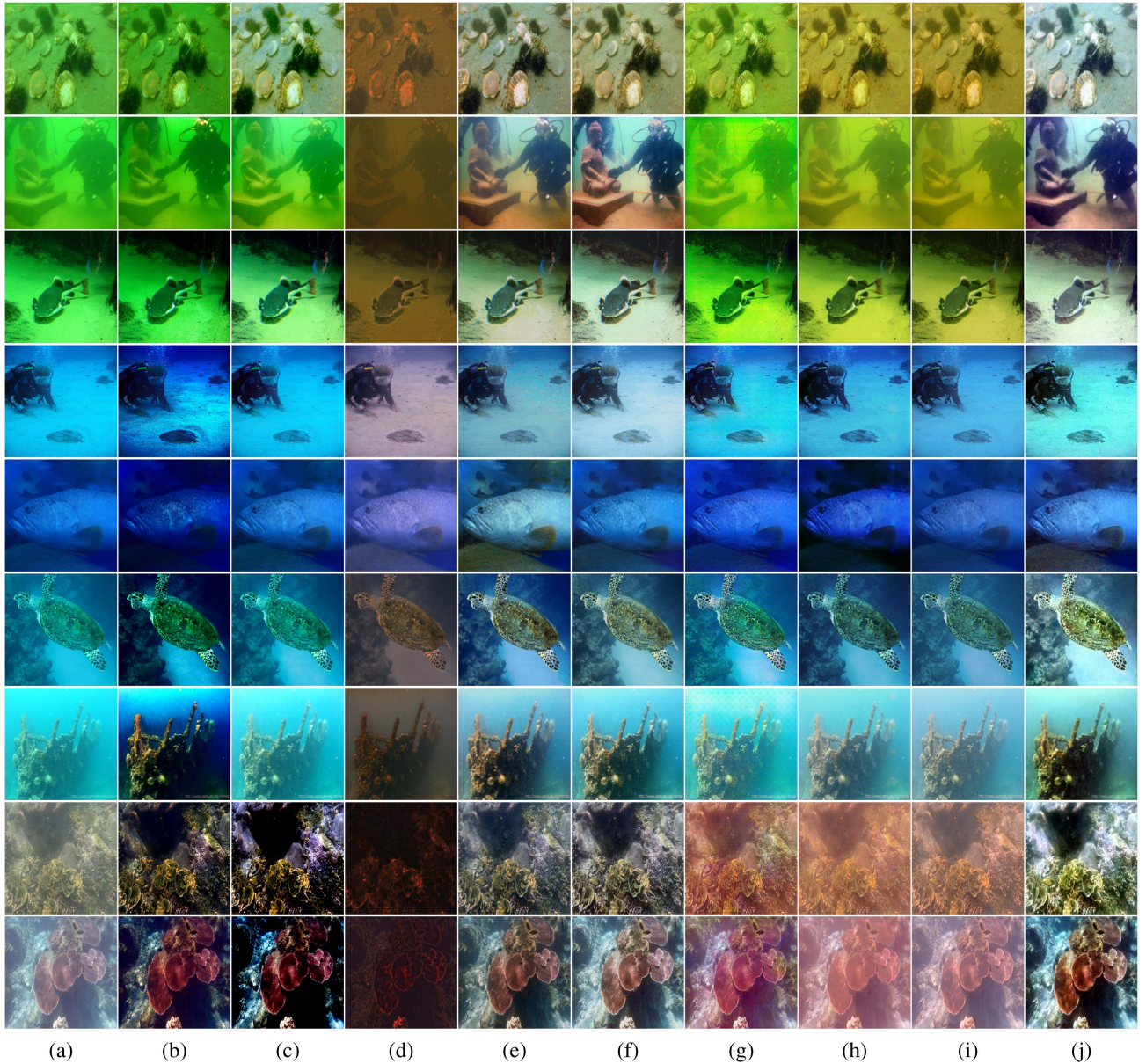


Fig. 5. Visual effects of eight different enhancement methods and our method on the TEST-U45 [46] dataset. (a) Raw. (b) UDPCP [19]. (c) UIBLA [51]. (d) UWCNN-typeII [23]. (e) WaterNet [24]. (f) Ucolor [25]. (g) FUnIEGAN [32]. (h) MLFcGAN [31]. (i) Shallow-uwnet [33]. (j) DiffWater.

iteration, the neural network is used to predict the noise, and a deterministic formula is employed to update the image x_t . While the computational load during training is not significant, it becomes time-consuming and increases the computational burden during testing, as the process needs to be repeated T times. This challenge is not exclusive to the proposed DiffWater method but is commonly encountered by all diffusion models.

E. Loss Function

To achieve controllable image generation, we introduce the conditional image y as additional side information and optimize a neural network denoising model f_θ [36] during the reverse diffusion process. The function $f_\theta(y, x_t, t)$, which takes the

conditional image y , the noisy target image x_t , and time step t as input, is trained to predict the noise vector ϵ . The loss function can, thus, be formulated as

$$\text{Loss} = \mathbb{E}_{(x,y)} \mathbb{E}_{\epsilon,t} \left\| f_\theta(y, \underbrace{\sqrt{\gamma_t}x_0 + \sqrt{1-\gamma_t}\epsilon}_{x_t}, t) - \epsilon \right\|_p^p \quad (17)$$

where $\epsilon \sim \mathcal{N}(0, I)$ and $p = 1$. By integrating the denoising model into the reverse diffusion process, we can control the details and diversity of the generated images. Algorithm 1 gives the pseudocode for the DiffWater training process.

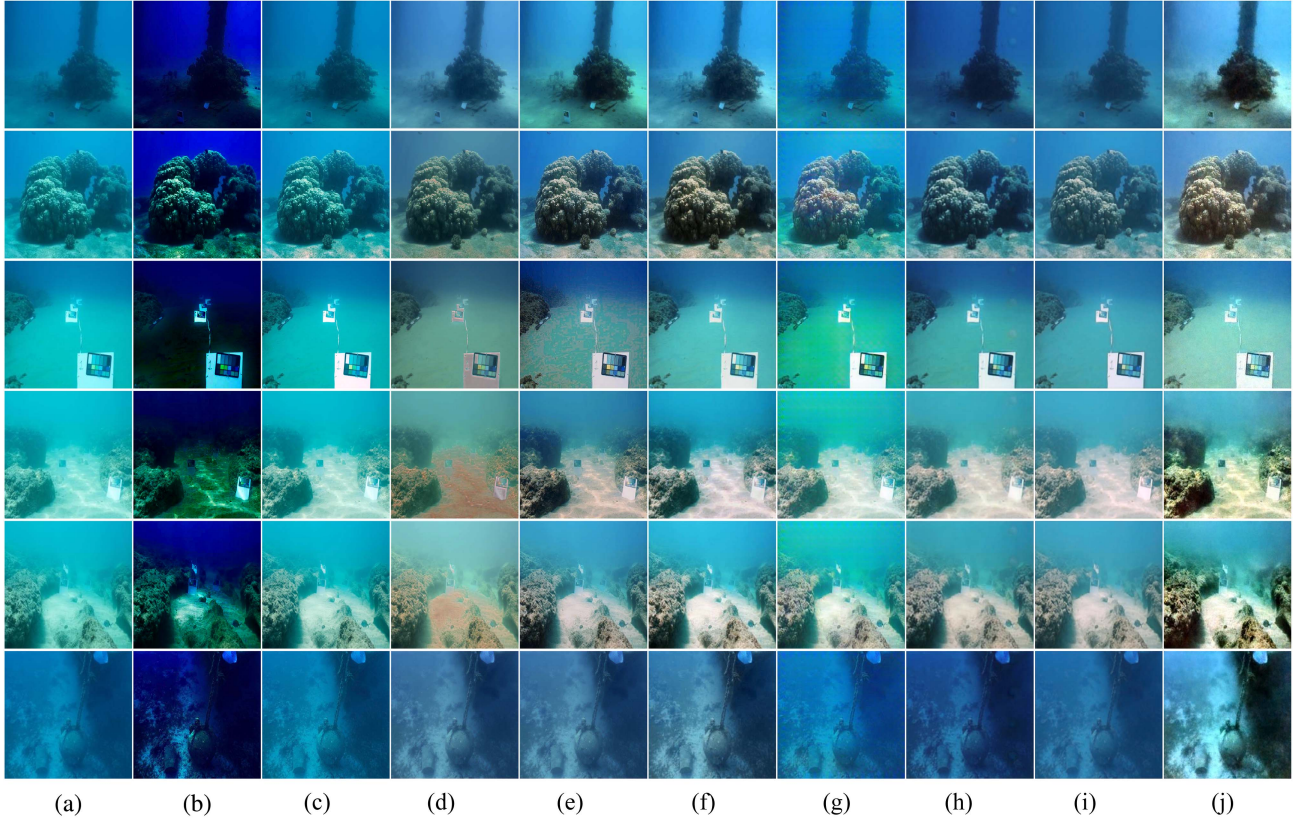


Fig. 6. Visual effects of eight different enhancement methods and our method on the TEST-S16 [20] dataset. (a) Raw. (b) UDCP [19]. (c) UIBLA [51]. (d) UWCNN-typeII [23]. (e) WaterNet [24]. (f) Ucolor [25]. (g) FUNIEGAN [32]. (h) MLFCGAN [31]. (i) Shallow-uwnet [33]. (j) DiffWater.

Algorithm 1: Training a Denoising Model f_θ .

Require: Paired underwater reference image \mathbf{x} and underwater degraded image $\tilde{\mathbf{y}}$, Color Channel Compensation (f_{3C}).

- 1: $\mathbf{y} = f_{3C}(\tilde{\mathbf{y}})$
- 2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
- 3: $t \sim \text{Uniform}(\{1, \dots, T\})$
- 4: $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
- 5: Take a gradient descent step on $\|\nabla_\theta f_\theta(\mathbf{y}, \sqrt{\gamma_t}\mathbf{x}_0 + \sqrt{1-\gamma_t}\epsilon, t) - \epsilon\|$
- 6: **until** converged

Algorithm 2: Inference in T Iterative Refinement Steps.

Require: Underwater degraded image $\tilde{\mathbf{y}}$, Color Channel Compensation (f_{3C}).

- 1: $\mathbf{y} = f_{3C}(\tilde{\mathbf{y}})$
- 2: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
- 3: **for** $t = T, \dots, 1$ **do**
- 4: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
- 5: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\gamma_t}}f_\theta(\mathbf{y}, \mathbf{x}_t, t)) + \frac{1-\gamma_{t-1}}{1-\gamma_t}\beta_t\mathbf{z}$
- 6: **end for**
- 7: **return** \mathbf{x}_0

IV. EXPERIMENTS

A. Implementation Details

The training of the proposed DiffWater method was based on the Pytorch 1.12.1 framework on NVIDIA RTX A5000 GPU. In all experiments, cropped 256×256 pixel patches were adopted as input, and the number of diffusion steps T was adopted as 2000. To improve method performance, the variance of the forward diffusion process was increased linearly from α_1 to α_T , in the range from 10^{-6} to 10^{-2} [36]. The initial learning rate was set as 3×10^{-6} , the Adam optimizer with hyperparameters $\beta \in (0.9, 0.999)$ was used and the batch size was set to 1 for training. The method was trained for one million iterations using

5004 paired underwater images from the LSUI dataset [45] and 800 paired underwater images from the UIEB dataset [24]. A total of 800 pairs of underwater images were excluded from the UIEB dataset for training purposes, leaving 90 pairs for further analysis. Subsequently, from this remaining set of images, three pairs were selected as the validation set. Finally, the pretrained model that exhibited the optimal performance on the validation set was employed for the evaluation of the remaining test sets.

B. Datasets and Evaluation Metrics

In this section, to demonstrate the effectiveness of the proposed DiffWater method, experiments were conducted on four public underwater image datasets as follows.

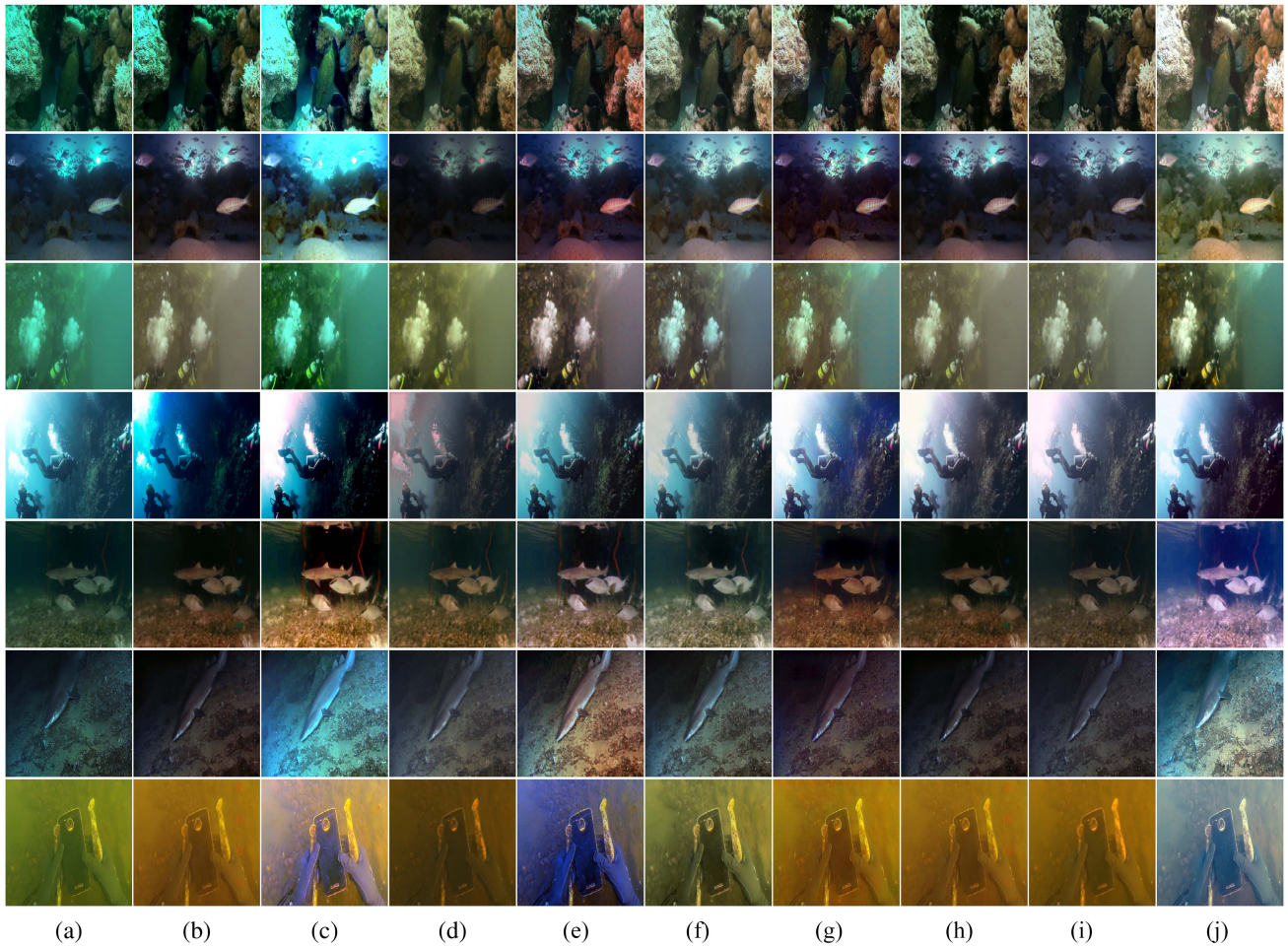


Fig. 7. Visual effects of eight different enhancement methods and our method on the TEST-C60 [24] dataset. (a) Raw. (b) UDCP [19]. (c) UIBLA [51]. (d) UWCNN-typeII [23]. (e) WaterNet [24]. (f) Ucolor [25]. (g) FUnIEGAN [32]. (h) MLFcGAN [31]. (i) Shallow-uwnet [33]. (j) DiffWater.

- 1) The UIEB dataset [24] is a benchmark dataset for UIE, containing 950 real underwater images of which 890 have corresponding high-quality reference images while the remaining 60 are challenging underwater images that cannot obtain satisfactory reference images. We used 800 image pairs in the UIEB dataset as the training set, 90 image pairs as TEST-U90, the 60 challenging underwater images as TEST-C60.
- 2) The LSUI dataset [45] is a large-scale underwater image dataset containing 5004 image pairs, with each pair consisting of a real underwater image and a corresponding high-quality reference image. This dataset involves richer underwater scenes including different illumination conditions, water types, and target categories, and has better visual quality than existing reference images.
- 3) The SQUID dataset [20] is a stereo underwater image dataset containing 57 stereo image pairs, with each image containing a color chart for evaluating the color restoration effect of underwater images. The benchmark test sampled 16 representative examples from the entire SQUID dataset, referred to as TEST-S16.
- 4) The U45 dataset [46] is a public underwater image test dataset containing 45 underwater images in different scenes, involving underwater degradation such as color cast, low contrast, and haze effects. The entire 45 images in the U45 dataset were used as the test set referred to as TEST-U45.

To evaluate the performance of different UIE methods, two full-reference metrics, PSNR [47] and SSIM [48], and two no-reference metrics, UCIQE [49] and UIQM [50], were utilized by us.

C. Comparing DiffWater With Other UIE Methods

In this article, as both diffusion models and GAN fall under the category of generative models, we specifically compare the proposed DiffWater method with GAN and diffusion models. Therefore, we will evaluate and compare the DiffWater method against two traditional methods (UDCP [19] and UIBLA [51]) as well as six deep-learning-based methods (UWCNN [23], Water-Net [24], Ucolor [25], MLFcGAN [31], FUnIEGAN [32], and Shallow-uwnet [33]). Among them, UWCNN, Shallow-uwnet, and Ucolor are CNN-based methods while MLFcGAN,

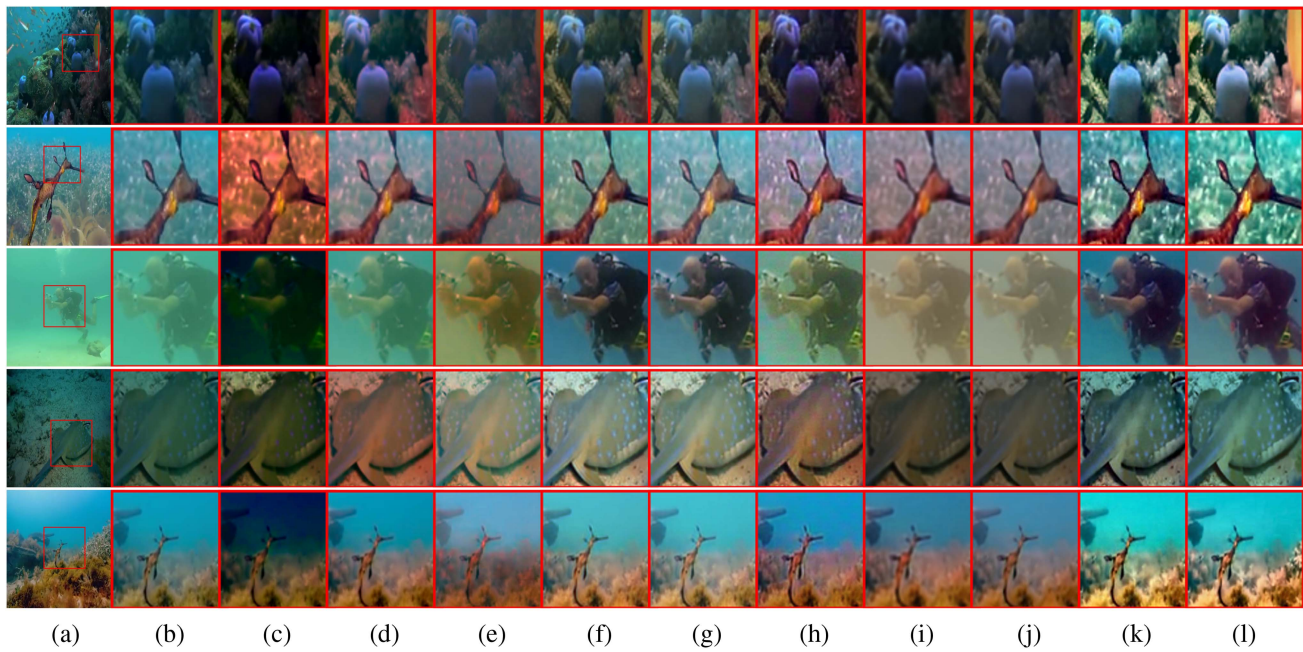


Fig. 8. Visual effect comparison of local image enlargement after applying eight different enhancement methods and our method was conducted on TEST-U90 [24] dataset. (a) Raw. (b) Raw-Locally enlarged. (c) UDCP [19]. (d) UIBLA [51]. (e) UWCNN-typeII [23]. (f) WaterNet [24]. (g) Ucolor [25]. (h) FUnIEGAN [32]. (i) MLFcGAN [31]. (j) Shallow-uwnet [33]. (k) DiffWater. (l) Reference.

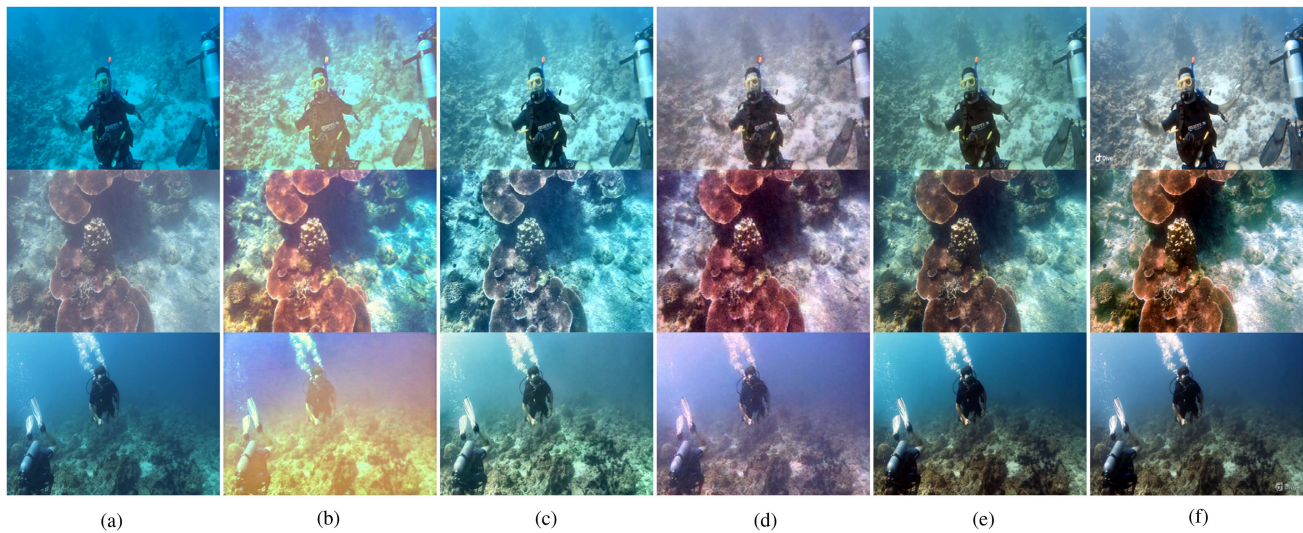


Fig. 9. Ablation study of the contributions of DiffWater. (a) RAW. (b) Result of the CDDPM-UDCP. (c) Result of the CDDPM-(3C+UDCP). (d) Result of the CDDPM- \tilde{g} . (e) Result of the DiffWater (proposed). (f) Result of the reference.

FUnIEGAN, and Water-Net are GAN-based methods. To ensure a fair comparison, both the proposed DiffWater method and the comparative methods were trained on two training datasets, LSUI and UIEB, and subsequently tested on other independent test datasets. Performances and comparison results are provided to ensure the robustness of our model’s performance across varying data distributions. Through comprehensive experimental comparisons of these methods, we validate the effectiveness of the proposed DiffWater method.

D. Qualitative Comparison

In this section, we used 800 pairs of images from the UIEB dataset [24] as the training set and 90 pairs of images as the test set, denoted as TEST-U90. The experimental results on the TEST-U90 set were evaluated using the full-reference image quality metrics PSNR [47] and SSIM [48], as shown in Table I.

The results indicate that the proposed DiffWater method has advantages in preserving the structural information and details of

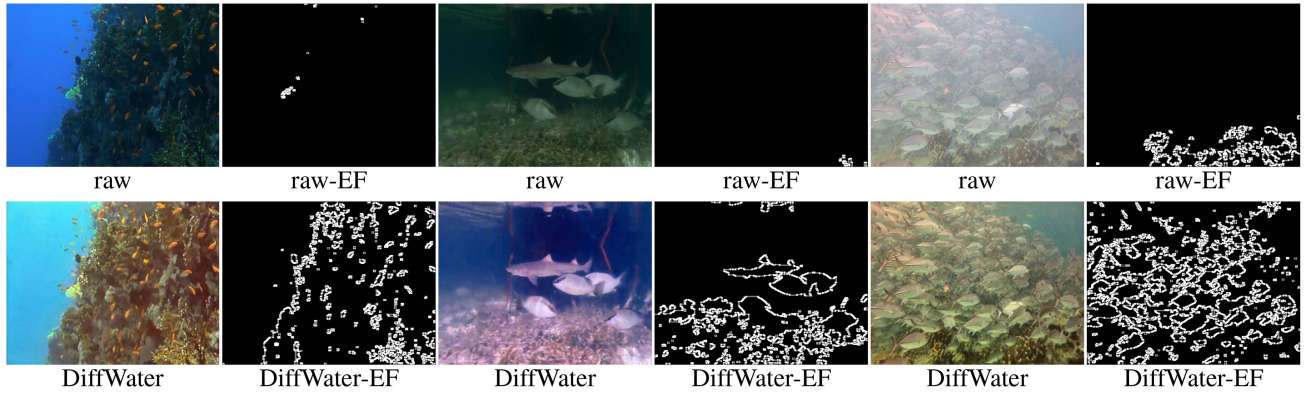


Fig. 10. Effect of DiffWater on edge extraction of input and enhanced images. (Top) Above is the Sobel extraction result of the input image. (Bottom) At the bottom is the Sobel extraction result enhanced by DiffWater.



Fig. 11. Effect of DiffWater for OD on input and enhanced images. (Top) The above figure shows the YOLOv7 OD result of the input image. (Bottom) The bottom is the target detection result of YOLOv7 enhanced by DiffWater.

TABLE I
PSNR AND SSIM AVERAGE SCORES OF DIFFERENT METHODS ON THE TEST-U90, AND THE BEST RESULTS ARE MARKED IN BOLD BLACK

	Method	TEST-U90			
		PSNR	SSIM	Train Time (s)	Test Time (s)
Traditional Method	UDCP [19]	11.3752	0.5156	/	31.842
	UIBLA [51]	15.1351	0.6454	/	42.531
CNN Method	UWCNN-typeII [23]	13.8223	0.6385	161.114	0.917
	Shallow-uwnet [33]	15.9740	0.7099	38.718	0.074
	Ucolor [25]	21.0005	0.8693	629.693	2.843
GAN Method	FUnIEGAN [32]	16.9189	0.7338	97.709	0.162
	MLFcGAN [31]	15.1004	0.6573	931.778	0.376
	Water-Net [24]	19.9223	0.8326	281.721	1.236
	DiffWater	20.9721	0.8951	80.356	48.517

the enhanced images compared to existing comparison methods, achieving the highest SSIM score. Compared to the second-best method, the proposed method achieved a relative increase of 2.58% in SSIM, despite a decrease of 2.84% in PSNR score. Although the proposed method's PSNR score is slightly lower than that of the Ucolor [25] method, it is still better than methods based on GANs and other comparison methods, indicating

that the proposed method has certain advantages in terms of fidelity. The DiffWater method emphasizes the preservation of image structure and texture details rather than just minimizing mean square error, highlighting its effectiveness in enhancing underwater images with structural sensitivity.

Fig. 4 shows the performance comparison of the proposed DiffWater method against other methods when enhancing images

TABLE II
AVERAGE SCORES OF UCIQE AND UIQM OF DIFFERENT METHODS ON THE TEST-U45 DATASET, AND THE BEST RESULTS ARE MARKED IN BOLD BLACK

	Method	TEST-U45	
		UIQM↑	UCIQE↑
	Raw	2.3821	0.3754
Traditional Method	UDCP [19]	3.3019	0.4546
	UIBLA [51]	2.6307	0.4557
CNN Method	UWCNN-typeII [23]	3.8481	0.3736
	Shallow-uwnet [33]	4.0523	0.3703
	Ucolor [25]	4.9504	0.4461
GAN Method	FUnIEGAN [32]	4.3502	0.3997
	MLFcGAN [31]	4.0229	0.3908
	Water-Net [24]	4.8603	0.4502
Diffusion Model	UW-DDPM [60]	4.8990	0.4454
	DiffWater	4.7306	0.4624

under different degraded underwater scenes. The experimental results demonstrate that the DiffWater method can effectively eliminate haze effects, significantly improve the clarity and visibility of image details, and outperform other methods in microstructure recovery. For blue scenes, DiffWater can effectively compensate for blue distortion, restore true colors, and preserve details and contrast, showing an obvious advantage in processing blue underwater images. Under low-light conditions, DiffWater can enhance image brightness and contrast, significantly improve image visibility, and has significant advantages in detail recovery and noise reduction. By comparing the visual enhancement results in Fig. 4, the effectiveness of DiffWater in enhancing degraded underwater images under different scenes is validated. Compared to other methods, DiffWater has obvious advantages in recovering image details, colors, and contrast. Additionally, as observed in Fig. 4, though some comparative methods perform better when processing hazy or yellowish images, none of the methods can provide satisfactory results for all applications. DiffWater demonstrates powerful generalization for different underwater images, is able to maintain the structural and textural details of images, and achieves photorealistic restoration.

E. Evaluation on Other Datasets

In this section, the underwater images are evaluated using the UCIQE and UIQM metrics. A total of 5004 images from the LSUI dataset are used as the training dataset [45], and the enhancement results of eight methods are compared under different underwater environment categories. To verify the effectiveness of these methods, 45 images from the U45 dataset, denoted as TEST-U45 [46], 16 typical images from the SQUID dataset, denoted as TEST-S16 [20], and 60 challenging underwater images from the UIEB dataset, denoted as TEST-C60 [24], are used for testing. The model’s performance will be evaluated using the test dataset, but retraining will not be performed on these test data. The enhancement results of eight methods are compared under different underwater environment categories.

The results in Table II show that on the TEST-U45 dataset, the DiffWater method outperforms the other eight methods in

TABLE III
AVERAGE SCORES OF UCIQE AND UIQM OF DIFFERENT METHODS ON THE TEST-S16 DATASET, AND THE BEST RESULTS ARE MARKED IN BOLD BLACK

	Method	TEST-S16	
		UIQM↑	UCIQE↑
	Raw	2.1131	0.3590
Traditional Method	UDCP [19]	1.4920	0.4431
	UIBLA [51]	1.4269	0.4042
CNN Method	UWCNN-typeII [23]	3.2716	0.3698
	Shallow-uwnet [33]	3.1680	0.3861
	Ucolor [25]	3.5754	0.4185
GAN Method	FUnIEGAN [32]	2.8500	0.3955
	MLFcGAN [31]	3.3020	0.4172
	Water-Net [24]	3.4989	0.4307
	DiffWater	4.5159	0.4503

TABLE IV
AVERAGE SCORES OF UCIQE AND UIQM OF DIFFERENT METHODS ON THE TEST-C60 DATASET, AND THE BEST RESULTS ARE MARKED IN BOLD BLACK

	Method	TEST-C60	
		UIQM↑	UCIQE↑
	Raw	2.9388	0.359
Traditional Method	UDCP [19]	2.7336	0.3926
	UIBLA [51]	3.0864	0.4081
CNN Method	UWCNN-typeII [23]	3.3557	0.3348
	Shallow-uwnet [33]	4.2963	0.3348
	Ucolor [25]	4.3337	0.3846
GAN Method	FUnIEGAN [32]	4.2946	0.3592
	MLFcGAN [31]	4.2923	0.3681
	Water-Net [24]	4.4453	0.4416
Diffusion Model	UW-DDPM [60]	4.8381	0.4288
	DiffWater	4.6559	0.4339

the UIQM metric. Through evaluation on the TEST-U45 dataset, the DiffWater method has shown excellent performance in the UIQM metric, highlighting its advantage in improving image quality. Although it is slightly inferior to the Ucolor method in the UCIQE metric, the DiffWater method exhibits comprehensive performance in overall visual effect and dual metric evaluation.

The results in Table III show that the DiffWater method exhibits relatively high performance on the TEST-S16 dataset, as evaluated by the UIQM and UCIQE metrics. The UIQM metric evaluates overall perceived image quality while the UCIQE metric focuses on color quality. The high scores of the DiffWater method in both metrics indicate its superior performance in handling the TEST-S16 dataset. The results of the UIQM and UCIQE metrics demonstrate that the DiffWater method has significant advantages in improving image quality and color restoration. These findings further validate the effectiveness and feasibility of the DiffWater method in the UIE task.

The results in Table IV indicate that through evaluation on the TEST-C60 dataset, and the DiffWater method outperforms the other eight methods in the UIQM metric, demonstrating its excellent performance in improving image quality. Although it is slightly inferior to the Water-Net method in the UCIQE metric, the DiffWater method exhibits comprehensive performance in overall visual effect and dual metric evaluation.

As shown in Figs. 5–7, the UDCP method [19] improved the image contrast by increasing artificial colors and reduced haziness in the image, but the enhanced image color tone became dark, resulting in unrealistic visual effects. The UIBLA method [51] improved the image quality by the local adaptive method, but its performance is limited by the characteristics of underwater imaging, and in some conditions cannot achieve ideal enhancement effects. The UWCNN method [23] has a certain color shift correction and detail recovery capability, but its enhancement effect is limited, and when processing complex underwater images with severe a color shift will produce red shift. The Water-Net method [24] has limited generalization ability for different underwater environments, and the generated images have artifacts and unnatural colors. The Ucolor method [25] can effectively enhance the image, but local overenhancement will still cause color shift in the image. The FUnIEGAN method [32], MLFcGAN method [31], and Shallow-wnet method [33] will produce images with yellow shift when processing images with green shift, resulting in color distortion and missing details.

In general, traditional UIE methods (UDCP [19] and UIBLA [51]) perform well in improving image brightness and saturation, but cannot effectively eliminate haziness, color shift, and chromatic aberration in underwater images, and are likely to cause overenhancement or distortion of images. CNN methods (UWCNN [23], Water-Net [24], and Ucolor [25]) perform best in maintaining image details and textures while also being able to restore the color and contrast of the image. However, the drawback of this method is that it cannot adapt well to different water quality and lighting conditions, which may cause the image to be too bright or too dark, or have color distortions and noise. GAN methods (FUnIEGAN [32], MLFcGAN [31], and Shallow-wnet [33]) perform best in restoring the real scene of the image, making the underwater image more realistic. However, the method performs relatively poorly in maintaining image details and textures, and the generated image has problems of sunlight or shadows underwater. In contrast, DiffWater shows higher quality in underwater image recovery, with higher clarity, contrast, color recovery, and naturalness. It is able to achieve consistent color and detail recovery of the real underwater scene with statistical characteristics of natural scenes, in line with human visual perception standards.

According to Fig. 8, the original underwater image may exhibit issues such as low contrast, noise, blurriness, distortion, and color shifts when locally magnified. The locally magnified images enhanced by the UDCP method show darker and color-shifted characteristics. The UIBLA method does not exhibit a significant improvement in defogging the locally magnified images. The UWCNN method results in globally reddish and blurry locally magnified images. The visually appealing results are observed in the locally magnified images enhanced by the Waternet and Ucolor methods; however, they still differ from the reference image. The FUnIEGAN, MLFcGAN, and Shallow-wnet methods result in locally magnified images with low contrast and a yellow color bias. In contrast, our proposed method demonstrates enhanced underwater images with higher contrast, improved visibility of details, and enhanced image

quality after local magnification, which closely resemble the reference image.

F. Ablation Study

In order to further validate the effectiveness of our proposed method, this study employed three conditional guidance methods and used 800 pairs of images from the UIEB dataset [25] as the training data, with TEST-U90 as the test set, to conduct ablation experiments and evaluate the impact of these four conditional guidance methods on the enhancement effect of underwater images.

Specifically, we conducted the following ablation experiments.

- 1) We removed the color correction of DiffWater, denoted as CDDPM-UDCP, and only used the degraded image y processed by UDCP as the conditional guidance for enhancing underwater images.
- 2) We removed the color correction of DiffWater, denoted as CDDPM- \tilde{y} , and only used the degraded image y as the conditional guidance for enhancing underwater images.
- 3) On the basis of DiffWater, we used the degraded image y processed by UDCP, denoted as CDDPM-(3C+UDCP), and added the two images together as the conditional guidance for enhancing underwater images.

Based on the results in Table V, we draw the following conclusions: Compared with the CDDPM-UDCP, CDDPM- y , and CDDPM-(3C+UDCP) methods, the DiffWater method exhibits better performance, achieving optimal results in all metrics. Fig. 9 provides a visual comparison of the four methods. Under the same experimental conditions, it can be observed that the images enhanced by the CDDPM-UDCP model are overall yellowish and distorted, with poor enhancement effect; the images enhanced by the CDDPM- \tilde{y} and CDDPM-(3C+UDCP) methods have higher brightness and more severe distortion, but with poor enhancement effect. Therefore, the proposed DiffWater method has a significant enhancement effect on underwater images. Through the previous experiments, it is demonstrated that under the guidance of degraded images that have undergone color correction, the DiffWater method can effectively improve the color fidelity and detail clarity of underwater images, thereby enhancing image quality.

G. Applicability Analysis

The enhanced images have brightness, high contrast, and color, which are very beneficial for advanced visual tasks and can effectively improve the applicability and performance of tasks, thereby bringing better results for underwater visual applications. To verify whether the enhanced images effectively enrich edge and feature information, various tests were conducted on the degraded and enhanced images. First, the Sobel operator was used to extract image edges [52] and the number of edge features before and after enhancement were compared. Second, we used the YOLOv7 algorithm for object detection (OD) tasks [53] to evaluate the impact of enhanced images on OD accuracy. Finally, we used a saliency detection (SD) method [54] to evaluate image quality and compared the number of salient features before and

TABLE V
 QUANTITATIVE RESULTS OF THE WAY OF THE GUIDANCE ON THE TESTING DATASETS WERE MEASURED BY THE AVERAGE PSNR AND SSIM VALUES

Ablation		Condition			
Baselines		full model (DiffWater)	CDDPM-UDCP	CDDPM- \bar{y}	CDDPM-(3C+UDCP)
TEST-U90	PSNR	20.9721	15.8436	16.5762	18.1253
	SSIM	0.8951	0.7658	0.8299	0.8210

The highest score is marked in bold black.

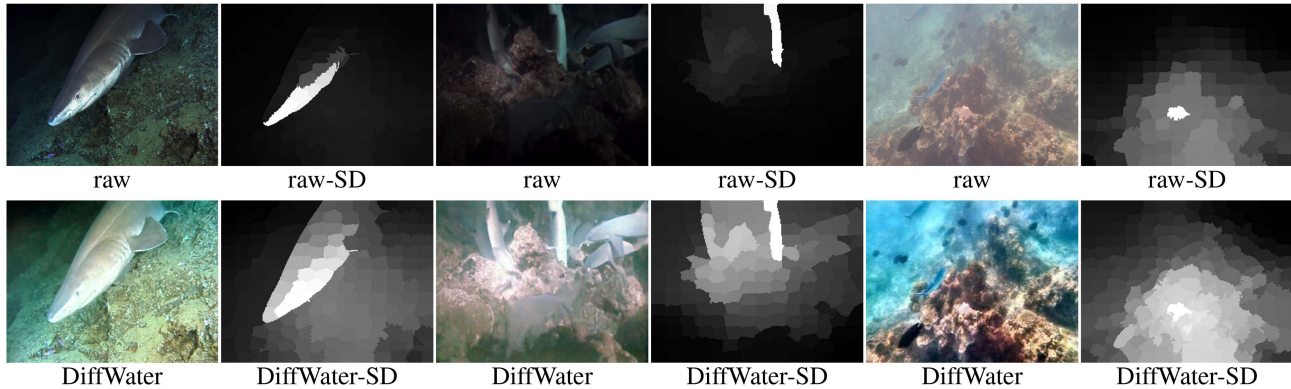


Fig. 12. Effect of DiffWater on SD of input and enhanced images. (Top) Above is the SD result of the input image. (Bottom) The bottom shows the enhanced SD results of DiffWater.

after enhancement. These tests comprehensively evaluated the applicability and effectiveness of enhanced images in advanced visual tasks.

The test results indicate that the DiffWater method has a significant improvement effect on UIE tasks. The enhanced images show significant improvements in edges and features, effectively enriching edge and feature information compared to degraded images, as demonstrated in Figs. 10–12. In addition, the method can significantly improve the accuracy of OD, further demonstrating its effectiveness in practical applications. These results suggest that the DiffWater method is a feasible and reliable UIE method, which can provide strong support and help for the field of underwater image processing.

V. CONCLUSION

In this article, we propose a novel UIE method based on the conditional DDPM (DiffWater). The proposed DiffWater method combines the benefits of DDPM and the characteristics of underwater imaging through introducing a color compensation method that can adapt to diverse water quality and illumination conditions to guide the diffusion and denoising process. It can generate clearer, more realistic, and naturally enhanced underwater imagery. The experimental results demonstrate that the proposed DiffWater method outperforms prevailing methods with respect to enhancement quality and efficacy, exhibiting better generalization capability and robustness. However, the proposed DiffWater method also possesses some limitations and shortcomings, such as reliance on large-scale data and slow sampling speed. In future, we will focus on ameliorating the algorithm, enhancing data quality, and optimizing the network architecture.

REFERENCES

- [1] W. Zhou, F. Zheng, G. Yin, Y. Pang, and J. Yi, “YOLOTrashCan: A deep learning marine debris detection network,” *IEEE Trans. Instrum. Meas.*, vol. 72, Nov. 2023, Art. no. 5002012.
- [2] Q. Qi et al., “Underwater image co-enhancement with correlation feature matching and joint learning,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1133–1147, Mar. 2022.
- [3] S. Sun, H. Guo, G. Wan, C. Dong, C. Zheng, and Y. Wang, “High-precision underwater acoustic localization of the black box utilizing an autonomous underwater vehicle based on the improved artificial potential field,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Feb. 2023, Art. no. 4202510.
- [4] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, “Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior,” *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5664–5677, Dec. 2016.
- [5] X. Cao, L. Ren, and C. Sun, “Dynamic target tracking control of autonomous underwater vehicle based on trajectory prediction,” *IEEE Trans. Cybern.*, vol. 53, no. 3, pp. 1968–1981, Mar. 2023.
- [6] H. Long, L. Shen, Z. Wang, and J. Chen, “Underwater forward-looking sonar images target detection via speckle reduction and scene prior,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Feb. 2023, Art. no. 5604413.
- [7] R. Hummel, “Image enhancement by histogram transformation,” *Comput. Graph Image Process.*, vol. 6, no. 2, pp. 184–95, 1975.
- [8] T. Ren et al., “Reinforced Swin-Convts transformer for simultaneous underwater sensing scene image enhancement and super-resolution,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2022, Art. no. 4209616.
- [9] Z. Liang, X. Ding, Y. Wang, X. Yan, and X. Fu, “GUDCP: Generalization of underwater dark channel prior for underwater image restoration,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 7, pp. 4879–4884, Jul. 2022.
- [10] Z. Shen et al., “UDAformer: Underwater image enhancement based on dual attention transformer,” *Comput. Graph.*, vol. 111, pp. 77–88, 2023.
- [11] Q. Jiang et al., “Two-step domain adaptation for underwater image enhancement,” *Pattern Recognit.*, vol. 122, 2022, Art. no. 108324.
- [12] I. Goodfellow et al., “Generative adversarial networks,” *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [13] Z. Fu et al., “Uncertainty inspired underwater image enhancement,” in *Proc. 17th Eur. Conf. Comput. Vis.*, 2022, pp. 465–482.
- [14] C. Fabbri, M. J. Islam, and J. Sattar, “Enhancing underwater imagery using generative adversarial networks,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 7159–7165.

- [15] J. Ho et al., "Denoising diffusion probabilistic models," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 6840–6851, 2020.
- [16] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 379–393, Jan. 2018.
- [17] Y. Jin et al., "Contrast enhancement by multiscale adaptive histogram equalization," *Proc. SPIE*, vol. 4478, pp. 206–213, 2001.
- [18] P. Li, Y. Huang, and K. Yao, "Multi-algorithm fusion of RGB and HSV color spaces for image enhancement," in *Proc. 37th Chin. Control Conf.*, 2018, pp. 9584–9589.
- [19] P. Drews, Jr. E. do Nascimento, F. Moraes, S. Botelho, and M. Campos, "Transmission estimation in underwater single images," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2013, pp. 825–830.
- [20] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2822–2837, Aug. 2021.
- [21] Z. Jiang, Z. Li, S. Yang, X. Fan, and R. Liu, "Target oriented perceptual adversarial fusion network for underwater image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 10, pp. 6584–6598, Oct. 2022.
- [22] Z. Wang et al., "Single underwater image enhancement using an analysis-synthesis network," 2021, *arXiv:2108.09023*.
- [23] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, pp. 107038–107049, 2020.
- [24] C. Li et al., "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, Nov. 2020.
- [25] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Trans. Image Process.*, vol. 30, pp. 4985–5000, May 2021.
- [26] X. Sun et al., "Deep pixel-to-pixel network for underwater image enhancement and restoration," *IET Image Process.*, vol. 13, no. 3, pp. 469–474, 2019.
- [27] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4217–4228, Dec. 2021.
- [28] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8107–8116.
- [29] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2242–2251.
- [30] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, "WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 387–394, Jan. 2018.
- [31] X. Liu, Z. Gao, and B. M. Chen, "MLFcGAN: Multilevel feature fusion-based conditional GAN for underwater image color correction," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 9, pp. 1488–1492, Sep. 2020.
- [32] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020.
- [33] A. Naik et al., "Shallow-UWnet: Compressed model for underwater image enhancement (student abstract)," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 18, pp. 15853–15854.
- [34] J. Sohl-Dickstein et al., "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proc. PMLR Int. Conf. Mach. Learn.*, 2015, pp. 2256–2265.
- [35] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *Proc. PMLR Int. Conf. Mach. Learn.*, 2021, pp. 8162–8171.
- [36] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4713–4726, Apr. 2023.
- [37] J. Ho et al., "Cascaded diffusion models for high fidelity image generation," *J. Mach. Learn. Res.*, vol. 23, no. 47, pp. 1–33, 2022.
- [38] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 10674–10685.
- [39] M. Mardani et al., "A variational perspective on solving inverse problems with diffusion models," 2023, *arXiv:2305.04391*.
- [40] O. Özdenizci and R. Legenstein, "Restoring vision in adverse weather conditions with patch-based denoising diffusion models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 10346–10357, Aug. 2023.
- [41] B. Kwar et al., "Denoising diffusion restoration models," *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 23593–23606, 2022.
- [42] Z. Zhang et al., "MetaUE: Model-based meta-learning for underwater image enhancement," 2023, *arXiv:2303.06543*.
- [43] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and M. Sbert, "Color channel compensation (3C): A fundamental pre-processing step for image enhancement," *IEEE Trans. Image Process.*, vol. 29, pp. 2653–2665, Nov. 2020.
- [44] M. Genkin et al., "Learning non-stationary Langevin dynamics from stochastic observations of latent trajectories," *Nature Commun.*, vol. 12, no. 1, 2021, Art. no. 5986.
- [45] L. Peng, C. Zhu, and L. Bian, "U-shape transformer for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 32, pp. 3066–3079, May 2023.
- [46] H. Li, J. Li, and W. Wang, "A fusion adversarial underwater image enhancement network with a public test dataset," 2019, *arXiv:1906.06819*.
- [47] J. Korhonen and J. You, "Peak signal-to-noise ratio revisited: Is simple beautiful?," in *Proc. 4th Int. Workshop Qual. Multimedia Experience*, 2012, pp. 37–38.
- [48] Z. Wang, 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.
- [49] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.
- [50] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE J. Ocean. Eng.*, vol. 41, no. 3, pp. 541–551, Jul. 2016.
- [51] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1579–1594, Apr. 2017.
- [52] N. Kanopoulos, N. Vasanthavada, and R. L. Baker, "Design of an image edge detection filter using the Sobel operator," *IEEE J. Solid-State Circuits*, vol. 23, no. 2, pp. 358–367, Apr. 1988.
- [53] C. Y. Wang et al., "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 7464–7475.
- [54] C. Yang, L. Zhang, and H. Lu, "Graph-regularized saliency detection with convex-hull-based center prior," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 637–640, Jul. 2013.
- [55] G. Han, M. Wang, H. Zhu, and C. Lin, "UIEGAN: Adversarial learning-based photo-realistic image enhancement for intelligent underwater environment perception," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, May 2023, Art. no. 5611514.
- [56] J. Yuan, Z. Cai, and W. Cao, "TEBCF: Real-world underwater image texture enhancement model based on blurriness and color fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2022, Art. no. 4204315.
- [57] H. Li, X. He, Y. Bai, F. Gong, D. Wang, and T. Li, "Restoration of wintertime ocean color remote sensing products for the high-latitude oceans of the Southern Hemisphere," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2022, Art. no. 4212312.
- [58] J. Yuan, W. Cao, Z. Cai, and B. Su, "An underwater image vision enhancement algorithm based on contour Bougie morphology," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8117–8128, Oct. 2021.
- [59] R. Chen, Z. Cai, and W. Cao, "MFFN: An underwater sensing scene image enhancement method based on multiscale feature fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2022, Art. no. 4205612.
- [60] S. Lu et al., "Underwater image enhancement method based on denoising diffusion probabilistic model," *J. Vis. Commun. Image Representation*, vol. 96, 2023, Art. no. 103926.
- [61] Y.-T. Peng, K. Cao, and P. C. Cosman, "Generalization of the dark channel prior for single image restoration," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2856–2868, Jun. 2018.
- [62] C. Li, J. Quo, Y. Pang, S. Chen, and J. Wang, "Single underwater image restoration by blue-green channels dehazing and red channel correction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2016, pp. 1731–1735.
- [63] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2822–2837, Aug. 2021.
- [64] S. Lu et al., "Speed-Up DDPM for real-time underwater image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: [10.1109/TCSVT.2023.3314767](https://doi.org/10.1109/TCSVT.2023.3314767).



Meisheng Guan is currently working toward the M.Sc. degree in mathematics with the School of Mathematics and Statistics, Ningbo University, Ningbo, China.
His research interests include computer vision and underwater image processing.



Yeyao Chen (Member, IEEE) received the B.S. degree in communication engineering and Ph.D. degree in information and communication engineering from Ningbo University, Ningbo, China, in 2017 and 2023, respectively.
He is currently a lecturer with the Faculty of Information Science and Engineering, Ningbo University, China. His current research interests include high dynamic range image/video processing, light field image processing, and deep learning.



Haiyong Xu (Member, IEEE) received the B.Sc. degree in applied mathematics from Jilin University, Changchun, China, in 2003, the M.Sc. degree in applied mathematics from Sun Yat-sen University, Guangzhou, China, in 2005, and the Ph.D. degree in information and communication engineering from Ningbo University, Ningbo, China, in 2020.
He is currently a Professor with the School of Mathematics and Statistics, Ningbo University. His research interests include underwater image processing and deep learning.



Ting Luo (Member, IEEE) received the B.Sc. degree in computer science from Ningbo University, Ningbo, China, in 2003, the M.Sc. degree in business information technology from Middlesex University, London, U.K., in 2004, and the Ph.D. degree in information and communication engineering from Ningbo University, in 2016.
He is currently with Ningbo University as a Professor. His research interests include multimedia security, image processing, data hiding, and pattern recognition.



Gangyi Jiang (Senior Member, IEEE) received the M.S. degree in electronic engineering from Hangzhou University, Hangzhou, China, in 1992, and the Ph.D. degree in electronic engineering from Ajou University, Suwon, South Korea, in 2000.
He is currently a Professor with the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. His research interests mainly include digital video compression and communications, multiview video coding, and image processing.



Xuebo Zhang (Senior Member, IEEE) received the B.Eng. degree in electronic engineering and the Ph.D. degree in underwater acoustic engineering from Naval University of Engineering, Wuhan, China, in 2009 and 2014, respectively.
He was with Underwater Antagonizing Laboratory, China, from 2014 to 2020, followed by a position with Northwest Normal University, China.
Dr. Zhang holds the role of Technology Committee Co-Chair for the IEEE Oceanic Engineering Society. He has served in various capacities, such as Session Chair, Regional Chair, Track Chair, Publicity Chair, Invited Speaker, and Special Session Chair, for several conferences. His contributions have earned him recognition, including the titles of “Outstanding Reviewer” and “Outstanding Editorial Board Member” for a prestigious journal, along with the Science and Technology Progress First Prize and other honors.



Mei Yu received the B.S. and M.S. degrees in computer science from the Hangzhou Institute of Electronics Engineering, Hangzhou, China, in 1990 and 1993, respectively, and the Ph.D. degree in computer science from Ajou University, Suwon, South Korea, in 2000.
She is currently a Professor with the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. Her research interests mainly include image/video coding and visual perception.