Adaptive Double-Branch Fusion Conditional Diffusion Model for Underwater Image Restoration

Yingbo Wang [iD] , *Member*, *IEEE*, Kun He [iD] , Qiang Qu [iD] , Xiaogang Du [iD] , Tongfei Liu [iD] , *Member*, *IEEE*, Tao Lei [iD] , *Senior Member*, *IEEE*, and Asoke K. Nandi [iD] , *Life Fellow*, *IEEE*

Abstract—Underwater images suffer from light absorption and scattering, impairs their visibility and applications. Existing underwater image restoration (UIR) methods based on generative models struggle are difficult to adapt to the complex and dynamic underwater environments characterized by illumination interference, low-light conditions, and non-uniform turbidity. To address these issues, we propose Water-CDM, a novel Adaptive Double-Branch Fusion Conditional Diffusion Model for underwater image restoration. Specifically, an adaptive double-branch fusion conditional diffusion model is presented utilizing a U-shaped full-attention network and Guided Multi-Scale Retinex with Brightness Correction (GMSRBC) to restore the challenging regions within underwater images. More precisely, to correct color casts and enhance the sharpness of underwater images, a U-shaped full-attention network incorporating Attention Blocks is designed for noise estimation during the reverse process of the conditional diffusion model. Concurrently, to mitigate overexposure during the enhancement of low-light underwater images under illumination interference, the GMSRBC method, featuring an Adaptive Brightness Correction Module, is proposed to efficiently adjust the brightness of underwater images. Experimental results demonstrate that the proposed Water-CDM significantly improves the quality of underwater images in challenging scenarios. Encouragingly, our proposed Water-CDM yields superior restoration outcomes compared to current state-of-the-art methods on three challenging publicly available datasets. Our codes will be released at: https://github.com/HKandWJJ/Water-CDM.

Index Terms—Underwater image restoration, conditional diffusion model, adaptive double-branch fusion, brightness enhancement.

I. INTRODUCTION

Yingbo Wang, Kun He, Qiang Qu, Xiaogang Du, Tongfei Liu and Tao Lei are with the Shaanxi Joint Laboratory of Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, China, and also with the School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, China (e-mail: wangyingbo@sust.edu.cn; 231612073@sust.edu.cn; 231612147@sust.edu.cn; duxiaogang@sust.edu.cn; liutongfei_home@hotmail.com; leitao@sust.edu.cn).

Asoke K. Nandi is with the Department of Electronic and Electrical Engineering, Brunel University of London, Uxbridge, Middlesex, UB8 3PH, United Kingdom, and also with the School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China. (E-mail: asoke.nandi@brunel.ac.uk)

UNDERWATER image plays a vital role in underwater biological detection, marine environmental protection, and other fields [1], [2], [3], [4]. However, due to the scattering and absorption of light in water during transmission, the quality of underwater images deteriorates severely [5], affecting tasks such as underwater object detection [6], recognition [7] and tracking [8]. Therefore, restoring underwater images to high quality has become a crucial issue [9], [10], [11]. Currently, UIR methods can be divided into those based on traditional mathematical modeling [12], [13] and those based on deep learning [14], [15]. Traditional UIR methods rely on physical models or digital image processing technologies. The core of those methods involves using prior knowledge or reasonable assumptions to reverse the degradation process of the image, thereby restoring clear underwater images, such as histogram equalization [16], white balance [17], and wavelet transform [18]. In contrast, UIR methods based on deep learning leverage their powerful nonlinear mapping capabilities to automatically learn the nonlinear mapping relationship between degraded images and clear images. They adaptively learn the image degradation laws of different underwater environments and improve the image quality [19]. These methods mainly include CNN-based methods [20], [21], Transformer-based methods [22], and Generative model-based methods [23], [24], [25], [26].

The UIR method based on CNNs mainly relies on their powerful ability in image feature extraction and mapping [27]. These CNNs extract high-level features from images through multi-layer convolution operations, thereby obtaining high-quality restored images that show excellent performance [28], [29]. The core of the Transformer-based UIR method [25] is the self-attention mechanism. This mechanism can dynamically calculate the correlation between any two elements in the input sequence, capture long-range dependencies, and assist in restoring degraded underwater images [30].

The UIR methods based on generative models include various types. Generative adversarial networks (GANs) [31], [32] utilize adversarial training between the generator and the discriminator, enabling the generator to generate clear images that are difficult to distinguish from real underwater images. Variational autoencoder (VAE)-based methods [24], [33], [34] leverage the advantages of VAEs in generating models and latent spatial representations to achieve restoration of degraded underwater images by learning and reconstructing the latent

features of underwater images. The Flow-based Models [25] achieve a mapping from a simple prior distribution to a complex data distribution through a series of reversible transformations, thereby restoring high-quality underwater images. With the development of the denoising diffusion probabilistic models (DDPMs) in the field of image generation [35], [36], DDPMs have attracted widespread attention from researchers due to their unique advantages. DDPMs gradually add noise to the data and train the model to remove the noise, enabling the model to learn the distribution characteristics of the sample data. This progressive training strategy not only improves the generation capacity of the model but also enhances its generalization ability. Although generative models have achieved remarkable results in underwater image restoration, they also face many challenges. For example, the training process of GANs is relatively complex and unstable, prone to problems such as mode collapse and gradient vanishing [37]. VAEs may over-smooth the data distribution when optimizing the reconstruction loss, resulting in blurred generated samples. Flow-based models have high computational complexity in high-dimensional space. DDPMs have certain limitations due to the lack of specific guidance conditions, which may cause the generated images to not match the expected images. Therefore, to improve the quality of underwater images, it is necessary to use effective conditions to guide the DDPM restoration process for underwater images.

In this paper, we propose a novel generative UIR method based on an adaptive double-branch fusion conditional diffusion model, named Water-CDM, designed to simultaneously solve the aforementioned issues. Water-CDM integrates degraded images as constraints within the forward diffusion process and incorporates a double-branch structure for the image restoration phase. The primary-branch focuses on the reverse sampling process, where we have developed the Attention Block of the Parallel Channel Spatial Attention Module (PCSAM). Additionally, we have constructed a U-shaped full-attention noise estimation network based on the Attention Block to further extract high-level image features. The secondary-branch features the Guided Multi-Scale Retinex with Brightness Correction (GMSRBC) method, which is designed to adaptively adjust image brightness. Finally, adaptive double-branch fusion is performed to achieve high-quality restoration of underwater degraded images. We conduct experiments and demonstrate the performance of our Water-CDM on three benchmark datasets compared to state-of-the-art (SOTA) UIR methods. Furthermore, we provide extensive computer vision tasks and ablation studies to show the effectiveness and potential value of the network structure designs.

The key contributions of this paper are as follows:

1) Unlike existing methods that aim to restore crisp images from degraded ones but often fall short in challenging scenarios, we propose a novel adaptive double-branch fusion conditional diffusion model named Water-CDM. This model is designed to generate clear images from the Gaussian noise images under the constraints posed by underwater degraded images. The proposed Water-CDM not only restores clear underwater images in challenging environments but also achieves superior results compared to SOTA methods.

2) Within the Water-CDM framework, we propose a U-shaped full-attention network based on Attention Blocks that leverages channel-space attention and gated deep-convolution operations. This network is specially designed to enhance the clarity and color casts of underwater images during the reverse process of the conditional diffusion model.

3) In the Water-CDM, a novel GMSRBC method is introduced to effectively suppress overexposure during the brightness enhancement of low-light images under illumination interference. Additionally, the brightness adaptive double-branch fusion strategy implemented in the HSL color space contributes to enhancing the overall brightness of underwater images.

II. RELATED WORKS

A. Underwater Image Restoration with Generative Models

In recent years, UIR methods based on generative models have been proven to be superior to others ; however, there are still numerous issues that need addressing. Islam *et al*. [37] present a simple yet efficient conditional GAN-based model for real-time underwater image enhancement (FUnIE-GAN). They design the generator network following the principles of U-net, and employ a Markovian Patch-GAN architecture for the discriminator. The adversarial training of the generator and discriminator aims to learn the mapping between the degraded images and clear images, thereby increasing image quality. However, the FUnIE-GAN is not very effective for enhancing severely degraded and texture-less images. Liu *et al*. [38] propose a generic multi-level features extraction and fusion network (MLFcGAN) under the framework of conditional GAN for underwater image color correction. Fu *et al*. [33] introduce a network based on probabilistic adaptive instance normalization, which combines a conditional VAE with AdaIn to construct an enhancement distribution (PUIE-net). This method improves image quality while also offering flexibility and robustness. Zhang *et al*. [25] propose a detection-driven heuristic normalizing flow underwater image enhancement, dubbed Water-flow, which integrates the UIR process into the normalizing flow to establish the invertible mappings between degraded and clear image through bilateral constraints.

With the continuous development of diffusion models, more researchers are actively exploring the potential application of image generation using these diffusion models. Sohl-Dickstein *et al*. [39] built a generative Markov chain that converts a simple known distribution through a target distribution using a diffusion process. Although the diffusion model was theoretically innovative, it did not receive widespread attention when first proposed due to deficiencies in image generation quality and sampling speed. It wasn't until 2020-year, Ho *et al*. [40] introduced the denoising diffusion probabilistic models (DDPM), which significantly improved the performance of the diffusion model and made significant progress in the field of image generation, surpassing previous generative models such as GANs and VAEs.
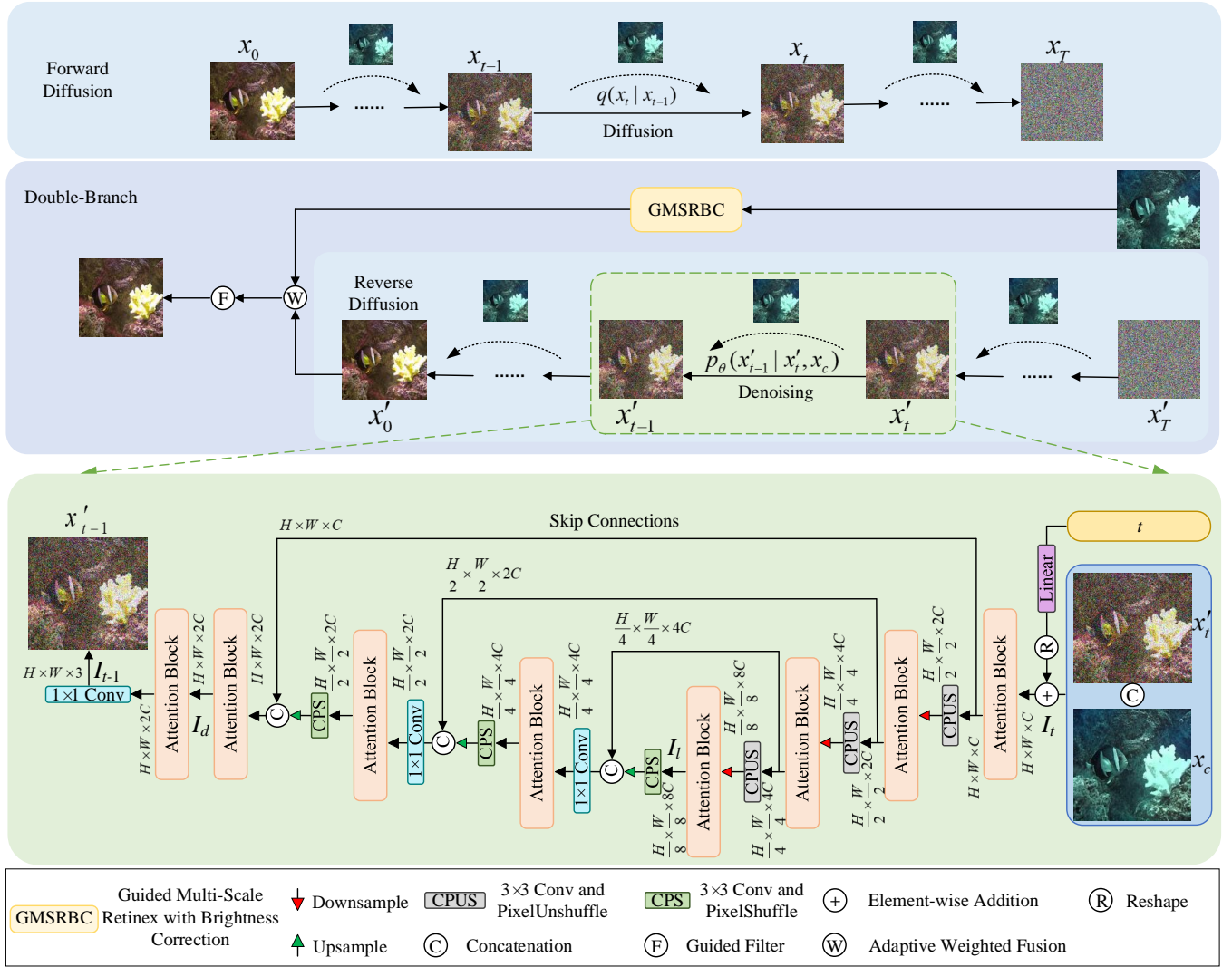
Fig. 1. Architecture of Water-CDM for underwater image restoration. The Water-CDM begins with forward diffusion and then restores the images using adaptive double-branch fusion. The core modules of Water-CDM are: (a) a U-shaped network based on Attention Blocks that performs the noise estimation. (b) The Guided Multi-Scale Retinex with Brightness Correction (GMSRBC) method, which performs the brightness compensation, *i.e.*, to suppress the overexposure in brightness enhancement.

B. Conditional Diffusion Model

The conditional diffusion model builds upon the diffusion model by introducing conditional information to guide the generation process, thereby achieving a more controllable and refined generation effect. Nichol *et al.* [41] proposed an improved DDPM that makes the variance learnable, allowing for the generation of variances more suitable for the data distribution. This approach significantly improves the log-likelihood of generated samples and enables the model to achieve near-optimal generation results with fewer sampling steps. Saharia *et al.* [42] introduced SR3 for image super-resolution via repeated refinement, and SR3 adapts DDPM to conditional image generation and performs super-resolution through a stochastic iterative denoising process. Ho *et al.* [43] proposed a cascaded diffusion model capable of generating high-fidelity images on class-conditional datasets. A cascaded diffusion model consists of a pipeline of multiple diffusion models that generate images of increasing resolution.

At present, conditional diffusion models are rapidly advancing in the field of underwater image restoration. Lu *et al.* [44] proposed a DDPM-based underwater image enhancement method called UW-DDPM, which employs a dual U-Net network for the image enhancement task and significantly improves the image quality. Lu *et al.* [45] introduced SU-DDPM, a method of real-time underwater image enhancement method based on DDPM. This method accelerates the inference process by modifying the initial sampling distribution and reducing the number of iterations during the reverse sampling process. Guan *et al.* [46] developed the DiffWater method, which is based on a conditional DDPM to enhance the quality of underwater images. DiffWater utilizes color channels as conditional guidance to improve the color appearance of image enhancement. Consequently, inspired by the significant impact of conditional diffusion models in image processing, this paper proposes a novel generative UIR method, dubbed Water-CDM—adaptive double-branch fusion conditional diffusion
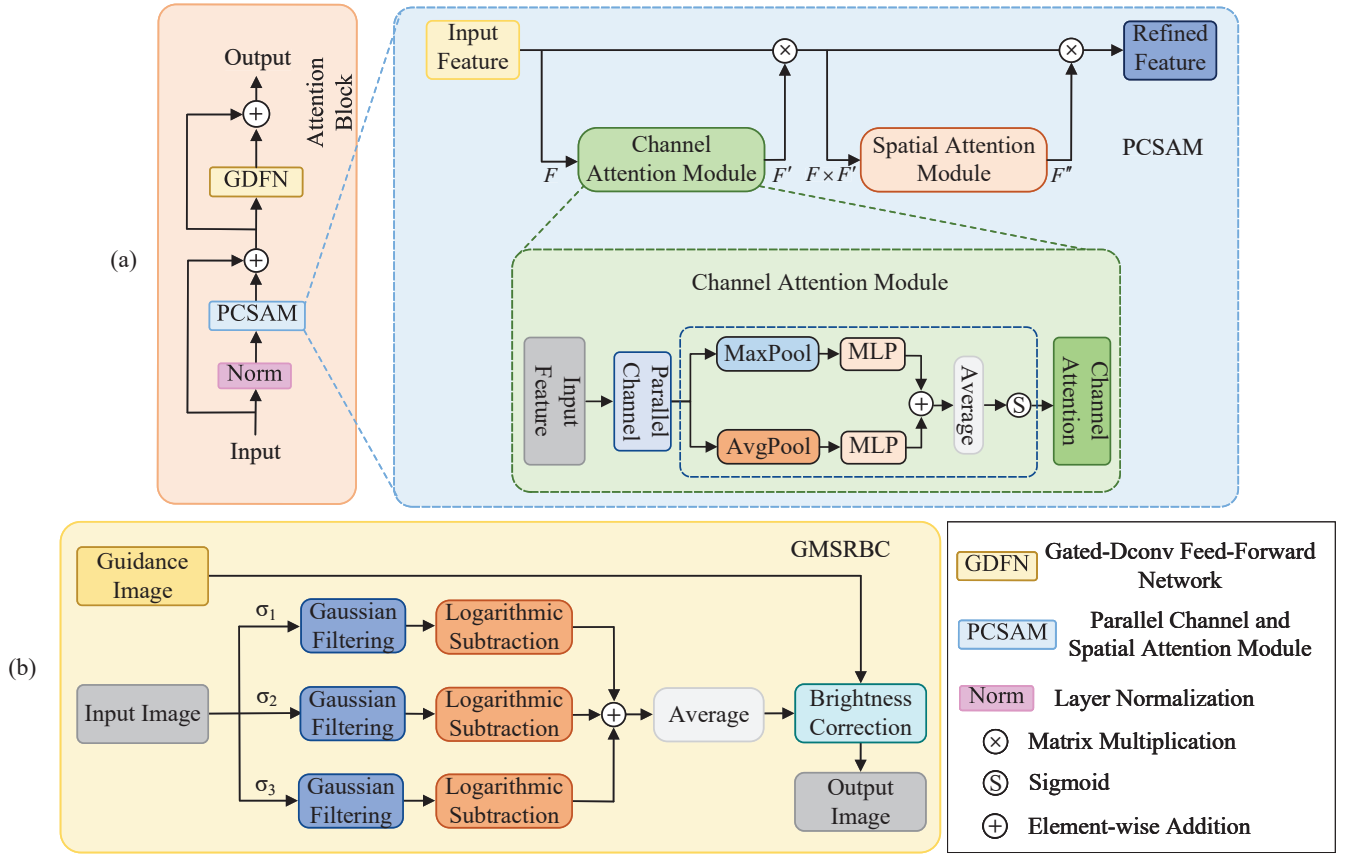
Fig. 2. Architecture of (a) the Attention Block with PCSAM and (b) the workflow of the GMSRBC.

model for underwater image restoration. The framework features a double-branch restoration architecture that adaptively fuses brightness information from both branches, significantly improving image quality in various underwater environments.

III. METHOD

This paper introduces the Water-CDM method, which aims to enhance the visibility of underwater images. The network structure of Water-CDM is detailed in section III-A. In the Water-CDM, the forward diffusion process is illustrated in section III-B. An adaptive double-branch fusion structure has been designed for image restoration. The primary-branch consists of a U-shaped full-attention noise estimation network based on Attention Blocks, which enhances the model's feature representation capabilities. The reverse denoising model incorporating these Attention Blocks is described in section III-C. The secondary-branch employs the GMSRBC method to mitigate the overexposure that occurs when brightness is enhanced using the MSR method. Finally, the overall brightness of the image through a double-branch adaptive brightness fusion mechanism, as presented in section III-D.

A. Network Structure of the Water-CDM

Water-CDM (Fig. 1) is a conditional diffusion model $p(x_0' \mid x_T', x_c)$, where x_c represents an underwater degraded image, $p(x_T') \sim N(0, I)$, and x_0' denotes the restored image. The forward diffusion process $q(x_t | x_{t-1})$ gradually adds Gaussian

noise following a normal distribution to reference image x_0 through the Markov process q to obtain x_T. The image restoration process follows the structure of adaptive double-branch fusion. The primary-branch involves the reverse inference process of the conditional diffusion model, which estimates the noise in images through a U-shaped network based on Attention Blocks. The secondary-branch employs a GMSRBC method to enhance image brightness. These two branches are then adaptively fused to achieve the final restored image.

B. Forward Diffusion

The theoretical basis of Water-CDM is the conditional diffusion model. In the forward diffusion process, a forward Markov process q is defined, which adds diagonal Gaussian noise to the clear image x_{t-1} at $t-1$. After one iteration, the noisy image x_t at t can be obtained, that is:

$$q(x_t | x_{t-1}) = N(x_t; \sqrt{\alpha_t} x_{t-1}, (1 - \alpha_t)I), \quad (1)$$

where $\alpha_t \in (0, 1)$ is a hyperparameter that determines the variance of the noise added at each iteration, with $\alpha_t = 1 - \beta_t, \beta_t \in (0, 1)$ is a variance. x_t is the diffusion result at time t, $t \in \{1, 2, ..., T\}$, and T is the total diffusion steps. x_t can be calculated by adding Gaussian noise using the Markovian property when the reference image x_0 is known, that is:

$$q(x_t | x_0) = N(x_t; \sqrt{\overline{\alpha_t}} x_0, (1 - \overline{\alpha_t})I), \quad (2)$$

where $\overline{\alpha_t} = \prod_{s=1}^{t} \alpha_s$. Given a sufficiently large T, the latent x_t is nearly an isotropic Gaussian distribution. The forward

diffusion process and constraint condition x_c are important components in training a noise estimation network. They also participate in the training of the U-shaped denoising network and the noise generation process.

C. Reverse Denoising Model with Attention Blocks

Inspired by [47], a U-shaped network based on Attention Blocks is designed as the noise estimation network during the reverse inference process. Given the conditional image x_c, time step t and sampling Gaussian noise image $x'_T \sim N(0, I)$ as the inputs to the reverse diffusion, Water-CDM first concatenates the x_c, x'_T and the reshaped t to obtain the shallow input $I_t \in \mathbb{R}^{H \times W \times C}$, where $H \times W$ denotes the spatial dimension and C is the number of channels. Then, this shallow input I_t passes through a 4-level symmetric encoder-decoder and is transformed into deep features $I_d \in \mathbb{R}^{H \times W \times 2C}$. Starting from the high-resolution input I_t, the encoder hierarchically reduces spatial size (by half) and expands channel capacity (by doubling) to obtain low-resolution features $I_l \in \mathbb{R}^{H/8 \times W/8 \times 8C}$. The decoder takes I_l as input and progressively recovers the high-resolution features to achieve global feature representations. For feature down-sampling and up-sampling, CPUS and CPS operations are applied respectively. CPUS consists of a 3×3 convolution and pixel-unshuffle, while CPS consists of a 3×3 convolution and pixel-shuffle [48]. To prevent feature loss during down-sampling, encoder features are concatenated with the decoder features via skip connection [49], and 1×1 convolutions are used to reduce channels in the decoder. After the up-sampling operation to obtain deep features I_d, an Attention Block is employed to further enrich the detail texture of the feature map at high spatial resolution. Finally, the enriched features are convolved to obtain the output image $I_{t-1} \in \mathbb{R}^{H \times W \times 3}$, and x'_{t-1} is extracted by channel. The data distribution of the reverse inference process can be expressed as $p(x'_0 | x'_T, x_c)$. At each iteration step t, the image x'_{t-1} is generated by gradually denoising the current noisy image x'_t with the underwater degraded image x_c as a constraint, they can be described as:

$$p_\theta(x'_{t-1} | x'_t, x_c) = N(x'_{t-1}; \mu_\theta(x'_t, t, x_c), \sigma_t^2 I), \qquad (3)$$

where σ_t is the time-dependent standard deviation, and μ_θ is the mean function [40] learned from the model parameters θ, which can be expressed as:

$$\mu_\theta(x'_t, t, x_c) = \frac{1}{\sqrt{\alpha_t}}(x'_t - \frac{\beta_t}{\sqrt{1 - \overline{\alpha_t}}} \varepsilon_\theta(x'_t, t, x_c)), \qquad (4)$$

where $\varepsilon_\theta(x'_t, t, x_c)$ is the noise estimation of the U-shaped network, α_t and β_t are parameters in the forward diffusion process. The U-shaped network parameters θ are optimized using the loss function, which can be computed as:

$$L(\theta) = \mathbb{E}_{x'_t, x_c, \varepsilon, t}[||\varepsilon - \varepsilon_\theta(x'_t, t, x_c)||^2], \qquad (5)$$

where ε represents the noise sampled from the standard normal distribution. Finally, the x'_0 obtained through the reverse inference process, as shown in Fig. 3(d), serves as the primary-branch restoration image and can be expressed as I_{CDM}.

The Attention Block in the reverse inference process is illustrated in Fig. 2(a), with its kernel component being the PCSAM. The PCSAM divides the number of channels into multiple parallel channels to learn independent channel attention matrices, thereby enhancing the feature representation and generalization performance of the model. The input-output characteristic relationship of the PCSAM module can be expressed as:

$$\begin{cases} F' = M_c(F) \otimes F, \\ F'' = M_s(F') \otimes F', \end{cases} \qquad (6)$$

where $F \in \mathbb{R}^{H \times W \times C}$ is the input feature of PCSAM, F' is the output feature after applying channel attention, and F'' is the final output feature of PCSAM. $M_c \in \mathbb{R}^{1 \times 1 \times C}$ represents the channel attention matrix, and $M_s \in \mathbb{R}^{H \times W \times 1}$ represents the spatial attention matrix. M_C and M_S can be defined as follows:

$$\begin{cases} M_c(F) = Parallel_i \begin{cases} \sigma[\frac{1}{2}[MLP(MaxPool(F)) \\ + MLP(AvgPool(F))]] \end{cases}, \\ M_s(F') = \sigma[f^{3 \times 3}[MaxPool(F'); AvgPool(F')]], \end{cases} \qquad (7)$$

where σ represents the Sigmoid activation function, MLP denotes a parameter-sharing multilayer perception, i indicates the parallel attention index and is less than the number of channels in F, and $Parallel\{\cdot\}$ refers to the parallel attention calculation performed after dividing the channels by i and subsequently merging the channel attention through a concatenate operation, $f^{3 \times 3}$ represents a convolution with a filter kernel size of 3×3.
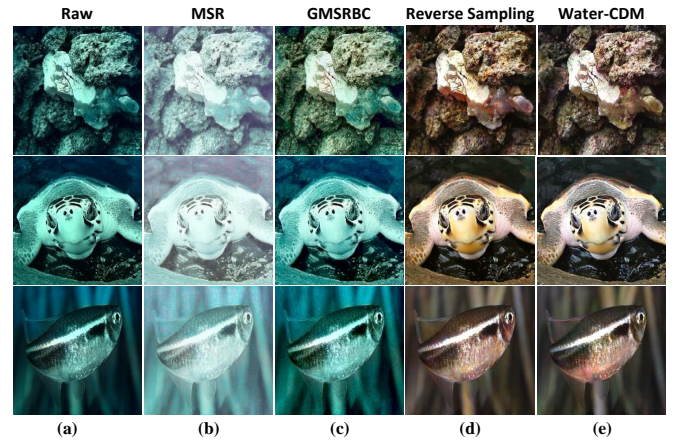

Fig. 3. Comparison of image restoration effects before and after brightness enhancement.

D. Brightness Compensation and Adaptive Fusion

Due to insufficient illumination, underwater images often exhibit a low-light phenomenon. The Multi-Scale Retinex (MSR) algorithm [50] is commonly used to enhance the brightness of such images. However, this method may cause overexposure in some areas [51], as shown in Fig. 3(b). We propose a Guided Multi-Scale Retinex with Brightness Correction (GMSRBC) method by integrating an adaptive brightness correction module into the MSR framework to address this limitation. As illustrated in Fig. 2(b), GMSRBC employs an underwater degraded image as a guidance image,
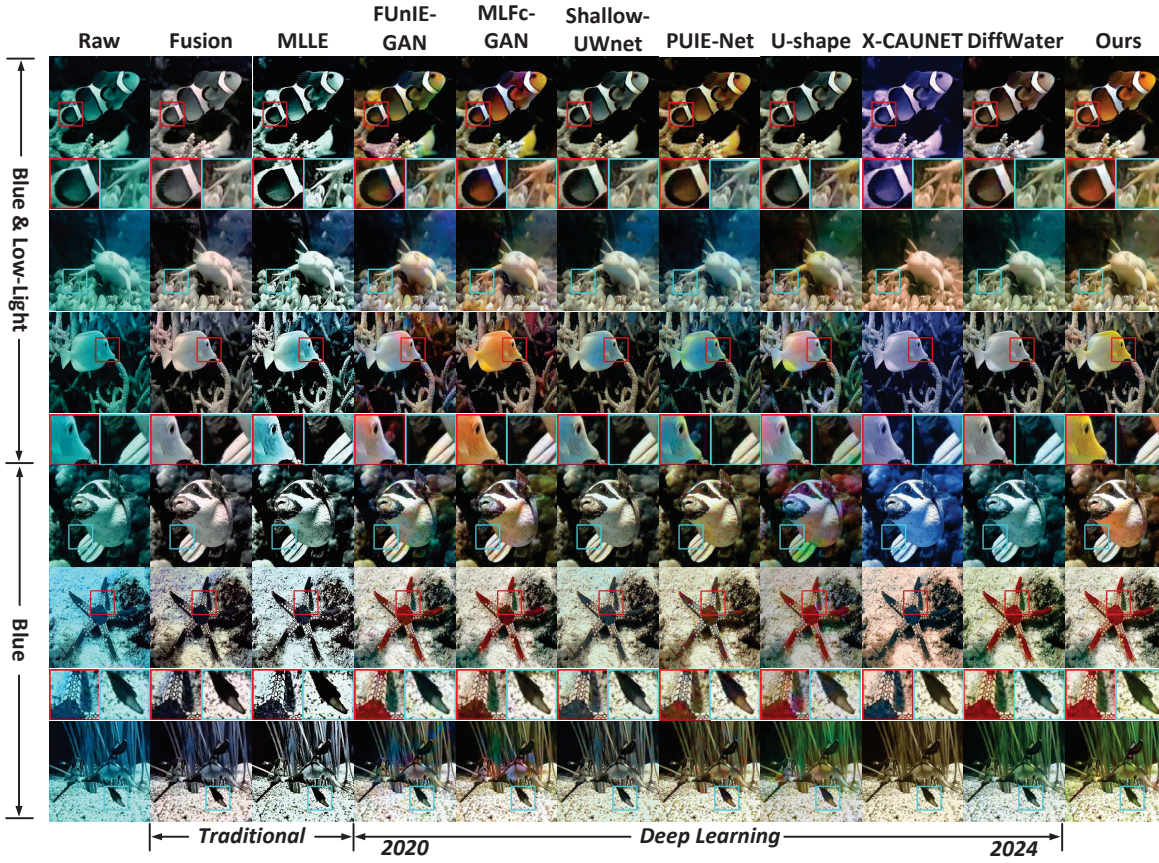
Fig. 4. Comparison of visual results on EUVP-Dark [37] Dataset. Raw underwater Images showed blueish and low light. Zoom in for the best view. Image insets are shown with solid boxes of the corresponding color in between for the zoomed 2 times in the images.

suppressing overexposure by utilizing the maximum brightness of the guidance image. This results in a restored image with overall brightness improvements, as demonstrated in Fig. 3(c). The adaptive brightness correction module can be expressed as:

$$
\begin{cases}
L_{MSR}(x,y) = max(L_{DI}(x,y)), \\
\qquad if L_{MSR}(x,y) \geq max(L_{DI}(x,y)); \\
L_{MSR}(x,y) = min(L_{DI}(x,y)), \\
\qquad if L_{MSR}(x,y) \leq min(L_{DI}(x,y)); \\
L_{MSR}(x,y) = L_{MSR}(x,y), \quad else.
\end{cases}
\tag{8}
$$

where L_{MSR} and L_{DI} represent the brightness information of I_{MSR} and I_{DI}, respectively, in the HSL color space. I_{MSR} refers to the underwater degraded image enhanced by the MSR algorithm, while I_{DI} is the original underwater degraded image. After processing through the adaptive brightness correction module, a brightness-compensated underwater degraded image is obtained, denoted as I_{GMSRBC}.

The adaptive double-branch fusion brightness is achieved by fusing the brightness L_{CDM} of the restored image I_{CDM} from the primary-branch and the brightness L_{GMSRBC} of the restored image I_{GMSRBC} from the secondary-branch, which can be expressed as:

$$
L' = \omega_{GMSRBC} L_{GMSRBC} + \omega_{CDM} L_{CDM}. \tag{9}
$$

The obtained brightness L' is combined with the hue and saturation of the primary-branch restored image I_{CDM} to obtain the final restored underwater image, shown in Fig. 3(d). The adaptive fusion weights are given by:

$$
\begin{cases}
\omega_{GMSRBC} = \dfrac{L'_{GMSRBC}}{L'_{GMSRBC} + L'_{CDM} + \delta}, \\
\omega_{CDM} = \dfrac{L_{CDM}}{L'_{GMSRBC} + L'_{CDM} + \delta},
\end{cases}
\tag{10}
$$

where δ is a hyperparameter set to $1e-5$, ω_{GMSRBC} and ω_{CDM} are the adaptively determined fusion weights, L'_{GMSRBC} and L'_{CDM} denote the brightness of L_{CDM} and L_{GMSRBC} after guided filtering, respectively.

IV. EXPERIMENTS

A. Experimental Settings

Implementation Details. Our method was implemented using PyTorch and trained on NVIDIA RTX 3090. We randomly crop the images to a size of 256×256, using these cropped patches as the training dataset for UIR. Initially, We trained the U-shaped network based on Attention Blocks for $1e^6$ iterations. The initial learning rate was set to $1e^{-2}$, decreasing linearly to an end learning rate of $1e^{-6}$. Adam was employed as our optimizer with a batch size of 8, and the total diffusion step size was set to 2000. Subsequently, we

TABLE I
QUANTITATIVE COMPARISONS OF DIFFERENT METHODS ON THREE DATASETS FOR UIR TASK IN TERMS OF PSNR(↑), SSIM(↑), UIQM(↑), AND PIQE(↓). ↑ DENOTES THOSE LARGE VALUES MEAN BETTER RESULTS. ↓ DENOTES THOSE SMALL VALUES MEAN BETTER RESULTS. THE BEST AND SECOND QUALITY SCORES OF THE EVALUATED METHODS ARE MARKED IN **BOLD** AND <u>UNDERLINE</u>.

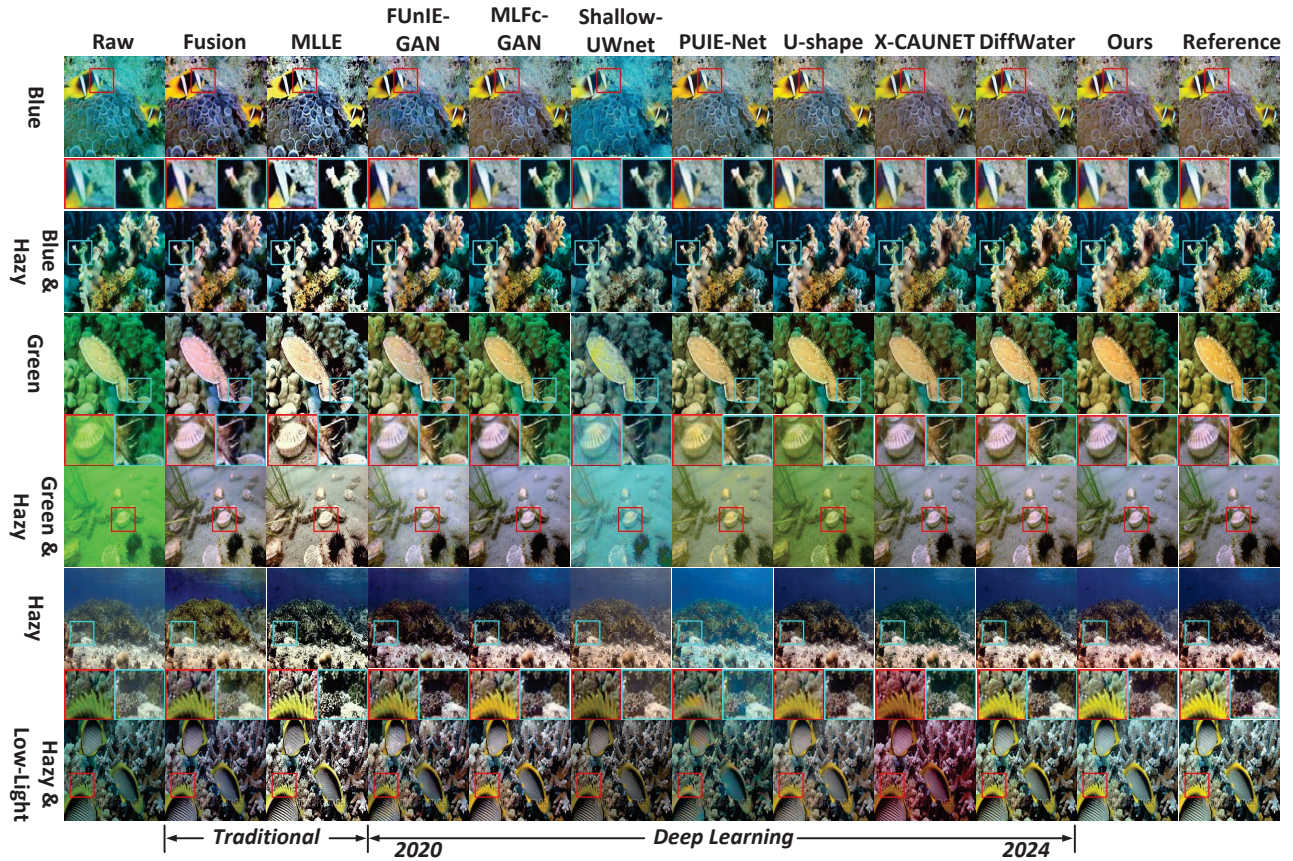| Method	EUVP-Dark		LSUI				UIEB			
	UIQM↑	PIQE↓	PSNR↑	SSIM↑	UIQM↑	PIQE↓	PSNR↑	SSIM↑	UIQM↑	PIQE↓
Fusion[52]	3.237	0.387	18.755	0.906	3.065	0.357	18.009	0.928	3.288	0.855
MLLE[53]	3.068	0.376	16.848	0.840	3.022	0.315	16.923	0.871	3.254	1.041
FUnIE-GAN[37]	3.076	0.414	20.862	0.927	2.688	0.263	18.875	0.909	2.885	0.765
MLFcGAN[38]	3.213	0.415	19.197	0.777	<u>3.203</u>	0.358	21.346	0.948	3.272	0.823
Shallow-UWnet[29]	<u>3.288</u>	0.389	23.782	0.827	2.923	0.429	18.421	0.878	**3.659**	0.845
PUIE-Net[33]	3.067	0.398	17.104	0.888	3.137	0.348	<u>22.956</u>	<u>0.962</u>	3.272	<u>0.705</u>
U-shape[30]	3.006	0.372	<u>24.945</u>	0.962	3.130	0.366	19.037	0.898	3.284	1.067
X-CAUNET[22]	3.046	0.400	21.022	0.932	2.338	0.414	17.734	0.913	3.118	0.858
DiffWater[46]	3.187	**0.306**	24.746	<u>0.964</u>	2.767	**0.124**	21.232	0.952	3.092	0.712
Ours	**3.354**	<u>0.366</u>	**28.008**	**0.978**	**3.281**	<u>0.214</u>	**23.231**	**0.980**	<u>3.417</u>	**0.696**

Fig. 5. Comparison of visual results on LSUI [30] Dataset. Raw underwater Images showed color cast, haziness, and low light. Zoom in for the best view.

applied the GMSRBC method to enhance the brightness of underwater images. Finally, we performed an adaptive fusion of double-branch brightness to obtain the restored images.

Training and Testing Datasets. Following previous works [44], [45], [46], we used three standard benchmark real underwater image datasets: EUVP-Dark [37], LSUI [30], and UIEB [28]. These datasets consider adverse underwater conditions such as low light, haziness, illumination, and color cast. The EUVP-Dark dataset contains low-light underwater images and comprises 5550 paired training images and 570 unpaired test images. The LSUI dataset covers a wide range of underwater scenes, including various lighting conditions, water

types, and target categories, and it consists of 4504 paired training images and 500 paired test images. The UIEB dataset involves underwater images disturbed by illumination issues and consists of 800 paired training images and 90 paired test images.

Comparison Methods and Evaluation Metrics. We compared our Water-CDM method with two traditional methods, Fusion [52] and MLLE [53], as well as seven deep learning-based methods: FUnIE-GAN [37], MLFcGAN [38], PUIE-Net [33], Shallow-UWnet [29], X-CAUNET [22], U-shape [30] and DiffWater [46]. The performance of Water-CDM on paired datasets was quantified using full-reference image quality as-
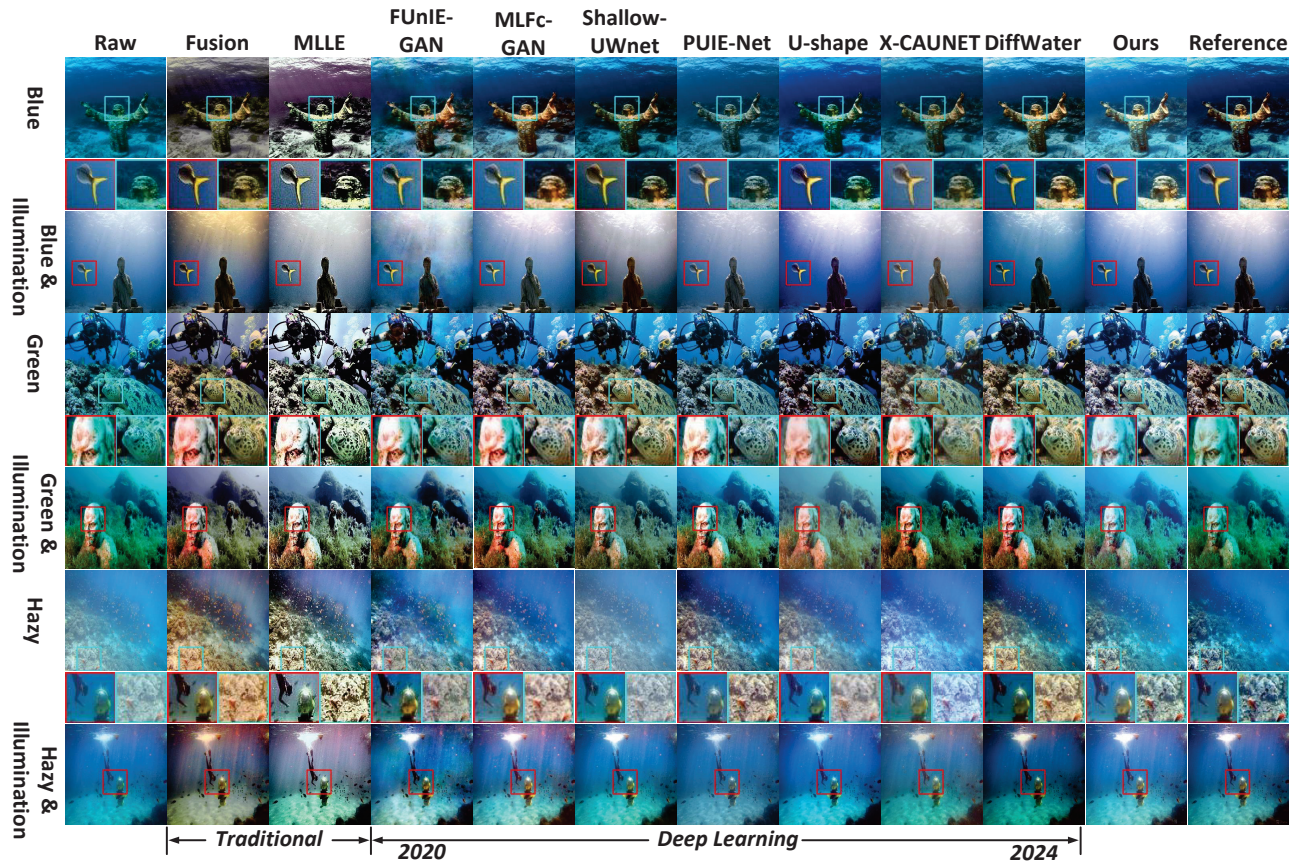
Fig. 6. Comparison of visual results on UIEB [28] Dataset. Raw underwater Images showed color cast, haziness, and illumination interference. Zoom in for the best view.

sessments, especially PSNR [54] and SSIM [55]. For unpaired datasets, we evaluated the restored images using UIQM [56] and PIQE [57]. PSNR and SSIM are the most commonly used metrics to indicate the similarity between restored and reference images. UIQM evaluates images based on color, sharpness, and contrast, while PIQE measures blockiness and noise.

B. Comparison of Experimental Results with State-of-the-Art Methods

In this section, we evaluate the effectiveness of the proposed method through qualitative and quantitative comparisons of benchmark datasets. Nine representative SOTA methods from recent years are compared to assess performance. Fig. 4 shows the qualitative results obtained by each method on the unpaired test images from the EUVP Dark dataset. For underwater images with low-light conditions and bluish tint, the Water-CDM method not only compensates for the red channel information and improves the overall brightness of the image but also preserves the structure and texture details of the underwater image. The quantitative metrics are shown in Table. I, Water-CDM achieved the best UIQM score and the second-best PIQE score. In particular, the Water-CDM method improved the UIQM score by 2.01% compared to the Shallow-UWnet method.

Fig. 5 presents the qualitative results of various methods on the LSUI paired test images. The Water-CDM method exhibits excellent color compensation capabilities while significantly preserving image details for underwater images affected by color cast and haziness. Table. I shows the quantitative results of each method, among which Water-CDM achieved the best scores in metrics PSNR, SSIM, and UIQM, demonstrating its advantages in improving image quality. Although Water-CDM achieved suboptimal scores on PIQE, compared to the methods that performed best on other metrics, the PSNR score of the proposed method increased by 12.28%, the SSIM score increased by 1.45%, and the UIQM score increased by 2.62%. These improvements fully validate the efficiency and superiority of the Water-CDM method in dealing with underwater degraded images.

Fig. 6 shows the qualitative results of various methods on the UIEB paired test images, which include color cast, haziness, and illumination interference. The Water-CDM method not only effectively corrects image color cast but also successfully suppresses overexposure while improving image sharpness, ensuring a natural transition of image brightness and detail retention. Compared to SOTA methods, it demonstrates significant advantages. The quantitative metrics for this dataset are shown in Table. I, where Water-CDM obtained the best scores in PSNR, SSIM, and PIQE. Although Water-CDM achieved suboptimal scores on UIQM, compared to the

TABLE II
COMPARISON OF TRAINING TIME (MIN), RUNTIIME (S), PARAMETERS (M), AND FLOPS (GFLOPS) AMONG DIFFERENT METHODS.

Efficiency	Water-CDM	DiffWater	X-CAUNET	U-shape	PUIE-Net	Shallow-UWnet	MLFc-GAN	FUnIE-GAN	MLLE	Fusion
Training time	2.61	7.98	12.88	4.02	**0.30**	1.08	2.91	<u>0.77</u>	-	-
Runtime	0.86	48.52	0.07	0.04	<u>0.03</u>	**0.02**	<u>0.03</u>	**0.02**	0.30	1.23
Params	18.62	155.33	<u>6.62</u>	31.59	16.12	**2.19**	52.22	7.02	-	-
FLOPs	68.99	155.24	61.19	**2.98**	19.33	21.63	15.89	<u>10.24</u>	-	-

methods that performed best on other metrics, the PSNR score of the proposed method increased by 1.20%, the SSIM score increased by 1.87%, and the PIQE score decreased by 1.28%.

In summary, the Water-CDM method has demonstrated excellent performance across three benchmark underwater image datasets, showing significant advantages over other SOTA methods in terms of color compensation, detail preservation, brightness enhancement, and sharpness improvement. Statistical analysis was conducted on quantitative metrics for three underwater datasets, as shown in Fig. 7, the Water-CDM achieved optimal scores in PSNR, SSIM, and UIQM. Although this method achieves suboptimal results in PIQE scoring, as mentioned in [27], [58], PIQE may be biased towards certain features and therefore cannot accurately reflect the true visual quality of the restored image alone.
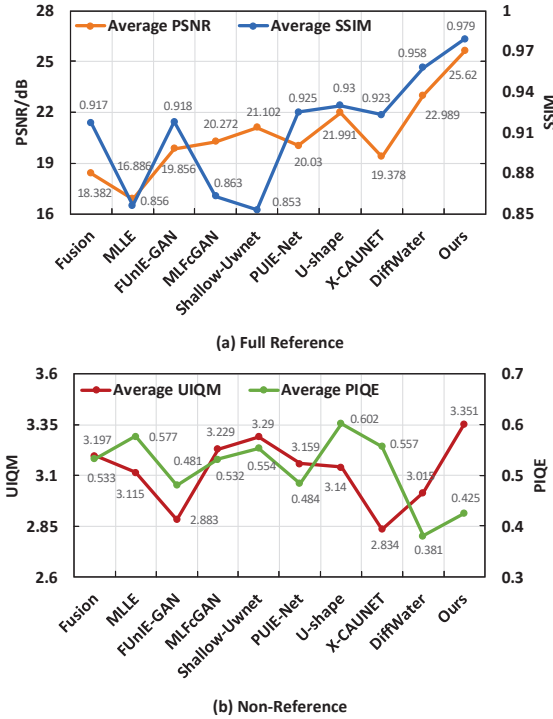

Fig. 7. Statistical analysis of quantitative metrics. (a) Comparison of average quantitative metrics of full-reference images on LSUI and UIEB, (b) comparison of average quantitative metrics of non-reference images on EUVP-dark, LSUI, and UIEB.

Additionally, we evaluated the proposed method on the EUVP-Dark dataset, comparing it with existing approaches regarding training time, runtime, the number of parameters, and FLOPS. The results are summarized in Table. II. While our model's training time per epoch is moderate, the image restora-

tion process requires 0.86s per image at 256×256 resolution owing to multiple reverse sampling iterations in the primary-branch. Despite the slightly increased model complexity resulting from multiple diffusion processes, our method maintains a moderate number of parameters that enable deployment on embedded devices. These characteristics demonstrate the model's suitability for embedded underwater image restoration systems.

C. Ablation Study

We conducted ablation tests on the UIEB dataset to verify the effectiveness of the key components of our designed model. In this section, we discuss the effectiveness of the double-branch structure, Attention Block, and GMSRBC module.
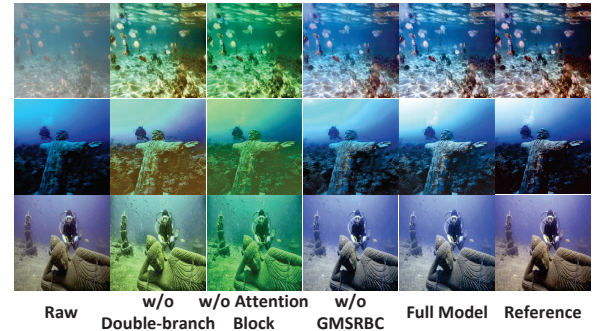

Fig. 8. Ablation results for visual comparison.

1) *Effectiveness of double-branch structure*: To demonstrate the contribution of the double-branch structure, we removed the double-branch component and used the conditional DDPM for the UIR task. We then compared the qualitative and quantitative evaluations of the two models (Fig. 8 and Table. III, w/o double-branch). When the conditional DDPM was applied directly to the UIR task without the double-branch structure, the images showed insufficient recovery and color distortion, and the evaluation metrics dropped significantly. This was due to the failure to extract global features of underwater images during the reverse inference process. The results indicate the effectiveness of the double-branch structure.

2) *Effectiveness of the Attention Block*: We removed the improved U-shaped network based on Attention Blocks and used a U-shaped network based on CNN as the primary-branch in the double-branch structure. The results are shown in Fig. 8 and Table. III (w/o Attention Block) for comparison. The restored images exhibited a greenish tint and poor quantitative metrics due to the failure to extract deeper features during

learning, resulting in model underfitting. The results demonstrate the effectiveness of the Attention Block.

3) *Effectiveness of the GMSRBC module*: We removed the GMSRBC module from the UIR task and used only the inverse inference process with the improved U-shaped network based on Attention Blocks to restore the image. The restoration results are shown in Fig. 8 and Table. III, w/o GMSRBC module, for comparison. The brightness of the low-light areas of the image was insufficient, and the sharpness of the restored image was poor, leading to a decrease in quantitative metrics. These results demonstrate that the GMSRBC module can improve the brightness while ensuring image quality does not deteriorate.

TABLE III
METRICS OF ABLATION EXPERIMENTS FOR COMPARISON ON THE UIEB DATASET.

Ablation	PSNR↑	SSIM↑	UIQM↑	PIQE↓
w/o double-branch	17.551	0.916	2.765	0.703
w/o Attention Block	16.103	0.918	3.032	0.703
w/o GMSRBC module	19.460	0.973	3.112	0.700
Ours	**23.231**	**0.980**	**3.417**	**0.696**

D. More Experiments and Analyses

To further demonstrate the contribution of Water-CDM to the UIR task, we performed application experiments as follows.

Feature matching: We used the SIFT [59] method for feature point matching of images before and after restoration, as shown in Fig. 9. By comparing the number of feature point matches before and after restoration, it can be observed that the number of feature point matches in the restored image significantly increases, and the distribution is more uniform and denser. This indicates that the proposed method significantly enriches the feature information of underwater images.
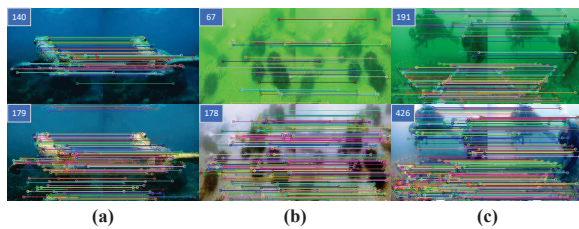

Fig. 9. Comparison of the results of SIFT feature matching on original underwater images and restored images. The number of feature matching is shown in the upper left corner.

Edge detection: We used the Canny [60] operator for edge extraction of images before and after restoration, shown in Fig. 10. It can be observed that the number of edge detections in the restored image has increased significantly. This indicates that the proposed method has effectively restored the edge details of underwater images, laying the foundation for advanced visual tasks in underwater scenes.

Object detection: We used the YOLOv7 [61] method for object detection on images before and after restoration, as

shown in Fig. 11. Thanks to the advantages of the proposed method in color correction, sharpness improvement, and brightness enhancement, the restored images achieved significant improvements in object detection accuracy and the number of detected targets. This further reduces the difficulty of developing object detection algorithms.
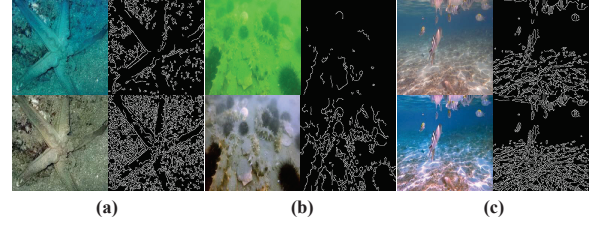

Fig. 10. Comparison of the result of Canny edge detection. We have chosen three images with different styles, which are: blueish, greenish, and haziness.
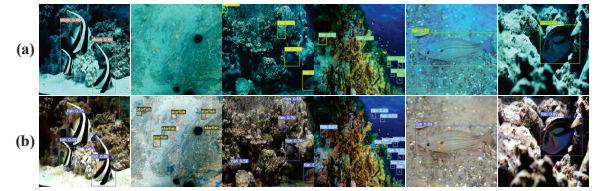

Fig. 11. Evaluation of object detection. (a) Degraded images and (b) restored images after the object detection with YOLOv7. It can be found that the proposed method achieves good performance in both accuracy and quantity.

Image segmentation: We used the MASK R-CNN [62] method for object detection and instance segmentation on images before and after restoration, as shown in Fig. 12. The underwater images restored by the proposed method exhibited higher accuracy in the instance segmentation task, with significantly improved target boundary integrity. These results further verify the effectiveness of the Water-CDM method in enhancing image quality and improving performance in computer vision tasks.

Fig. 12. Evaluation of object detection and instance segmentation. (a) Degraded images and (b) restored images after the instance segmentation with MASK R-CNN. It can be found that the proposed method is suitable for underwater instance segmentation.

The effectiveness of Water-CDM resulted in significant improvements in feature matching, edge detection, object detection, and image segmentation of the underwater images. These enhancements contributed to their application in high-level vision tasks. Overall, Water-CDM made a substantial contribution to the UIR task.

V. DISCUSSION AND FUTURE WORK

Water-CDM is a supervised learning-based UIR model, and its efficacy is significantly influenced by the quality and

quantity of reference images. As a pivotal component in the model's training process, the caliber of the reference images is crucial for determining the model's restoration effectiveness on underwater imagery. However, procuring high-quality reference images is often fraught with challenges, including the complexities of underwater environments, variable lighting conditions, and the limitations of imaging equipment. These factors can compromise the quality of reference images, thereby limiting the model's restoration capabilities. Therefore, to further improve the practicality and applicability of the Water-CDM method, it is imperative to address the urgent issue of reducing the model's reliance on high-quality labeled data. Future research endeavors will focus on exploring more efficient and robust techniques to diminish the model's dependence on reference images. This will bolster the model's autonomy and generation capability, paving the way for broader and more effective use in underwater image restoration.

VI. Conclusion

To address the problems of color cast, haziness, and dimness in underwater images within challenging underwater scenarios, we have developed Water-CDM, a high-quality UIR method based on an adaptive double-branch fusion conditional diffusion model. We first designed a novel conditional diffusion model with an adaptive double-branch fusion structure, which enhances the model's feature extraction capabilities and improves its robustness. Then, in the primary-branch of the UIR process, we introduced a U-shaped full-attention network based on Attention Blocks for noise estimation during the reverse diffusion process. This network effectively corrects color casts and enhances the clarity of underwater images. The secondary-branch uses the GMSRBC method to adaptively adjust the image brightness. Finally, the underwater image restoration result is obtained by brightness-adaptive fusion of two-branch restored images. This method not only significantly improves the brightness of low-light regions but also mitigates overexposure issues that can arise from brightness enhancement. Experimental results demonstrate that Water-CDM achieves an average UIQM gain of 1.854% over the previous SOTA method for UIR in three underwater datasets. Additionally, our method's performance on four computer vision tasks indicates that it provides valuable technical support for the advancement and application of sophisticated computer vision techniques in underwater imaging scenarios.

Acknowledgment

References

[1] Z. Zhao, Y. Liu, X. Sun, J. Liu, X. Yang, and C. Zhou, "Composited fishnet: Fish detection and species recognition from low-quality underwater videos," *IEEE Trans. Image Process.*, vol. 30, pp. 4719–4734, Apr. 2021.

[2] M. Jahanbakht, W. Xiang, L. Hanzo, and M. Rahimi Azghadi, "Internet of underwater things and big marine data analytics—a comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 904–956, Jan. 2021.

[3] Y. Yang, Y. Xiao, and T. Li, "A survey of autonomous underwater vehicle formation: Performance, formation control, and communication capability," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 815–841, Feb. 2021.

[4] H. Song, L. Chang, Z. Chen, and P. Ren, "Enhancement-registration-homogenization (ERH): A comprehensive underwater visual reconstruction paradigm," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 44, no. 10, pp. 6953–6967, Jul. 2022.

[5] 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, Sep. 2016.

[6] W.-H. Lin, J.-X. Zhong, S. Liu, T. Li, and G. Li, "Roimix: Proposal-fusion among multiple images for underwater object detection," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2020, pp. 2588–2592.

[7] P. Zhu, J. Isaacs, B. Fu, and S. Ferrari, "Deep learning feature extraction for target recognition and classification in underwater sonar images," in *Proc. IEEE Conf. Decis. Control(CDC)*, Dec. 2017, pp. 2724–2731.

[8] K. Panetta, L. Kezebou, V. Oludare, and S. Agaian, "Comprehensive underwater object tracking benchmark dataset and underwater image enhancement with GAN," *IEEE J. Oceanic Eng.*, vol. 47, no. 1, pp. 59–75, Jul. 2022.

[9] W. Zhang, L. Zhou, P. Zhuang, *et al.*, "Underwater image enhancement via weighted wavelet visual perception fusion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 4, pp. 2469–2483, Apr. 2024.

[10] Y. Kang, Q. Jiang, C. Li, W. Ren, H. Liu, and P. Wang, "A perception-aware decomposition and fusion framework for underwater image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 3, pp. 988–1002, Mar. 2023.

[11] J. Zhou, S. Wang, Z. Lin, Q. Jiang, and F. Sohel, "A pixel distribution remapping and multi-prior retinex variational model for underwater image enhancement," *IEEE Trans. Multimedia*, vol. 26, pp. 7838–7849, Mar. 2024.

[12] 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.

[13] P. Zhuang, J. Wu, F. Porikli, and C. Li, "Underwater image enhancement with hyper-laplacian reflectance priors," *IEEE Trans. Image Process.*, vol. 31, pp. 5442–5455, Aug. 2022.

[14] Y. Guo, H. Li, and P. Zhuang, "Underwater image enhancement using a multiscale dense generative adver-

sarial network," *IEEE J. Oceanic Eng.*, vol. 45, no. 3, pp. 862–870, Jul. 2020.

[15] C.-H. Yeh, C.-H. Lin, L.-W. Kang, *et al.*, "Lightweight deep neural network for joint learning of underwater object detection and color conversion," *IEEE T. Neur. Net. Lear.*, vol. 33, no. 11, pp. 6129–6143, Nov. 2022.

[16] V. Voronin, E. Semenishchev, S. Tokareva, A. Zelenskiy, and S. Agaian, "Underwater image enhancement algorithm based on logarithmic transform histogram matching with spatial equalization," in *Proc. IEEE Int. Conf. Signal Process. (ICSP)*, Aug. 2018, pp. 434–438.

[17] M. Kumar and A. K. Bhandari, "Contrast enhancement using novel white balancing parameter optimization for perceptually invisible images," *IEEE Trans. Image Process.*, vol. 29, pp. 7525–7536, Jun. 2020.

[18] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, "Multi-level Wavelet-CNN for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops(CVPRW)*, Jun. 2018, pp. 886–88 609.

[19] L. Chen, Z. Jiang, L. Tong, *et al.*, "Perceptual underwater image enhancement with deep learning and physical priors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 8, pp. 3078–3092, Aug. 2021.

[20] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, no. 107038, Feb. 2020.

[21] Z. Wang, L. Shen, M. Xu, M. Yu, K. Wang, and Y. Lin, "Domain adaptation for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 32, pp. 1442–1457, Feb. 2023.

[22] A. Pramanick, S. Sarma, and A. Sur, "X-CAUNET: Cross-color channel attention with underwater image-enhancing transformer," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Apr. 2024, pp. 3550–3554.

[23] R. Liu, Z. Jiang, S. Yang, and X. Fan, "Twin adversarial contrastive learning for underwater image enhancement and beyond," *IEEE Trans. Image Process.*, vol. 31, pp. 4922–4936, Jul. 2022.

[24] S. Lu, F. Guan, and H. Lai, "Underwater image enhancement based on global features and prior distribution guided," *Image Vision Comput.*, vol. 148, no. 105101, pp. 1–15, Aug. 2024.

[25] Z. Zhang, Z. Jiang, J. Liu, X. Fan, and R. Liu, "Waterflow: Heuristic normalizing flow for underwater image enhancement and beyond," in *Proc. ACM Inter. Conf. Multimedia (ACM Multimedia)*, Oct. 2023, pp. 7314–7323.

[26] F. Bao, S. Nie, K. Xue, *et al.*, "All are worth words: A vit backbone for diffusion models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 22 669–22 679.

[27] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4861–4875, Dec. 2020.

[28] C. Li, C. Guo, W. Ren, *et al.*, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, Nov. 2020.

[29] A. Naik, A. Swarnakar, and K. Mittal, "Shallow-UWnet: Compressed model for underwater image enhancement," *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 18, pp. 15 853–15 854, May 2021.

[30] 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.

[31] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, Dec. 2014, pp. 2672–2680.

[32] 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.

[33] Z. Fu, W. Wang, Y. Huang, X. Ding, and K.-K. Ma, "Uncertainty inspired underwater image enhancement," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Nov. 2022, pp. 465–482.

[34] Q. Qi, Y. Zhang, F. Tian, *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.

[35] W. Peebles and S. Xie, "Scalable diffusion models with transformers," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 4172–4182.

[36] Y. Zhang, J. Yuan, and Z. Cai, "Dcgf: Diffusion-color-guided framework for underwater image enhancement," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, pp. 1–12, Dec. 2024.

[37] 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.

[38] X. Liu, Z. Gao, and B. M. Chen, "MLFcGAN: Multi-level feature fusion-based conditional gan for underwater image color correction," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 9, pp. 1488–1492, Sep. 2020.

[39] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Jul. 2015, pp. 2256–2265.

[40] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, Dec. 2020, pp. 6840–6851.

[41] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Feb. 2021, pp. 8162–8171.

[42] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 45, no. 4, pp. 4713–4726, Apr. 2023.

[43] J. Ho, C. Saharia, W. Chan, D. J. Fleet, M. Norouzi, and T. Salimans, "Cascaded diffusion models for high

fidelity image generation," *J. Mach. Learn. Res.*, vol. 23, no. 47, pp. 1–33, Jan. 2022.

[44] S. Lu, F. Guan, H. Zhang, and H. Lai, "Underwater image enhancement method based on denoising diffusion probabilistic model," *J. Vis. Commun. Image Represent.*, vol. 96, no. 103926, pp. 1–10, Oct. 2023.

[45] S. Lu, F. Guan, H. Zhang, and H. Lai, "Speed-up ddpm for real-time underwater image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 5, pp. 3576–3588, May 2024.

[46] M. Guan, H. Xu, G. Jiang, *et al.*, "DiffWater: Underwater image enhancement based on conditional denoising diffusion probabilistic model," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 17, pp. 2319–2335, Dec. 2024.

[47] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5718–5729.

[48] W. Shi, J. Caballero, F. Huszár, *et al.*, "Real-Time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.

[49] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assisted Interv. (MICCAI)*, May 2015, pp. 234–241.

[50] D. Jobson, Z. Rahman, and G. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.

[51] J. Zhou, Q. Gai, D. Zhang, K.-M. Lam, W. Zhang, and X. Fu, "Iacc: Cross-illumination awareness and color correction for underwater images under mixed natural and artificial lighting," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–15, Jan. 2024.

[52] 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.

[53] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li, "Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement," *IEEE Trans. Image Process.*, vol. 31, pp. 3997–4010, Jun. 2022.

[54] J. Korhonen and J. You, "Peak signal-to-noise ratio revisited: Is simple beautiful?" In *Proc. Int. Workshop on Qual. of Multim. Experience (QoMEX)*, Jul. 2012, pp. 37–38.

[55] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[56] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE J. Oceanic Eng.*, vol. 41, no. 3, pp. 541–551, Jul. 2016.

[57] V. N, P. D, M. C. Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *Proc. Nat. Conf. Commun. (NCC)*, Feb. 2015, pp. 1–6.

[58] S. Huang, K. Wang, H. Liu, J. Chen, and Y. Li, "Contrastive semi-supervised learning for underwater image restoration via reliable bank," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 18 145–18 155.

[59] D. Lowe, "Distinctive image features from scale-invariant key points," *Int. J. Comput. Vis.*, vol. 20, pp. 91–110, Nov. 2003.

[60] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.

[61] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.

[62] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

Yingbo Wang (Member, IEEE) received the PH.D. degree in the School of Optics and Photonics from Beijing Institute of Technology, Beijing, China, in 2021. He is currently a Lecturer at the School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology. His current research interests include image restoration, computer vision, and deep learning.

Kun He received a bachelor's degree in software engineering from Qingdao University of Science and Technology, Qingdao, China, in 2022. He is currently working toward a master's degree at Shaanxi University of Science and Technology, Xi'an, China. His research interests include underwater image processing, deep learning and its application to marine information processing.

Qiang Qu received a bachelor's degree in software engineering from Xi'an Aeronautical Institute, China, in 2021. He is currently pursuing a master's degree with the School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, China. His research interests include image processing, computer vision, and particularly in the domains of image restoration and enhancement.

Xiaogang Du received the B.S., M.S., and Ph.D. degrees from Lanzhou Jiaotong University, Lanzhou, China, in 2007, 2010, and 2017, respectively. He is currently an Associate Professor at the School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology. His research interests include computer vision, machine learning, and deep learning.

Tongfei Liu (Member, IEEE) received the master's degree from the Xi'an University of Technology, Xi'an, China, in 2020, and the Ph.D. degree from the School of Electronic Engineering, Xidian University, Xi'an, in 2023. He is currently a Lecturer at the School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an. His research interests include deep learning, spatial-spectral feature extraction, pattern recognition, ground target detection, and land cover/land use change detection and classification, through VHR remote sensing images (including satellite and aerial images).

Tao Lei (Senior Member, IEEE) received the Ph.D. degree in Information and Communication Engineering from Northwestern Polytechnical University, Xi'an, China, in 2011. From 2012 to 2014, he was a Postdoctoral Research Fellow with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China. From 2015 to 2016, he was a Visiting Scholar with the Quantum Computation and Intelligent Systems group at University of Technology Sydney, Australia. From 2017.07-2017.10, he was a Research Fellow with the College of Electronic and Computer Engineering, Brunel University of London, UK. From 2023.11-2024.04, he was a Research Fellow with the School of Information Science and Technology, Aichi University, Japan. He has authored and co-authored more than 100 research papers. He is currently a Professor with the Shaanxi Joint Laboratory of Artificial Intelligence, Shaanxi University of Science and Technology. His current research interests include image processing, pattern recognition, and machine learning.

Asoke k. Nandi (Life Fellow, IEEE) received the Ph.D. degree in physics from the University of Cambridge (Trinity College). He held academic positions in several universities, including Oxford, Oxford, Imperial College London, Strathclyde, and Liverpool as well as Finland Distinguished Professorship. In 2013 he moved to Brunel University of London. In 1983 Professor Nandi co-discovered the three fundamental particles known as W^+, W^-, and Z^0, providing the evidence for the unification of the electromagnetic and weak forces, for which the Nobel Committee for Physics in 1984 awarded the prize to two of his team leaders for their decisive contributions. His current research interests lie in signal processing and machine learning, with applications to machine health monitoring, functional magnetic resonance data, gene expression data, communications, and biomedical data. He made fundamental theoretical and algorithmic contributions to many aspects of signal processing and machine learning. He has much expertise in "Big Data". Professor Nandi has authored over 650 technical publications, including 310 journal papers as well as six books, entitled Image Segmentation: Principles, Techniques, and Applications (Wiley, 2022), Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines (Wiley, 2020), Automatic Modulation Classification: Principles, Algorithms and Applications (Wiley, 2015), Integrative Cluster Analysis in Bioinformatics (Wiley, 2015), Blind Estimation Using Higher-Order Statistics (Springer, 1999), and Automatic Modulation Recognition of Communications Signals (Springer, 1996). The H-index of his publications is 91 (Google Scholar) and ERDOS number is 2. Professor Nandi is a Fellow of the Royal Academy of Engineering and a Fellow of six other institutions including the IEEE. In 2023, he has been honoured by the Academia Europaea and the Academia Scientiarum et Artium Europaea. He has received many awards, including the IEEE Heinrich Hertz Award in 2012, the Glory of Bengal Award for his outstanding achievements in scientific research in 2010, the Water Arbitration Prize of the Institution of Mechanical Engineers in 1999, and the Mountbatten Premium of the Institution of Electrical Engineers in 1998. Professor Nandi was an IEEE Distinguished Lecturer (EMBS, 2018-2019).