SDUIE: Semi-Supervised Diffusion for Underwater Image Enhancement with Quant-Text Dual Control

Xiaofeng Cong[1] Yu-Xin Zhang[1] Hao Shen[2] Yeying Jin[3] Junming Hou[1] Jie Gui[1]*
[1]Southeast University [2]Anhui University of Science and Technology [3]Tecent
cxf_svip@163.com, {yuxinzhang,junming_hou}@seu.edu.cn
haoshenhs@gmail.com, jinyeying@u.nus.edu, guijie@ustc.edu

Abstract

Underwater images often exhibit dominant blue-green hues due to wavelength-dependent light attenuation. While existing enhancement methods have achieved promising performance, they typically overlook the subjective nature of visual preferences. To address this gap, we propose SDUIE, a level-aware Semi-supervised Diffusion framework for Underwater Image Enhancement that enables dual control through both quantitative and textual inputs. SDUIE-Quant allows continuous, numerical adjustment of enhancement levels via low-rank adaptation weight merging within a dual-branch diffusion model. This model comprises a supervised branch trained on synthetic underwater-terrestrial pairs and a self-supervised branch designed to preserve the natural hues of real-world underwater scenes. Building on this, SDUIE-Text introduces intuitive, language-guided control by aligning semantic prompts with visual enhancement effects, leveraging the learned fusion weights. This dual-modality design offers both precise control and flexible, user-preferred enhancement. Experimental results demonstrate that SDUIE achieves state-of-the-art results while better preserving the aesthetic qualities often missed by conventional methods. The source code is in https://github.com/Xiaofeng-life/SDUIE.

1. Introduction

Underwater photography is an essential topic for ocean exploration and documentation. However, due to the influence of water molecules and suspended particles, underwater images often suffer from color cast [19, 23, 25, 31, 41, 48, 49, 51, 56]. Underwater light transmission exhibits strong wavelength dependence, where rapid attenuation of red and yellow wavelengths contrasts with relatively stable green and blue transmission, resulting in the characteristic blue-green hue of underwater imagery [17, 22, 33, 54].

Currently, underwater image enhancement methods mainly divide into two types, namely the non-deep learning-based [2, 10, 32, 39, 40, 66, 70] and deep learning-based algorithms [1, 16, 45]. The non-deep learning-based methods primarily adopt digital image processing algorithms, imaging models, statistical observations, and prior assumptions [34, 60, 70]. However, due to the complexity of the real-world underwater scenes, the color correction performance of prior-based methods may be limited [5, 67].

The deep learning-based algorithms mainly employ neural networks to learn the mapping from distorted to clean images [43, 47]. Various network architectures have been carefully designed. For instance, approaches based on the CNN [28], Transformer [31, 41], GAN [50], Flow [61], and Diffusion mechanisms [42, 62]. Meanwhile, objective metrics [8, 29] for evaluation have been explored for the assessment of enhancement algorithms [21, 27, 65]. Quantitative and visual evaluations indicate that underwater vision research has been advanced by these well-designed algorithms. However, as stated in [4, 5, 18], algorithms that only provide fixed outputs may be inadequate for addressing the inherently ill-posed nature of underwater imaging.

To provide diverse enhancement results, CECF [5] and PWAE [18] explore style controllable outputs under different guidance images. However, they lack explicit control over the enhancement degree, which is a key factor influencing human perception. As established by [4], preservation of appropriate blue-green tones is crucial for maintaining underwater image characteristics, given the substantial variation in human preferences. For enhancement methods with the fixed-level enhancement result, they may suffer from the limitation of under-enhancement or over-enhancement [4]. Therefore, [4] designs a multi-domain translation method to generate images with different enhancement levels, but its latent space represents a broad range of style informations. Besides, it cannot accept text-mode instructions. These constraints highlight the ongoing challenge, which can adaptively adjust enhancement levels according to human perceptual requirements.

*Corresponding author

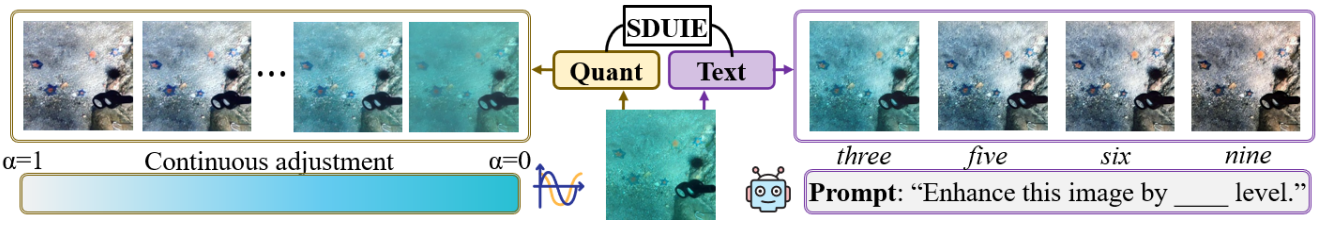


Figure 1. The enhancement results obtained by SDUIE-Quant under different fusion factors α and SDUIE-Text under different prmopts. Unlike existing methods that provide fixed outputs, SDUE enables dual control through both quantitative and textual inputs.

To this end, we propose a level-aware Semi-supervised Diffusion framework for Underwater Image Enhancement (SDUIE) with dual-modality controllability for balancing enhancement effects as shown in Fig. 1. The SDUIE-Quant enables continuous adjustment of enhancement intensity via low-rank adaptation weight merging [11, 59], which adopts a semi-supervised training strategy under a shared latent space diffusion model. The SDUIE-Text can perceive human instructions to generate enhanced images with different levels by utilizing the level-aware image-text pairs obtained by SDUIE-Quant. This dual-modality approach provides flexible enhancement tuning while maintaining robust performance across diverse underwater scenes. The overall contributions of this paper include

- We design a diffusion-based semi-supervised underwater image enhancement framework SDUIE, which supports both precise *quantitative* control (SDUIE-Quant) and intuitive *text* manipulation (SDUIE-Text) of enhancement levels. Quantitative and visual evaluation on real-world underwater images verifies its effectiveness.
- We propose a dual-branch latent space diffusion network for precise numerical control of enhancement levels, namely SDUIE-Quant, where one branch maintains hue consistency across domains while the other performs enhancement. Our synthetic-to-real adaptation strategy combines physically-based underwater image synthesis with shared latent space training on both synthesized and real-world data for better generalization.
- We introduce a prompt-guided modulation method for semantic manipulation of enhancement levels, namely SDUIE-Text, which establishes direct mappings between natural language instructions and visual enhancement effects. It leverages the learned fusion representations and image-prompt pairs that provided by SDUIE-Quant, extending our framework's controllable enhancement capabilities to semantic-level operation.

2. Related Work

Non-deep learning-based algorithms. Underwater image enhancement represents a long-standing challenge in computer vision research [39, 68]. Based on image processing algorithms, statistical assumptions, and prior knowledge of underwater imaging, a variety of effective algorithms have been proposed [34, 60, 70]. ULAP [39] designs an underwater light attenuation prior, which can facilitate underwater depth estimation and thus solve for clear images from imaging models. HLRP [70] proposes a hyper-laplacian reflectance prior, which turns a complex underwater image enhancement issue into simple subproblems. However, due to the complexity of the real world, prior-based methods may have limited correction capabilities [5, 67].

Deep learning-based algorithms. The data-driven learning capacity of deep neural networks has propelled substantial progress in underwater image enhancement, with diverse network architectures being developed [19, 23, 25, 31, 41, 48, 49, 51, 56, 69]. Shallow-UWNet [28] uses naive convolution and cross-layer connections to build an end-to-end underwater image enhancement network. Semi-UIR [13] utilizes the teacher-student knowledge-guided strategy to update the parameters of the asymmetric illumination-aware multi-scale network. SMDR-IS [55] proposes synergistic multi-scale detail refinement based on the intrinsic supervision, which can achieve multi-scale detail refinement and minimize the interference from irrelevant information. HCLR-Net [67] designs a hybrid attention module to extract and process underwater distortion features, along with a contrast regularization strategy that can promote high-quality feature learning. CPDM [38] takes the diffusion model as its fundamental model and builds a content-preserving framework to improve the scene restoration effect through content-aware training. DM-UIE [42] applies the Transformer architecture to conditional denoising diffusion probabilistic models and designs a non-uniform sampling strategy. WF-Diff [62] combines the Fourier transform and diffusion model [37, 44, 46] to improve the ability to restore details. Existing image enhancement algorithms have achieved impressive quantitative evaluation performance on baseline datasets. However, a topic that has been rarely studied is how to provide users with diverse enhancement results with specific semantics, thereby satisfying the subjective evaluation preferences of different users.

Algorithms that can provide diverse outputs. Underwater image enhancement is an ill-posed problem, where the perfect enhancement result may not exist. In addition

Algorithm 1 SDUIE-Quant Training & Inference

Initialize:
LoRA weights: $\theta = \{\theta_{\mathcal{A}}, \theta_{\mathcal{B}}\}$
Encoder Paths: $\{\mathcal{E}_{ie}, \mathcal{E}_{ir}\}$ equip $\{\theta^{ie}_{\mathcal{E}}, \theta^{ir}_{\mathcal{E}}\}$
Diffusion Path: $\mathcal{U}$ equip $\theta_{\mathcal{U}}$
Decoder Paths: $\{\mathcal{D}_{ie}, \mathcal{D}_{ir}\}$ equip $\{\theta^{ie}_{\mathcal{D}}, \theta^{ir}_{\mathcal{D}}\}$
Discriminate Paths: $\mathcal{K}_{ie}$ by $\Theta^{ie}_{\mathcal{K}}$, $\mathcal{K}_{ir}$ by $\Theta^{ir}_{\mathcal{K}}$
Training Phase:
for step $= 1$ **to** Total Steps **do**
  Sample $\{y^i\}_{i=1}^B \sim p(y)$, $\{x_r^i\}_{i=1}^B \sim p(x_r)$
  # Image Enhancement:
  Compute $\{y^i\}_{i=1}^B \rightarrow \{x_s^i\}_{i=1}^B$ *# Eq. 3*
  Compute $x_s \rightarrow \mathcal{E}_{ir} \rightarrow \mathcal{U} \rightarrow \mathcal{D}_{ie} \rightarrow \hat{y}_{x_s}$
  Update $\theta^{ir}_{\mathcal{E}}, \theta_{\mathcal{U}}, \theta^{ie}_{\mathcal{D}}$ via $\nabla \mathcal{L}^{ie}_{p}$ *# Eq. 4*
  Compute $y \rightarrow \mathcal{K}_{ie}$ and $\hat{y}_{x_s} \rightarrow \mathcal{K}_{ie}$
  Update $\theta^{ir}_{\mathcal{E}}, \theta_{\mathcal{U}}, \theta^{ie}_{\mathcal{D}}, \Theta^{ie}_{\mathcal{K}}$ via $\nabla \mathcal{L}^{ie}_{a}$ *# Eq. 5*
  # Hue Preservation:
  for $x \in \{x_s, x_r\}$ **do**
    Compute $x \rightarrow \mathcal{E}_{ir} \rightarrow \mathcal{U} \rightarrow \mathcal{D}_{ir} \rightarrow \bar{x}$
    Compute $x \rightarrow \mathcal{K}_{ir}$ and $\bar{x} \rightarrow \mathcal{K}_{ir}$
    Update $\theta^{ir}_{\mathcal{E}}, \theta_{\mathcal{U}}, \theta^{ir}_{\mathcal{D}}, \Theta^{ir}_{\mathcal{K}}$ via $\nabla(\sum \mathcal{L}^{ir}_{p,x} + \lambda_2 \sum \mathcal{L}^{ir}_{a,x})$ *# Eq. 6, 8, 9, 11*
  end for
  Compute $y \rightarrow \mathcal{E}_{ie} \rightarrow \mathcal{U} \rightarrow \mathcal{D}_{ie} \rightarrow \bar{y}$
  Compute $y \rightarrow \mathcal{K}_{ie}$ and $\bar{y} \rightarrow \mathcal{K}_{ie}$
  Update $\theta^{ie}_{\mathcal{E}}, \theta_{\mathcal{U}}, \theta^{ie}_{\mathcal{D}}, \Theta^{ie}_{\mathcal{K}}$ via $\nabla(\sum \mathcal{L}^{ir}_{p,y} + \lambda_2 \sum \mathcal{L}^{ir}_{a,y})$ *# Eq. 7, 10*
end for
Inference Phase:
Require image x_r, fusion factor $\alpha \in [0, 1]$
Compute $\mathcal{S}(\theta^{ie}_{\mathcal{D}}, \theta^{ir}_{\mathcal{D}}; \alpha) \rightarrow \mathcal{D}_m$ *# Eq. 13*
Compute $x_r \rightarrow \mathcal{E}_{ir} \rightarrow \mathcal{U} \rightarrow \mathcal{D}_m \rightarrow \hat{y}_\alpha$
return $\hat{y}_\alpha$

to the above studies that provide a single enhancement result, recent studies have attempted to design algorithms that can provide diverse visual results. PWAE [18] designs a Wasserstein autoencoder network that can obtain a 2D latent space style representation, which generates restoration results of corresponding styles through different style guidance images. Utilizing the separation of color and content, CECF [5] uses images with long-wavelength color as guidance to obtain color-adjustable underwater enhancement effects of organisms. UIESS [4] proposes to control the degree of image enhancement by manipulating the style latent space. However, the latent space of UIESS represents a wide range of style information. Different from the research of UIESS, PWAE, and CECF, this paper explores how to utilize the prior knowledge of the pre-trained diffusion model to implement both precise numerical control and intuitive semantic manipulation of the enhancement level during the enhancement process.

3. Methods

A. Overall framework. Fig. 2 presents the overall framework of our SDUIE. SDUIE-Quant employs a dual-branch design with separate reconstruction and enhancement decoders for preserving natural tones while enhancing the image. It bridges the synthetic-real domain gap by combining physically-guided underwater image simulation with shared latent space training. The synthetic and real-world underwater dataset are denoted as $\mathcal{S}_{x_s} = \{x_s^i\}_{i=1}^{\mathcal{N}}$ and $\mathcal{S}_{x_r} = \{x_r^i\}_{i=1}^{\mathcal{M}}$. The real-world terrestrial dataset is $\mathcal{S}_y = \{y^i\}_{i=1}^{\mathcal{N}}$. And $x_s, x_r, y \in \mathbb{R}^{H \times W \times C}$, where H, W, and C denote height, width, and channels, respectively. The encoder $\mathcal{E}_{ie}$ is responsible for the feature encoding of y, while the $\mathcal{E}_{ir}$ is used for the encoding of x_s and x_r. The decoder $\mathcal{D}_{ie}$ is used for image enhancement of simulated underwater images (x_s) and hue preservation of terrestrial images (y). $\mathcal{D}_{ir}$ is employed to the hue preservation of simulated (x_s) and real-world (x_r) underwater images. Encoders $\{\mathcal{E}_{ie}, \mathcal{E}_{ir}\}$ and decoders $\{\mathcal{D}_{ie}, \mathcal{D}_{ir}\}$ share the same UNet $\mathcal{U}$ in latent space. The fixed prompt of SDUIE-Quant is denoted as p, while the prompt representing different enhancement levels of SDUIE-Text is marked as p_α.

The encoder $\mathcal{E}'_{ir}$, latent space diffusion model $\mathcal{U}'$, and decoder $\mathcal{D}'_{ie}$ of SDUIE-Text are initialized from the SDUIE-Quant. Building on the enhanced images obtained by SDUIE-Quant and the constructed human instructions pairs, SDUIE-Text learns text-semantic to enhancement-level mappings through image-text matching fine-tuning, enabling natural language-based control of enhancement.

B. Preliminary. The proposed SDUIE takes the diffusion model [30, 36] as the backbone, and fine-tunes the pre-trained model by low-rank adaptation (LoRA) [6]. The latent space diffusion model [35] is trained by

$$\mathcal{L} = \mathbb{E}_{\mathcal{E}(x),p,\epsilon\sim N(0,1),t}[\|\epsilon - \epsilon_\theta(z_t, t, \tau_\theta(p))\|_2^2], \quad (1)$$

where t, z_t, ϵ, ϵ_θ, and τ_θ are the time step, latent variable, noise from Gaussian $N(0, 1)$, estimated noise, and domain specific encoder, respectively. $\mathcal{E}(x)$ is the encoder for the input image x. The weight matrix $\mathcal{W} \in \mathbb{R}^{n \times m}$ of the pre-trained model can be updated to $\mathcal{W}^* = \mathcal{W} + \mathcal{B}\mathcal{A}$ by LoRA, where $\mathcal{B} \in \mathbb{R}^{n \times r}$, $\mathcal{A} \in \mathbb{R}^{r \times m}$ and r denotes the rank. The merge of LoRAs [63] with balance factors ζ_{ie} and ζ_{ir} is

$$\mathcal{W}^* = \mathcal{W} + (\zeta_{ie} \times \mathcal{B}_{ie}\mathcal{A}_{ie} + \zeta_{ir} \times \mathcal{B}_{ir}\mathcal{A}_{ir}). \quad (2)$$

C. Training and inference of SDUIE-Quant. During SDUIE-Quant training, the model's parameter updates are guided by both simulated (x_s) and real-world (x_r) underwater images. The framework employs a dual-path architecture: (1) an image enhancement path, which learns to transform simulated underwater images into ground-truth-like outputs, and (2) a hue preservation path, which is dedicated to self-reconstruction tasks. The complete training and inference workflow operates as follows.

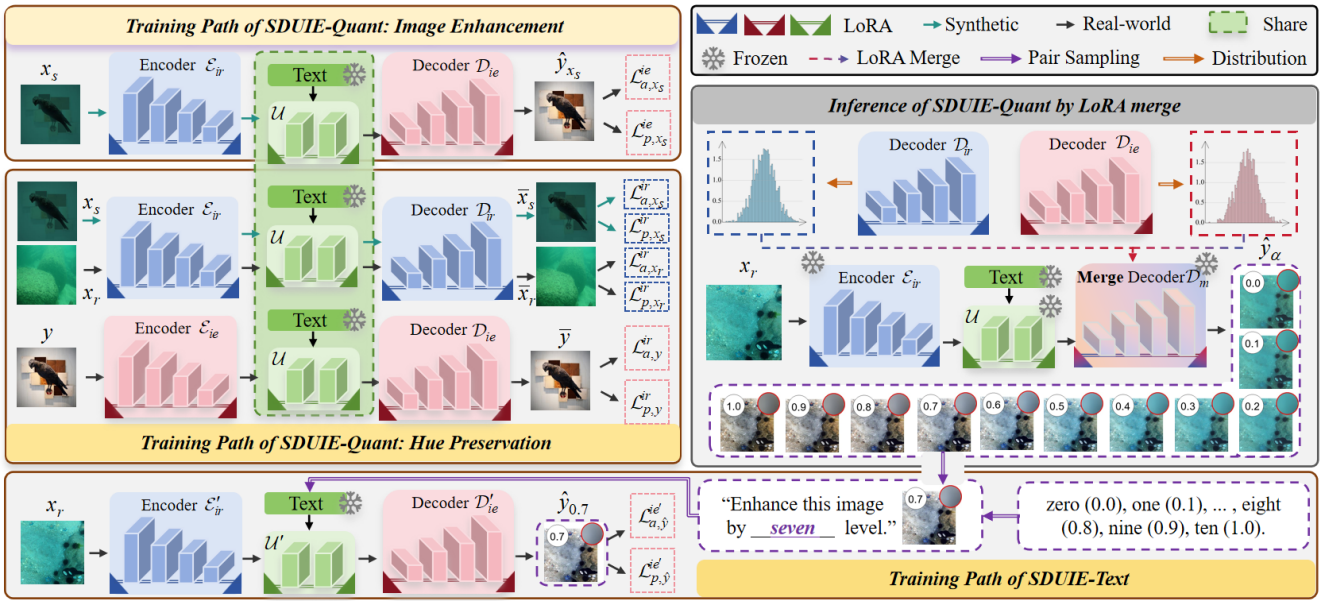


Figure 2. Training and inference of the Proposed SDUIE Framework. Our framework comprises two key components: SDUIE-Quant for precise numerical control, and SDUIE-Text for semantic-level control via natural language. They are built upon a shared latent space and a synthetic-to-real adaptation strategy, which utilizes both synthetic and real-world data to achieve robust enhancement.

C.1 Image enhancement path. For training the image enhancement decoder, high-quality reference images are essential. We utilize real-world terrestrial images as ground truth labels, generating corresponding underwater images through a physics-based simulation pipeline. This employs a color distortion-aware imaging model [58] built upon established underwater imaging principles [7, 9], mathematically formulated as:

$$x_s(c) = \eta(c) \circ [y(c) \circ e^{-\beta d} + L(c) \circ (1 - e^{-\beta d})], \quad (3)$$

where $\circ$ represents pixel-wise multiplication and c indicates the channel index. The model incorporates four key parameters: depth map $d \in \mathbb{R}^{H \times W}$, scattering coefficient $\beta \in \mathbb{R}^1$, environmental light $L \in \mathbb{R}^3$, and color distortion vector $\eta \in \mathbb{R}^3$ (derived from manually selected underwater color patches). Following standard underwater enhancement approaches, SDUIE-Quant performs color correction through its enhancement path $\mathcal{E}_{ir} \rightarrow \mathcal{U} \rightarrow \mathcal{D}_{ie}$ when processing simulated images. The transformation is optimized using a pixel-wise (subscript p) loss function between enhanced outputs and ground truth references, formulated as:

$$\mathcal{L}^{ie}_{p,x_s} = \mathbb{E}_{(x_s,y)\sim p(x_s,y)} ||\mathcal{D}_{ie}(\mathcal{U}(\tau_\theta(p), \mathcal{E}_{ir}(x_s))) - y||_1. \quad (4)$$

Moreover, to optimize the quality of enhanced outputs, we implement joint training of the discriminator $\mathcal{K}_{ie}$ and diffusion model, with the adversarial (subscript a) loss formulated as:

$$\mathcal{L}^{ie}_{a,x_s} = \mathbb{E}_y \log[\mathcal{K}_{ie}(y)] + \mathbb{E}_{x_s} \log[1 - \mathcal{K}_{ie}(\mathcal{D}_{ie}(\mathcal{H}^s_{ir}))], \quad (5)$$

where $\mathcal{H}^s_{ir} = \mathcal{U}(\tau_\theta(p), \mathcal{E}_{ir}(x_s))$.

C.2 Hue preservation path. While the image enhancement path demonstrates strong performance on synthetic data, its generalization to real-world underwater images may be constrained by the domain gap. To bridge this limitation, we incorporate a self-reconstruction mechanism that enables the encoder and diffusion model in latent space to learn authentic underwater patterns through dual processing pathways: $\mathcal{E}_{ir} \rightarrow \mathcal{U} \rightarrow \mathcal{D}_{ir}$ and $\mathcal{E}_{ie} \rightarrow \mathcal{U} \rightarrow \mathcal{D}_{ie}$. The reconstruction losses for both simulated (x_s) and real-world (x_r, y) images are formally defined as:

$$\begin{cases} \mathcal{L}^{ir}_{p,x_s} = \mathbb{E}_{x_s\sim p(x_s)} ||\mathcal{D}_{ir}(\mathcal{U}(\tau_\theta(p), \mathcal{E}_{ir}(x_s))) - x_s||_1, & (6) \\ \mathcal{L}^{ir}_{p,y} = \mathbb{E}_{y\sim p(y)} ||\mathcal{D}_{ie}(\mathcal{U}(\tau_\theta(p), \mathcal{E}_{ie}(y))) - y||_1, & (7) \\ \mathcal{L}^{ir}_{p,x_r} = \mathbb{E}_{x_r\sim p(x_r)} ||\mathcal{D}_{ir}(\mathcal{U}(\tau_\theta(p), \mathcal{E}_{ir}(x_r))) - x_r||_1. & (8) \end{cases}$$

By sharing the diffusion model $\mathcal{U}$ across all encoder-decoder pairs, our framework achieves latent space alignment between synthetic and real-world image representations. Furthermore, we employ adversarial training to optimize the reconstruction process across various domains, with the corresponding loss formulated as:

$$\begin{cases} \mathcal{L}^{ir}_{a,x_s} = \mathbb{E}_{x_s} \log[\mathcal{K}_{ir}(x_s)] + \mathbb{E}_{x_s} \log[1 - \mathcal{K}_{ir}(\mathcal{D}_{ir}(\mathcal{H}^s_{ir}))], & (9) \\ \mathcal{L}^{ir}_{a,y} = \mathbb{E}_y \log[\mathcal{K}_{ie}(y)] + \mathbb{E}_y \log[1 - \mathcal{K}_{ie}(\mathcal{D}_{ie}(\mathcal{H}^r_{ie}))], & (10) \\ \mathcal{L}^{ir}_{a,x_r} = \mathbb{E}_{x_r} \log[\mathcal{K}_{ir}(x_r)] + \mathbb{E}_{x_r} \log[1 - \mathcal{K}_{ir}(\mathcal{D}_{ir}(\mathcal{H}^r_{ir}))], & (11) \end{cases}$$

where $\mathcal{H}^r_{ie} = \mathcal{U}(\tau_\theta(p), \mathcal{E}_{ie}(y))$, $\mathcal{H}^r_{ir} = \mathcal{U}(\tau_\theta(p), \mathcal{E}_{ir}(x_r))$, and $\mathcal{K}_{ir}$ denote the discriminator for hue preservation path.

C.3 Total losses. The overall loss function is derived from the joint optimization objectives of both the image enhancement and hue preservation pathways, formulated as:

$$\begin{aligned} \mathcal{L}_{all} = &\mathcal{L}^{ie}_{p,x_s} + \mathcal{L}^{ie}_{a,x_s} + \lambda_1(\mathcal{L}^{ir}_{p,x_s} + \mathcal{L}^{ir}_{p,y} + \mathcal{L}^{ir}_{p,x_r}) \\ &+ \lambda_2(\mathcal{L}^{ir}_{a,x_s} + \mathcal{L}^{ir}_{a,y} + \mathcal{L}^{ir}_{a,x_r}), \end{aligned} \quad (12)$$

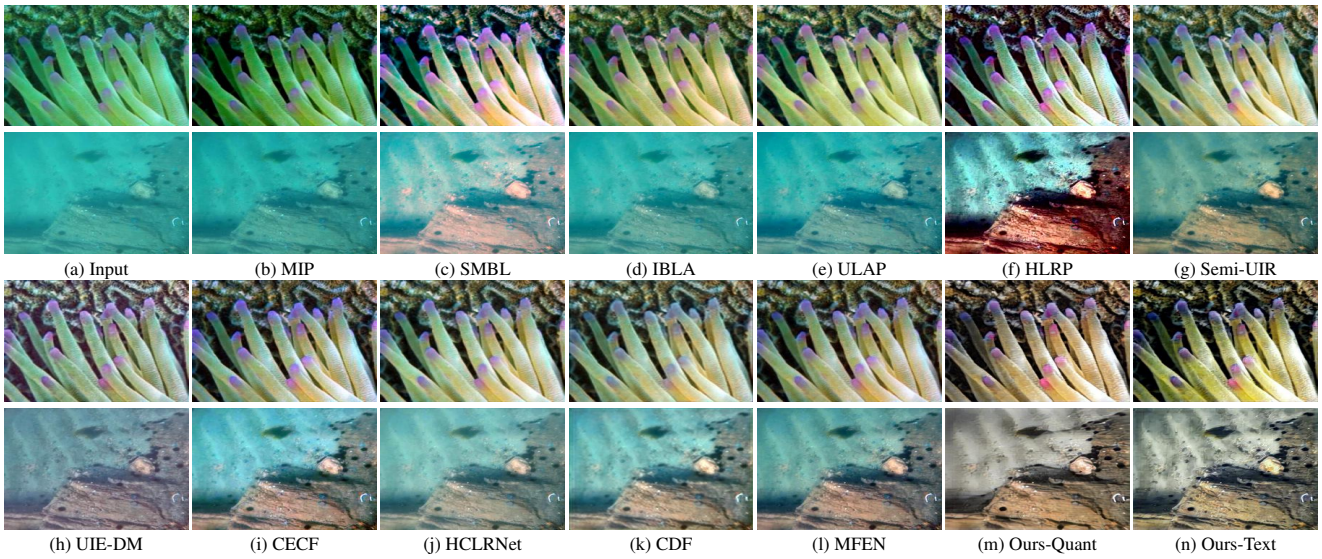


Figure 3. Visual comparisons of SDUIE with other underwater image enhancement methods.

where λ_1 and λ_2 are the weight factors. During the training process, we update the parameters of different networks alternately. The details of the training process are given in Algorithm 1.

C.4 Inference stage. During inference, SDUIE-Quant processes the real-world underwater image x_r to extract latent space features $z^{ie}_{x_r} = \mathcal{U}(\tau_\theta(p), \mathcal{E}_{ir}(x_r))$. Enhancement degree control is achieved through LoRA weight merging between the two decoders, balancing feature strengths from both the image enhancement and hue preservation. Spherical interpolation [15] is used to replaces the second term in Eq. 2 for specific layers. Given layer weights ω_{ie} in $\mathcal{D}_{ie}$ and ω_{ir} in $\mathcal{D}_{ir}$, the merging weight $\mathcal{D}_m$ is obtained by:

$$\mathcal{S}(\omega_{ie}, \omega_{ir}; \alpha) = \omega_{ir} \frac{\sin((1-\alpha)\theta)}{\sin\theta} + \omega_{ie} \frac{\sin(\alpha\theta)}{\sin\theta}, \quad (13)$$

where $\theta = \arccos(\omega_{ie} \cdot \omega_{ir})$. The fusion factor $\alpha \in [0, 1]$ controls the ratio between components, with $\alpha = 0$ corresponding to pure hue preservation by $\mathcal{D}_{ir}$ and $\alpha = 1$ representing complete enhancement by $\mathcal{D}_{ie}$.

D. Training and inference of SDUIE-Text. Building on SDUIE-Quant, SDUIE-Text extends the framework by adding prompt-conditioned enhancement capability. SDUIE-Text consists of three components initialized from SDUIE-Quant: encoder $\mathcal{E}'_{ir}$, diffusion model $\mathcal{U}'$, and decoder $\mathcal{D}'_{ie}$. The training process employs an interpolated dataset $\mathcal{S}_{\hat{y}} = \{\hat{y}_\alpha, p_\alpha\}$ with enhancement levels generated via LoRA merge, where text prompts p_α follow the format "Enhance this image by ___ level.". This design enables SDUIE-Text to associate text descriptions with specific enhancement effects. The training and inference workflow proceeds as follows.

D.1 Training stage. During the training of SDUIE-Text, x_r is sampled from $\mathcal{S}_{x_r}$ and an enhancement level $\hat{y}_\alpha$ is randomly sampled from $\mathcal{S}_{\hat{y}}$. Through this enhancement process, SDUIE-Text is able to learn different levels of enhancement degrees represented by the prompt p_α. The pixel wise loss and adversarial loss under the discriminator $\mathcal{K}'_{ie}$ are formulated as

$$\begin{cases} \mathcal{L}^{ie'}_{p,\hat{y}} = \mathbb{E}_{(x_r,\hat{y}_\alpha)} ||\mathcal{D}'_{ie}(\mathcal{H}'_{ir}) - \hat{y}_\alpha||_1, & (14) \\ \mathcal{L}^{ie'}_{a,\hat{y}} = \mathbb{E}_{\hat{y}_\alpha} \log[\mathcal{K}'_{ie}(\hat{y}_\alpha)] + \mathbb{E}_{x_r} \log[1 - \mathcal{K}'_{ie}(\mathcal{D}'_{ie}(\mathcal{H}'_{ir}))], & (15) \end{cases}$$

where $\mathcal{H}'_{ir} = \mathcal{U}'(\tau_\theta(p_\alpha), \mathcal{E}'_{ir}(x_r))$.

D.2 Inference stage. During the inference process of SDUIE-Text, users can change the degree of color enhancement directly through the prompt template, such as "Enhance this image by ___ level.". Other calculations during inference are consistent with SDUIE-Quant.

4. Experiments

Data preparation. The real-world indoor images are selected as labels [64]. 1001 pairs of images are synthesized using Eq. 3. The datasets used for comparison enhancement results include UCCS [24], EUVP [14] (the test set), U45 [20], and challenging-60 [19].

Comparisons and evaluations. Comparisons methods include MIP [2], SMBL [40], IBLA [32], ULAP [39], HLRP [70], Semi-UIR [13], UIE-DM [42], CECF [5], HCLRNet [67], CDF [57], and MFEN [3]. The training data of the learning-based algorithms are from [19]. Since CECF and HLRP can only perform fixed-size inference, the calculation of metrics is unified to 256×256. The widely adopted UIQM [29], UCIQE [52], and URANKER [8] are used as evaluation metrics.

Implementation details. The batch size is set to 1. The Adam optimizer is used. The LoRA ranks [12, 26, 53] of UNet $\mathcal{U}$ and VAE (encoders and decoders) are set to 8 and 4, respectively. The weight factors λ_1 and λ_2 are set to

Table 1. Quantitative evaluation results, where red and blue denote the best and second best.

Method	Challenge-60			U45			UCCS			EUVP		
	UIQM	UCIQE	URANKER	UIQM	UCIQE	URANKER	UIQM	UCIQE	URANKER	UIQM	UCIQE	URANKER
MIP	3.627	0.537	0.099	3.875	0.570	0.110	2.465	0.506	-0.597	2.904	0.562	0.495
SMBL	3.863	0.579	0.870	3.544	0.602	0.771	4.048	0.565	1.853	3.470	0.608	1.298
IBLA	3.474	0.577	0.441	2.770	0.592	0.522	2.480	0.531	-0.216	2.898	0.579	0.929
ULAP	4.152	0.567	-0.081	3.550	0.596	-0.126	2.810	0.517	-0.086	3.705	0.587	0.801
HLRP	4.392	0.644	1.370	4.908	0.601	1.364	4.760	0.646	1.299	4.491	0.655	1.917
Semi-UIR	4.098	0.576	1.417	4.431	0.599	2.032	4.049	0.549	1.335	3.660	0.597	1.818
UIE-DM	3.874	0.563	1.531	4.097	0.597	1.983	4.166	0.567	1.468	3.652	0.598	1.897
CECF	4.127	0.572	1.501	4.354	0.591	2.013	4.286	0.553	1.474	3.681	0.590	1.792
HCLRNet	4.100	0.565	1.253	4.293	0.602	1.970	3.935	0.537	1.088	3.680	0.586	1.706
CDF	3.757	0.571	1.467	3.851	0.593	1.873	3.971	0.552	1.549	3.490	0.588	1.880
MFEN	4.263	0.578	1.400	4.427	0.605	1.994	4.280	0.557	1.408	3.940	0.595	1.790
SDUIE-Quant	5.010	0.598	1.850	5.501	0.600	2.478	5.351	0.563	2.481	4.966	0.604	2.292
SDUIE-Text	4.750	0.611	1.895	4.991	0.620	2.492	5.320	0.558	2.297	4.728	0.609	2.373

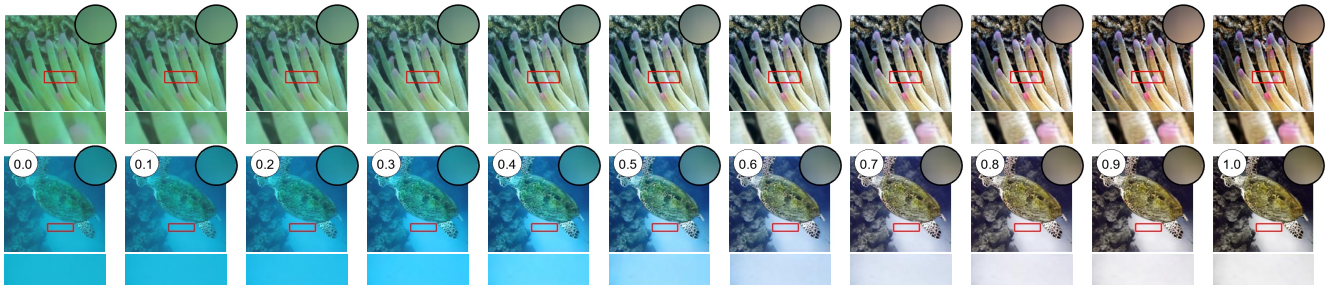


Figure 4. Visual results of SDUIE-Quant under different fusion factor α.

1. The total steps for SDUIE-Quant and SDUIE-Text are 10000 and 5000, respectively. The learning rate is 0.0001.

4.1. Comparison with other enhancement methods

The quantitative results in Table 1 demonstrate that the performance of SDUIE is competitive compared to existing underwater image enhancement algorithms. Meanwhile, the results obtained by SDUIE-Quant and SDUIE-Text are quite close (for fair comparison, the enhancement level of SDUIE-Quant and SDUIE-Text are set to 1.0 and "ten" in the experiments). Fig. 3 compares the visual effects of the proposed SDUIE with other underwater image enhancement algorithms. More visualization results are in the Supplementary Materials. The visual results show that MIP, IBLA, and ULAP may not effectively deal with the blue effect of underwater images. SMBL, HLRP, and CECF tend to be with under-enhancement. The brightness of the enhanced images for blue scenes by UIE-DM and HCLR-Net is reduced. The color correction capabilities of existing methods have certain limitations. Overall, the proposed SDUIE has a better enhancement effect on color and details. Quantitative and visual results verify the effectiveness of the proposed method.

4.2. Level-aware enhancement results

Analysis of SDUIE-Quant. Fig. 4 illustrates the progressive enhancement effects achieved by varying the fusion factor α. Two key observations emerge: (1) enhancement intensity scales proportionally with α values, and (2) the dominant blue-green hue systematically transitions toward terrestrial images. These visual results confirm SDUIE-Quant's capability to generate controllable enhancement outputs across a continuous adjustment. Table 2 presents the corresponding quantitative analysis using UIQM, UCIQE, and URANKER metrics. The data reveals a strong positive correlation between α values and objective scores, demonstrating that enhancement intensity can be effectively calibrated. This dual validation (visual and metric-based) confirms the SDUIE-Quant's effectiveness in producing tunable enhancement results.

Analysis of SDUIE-Text. The text-guided enhancement capability of SDUIE-Text is demonstrated in Fig. 5, where varying prompts produce corresponding enhancement effects. This visual evidence confirms the model's successful interpretation of semantic instructions for enhancement. Quantitative analysis in Table 3 reveals a consistent positive correlation between the enhancement level specified in the prompts and the measured objective metrics. The progressive improvement in evaluation scores with increasing enhancement intensity further validates the effectiveness of SDUIE-Text's prompt-based control mechanism.

Objective assessment. As mentioned in the section Introduction, there may be subjective diversity when evaluating the degree of enhancement levels, resulting in different preferences for over-enhancement and under-enhancement. To this end, we select different real-world underwater scenes

Table 2. Evaluation results under different fusion factors of SDUIE-Quant.

Data	Metrics	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
UCCS	UIQM ↑	1.501	1.662	1.905	2.252	2.740	3.395	4.073	4.609	4.996	5.239	**5.351**
	UCIQE ↑	0.488	0.488	0.490	0.496	0.506	0.519	0.529	0.538	0.550	0.558	**0.563**
	URANKER ↑	-1.415	-1.228	-0.924	-0.469	0.157	0.905	1.545	1.993	2.274	2.422	**2.481**
EUVP	UIQM ↑	2.135	2.259	2.483	2.798	3.208	3.697	4.172	4.529	4.810	4.950	**4.966**
	UCIQE ↑	0.506	0.508	0.515	0.527	0.544	0.561	0.572	0.581	0.591	0.599	**0.604**
	URANKER ↑	-0.520	-0.341	-0.031	0.391	0.900	1.429	1.798	2.005	2.132	2.225	**2.292**

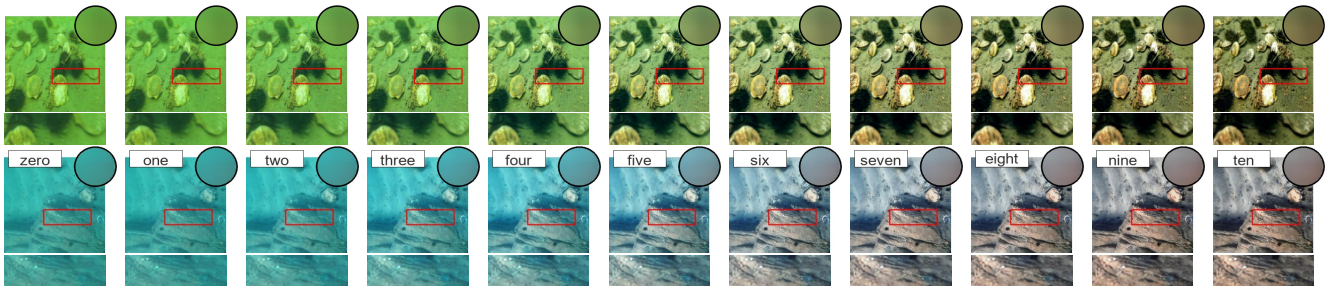


Figure 5. Visual results of SDUIE-Text under prompts with different semantics of enhancement levels.

Table 3. Evaluation results under different levels of SDUIE-Text on the EUVP dataset.

Metrics	one	two	three	four	five
UIQM ↑	2.659	2.784	3.071	3.362	3.873
UCIQE ↑	0.535	0.546	0.564	0.577	0.591
URANKER↑	0.382	0.670	1.154	1.544	1.941
Metrics	six	seven	eight	nine	ten
UIQM ↑	4.182	4.369	4.547	4.703	**4.728**
UCIQE ↑	0.597	0.601	0.604	0.607	**0.609**
URANKER↑	2.149	2.247	2.306	2.348	**2.373**

with various color cast effects and employ SDUIE-Text to generate the enhanced images with different enhancement levels. Four volunteers with no image processing experience participated in the image quality evaluation (because the purpose of this experiment is to verify the subjective inconsistency of different volunteers, and the results in Fig. 6 have verified this, we did not look for more volunteers). The results in Fig. 6 show that human subjective choices are not as stable as objective evaluation metrics. Humans may subjectively support that appropriate retention of blue-green tones can maintain the beauty of the image.

The change of red channel under different enhancement levels. As the underwater depth increases, the proportion of the red component in the image gradually decreases. Therefore, an important purpose of the underwater image enhancement algorithm is to enhance the intensity of the red channel. Fig. 7 calculates the change in the mean pixel value of the red channel. The results show that with the increase of the enhancement levels, the red component can be effectively enhanced.

Summary. Quantitative and visual results verify that SDUIE-Quant and SDUIE-Text can effectively control the enhancement levels of underwater images.

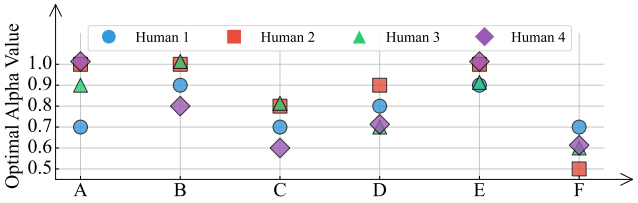


Figure 6. Subjective scores of the best enhancement effects for different scenes (x-axis).

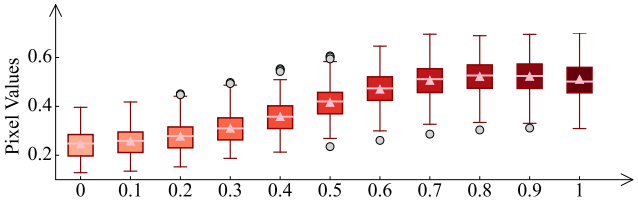


Figure 7. The pixel mean of red channel under different α (x-axis). With the increase of the enhancement levels, the red component can be effectively enhanced.

Table 4. Quantitative ablation results.

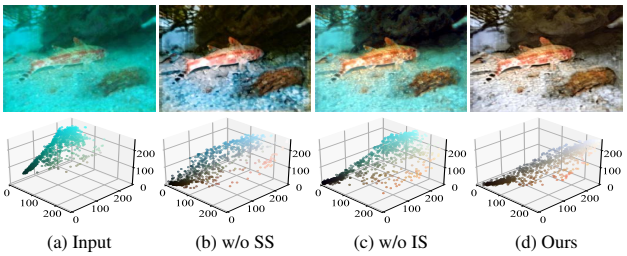
Metrics	w/o SS	w/o IS	w/o AT	Ours
UIQM ↑	5.220	4.152	5.173	**5.351**
UCIQE ↑	0.560	**0.563**	0.559	**0.563**
URANKER ↑	2.204	1.606	1.771	**2.481**

4.3. Ablation study and discussions

Analysis of the semi-supervised processes. In the training process of SDUIE, a semi-supervised method combining synthetic and real-world images is adopted. Fig. 8 shows that after removing the semi-supervised training (w/o SS),

Table 5. Inference speeds of different methods.

Method	MIP	SMBL	IBLA	ULAP	HLRP
Time (ms)	312.5	306.3	391.8	549.2	163.3
Method	HCLR	UIE-DM	CDF	MFEN	SDUIE
Time (ms)	83.5	205.4	42.3	34.6	83.4



(a) Input (b) w/o SS (c) w/o IS (d) Ours

Figure 8. Ablation study of the semi-supervised (w/o SS) and image synthesis (w/o IS).

the color correction ability is obviously limited, and the clarity of the scene is reduced. Meanwhile, Table 4 also verifies the importance of the semi-supervised training.

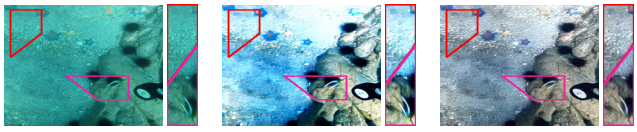
Analysis of our synthetic dataset. To make the enhancement results close to terrestrial imaging, we construct a simulated dataset using terrestrial images. To verify the effectiveness of this data simulation process, we train SDUIE on the UIEB dataset [19] rather than our image synthesis (w/o IS) pipeline. The visual results in Fig. 8 show that without IS, the color cast with blue effects in the image cannot be effectively suppressed. Moreover, quantitative results in Table 4 also demonstrate the necessity of IS.

Analysis of adversarial losses. We remove the $\mathcal{K}_{ie}$ and $\mathcal{K}_{ir}$ for adversarial training (w/o AT) and display the quantitative results in Table 4, which suggests that adversarial loss is beneficial to the quality of generated images. The visual results is in Supplementary Materials.

Spherical interpolation and linear interpolation. The interpolation process of SDUIE-Quant uses spherical interpolation to merge the two LoRA weights for convolution layers. In order to verify which method is better, we compare the visual effects obtained by spherical interpolation and linear interpolation as shown in Fig. 9. Visual results show that linear interpolation may cause overexposure in local areas of the image, while spherical interpolation exhibits better smoothness for brightness gradients. More visual results under different underwater image datasets are in the Supplementary Materials.

Evaluation of the inference speed. The evaluation results of inference speeds for different underwater image enhancement algorithms are presented in Table 5. SDUIE achieves a competitive inference speed, which is attributable to its efficient single-step diffusion design.

Diversified level description templates. To further investigate the impact of prompts on SDUIE-Text's performance, the SDUIE-Text is trained using 100 description templates,



(a) Input (b) Linear (c) Spherical

Figure 9. Comparison of the visual effects of different interpolation methods, where $\alpha = 0.6$.

Table 6. Results under different prompts on U45 dataset.

Metrics	T0	T1	T2	T3	T4	T5
UIQM ↑	**4.991**	4.968	4.957	4.986	4.916	4.945
UCIQE ↑	0.620	**0.623**	0.619	0.620	0.622	0.620
URANKER↑	2.492	2.325	2.577	**2.694**	2.435	2.497

with additional tested prompt-level templates is constructed for evaluation, including "T0: Enhance this image by ___ level." (Table 1), "T1: Refine the underwater picture to achieve ___ levels of enhancement.", and "T2: Raise underwater image clarity to level ___ for a better look.". T3, T4 and T5 are in the Supplementary Materials. Correspondingly, we evaluate the level control performance of SDUIE-Text. Table 6 shows the metric values obtained with different prompt templates (with the level of ten). The results demonstrate that SDUIE-Text can effectively perform the enhancement process under different prompts. We further experiment with two additional large language models to construct the prompt set. The UIQM, UCIQE, and URANKER metrics on Challenge-60 are (1) 4.744, 0.610, and 1.894, and (2) 4.749, 0.611, and 1.896, respectively. The results indicate that the performance across different language models is comparable.

5. Conclusions

In this paper, different from existing studies, we study how to control the degree of enhancement of underwater images. A novel semi-supervised diffusion framework for underwater image enhancement, named SDUIE, is proposed. The SDUIE-Quant enables continuous enhancement adjustment through low-rank adaptation weight merging. The dual-branch architecture of SDUIE-Quant combines synthetic data learning with real-world image reconstruction through a shared latent space, effectively maintaining underwater scene characteristics. The SDUIE-Text is trained under the fusion results obtained by low-rank weight merging and provides semantic-level control via natural language instructions. Experimental results demonstrate the framework's effectiveness in producing high-quality enhancements while preserving natural underwater color tones.

References

[1] Xuewen Bing, Wenqi Ren, Yang Tang, Gary G Yen, and Qiyu Sun. Domain adaptation for in-air to underwater image enhancement via deep learning. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2023. 1

[2] Nicholas Carlevaris-Bianco, Anush Mohan, and Ryan M Eustice. Initial results in underwater single image dehazing. In *Oceans 2010 Mts/IEEE Seattle*, pages 1–8, 2010. 1, 5

[3] Ashutosh Chauhan, Dakshi Goel, Meghna Kapoor, Badri Narayan Subudhi, and Vinit Jakhetiya. Illumination aware multi-scale attention fusion model for underwater image enhancement. *IEEE Transactions on Artificial Intelligence*, 2025. 5

[4] Yu-Wei Chen and Soo-Chang Pei. Domain adaptation for underwater image enhancement via content and style separation. *IEEE Access*, 10:90523–90534, 2022. 1, 3

[5] Xiaofeng Cong, Jie Gui, and Junming Hou. Underwater organism color fine-tuning via decomposition and guidance. In *AAAI Conference on Artificial Intelligence*, pages 1389–1398, 2024. 1, 2, 3, 5

[6] Guanqi Ding, Chengyu Yang, Shuhui Wang, Xincheng Li, Jinzhe Zhang, Xin Jin, and Qingming Huang. Dis^{2}booth: Learning image distribution with disentangled features for text-to-image diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2744–2752, 2025. 3

[7] Zhenqi Fu, Huangxing Lin, Yan Yang, Shu Chai, Liyan Sun, Yue Huang, and Xinghao Ding. Unsupervised underwater image restoration: From a homology perspective. In *AAAI Conference on Artificial Intelligence*, pages 643–651, 2022. 4

[8] Chunle Guo, Ruiqi Wu, Xin Jin, Linghao Han, Weidong Zhang, Zhi Chai, and Chongyi Li. Underwater ranker: Learn which is better and how to be better. In *AAAI Conference on Artificial Intelligence*, pages 702–709, 2023. 1, 5

[9] Junyu Hao, Hongbo Yang, Xia Hou, and Yang Zhang. Two-stage underwater image restoration algorithm based on physical model and causal intervention. *IEEE Signal Processing Letters*, 30:120–124, 2022. 4

[10] Guojia Hou, Nan Li, Peixian Zhuang, Kunqian Li, Haihan Sun, and Chongyi Li. Non-uniform illumination underwater image restoration via illumination channel sparsity prior. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(2):799–814, 2024. 1

[11] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021. 2

[12] Jia-Hong Huang, Yixian Shen, Hongyi Zhu, Stevan Rudinac, and Evangelos Kanoulas. Gradient weight-normalized low-rank projection for efficient llm training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 24123–24131, 2025. 5

[13] Shirui Huang, Keyan Wang, Huan Liu, Jun Chen, and Yunsong Li. Contrastive semi-supervised learning for underwater image restoration via reliable bank. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 18145–18155, 2023. 2, 5

[14] Md Jahidul Islam, Youya Xia, and Junaed Sattar. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2):3227–3234, 2020. 5

[15] Young Kyun Jang, Dat Huynh, Ashish Shah, Wen-Kai Chen, and Ser-Nam Lim. Spherical linear interpolation and text-anchoring for zero-shot composed image retrieval. In *European Conference on Computer Vision*, pages 239–254, 2024. 5

[16] Qun Jiang, Yunfeng Zhang, Fangxun Bao, Xiuyang Zhao, Caiming Zhang, and Peide Liu. Two-step domain adaptation for underwater image enhancement. *Pattern Recognition*, 122:108324, 2022. 1

[17] Qiuping Jiang, Yaozu Kang, Zhihua Wang, Wenqi Ren, and Chongyi Li. Perception-driven deep underwater image enhancement without paired supervision. *IEEE Transactions on Multimedia*, 2023. 1

[18] Guisik Kim, Sung Woo Park, and Junseok Kwon. Pixel-wise wasserstein autoencoder for highly generative dehazing. *IEEE Transactions on Image Processing*, 30:5452–5462, 2021. 1, 3

[19] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29:4376–4389, 2019. 1, 2, 5, 8

[20] Hanyu Li, Jingjing Li, and Wei Wang. A fusion adversarial underwater image enhancement network with a public test dataset. *arXiv preprint arXiv:1906.06819*, 2019. 5

[21] Kunqian Li, Hongtao Fan, Qi Qi, Chi Yan, Kun Sun, and QM Jonathan Wu. Tctl-net: Template-free color transfer learning for self-attention driven underwater image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023. 1

[22] Yinyi Li, Liquan Shen, Mengyao Li, Zhengyong Wang, and Lihao Zhuang. Ruiesr: Realistic underwater image enhancement and super resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023. 1

[23] Qiong Liu, Qi Zhang, Wei Liu, Wenbai Chen, Xinwang Liu, and Xiangke Wang. Wsds-gan: A weak-strong dual supervised learning method for underwater image enhancement. *Pattern Recognition*, page 109774, 2023. 1, 2

[24] Risheng Liu, Xin Fan, Ming Zhu, Minjun Hou, and Zhongxuan Luo. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE transactions on circuits and systems for video technology*, 30(12): 4861–4875, 2020. 5

[25] Risheng Liu, Zhiying Jiang, Shuzhou Yang, and Xin Fan. Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Transactions on Image Processing*, 31:4922–4936, 2022. 1, 2

[26] Xiao-Yang Liu, Rongyi Zhu, Daochen Zha, Jiechao Gao, Shan Zhong, Matt White, and Meikang Qiu. Differentially private low-rank adaptation of large language model using federated learning. *ACM Transactions on Management Information Systems*, 16(2):1–24, 2025. 5

[27] Pan Mu, Hanning Xu, Zheyuan Liu, Zheng Wang, Sixian Chan, and Cong Bai. A generalized physical-knowledge-guided dynamic model for underwater image enhancement. In *ACM International Conference on Multimedia*, pages 7111–7120, 2023. 1

[28] Ankita Naik, Apurva Swarnakar, and Kartik Mittal. Shallow-uwnet: Compressed model for underwater image enhancement (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 15853–15854, 2021. 1, 2

[29] Karen Panetta, Chen Gao, and Sos Agaian. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3):541–551, 2015. 1, 5

[30] Gaurav Parmar, Taesung Park, Srinivasa Narasimhan, and Jun-Yan Zhu. One-step image translation with text-to-image models. *arXiv preprint arXiv:2403.12036*, 2024. 3

[31] Lintao Peng, Chunli Zhu, and Liheng Bian. U-shape transformer for underwater image enhancement. *IEEE Transactions on Image Processing*, 2023. 1, 2

[32] Yan-Tsung Peng and Pamela C Cosman. Underwater image restoration based on image blurriness and light absorption. *IEEE transactions on image processing*, 26(4):1579–1594, 2017. 1, 5

[33] Hao Qi, Huiyu Zhou, Junyu Dong, and Xinghui Dong. Deep color-corrected multi-scale retinex network for underwater image enhancement. *IEEE Transactions on Geoscience and Remote Sensing*, 2023. 1

[34] Yuan Rao, Wenjie Liu, Kunqian Li, Hao Fan, Sen Wang, and Junyu Dong. Deep color compensation for generalized underwater image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023. 1, 2

[35] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022. 3

[36] Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. In *ECCV*, pages 87–103, 2024. 3

[37] Fei Shen, Hu Ye, Sibo Liu, Jun Zhang, Cong Wang, Xiao Han, and Yang Wei. Boosting consistency in story visualization with rich-contextual conditional diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 6785–6794, 2025. 2

[38] Xiaowen Shi and Yuan-Gen Wang. Cpdm: Content-preserving diffusion model for underwater image enhancement. *Scientific Reports*, 14(1):31309, 2024. 2

[39] Wei Song, Yan Wang, Dongmei Huang, and Dian Tjondronegoro. A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration. In *Advances in Multimedia Information Processing Pacific-Rim Conference on Multimedia*, pages 678–688, 2018. 1, 2, 5

[40] Wei Song, Yan Wang, Dongmei Huang, Antonio Liotta, and Cristian Perra. Enhancement of underwater images with statistical model of background light and optimization of transmission map. *IEEE Transactions on Broadcasting*, 66(1): 153–169, 2020. 1, 5

[41] Yi Tang, Takafumi Iwaguchi, Hiroshi Kawasaki, Ryusuke Sagawa, and Ryo Furukawa. Autoenhancer: Transformer on u-net architecture search for underwater image enhancement. In *Proceedings of the Asian Conference on Computer Vision*, pages 1403–1420, 2022. 1, 2

[42] Yi Tang, Hiroshi Kawasaki, and Takafumi Iwaguchi. Underwater image enhancement by transformer-based diffusion model with non-uniform sampling for skip strategy. In *ACM International Conference on Multimedia*, pages 5419–5427, 2023. 1, 2, 5

[43] Hao Wang, Weibo Zhang, Lu Bai, and Peng Ren. Metalantis: A comprehensive underwater image enhancement framework. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 1

[44] Weilun Wang, Jianmin Bao, Wengang Zhou, Dongdong Chen, Dong Chen, Lu Yuan, and Houqiang Li. Sindiffusion: Learning a diffusion model from a single natural image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 2

[45] Zhengyong Wang, Liquan Shen, Mai Xu, Mei Yu, Kun Wang, and Yufei Lin. Domain adaptation for underwater image enhancement. *IEEE Transactions on Image Processing*, 32:1442–1457, 2023. 1

[46] Yongliang Wu, Shiji Zhou, Mingzhuo Yang, Lianzhe Wang, Heng Chang, Wenbo Zhu, Xinting Hu, Xiao Zhou, and Xu Yang. Unlearning concepts in diffusion model via concept domain correction and concept preserving gradient. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8496–8504, 2025. 2

[47] Zhiheng Wu, Zhengxing Wu, Xingyu Chen, Yue Lu, and Junzhi Yu. Self-supervised underwater image generation for underwater domain pre-training. *IEEE Transactions on Instrumentation and Measurement*, 2024. 1

[48] Yaofeng Xie, Lingwei Kong, Kai Chen, Ziqiang Zheng, Xiao Yu, Zhibin Yu, and Bing Zheng. Uveb: A large-scale benchmark and baseline towards real-world underwater video enhancement. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2024. 1, 2

[49] Xinwei Xue, Zexuan Li, Long Ma, Qi Jia, Risheng Liu, and Xin Fan. Investigating intrinsic degradation factors by multi-branch aggregation for real-world underwater image enhancement. *Pattern Recognition*, 133:109041, 2023. 1, 2

[50] Haorui Yan, Zhenwei Zhang, Jing Xu, Tingting Wang, Ping An, Aobo Wang, and Yuping Duan. Uw-cyclegan: Model-driven cyclegan for underwater image restoration. *IEEE Transactions on Geoscience and Remote Sensing*, 2023. 1

[51] Shuaizheng Yan, Xingyu Chen, Zhengxing Wu, Min Tan, and Junzhi Yu. Hybrur: A hybrid physical-neural solution for unsupervised underwater image restoration. *IEEE Transactions on Image Processing*, 2023. 1, 2

[52] Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12):6062–6071, 2015. 5

[53] Yaming Yang, Dilxat Muhtar, Yelong Shen, Yuefeng Zhan, Jianfeng Liu, Yujing Wang, Hao Sun, Weiwei Deng, Feng Sun, Qi Zhang, et al. Mtl-lora: Low-rank adaptation for

multi-task learning. In *AAAI Conference on Artificial Intelligence*, pages 22010–22018, 2025. 5

[54] Meng Yu, Liquan Shen, Zhengyong Wang, and Xia Hua. Task-friendly underwater image enhancement for machine vision applications. *IEEE Transactions on Geoscience and Remote Sensing*, 2023. 1

[55] Dehuan Zhang, Jingchun Zhou, Chunle Guo, Weishi Zhang, and Chongyi Li. Synergistic multiscale detail refinement via intrinsic supervision for underwater image enhancement. In *Proceedings of the AAAI conference on artificial intelligence*, pages 7033–7041, 2024. 2

[56] Fan Zhang, Shaodi You, Yu Li, and Ying Fu. Atlantis: Enabling underwater depth estimation with stable diffusion. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2024. 1, 2

[57] Haopeng Zhang, Hongli Xu, Xiaosheng Yu, Xiangyue Zhang, Xiujing Gao, and Chengdong Wu. Cdf-uie: Leveraging cross-domain fusion for underwater image enhancement. *IEEE Transactions on Geoscience and Remote Sensing*, 2025. 5

[58] Jing Zhang, Yang Cao, Zheng-Jun Zha, and Dacheng Tao. Nighttime dehazing with a synthetic benchmark. In *ACM international conference on multimedia*, pages 2355–2363, 2020. 4

[59] Kaiwen Zhang, Yifan Zhou, Xudong Xu, Bo Dai, and Xingang Pan. Diffmorpher: Unleashing the capability of diffusion models for image morphing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 7912–7921, 2024. 2

[60] Weidong Zhang, Peixian Zhuang, Hai-Han Sun, Guohou Li, Sam Kwong, and Chongyi Li. Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Transactions on Image Processing*, 31: 3997–4010, 2022. 1, 2

[61] Zengxi Zhang, Zhiying Jiang, Jinyuan Liu, Xin Fan, and Risheng Liu. Waterflow: heuristic normalizing flow for underwater image enhancement and beyond. In *ACM International Conference on Multimedia*, pages 7314–7323, 2023. 1

[62] Chen Zhao, Weiling Cai, Chenyu Dong, and Chengwei Hu. Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8281–8291, 2024. 1, 2

[63] Ming Zhong, Yelong Shen, Shuohang Wang, Yadong Lu, Yizhu Jiao, Siru Ouyang, Donghan Yu, Jiawei Han, and Weizhu Chen. Multi-lora composition for image generation. *arXiv preprint arXiv:2402.16843*, 2024. 3

[64] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 40(6):1452–1464, 2017. 5

[65] Jingchun Zhou, Boshen Li, Dehuan Zhang, Jieyu Yuan, Weishi Zhang, Zhanchuan Cai, and Jinyu Shi. Ugif-net: An efficient fully guided information flow network for underwater image enhancement. *IEEE Transactions on Geoscience and Remote Sensing*, 2023. 1

[66] Jingchun Zhou, Qilin Gai, Dehuan Zhang, Kin-Man Lam, Weishi Zhang, and Xianping Fu. Iacc: Cross-illumination awareness and color correction for underwater images under mixed natural and artificial lighting. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024. 1

[67] Jingchun Zhou, Jiaming Sun, Chongyi Li, Qiuping Jiang, Man Zhou, Kin-Man Lam, Weishi Zhang, and Xianping Fu. Hclr-net: Hybrid contrastive learning regularization with locally randomized perturbation for underwater image enhancement. *International Journal of Computer Vision*, pages 1–25, 2024. 1, 2, 5

[68] Jingchun Zhou, Zongxin He, Dehuan Zhang, Siyuan Liu, Xianping Fu, and Xuelong Li. Spatial residual for underwater object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 2

[69] Jingchun Zhou, Chunjiang Liu, Dehuan Zhang, Zongxin He, Ferdous Sohel, and Qiuping Jiang. Rsuia: Dynamic no-reference underwater image assessment via reinforcement sequences. *IEEE Transactions on Multimedia*, 2025. 2

[70] Peixian Zhuang, Jiamin Wu, Fatih Porikli, and Chongyi Li. Underwater image enhancement with hyper-laplacian reflectance priors. *IEEE Transactions on Image Processing*, 31:5442–5455, 2022. 1, 2, 5