water

MDPI

Article

An Underwater Image Enhancement Method Based on Diffusion Model Using Dual-Layer Attention Mechanism

Hong Zhang *, Ran He and Wei Fang

School of Information Engineering, Minzu University of China, Beijing 100080, China;
heran_1024@163.com (R.H.); 22302092@muc.edu.cn (W.F.)
* Correspondence: zhanghong751103@muc.edu.cn

Abstract: Diffusion models have been increasingly utilized in various image-processing tasks, such as segmentation, denoising, and enhancement. These models also show exceptional performance in enhancing underwater images. However, conventional models for underwater image enhancement often face the challenge of simultaneously improving color restoration and super-resolution. This paper introduces a dual-layer attention mechanism that integrates spatial and channel attention to enhance color restoration, while preserving critical image features. Additionally, specific scale factors and interpolation methods are employed during the upsampling process to increase resolution. The proposed DL-UW method achieves significant enhancements in color, illumination, and resolution for low-quality underwater images, resulting in high PSNR, SSIM, and UIQM values. The model demonstrates a robust performance on different datasets, confirming its broad applicability and effectiveness.

Keywords: underwater image enhancement; diffusion model; dual-layer attention mechanism

check for updates

Citation: Zhang, H.; He, R.; Fang, W. An Underwater Image Enhancement Method Based on Diffusion Model Using Dual-Layer Attention Mechanism. *Water* **2024**, *16*, 1813. https://doi.org/10.3390/w16131813

Academic Editor: Long Ho

Received: 22 April 2024
Revised: 24 May 2024
Accepted: 5 June 2024
Published: 26 June 2024

1. Introduction

In the realm of computer vision, while extensive exploration of land has been conducted, our understanding of oceans and lakes remains relatively limited due to the formidable challenges posed by underwater exploration. Scientists strive to fathom the underwater world using various technological methods, such as underwater optical imaging techniques [1], UAI (underwater acoustic imaging), and Quantum LiDAR technology [2]. The acquisition and imaging of underwater optical images hold significant potential applications across diverse fields, including oceanographic research, underwater resource exploration, mapping submarine topography, and underwater archaeology [3]. However, natural factors like light scattering and absorption in water significantly impact image visibility and contrast, often leading to captured images falling short of meeting stringent quality requirements for other application scenarios.

At the beginning of the 21st century, researchers conducted extensive studies on underwater image restoration focused on physical models. However, the intricate nature of underwater physical environments posed significant challenges to accurate modeling, thereby limiting the effectiveness of these physical model-based approaches. Consequently, researchers have explored image enhancement techniques that do not rely on physical models. These methodologies have introduced novel perspectives for underwater image processing and have, to a certain extent, enhanced the quality of underwater images [4]. Nevertheless, as technology continues to advance and applications deepen, there is still a need to explore more advanced and effective methods for processing underwater images to meet higher demands for image quality and clarity in practical applications.

In recent years, the field of computer vision, particularly image processing, has witnessed significant breakthroughs attributed to neural networks. In 2020, Ho et al. [5] introduced the Denoising Diffusion Probabilistic Model (DDPM), which has garnered widespread attention. The model has demonstrated a remarkable performance in various aspects, including the denoising, restoration, super-resolution, generation, and enhancement of images. As a latent variable model, DDPM builds a bridge between the data distribution and simpler distributions, such as Gaussian distribution, progressively transforming data into pure Gaussian noise. This process follows the principles of the Markov process and involves reverse data reconstruction and training across a weighted variational boundary, ultimately enabling the generation of high-quality synthetic images. The successful deployment of DDPM not only pioneers new methodologies in the realm of computer vision but also provides a potent tool for image-processing tasks, thereby propelling further advancements in this field.

Building upon prior research, we have observed that conventional underwater image enhancement tasks frequently encounter challenges, such as unbalanced color distribution, inadequate super-resolution effects, and constraints in enhancing individual underwater images [6]. Inspired by these challenges, this paper proposes a novel underwater image enhancement method based on the Denoising Diffusion Probabilistic Model, aimed at addressing these issues. Underwater image enhancement has stricter requirements for color, saturation, contrast, and resolution compared to general image enhancement. The objectives of this experiment are not only to achieve super-resolution of underwater images but also to significantly enhance the overall quality of the images while preserving key features, ultimately producing high-quality images that meet human visual perception standards. This method is expected to overcome the shortcomings of traditional underwater image enhancement techniques, offering new perspectives and tools for research and applications in related fields.

Our main contributions are summarized as follows:

1. We introduced a denoising neural network framework based on the probabilistic diffusion model, which has shown remarkable performance in enhancing the resolution and color restoration of low-quality underwater images. When compared with other models performing the same tasks, the model proposed in this paper demonstrated superior competitive strength.
2. We propose an innovative dual-layer attention mechanism. Specifically, the first layer of attention focuses on the interrelationships of spatial locations, effectively capturing spatial correlations in the input data. This helps the model better understand and utilize feature information from different positions. The second layer implements inter-channel attention weighting by adaptively learning attention coefficients to weight the input channel features. This allows the model to more accurately focus on important channel features, thereby enhancing the effectiveness of feature expression and optimizing image enhancement results. Compared to the baseline model, the PSNR index improved by 32.2%, and the SSIM index increased by 7.9%. The UIQM is improved by 17.4% compared to the input image.
3. We adopted a sub-pixel interpolation upsampling strategy, utilizing a 1.25× scale factor to augment the number of sub-pixels, in conjunction with bilinear interpolation. This strategy preserves the details of the image while preventing the generated images from becoming overly smooth. This approach significantly enhanced the image resolution. Compared to models not using this strategy, the restored images showed a 23.39% improvement in the PSNR and a 7.3% increase in the SSIM.

2. Related Work

2.1. Underwater Image Super-Resolution

Contrasted with atmospheric optical imaging, underwater optical imaging is constrained by its distinctive imaging environment, frequently resulting in a less-than-optimal image quality. This is primarily due to various noise interferences present in underwater environments, such as scattering, absorption, and background light, which collectively degrade the overall image quality. More complexly, the propagation of visible light in water is wavelength-dependent, leading to underwater images often displaying bluish-green hues, significantly reduced contrast, and distorted colors, making it difficult to reveal more image details. Hence, accomplishing accurate, swift, and precise restoration of details and features within such intricate environments continues to pose a pressing challenge for ensuing underwater tasks.

To address the aforementioned complex issues in underwater imaging, researchers have proposed numerous underwater image enhancement and restoration methods, which can be broadly classified into two major categories: physical model-based restoration methods and non-physical model-based image enhancement methods.

The underwater image restoration methods based on physical models are committed to simulating the degradation process of underwater images through mathematical modeling and then estimating model parameters to restore clear underwater images [7]. These methods must fully consider the physical laws of underwater imaging, including factors such as light absorption, scattering, and color distortion. For instance, there is the self-tuning underwater image restoration filter based on the Jaffe–McGlamery model [8], as well as algorithms that estimate scene depth to eliminate the effects of light scattering in underwater images [9]. Although these methods have improved the quality of underwater images to a certain extent, there are still some limitations. This is because existing modeling methods are unable to fully reproduce the imaging process of underwater images, making precise modeling difficult.

Non-physical model approaches focus on directly improving the visual quality of underwater images through image-processing techniques, rather than relying on mathematical models of underwater optical imaging. These methods usually do not involve complex calculations of physical parameters but rather adjust the pixel values of the image directly to enhance image quality. In early research, researchers often directly applied traditional image enhancement methods to underwater images. Jung et al. [10] proposed an adaptive joint tri-directional filter (AJTF) that could enhance the clarity of images and depth maps. Wang S. H. et al. [11] focused on processing images under uneven illumination conditions. With the increasing demand for innovation in underwater image enhancement techniques, more and more image enhancement methods that can adapt to underwater conditions have been proposed. Among underwater image enhancement, color distortion is a particularly prominent issue. In 2014, Kan et al. [12] utilized the water absorption spectrum to estimate changes in tristimulus values and proposed a color-restoration method based on the water absorption spectrum. Additionally, in 2015, Li et al. [13] successfully improved the contrast, brightness, color, and visibility of underwater images by comprehensively utilizing methods such as color compensation, histogram equalization, saturation, and intensity stretching. Although these methods have improved the visual quality of underwater images to a certain extent, they lack generalizability. However, optimistically, these methods provide more reliable data support for subsequent underwater image analysis and applications.

In recent years, learning-based data-driven methods have achieved significant progress in the field of underwater image enhancement. In 2017, Wang et al. [14] proposed the UIE-Net end-to-end underwater image enhancement network based on the attenuation model, which utilized a simple convolutional neural network to estimate the transmission image and the attenuation coefficients of three channels, achieving the effect of image color enhancement. In 2017, Li et al. proposed a generative adversarial network named CycleGAN [15], which generated a model capable of performing multiple types of image-to-image transformations through adversarial training. Based on this, Li et al. proposed WaterGAN [16] the following year, a two-stage algorithm using two fully convolutional networks, which split underwater image enhancement into two sub-tasks: synthesizing relative depth maps and color restoration. These models effectively improved the quality of underwater images through generative adversarial methods. However, these methods did not comprehensively address the practical application needs of underwater images. Early research focused on restoring image colors, while the most widely used GAN-based methods in recent years often have issues such as uncontrollable output results and severe image degradation. Our proposed method can effectively address these issues.

2.2. Diffusion Model

Since the advent of diffusion models based on non-physical models in 2020, a large number of research works have emerged in the field of computer vision adopting this novel modeling method. The distinctiveness of this model lies in its construction of a Markov chain that links a sample, x_0, with a pure Gaussian distribution, x_T (a simple distribution), thereby facilitating a Markov process. Transitions of this chain are learned to reverse a diffusion process, which is a Markov chain that gradually adds noise to the data in the opposite direction of sampling until the signal is destroyed [5]. By studying the reverse process of this chain, the parameterization of a relatively simple neural network can be achieved. Figure 1 clearly demonstrates the process of adding noise and denoising in the diffusion model. In this process, the gradual addition of noise to the sample is known as the forward process, $q(\mathrm{x}_t|\mathrm{x}_{t-1})$, while the process of progressively denoising from pure Gaussian noise is termed the unknown reverse process, $p_\theta(\mathrm{x}_{t-1}|\mathrm{x}_t)$. To approximate this reverse process, neural networks are employed with the minimization of cross-entropy as the training objective, specifically defined by Equation (1):

$$\min_{\theta}\mathbb{E}_{\mathrm{q}(\mathrm{x}_0:\mathrm{x}_\mathrm{T})}\left[\log\frac{\mathrm{q}(\mathrm{x}_0:\mathrm{x}_\mathrm{T})}{\mathrm{p}_\theta(\mathrm{x}_0:\mathrm{x}_\mathrm{T})}\right]\rightarrow\min_{\theta}\left[\sum_{(\mathrm{x}_0:\mathrm{x}_\mathrm{T})\in x}\log\frac{\mathrm{q}(\mathrm{x}_0:\mathrm{x}_\mathrm{T})}{\mathrm{p}_\theta(\mathrm{x}_0:\mathrm{x}_\mathrm{T})}\right] \tag{1}$$

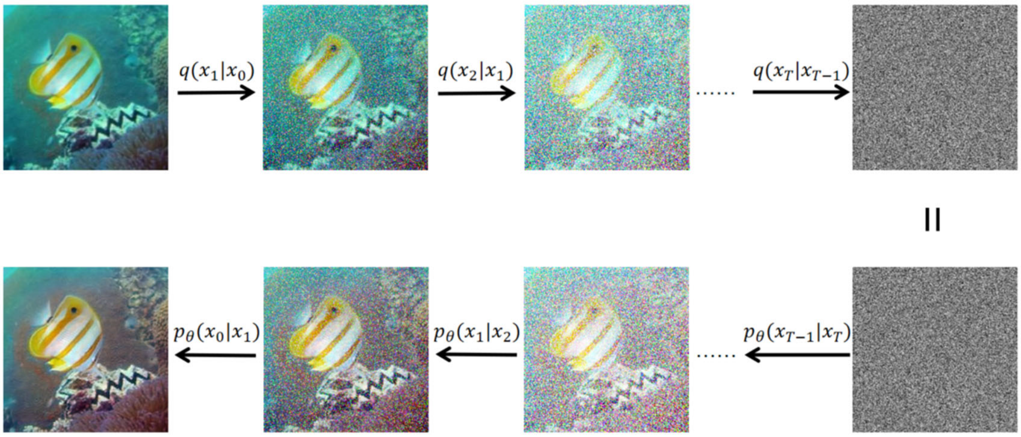


Figure 1. Schematic diagram of the forward and reverse processes of the diffusion model.

When the trained model, $p_\theta(x_0 : x_T|c)$, closely approximates the true distribution, $q(x_0 : x_T)$, it can be inferred that $p_\theta(x_{t-1}|x_t)$ also closely approximates $q(x_{t-1}|x_t)$. As previously mentioned, the forward process can be considered a Markov process; hence, its conditional probability, $q(x_t|x_{t-1})$, follows a Gaussian distribution, specifically expressed as $\mathcal{N}(x_t; \sqrt{\alpha_t}x_{t-1}, \beta_t I)$, where t ranges from 1 to T. Here, $\mathcal{N}(x; \mu, \sigma^2)$ denotes a Gaussian distribution with mean, μ, and variance, σ^2, and its probability density function is $\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/(2\sigma^2)}$. In diffusion models, β_t is typically chosen such that $0 \leq \beta_1 < \beta_2 < \cdots < \beta_T < 1$, with $\beta_t I$ representing the variance of the Gaussian distribution, where I is the identity matrix. Here, $0 \leq \beta_t < 1$ is a scalar hyperparameter that describes the magnitude of variance, i.e., the intensity of the "Gaussian noise". Additionally, $\sqrt{\alpha_t}x_{t-1}$ is the mean of this conditional probability, where $\alpha_t \equiv 1 - \beta_t$. Consequently, x_t can be viewed as a linear combination of the random variable, x_{t-1}, and a standard normal distribution, ε_t: $x_t = \sqrt{\alpha_t}x_{t-1} + \sqrt{\beta_t}\varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, I)t = 1, 2, \cdots T$.

Given that the objective function is the variational lower bound under Markov conditions, the objective function can be expressed using Bayes' theorem as Equation (2).

The terms L_0, L_t, and L_T are further specified in Equations (2)–(5), respectively, where $t = 1, 2, \cdots T-1$. Based on these, Equation (6) is subsequently derived.

$$L(\theta) = \mathbb{E}_{\mathrm{q}}\left[\log\frac{q(x_T, x_0)}{p_\theta(x_T)} + \sum_{\mathrm{t=2}}^{\mathrm{T}}\log\frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)} + \log\frac{1}{p_\theta(x_0|x_1)}\right] \tag{2}$$

$$\mathrm{L}_0 \equiv \mathbb{E}_{\mathrm{q}}\left[\log\frac{1}{\mathrm{p}_\theta(\mathrm{x}_0|\mathrm{x}_1)}\right] \tag{3}$$

$$L_t \equiv \mathbb{E}_{\mathrm{q}}\left[\log\frac{q(x_t|x_{t+1}, x_0)}{p_\theta(x_t|x_{t+1})}\right] \tag{4}$$

$$L_T \equiv \mathbb{E}_{\mathrm{q}}\left[\log\frac{q(x_T, x_0)}{p_\theta(x_T)}\right] \tag{5}$$

$$L = L_0 + \sum_{t=1}^{T-1} L_t + L_T \tag{6}$$

Since the state at time, T, is pure Gaussian noise, and L_T does not depend on model parameters, L_T can be ignored in the derivation process. Utilizing Bayes' theorem, the posterior distribution can be represented through $\widetilde{\beta}_t$ and $\widetilde{\mu}_t(x_t, x_0)$ as follows: $q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \widetilde{\mu}_t(x_t, x_0), \widetilde{\beta}_t I)$. Here, $\widetilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$ and $\widetilde{\mu}_t(x_t, x_0) = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\sqrt{\alpha_t}x_t + \frac{\beta_t}{1-\bar{\alpha}_t}\sqrt{\bar{\alpha}_{t-1}}x_0$.

3. Methodology

In image enhancement tasks based on diffusion models, the input, x_0, typically consists of the original low-resolution image, I^{LR}, with the corresponding high-quality image, I^{HR}, serving as the reference output. This design establishes a clear supervised-learning framework, clearly defining the model's learning objective—the mapping process from the low-quality image, I^{LR}, to the high-quality image, I^{HR}.

For the specific task of underwater image enhancement, which this paper focuses on, it is essential not only to enhance the image details but also to improve the coloration. Therefore, the low-quality original underwater image, I_{UW}^{LR}, is used as the model's input, x_0, with the corresponding high-quality underwater image, I_{UW}^{HR}, selected as the reference output, $\widetilde{x}_0$. During the data-preprocessing stage, the Bicubic interpolation method is intentionally used to process low-quality original underwater images. Compared to the general resize method, this approach can more effectively maintain the details and clarity of the image, significantly reducing distortion and aliasing effects. The implementation of this preprocessing strategy aims to enhance the model's ability to perform super-resolution on low-resolution images, thereby further improving the results of underwater image enhancement.

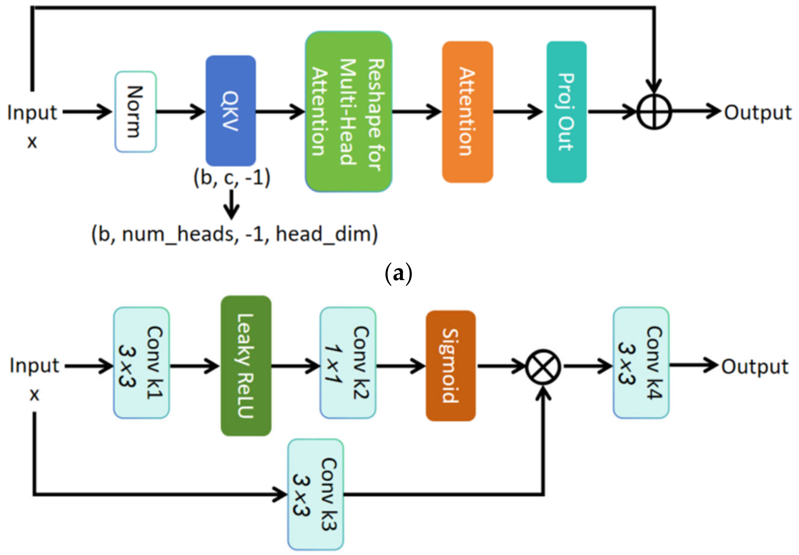
The approach employed in this study is based on an improved denoising probabilistic model, which holds more distinct advantages compared to the underwater image enhancement methods based on physical models and deep learning mentioned in Section 2.1. Physical model methods are prone to being influenced by changes in imaging conditions, while GAN methods based on deep learning require a delicate balance between the generator and discriminator during the training process, making their training process less stable. In contrast, the DDPM boasts a simple loss design that provides sufficient theoretical support for training. Moreover, the Markov chain process formed during the DDPM training process means that the current state is only related to the previous moment, making the training process more stable. However, in practical applications, image degradation issues also seem to plague traditional denoising probabilistic models. To address this issue, this study introduces a dual-layer attention mechanism into the model, a strategy that considers both spatial and channel information, enabling it to better capture local and global image features. Additionally, the unique sub-pixel interpolation upsampling strategy of this study can significantly reduce image degradation during the upsampling phase.

3.1. Dual-Layer Attention Mechanism

Similar to the traditional DDPM [5], the network structure in this experiment employs a modified U-Net model as a crucial component of the generative network to address the dual challenges of underwater image enhancement and denoising. Given the outstanding performance of U-Net in image segmentation and other related tasks, specific modifications have been made to adapt it to the unique requirements of underwater image enhancement.

Due to the distinctive conditions of underwater environments, such as scattering and light attenuation, the network input is specially designed to concatenate x_0 and $\widetilde{x}_0$ along dimension 1 and embed timestep information to enhance the network's perception of the temporal dimension. Initially, the network processes the input data through a convolutional layer to extract fundamental feature information. This is followed by a series of residual blocks, each containing multiple convolutional layers, normalization operations, and residual connections. This structure not only aids in capturing more advanced features within the images but also mitigates the issue of gradient vanishing to some extent, thereby improving the training performance of the network. Additionally, the output channel numbers of each residual block are flexibly adjusted according to the specific parameters of the network design, meeting the needs of different levels of feature extraction. During the downsampling phase, a single-layer attention module is introduced, which helps the model focus more on the key information within the image, enhancing attention to specific regions. Conversely, the upsampling block employs a structure symmetrical to the downsampling block, gradually enlarging the feature maps to restore the processed features to the same dimensions as the original input image, and outputs the enhanced or denoised result.

The middle block, as a crucial component connecting the encoder and decoder in the U-Net structure, receives feature maps output from the encoder as input. Through a series of processing operations, it extracts more advanced and abstract feature information. This provides valuable information for the decoder to reconstruct the original input more accurately. During the processing of the input feature maps, an initial integration of features takes place, merging low-level and high-level features, which not only aids in improving feature representation but also promotes gradient flow during the training process. Due to the uniqueness of underwater image enhancement, special attention needs to be paid to the dependencies between image features and multi-channel color spaces. Thus, a dual-layer attention mechanism is introduced in the middle block. This dual-layer attention mechanism consists of two key components: the Attention Block and the Attention Branch. Figure 2a demonstrates the structure of a multi-head attention mechanism. The input, x, first undergoes normalization and then generates queries, keys, and values. These generated QKVs are reshaped to fit the multi-head attention computation. After being processed by the attention layer, the results are passed through the projection output layer and added to the residual connection, ultimately producing an output feature map with the same shape as the input. Figure 2b demonstrates a convolutional neural network branch structure with an attention mechanism. The input, x, first undergoes feature extraction through a 3 × 3 convolutional layer and then is processed by a Leaky ReLU activation function for non-linear transformation. Next, it passes through a 1 × 1 convolutional layer to change the number of channels and is then normalized by the sigmoid function to generate attention coefficients. The input, x, also passes through another 3 × 3 convolutional layer, and its output is element-wise multiplied with the attention coefficients. Finally, the result is further processed by another 3 × 3 convolutional layer, producing an output feature map with the same shape as the input. The Attention Block focuses on enhancing the model's feature-representation capability by performing adaptive spatial attention operations in the middle layers. The Attention Branch, based on the learned attention weights, selectively enhances relevant features in the input data, further boosting the model's perceptual abilities. This specialized dual-layer attention structure works collaboratively to capture complex relationships within the input data, effectively utilizing feature information across different scales, thereby achieving more accurate and comprehensive performance enhancements in underwater image enhancement tasks.



(**a**)

(**b**)

Figure 2. Structure of dual-layer attention mechanism. (**a**) The structure and data flow of Attention Block. (**b**) The structure and data flow of the Attention Branch.

3.2. Improved DDPM

The loss function of the traditional denoising probability model DDPM is based on the optimization of the negative log probability ELBO, expressed as follows:

$$L = \mathbb{E}_{\mathrm{q}}\left[DKL(q(x_T|x_0)||p(x_T)) + \sum_{t=2}^{T} DKL(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t)) - \log p_\theta(x_0|x_t)\right] \tag{7}$$

where $\mathbb{E}_{\mathrm{q}}$ denotes the expectation, DKL denotes the KL divergence, q is the probability density function of the reverse process, p is the probability density function of the forward process, x_t represents the state at time t, and T is the total number of time steps. To simplify calculations, the loss function is often simplified to $L_{simple} = \mathbb{E}_{t,x_0,\epsilon}[||\epsilon - \epsilon_\theta(x_t, t)||^2]$. However, Nichol et al. [17] pointed out that, in the loss function of DDPM [5], the variance term, $\sum_\theta (x_t, t) = \sigma^2 I$, is treated as a fixed constant rather than a learnable parameter, which actually limits the model's expressive capacity. Therefore, to enhance the model's flexibility and performance, a penalty term, L_{vlb}, is added when computing the loss, optimizing the variational lower bound, thereby making the variance a learnable term. When $t = 0$, the variational lower bound is the negative log likelihood, $L_0 := -\log p_\theta(x_0|x_1)$. When $t \neq 0$, the variational lower bound is the KL divergence between the predicted x_t and the true x_t: $L_{t-1} := D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))$. Thus, the variational lower bound can be represented as $L_{vlb} := L_0 + L_1 + \cdots + L_{T-1} + L_T$. Through this approach, a composite loss function is constructed, as shown in Equation (8).

$$L_{hybrid} = L_{simple} + \lambda L_{vlb} \tag{8}$$

In this framework, L_{simple} is responsible for capturing the main distributional characteristics of the data, while L_{vlb} serves as a penalty term, aiding the model in learning finer details and the intrinsic structure of the data. λ is the weight coefficient, and in IDDPM [17], $\lambda = 0.001$ is recommended. Compared to directly optimizing the log-likelihood, this hybrid learning objective offers significant advantages. It not only retains L_{simple} as the primary source of the loss, ensuring that the model can stably learn the main features of the data, but also enhances the model's fine perception of the data distribution through the introduction of L_{vlb}, thereby improving the convergence speed and effectiveness during model training.

When training the denoising probability model, the core objective is to find a function, f, that minimizes the discrepancy between the predicted image, $f(x_i)$, and the actual high-resolution image, y_i, while preventing the model from overfitting. This process can be described through the following optimization problem formulation: $\min_f \sum_{i=1}^{N} L(y_i, f(x_i)) + \mu R(f)$. Here, L represents the loss function, which measures the discrepancy between the model prediction, y_i, and the actual $f(x_i)$; $R(f)$ is a regularization term to prevent overfitting during training; and μ is a hyperparameter used to adjust the relative weight between the loss function and the regularization term, aiming to achieve an optimal balance between model performance and generalization ability.

3.3. Sub-Pixel Interpolation Upsampling Strategy

Historically, researchers often treated super-resolution tasks with different scaling factors as independent tasks, designing and training separate models for each scaling factor [18]. However, this strategy was not only inefficient but also limited the versatility and flexibility of the models. To overcome this limitation, Hu et al. [19] introduced the Meta-SR method, which enables super-resolution processing for arbitrary scaling factors, including non-integer values, through a single model. Inspired by this, Lim et al. [6] further enhanced super-resolution performance by scaling up the model size. Motivated by these studies, we explored the impact of different scaling factors on super-resolution tasks and discovered through experiments that models exhibit varying super-resolution effects when faced with different scaling factors.

Diffusion models fundamentally involve the precise manipulation of individual pixels in the original input image, aiming to minimize the discrepancy between the predicted and the real images. The relationship between the original noise size and the expected output shape is critical in the model's upsampling process. When the noise size is less than or equal to the expected output, the limited number of pixels restricts the insertion of details, making it difficult to display more intricacies. Conversely, when the noise size exceeds the expected output, the increased number of sub-pixels provides more room for detail insertion. However, an increase in sub-pixel quantity does not always result in better outcomes. Experiments show that interpolation at different scaling factors yields varied results. If an inappropriate interpolation method is chosen, interpolating among a large number of sub-pixels might prevent effective connectivity between adjacent pixels, leading to an overly smooth image lacking in edge information between instances. In such cases, the final generated image does not appear to have undergone effective super-resolution processing. Therefore, after balancing image details, clarity, and naturalness through experimental comparisons, we chose a 1.25× scaling factor for the upsampling size. This scaling factor effectively preserves image details, while avoiding excessive sharpening, achieving a good balance between clarity and naturalness.

Simultaneously, the interpolation method used in the upsampling process is crucial for the impact on image edge details. In the traditional U-Net structure handling super-resolution tasks, the upsampling block often utilizes nearest-neighbor interpolation. This method is simple and efficient, as it directly assigns the value of each pixel in the target image to the value of the closest pixel in the source image. However, a significant drawback of this method is that it can lead to jagged edges, particularly evident when images are enlarged. To overcome the deficiencies of nearest-neighbor interpolation, this paper opts for bilinear interpolation. Based on the concept of linear interpolation [20], bilinear interpolation calculates the weighted average of four neighboring pixels to determine the value of a new pixel. This method is more refined and better preserves the edge details of images. As illustrated in Figure 3, which shows a schematic of bilinear interpolation, assuming the values of four adjacent pixels are known, bilinear interpolation precisely calculates the value of the sub-pixel, $f(x, y)$.

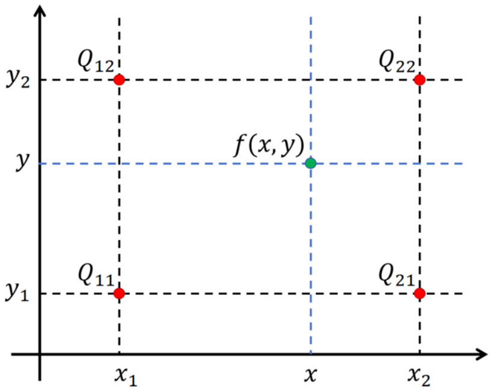


Figure 3. Bilinear interpolation diagram.

Initially, linear interpolation is performed in the x, resulting in the following two formulas:

$$f(x,y_1) \approx \frac{x-x_1}{x_2-x_1}[f(Q_{21})-f(Q_{11})]+f(Q_{11}) = \frac{x_2-x}{x_2-x_1}f(Q_{11})+\frac{x-x_1}{x_2-x_1}f(Q_{21}) \quad (9)$$

$$f(x,y_2) \approx \frac{x-x_1}{x_2-x_1}[f(Q_{22})-f(Q_{12})]+f(Q_{12}) = \frac{x_2-x}{x_2-x_1}f(Q_{12})+\frac{x-x_1}{x_2-x_1}f(Q_{22}) \quad (10)$$

Subsequently, linear interpolation is carried out in the y, resulting in the following:

$$f(x,y) \approx f(x,y_1)+\frac{y-y_1}{y_2-y_1}[f(x,y_2)-f(x,y_1)] = \frac{y_2-y}{y_2-y_1}f(x,y_1)+\frac{y-y_1}{y_2-y_1}f(x,y_2) \quad (11)$$

The expressions for $f(x,y_1)$ and $f(x,y_2)$ previously obtained are substituted into the above formulas to ultimately derive the following:

$$f(x,y) = w_{11}f(Q_{11})+w_{21}f(Q_{21})+w_{12}f(Q_{12})+w_{22}f(Q_{22}) \quad (12)$$

where w_{xy} represents the weight of each integer point, namely, the bilinear interpolation kernel.

Consequently, for any non-integer point, p, on the feature map, X, its value can be calculated using bilinear interpolation, expressed as $X(p) = \sum_q G(q,p) \cdot X(q)$. Here, the coordinates of the non-integer point, p, are (x,y), while q represents the four adjacent integer values, Q, on the feature map X. The bilinear interpolation kernel function, $G(q,p) = g(q_x,p_x) \cdot g(q_y,p_y)$, measures the correlation between the integer point, q, and the non-integer point, p. The function $g(a,b) = \max(0.1 - |a-b|)$ represents the distance weight between the diagonal integer points and the target point, p, ensuring that the interpolation process considers only adjacent integer points and that the weights decrease as the distance increases.

4. Experiments

4.1. Experimental Settings

- Datasets:

For a long time, obtaining genuine and paired underwater image datasets has been a major challenge in the field of underwater image research. The imaging process of underwater images is influenced by multiple factors, such as water absorption and attenuation of different wavelengths of light, target distance, and spectral distribution, making data collection particularly complex. Although the use of artificial light sources can increase the underwater visibility to some extent, it often leads to irregular light spots in images, further complicating data processing [21]. In light of this, this study selected the EUVP underwater image dataset released in 2020 [22] as the research foundation. This dataset not only contains paired image samples of poor and good perceptual quality but also boasts a rich source of data, with poor-quality underwater images collected by seven different cameras during oceanic expeditions. To ensure the quality and validity of the data, this experiment prepared these paired datasets following the steps recommended in [23]. In this study, the Underwater Scenes paired dataset from the EUVP dataset was primarily used as the training set, which includes 2185 training image pairs, 139 validation image pairs, and 515 test samples. Additionally, to further demonstrate the model's generalizability, 120 paired test samples from the UFO-120 dataset were also used for validation.

- Evaluation metrics:

In underwater image enhancement tasks, a series of full-reference metrics are commonly used to assess the effectiveness of the enhancement, including Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [24]. PSNR is a metric that measures the signal-to-noise ratio between the enhanced image and the original image, primarily used to evaluate the quality of image reconstruction, typically expressed in decibels (dB). For images I^{HR} and I^{SR} of size m × n, the Mean Squared Error (MSE) is first calculated, defined as $MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I^{HR}(i,j) - I^{SR}(i,j)\right]^2$. Subsequently, PSNR can be calculated based on MSE, with the formula $PSNR = 10 \cdot \log_{10}\left(\frac{MSE_I^2}{MSE}\right)$, where MSE_I^2 represents the maximum possible pixel value of the image.

SSIM is a metric that measures the structural similarity between the enhanced image and the original image, considering the similarity in luminance, l; contrast, c; and structure, s. Similar to PSNR, SSIM is also calculated based on three comparative dimensions between images I^{HR} and I^{SR}, with the formula $SSIM\left(I^{HR}, I^{SR}\right) = \left[l\left(I^{HR}, I^{SR}\right) \cdot \alpha\right] \cdot \left[c\left(I^{HR}, I^{SR}\right) \cdot \beta\right] \cdot \left[s\left(I^{HR}, I^{SR}\right) \cdot \gamma\right]$.

When using these evaluation metrics, the processed images must be compared with their ground truth.

Since the high-quality images in the currently used dataset were synthesized through a series of image-processing techniques, in addition to calculating the difference between the generated images and the high-quality images using the above metrics, we also calculated the Underwater Image Quality Metric (UIQM) for each generated image. The UIQM is a no-reference metric designed based on the principles of human visual system excitation. This metric aims to comprehensively evaluate the color, clarity, and contrast of the image. Specifically, UIQM combines the Color Measurement Index (UICM), the Sharpness Measurement Index (UISM), and the Contrast Measurement Index (UIConM) through a linear combination of these sub-metrics to arrive at the final quality evaluation. The formula used is $UIQM = c_1 \cdot UICM + c_2 \cdot UISM + c_3 \cdot UIConM$. This study employed the same UIQM calculation method and sub-metric weights as [22] to ensure the accuracy and consistency of the evaluation results.

- Implementation details:

This experiment was conducted using the Pytorch 1.8 [25] framework to build and implement the underwater image enhancement model. During the model training phase, the input image size was set to 128 × 128 pixels. For low-resolution images, they were first resized to 64 × 64 pixels using the resize function and then enlarged back to 128 × 128 pixels to meet the model's input requirements. Additionally, the batch size was set to 16, with an initial learning rate of 0.0001, and the plan was to conduct training over 100,000 epochs. The AdamW optimizer [26] was chosen for optimizing the neural network parameters. In the settings for the diffusion model, the diffusion steps were set at 100, and a linear noise schedule type was chosen to ensure that the model could learn the intrinsic structure of the image data. Within the U-Net structure, a time embedding module was introduced to enhance the model's ability to process temporal information. SiLu was selected as the activation function to improve the model's non-linear expressive capability and feature learning capacity. The experimental setup utilized a computer equipped with a 16 vCPU Intel(R) Xeon(R) Gold 6430 CPU and an RTX 4090 (24GB) GPU for training and testing.

4.2. Comparison with the State of the Art

This experiment considered three commonly used evaluation metrics in the field of underwater image processing: PSNR, SSIM, and UIQM. To verify the performance of the proposed method, we compared various underwater image enhancement approaches on the EUVP test set, including adversarial-based UGAN-P [23], CycleGAN [15], FunIE-GAN [22], TACL [27], physics-based Uw-HL [28], transformer-based U-shape [29], and another diffusion model-based DM_uw [30], totaling seven different methods. The experimental results, as shown in Table 1, clearly demonstrate that the method proposed in this paper achieved improvements in both PSNR and UIQM, especially showing significant enhancement in the SSIM metric. Notably, this method does not only excel in a single metric but provides balanced and outstanding performance across all three metrics.

Table 1. Quantitative comparison of EUVP datasets.

Method	PSNR	SSIM	UIQM (Input: 2.57)
UGAN-P (2018) [23]	19.59	0.6685	2.72
CycleGAN (2017) [15]	17.14	0.6400	2.44
FunIE-GAN (2020) [22]	21.92	0.8876	2.78
Uw-HL (2020) [28]	18.85	0.7722	2.62
TACL (2022) [27]	18.92	0.8699	3.01
U-shape (2023) [29]	25.03	0.9045	3.01
DM_uw (2023) [30]	19.97	0.8852	2.97
DL-UW (ours)	25.48	0.9431	3.03

Additionally, to further validate the generalizability and adaptability of the proposed method, we conducted comprehensive comparative experiments on the UFO-120 test set, involving multiple methods, such as FunIE-GAN [22], TACL [27], U-shape [29], and DM_uw [30]. The experimental results, presented in Table 2, show that the method not only excelled in the key metrics of PSNR, SSIM, and UIQM but also demonstrated superior generalization capabilities in comparison with other methods.

Table 2. Quantitative comparison of UFO-120 datasets.

Method	PSNR	SSIM	UIQM (Input: 2.64)
FunIE-GAN (2020) [22]	24.09	0.9314	3.05
TACL (2022) [27]	18.81	0.8687	3.01
U-shape (2023) [29]	22.83	0.9231	3.06
DM_uw (2023) [30]	25.08	0.9436	2.93
DL-UW (ours)	25.26	0.9468	3.01

4.3. Qualitative Analysis

During the testing phase, low-quality images were used as the sole input to the model to evaluate its ability to enhance image details and correct colors. As shown in Figure 4a,b, the model proposed in this paper demonstrated an outstanding performance, effectively removing the whitening, graying, and bluing phenomena common in underwater images, while significantly improving color restoration. Notably, the model was also capable of restoring the glossiness of objects. Additionally, our model excelled in detail restoration. As illustrated in Figure 4c,d, for the complex and varied linear and arcuate textures in underwater images, the model accurately captured and restored these details, achieving a visually perceptible super-resolution effect.

The experiment further compared the performance of other underwater image enhancement methods in practical applications, including those based on diffusion models such as IDDPM [17] and DM_uw [30], as well as FunIE-GAN [22] and TACL [27]. To ensure a fair comparison, a representative set of underwater images was selected as a uniform input. As shown in Figure 5, we found that some models had deficiencies in color restoration, specifically manifesting as color oversaturation or insufficient image brightness.

Moreover, in the critical sub-task of image super-resolution, the performance of some methods was less than satisfactory, struggling to simultaneously address the important goals of image super-resolution and color restoration. Among them, the DM_uw method based on skip sampling did not include super-resolution within the scope of improvement, and the TACL method based on GAN showed significant deficiencies in color-restoration capabilities. In contrast, although the method proposed in this paper still has some gaps in achieving high-definition super-resolution effects, overall, its performance in underwater image enhancement tasks remains competitive.

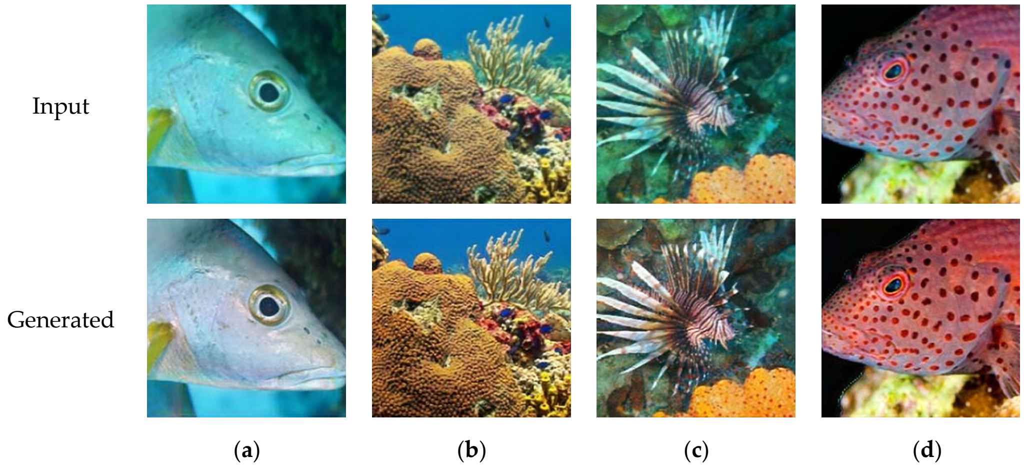


Figure 4. The enhancement effect of our model on underwater images. Among them, (**a**,**b**) significantly improve the color richness of the original image, and (**c**,**d**) improve the image resolution.

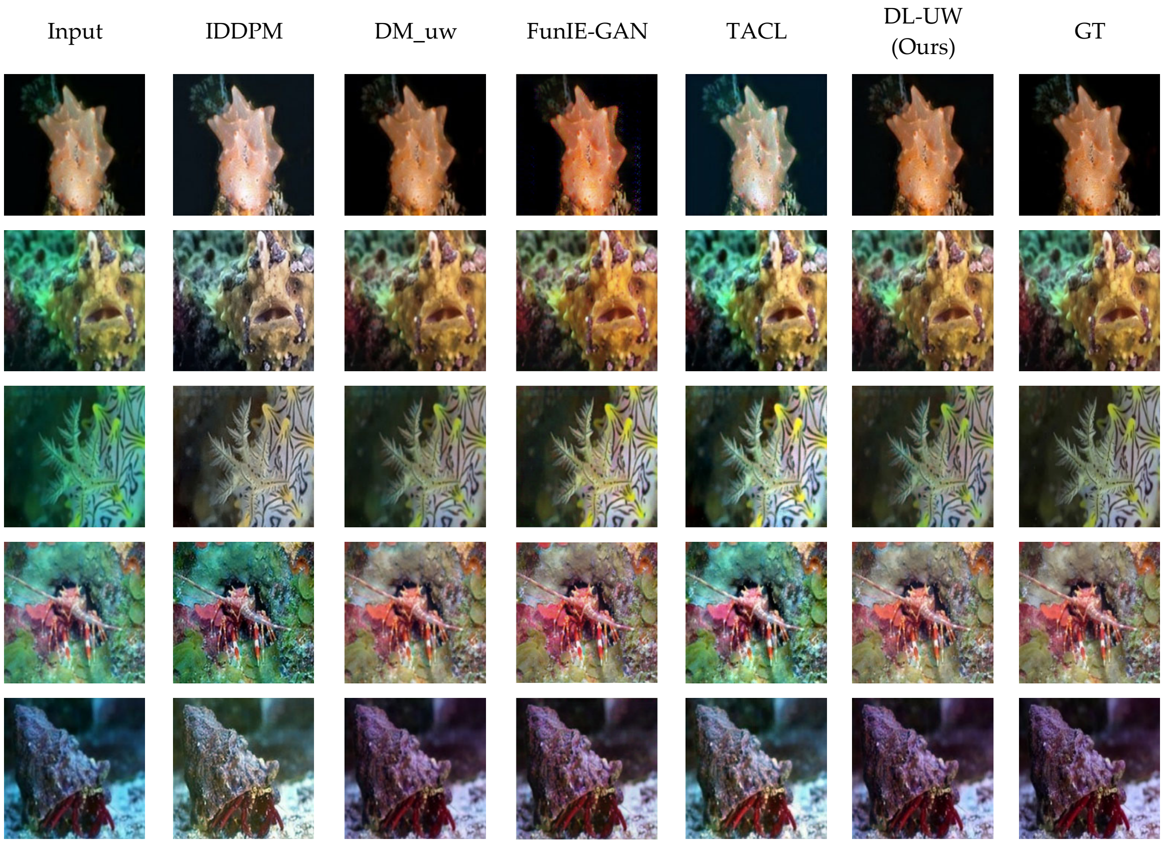


Figure 5. Comparison of underwater image enhancement effects by different methods. The basic facts are shown in the last column.

4.4. Quantitative Analysis

To comprehensively assess the performance of the proposed dual-layer attention mechanism model, we conducted quantitative analyses on the test sets of the EUVP and UFO-120 datasets. Comparative experiments revealed that the dual-layer attention mechanism model demonstrated significant advantages in the key metrics of PSNR and SSIM, showing marked improvements over the baseline model, which uses a single-layer attention mechanism. This result fully substantiates the effectiveness of the dual-layer attention mechanism in enhancing the quality of underwater images. In terms of the UIQM metric, our model also showed improvements on the UFO-120 dataset, but the differences were not significant on the EUVP dataset. This may be due to variations in image characteristics and quality assessment standards between the two datasets. Specific numerical comparison results are detailed in Table 3, where the best results are highlighted in red.

Table 3. Comparison experiment with baseline model using double-layer attention mechanism.

Dataset	Method	PSNR	SSIM	UIQM (Input: 2.58)
EUVP	Baseline model	19.28	0.8744	3.04
	DL-UW (ours)	25.48	0.9431	3.03
UFO-120	Baseline model	18.61	0.8526	2.92
	DL-UW (ours)	25.26	0.9468	3.01

Furthermore, to delve deeper into the effects of interpolation methods and scaling factors on underwater image enhancement, we specifically selected 20 random images from the EUVP test set to create a dedicated test set for quantitative experiments on interpolation methods and scaling factors. When the scaling factor was set to 1.00× and the nearest neighbor interpolation method was used, the generated images achieved a PSNR value of 22.81 and an SSIM value of 0.9008. In this experiment, a scaling factor of 1.25× was set as the target divisor scaling factor, and the performance of three different interpolation methods—nearest neighbor, bicubic, and bilinear interpolation—during the upsampling phase was investigated. As shown in Table 4, compared to the other two interpolation methods, bilinear interpolation exhibited significant improvements across the key evaluation metrics of PSNR, SSIM, and UIQM. This finding not only validates the effectiveness of bilinear interpolation in underwater image enhancement tasks but also provides valuable insights for subsequent research.

Table 4. Comparison results of different interpolation methods.

Method	PSNR	SSIM	UIQM (Input: 2.58)
Nearest	27.44	0.9612	2.84
Bicubic	27.76	0.9641	2.86
Bilinear	28.15	0.9663	2.89

Therefore, to determine the optimal upsampling interpolation method and scaling factor, we conducted a series of experiments. After the comparative analysis, we selected bilinear interpolation as the interpolation method for the upsampling process and studied different scaling factors as the sole variable. As illustrated in Figure 6, we observed that both PSNR and SSIM metrics reached their highest values when the scaling factor was set to 1.25×, while the UIQM metric performed best at a scaling factor of 1.375×. However, in comparison, the improvement in UIQM values is not significant, as the UIQM metric takes into account three aspects of an image: color, sharpness, and contrast. While blurring reduces the sharpness of an image, the overall image quality is still influenced by other factors. Nevertheless, as the scale factor increases, the trend of UIQM value changes is roughly the same as that of PSNR and SSIM. Further observation in Figure 7 revealed that when the scaling factor was set to 1.00×, the generated images exhibited severe sharpening effects, leading to lower levels in three evaluation metrics. However, when the scaling factor was increased to 1.25, the overall quality of the images achieved a balanced state, with better handling of detail retention and color restoration. Nevertheless, as the scaling factor was further increased, the images began to show signs of excessive smoothness. This occurs because, as the number of sub-pixels between features increases, the interpolation method starts adopting a "conservative" approach to feature restoration, trying to evenly fill the sub-pixels between adjacent features, often resulting in the loss of image details.

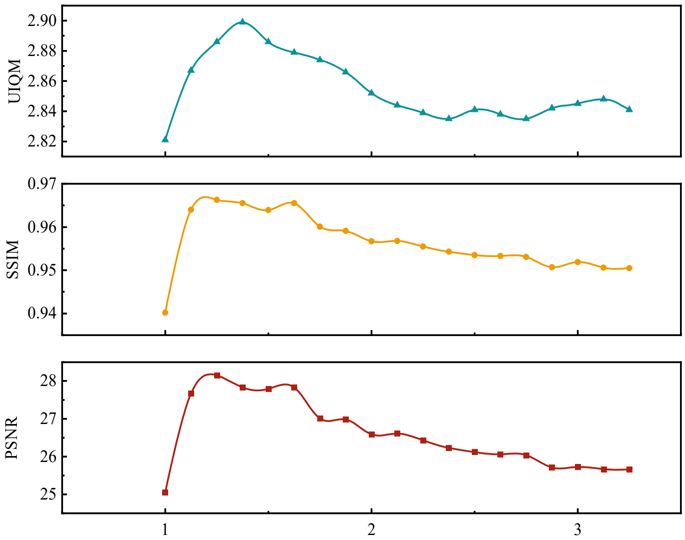
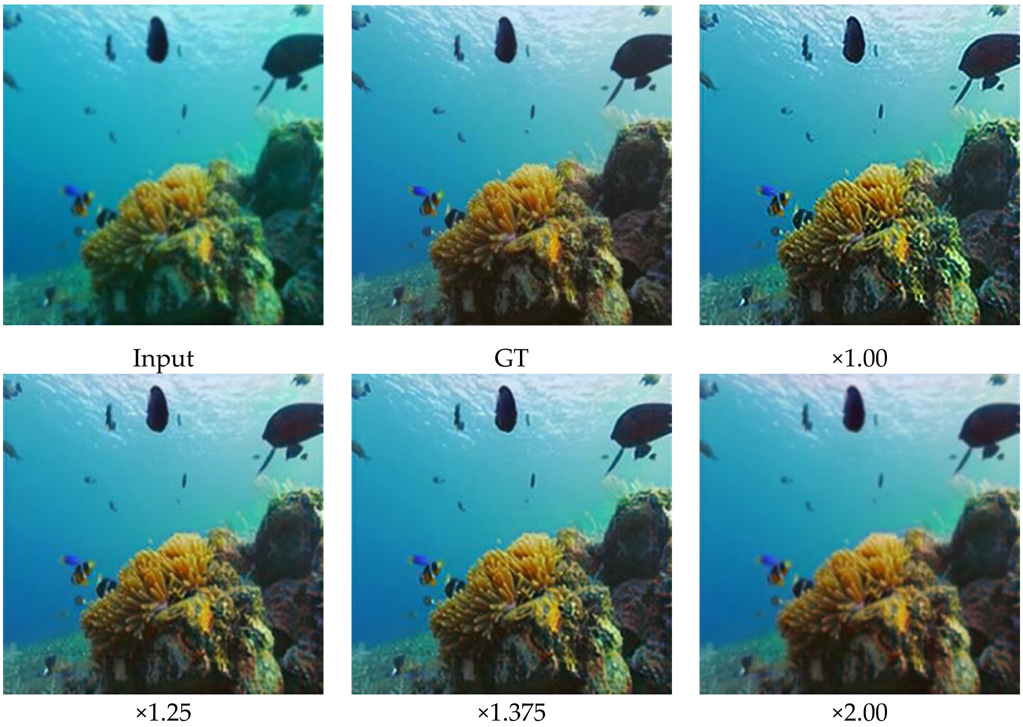


Figure 6. In the case of using different scale factors, PSNR, SSIM, and PSNR evaluation indexes are generated.



Input GT ×1.00

×1.25 ×1.375 ×2.00

Figure 7. When 256 × 256 underwater images are fed to the model, enhanced images are generated using different scale factors.

In summary, the following conclusion can be drawn: choosing an appropriate scaling factor is crucial for enhancing the effectiveness of underwater image enhancement when using bilinear interpolation for upsampling. In this experiment, we found that scaling factors of 1.25× and 1.375× showed better performance in terms of PSNR, SSIM, and UIQM metrics.

5. Conclusions

This paper introduces an innovative underwater image enhancement model based on a denoising neural network using probabilistic diffusion, combined with a composite loss diffusion model for underwater image enhancement applications. The proposed model effectively generates enhanced images from low-quality underwater inputs, supporting both single and batch image processing. The experimental results demonstrate that the model excels in the core sub-tasks of underwater image super-resolution and color correction. The dual-layer attention mechanism introduced in this study plays a crucial role in color correction, outperforming traditional single-layer attention mechanisms in handling complex color variations in underwater images, significantly enhancing color restoration and detail retention. Furthermore, we explored the impact of sub-pixel completion strategies during the upsampling process on super-resolution outcomes. Our experiments confirm that combining bilinear interpolation with a 1.25 scaling factor significantly mitigates the overly "conservative" issue of sub-pixel completion strategies, enabling our model to achieve super-resolution effects more rapidly and thereby enhancing the visual quality of underwater images.

Looking forward, we aim to explore the possibility of jointly processing multiple scales of super-resolution tasks within a single network, based on the research presented in this paper. This would contribute to the development of a more efficient and flexible underwater image enhancement system, providing stronger technical support for practical applications. We believe that through continuous research and innovation, we can advance underwater image enhancement technologies, contributing further to the research and applications in related fields.

Author Contributions: Conceptualization, H.Z.; methodology, R.H.; software, R.H.; writing—original draft preparation, R.H.; writing—review and editing, W.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors would like to acknowledge the anonymous reviewers and editors whose thoughtful comments helped to improve this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jin, X. The development of research in marine geophysics and acoustic technology for submarine exploration. *Prog. Geophys.* **2007**, *22*, 1243–1249.
2. Guo, J.C.; Li, C.Y. Research progress of underwater image enhancement and restoration methods. *J. Image Graph.* **2017**, *22*, 273–287.
3. Liu, F.; Sun, S.J.; Han, P.L. Development of Underwater Polarization Imaging Technology. *Laser Optoelectron. Prog.* **2021**, *58*, 0600001.
4. Yu, W.Y.; Chen, X.G. Study on Image Fusion Algorithm Based on Wavelet Transform. *Trans. Beijing Inst. Technol.* **2014**, *34*, 1262–1266.
5. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840–6851.
6. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.
7. Chiang, J.Y.; Chen, Y.-C. Underwater Image Enhancement by Wavelength Compensation and Dehazing. *IEEE Trans. Image Process.* **2011**, *21*, 1756–1769. [CrossRef]
8. Trucco, E.; Olmos-Antillon, A.T. Self-Tuning Underwater Image Restoration. *IEEE J. Ocean. Eng.* **2006**, *31*, 511–519. [CrossRef]

9. McGlamery, B.L. A Computer Model for Underwater Camera Systems. In Proceedings of the Ocean Optics VI, Monterey, CA, USA, 23–25 October 1979; Volume 208, pp. 221–231.
10. Jung, S.-W. Enhancement of Image and Depth Map Using Adaptive Joint Trilateral Filter. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *23*, 258–269. [CrossRef]
11. Wang, S.; Zheng, J.; Hu, H.-M.; Li, B. Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images. *IEEE Trans. Image Process.* **2013**, *22*, 3538–3548. [CrossRef]
12. Kan, L.; Yu, J.; Yang, Y.; Liu, H.; Wang, J. Color Correction of Underwater Images Using Spectral Data. In Proceedings of the Optoelectronic Imaging and Multimedia Technology III, Beijing, China, 9–11 October 2014; Volume 9273, pp. 48–54.
13. Li, C.; Guo, J. Underwater Image Enhancement by Dehazing and Color Correction. *J. Electron. Imaging* **2015**, *24*, 033023. [CrossRef]
14. Wang, Y.; Zhang, J.; Cao, Y.; Wang, Z. A Deep CNN Method for Underwater Image Enhancement. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 1382–1386.
15. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
16. Li, J.; Skinner, K.A.; Eustice, R.M.; Johnson-Roberson, M. WaterGAN: Unsupervised Generative Network to Enable Real-Time Color Correction of Monocular Underwater Images. *IEEE Robot. Autom. Lett.* **2017**, *3*, 387–394. [CrossRef]
17. Nichol, A.Q.; Dhariwal, P. Improved Denoising Diffusion Probabilistic Models. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 8162–8171.
18. Shi, W.; Caballero, J. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.
19. Hu, X.; Mu, H.; Zhang, X.; Wang, Z.; Tan, T.; Sun, J. Meta-SR: A Magnification-Arbitrary Network for Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1575–1584.
20. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable Convolutional Networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 764–773.
21. Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; Tao, D. An Underwater Image Enhancement Benchmark Dataset and Beyond. *IEEE Trans. Image Process.* **2019**, *29*, 4376–4389. [CrossRef] [PubMed]
22. Islam, M.J.; Xia, Y.; Sattar, J. Fast Underwater Image Enhancement for Improved Visual Perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [CrossRef]
23. Fabbri, C.; Islam, M.J.; Sattar, J. Enhancing Underwater Imagery Using Generative Adversarial Networks. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; pp. 7159–7165.
24. Hore, A.; Ziou, D. Image Quality Metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.
25. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Volume 32.
26. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101.
27. Liu, R.; Jiang, Z.; Yang, S.; Fan, X. Twin Adversarial Contrastive Learning for Underwater Image Enhancement and Beyond. *IEEE Trans. Image Process.* **2022**, *31*, 4922–4936. [CrossRef] [PubMed]
28. Berman, D.; Levy, D.; Avidan, S.; Treibitz, T. Underwater Single Image Color Restoration Using Haze-Lines and a New Quantitative Dataset. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 2822–2837. [CrossRef] [PubMed]
29. Peng, L.; Zhu, C.; Bian, L. U-Shape Transformer for Underwater Image Enhancement. *IEEE Trans. Image Process.* **2023**, *32*, 3066–3079. [CrossRef] [PubMed]
30. Tang, Y.; Kawasaki, H.; Iwaguchi, T. Underwater Image Enhancement by Transformer-Based Diffusion Model with Non-Uniform Sampling for Skip Strategy. In Proceedings of the 31st ACM International Conference on Multimedia, Ottawa, ON, Canada, 29 October–3 November 2023; pp. 5419–5427.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.