

## Article

# LM-CycleGAN: Improving Underwater Image Quality Through Learned Perceptual Image Patch Similarity and Multi-Scale Adaptive Fusion Attention

Jiangyan Wu <sup>1,2</sup>, Guanghui Zhang <sup>1,2,\*</sup> and Yugang Fan <sup>1,2</sup>

<sup>1</sup> Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China

<sup>2</sup> Yunnan Key Laboratory of Intelligent Control and Application, Kunming University of Science and Technology, Kunming 650500, China

\* Correspondence: ghzhang@kust.edu.cn

**Abstract:** The underwater imaging process is often hindered by high noise levels, blurring, and color distortion due to light scattering, absorption, and suspended particles in the water. To address the challenges of image enhancement in complex underwater environments, this paper proposes an underwater image color correction and detail enhancement model based on an improved Cycle-consistent Generative Adversarial Network (CycleGAN), named LPIPS-MAFA CycleGAN (LM-CycleGAN). The model integrates a Multi-scale Adaptive Fusion Attention (MAFA) mechanism into the generator architecture to enhance its ability to perceive image details. At the same time, the Learned Perceptual Image Patch Similarity (LPIPS) is introduced into the loss function to make the training process more focused on the structural information of the image. Experiments conducted on the public datasets UIEB and EUVP demonstrate that LM-CycleGAN achieves significant improvements in Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), Average Gradient (AG), Underwater Color Image Quality Evaluation (UCIQE), and Underwater Image Quality Measure (UIQM). Moreover, the model excels in color correction and fidelity, successfully avoiding issues such as red checkerboard artifacts and blurred edge details commonly observed in reconstructed images generated by traditional CycleGAN approaches.



**Citation:** Wu, J.; Zhang, G.; Fan, Y.

LM-CycleGAN: Improving Underwater Image Quality Through Learned Perceptual Image Patch Similarity and Multi-Scale Adaptive Fusion Attention. *Sensors* **2024**, *24*, 7425. <https://doi.org/10.3390/s24237425>

Academic Editor: Jan Cornelis

Received: 26 September 2024

Revised: 6 November 2024

Accepted: 18 November 2024

Published: 21 November 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** underwater image enhancement; cycle-consistent generative adversarial networks; multi-scale adaptive fusion attention; learned perceptual image patch similarity

## 1. Introduction

Underwater imaging plays a crucial role across various fields, including marine ecology research, underwater archaeology, underwater engineering, and marine resource exploration [1–4]. However, underwater images often degrade during acquisition and transmission due to multiple factors [5]. For instance, the optical properties of water and suspended particles can cause light scattering and absorption, reducing contrast and color accuracy in underwater images. Additionally, unstable conditions such as water currents and waves can affect the stability of image acquisition, further complicating underwater imaging. Therefore, Underwater Image Enhancement (UIE) is particularly important.

In recent years, traditional and deep learning-based methods have made significant advancements in the field of UIE. However, these methods still exhibit certain limitations when faced with complex underwater environments. On the one hand, physical model-based UIE methods may struggle to obtain sufficient prior information in challenging environments, leading to issues such as color distortion or overcompensation. On the other hand, non-physical model-based methods often rely on fixed parameter settings or heuristic rules, which tend to demonstrate poor generalization capabilities. Although deep learning-based methods have partially addressed these issues, they either fall short in

precise feature extraction, suffer from inadequate generalization performance, or typically require a substantial amount of labeled data for training. Therefore, the development of a UIE method that possesses strong generalization capabilities, high robustness, and efficient real-time performance is particularly critical. To this end, this paper proposes a novel model, LM-CycleGAN, with strong adaptability and superior enhancement. Specifically, a Multi-scale Adaptive Fusion Attention (MAFA) mechanism is designed and integrated into the generator, enhancing the generator's ability to perceive detailed image features. Additionally, the Learned Perceptual Image Patch Similarity (LPIPS) [6] is incorporated into the loss function, effectively introducing deep structural information about the image into the model's training process. In summary, the main contributions of this study can be outlined as follows:

A Multi-scale Adaptive Fusion Attention Mechanism was designed, enabling multi-scale adaptive fusion across different heads in multi-head attention, significantly enhancing the generator's capability to extract detailed image features.

The computation of LPIPS was optimized and integrated into the loss function of CycleGAN, improving the model's ability to assess the similarity between two images and enhancing the quality of image reconstruction results.

The proposed LM-CycleGAN model demonstrated superior performance on the UIEB [7], EUVP [8], and RUIE [9] datasets, validating its effectiveness in the field of UIE.

## 2. Related Work

Depending on the realization principle, underwater image enhancement methods can be mainly classified into two main categories: traditional methods and deep learning-based methods. Among them, the traditional methods are further categorized into physical model-based methods and non-physical model-based methods.

### 2.1. Physical Model-Based Methods

Physical model-based methods construct models by considering image degradation factors in specific environments, thereby reversing or compensating for degradation effects and improving image quality [10–12]. For instance, Liu et al. [11] developed a method for nighttime foggy images using a nonlinear and variational Retinex model to estimate illumination and reflectance, effectively removing haze and improving image quality. He et al. [12] introduced the Dark Channel Prior (DCP) algorithm, which clears haze by analyzing dark channel information, enhancing image clarity. Inspired by the successful application of the DCP in removing the "fog" effect from images, several studies [13–15] have applied similar principles to UIE. Chiang et al. [14] achieved color enhancement of underwater images by compensating for color channels through a model that accounts for the attenuation of light energy with depth. Similarly, the Underwater Dark Channel Prior (UDCP) algorithm [15] enhances underwater images by ignoring the red channel and relying on information from the blue-green channels. Although physical model-based UIE methods have achieved some success in improving degraded image quality, they tend to exhibit weak generalization performance when dealing with complex and variable underwater environments.

### 2.2. Non-Physical Model-Based Methods

Unlike physical model-based methods, non-physical model-based methods typically enhance underwater images by directly adjusting pixel values. For example, Garg et al. [16] applied the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm to enhance underwater images. Hu et al. [17] adopted an adaptive color correction method, the CLAHE algorithm, and multi-scale brightness fusion technology to enhance the details and contrast of images. Iqbal et al. introduced the UCM algorithm [18], which first enhances colors in the RGB color space and then adjusts contrast in the HSV color space to achieve image color correction and enhancement. However, methods relying on fixed parameters or heuristic rules often introduce issues such as overcompensation, color distortions,

tion, or insufficient detail processing in image reconstruction. Therefore, more adaptive and intelligent algorithms are needed to meet the challenges of UIE.

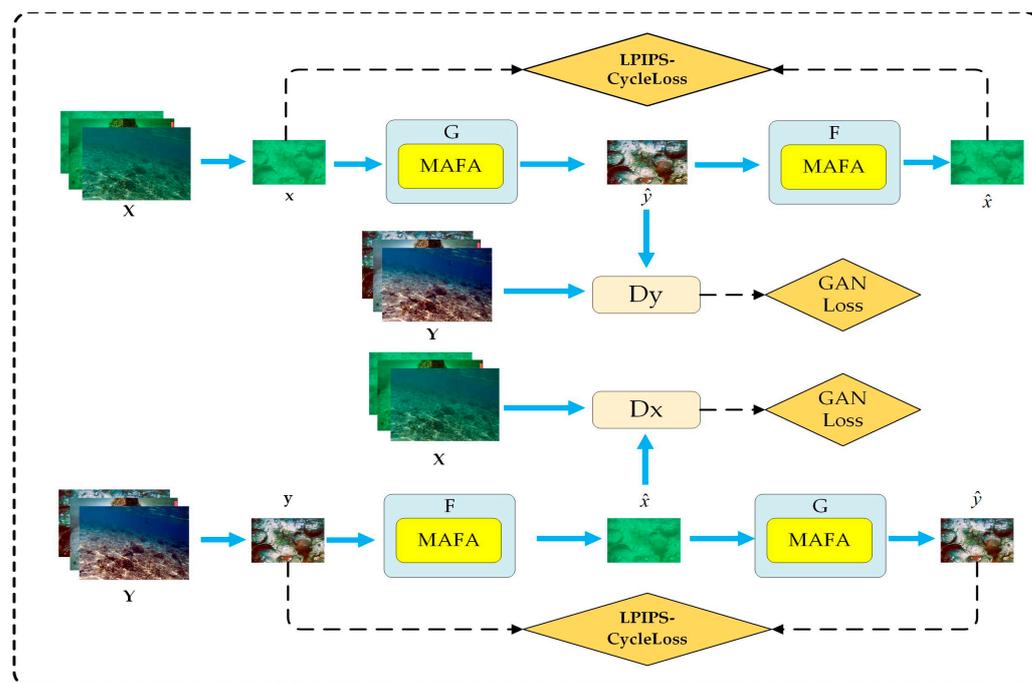
### 2.3. Deep Learning-Based Methods

With the rapid development of deep learning technologies, the field of image enhancement has seen new opportunities. Deep learning models, through large-scale training data, can learn deep features and structural information of images, making them more adaptive and robust in handling complex and dynamic environments. Among them, image enhancement techniques based on Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) [19] have demonstrated outstanding performance in various image-processing tasks. Among CNN-based image enhancement applications, Li et al. [20] proposed a CNN-based image enhancement model, named UWCNN, which utilizes a synthetic database for training and improves the visibility of underwater images through an end-to-end data-driven mechanism. Saleh et al. [21] introduced an unsupervised UIE framework (UDnet), which employs a conditional variational autoencoder combined with probability-adaptive instance normalization and a statistically guided multi-color space stretching method to generate realistic underwater images. On the other hand, in the application of image enhancement using GAN as well as Cycle Consistent Generative Adversarial Networks [22] (CycleGAN). Li et al. [23] developed an enhancement model based on CycleGAN (SSIMCycleGAN). This model calculates the Structural Similarity Index (SSIM) between degraded and generated images and integrates it into the loss function, thereby improving the contrast of the generated images. Another model, SESSCycleGAN [24] employs the Nobuyuki Otsu method (OTSU) [25] to extract edge images from both the degraded image and generated high-quality underwater images and constrains the model by calculating the L1 distance between the two edge images. In addition, Bakht et al. [26] integrate multi-level attention mechanisms within the GAN architecture (MuLAGAN), enhancing the model's ability to capture the details of underwater images. Cong et al. [27] proposed a physics model-guided GAN model (PUGAN) that adopts a dual-discriminator scheme to generate images that are both realistic and visually comfortable.

## 3. Materials and Methods

### 3.1. Architecture of LM-CycleGAN

As illustrated in Figure 1, the LM-CycleGAN model consists of two sets of generators and discriminators. Generators  $G$  and  $F$  share the same network architecture as the discriminators  $D_X$  and  $D_Y$ . Specifically, generator  $G$  transforms input images from the source domain  $X$  to the target domain  $Y$ , while generator  $F$  maps images from  $Y$  back to source domain  $X$ . In the discriminator module,  $D_X$  and  $D_Y$  are used to distinguish between real and generated images in domains  $X$  and  $Y$ , respectively. During the training process, a degraded image  $x$  is processed by generator  $G$  to produce a high-quality image  $\hat{y}$ . This image  $\hat{y}$  is then combined with a randomly sampled real image  $y$  from the domain  $Y$  and fed into discriminator  $D_Y$  for real-fake discrimination, generating the GAN Loss based on the discriminator's output. The objective of the generator  $G$  is to maximize this loss, while the discriminator aims to minimize it. This adversarial training encourages the generator to produce images that closely resemble real images. Subsequently,  $\hat{y}$  is passed through generator  $F$  to reconstruct the underwater degraded image  $\hat{x}$ . This reconstructed image is compared to the original  $x$  to compute the consistency loss. This process can be analogized to text translation tasks: Translate a sentence from Chinese to French, then back to Chinese. Compare the original sentence with the retranslated version to identify any differences. Enhance the performance of the translation system by minimizing these differences. Additionally, the literature [22] indicates that employing cycle consistency loss effectively mitigates the "mode collapse" problem, where the generator produces the same image no matter what the input image is.



**Figure 1.** Network structure of LM-CycleGAN.  $X$  denotes the underwater degraded image domain and  $Y$  denotes the underwater high-quality image domain.

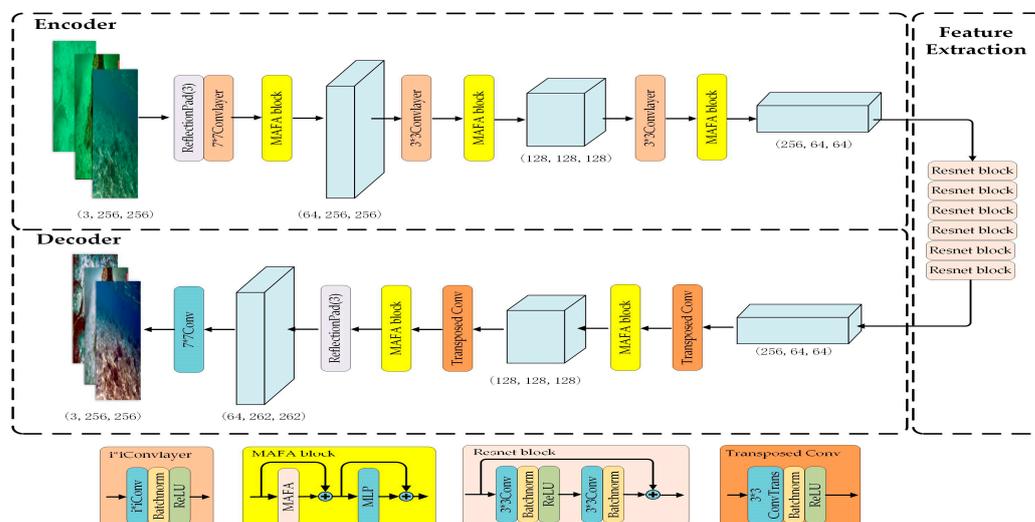
To enhance the model's ability to capture and restore image details, this study integrates the MAFA module into CycleGAN's generator. Additionally, to ensure consistency between the reconstructed and original images, the LPIPS–Cycle Loss is employed. The LPIPS–Cycle Loss ensures that the reconstructed images  $\hat{x} = F(G(x))$  and  $\hat{y} = G(F(y))$  are as consistent as possible with the original input images  $x$  and  $y$ , respectively. By incorporating the LPIPS Loss, the model captures semantic similarities between images more effectively, compensating for the limitations of traditional L1 loss in capturing high-level visual information, thus significantly improving the perceptual quality of the generated images.

### 3.2. Generator Structure Based on MAFA

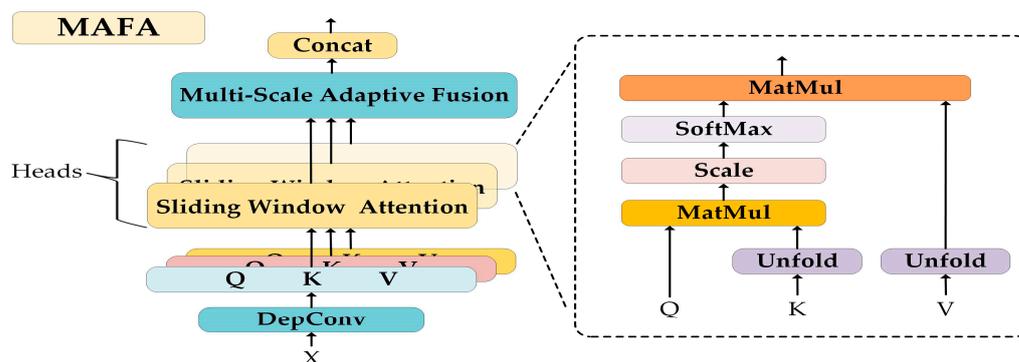
As illustrated in Figure 2, this study adopts the generator network architecture proposed by Johnson et al. [28], which consists of three convolutional layers, six residual blocks, three transposed convolutional layers, two ReflectionPad layers, and a convolutional layer that maps features to RGB. A key distinction of our approach is the incorporation of the efficient and robust feature extraction module MAFA into the backbone of the generator. The MAFA mechanism facilitates the adaptive fusion of multi-scale features, thereby effectively capturing details at various levels. This capability is particularly important for enhancing edge details in blurred underwater images.

Building upon the Multi-Scale Dilated Attention mechanism (MSDA) [29], this paper introduces a novel Multi-scale Adaptive Fusion Attention mechanism (MAFA). Specifically, the mechanism first utilizes depthwise separable convolutions (DepConv) to transform the input feature maps, generating Query (Q), Key (K), and Value (V) feature maps. Thereafter, these feature maps are divided into different “heads” along the channel dimension. Each “head” employs the ‘nn.Unfold’ operation from PyTorch to perform sliding window attention at varying dilation rates, thus capturing local information at multiple scales. Additionally, each “head” is assigned a learnable weight vector. Following independent attention computations for each head, the output feature maps are weighted by their respective weight vectors. Finally, all weighted feature maps are aggregated along the channel dimension to form a comprehensive feature map. The design of the MAFA mechanism

enables the model to dynamically adjust the importance of each head during training, effectively facilitating the adaptive fusion of features across different scales. Figure 3 illustrates the specific structure of the MAFA module.



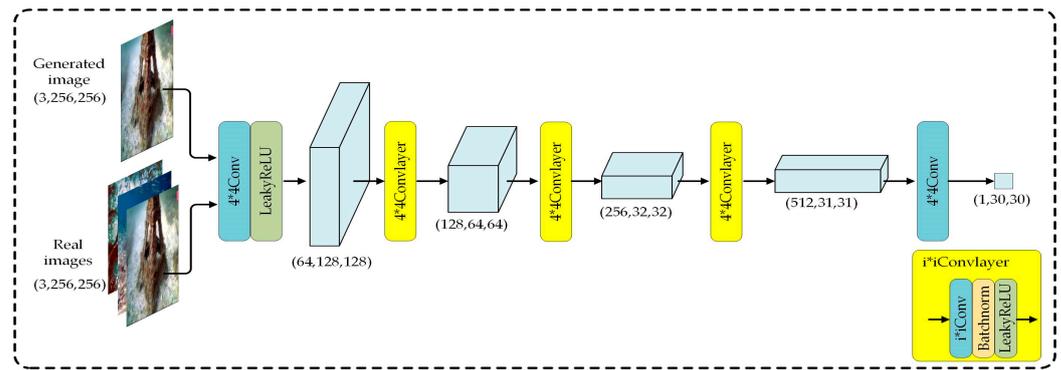
**Figure 2.** The network structure of the LM-CycleGAN generator, where the MLP is the Multi-Layer Perceptron, “n\*nConv” denotes an operation that involves processing with a single convolutional kernel, and “⊕” denotes element-wise addition.



**Figure 3.** Network structure of Multi-scale Adaptive Fusion Attention (MAFA), where the dilation rates are set to [1–4], and the number of ‘heads’ is set to 8.

### 3.3. Discriminator Network Structure

The network architecture of the LM-CycleGAN discriminator adopts the PatchGAN [30] structure, which is a fully convolutional network. Unlike traditional GAN discriminators that map the entire input image to a single probability value to determine whether it is real, PatchGAN employs a more localized and detailed evaluation strategy. It processes the input image through a fully convolutional network, ultimately producing an  $N \times N$  feature map, where each element corresponds to a small patch of the input image. The value of each element reflects the likelihood that the corresponding patch belongs to a real image. This approach enables PatchGAN to track and quantify the realism of specific regions in the image, influencing the overall discrimination decision. The architecture of the LM-CycleGAN discriminator network, as shown in Figure 4, comprises five convolutional layers.



**Figure 4.** Network structure of LM-CycleGAN discriminator. Assuming the input is an RGB image with dimensions  $3 \times 256 \times 256$  pixels, the final output will be a tensor of size  $1 \times 30 \times 30$ .

### 3.4. Loss Function

In LM-CycleGAN, the design of adversarial loss (GAN Loss) and cycle consistency loss (Cycle Loss) is consistent with the loss functions presented in CycleGAN [22]. To further enhance the model's performance, structural consistency loss (LPIPS Loss) has also been introduced in this study to improve the perceptual quality and detail preservation of the generated images.

#### (1) GAN Loss

The GAN Loss consists of two parts: the forward process, which generates a high-quality image from a low-quality underwater image, and the reverse process, which generates a low-quality image from a real high-quality underwater image. The formula for the GAN Loss is given in Equation (1):

$$GAN_{Loss} = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) \quad (1)$$

For the forward process, the GAN Loss function is expressed as

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim P_{data}(Y)} \left[ \left( D_Y(y) - 1 \right)^2 \right] + E_{x \sim P_{data}(X)} \left[ D_Y(G(x))^2 \right] \quad (2)$$

For the inverse process, the adversarial loss function is expressed as

$$(F, D_X, Y, X) = E_{x \sim P_{data}(X)} \left[ \left( D_X(x) - 1 \right)^2 \right] + E_{y \sim P_{data}(Y)} \left[ D_X(F(y))^2 \right] \quad (3)$$

#### (2) Cycle Loss

The Cycle Loss is designed to ensure that the image is successfully returned to the original domain after being transformed through the two generators, thus ensuring the stability and accuracy of the generated image. The loss function for this part is denoted as:

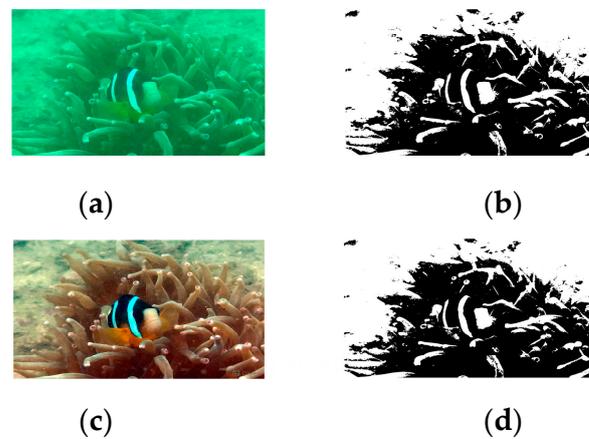
$$L_{cyc}(G, F) = E_{x \sim P_{data}(x)} [\|F(G(x)) - x\|_1] + E_{y \sim P_{data}(y)} [\|G(F(y)) - y\|_1] \quad (4)$$

where  $\|\cdot\|_1$  denotes the L1 norm, which is used to calculate the pixel-by-pixel difference between two images.

#### (3) LPIPS Loss

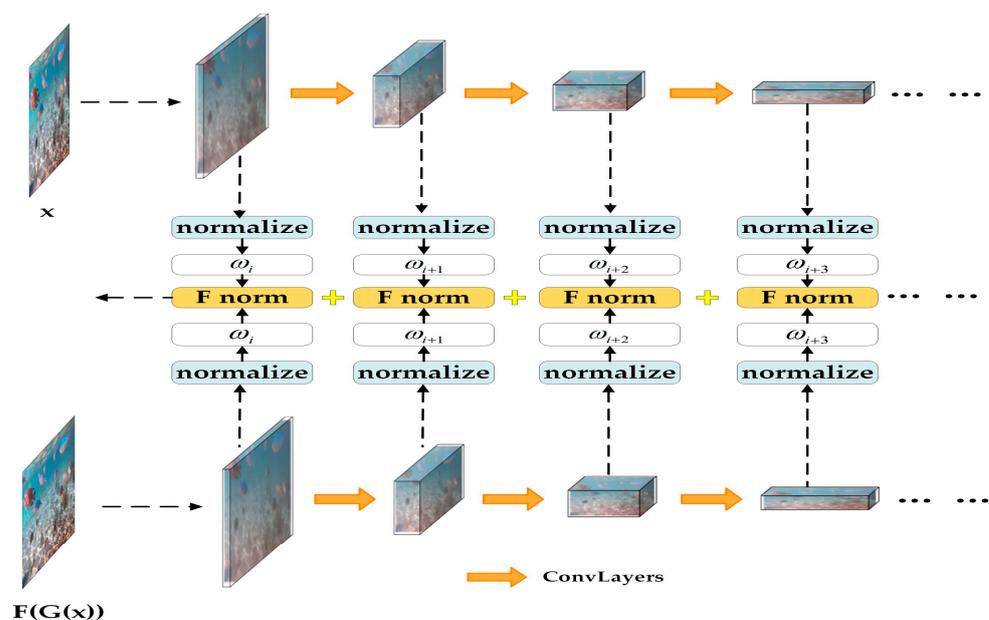
In CycleGAN, the Cycle Loss calculates the similarity between two images using the L1 distance, which computes pixel-wise differences without accounting for the structural information between the images. To address this limitation, SESS-CycleGAN introduces edge constraints during the generation process from the source domain to the target domain. This approach utilizes the OTSU algorithm to obtain the edge images of both the original and generated images with the expectation of minimizing the L1 distance between them.

The edge images corresponding to both the original and generated images are shown in Figure 5.



**Figure 5.** Original image and its edge image and generated image and its edge image: (a) Underwater degraded image; (b) Edge image corresponding to the underwater degraded image; (c) Generated underwater high-quality image; (d) Edge image corresponding to the underwater high-quality image.

By adopting this strategy, structural information can be integrated into the CycleGAN loss function, ensuring it does not negatively impact color correction. However, the OTSU algorithm is sensitive to image noise, and its performance may be affected when there is a significant size disparity between the target and background or when dealing with multi-target underwater images. For instance, in Figure 5b,d, the edges of the fish blend with those of other elements, such as coral, leading to information loss. Since the presence of fish and vegetation is common in underwater tasks, the current method still faces limitations when handling complex scenes. To more effectively incorporate structural information into the training process of CycleGAN, this paper integrates LPIPS into the loss function. The architecture of LPIPS is shown in Figure 6.



**Figure 6.** The network structure of Learned Perceptual Image Patch Similarity (LPIPS),  $\omega_i$  denotes the specific weight layer corresponding to the  $i$ th output layer.

The computation steps for the LPIPS [6] metric are as follows: First, the two images to be compared are passed through a pre-trained feature extraction network (such as AlexNet [31], SqueezeNet [32], or VGG [33]) to obtain feature representations. Then, the feature maps output by each layer are normalized along the channel dimension, with the normalized feature maps denoted as  $x_i^*$  and  $F(G(x))_i^*$ , respectively. Next, the feature map is scaled using a specific weight layer. Following this, the Frobenius norm between the scaled feature maps is computed and averaged over the spatial dimensions. Finally, the Frobenius norms from all layers are summed to provide a similarity score, where smaller values indicate higher similarity. This process can be expressed by the following equation:

$$LPIPS(x, F(G(x))) = \sum_i \frac{1}{H_i W_i} \|\omega_i \odot (x_i^* - F(G(x))_i^*)\|_F \quad (5)$$

Here,  $H_i$  and  $W_i$  represent the corresponding width and height of the  $i$ -th layer feature map,  $\omega_i$  represents the corresponding specific weight layer of the  $i$ -th layer, and  $\odot$  multiplies each feature map by the corresponding  $\omega_i$ . When denoting a matrix as  $A \in R^{m \times n}$ , the formula for calculating the Frobenius norm is as follows:

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} \quad (6)$$

In this paper, AlexNet is selected as the feature extraction network. During the forward pass of CycleGAN, the original image  $x$  and the generated image  $F(G(x))$  are input into the LPIPS network for similarity computation, with the generators  $G$  and  $F$  aiming to minimize Equation (5). This strategy introduces structural information during the model training process, further enhancing the performance of UIE. The loss function for this component is expressed as follows:

$$L_{lrips}(G, F) = LPIPS(x, F(G(x))) + LPIPS(y, G(F(y))) \quad (7)$$

(4) Total loss function of LM-CycleGAN

In summary, the total loss function of LM-CycleGAN is expressed as

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \alpha L_{cyc}(G, F) + \beta L_{lrips}(G, F) \quad (8)$$

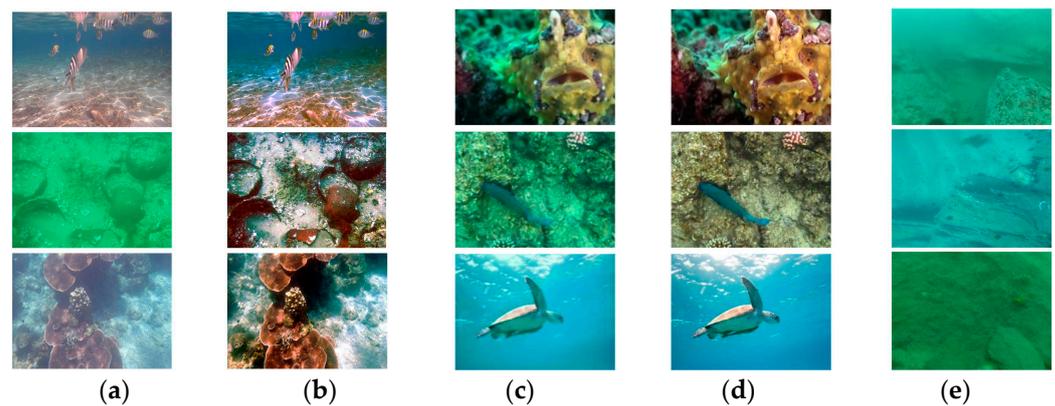
where ' $\alpha$ ' and ' $\beta$ ' represent the weight factors of Cycle Loss and LPIPS Loss, respectively. These factors are determined through hyperparameter tuning. In this paper, ' $\alpha$ ' is set to 10, and ' $\beta$ ' is set to 5.

#### 4. Experimental Results and Analysis

To comprehensively evaluate the proposed image enhancement model from multiple perspectives, Sections 4.4 and 4.5 of this paper present a comparative analysis between LM-CycleGAN and other UIE algorithms, including traditional algorithms such as UDCP [15], CLAHE [16], and UCM [18], as well as deep learning-based methods such as UWCNN [20], UDnet [21], CycleGAN [22], SSIM-CycleGAN [23], SESS-CycleGAN [24], MuLA-GAN [26], and PUGAN [27]. In addition, the ablation study and real-time performance analysis are conducted in Sections 4.6 and 4.7, respectively.

##### 4.1. Dataset Introduction

The UIEB dataset consists of 890 paired images divided into 800 training pairs and 90 test pairs. The EUVP test set consists of 100 paired images randomly selected from the EUVP-515 dataset. The RUIE test set consists of 90 challenging, no-reference, underwater degraded images characterized by images with blue, green, and blue-green casts. For a fair comparison, all models were trained on the UIEB dataset and tested on the UIEB, EUVP, and RUIE datasets (see Figure 7).



**Figure 7.** Sample images from the UIEB, EUVP, and RUIE datasets: (a,b) underwater degraded images and their corresponding reference images from the UIEB dataset; (c,d) underwater degraded images and their corresponding reference images from the EUVP dataset; (e) underwater degraded images from the RUIE dataset.

#### 4.2. Experimental Settings

The experiments were conducted in a Windows 11 environment using a server manufactured by VirtAITech (Shanghai, China), which is equipped with an NVIDIA GeForce RTX 4090 GPU with 24 GB of VRAM. The deep learning framework employed was PyTorch 2.1.0, along with Python 3.9, PyCharm 2023, and CUDA version 11.8. The network was trained for 300 epochs on the UIEBD datasets, with a batch size of 1. The Adam optimizer was utilized, with an initial learning rate of 0.0002 and an exponential decay rate that reduced the learning rate to 0 after 150 epochs.

#### 4.3. Evaluation Indicators

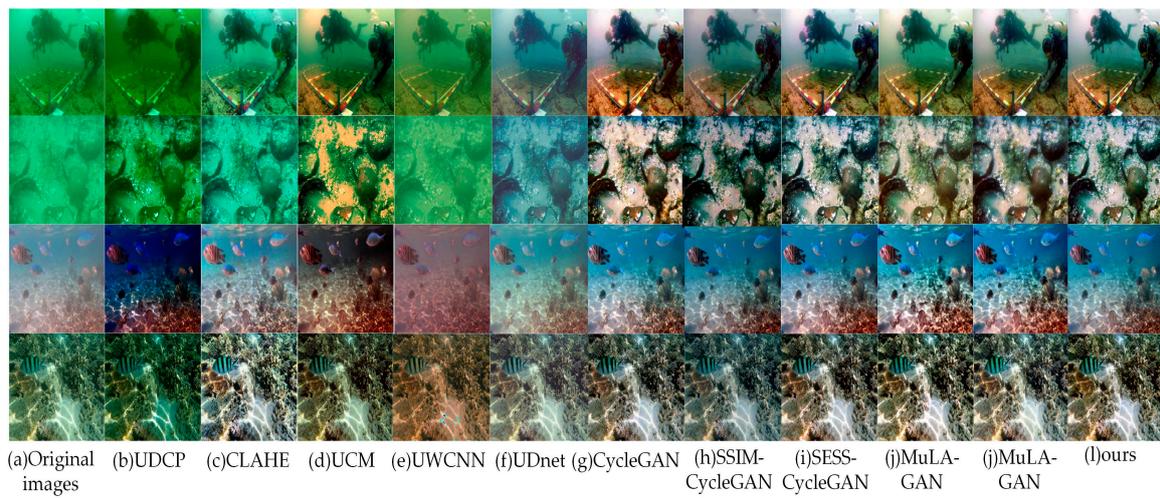
In the evaluation process, five commonly used metrics in UIE were employed to assess the quality of underwater reconstructed images: the Underwater Color Image Quality Evaluation (UCIQE) [34], the Underwater Image Quality Measurement (UIQM) [35], the Structural Similarity Index Measure (SSIM), Peak Signal-to-Noise Ratio (PSNR), and Average Gradient (AG).

UCIQE is a linear combination of chroma, contrast, and saturation values. UIQM is a linear combination of underwater image colorfulness measure, underwater image sharpness measure, and underwater image contrast measure. Higher scores for both UCIQE and UIQM indicate better performance in these aspects. SSIM quantifies the similarity between two images in terms of luminance, contrast, and structural information, with higher SSIM values indicating greater similarity. PSNR measures the ratio between the peak signal power and the noise power. Higher PSNR values indicate less image distortion and better quality. AG is calculated based on the mean gradient of an image. Higher AG values represent images with sharper edges and finer details.

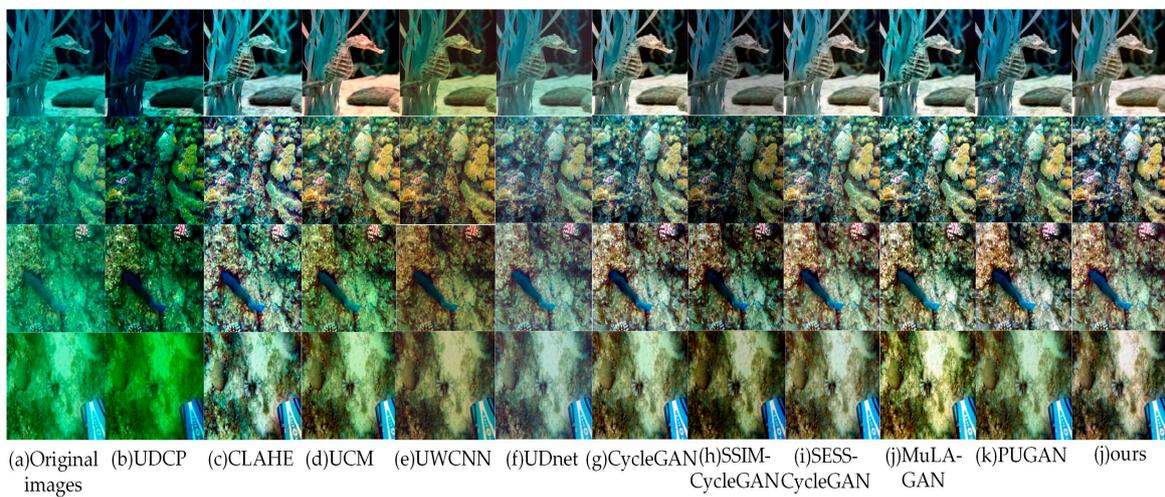
Among these metrics, SSIM and PSNR are full-reference measures, requiring a comparison between the generated high-quality underwater image and the corresponding ground-truth image. In contrast, UCIQE, UIQM, and AG are no-reference metrics, which can be directly calculated from the images without the need for external reference images.

#### 4.4. Visual Comparison with Other Methods

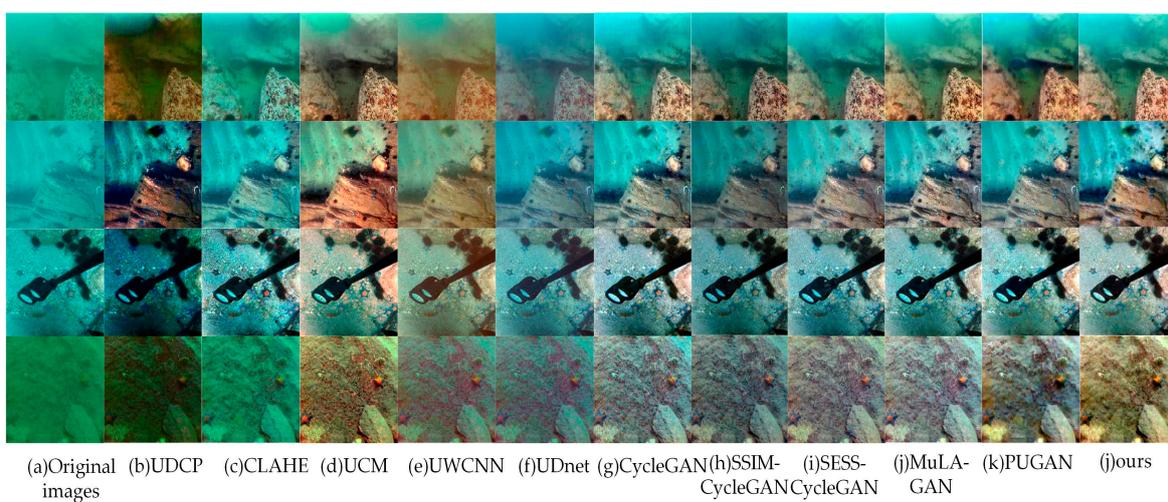
Different UIE algorithms were evaluated on the UIEB, EUVP, and RUIE test sets, with the results presented in Figures 8–10, respectively.



**Figure 8.** Visual comparison of image enhancement algorithms on the UIEB dataset.



**Figure 9.** Visual comparison of image enhancement algorithms on the EUVP dataset.



**Figure 10.** Visual comparison of image enhancement algorithms on the RUIE dataset.

From Figures 8–10, the UDCP, CLAHE, and UCM algorithms enhance image contrast to some extent. However, their performance in color correction is not particularly impressive. The processed images still display significant color distortion and fail to effectively

eliminate the “haze”. In contrast, GAN-based algorithms excel in color correction, accurately matching the color distributions of the target images. Nevertheless, images enhanced by CycleGAN exhibit localized red checkerboard artifacts and blurred edge details, which may be attributed to the limited ability of the CycleGAN generator to extract detailed features from images. Images processed by SSIM-CycleGAN tend to be darker, likely due to its requirement for consistency in SSIM metrics (brightness, contrast, and structural information) between the original and generated images while overlooking the inherent brightness differences that exist between them, resulting in unintended negative effects. The SESS-CycleGAN, MuLA-GAN, PUGAN, and the method proposed in this paper all achieve satisfactory results on the UIEB and EUVP datasets, with only slight differences in color distribution at the subjective visual level. Notably, on the challenging no-reference RUIE dataset, the images generated by our proposed method exhibit higher contrast and improved clarity. This indicates that our model demonstrates exceptional capability in capturing image details and possesses stronger generalization performance.

#### 4.5. Objective Comparison with Other Methods

Tables 1–3 present the performance metrics of various UIE methods on the UIEB, EUVP, and RUIE datasets, respectively. Values in bold indicate the best performance, and values underlined indicate the second-best performance.

**Table 1.** Performance comparison of various algorithms on the UIEB dataset (↑ indicates that higher values are more desirable).

Model/Method	UCIQE ↑	UIQM ↑	SSIM ↑	PSNR ↑	AG ↑
UDCP	<b>0.5230</b>	<b>1.3371</b>	0.6439	28.2402	9.8035
CLAHE	0.4429	<u>1.0293</u>	0.7190	28.2846	10.5692
UCM	<u>0.4927</u>	1.0138	0.7177	28.6008	10.9695
UWCNN	0.3707	0.5164	0.7066	28.0678	6.7591
UDnet	0.3554	0.4492	0.7665	28.1062	7.5259
CycleGAN	0.4504	0.7244	0.7554	28.2691	11.1365
SSIM-CycleGAN	0.4618	0.8319	0.7717	<u>28.9576</u>	11.3564
SESS-CycleGAN	0.4608	0.8330	0.7593	28.8892	<u>11.3594</u>
MuLA-GAN	0.4542	0.8325	<u>0.7759</u>	28.8459	10.0260
PUGAN	0.4628	0.7921	0.7714	28.9979	9.2452
ours	0.4842	0.8936	<b>0.7933</b>	<b>29.2120</b>	<b>11.6876</b>

**Table 2.** Performance comparison of various algorithms on the EUVP dataset (↑ indicates that higher values are more desirable).

Model/Method	UCIQE ↑	UIQM ↑	SSIM ↑	PSNR ↑	AG ↑
UDCP	<b>0.5204</b>	<b>1.1924</b>	0.5799	28.0624	7.3796
CLAHE	0.4597	<u>1.0586</u>	0.7295	28.3214	8.3591
UCM	<u>0.4745</u>	0.9554	0.7573	28.6456	8.6528
UWCNN	0.3958	0.5329	0.7437	28.0347	6.0778
UDnet	0.3649	0.4618	0.7535	28.1286	6.0875
CycleGAN	0.4457	0.7643	0.7635	28.2800	9.0708
SSIM-CycleGAN	0.4482	0.7989	<u>0.7771</u>	28.4973	9.5739
SESS-CycleGAN	0.4454	0.7775	0.7747	28.6656	9.4611
MuLA-GAN	0.4571	0.7819	0.7730	28.4388	<u>9.9494</u>
PUGAN	0.4495	0.8178	0.7647	<u>28.8436</u>	9.0193
ours	0.4669	0.8358	<b>0.7883</b>	<b>29.1524</b>	<b>10.7343</b>

By analyzing the experimental data shown in Tables 1–3, the proposed LM-CycleGAN algorithm achieved the best scores in SSIM, PSNR, and AG metrics. This suggests that the algorithm can not only accurately restore the details of target images, and reduce image distortion and noise, but also enhance visual quality, improving clarity and contrast. In

the UCIQE and UIQM metrics, the UDCP algorithm achieved the optimal value, while the CLAHE algorithm and the UCM algorithm achieved the sub-optimal value. When considering the subjective visual representations provided in Section 4.4, conclusions similar to those found in the literature [9], ref. [36] can be drawn. Given that the UCIQE and UIQM metrics primarily focus on linear combinations of low-level features such as contrast and saturation, they neglect higher-level semantic information or prior knowledge related to human visual perception. Additionally, these measures do not evaluate whether the intensity values of the entire image fall within a reasonable range. Consequently, although the UDCP, CLAHE, and UCM algorithms score higher in UCIQE and UIQM values, there is still a significant gap between the image quality produced by these algorithms and human visual perception. Notably, compared to deep learning-based methods, the proposed LM-CycleGAN algorithm not only achieved optimal values in these two metrics but also generated images that better align with human visual perception.

**Table 3.** Performance comparison of various algorithms on the RUIE dataset ( $\uparrow$  indicates that higher values are more desirable).

Model/Method	UCIQE $\uparrow$	UIQM $\uparrow$	AG $\uparrow$
UDCP	<b>0.4227</b>	<b>0.9736</b>	6.1208
CLAHE	0.3409	0.4619	7.8120
UCM	<u>0.3932</u>	<u>0.7004</u>	8.6002
UWCNN	0.3707	0.5164	6.7591
UDnet	0.3275	0.3614	4.5951
CycleGAN	0.3618	0.5765	7.8345
SSIM-CycleGAN	0.3638	0.6150	8.2624
SESS-CycleGAN	0.3627	0.6209	8.4896
MuLA-GAN	0.3688	0.6394	8.3231
PUGAN	0.3715	0.6288	<u>8.7618</u>
ours	0.3890	0.6569	<b>8.9634</b>

#### 4.6. Ablation Experiment

A series of ablation experiments were conducted to evaluate the effectiveness of the components in the proposed LM-CycleGAN model. Table 4 presents the performance comparison between the traditional CycleGAN model and the models augmented with different improvement modules. These improvement models include the CycleGAN model integrating the MSDA mechanism into the generator structure (hereafter referred to as MSDA), the CycleGAN model embedding the MAFA mechanism into the generator structure (hereafter referred to as MAFA), the CycleGAN model incorporating LPIPS Loss (hereafter referred to as LPIPS), the CycleGAN model with the joint MSDA mechanism and LPIPS Loss (hereafter referred to as MSDA + LPIPS), and the CycleGAN model with the joint MAFA mechanism and LPIPS Loss (the proposed methods).

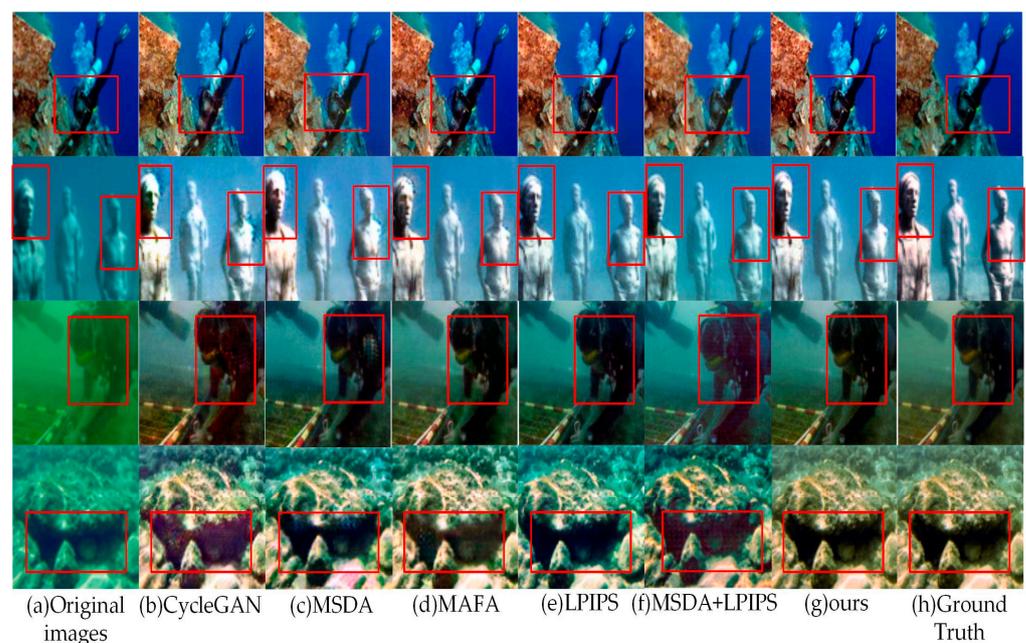
**Table 4.** Performance evaluation of CycleGAN models with various improvements on UIEB ( $\uparrow$  indicates that higher values are more desirable and “ $\checkmark$ ” indicates that the corresponding module has been added to CycleGAN).

Experiments	MSDA	MAFA	LPIPS	UCIQE $\uparrow$	UIQM $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	AG $\uparrow$
T1	—	—	—	0.4504	0.7244	0.7554	28.2691	11.1365
T2	$\checkmark$	—	—	0.4534	0.7708	0.7661	28.6008	11.3141
T3	—	$\checkmark$	—	0.4587	0.8491	0.7765	29.0604	11.4885
T4	—	—	$\checkmark$	0.4629	0.8501	0.7839	28.9574	11.4448
T5	$\checkmark$	—	$\checkmark$	0.4784	0.8763	0.7896	29.9013	11.5635
T6	—	$\checkmark$	$\checkmark$	<b>0.4842</b>	<b>0.8936</b>	<b>0.7933</b>	<b>29.2120</b>	<b>11.6876</b>

Table 4 shows that compared to the control group (T1), groups T2, T3, and T4 exhibit significant improvements across all five key performance metrics. This indicates that intro-

ducing the MSDA mechanism, MAFA mechanism, and LPIPS Loss effectively suppresses noise while revealing richer and more intricate texture details. Furthermore, the comparison between groups T2 and T3 demonstrates the effectiveness of the multi-scale feature adaptive fusion strategy. The results for groups T5 and T6 indicate that the combined model of MAFA and LPIPS outperforms the combined model of MSDA and LPIPS, further underscoring the superiority of the MAFA mechanism. Overall, the LM-CycleGAN model that integrates MAFA and LPIPS strategies (group T6) consistently outperforms the other models across all five evaluation metrics.

Figure 11 demonstrates the image enhancement effects of various improvement models on the UIEB dataset. Traditional CycleGAN-reconstructed images exhibit significant red checkerboard artifacts and blurred edge details. Images processed with MSDA have partially eliminated the red artifact but still exhibit color bias, localized detail blur, and slight grid effects. In contrast, images processed with MAFA have more realistic colors but still exhibit issues with localized detail blur. Images processed with LPIPS demonstrate clearer edge details, although they still present some color bias. The model that unites MAFA and LPIPS has made improvements in addressing localized detail blur but still has room for enhancement in color correction. Notably, the LM-CycleGAN model achieved the desired effects: images processed by it not only eliminated the red checkerboard artifacts and enhanced the clarity of texture details in the reconstructed underwater images but also maintained color accuracy.



**Figure 11.** Comparison of enhancement effects from different strategies on the UIEB dataset, where the red box highlights local enhancement areas.

#### 4.7. Real-Time Analysis and Discussion

Given that UIE techniques are typically deployed on resource-constrained devices, this section will examine and analyze the real-time performance of 11 UIE methods. To better simulate the real-world effects of deploying UIE technologies on such devices, we chose the relatively low-performance NVIDIA GeForce MX130 GPU for this experiment. In this analysis, three key performance indicators are considered: parameters, FLOPs, and FPS. Parameters refer to the number of trainable variables in a model, FLOPs (Floating Point Operations Per Second) indicate the computational complexity of a model, and FPS (Frames Per Second) measures the number of frames processed in one second, reflecting the real-time performance of an image processing method.

Through the analysis of Table 5, traditional UIE algorithms exhibit fast processing speeds (FPSs) due to their direct application of fixed rules or mathematical formulas for image processing. Among deep learning methods, although UWCNN has the smallest network size, its real-time performance is compromised as it requires physical models to generate additional images. MuLA-GAN achieves the highest FPS with fewer parameters and lower computational complexity. In contrast, the method proposed in this paper achieves an FPS of 94.3. Although this is not the optimal performance, it is sufficient to meet the real-time requirements for underwater operations.

**Table 5.** Real-time performance analysis of 11 underwater image enhancement methods ( $\uparrow$  indicates that higher values are more desirable, while  $\downarrow$  is the opposite).

Model/Method	Params (M) $\downarrow$	FLOPs (G) $\downarrow$	FPS (Hz) $\uparrow$
UDCP	—	—	142.3
CLAHE	—	—	154.2
UCM	—	—	128.4
UWCNN	1.1	3.08	42.3
UDnet	16.1	96.3	54.3
CycleGAN	11.4	58.2	99.2
SSIM-CycleGAN	13.6	63.3	89.4
SESS-CycleGAN	13.1	62.7	90.9
MuLA-GAN	17.3	30.2	117.8
PUGAN	95.7	72.5	62.5
ours	12.6	60.1	94.3

## 5. Conclusions

This paper proposes an advanced UIE method, LM-CycleGAN, which has achieved excellent experimental results on publicly available UIEB, EUVP, and RUIE datasets. By applying the multi-scale feature adaptive fusion strategy, LM-CycleGAN can effectively capture important detail features in underwater images, thereby enhancing the model's adaptability in diverse environments. Additionally, by introducing LPIPS Loss, the model pays more attention to structural consistency during the reconstruction process, which further enhances the ability to recover details in complex underwater images. However, the model performs poorly when handling high-saturation blue and green images, which are very common in underwater environments. A possible reason for this limitation is the insufficient number of such samples in the training dataset, which restricts the model's generalization ability in these scenarios, leading to issues such as color bias and detail loss. Future research will further explore directions such as dataset augmentation and model lightweight to provide more robust and efficient solutions for underwater image enhancement.

**Author Contributions:** Conceptualization, J.W. and G.Z.; methodology, J.W.; software, J.W.; validation, J.W. and Y.F.; formal analysis, J.W.; investigation, J.W.; resources, G.Z.; data curation, J.W.; writing—original draft preparation, J.W.; writing—review and editing, G.Z.; visualization, J.W. and Y.F.; supervision, G.Z. and Y.F.; project administration, G.Z.; funding acquisition, G.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Yunnan Provincial Major Science and Technology Project (Research on the blockchain-based agricultural product traceability system and demonstration of platform construction; Project No. 202102AD080002).

**Institutional Review Board Statement:** Not Applicable.

**Informed Consent Statement:** Not Applicable.

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

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

## References

1. Li, J.; Xu, W.; Deng, L.; Xiao, Y.; Han, Z.; Zheng, H. Deep learning for visual recognition and detection of aquatic animals: A review. *Rev. Aquac.* **2023**, *15*, 409–433. [[CrossRef](#)]
2. Guyot, A.; Lennon, M.; Thomas, N.; Gueguen, S.; Petit, T.; Lorho, T.; Cassen, S.; Hubert-Moy, L. Airborne Hyperspectral Imaging for Submerged Archaeological Mapping in Shallow Water Environments. *Remote Sens.* **2019**, *11*, 2237. [[CrossRef](#)]
3. Lee, D.; Kim, G.; Kim, D.; Myung, H.; Choi, H.-T. Vision-based object detection and tracking for autonomous navigation of underwater robots. *Ocean Eng.* **2012**, *48*, 59–68. [[CrossRef](#)]
4. Bell, K.L.; Chow, J.S.; Hope, A.; Quinzin, M.C.; Cantner, K.A.; Amon, D.J.; Cramp, J.E.; Rotjan, R.D.; Kamalu, L.; de Vos, A.; et al. Low-cost, deep-sea imaging and analysis tools for deep-sea exploration: A collaborative design study. *Front. Mar. Sci.* **2022**, *9*, 873700. [[CrossRef](#)]
5. Zhou, J.; Yang, T.; Zhang, W. Underwater vision enhancement technologies: A comprehensive review, challenges, and recent trends. *Appl. Intell.* **2023**, *53*, 3594–3621. [[CrossRef](#)]
6. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 586–595.
7. 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](#)]
8. Islam, M.J.; Xia, Y.; Sattar, J. Fast underwater image enhancement for improved visual perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [[CrossRef](#)]
9. Liu, R.; Fan, X.; Zhu, M.; Hou, M.; Luo, Z. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 4861–4875. [[CrossRef](#)]
10. Li, C.; Hu, E.; Zhang, X.; Zhou, H.; Xiong, H.; Liu, Y. Visibility restoration for real-world hazy images via improved physical model and Gaussian total variation. *Front. Comput. Sci.* **2024**, *18*, 181708. [[CrossRef](#)]
11. Liu, Y.; Yan, Z.; Tan, J.; Li, Y. Multi-purpose oriented single nighttime image haze removal based on unified variational retinex model. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *33*, 1643–1657. [[CrossRef](#)]
12. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 2341–2353. [[PubMed](#)]
13. Chao, L.; Wang, M. Removal of water scattering. In Proceedings of the 2010 2nd International Conference on Computer Engineering and Technology, Chengdu, China, 16–18 April 2010; pp. V2–V35.
14. Chiang, J.Y.; Chen, Y.C. Underwater image enhancement by wavelength compensation and dehazing. *IEEE Trans. Image Process.* **2011**, *21*, 1756–1769. [[CrossRef](#)] [[PubMed](#)]
15. Drews, P.; Nascimento, E.; Moraes, F.; Botelho, S.; Campos, M. Transmission estimation in underwater single images. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Sydney, NSW, Australia, 1–8 December 2013; pp. 825–830.
16. Garg, D.; Garg, N.K.; Kumar, M. Underwater image enhancement using blending of CLAHE and percentile methodologies. *Multimed. Tools Appl.* **2018**, *77*, 26545–26561. [[CrossRef](#)]
17. Hu, H.; Xu, S.; Zhao, Y.; Chen, H.; Yang, S.; Liu, H.; Zhai, J.; Li, X. Enhancing Underwater Image via Color-Cast Correction and Luminance Fusion. *IEEE J. Ocean. Eng.* **2023**, *49*, 15–29. [[CrossRef](#)]
18. Iqbal, K.; Odetayo, M.; James, A.; Salam, R.A.; Talib, A.Z.H. Enhancing the low quality images using unsupervised colour correction method. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10–13 October 2010; pp. 1703–1709.
19. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, USA, 8–13 December 2014; p. 27.
20. Anwar, S.; Li, C.; Porikli, F. Deep underwater image enhancement. *arXiv* **2018**, arXiv:1807.03528.
21. Saleh, A.; Sheaves, M.; Jerry, D.; Azghadi, M.R. Adaptive uncertainty distribution in deep learning for unsupervised underwater image enhancement. *arXiv* **2022**, arXiv:2212.08983.
22. 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.
23. Li, C.; Guo, J.; Guo, C. Emerging from water: Underwater image color correction based on weakly supervised color transfer. *IEEE Signal Process. Lett.* **2018**, *25*, 323–327. [[CrossRef](#)]
24. Li, Q.Z.; Bai, W.X.; Niu, J. Underwater image color correction and enhancement based on improved cycle-consistent generative adversarial networks. *Acta Autom. Sin.* **2020**, *46*, 1–11.
25. Chen, B.; Zhang, X.; Wang, R.; Li, Z.; Deng, W. Detect concrete cracks based on OTSU algorithm with differential image. *J. Eng.* **2019**, *23*, 9088–9091. [[CrossRef](#)]
26. Bakht, A.B.; Jia, Z.; Din, M.U.; Akram, W.; Saoud, L.S.; Seneviratne, L.; Lin, D.; He, S.; Hussain, I. MuLA-GAN: Multi-Level Attention GAN for Enhanced Underwater Visibility. *Ecol. Inform.* **2024**, *81*, 102631. [[CrossRef](#)]
27. Cong, R.; Yang, W.; Zhang, W.; Li, C.; Guo, C.-L.; Huang, Q.; Kwong, S. PUGAN: Physical model-guided underwater image enhancement using gan with dual-discriminators. *IEEE Trans. Image Process.* **2023**, *32*, 4472–4485. [[CrossRef](#)] [[PubMed](#)]

28. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In Proceedings of the 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 694–711.
29. Jiao, J.; Tang, Y.M.; Lin, K.Y.; Gao, Y.; Ma, A.J.; Wang, Y.; Zheng, W.-S. Dilateformer: Multi-scale dilated transformer for visual recognition. *IEEE Trans. Multimed.* **2023**, *25*, 8906–8919. [[CrossRef](#)]
30. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
31. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–8 December 2012; pp. 1097–1105.
32. Iandola, F.N. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. *arXiv* **2016**, arXiv:1602.07360.
33. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
34. Yang, M.; Sowmya, A. An underwater color image quality evaluation metric. *IEEE Trans. Image Process.* **2015**, *24*, 6062–6071. [[CrossRef](#)]
35. Panetta, K.; Gao, C.; Agaian, S. Human-visual-system-inspired underwater image quality measures. *IEEE J. Ocean. Eng.* **2015**, *41*, 541–551. [[CrossRef](#)]
36. Wang, Y.; Guo, J.; He, W.; Gao, H.; Yue, H.; Zhang, Z.; Li, C. Is underwater image enhancement all object detectors need? *IEEE J. Ocean. Eng.* **2024**, *49*, 606–621. [[CrossRef](#)]

**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.