

Article

MSFE-UIENet: A Multi-Scale Feature Extraction Network for Marine Underwater Image Enhancement

Shengya Zhao ^{1,2,†}, Xinkui Mei ^{1,*},, Xiufen Ye ¹ and Shuxiang Guo ³¹ College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China; zhaosy@hrbeu.edu.cn (S.Z.); yexiufen@hrbeu.edu.cn (X.Y.)² Deep Sea Technology Department, National Deep Sea Center, Qingdao 266037, China³ Department of Electronic and Electrical Engineering, Southern University of Science and Technology, Shenzhen 518055, China; guo.shuxiang@sustech.edu.cn

* Correspondence: meixinkui@hrbeu.edu.cn

† These authors contributed equally to this work.

Abstract: Underwater optical images have outstanding advantages for short-range underwater target detection tasks. However, owing to the limitations of special underwater imaging environments, underwater images often have several problems, such as noise interference, blur texture, low contrast, and color distortion. Marine underwater image enhancement addresses degraded underwater image quality caused by light absorption and scattering. This study introduces MSFE-UIENet, a high-performance network designed to improve image feature extraction, resulting in deep-learning-based underwater image enhancement, addressing the limitations of single convolution and upsampling/downsampling techniques. This network is designed to enhance the image quality in underwater settings by employing an encoder–decoder architecture. In response to the underwhelming enhancement performance caused by the conventional networks' sole downsampling method, this study introduces a pyramid downsampling module that captures more intricate image features through multi-scale downsampling. Additionally, to augment the feature extraction capabilities of the network, an advanced feature extraction module was proposed to capture detailed information from underwater images. Furthermore, to optimize the network's gradient flow, forward and backward branches were introduced to accelerate its convergence rate and improve stability. Experimental validation using underwater image datasets indicated that the proposed network effectively enhances underwater image quality, effectively preserving image details and noise suppression across various underwater environments.

Keywords: underwater image enhancement; multi-scale feature extraction; pyramid downsampling module; forward and backward branches



Citation: Zhao, S.; Mei, X.; Ye, X.; Guo, S. MSFE-UIENet: A Multi-Scale Feature Extraction Network for Marine Underwater Image Enhancement. *J. Mar. Sci. Eng.* **2024**, *12*, 1472. <https://doi.org/10.3390/jmse12091472>

Academic Editor: Sergei Chernyi

Received: 26 July 2024

Revised: 17 August 2024

Accepted: 22 August 2024

Published: 23 August 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/>).

1. Introduction

Advancements in science and technology, along with a better understanding of the ocean, have led researchers to explore and develop marine resources, especially in the deep sea. The deep sea harbors abundant energy, mineral, biological, and genetic resources. Humans use underwater robots equipped with various detection devices to explore resources and monitor deep-sea environments. Using visual systems for underwater optical detection, as a main method, plays a key role in deep-sea exploration, much like their terrestrial counterparts [1]. Due to the minimal impact of the air medium on light propagation, optical imaging in air is relatively straightforward. However, imaging principles become significantly more intricate in complex media [2,3]. The underwater environment medium is exceptionally intricate; within this context, significant differences exist between underwater and aerial environments owing to the varying light absorption effects. Although negligible in air [4], this effect becomes pronounced underwater, with red light wavelengths experiencing substantial attenuation in water, whereas blue-green wavelengths are less

affected. Consequently, underwater images often have blue-green tints and blurred textures owing to small waterborne particles. These low-quality images pose significant challenges to marine resource development, ocean fisheries, naval operations, marine engineering, and scientific research, making the acquisition of high-quality underwater images a priority. Figure 1 illustrates the principle of underwater light propagation and demonstrates the effects of underwater image enhancement (UIE).

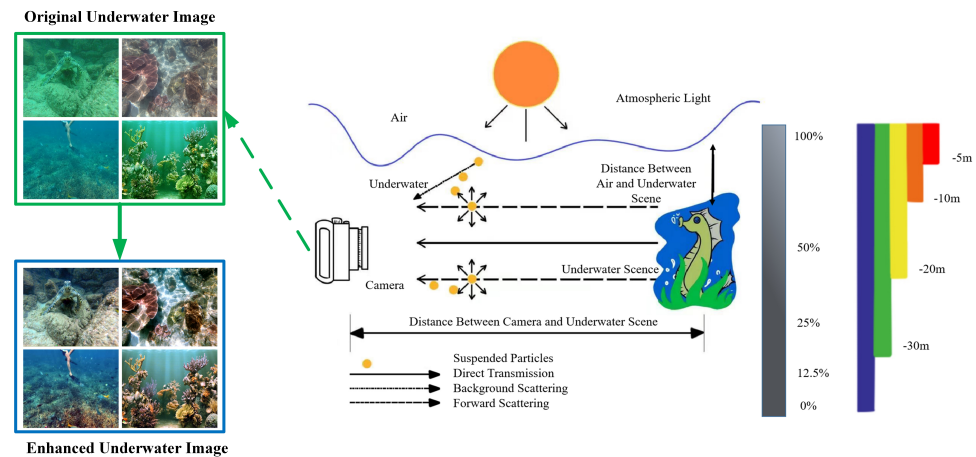


Figure 1. Schematic of the underwater optical imaging principle.

In recent years, several strategies have been proposed to enhance underwater imagery. These methods are broadly divided into three primary categories: model-free, model-based, and deep-learning-based [5]. Traditional UIE techniques, which encompass both model-free and model-based methods, exhibit specific enhancement and restoration effects. However, these techniques have significant limitations in terms of their real-time performance and generalization capabilities. With the rapid progression of deep-learning technology, deep-learning-based UIE methods have surfaced as a key area of research. Compared to traditional methods, deep-learning-based methods have demonstrated superior modeling capabilities and robustness [6]. Deep neural networks (DNNs) employ multiple layers of processing units to abstract high-level features from low-level input information. The objective of feature extraction is to discern valuable insights from raw data and convert them into more effective representations. This procedure leverages the inherent robust nonlinear fitting and generalization capabilities of DNNs, thereby facilitating the adaptive learning of semantics, robustness, and data relationships. Consequently, the efficacy of DNNs is substantially influenced by the quality of feature extraction, a critical factor for most UIE methods based on deep learning.

Nevertheless, it is crucial to underscore that prevalent learning-based methods frequently augment their feature extraction capabilities by amplifying the number of network layers. However, they often overlook the segregation and amalgamation of features at varying depths. Specifically, an overabundance of network layers can lead to the loss of pertinent information during the transition from low-level to high-level features, consequently impacting the overall quality of feature extraction. The existing deep-learning-based UIE approaches process all scale features of degraded images using a uniform receptive field size. For features at different scales, a single receptive field cannot accommodate all the details, frequently resulting in challenges such as information loss, diminished contrast, and inadequate feature-extraction capability [7]. Furthermore, in terms of network structure design, deep-learning-based methods continue to face training difficulties and the problem of gradient explosion.

Moreover, an excess of network layers often learns more irrelevant or arbitrary features, resulting in a substantial augmentation of the network's computational burden. To mitigate this problem, we have devised a multi-scale feature extraction network (MSFE-UIENet)

to bolster underwater image enhancement via efficient feature extraction. The principal contributions of this study are enumerated below:

1. This study introduces a high-performance UIE network based on multi-scale feature extraction. The network incorporates two fundamental modules: the feature extraction module (FEM) and the multi-scale spatial pyramid pooling features block (MSPPF). These modules effectively amplify the feature extraction capability of the network and minimize the insufficient enhancement effects typically observed in traditional enhancement networks.
2. Forward and backward branches are incorporated to improve the gradient flow of the network. After processing the source feature using this module, the desired shape can be attained, thereby facilitating the integration of the target feature.
3. Comprehensive evaluations indicate that the proposed network outperforms several state-of-the-art methods in terms of enhancement effects and computational complexity on widely utilized public underwater datasets.

2. Related Works

Considering the challenges associated with the enhancement of degraded underwater images, extensive research has been conducted. In this section, we discuss the three main method categories.

2.1. Model-Free Methods

Various model-free methods based on multiple fusion frameworks have been established as effective strategies for enhancing underwater images. These methods typically consist of a sequential process in which fusion images are generated from the input image, which are determined by factors such as contrast, detail, and exposure. Finally, the fusion images and fusion weights are combined to produce the enhanced output image.

Iqbal et al. [8] proposed an unsupervised color correction method for UIE that relies on color balancing and contrast correction in both the RGB and HSI color models. Hitam et al. [9] introduced a UIE method called CLAHE, which combines the results from RGB and HSV color models using Euclidean norms. Fu et al. [10] proposed a widely used classical method based on Retinex, whereas Zhang et al. [11] developed an LAB-MSR method based on the Retinex model to enhance underwater images. Li et al. [12] combined the traditional Retinex model with noise maps to create a robust Retinex model. Ancuti et al. [13] presented a novel strategy for enhancing underwater videos and images based on fusion principles, whereas Gao et al. [14] proposed a method for UIE using adaptive retinal mechanisms. Yuan et al. [15] introduced a method based on contour bougie morphology. Generally, current model-free methods can improve the global contrast and edge sharpness of underwater images. However, color casts may persist, and excessive enhancement remains a potential issue.

2.2. Model-Based Methods

Underwater image processing involves eliminating the haze caused by scattering and restoring color distortion. Several methods have recently incorporated physical imaging processes into UIE by constructing models that reverse the degradation process to restore clear underwater scenes. These physics-based models estimate underwater imaging parameters using prior information.

He et al. [16] proposed a dark channel prior (DCP) that performed well in image de-fogging. Some researchers have combined DCP with wavelength-dependent compensation algorithms to restore underwater images. Drews et al. [17] introduced underwater DCP (UDCP), which applies adaptive DCP to estimate the transmission of the underwater scene. Peng et al. [18] developed a generalized DCP (GDCP) for image restoration that integrates adaptive color correction into the image formation model. They also proposed a method called the IBLA [19], which is based on light absorption and image blurriness for depth estimation. Galdrán et al. [20] enhanced the contrast of underwater images by recovering the

colors associated with short wavelengths. Berman et al. [21] proposed a method to restore underwater images using a haze-lines model based on wavelength-dependent attenuation. Mei et al. [22] proposed a method based on optical geometric properties. Although these methods improve the details of underwater images, they often excessively amplify the noise and cause color distortions, occasionally resulting in overenhancement.

2.3. Deep-Learning-Based Methods

With the advancements in deep learning, its applications have significantly expanded to the field of UIE.

Wang et al. [23] proposed UICE2-Net, a model that efficiently and effectively integrates both RGB and HSV color spaces within a single convolutional neural network (CNN). Li et al. [24] introduced Water-Net as a baseline for training CNNs and developed the UWCNN [25] model, which directly reconstructs clear underwater images. This model benefits from underwater image training data obtained through artificial synthesis using prior information. Prasen et al. [26] proposed DeepWaveNet, a framework optimized using pixel-wise and feature-cost functions. Li et al. [27] introduced a generative adversarial network (GAN)-based model called WaterGAN to generate underwater image datasets from air images and depth pairings that were used for underwater image color correction through an unsupervised pipeline. Mei et al. [28] proposed UIR-Net, a lightweight baseline model that recovers and enhances underwater images simultaneously. Although it exhibits good recovery effects, it still presents deficiencies in terms of color correction. Wang et al. [29] developed a self-supervised model called UWdepth to obtain depth information from underwater scenes using monocular sequences, which were then employed in UIE processes. Fabbri et al. [30] proposed UGAN, a GAN-based model, to improve the quality of underwater images. Alik Pramanick et al. [31] proposed a framework called X-CAUNET to enhance underwater images by using cross-attention transformers. Huang et al. [32] proposed a mean-teacher-based semi-UIR framework to incorporate the unlabeled data into network training. Ankita Naik et al. [33] introduced a shallow neural network architecture that maintains performance while reducing parameters compared to the latest models. Owing to the scarcity of underwater datasets, supervised methods often do not yield satisfactory results. Instead, they produce outcomes characterized by oversaturation, lack of clarity in detail, and unnatural background colors.

3. Method

3.1. Framework of MSFE-UIENet

To address the shortcomings of the traditional single downsampling method, which often leads to subpar enhancement effects, MSFE-UIENet introduces an innovative pyramid downsampling module called MSPPF. By employing multi-scale downsampling, this module captures a more comprehensive array of image information, thereby amplifying the image enhancement outcomes. This study presents a high-performance FEM to further enhance the feature extraction capability of a network. With a robust feature extraction algorithm and structure, this module can accurately identify key features in underwater images, thereby providing a superior input for subsequent enhancement procedures. Moreover, forward and backward branches were introduced to optimize the gradient flow of the network. The UIE network proposed in this study significantly improves the performance and effectiveness of UIE by integrating a pyramid downsampling module, a high-performance FEM, and forward and backward branches.

Figure 2 illustrates the comprehensive architecture of MSFE-UIENet, encompassing six FEMs, two MSPPF modules, and both forward and backward branch structures. The underwater image intended for enhancement was utilized as the input and processed using diverse module structures of the network. The derived feature information undergoes procedures such as sampling and channel merging, ultimately facilitating efficient tasks related to UIE. The main process is as follows:

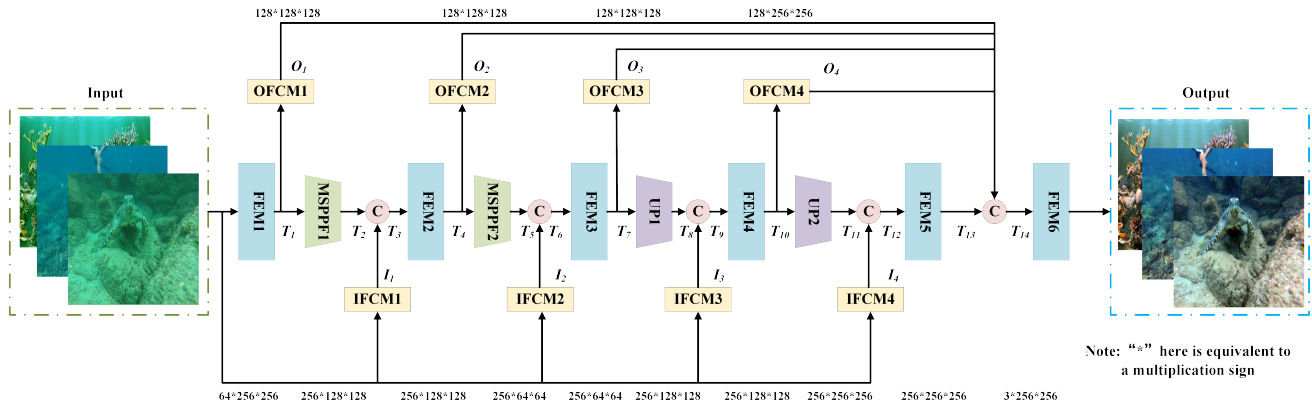


Figure 2. Framework diagram of the algorithm.

The underwater image to be enhanced serves as input x , and we can sequentially extract features using six FEMs. The formula is as follows:

$$\begin{cases} T_1 = \mathcal{F}(x) \\ T_2 = \mathcal{F}(\text{cat}[\mathcal{M}(T_1), I_1]) \\ T_3 = \mathcal{F}(\text{cat}[\mathcal{M}(T_2), I_2]) \\ T_4 = \mathcal{F}(\text{cat}[\mathcal{U}(T_3), I_3]) \\ T_5 = \mathcal{F}(\text{cat}[\mathcal{U}(T_4), I_4]) \end{cases} \quad (1)$$

where $\mathcal{F}(\cdot)$ denotes the FEM operation, $\mathcal{M}(\cdot)$ symbolizes the downsampling operation using the MSPPF module, $\text{cat}[\cdot]$ signifies the channel merging operation, $\mathcal{U}(\cdot)$ represents the upsampling operation, and I_i represents the i -th forward features.

The feature T_1 , with a shape of $64 \times 256 \times 256$, is input into the MSPPF module for downsampling, yielding $\mathcal{M}(T_1)$ with the shape of $256 \times 128 \times 128$. Through subsequent FEM feature extraction operations, MSPPF module downsampling operations, and channel merging upsampling operations, we obtain features T_2 , T_3 , T_4 , and T_5 , each with a shape of $256 \times 128 \times 128$.

$$\text{Output} = \mathcal{F}(\text{cat}[T_5, O_1, O_2, O_3, O_4]) \quad (2)$$

Finally, *Output* is derived, where O_i represent the i -th backward features and *Output* represents the final output image.

3.2. Feature Extraction Module (FEM)

In recent years, FEMs have proliferated in the field of object detection based on deep learning. Conversely, in the enhancement domain, encoder–decoder networks predominantly serve task learning. To address the problem that traditional enhancement networks cannot fully harness image features using a single convolutional structure, this study introduces a high-performance FEM. The primary focus was on amalgamating the residual structure method and channel merging to thoroughly extract the features of the image. Figure 3 shows the structure of the FEM.

FEM first uses two CIS modules to extract features from the input features, and then one of the features passes through the two RFBS. To improve the feature extraction ability, the feature is output through a CIS module after merging the channel with another feature.

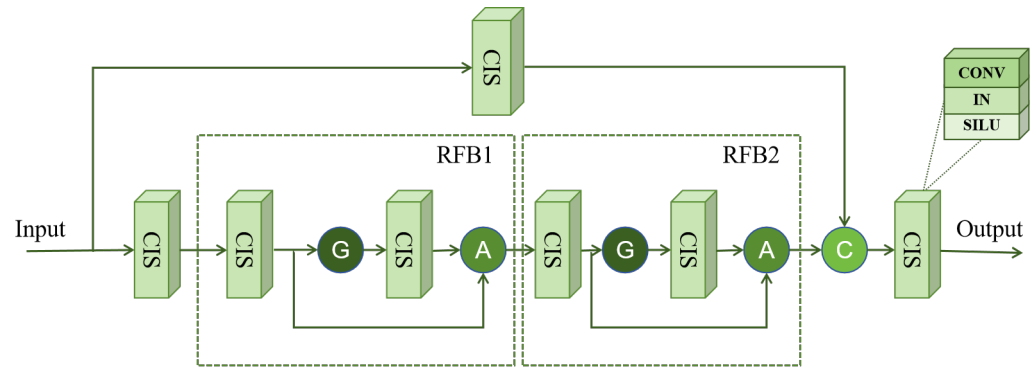


Figure 3. Feature extraction module. CIS comprises convolution, instance normalization, and SiLU activation function. The convolution kernel is 3. Two RFBs are used in this part.

3.3. Multi-Scale Spatial Pyramid Pooling Features (MSPPF)

In traditional encoder–decoder networks for enhancement, the downsampling operation often uses the simplest max-pooling for downsampling. Pyramid-type downsampling has been proven to improve the network feature extraction ability, which improves spatial perception and enhances robustness. Therefore, the MSPPF module shown in Figure 4 was constructed in this study.

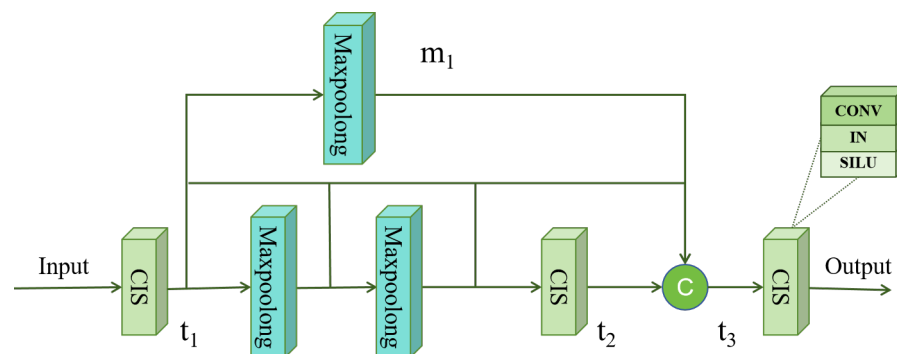


Figure 4. Schematic of MSPPF module. CIS comprises convolution, instance normalization, and SiLU activation function. The convolution kernel is 3.

The MSPPF module can downsample the input features and is expressed as follows:

$$\begin{cases} t_1 = \mathcal{C}(x) \\ t_2 = \mathcal{C}(\mathcal{MP}_2(t_1)) \\ t_3 = \text{cat}[\mathcal{MP}(t_1), t_2] \end{cases} \quad (3)$$

where x is the input; $\mathcal{C}(\cdot)$ represents a feature extraction unit consisting of a convolution, instance normalization (IN), and the SiLU activation function; $\mathcal{MP}(\cdot)$ denotes the max-pooling operation; and $\mathcal{MP}_2(\cdot)$ denotes two consecutive max-pooling operations

$$\text{Output} = \mathcal{C}(t_3) \quad (4)$$

The feature sampling information *Output* obtained by the MSPPF module captures a more comprehensive array of image information, thereby amplifying the image enhancement outcomes.

The structural superiority of the multi-scale pyramid pooling features (MSPPF) module is attributed to its ability for multi-scale feature extraction and its efficient computational performance. By implementing pooling operations at varying scales, the MSPPF module is capable of capturing a richer set of feature information, optimizing the computational process, reducing redundancy, and consequently enhancing performance.

3.4. Forward and Backward Branches

In encoder–decoder-styled enhancement networks, skip connections serve as a technique linking low-level features with high-level counterparts. Its function is to directly transmit low-level detailed information to a higher level, thereby providing a more comprehensive feature representation and superior information flow. Nevertheless, previous networks (such as Unet and UNet++) could only establish skip connections between the features of specific layers. In this study, we critically examined this structure and proposed a method capable of establishing skip connections in any layer.

We designed a feature-processing module for both forward and backward branches. The features preceding the jump are referred to as source features, whereas those following the jump are referred to as target features. Once the source feature is processed by this module, the anticipated shape can be acquired, allowing the merging of the target feature.

3.4.1. Forward Calculation Module (FCM)

As shown in Figure 2, the forward branch generates the forward calculation module (FCM). The FCM processes the output image (source features) to obtain four features, I_1 , I_2 , I_3 , and I_4 , and then merges them with the target features. The structure of the FCM module, which primarily modifies the shape of the source features through resizing and CIS convolution units, is shown in Figure 5.

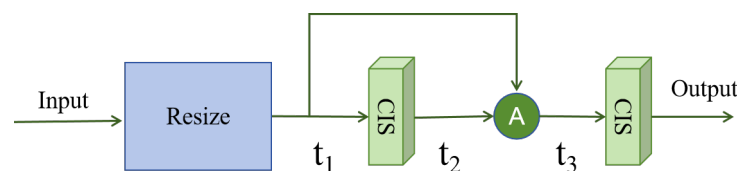


Figure 5. FCM. Input is the source feature (the input image). Output is the target feature.

The formula is as follows:

$$\begin{cases} t_1 = \mathcal{R}(x) \\ t_2 = \mathcal{C}(t_1) \\ t_3 = \text{Add}[t_1, t_2] \end{cases} \quad (5)$$

where x is the input, $\mathcal{R}(\cdot)$ is an operation that modifies the length and width of the source feature to match the length and width of the target feature, and $\text{Add}[\cdot, \cdot]$ is an operation that adds two features.

$$\text{Output} = \mathcal{C}(t_3) \quad (6)$$

3.4.2. Backward Calculation Module (BCM)

As shown in Figure 2, the backward branch enables the backward calculation module (BCM) to process the source features to obtain four features, O_1 , O_2 , O_3 , and O_4 , and then merge them with the target feature. The structure of the BCM is shown in Figure 6, where the shape of the source features is modified by the interpolation and CIS convolution units.

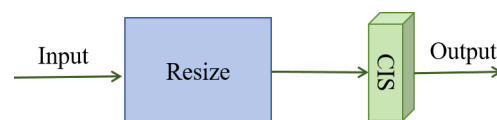


Figure 6. BCM. Input is the source feature. Output is the target feature.

The formula is as follows:

$$\text{Output} = \mathcal{C}(\mathcal{R}(x)) \quad (7)$$

where x is the input.

3.5. Loss Function

Currently, most deep-learning tasks are trained based on empirical risk minimization, and the loss function is crucial for the final outcome of the task. Therefore, it is crucial to examine the impact of the loss function on the image enhancement results in deep-learning-based UIE studies. After an extensive experimental analysis, we used L_{ALL} in (12) as the loss function.

The L_1 Loss, also known as the mean absolute error, represents the average distance between the predicted values, x , and the true values, y , of the model. It is utilized to quantify the pixel-level discrepancy between the reference network and the training outcomes, and it can be computed as follows:

$$L_1 = \frac{1}{N} \sum |x - y| \quad (8)$$

The L_{SSIM} Loss takes into account luminance, contrast, and structural factors, all of which significantly influence human visual perception. It can be calculated as follows:

$$\begin{aligned} SSIM(x, y) &= [l(x, y)^\alpha * c(x, y)^\beta * s(x, y)^\lambda] \\ &= \frac{(2\mu_x\mu_y + c_1)(2\delta_x\delta_y + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\delta_x^2 + \delta_y^2 + c_2)} \end{aligned} \quad (9)$$

where $l(x, y)$, $c(x, y)$, $s(x, y)$ are as follows:

$$\begin{aligned} l(x, y) &= \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \\ c(x, y) &= \frac{2\delta_x\delta_y + c_2}{\delta_x^2 + \delta_y^2 + c_2} \\ s(x, y) &= \frac{\delta_{xy} + c_3}{\delta_x + \delta_y + c_3} \end{aligned} \quad (10)$$

Therefore, we can obtain L_{SSIM} as follows:

$$L_{SSIM} = 1 - SSIM(x, y) \quad (11)$$

To sum up, L_{ALL} Loss can be obtained as follows:

$$L_{ALL} = k_1 L_1 + k_2 L_{SSIM} \quad (12)$$

where $k_1 = 0.8$, $k_2 = 0.2$.

4. Experiments and Discussion

4.1. Datasets and Settings

The experimental data were procured from the widely utilized open-source UIEB datasets [24] in this domain. UIEB images frequently display color distortion, diminished contrast, loss of detail, and a bias towards green or blue tints. The UIEB consists of 890 pairs of underwater images, encompassing both the original and reference images. During the experiment, we selected 800 images for training and 90 images for testing; all other deep-learning-based comparative methods have undergone training and testing using UIEB datasets with the same train ratio. In addition, we also selected 60 no-reference challenging images from the UIEB dataset for testing to verify the superiority of the proposed method.

The experiments implemented the Adam optimizer on a system running Windows 11, equipped with an NVIDIA RTX 4090 GPU (NVIDIA, Santa Clara, CA, USA). The batch size was configured to 4. Learning rate was established at 0.001. The number of epochs was determined to be 200. For the convenience of training, all image sizes are uniformly set to 256×256 , and the entire training time was controlled within 6 h.

4.2. Comparison with State-of-the-Art Methods for UIE

Our method is compared with eight current state-of-the-art (SOTA) UIE methods, including three traditional restoration and enhancement methods, IBLA [19], DCP [16], and Retinex [10], and five deep-learning-based methods, WaterNet [24], UGAN [30], Shallow [33], Deepwave [26], and FunieGan [34].

4.2.1. Qualitative Evaluations

We evaluated our method against the aforementioned techniques using underwater images captured in challenging scenes. Figure 7 presents a comprehensive comparison, primarily featuring bluish and greenish images. The results presented in Figure 7 reveal that our method improves image contrast, enhances detail, and eliminates color distortion, surpassing other existing methods. Specifically, our method circumvents incorrect color correction—as observed in IBLA [19] and DCP [16]—and excessive enhancement, as seen in Retinex [10]. Although other methods have achieved UIE to a certain extent, they still have shortcomings. Furthermore, the augmented images generated using our approach display more realistic color information and finer details, rendering them more congruent with human visual perception. Notably, our method consistently delivers a strong performance across all tested images.

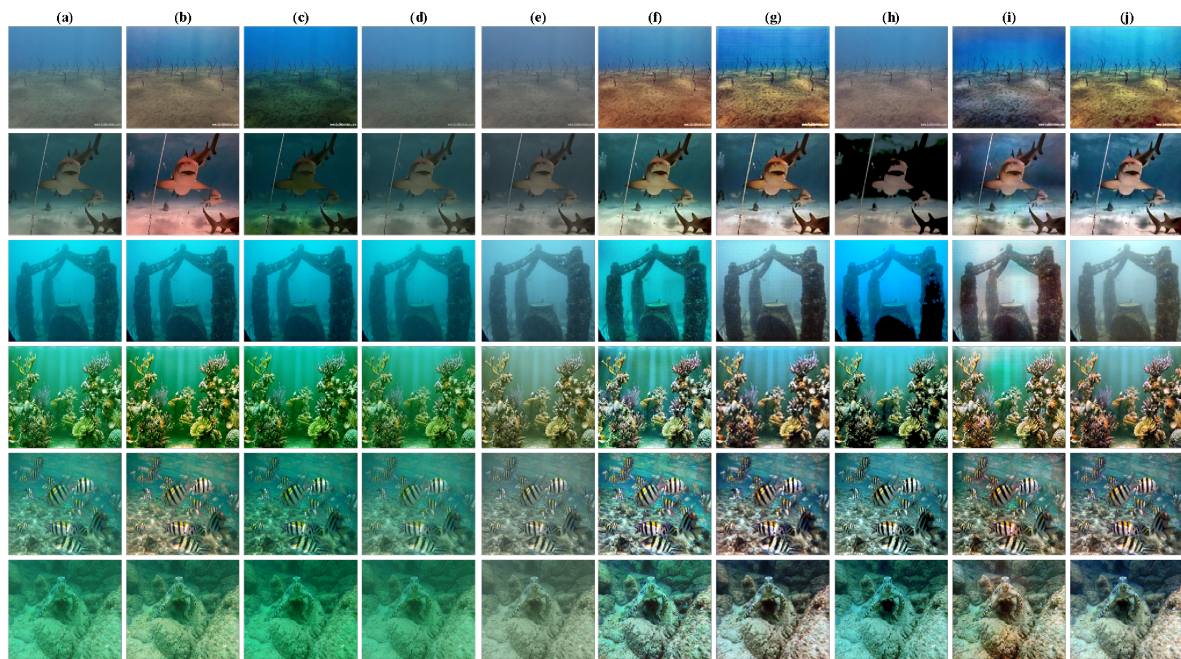


Figure 7. Comparison of enhancement results for bluish and greenish images of UIEB: (a) Raw. (b) IBLA [19]. (c) DCP [16]. (d) Retinex [10]. (e) Shallow [33]. (f) WaterNet [24]. (g) UGAN [30]. (h) Deepwave [26]. (i) FunieGan [34]. (j) Ours.

To demonstrate the superior processing capability of our method for challenging underwater images, we present and analyze the texture details of image enhancement effects, which are vital for underwater images with wide scenes containing abundant texture details. However, several existing UIE techniques tend to overlook essential information. In contrast, our method retains detailed information while enhancing the images. This is evident from the comparison of the enhancement results for the texture-rich images presented in Figure 8. Although some methods, such as Retinex [10], Shallow [33], and Deepwave [26], appear to achieve image enhancement, closer inspection of the details reveals a loss of texture information. This is a prevalent issue in most current underwater image-enhancement techniques. In contrast, our method maximizes the preservation of detail and texture information during the UIE process.

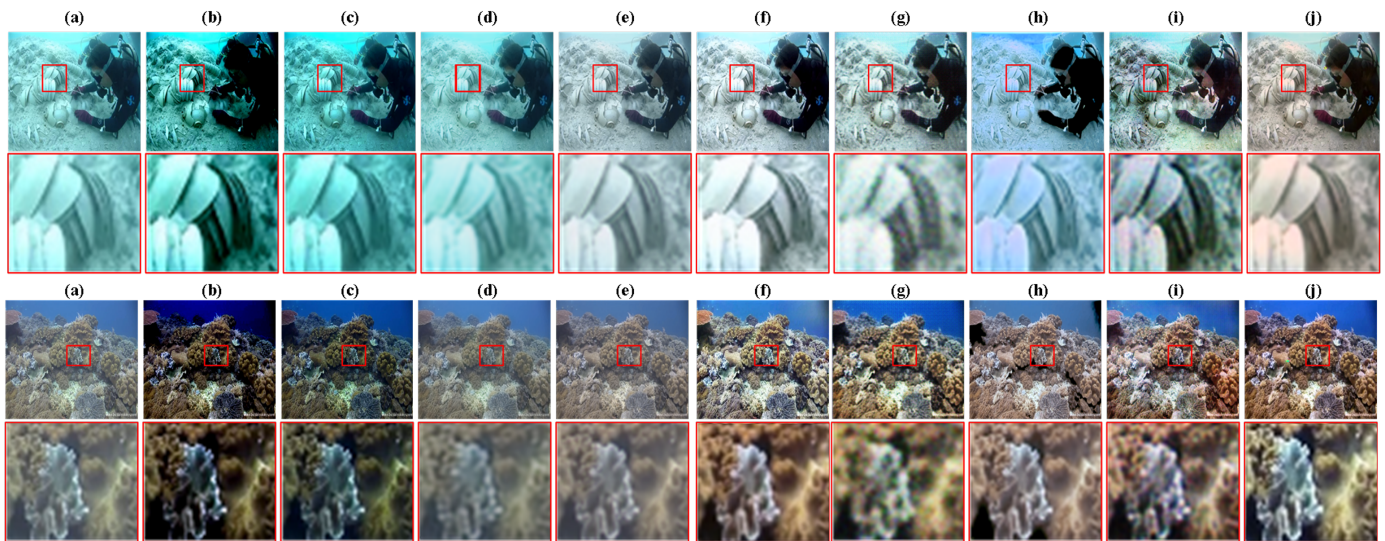


Figure 8. Comparison of enhancement results for detailed information: (a) Raw. (b) IBLA [19]. (c) DCP [16]. (d) Retinex [10]. (e) Shallow [33]. (f) WaterNet [24]. (g) UGAN [30]. (h) Deepwave [26]. (i) FunieGan [34]. (j) Ours.

To further substantiate the superiority of our proposed method, we conducted tests on 60 no-reference challenging images from the UIEB dataset. The test result is shown in Figure 9; our method also has good performance.

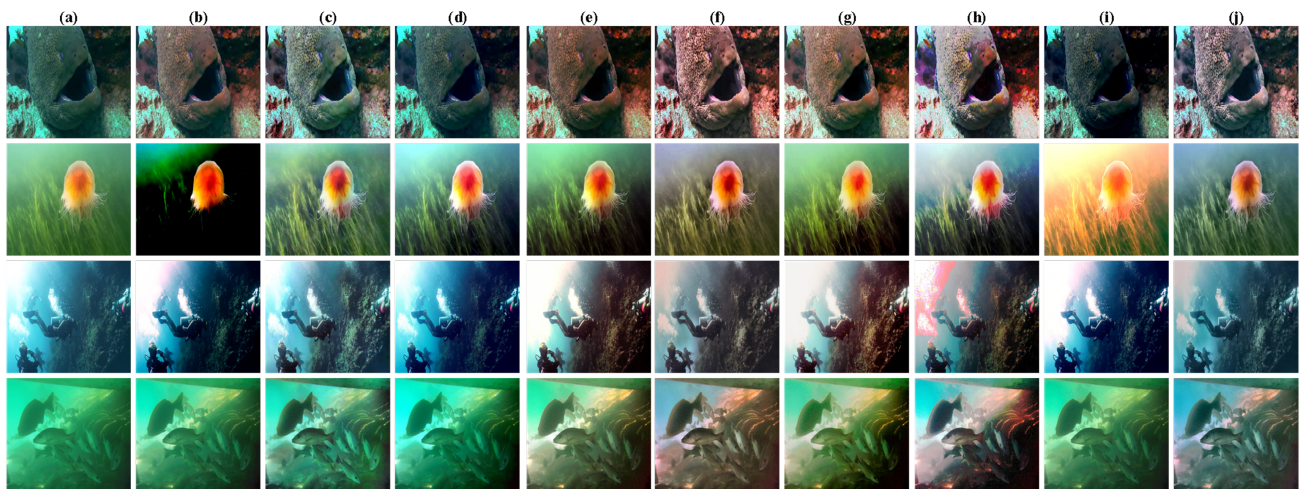


Figure 9. Comparison of enhancement results of 60 challenging UIEB images: (a) Raw. (b) IBLA [19]. (c) DCP [16]. (d) Retinex [10]. (e) Shallow [33]. (f) WaterNet [24]. (g) UGAN [30]. (h) Deepwave [26]. (i) FunieGan [34]. (j) Ours.

4.2.2. Quantitative Evaluation

To conduct a comprehensive quantitative analysis, we utilized standard full-reference metrics (PSNR, SSIM, and RMSE [35]) and non-reference metrics (UCIQE [36] and UIQM [37]) that are widely applied in underwater image quality assessments.

Table 1 shows a comparison with eight other methods. It demonstrates that our method, which utilizes physical models and mathematical geometric properties, achieves the highest scores in the PSNR, SSIM, RMSE, and UICQE metrics. Although our method did not achieve optimal solutions in UIQM, the subjective evaluations presented in Figures 7–9 support our findings. These findings indicate that the proposed algorithm outperforms the tested methods, resulting in significantly improved image quality compared to other algorithms.

Table 1. Comparison results of MSFE-UIENet with other SOTA methods on UIEB.

Method	PSNR ↑	SSIM ↑	RMSE ↓	UIQM ↑	UICQE ↑
Input	18.90	0.66	19.55	3.21	0.55
IBLA [19]	16.67	0.59	24.82	5.15	0.62
DCP [16]	14.66	0.58	24.73	3.04	0.59
Retinex [10]	18.08	0.62	20.59	3.22	0.54
Shallow [33]	19.74	0.71	17.02	4.20	0.53
WaterNet [24]	20.15	0.79	16.24	4.31	0.54
UGan [30]	21.15	0.72	15.85	4.63	0.62
Deepwave [26]	16.72	0.56	24.72	4.00	0.60
FUnIE-GAN [34]	19.27	0.71	17.83	4.76	0.62
Ours	25.85	0.88	9.75	4.33	0.63

The best score is highlighted in blue and the second highest score is highlighted in green. "↑" means that the higher value, the better. "↓" means that the lower value, the better.

Owing to the absence of reference truth images for the 60 challenging images within the UIEB dataset, we were restricted to the use of non-reference metrics, UCIQE and UIQM. As delineated in Table 2, our method continued to yield commendable results for these 60 challenging images.

Table 2. Comparison results of 60 challenging UIEB images.

Method	UIQM ↑	UICQE ↑	Method	UIQM ↑	UICQE ↑
Input	3.04	0.48	WaterNet [24]	4.13	0.53
IBLA [19]	4.10	0.57	UGan [30]	4.21	0.55
DCP [16]	3.54	0.54	Deepwave [26]	3.98	0.54
Retinex [10]	4.08	0.53	FunieGan [34]	4.24	0.53
Shallow [33]	4.02	0.49	Ours	4.27	0.56

The best score is highlighted in blue and the second highest score is highlighted in green. "↑" means that the higher value, the better.

4.2.3. Application Test

Additionally, we conducted a comparative analysis of the Canny edge detection and scale-invariant feature transform (SIFT) detection of the enhancement results using different methods, underscoring the superiority of our approach.

The Canny edge detection algorithm [38], a prevalent tool in image processing, identifies edges or contours in images. The edges detected by Canny were relatively fine and clear, which aided the subsequent analyses. As illustrated in Figure 10, the enhanced images extract more edge information than the unprocessed images. Figure 10 demonstrates that, under identical parameter conditions, our method extracts more coherent edge information from the processed image, significantly surpassing the performance of alternative algorithms. Our method significantly outperforms the other algorithms.

Furthermore, to visually substantiate the efficacy of Canny edge detection, we utilized the edge detection results from standard ground-truth images to compute the true positive rate (TPR) and false positive rate (FPR). TPR represents the ratio of accurately detected edges to the total real edges, whereas FPR reflects the ratio of pixels erroneously identified as edges to all non-edge pixels. These two metrics offer insightful data on edge detection performance, as detailed in the subsequent Table 3. Our method achieved the highest true positive rate (TPR) value for Canny edge detection. Despite not securing the best false positive rate (FPR), it still ranked among the top performers.

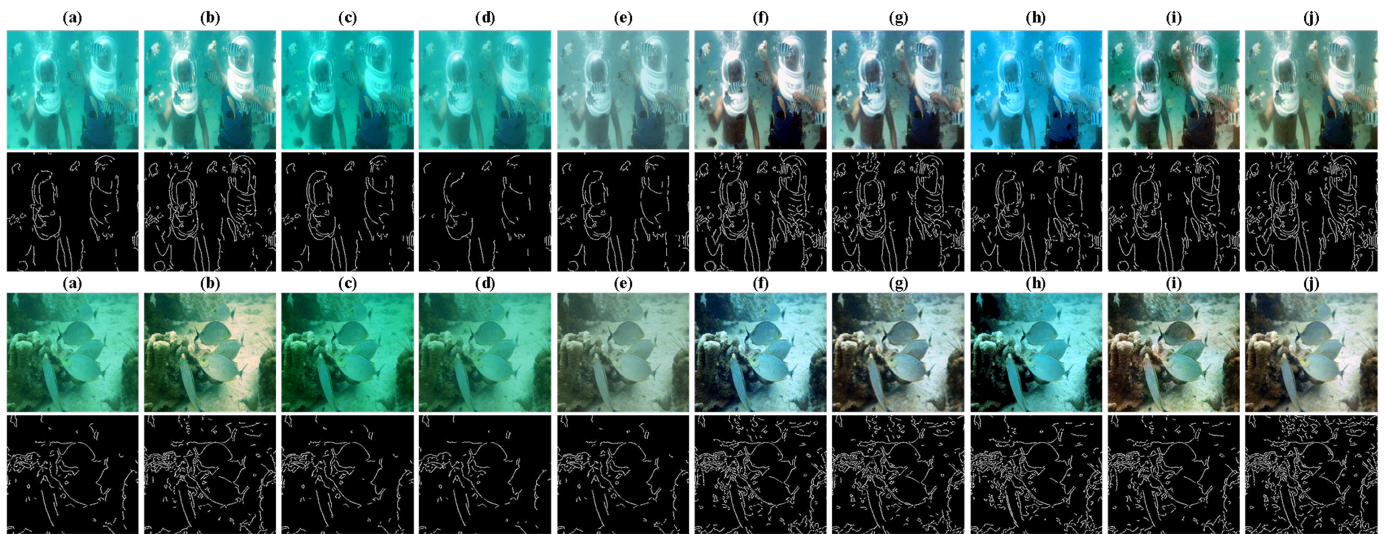


Figure 10. Comparison of results for Canny edge detection: (a) Raw. (b) IBLA [19]. (c) DCP [16]. (d) Retinex [10]. (e) Shallow [33]. (f) WaterNet [24]. (g) UGAN [30]. (h) Deepwave [26]. (i) FunieGan [34]. (j) Ours.

Table 3. Comparison results of Canny edge detection quantification.

Pic1	TPR ↑	FPR ↓	Pic2	TPR ↑	FPR ↓
Input	0.38	0.07	Input	0.38	0.08
IBLA [19]	0.56	0.11	IBLA [19]	0.52	0.12
DCP [16]	0.39	0.07	DCP [16]	0.38	0.09
Retinex [10]	0.24	0.05	Retinex [10]	0.32	0.07
Shallow [33]	0.39	0.08	Shallow [33]	0.39	0.08
WaterNet [24]	0.63	0.14	WaterNet [24]	0.54	0.18
UGan [30]	0.55	0.17	UGan [30]	0.51	0.19
Deepwave [26]	0.50	0.11	Deepwave [26]	0.52	0.17
FUnIE-GAN [34]	0.60	0.17	FUnIE-GAN [34]	0.54	0.23
Ours	0.66	0.09	Ours	0.58	0.10

The best score is highlighted in blue and the second highest score is highlighted in green. "↑" means that the higher value, the better. "↓" means that the lower value, the better.

SIFT [39] is a tool for detecting and describing local features in images. The detection and matching of SIFT keypoints are widely used in image mosaics, target tracking, and iterative reconstruction. As shown in Figure 11, our method's enhancement of the underwater imagery yielded a larger number of keypoints compared to the original image. Despite our method not detecting the highest number of SIFT feature points, it outperformed competing algorithms. As shown in Table 4, methods such as UGAN [30] and FunieGan [34], while capable of detecting more SIFT feature points than others, significantly underperform in SIFT matching. The results suggest that our method enhances the visual perceptual quality of underwater images and retains their semantic information.

Table 4. Comparison results of SIFT quantification.

Method	Points	Pairs	Accuracy	Method	Points	Pairs	Accuracy
Input	628	367	0.58	WaterNet [24]	882	497	0.56
IBLA [19]	757	408	0.54	UGan [30]	930	324	0.35
DCP [16]	592	317	0.54	Deepwave [26]	710	222	0.31
Retinex [10]	772	364	0.47	FunieGan [34]	959	442	0.46
Shallow [33]	659	424	0.64	Ours	894	580	0.68

The best score is highlighted in blue and the second highest score is highlighted in green.

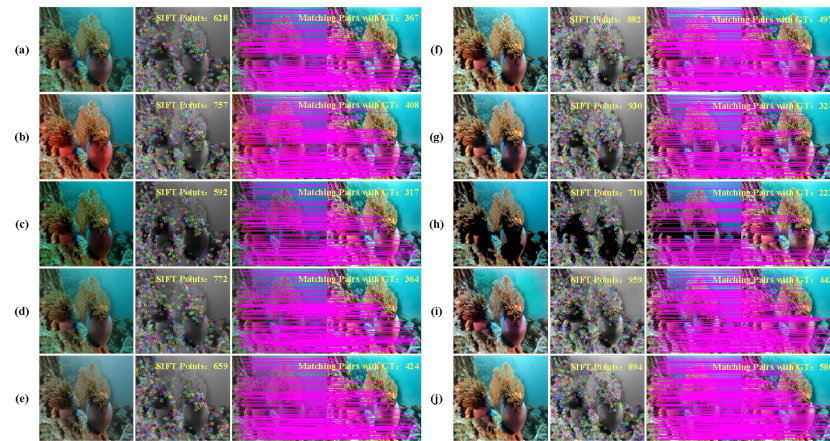


Figure 11. Comparison of results for detection and matching of sift key points: (a) Raw. (b) IBLA [19]. (c) DCP [16]. (d) Retinex [10]. (e) Shallow [33]. (f) WaterNet [24]. (g) UGAN [30]. (h) Deepwave [26]. (i) FunieGan [34]. (j) Ours.

4.3. Ablation Study

This study introduces forward and backward branches to augment the gradient flow of the network and improve the convergence speed and stability of the network. Specifically, the forward and backward branches were individually removed, whereas the remaining components were retained. Subsequently, an analysis and statistical evaluation of the objective results post-training were conducted.

As shown in Table 5, the five objective metrics for both the forward and backward branches were below the final outcome. This observation underscores the significant contribution of the forward and backward branch strategies proposed in this study for performance enhancement.

Table 5. Comparison results of the ablation study.

Method	PSNR $\uparrow$	SSIM $\uparrow$	RMSE $\downarrow$	UIQM $\uparrow$	UICQE $\uparrow$
without FCM	23.04	0.62	21.65	4.21	0.50
without BCM	24.46	0.73	10.71	4.30	0.62
ALL	25.85	0.88	9.75	4.33	0.63

" $\uparrow$ " means that the higher value, the better. " $\downarrow$ " means that the lower value, the better.

To ascertain the optimal values for the coefficients k_1 and k_2 in the loss function, we randomly selected 25% of the training set for ablation experiments and assessed various combinations of k_1 and k_2 values. As indicated in Table 6, the experiment yielded the most favorable results when k_1 approximated 0.8 and k_2 approximated 0.2. Drawing on these experimental results, combining past experimental training experience, we established k_1 and k_2 at 0.8 and 0.2, respectively.

Table 6. Comparison results of different k_1 and k_2 combinations.

k_1 k_2	$k_2 = 0.0$	$k_2 = 0.2$	$k_2 = 0.4$	$k_2 = 0.6$	$k_2 = 0.8$	$k_2 = 1.0$
$k_1 = 0.0$	(22.73, 0.62)	(23.02, 0.66)	(22.51, 0.61)	(22.42, 0.59)	(22.31, 0.60)	(22.11, 0.58)
$k_1 = 0.2$	(22.82, 0.63)	(23.09, 0.65)	(23.10, 0.63)	(22.98, 0.61)	(22.87, 0.62)	(22.79, 0.60)
$k_1 = 0.4$	(23.23, 0.74)	(23.42, 0.75)	(23.31, 0.74)	(23.22, 0.72)	(23.20, 0.71)	(23.08, 0.69)
$k_1 = 0.6$	(23.81, 0.78)	(23.89, 0.81)	(23.13, 0.80)	(23.71, 0.77)	(23.67, 0.75)	(23.54, 0.73)
$k_1 = 0.8$	(24.23, 0.82)	(24.96, 0.85)	(24.81, 0.83)	(24.73, 0.81)	(24.42, 0.79)	(24.05, 0.80)
$k_1 = 1.0$	(24.11, 0.80)	(24.67, 0.81)	(24.45, 0.80)	(24.36, 0.75)	(24.13, 0.76)	(24.08, 0.73)

The values in cell are (PSNR, SSIM). The best score is highlighted in blue.

5. Conclusions

To solve the problem of underwater image distortion, this study presents a deep-learning-driven UIE algorithm designed to augment the quality of images in underwater

environments. This study presents a high-performance UIE network named MSFE-UIENet, which utilizes the encoder–decoder architecture to augment image enhancement. To address the issue of subpar enhancement outcomes attributed to the sole downsampling method in traditional enhancement networks, this study introduces a pyramid downsampling module that employs multi-scale downsampling to extract richer image features. Subsequently, to enhance the feature extraction capability of the network, a high-performance FEM was introduced, enabling the more accurate capture of detailed information in underwater images. Finally, to optimize the network's gradient flow, forward and backward branches were implemented to enhance the network's convergence speed and stability. Experimental validation of the underwater image datasets confirmed that the proposed network can effectively enhance the quality of underwater images, exhibiting improved image detail preservation and noise suppression effects in diverse underwater scenarios.

Author Contributions: Conceptualization, software, and methodology, S.Z. and X.M.; data curation, S.Z. and X.M.; writing—original draft preparation, S.Z. and X.M.; writing—review and editing, X.Y. and S.G.; visualization, S.Z. and X.M.; supervision, project administration, and funding acquisition, X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (Grant No. 42276187).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in UIEB at https://li-chongyi.github.io/proj_benchmark.html, accessed on 6 May 2024.

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

References

1. Wang, J.; Ye, X.; Mei, X.; Wei, X. Learning mapping by curve iteration estimation for real-time underwater image enhancement. *Opt. Express* **2024**, *32*, 9931–9945. [CrossRef] [PubMed]
2. Bertolotti, J.; Van Putten, E.G.; Blum, C.; Lagendijk, A.; Vos, W.L.; Mosk, A.P. Non-invasive imaging through opaque scattering layers. *Nature* **2012**, *491*, 232–234. [CrossRef] [PubMed]
3. Cecconi, V.; Kumar, V.; Bertolotti, J.; Peters, L.; Cutrona, A.; Olivieri, L.; Peccianti, M. Terahertz spatiotemporal wave synthesis in random systems. *ACS Photonics* **2024**, *11*, 362–368. [CrossRef] [PubMed]
4. Vellekoop, I.M.; Mosk, A.P. Focusing coherent light through opaque strongly scattering media. *Opt. Lett.* **2007**, *32*, 2309–2311. [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. Hu, K.; Weng, C.; Zhang, Y.; Jin, J.; Xia, Q. An overview of underwater vision enhancement: From traditional methods to recent deep learning. *J. Mar. Sci. Eng.* **2022**, *10*, 241. [CrossRef]
7. Wei, X.; Ye, X.; Mei, X.; Wang, J.; Ma, H. Enforcing high frequency enhancement in deep networks for simultaneous depth estimation and dehazing. *Appl. Soft Comput.* **2024**, *163*, 11873. [CrossRef]
8. 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 2010 IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10–13 October 2010; pp. 1703–1709.
9. Hitam, M.S.; Awalludin, E.A.; Yussof, W.N.J.H.W.; Bachok, Z. Mixture contrast limited adaptive histogram equalization for underwater image enhancement. In Proceedings of the 2013 International Conference on Computer Applications Technology (ICCAT), Sousse, Tunisia, 20–22 January 2013; pp. 1–5.
10. Fu, X.; Zhuang, P.; Huang, Y.; Liao, Y.; Zhang, X.P.; Ding, X. A retinex-based enhancing approach for single underwater image. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 4572–4576.
11. Zhang, S.; Wang, T.; Dong, J.; Yu, H. Underwater image enhancement via extended multi-scale Retinex. *Neurocomputing* **2017**, *245*, 1–9. [CrossRef]
12. Li, M.; Liu, J.; Yang, W.; Sun, X.; Guo, Z. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Trans. Image Process.* **2018**, *27*, 2828–2841. [CrossRef]
13. Ancuti, C.; Ancuti, C.O.; Haber, T.; Bekaert, P. Enhancing underwater images and videos by fusion. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 81–88.

14. Gao, S.B.; Zhang, M.; Zhao, Q.; Zhang, X.S.; Li, Y.J. Underwater image enhancement using adaptive retinal mechanisms. *IEEE Trans. Image Process.* **2019**, *28*, 5580–5595. [[CrossRef](#)] [[PubMed](#)]
15. Yuan, J.; Cao, W.; Cai, Z.; Su, B. An underwater image vision enhancement algorithm based on contour bougie morphology. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 8117–8128. [[CrossRef](#)]
16. 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](#)]
17. Drews, P.L.; Nascimento, E.R.; Botelho, S.S.; Campos, M.F.M. Underwater depth estimation and image restoration based on single images. *IEEE Comput. Graph. Appl.* **2016**, *36*, 24–35. [[CrossRef](#)] [[PubMed](#)]
18. Peng, Y.T.; Cao, K.; Cosman, P.C. Generalization of the dark channel prior for single image restoration. *IEEE Trans. Image Process.* **2018**, *27*, 2856–2868. [[CrossRef](#)] [[PubMed](#)]
19. Peng, Y.T.; Cosman, P.C. Underwater image restoration based on image blurriness and light absorption. *IEEE Trans. Image Process.* **2017**, *26*, 1579–1594. [[CrossRef](#)] [[PubMed](#)]
20. Galdran, A.; Pardo, D.; Picón, A.; Alvarez-Gila, A. Automatic red-channel underwater image restoration. *J. Vis. Commun. Image Represent.* **2015**, *26*, 132–145. [[CrossRef](#)]
21. 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](#)]
22. Mei, X.; Ye, X.; Wang, J.; Wang, X.; Huang, H.; Liu, Y.; Jia, Y.; Zhao, S. UIEOGP: An underwater image enhancement method based on optical geometric properties. *Opt. Express* **2023**, *31*, 36638–36655. [[CrossRef](#)] [[PubMed](#)]
23. Wang, Y.; Guo, J.; Gao, H.; Yue, H. UIEC²-Net: CNN-based underwater image enhancement using two color space. *Signal Process. Image Commun.* **2021**, *96*, 116250. [[CrossRef](#)]
24. 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](#)]
25. Li, C.; Anwar, S.; Porikli, F. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognit.* **2020**, *98*, 107038. [[CrossRef](#)]
26. Sharma, P.; Bisht, I.; Sur, A. Wavelength-based attributed deep neural network for underwater image restoration. *ACM Trans. Multimed. Comput. Commun. Appl.* **2023**, *19*, 1–23. [[CrossRef](#)]
27. 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](#)]
28. Mei, X.; Ye, X.; Zhang, X.; Liu, Y.; Wang, J.; Hou, J.; Wang, X. UIR-Net: A Simple and Effective Baseline for Underwater Image Restoration and Enhancement. *Remote Sens.* **2022**, *15*, 39. [[CrossRef](#)]
29. Wang, J.; Ye, X.; Liu, Y.; Mei, X.; Hou, J. Underwater self-supervised monocular depth estimation and its application in image enhancement. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105846. [[CrossRef](#)]
30. 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.
31. Pramanick, A.; Sarma, S.; Sur, A. X-CAUNET: Cross-Color Channel Attention with Underwater Image-Enhancing Transformer. In Proceedings of the ICASSP 2024—2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Republic of Korea, 14–19 April 2024; pp. 3550–3554.
32. Huang, S.; Wang, K.; Liu, H.; Chen, J.; Li, Y. Contrastive semi-supervised learning for underwater image restoration via reliable bank. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 18145–18155.
33. Naik, A.; Swarnakar, A.; Mittal, K. Shallow-uwnet: Compressed model for underwater image enhancement (student abstract). *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 15853–15854. [[CrossRef](#)]
34. Islam, M.J.; Xia, Y.; Sattar, J. Fast underwater image enhancement for improved visual perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [[CrossRef](#)]
35. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]
36. Yang, M.; Sowmya, A. An underwater color image quality evaluation metric. *IEEE Trans. Image Process.* **2015**, *24*, 6062–6071. [[CrossRef](#)]
37. Panetta, K.; Gao, C.; Agaian, S. Human-visual-system-inspired underwater image quality measures. *IEEE J. Ocean. Eng.* **2015**, *41*, 541–551. [[CrossRef](#)]
38. Bao, P.; Zhang, L.; Wu, X. Canny edge detection enhancement by scale multiplication. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1485–1490. [[CrossRef](#)] [[PubMed](#)]
39. Wang, F.; You, H.; Fu, X. Adapted anisotropic Gaussian SIFT matching strategy for SAR registration. *IEEE Geosci. Remote Sens. Lett.* **2014**, *12*, 160–164. [[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.