electronics

Article

DeepSeaNet: An Efficient UIE Deep Network

Jingsheng Li [1], Yuanbing Ouyang [2], Hao Wang [2], Di Wu [3,*] and Yushan Pan [4,*]

[1] School of Artificial Intelligence, Xidian University, Xi'an 710126, China; jingsheng.li@stu.xidian.edu.cn
[2] School of Cyber Engineering, Xidian University, Xi'an 710126, China; yuanbing.oy@stu.xidian.edu.cn (Y.O.); wanghao@xidian.edu.cn (H.W.)
[3] Department of ICT and Natural Sciences, Norwegian University of Science and Technology, 6009 Aalesund, Norway
[4] Department of Computing, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China
[*] Correspondence: di.wu@ntnu.no (D.W.); yushan.pan@xjtlu.edu.cn (Y.P.); Tel.: +47-70161647 (D.W.); +86-0512-89165347 (Y.P.)

Abstract: Underwater image enhancement and object recognition are crucial in multiple fields, like marine biology, archeology, and environmental monitoring, but face severe challenges due to low light, color distortion, and reduced contrast in underwater environments. DeepSeaNet re-evaluates the model guidance strategy from multiple dimensions, enhances color recovery using the MCOLE score, and addresses the problem of inconsistent attenuation across different regions of underwater images by integrating a feature extraction method guided by a global attention mechanism by ViT. Comprehensive tests on diverse underwater datasets show that DeepSeaNet achieves a maximum PSNR of 28.96 dB and an average SSIM of 0.901, representing a 20–40% improvement over baseline methods. These results highlight DeepSeaNet's superior performance in enhancing image clarity, color richness, and contrast, making it a remarkably effective instrument for underwater image processing and analysis.

Keywords: underwater environment; image enhancement; UDnet; ViT

Received: 10 May 2025
Revised: 5 June 2025
Accepted: 10 June 2025
Published: 12 June 2025

Citation: Li, J.; Ouyang, Y.; Wang, H.; Wu, D.; Pan, Y. DeepSeaNet: An Efficient UIE Deep Network. *Electronics* **2025**, *14*, 2411. https://doi.org/10.3390/electronics14122411

1. Introduction

Improving and recognizing underwater images is vital for diverse underwater exploration and research activities, including underwater robotics, marine studies, resource discovery, and archeological investigations [1,2]. The underwater environment degrades image quality due to water absorption and scattering, reducing contrast, color, and sharpness [3]. This poses a challenge to traditional processing methods. Underwater image enhancement (UIE) [4,5] aims to restore color, sharpness, and detail, thereby enhancing the precision of tasks like object recognition [6] and trajectory planning.

Key challenges in UIE are light refraction and color distortion, leading to blue/green hues and changed contrast. Enhancement improves color realism and object recognition, crucial for underwater tasks. Integrating enhancement with recognition boosts system robustness and automation, allowing underwater robots [7] to navigate and work autonomously in complex scenarios, thus enhancing efficiency and application scope.

Underwater images often suffer from poor quality, manifested as color distortion, low contrast, and structural degradation such as blurred details. This degradation primarily stems from the absorption and scattering of light caused by impurities in water. Crucially, the attenuation of light underwater is uneven across different color channels and spatial regions. For instance, red light typically attenuates faster than green and blue light. This uneven attenuation is the fundamental cause of the observed color distortion and loss of image detail.

Several studies have explored the challenges of UIE [8–15]. UDnet [13] is an unsupervised UIE framework that combines U-Net and PAdaIN [16], employing a convolutional neural network as well as Probabilistic Adaptive Instance Normalization (PAdaIN). The framework is able to transform global enhancement statistics, encode uncertainty, and introduce a multi-color spatial stretching method based on the guidance of multi-scale statistical information as a way to enhance contrast and optimize color performance. Nonetheless, there are still areas for enhancement regarding color and visualization of UDnet-enhanced images compared to the reference image. DewaterNet [17] is an advanced underwater image enhancement network that improves image quality by integrating Generative Adversarial Networks (GANs) [18] and multi-term objective functions. This approach achieves good results in color correction and contrast enhancement, effectively addressing common color distortion issues in underwater images and enhancing their visual contrast. Nonetheless, U-Net-based models do not adequately enhance parts of underwater images that suffer from severe attenuation across color channels and spatial regions. Compared to reference images, there is still room for improvement in terms of the color and visualization of the enhanced images by UDnet. It may not fully achieve the desired level of naturalness and clarity when processing images with rich colors and complex structures. The U-Shape Transformer [19] integrates ViT and specially designed multiscale feature fusion transformers and global feature modeling transformers for the UIE task. This approach demonstrates richer color representation and higher accuracy when dealing with underwater images that suffer from inconsistent attenuation across different color channels and spatial regions. Moreover, it provides more ideal detail recovery and rendering effects in the processing of complex structures. Despite this, the model still has limitations in terms of color and semantic structure for underwater images under specific scenarios.

To address the structural issues of color distortion, low contrast, and blurred details, we introduce a bootstrap evaluation framework that evaluates underwater image quality through multi-branch cooperative learning mechanisms. We incorporate the pre-trained MCOLE [20] model to process the original images for bootstrapping training, ensuring that the generated enhanced images exhibit greater similarity to reference images in terms of color accuracy and visual clarity.

To further tackle the issue of inconsistent attenuation across different color channels and spatial regions in underwater images, we integrate a visual transformer (ViT) [21] in the feature extraction module. The ViT segments the image into fixed-size blocks and treats these blocks as continuous inputs to model the global information through the self-attentive mechanism. Unlike U-Net networks and GANs, it can effectively capture long-range dependencies and global features in images and better address the issue of inconsistent attenuation in different regions of underwater images. We derive hierarchical features through parallel analysis of distinct quality attributes such as chromatic properties and structural visibility to improve assessment accuracy and interpretability. The method enables an objective assessment of enhancement techniques and helps to accomplish tasks such as underwater image defogging, color correction, and contrast enhancement. Our main contributions are as follows:

Color-guided evaluation construction: An innovative loss function framework is proposed to enhance image quality by synergistically combining three distinct components: MCOLE-driven perceptual loss, color space constraints, and structural feature MSE. This method addresses the limitations of traditional VGG-16 [22] perceptual loss and MSE by optimizing color and structural similarity in a unified manner. Experimental evaluations show that this multi-faceted approach achieves advanced results in PSNR and SSIM, outperforming baseline methods by a significant margin.

Optimizing the design of feature extraction structure: ViT is innovatively integrated into the feature extraction module, replacing the conventional method and accurately capturing global features. This approach effectively addresses the challenge of inconsistent attenuation in different areas of underwater images. Additionally, the multiscale feature fusion strategy is employed, allowing the model to thoroughly extract image features from various levels and scales. As a result, the model can acquire image information more comprehensively, significantly improving its adaptability to complex scenes. In non-referenced evaluation, this feature extraction structure enables the model to deliver outstanding performance, effectively enhancing the image's clarity, color richness, and contrast.

2. Related Work

Although difficult, UIE is a rewarding endeavor that addresses issues such as color imbalance, diminished contrast, reduced brightness, and increased noise levels [3]. There are three main approaches to tackling these problems: model-free methods, deep learning-based methods, and probability-based methods.

2.1. Model-Free Methods

Model-free methods optimize underwater imagery through pixel-level manipulation paradigms, bypassing dependency on preformulated computational frameworks while maintaining photometric integrity through direct signal transformation operations. The SPDF framework [23] generates two complementary versions of an image—one with corrected contrast and the other with sharpened details—through preprocessing. These are divided into three separate elements: mean intensity, contrast, and structure, which are then fused in a perceptually aware image space and integrated through the inverse decomposition process to rebuild the enhanced image. The MLLE method [24] dynamically enhances contrast by locally modifying color and fine details through the computation of local block statistics (mean and variance). It introduces a strategy for color equilibrium in the CIELAB color space, significantly improving color vividness, contrast, and detail. TOPAL [25] enhances visual contrast and performs color correction using multiscale dense boosting and advanced aesthetic rendering modules, and integrates details with a dual-channel attention module. This approach utilizes a multiscale adversarial framework to reduce discrepancies between synthetic and authentic visual data, integrating perceptual cues to enhance scene understanding.

Model-free UIE methods are efficient in computation and easy to integrate, thereby enabling live processing and environments with limited hardware resources. They do not rely on prior knowledge, offering better generalizability and lightweight characteristics, which allow for rapid deployment and effective image enhancement. However, these methods have limited adaptability and struggle to handle complex underwater environments and dynamic scenes. They are prone to over-enhancement and lack global consistency, and they cannot effectively model the uncertainties in the underwater imaging process, leading to instability in complex scenes. Thus, those methods are only applied for image enhancement after generating three reference images in our approach, and it introduces uncertainty to help the resulting virtual reference images more closely approximate the real reference images.

2.2. Deep Learning-Based Methods

Deep learning-based methods improve UIE by learning from training datasets. FloodNet [26] employs a multiscale feature fusion and enhancement method to achieve feature extraction and hierarchical fusion. It utilizes adaptive local–global residual learning to generate high-quality restored images. ADMNNet [27], with its attention-guided dynamic multi-branch structure, overcomes the limitations of traditional convolutional neural networks by incorporating attention mechanisms and dynamic fusion of multiscale features. It employs a dynamic feature optimization method to enhance feature representation by adjusting receptive field sizes and channel attention. WaveNet [28] dynamically adjusts receptive field sizes based on color channel propagation and introduces an attention-based skip mechanism for better performance. LiteEnhanceNet [14], a lightweight network, reduces computational complexity with depthwise separable convolution and one-shot aggregation connections while maintaining high image enhancement performance through activation functions and a squeeze-and-excitation module. SNR-Net [29] presents a dual-branch approach for UIE by merging transformer models, which are based on the Signal-to-Noise Ratio (SNR), with convolutional networks. By dynamically enhancing pixel quality and strengthening multiscale feature perception, it effectively improves color imbalance, underexposure, and blurriness in underwater images. CE-CGAN [30] enhances image quality by using a generator to map input image features to high-contrast images and employing a discriminator to classify both generated and real images. UDAformer [31] is a method for UIE that utilizes a dual-attention Transformer. It efficiently encodes and decodes underwater image features through a dual-attention feature encoding and decoding method. The method uses residual connections to restore underwater images, significantly improving the enhancement results. Phaseformer [32] proposes a lightweight phase-based Transformer framework for underwater image reconstruction. The method extracts clean features through a phase self-attention mechanism and restores structural information by propagating salient features using an optimized phase attention module.

These methods leverage the powerful capabilities of convolutional neural networks (CNNs) [33], GANs, and Transformers [34] to adapt to various underwater environments. Substantial advancements have been achieved in enhancing underwater images through deep learning-based approaches. However, they face challenges such as high data dependency, limited generalization ability, high computational complexity, and insufficient modeling of the underwater imaging process. Although Transformers demonstrate notable strengths in capturing long-range dependencies and enabling parallel computation, enabling them to better capture global features in underwater images, they also have drawbacks such as elevated computational resource utilization, long training times, and a high demand for large-scale data. These issues limit their practical application in complex and dynamic underwater environments. Therefore, in subsequent methods, we skillfully introduce uncertainty into the U-Net-based network and employ ViT to capture global features, thereby simulating diverse underwater environments and enhancing the model's generalization ability.

2.3. Probabilistic-Based Methods

Probability-based deep learning methods, especially Conditional Variational Autoencoders (cVAEs) [35], incorporate uncertainty modeling to manage perturbations, modeling errors and inherent uncertainties in underwater environments. These methods utilize Variational Autoencoders (VAEs) [36], which map input data to a compact feature space and restore it through a decoding process. Unlike traditional encoders, VAEs model the probability distributions of latent variables, enabling better handling of diverse data characteristics. During training, regularization and reconstruction losses ensure effective data representation by minimizing discrepancies between posterior and prior distributions.

Recent research, such as PUIE-Net [16], combines cVAEs with adaptive instance normalization to formulate an improvement distribution for degraded underwater images.

By utilizing a consensus approach to anticipate predictable conclusions, it addresses the ambiguity of reference maps and reduces bias. This method improves adaptability to labeling biases while maintaining result stability. Experimental results demonstrate its competitive performance on multiple real-world datasets. Building on this, UDNet [13] establishes an end-to-end framework that synergistically combines an adaptive uncertainty quantification module with a stochastic reference sample selection strategy during training, systematically enhancing cross-domain generalization performance in visual computing systems.

Probabilistic methods, such as cVAEs and VAEs, effectively handle disturbances and uncertainties in underwater environments by integrating uncertainty modeling. These methods capture diverse data characteristics through probability distributions, generate varied enhanced results, and improve model generalization via techniques like PAdaIN. However, their training process is complex, requiring optimization of reconstruction loss and regularization terms, which may lead to unstable training. In the subsequent methods, we optimized the loss functions and regularization terms of the probabilistic methods, making them more stable and efficient.

3. Methods

To alleviate the inconsistency of the model in color attenuation regions, we have integrated ViT as the feature extractor into the UDnet framework. ViT is capable of effectively capturing the characteristics of spatial regions in images, thereby enabling the entire network to better handle complex underwater scenes and providing a more robust foundation for multi-color space stretching and uncertainty modeling. Meanwhile, to enhance the capacity of the model to perceive images, we have incorporated the MCOLE score module into the loss function. The MCOLE score module can comprehensively evaluate the enhanced images in terms of color and visual features, thereby further optimizing the model's performance. We present the overall architecture in Figure 1.

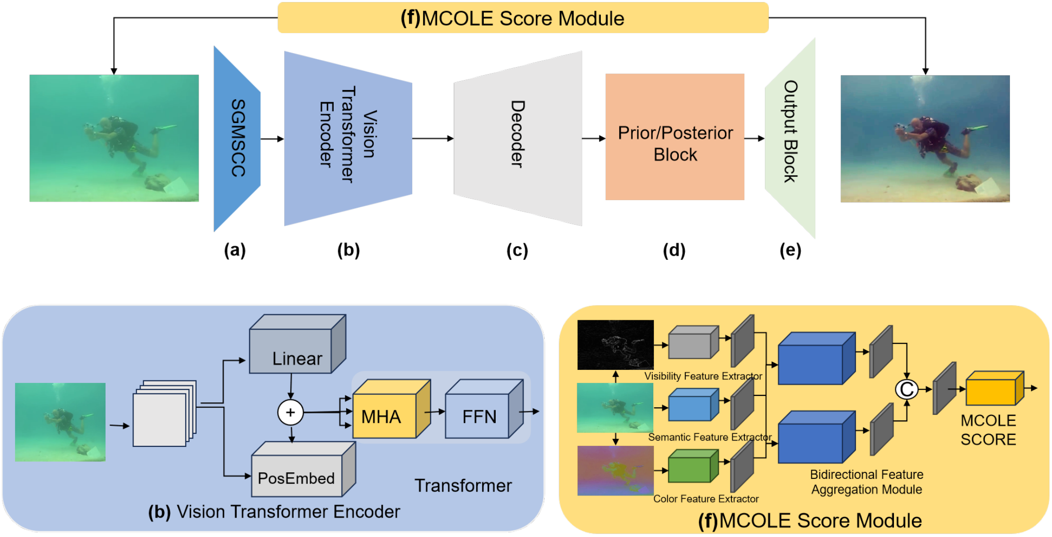

Figure 1. Framework: the figure illustrates the overall architecture of our model, where (**a**) represents the Statistically Guided Multi-Color Space Stretch (SGMCSS), (**b**) represents the feature extraction, (**c**) represents the decoder, (**d**) represents the Prior/Posterior Block, (**e**) represents the output block and (**f**) represent the loss function, respectively.

3.1. Network Structure

The original image first goes through the Statistically Guided Multi-Color Space Stretch (SGMCSS) (a) to generate a reference image. Then the ViT (b) is used to replace the initial feature descriptors of UDnet, allowing the network to more effectively capture global information and increase its focus on areas with low visibility or significant degradation. The input feature map has dimensions of $H \times W \times C$. For the sequence compatible with transformer architectures, the feature map of the input data is decomposed into a sequence of flattened 2D patches $\{x_i\}_{i=1}^{N} \in \mathbb{R}^{L \times D}$. Here, L refers to the patch resolution, while N stands for the number of patches obtained. To retain the valuable positional details of each area, learnable position embeddings are directly incorporated, which can be formulated as follows:

$$X = \text{Linear}(x_i) + \text{PosEmbed}(x_i), \tag{1}$$

where Linear denotes a linear projection process, and PosEmbed signifies a position embedding process.

The feature sequence X is introduced into the transformer encoder block, which is composed of L transformer encoder layers. The transformer encoder layer processes the sequence through multi-head self-attention (MSA) and multi-layer perceptron (MLP) blocks,

$$\begin{aligned} S_l' &= \text{MHA}(\text{LN}(S_{l-1})) + S_{l-1} \\ S_l &= \text{FFN}(\text{LN}(S_l')) + S_l' \end{aligned} \tag{2}$$

This formulation employs LN to represent layer normalization operations, with S_l corresponding to the feature sequence generated from the l-th transformer layer's computational process.

The final transformer block outputs a feature sequence S_{output}, which contains spatial and color information. This sequence is passed into the prior and posterior processing module. After passing through the decoder (c), the sequence is then transferred to the Prior/Posterior Block (d). This module aims to calculate the mean and standard deviation distributions, which helps determine the potential enhancements for the image. Finally, the enhanced image is output through the Output Block (e).

3.2. MCOLE Score Module

The MCOLE (f) method enhances underwater image quality through multi-level feature fusion. It first extracts three key features—color (YCbCr), structural visibility (Gradient), and semantic (RGB) features—from the image using different layers of the VGG-11 [22] network. These features are then fused via the Bidirectional Feature Aggregation Module (BFAM), which includes the Global Context Interaction Module (GCIM) and the Bidirectional Visual Fusion Module (BVFM) for feature compression, fusion, and aggregation.

MCOLE relies on VGG-11 to extract color, structural, and semantic features. The YCbCr color space isolates luminance (Y) from chrominance (Cb, Cr), which enhances color accuracy. The gradient map highlights structural details, especially background elements that are often overlooked. Therefore, we first preprocess the image by converting RGB to YCbCr and generating a gradient map for feature extraction. In our implementation, we use the Scharr operator [37] to generate the gradient map, which effectively obtains the structural details of the image.

The YCbCr color space and gradient map are processed using VGG-11 as the backbone network to extract color features and structural visibility features. After optimization with a dataset, the color features $\mathcal{F}_c$, structural visibility features $\mathcal{F}_v$, and semantic features $\mathcal{F}_s$ can be obtained as follows:

$$\begin{aligned} \mathcal{F}_c(x) &= [fc(1), \ldots, fc(5)] \\ \mathcal{F}_v(g) &= [fv(1), \ldots, fv(5)] \\ \mathcal{F}_s(x) &= [fs(1), \ldots, fs(5)] \end{aligned} \tag{3}$$

Then the GCIM method adopts the following workflow after feature extraction: First, a 1 × 1 convolutional layer is applied to reduce the dimensionality of feature channels. Subsequently, the processed features undergo non-linear transformation through the Sigmoid activation function. This pipeline ultimately generates spatial attention-guided weight distribution maps that capture position-wise importance across feature representations. Specifically, the GCIM operation is articulated as follows:

$$f_{cs}(i) = FG(fc(i), fs(i)), \quad f_{vs}(i) = FG(fv(i), fs(i)) \tag{4}$$

Next, the BVFM progressively aggregates these fused features along both bottom–up and top–down paths, reducing computational burden and extracting key information:

$$\hat{f}_{cs}^{(1)} = FB(f_{cs}^{(2)}; f_{cs}^{(3)}), \quad \hat{f}_{cs}^{(2)} = FB(\hat{f}_{cs}^{(1)}; f_{cs}^{(4)}), \quad f_{cs}^{(up)} = FB(\hat{f}_{cs}^{(2)}; f_{cs}^{(5)}), \tag{5}$$

where FB denotes the BVFM, indicating the feature aggregation process using convolution, max-pooling, the position attention module (PAM) [38], and the channel attention module (CAM) [38] to enhance the features.

Finally, MCOLE concatenates the bottom–up and top–down features and passes them through three fully connected layers to yield the final quality score:

$$\hat{q} = F_Q(\text{Concat}(f_{cs}^{(up)}, f_{vs}^{(up)}, f_{cs}^{(dn)}, f_{vs}^{(dn)})), \tag{6}$$

where F_Q represents the quality prediction function. Through this multi-level, bidirectional feature aggregation and quality prediction process, MCOLE effectively enhances underwater image quality and provides accurate quality assessment.

3.3. Loss Function

The DeepSeaNet model is constructed based on the cVAEs, and its training process optimizes the variational lower bound to achieve feature learning. Compared with traditional deterministic enhancement models, the innovation of this system lies in the probabilistic modeling of enhancement statistics in the latent space. Specifically, a dual-channel posterior inference module is designed in the network architecture. Through deep feature analysis, the input image is mapped to a parameterized Gaussian distribution, and latent variables are sampled from the mean μ and standard deviation σ of this distribution for image reconstruction. Thus, to improve the enhancement performance of the model, the total loss function we designed consists of the enhancement loss and the Kullback–Leibler (KL) [39] divergence loss. The enhancement loss is composed of the MSE loss and the perceptual loss, while the KL divergence loss is made up of the KL divergence losses based on variance and mean.

In terms of loss function design, the system adopts a multi-dimensional error joint optimization strategy:

$$\mathcal{L}_e = \underbrace{\mathcal{L}_{Vmse}}_{\text{visual fidelity}} + \underbrace{\mathcal{L}_{Cmse}}_{\text{color preservation}} + \lambda \underbrace{\mathcal{L}_{MCOLE}}_{\text{perceptual optimization}} \tag{7}$$

The first two terms measure the structural similarity in the visible light band and the MSE in the color space. The innovative use of the MCOLE scoring mechanism to construct the perceptual loss involves calculating the score $\hat{q} \in (0, 1)$ of the enhanced image through a pre-trained quality evaluation network and establishing the negative log-likelihood loss

$$\mathcal{L}_{MCOLE} = -\log(\hat{q}) \tag{8}$$

This design guides the model to generate images with superior quality that conform to human visual perception through backpropagation. The hyperparameter λ employs a dynamic adjustment strategy to adaptively balance the relationship between pixel precision and perceptual quality in the training process.

In terms of probabilistic modeling, the system uses KL divergence to constrain the matching of latent variable distributions. The KL divergence loss based on variance and mean is as follows:

$$\mathcal{L}_m = D_{KL}(\mathcal{P}_m(x) \parallel \mathcal{Q}_m(y,x)), \quad \mathcal{L}_s = D_{KL}(\mathcal{P}_s(x) \parallel \mathcal{Q}_s(y,x)) \tag{9}$$

Here, $\mathcal{P}_m$ represents the mean and $\mathcal{P}_s$ represents the variance of the prior distribution, while $\mathcal{Q}_m$ and $\mathcal{Q}_s$ are the posterior estimates. This normalization strategy guarantees that the hidden space retains the critical features of the data while avoiding overfitting.

The final objective function integrates the enhancement loss and the distribution alignment term:

$$\mathcal{L}_{total} = \mathcal{L}_e + \beta(\mathcal{L}_m + \mathcal{L}_s) \tag{10}$$

The balance coefficient β is determined through grid search. This hybrid optimization mechanism enables the model to maintain its capability to restore details while effectively enhancing its adaptability to underwater optical distortions.

4. Experiments

This section elaborates on the experimental evaluation of our method. We begin by thoroughly presenting the datasets used, providing comprehensive insights into their composition and relevance. Following this, we delve into the evaluation metrics employed to gauge performance, alongside a detailed account of the implementation specifics that underpin the experimental setup.

4.1. Datasets

We evaluated our model's performance across six openly accessible datasets. For full-reference evaluation, we utilized UIEBD [40], EUVP [7], and UFO [15]. These benchmark datasets comprise paired collections of real-world underwater scenes with their ground-truth counterparts, providing a standardized benchmark for the quantitative assessment of enhancement outputs against their corresponding ground-truth counterparts. On the non-reference side, we employed DeepFish [41], RUIE [42], and SUIM [43]. DeepFish is built from underwater video screenshots, while RUIE and SUIM offer a wealth of authentic underwater photography. This experimental configuration facilitated reference-free evaluation of image restoration performance through standardized datasets containing non-reference-annotated visual samples, providing a comprehensive overview of our model's effectiveness in diverse scenarios. Paired images refer to those with ground-truth labels, whereas the remaining datasets consist only of unpaired images. In our experiments, we utilized solely UIEBD [40] for unsupervised learning (without ground-truth labels), with the remaining datasets reserved for performance assessment.

4.2. Evaluation Metrics

The enhanced images undergo comprehensive numerical evaluation using both full-reference and non-reference metrics. Three key full-reference metrics are the Peak Signal-to-Noise Ratio (PSNR) [44], the Structural Similarity Index (SSIM) [44], and the mean squared error (MSE). Higher values (for PSNR and SSIM) or lower values (for MSE) of these metrics indicate that the enhanced image closely resembles the ground-truth image, which is an important indicator of high-quality image restoration. Additionally, three non-reference metrics are employed: the Color Measurement Index (UICM) [45], the Sharpness Measurement Index (UISM) [45], and the Contrast Measurement Index (UIConM) [45]. Elevated scores in these indices suggest superior color balance, enhanced sharpness, and improved contrast in the image.

4.3. Implementation Details

The input resolution for our model's training is 224 × 224 pixels. To optimize the training process, we initiate the learning rate at 1×10^{-4}. Throughout the training, which spans 250 epochs, each batch consists of six samples. Our network is trained on a Linux host equipped with a single RTX4090 GPU. The training employs the ADAM optimizer, and the loss functions include mean squared error (MSE) $\mathcal{L}_{Cmse}$, $\mathcal{L}_{Vmse}$, MCOLE loss, and Kullback–Leibler (KL) divergence [39]. Our model has average inference times of 0.658 s, 0.011 s, and 0.177 s on the UIEBD, EUVP, and UFO-120 test datasets, respectively. On the DeepFish, RUIE, and SUIM test datasets, the average inference times are 0.765 s, 0.010 s, and 0.027 s, respectively. The model size is 2.4 gigabytes. To boost the model's generalization ability, we utilize data expansion techniques to augment the training dataset. These strategies encompass image rotations, horizontal flipping, and vertical flipping. Subsequently, we utilize 1 × 1 convolution operations to adjust the samples to the required number of channels. After this adjustment, the samples are input into the AdaIN layer, which features a 20-dimensional latent space N, to further facilitate the improvement of the network's generalization capacity.

5. Results

5.1. Comparison Study

We selected eight popular UIE models for comparison with our model, including Histoformer (2025) [46], WaterNet (2019) [40], Deep SESR (2020) [15], Deep WaveNet (2023) [28], LiteEnhanceNet (2024) [14], UDnet (2025) [13], Funie-GAN (2020) [7], and U-Shape Transformer (2023) [19]. Among them, WaterNet, Deep SESR, and Funie-GAN are classical models trained using UIEBD, UFO-120, and EUVP as training sets, respectively. Histoformer, Deep WaveNet, LiteEnhanceNet, UDnet, and U-Shape Transformer are recently published models with good performance. Evaluating our model against these models can reveal the improvement in efficacy of our model. The parameter configurations of all comparison models were in accordance with the settings outlined in the original papers, with adjustments made solely to accommodate image size requirements to ensure a fair comparison.

5.1.1. Full-Reference Results

In the research of underwater image enhancement (UIE), PSNR, SSIM, and MSE are three important full-reference metrics used for comprehensively evaluating image quality. PSNR, a well-known objective metric, has gained widespread application in the comprehensive evaluation of image and video quality. The method assesses discrepancies between enhanced and reference images by calculating the MSE of pixel intensities. The PSNR acts as a vital metric for image quality assessment, with higher values indicating superior performance in noise reduction, detail recovery, and precise image reconstruction. SSIM offers a thorough assessment of image similarity by analyzing structural, luminance, and contrast features, ensuring that multiple aspects of the image are considered. The SSIM functions as a quantitative indicator for evaluating image fidelity, where increased scores demonstrate that processed visual data achieves greater approximation to source imagery while simultaneously preserving essential architectural coherence. At the same time, the processing of brightness and contrast is closer to the original image, making the image look more natural and realistic. MSE directly measures the mean squared error between the enhanced image and the true image; a lower MSE value indicates better image quality.

We used full-reference metrics to compare and evaluate our proposed method with current cutting-edge technologies on diverse underwater datasets. The performance superiority of our method is clearly demonstrated through the PSNR and SSIM scores presented in Table 1. And our approach achieves significantly higher results in both metrics compared to other methods. Specifically, in terms of PSNR, our method attains a maximum score of 28.96 dB, representing an improvement of 20–40% over the baseline methods. This is due to the construction of a more accurate underwater image degradation model, which meticulously models the complex physical processes such as light absorption and scattering during its propagation through water, enabling more accurate pixel-level restoration. Regarding SSIM, the average value of our method reached 0.901, and it achieved significantly superior results on three datasets. Compared with the baseline, we also achieved a 10–40% improvement in performance. The success of this approach is largely due to the multiscale feature extraction module we developed. This module can effectively capture the structural information of images across different scales. By integrating these multiscale features, the model ensures a higher degree of structural similarity in the output. This comprehensive approach allows for faithful preservation of both large-scale object boundaries and fine-scale textural elements, ensuring that all relevant structural details are prominently retained in the enhanced imagery. In terms of MSE, our method achieved the lowest error values on both the UFO and UIEBD datasets, indicating that the image quality is closer to the true images. Our method scored 0.27×10^3 on the UFO dataset and 0.33×10^3 on the UIEBD dataset, significantly outperforming other methods. Visual comparisons can be found in Figure 2.

Table 1. Comparison of DeepSeaNet with WaterNet, Funie-GAN, Deep SESR, Deep WaveNet, LiteEnhanceNet, and UDnet on the EUVP, UFO, and UIEBD datasets. The performance metrics for UIE are based on the average PSNR, SSIM, and MSE values.

Method	**EUVP**			**UFO**			**UIEBD**		
	PSNR ↑	**SSIM ↑**	**MSE ($\times 10^3$) ↓**	**PSNR ↑**	**SSIM ↑**	**MSE ($\times 10^3$) ↓**	**PSNR ↑**	**SSIM ↑**	**MSE ($\times 10^3$) ↓**
WaterNet	23.06	0.796	0.72	22.47	0.766	0.95	19.11	0.792	0.79
Funie-GAN	26.22	0.793	0.39	25.64	0.767	0.62	17.13	0.744	0.68
Deep SESR	27.08	0.805	0.34	25.70	0.751	0.35	16.63	0.443	1.70
Deep WaveNet	28.62	0.832	0.29	25.71	**0.770↑**	0.30	21.57	0.800	0.60
LiteEnhanceNet	20.65	0.763	**0.15 ↓**	20.24	0.726	1.28	23.23	0.894	0.41
UDnet	22.96	0.771	0.85	22.43	0.738	0.74	22.23	0.812	0.86
DeepSeaNet (ours)	**28.96 ↑**	**0.856 ↑**	0.28	**28.70 ↑**	0.756	**0.27 ↓**	**28.57 ↑**	**0.901 ↑**	**0.33 ↓**

The symbol "↑" indicates that a higher value is better. The symbol "↓" indicates that a lower value is better.

5.1.2. Non-Reference Results

In the study of UIE, non-reference metrics play a vital role in evaluating the quality of enhanced images, especially when the original reference image is not accessible. In this paper, UICM, UISM, and UIConM are selected as the non-reference evaluation metrics. The results show that DeepSeaNet performs outstandingly in multiple non-reference metrics. UICM serves as a critical quantitative indicator for assessing color balance within images. A higher UICM value correlates with images exhibiting richer and more harmonious color distributions, while also reflecting enhanced correction of color distortion. UISM serves as a metric for evaluating image sharpness. A higher UISM value correlates with enhanced edge definition and detail clarity, thereby diminishing the extent of blurring in the image. UIConM evaluates the level of contrast in an image. When the value is higher, it signifies that the contrast between various regions is more appropriate, the image has a more pronounced hierarchical structure, and objects stand out more distinctly against the background.

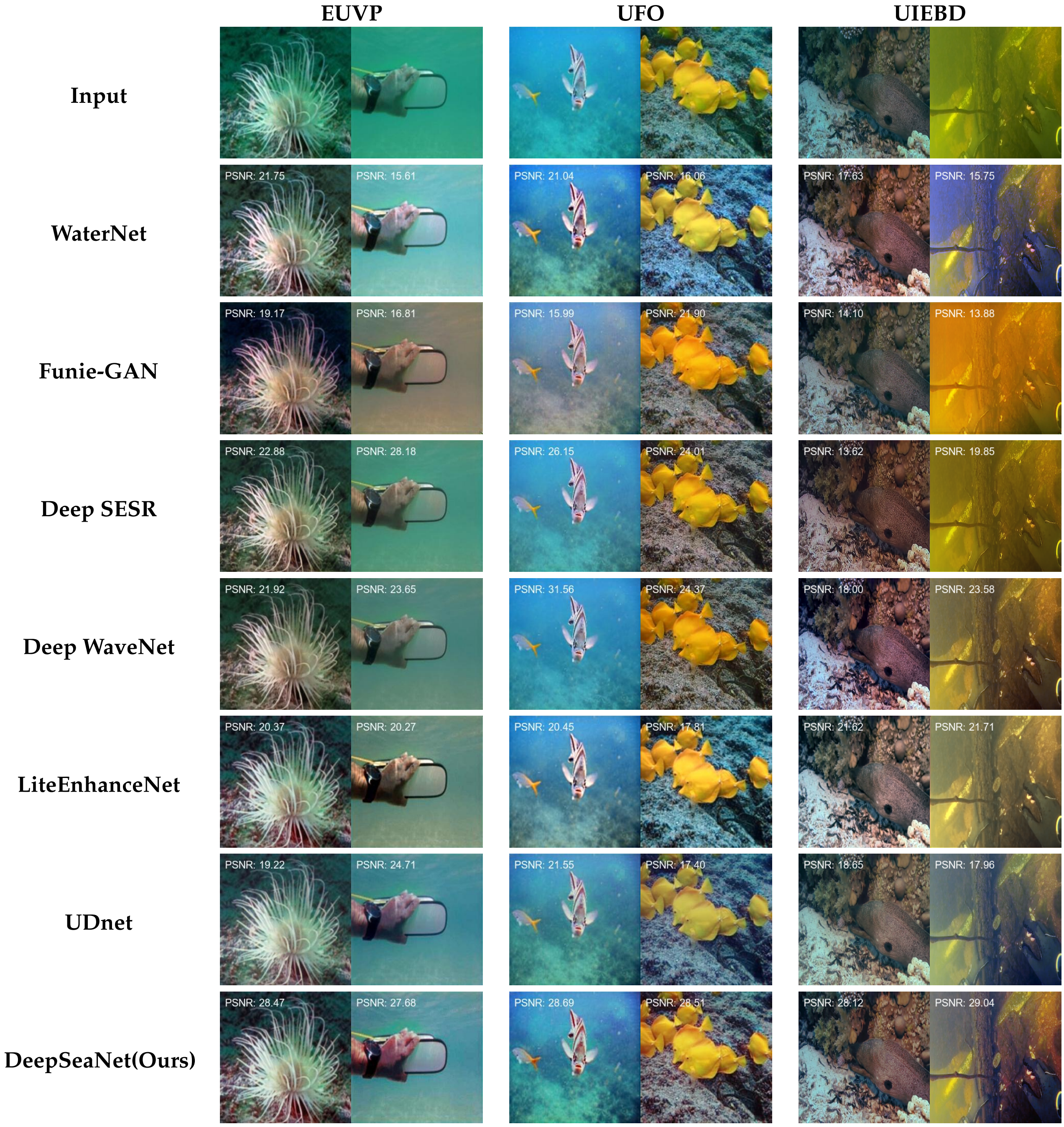

Figure 2. Visual comparison of underwater images on EUVP, UFO, and UIEBD, with the method used listed to the right of each row.

In terms of the UISM metric, DeepSeaNet scores 6.537 on the DeepFish dataset, 6.909 on the RUIE dataset, and 6.284 on the SUIM dataset, all higher than the other models in Table 2. This benefits from the application of ViT as a feature extractor. By leveraging the self-attention mechanism, ViT is capable of effectively modeling long-distance dependencies and extracting global features within images, thereby enhancing its ability to understand complex visual patterns. This architecture exhibits superior edge conservation precision and high-frequency detail retention throughout the enhancement pipeline, generating perceptually optimized visual outputs with elevated sharpness indices that demonstrate measurable improvements in UISM evaluation scores. At the same time, the data augmentation strategies adopted during the model training process, such as rotation, horizontal flipping, and vertical flipping, also help the model learn image features from different angles and transformations, further improving the processing ability of image sharpness.

In terms of the UICM metric, DeepSeaNet scores as high as 5.878 on the DeepFish dataset, 6.562 on the RUIE dataset, and 7.475 on the SUIM dataset, all achieving leading results. For the UIConM metric, DeepSeaNet scores 0.310 on the DeepFish dataset, 0.334 on the RUIE dataset, and 0.311 on the SUIM dataset, showing advantages in comparison with other models. The success of DeepSeaNet is attributable to its comprehensive structural layout and the careful enhancement of its loss function. The model's architecture is designed to seamlessly integrate information from multiple sources, enabling it to dynamically fine-tune the contrast across different regions of an image during the UIE process. Additionally, the loss function combines an MCOLE-driven perceptual loss with MSE losses for color-structural features, directing the model to prioritize balanced contrast adjustments. This ensures that enhancements are neither excessive nor insufficient, maintaining visual harmony while optimizing image quality. At the same time, it can also restore better color features, thus achieving good results in both the UIConM and UICM metrics. Visual comparisons can be found in Figure 3. A comparison of image detail can be found in Figure 4.

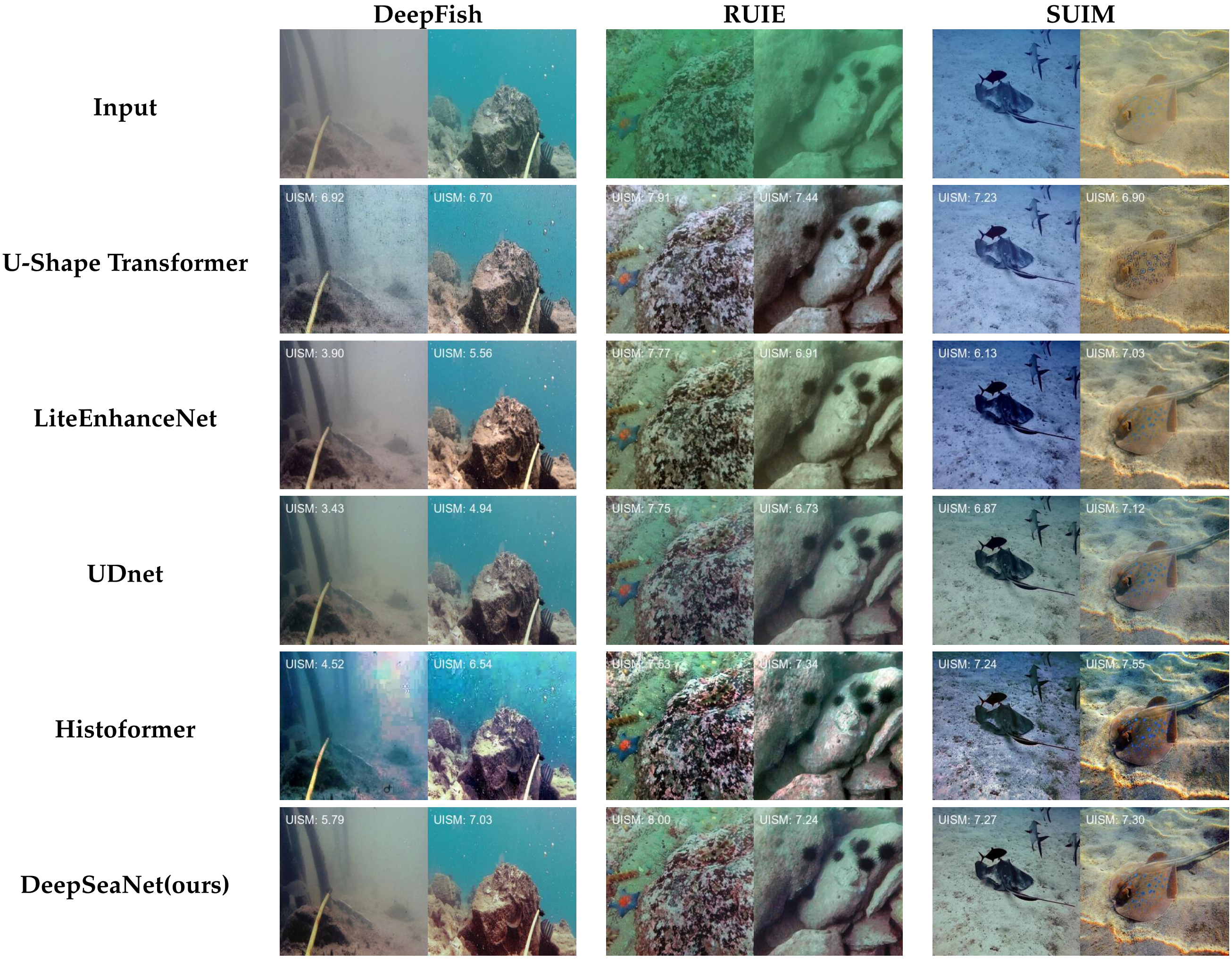

Figure 3. Visual comparison of underwater images on DeepFish, RUIE, and SUIM, with the method used listed to the right of each row.

Table 2. A comparison between DeepSeaNet, U-Shape Transformer, LiteEnhanceNet, UDnet, and Histoformer on the DeepFish, RUIE, and SUIM datasets. The performance metrics for UIE are based on the average UICM, UISM, and UIconM values.

Method	DeepFish			RUIE			SUIM		
	UICM ↑	UISM ↑	UIconM ↑	UICM ↑	UISM ↑	UIconM ↑	UICM ↑	UISM ↑	UIconM ↑
U-Shape Transformer	4.739	6.425	**0.322 ↑**	3.608	6.840	0.308	6.463	5.301	0.300
LiteEnhanceNet	4.725	6.369	0.314	2.973	6.668	0.323	7.179	6.245	0.274
UDnet	4.065	6.122	0.320	4.354	6.426	**0.335 ↑**	5.455	5.866	**0.319 ↑**
Histoformer	5.732	6.094	0.307	5.788	6.472	0.299	7.392	6.154	0.297
DeepSeaNet (ours)	**5.878 ↑**	**6.537 ↑**	0.310	**6.562 ↑**	**6.909 ↑**	0.334	**7.475 ↑**	**6.284 ↑**	0.311

The symbol "↑" indicates that a higher value is better.

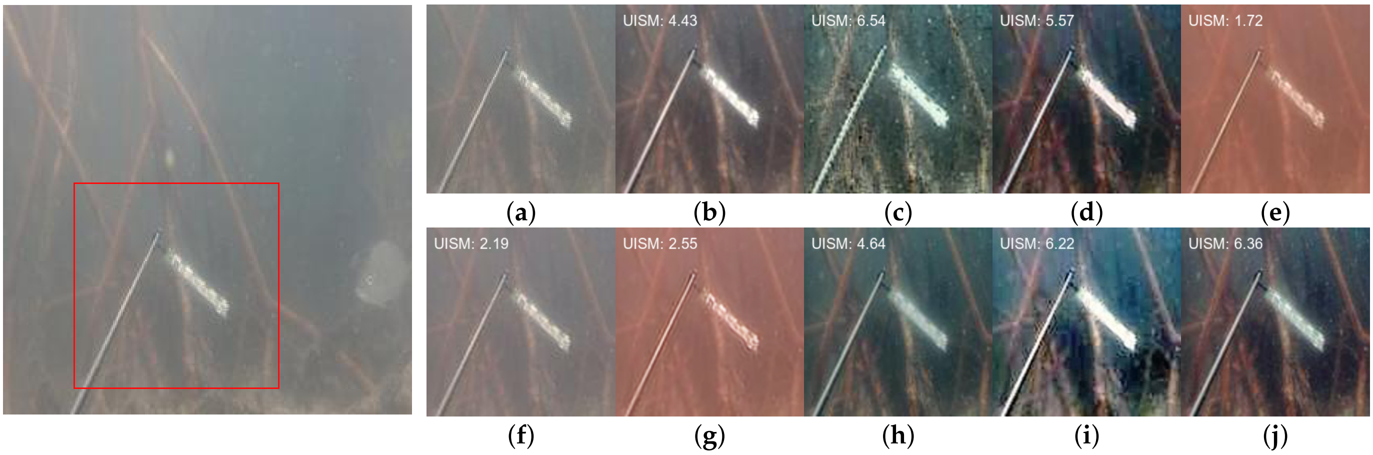

Figure 4. Comparison of image detail recovery on DeepFish dataset. (**a**) represents raw image, (**b**) represents LiteEnhanceNet, (**c**) represents U-Shape Transformer, (**d**) represents WaterNet, (**e**) represents Deep WaveNet, (**f**) represents Funie-GAN, (**g**) represents Deep SESR, (**h**) represents UDnet, (**i**) represents Histoformer, and (**j**) represents DeepSeaNet.

5.2. Ablation Study

To clarify the role of each part of the loss function, we evaluated their respective contributions by analyzing their impact on the total loss. Specifically, we deliberately removed certain components while retaining the rest and observed the corresponding changes in evaluation metrics. The greater the change caused by the removal of a particular component, the more significant its contribution to the overall loss. Table 3 shows that removing $\mathcal{L}_e$ leads to the largest drop in both PSNR and SSIM values, while removing $\mathcal{L}_{KL}$ results in a decrease in PSNR. Actually, removing the sub-components of $\mathcal{L}_e$ also leads to varying degrees of decline in PSNR and SSIM values.

Table 3. Results of ablation study. "×" indicates that the model does not contain the component, while "✓" indicates that the model contains the component.

$\mathcal{L}_{Vmse}$	$\mathcal{L}_{Cmse}$	$\mathcal{L}_{MCOLE}$	$\mathcal{L}_{KL}$	ViT	PSNR ↑	SSIM ↑
×	✓	✓	✓	✓	25.77	0.877
✓	×	✓	✓	✓	26.03	0.884
✓	✓	×	✓	✓	23.87	0.894
×	×	×	✓	✓	22.30	0.873
✓	✓	✓	×	✓	23.72	0.903
✓	✓	✓	✓	×	25.19	0.858
✓	✓	✓	✓	✓	**28.57 ↑**	**0.901 ↑**

The symbol "↑" indicates that a higher value is better.

We also conducted an ablation study comparing the contributions of different loss function components to model performance, and the results can be found in Table 3. It is worth noting that when all loss function components are included, the PSNR value reaches 28.57 and the SSIM value reaches 0.901, indicating that combining all components can significantly improve model performance.

Specifically, removing $\mathcal{L}_{Vmse}$ leads to a drop in PSNR to 25.77 and SSIM to 0.877; removing $\mathcal{L}_{Cmse}$ leads to a drop in PSNR to 26.03 and SSIM to 0.884; removing $\mathcal{L}_{MCOL}$ leads to a drop in PSNR to 23.87 and SSIM to 0.894; removing $\mathcal{L}_{KL}$ leads to a drop in PSNR to 22.30 and SSIM to 0.873. This shows that $\mathcal{L}_{KL}$ has the greatest impact on PSNR, while $\mathcal{L}_{Vmse}$ has the greatest impact on SSIM.

Furthermore, we compared the contributions of ViT and U-Net encoders to model performance. When using only the U-Net encoder, the PSNR value is 25.19 and the SSIM value is 0.858; when using the ViT encoder, the PSNR value reaches 28.57 and the SSIM value reaches 0.901. This indicates that the ViT encoder performs better in image restoration tasks, being more effective at restoring image details and structural information.

6. Discussion

The DeepSeaNet model innovatively introduces the MCOLE scoring module in the UDnet framework and the global attention mechanism in the encoder, which exhibits significant performance improvement in UIE. Performing effective capture of global features and long-range dependencies in underwater images is essential to solving the problem of inconsistent region attenuation. Especially when dealing with underwater images with blue, green, and yellow tones, our model can effectively filter these tones and restore the real colors more accurately, whereas other models such as Funie-GAN may perform well on some tones but may suffer from color distortion on other tones.

In addition, DeepSeaNet performs well in handling noise introduced by scattered light. DeepSeaNet demonstrates superior capabilities in systematic noise suppression and structural refinement compared to conventional approaches like U-Shape Transformer, which tend to introduce high-frequency artifacts and amplify structural artifacts during enhancement procedures. Through multistage feature purification, our architecture achieves a balanced preservation of critical edge information while enhancing textural fidelity in processed visual outputs. Compared with brightness enhancement models such as LiteEnhanceNet, our model not only heightens the image brightness but also sharpens the contrast, making the image brighter and more vivid.

Our analysis highlights the significant impact of different encoders on model performance. Unlike the U-Net encoder, the ViT encoder shows a significant enhancement in PSNR, SSIM, and MSE metrics, highlighting its superior performance in image restoration tasks. This advantage arises from the ViT encoder's use of the Transformer framework, which allows it to efficiently grasp long-range dependencies within images—a key factor in restoring overall structure and fine details. In contrast, while the U-Net encoder performs reasonably well in extracting local features, its ability to handle complex image scenarios is relatively constrained.

By conducting ablation studies, we were able to identify the critical contributions of the loss function's various components to how the model performs. Findings indicate that eliminating the enhancement loss results in the most significant reduction in PSNR and SSIM values, which highlights its crucial role in improving image quality and structural similarity. Removing the KL scattering loss, on the other hand, mainly affects PSNR values, highlighting its importance in optimizing image brightness and contrast. These findings demonstrate the important role of these loss function components in driving model performance. Despite DeepSeaNet's excellent performance in image enhancement, there are still some limitations. For example, although $\mathcal{L}_e$ and $\mathcal{L}_{KL}$ contribute significantly to performance, the model may not be able to completely remove all distortions in complex scenarios such as high-noise or extreme-lighting conditions. In addition, the current loss function design may not be robust enough for certain types of images such as low-contrast or high-dynamic-range images. These limitations suggest that although the current design is effective in most cases, further optimization is needed to cope with more challenging scenarios and enhance the model's ability.

7. Conclusions

Marked by light attenuation, color distortion, diminished contrast, heightened noise, and blurring effects, underwater images present considerable challenges for enhancement techniques. We proposed DeepSeaNet, which innovatively integrates a multi-channel color evaluation mechanism and a global attention encoder into the UDnet framework, effectively addressing the issue of inconsistent attenuation across regions and channels in underwater images. By incorporating multiscale feature fusion and leveraging global information, DeepSeaNet generates enhanced images with superior color expression and clearer structural features. DeepSeaNet holds great promise in object recognition, exploration, and archeology in the deep-sea field.

Author Contributions: Conceptualization, data curation, and validation, J.L.; writing—original draft preparation, J.L. and Y.O.; writing—review and editing, H.W., D.W. and Y.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data will be available upon request.

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

References

1. Chen, X.; Yu, J.; Kong, S.; Wu, Z.; Fang, X.; Wen, L. Towards real-time advancement of underwater visual quality with GAN. *IEEE Trans. Ind. Electron.* **2019**, *66*, 9350–9359. [CrossRef]
2. Wang, Y.; Tang, C.; Cai, M.; Yin, J.; Wang, S.; Cheng, L.; Wang, R.; Tan, M. Real-time underwater onboard vision sensing system for robotic gripping. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 5002611
3. Schettini, R.; Corchs, S. Underwater image processing: State of the art of restoration and image enhancement methods. *EURASIP J. Adv. Signal Process.* **2010**, *2010*, 746052. [CrossRef]
4. Yang, M.; Hu, J.; Li, C.; Rohde, G.; Du, Y.; Hu, K. An in-depth survey of underwater image enhancement and restoration. *IEEE Access* **2019**, *7*, 123638–123657. [CrossRef]
5. Sahu, P.; Gupta, N.; Sharma, N. A survey on underwater image enhancement techniques. *Int. J. Comput. Appl.* **2014**, *87*, 19–23. [CrossRef]
6. Jiang, L.; Wang, Y.; Jia, Q.; Xu, S.; Liu, Y.; Fan, X.; Li, H.; Liu, R.; Xue, X.; Wang, R. Underwater species detection using channel sharpening attention. In Proceedings of the 29th ACM International Conference on Multimedia, Virtual, 24–26 October 2021; pp. 4259–4267.
7. Islam, M.J.; Xia, Y.; Sattar, J. Fast underwater image enhancement for improved visual perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [CrossRef]
8. Li, C.; Anwar, S.; Porikli, F. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognit.* **2020**, *98*, 107038. [CrossRef]
9. Liu, P.; Wang, G.; Qi, H.; Zhang, C.; Zheng, H.; Yu, Z. Underwater image enhancement with a deep residual framework. *IEEE Access* **2019**, *7*, 94614–94629. [CrossRef]
10. Zhao, C.; Cai, W.; Dong, C.; Zeng, Z. Toward sufficient spatial-frequency interaction for gradient-aware underwater image enhancement. 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; IEEE: Piscataway, NJ, USA, 2024; pp. 3220–3224.

11. Ma, Z.; Oh, C. A wavelet-based dual-stream network for underwater image enhancement. In Proceedings of the ICASSP 2022—2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 22–27 May 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 2769–2773.
12. Wang, Y.; Hu, S.; Yin, S.; Deng, Z.; Yang, Y.H. A multi-level wavelet-based underwater image enhancement network with color compensation prior. *Expert Syst. Appl.* **2024**, *242*, 122710. [CrossRef]
13. Saleh, A.; Sheaves, M.; Jerry, D.; Rahimi Azghadi, M. Adaptive deep learning framework for robust unsupervised underwater image enhancement. *Expert Syst. Appl.* **2025**, *268*, 126314. [CrossRef]
14. Zhang, S.; Zhao, S.; An, D.; Li, D.; Zhao, R. LiteEnhanceNet: A lightweight network for real-time single underwater image enhancement. *Expert Syst. Appl.* **2024**, *240*, 122546. [CrossRef]
15. Islam, M.J.; Luo, P.; Sattar, J. Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. *arXiv* **2020**, arXiv:2002.01155.
16. Fu, Z.; Wang, W.; Huang, Y.; Ding, X.; Ma, K.K. Uncertainty inspired underwater image enhancement. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 465–482.
17. Li, H.; Zhuang, P. DewaterNet: A fusion adversarial real underwater image enhancement network. *Signal Process. Image Commun.* **2021**, *95*, 116248. [CrossRef]
18. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems 27 (NIPS 2014), Montreal, QC, Canada, 8–13 December 2024.
19. Peng, L.; Zhu, C.; Bian, L. U-shape transformer for underwater image enhancement. *IEEE Trans. Image Process.* **2023**, *32*, 3066–3079. [CrossRef]
20. Jiang, Q.; Yi, X.; Ouyang, L.; Zhou, J.; Wang, Z. Towards dimension-enriched underwater image quality assessment. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *35*, 1385–1398. [CrossRef]
21. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
22. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
23. Kang, Y.; Jiang, Q.; Li, C.; Ren, W.; Liu, H.; Wang, P. A Perception-Aware Decomposition and Fusion Framework for Underwater Image Enhancement. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 988–1002. [CrossRef]
24. Zhang, W.; Zhuang, P.; Sun, H.H.; Li, G.; Kwong, S.; Li, C. Underwater Image Enhancement via Minimal Color Loss and Locally Adaptive Contrast Enhancement. *IEEE Trans. Image Process.* **2022**, *31*, 3997–4010. [CrossRef]
25. Jiang, Z.; Li, Z.; Yang, S.; Fan, X.; Liu, R. Target Oriented Perceptual Adversarial Fusion Network for Underwater Image Enhancement. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 6584–6598. [CrossRef]
26. Gangisetty, S.; Rai, R.R. FloodNet: Underwater image restoration based on residual dense learning. *Signal Process. Image Commun.* **2022**, *104*, 116647. [CrossRef]
27. Yan, X.; Qin, W.; Wang, Y.; Wang, G.; Fu, X. Attention-guided dynamic multi-branch neural network for underwater image enhancement. *Knowl. Based Syst.* **2022**, *258*, 110041. [CrossRef]
28. 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]
29. Fan, G.; Zhou, J.; Xu, C.; Cheng, Z. Deep dive into clarity: Leveraging signal-to-noise ratio awareness and knowledge distillation for underwater image enhancement. *Expert Syst. Appl.* **2025**, *269*, 126317. [CrossRef]
30. Agarwal, A.; Gupta, S.; Vashishath, M. Contrast enhancement of underwater images using conditional generative adversarial network. *Multimed. Tools Appl.* **2024**, *83*, 41375–41404. [CrossRef]
31. Shen, Z.; Xu, H.; Luo, T.; Song, Y.; He, Z. UDAformer: Underwater image enhancement based on dual attention transformer. *Comput. Graph.* **2023**, *111*, 77–88. [CrossRef]
32. Khan, M.; Negi, A.; Kulkarni, A.; Phutke, S.S.; Vipparthi, S.K.; Murala, S. Phaseformer: Phase-based Attention Mechanism for Underwater Image Restoration and Beyond. *arXiv* **2024**, arXiv:2412.01456.
33. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
34. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
35. Sohn, K.; Lee, H.; Yan, X. Learning structured output representation using deep conditional generative models. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 3483–3491.
36. Kingma, D.P.; Welling, M. An introduction to variational autoencoders. *Found. Trends® Mach. Learn.* **2019**, *12*, 307–392. [CrossRef]
37. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386. [CrossRef] [PubMed]

38. Li, H.; Cen, Y.; Liu, Y.; Chen, X.; Yu, Z. Different input resolutions and arbitrary output resolution: A meta learning-based deep framework for infrared and visible image fusion. *IEEE Trans. Image Process.* **2021**, *30*, 4070–4083. [CrossRef] [PubMed]
39. Contreras-Reyes, J.E.; Arellano-Valle, R.B. Kullback–Leibler divergence measure for multivariate skew-normal distributions. *Entropy* **2012**, *14*, 1606–1626. [CrossRef]
40. 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]
41. Saleh, A.; Laradji, I.H.; Konovalov, D.A.; Bradley, M.; Vazquez, D.; Sheaves, M. A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis. *Sci. Rep.* **2020**, *10*, 14671. [CrossRef]
42. 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]
43. Islam, M.J.; Edge, C.; Xiao, Y.; Luo, P.; Mehtaz, M.; Morse, C.; Enan, S.S.; Sattar, J. Semantic segmentation of underwater imagery: Dataset and benchmark. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1769–1776.
44. 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]
45. Panetta, K.; Gao, C.; Agaian, S. Human-visual-system-inspired underwater image quality measures. *IEEE J. Ocean. Eng.* **2015**, *41*, 541–551. [CrossRef]
46. Peng, Y.T.; Chen, Y.R.; Chen, G.R.; Liao, C.J. Histoformer: Histogram-Based Transformer for Efficient Underwater Image Enhancement. *IEEE J. Ocean. Eng.* **2025**, *50*, 164–177. [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.