Journal of *Marine Science and Engineering*

MDPI

Article

Agent-Guided Non-Local Network for Underwater Image Enhancement and Super-Resolution Using Multi-Color Space

Rong Wang *, Yonghui Zhang * and Yulu Zhang

School of Information and Communication Engineering, Hainan University, Haikou 570228, China; ylzhang@hainanu.edu.cn

* Correspondence: wrong@hainanu.edu.cn (R.W.); yhzhang@hainanu.edu.cn (Y.Z.)

Abstract: The absorption and scattering of light in water usually result in the degradation of underwater image quality, such as color distortion and low contrast. Additionally, the performance of acquisition devices may limit the spatial resolution of underwater images, resulting in the loss of image details. Efficient modeling of long-range dependency is essential for understanding the global structure and local context of underwater images to enhance and restore details, which is a challenging task. In this paper, we propose an agent-guided non-local attention network using a multi-color space for underwater image enhancement and super-resolution. Specifically, local features with different receptive fields are first extracted simultaneously in the RGB, Lab, and HSI color spaces of underwater images. Then, the designed agent-guided non-local attention module with high expressiveness and lower computational complexity is utilized to model long-range dependency. Subsequently, the results from the multi-color space are adaptively fused with learned weights, and finally, the reconstruction block composed of deconvolution and the designed non-local attention module is used to output enhanced and super-resolution images. Experiments on multiple datasets demonstrated that our method significantly improves the visual perception of degraded underwater images and efficiently reconstructs missing details, and objective evaluations confirmed the superiority of our method over other state-of-the-art methods.

Keywords: underwater image enhancement; super-resolution; non-local attention; deep learning

Citation: Wang, R.; Zhang, Y.; Zhang, Y. Agent-Guided Non-Local Network for Underwater Image Enhancement and Super-Resolution Using Multi-Color Space. *J. Mar. Sci. Eng.* **2024**, *12*, 358. https://doi.org/10.3390/jmse12020358

Academic Editor: Mihalis Golias

Received: 24 January 2024
Revised: 10 February 2024
Accepted: 18 February 2024
Published: 19 February 2024

1. Introduction

As a valuable information medium in the deep ocean, underwater images provide a wealth of data for marine science, resource exploration, and ecological protection. However, due to optical effects such as water absorption and scattering during the propagation of light, underwater images are often accompanied by challenging problems such as color distortion, contrast degradation, lack of clarity, and blurring of details. Moreover, the acquired underwater images may suffer from low spatial resolution due to the performance limitations of the acquisition equipment. In our study, we focus on addressing these challenges through two key explorations: underwater image enhancement (UIE) and underwater image super-resolution (SR).

Underwater-image-enhancement techniques [1,2] focus on adjusting the pixel values of images to improve their visual quality, thereby increasing the usability of underwater images for various underwater activities and applications. Existing underwater-image-enhancement methods are mainly categorized into traditional enhancement methods [3–11] and learning-based enhancement methods [12–24]. Traditional underwater-image-enhancement methods include both physical-model-based methods [4–7] and non-physical-model-based methods [8–11]. Most physical-model-based methods utilize prior knowledge to invert the underwater imaging model for computing the real scene, and their robustness is influenced by the accuracy of the prior knowledge. Non-physical-model-based methods directly modify the pixel values of images using signal-processing methods to improve image

quality, with no involvement in the imaging model. In recent research, learning-based underwater-image-enhancement methods have become mainstream, such as methods based on convolutional neural networks (CNNs) [12–17], generative adversarial networks (GANs) [18–22], and Vision Transformer (ViT) [23,24]. CNN-based methods achieve image enhancement by learning the mapping relationship between the raw and reference images, while GAN-based methods employ adversarial learning to generate realistic images, focusing on enhancing the overall perceptual quality of underwater images. Currently, underwater-image-enhancement methods still fall short of satisfactory results. On the one hand, traditional methods either overlook imaging mechanisms leading to excessive or insufficient enhancement or rely on specific prior knowledge, making them unsuitable for diverse underwater scenes. On the other hand, most learning-based methods concentrate on extracting the local features of images, neglecting the importance of capturing non-local correlation. Moreover, they predominantly use the RGB color space, which cannot directly reflect some critical attributes of images, hindering further improvement of issues such as low contrast, saturation, and brightness.

The basic goal of image super-resolution is to generate a corresponding high-resolution (HR) image from a low-resolution (LR) counterpart, thereby providing the image with more details and clarity. Consequently, underwater image super-resolution is crucial for improving image quality and enhancing visual perception. In recent years, deep-learning-based approaches, especially CNNs and GANs, have achieved significant advancements in image super-resolution tasks [25–30]. For instance, Dong et al. [25] were the first to propose using CNNs to learn the mapping relationship between LR and HR images, generating high-quality images. The super-resolution algorithm SRGAN introduced by Ledig et al. [28], incorporating a GAN framework, improves image resolution in a more natural and realistic manner. While these methods have demonstrated satisfactory performance in terrestrial images, they may not generalize well to underwater scenes due to significant differences in underwater optical characteristics compared to terrestrial environments. Additionally, uneven lighting and varying water quality in different aquatic environments pose challenges in underwater scenes. To more effectively address the image super-resolution requirements in underwater scenes, several methods and datasets [31–33] tailored to underwater super-resolution tasks have been proposed. For example, Islam et al. [31] constructed an underwater image super-resolution dataset (USR-248) and introduced two methods, SRDRM and SRDRM-GAN, based on fully convolutional deep residual networks, aimed at enhancing the spatial resolution of reconstructed underwater images. While these methods are effective at improving the quality of underwater images, there is still room for further optimization.

In addressing the aforementioned issues, this paper proposes an agent-guided non-local attention-based network for underwater image enhancement and super-resolution. Specifically, we simultaneously extracted shallow features of underwater images in the RGB, Lab, and HSI color spaces and utilized agent-guided non-local attention modules to capture long-range dependency in underwater images. The learned non-local context residuals were then added to the inputs of the corresponding color spaces to obtain enhanced results in different color spaces, and then, they were adaptively fused with the learned weights. The reconstruction block, composed of deconvolution and the designed non-local attention module, produces enhanced underwater images and super-resolution images. In summary, the main contributions of this paper are as follows:

1. An agent-guided non-local attention network using the multi-color space is proposed for underwater image enhancement and super-resolution. The network extracts feature maps with different receptive fields and captures long-range dependency in the RGB, Lab, and HSI color spaces simultaneously, then the adaptive fusion of outputs from different color spaces is achieved with learned weight coefficients. Finally, enhanced images and high-resolution underwater images are generated by the reconstruction block.

2. An agent-guided non-local attention module is designed to capture long-range dependency in underwater images, enhancing the understanding of the global structure and contextual information. The module performs two non-local attention computations, demonstrating high expressive power while maintaining significantly lower computational complexity compared to the typical non-local block.
3. The proposed network outperforms state-of-the-art methods in several objective quality assessment metrics on multiple datasets, both for underwater image enhancement and underwater image super-resolution. Qualitative comparisons also demonstrate that our network produces visually pleasing results.

The rest of this paper is organized as follows: Section 2 reviews related work on underwater-image-enhancement methods and underwater-image-super-resolution methods and revisits the non-local block. Section 3 illustrates the network architecture and relevant internal structures of our method. Section 4 presents the experimental details, results and analysis, and ablation study. Section 5 provides the conclusion of this work.

2. Related Works

2.1. Underwater Image Enhancement

The underwater-image-enhancement methods based on physical models primarily utilize prior knowledge to estimate optical parameters in the imaging model and invert the model to obtain a clean image. For instance, Drews et al. [4] extended the dark channel prior (DCP) [3] to the underwater dark channel prior (UDCP), using the priors of the blue and green channels to estimate the depth map, addressing issues caused by the severe attenuation of red light in water, leading to underwater image distortion and blue–green color problems. Song et al. [5] employed underwater light attenuation prior information to calculate the depth map, thereby estimating optical parameters such as background light and the transmission map for underwater image restoration. Liang et al. [7] used priors from the red, green, and blue channels to estimate transmission map parameters, incorporating image brightness, blurriness, and wavelength-related attenuation as priors to estimate background light and inversely reconstruct the underwater image. Non-physical-model-based underwater-image-enhancement methods mainly modify the pixel values of images to achieve enhancement. Jyoti et al. [8] proposed histogram equalization (HE), utilizing pixel value transformations to adjust the distribution of pixel values throughout the entire brightness range for a more-uniform distribution and enhanced visual effect.

In recent years, deep-learning-based underwater-image-enhancement methods have gained extensive attention. Li et al. [12], combining underwater optical characteristics and the imaging model, provided ten synthetic underwater image datasets and trained the proposed lightweight underwater-image-enhancement model (UWCNN). Li et al. [13] constructed a real-world underwater image dataset (UIEBD) used to train the proposed CNN-based underwater-image-enhancement network (Water-Net), demonstrating the suitability of this dataset for training CNNs. Li et al. [15], leveraging the advantages of physical models and deep learning, introduced a deep network with multi-color space embedding (Ucolor), incorporating medium transmission as prior knowledge to guide the network in generating enhanced images with higher contrast and color correction. Fabbri et al. [18], using collected real-world clear and degraded underwater images, trained the CycleGAN [34] to generate paired images, employed to train the proposed GAN-based underwater-image-enhancement network. Islam et al. [19] adopted the paired image-generation proposed in [18] to create the underwater dataset (EUVP) and designed a network (FUnIE-GAN) for fast underwater image enhancement, achieving improved enhancement performance through a mix of supervised and unsupervised training. Peng et al. [24] introduced a U-shaped network based on Vision Transformer [35], successfully incorporating the Transformer with the self-attention mechanism into the underwater-image-enhancement task.

Physical-model-based underwater-image-enhancement methods have achieved good performance in certain specific underwater scenes. However, due to the assumed imaging models that may not fit all underwater scenes and the difficulty in accurately estimating

multiple imaging parameters, these methods fail to maintain robust performance when encountering diverse underwater environments. Non-physical-model-based underwater-image-enhancement methods can improve visual quality to some extent, but they usually produce over-/under-enhanced results due to the neglect of underwater imaging mechanisms. In comparison to traditional methods, existing learning-based underwater-image-enhancement methods have made significant progress. Nevertheless, these deep learning methods based on CNNs and GANs exhibit limitations in robustness and generalization performance because they overlook the capture of long-range dependency in degraded images, which is crucial for understanding the global structure of underwater scenes. While Transformer-based methods emphasize attention to global information, they discard convolutions that are advantageous for capturing local features, resulting in unsatisfactory image quality. Additionally, these methods predominantly rely on the RGB color space, making them insensitive to attributes such as brightness and saturation, thereby limiting their performance.

2.2. Underwater Image Super-Resolution

In recent years, extensive research has demonstrated significant advantages in applying deep learning to address low-level visual tasks, with image super-resolution being a beneficiary of these advancements. After Dong et al. [25] introduced the three-layer convolutional neural network (SRCNN) for image super-resolution, newer and deeper networks have been continually proposed. Mao et al. [26] presented the deeply recursive convolutional autoencoder network (DSRCNN) to enhance spatial resolution for restoring image details. The performance of the residual learning-based deep network (EDSR), proposed by Lim et al. [27], surpassed the SRCNN in single-image super-resolution. Additionally, Liang et al. [29] applied the Swin Transformer [36] with window self-attention to single-image super-resolution, achieving significant performance improvements. Recently, Choi et al. [30] introduced the N-Gram context for adjacent local window interaction, enhancing the network proposed in [29]. While these methods have achieved satisfactory performance in terrestrial images, they may not be suitable for underwater super-resolution tasks because the underwater environment presents optical properties and challenges that are significantly different from those of the terrestrial environment. These challenges include blurring and loss of detail due to light attenuation and scattering, color distortion due to light absorption in the water, and blurring and noise due to currents, particulate matter, and contaminants. Therefore, studies on super-resolution methods based on underwater scenes have been initiated. Unlike Islam et al. [31], who proposed a GAN-based fully convolutional deep residual network for underwater super-resolution, Chen et al. [32] introduced a CNN model based on channel attention to learn the mapping relationship between low-resolution and high-resolution images. In addition, Islam et al. [33] proposed a residual CNN network with a multi-modal loss function (Deep-SESR), capable of simultaneously handling enhancement and super-resolution tasks, and constructed an underwater image benchmark dataset (UFO-120) for enhancement and super-resolution.

Although these super-resolution methods have yielded encouraging results in improving image quality, there is room for improvement in their ability to enhance visual perception, as they primarily focus on reconstructing image details without fully considering varying degrees of color distortion. While these methods that achieve both enhancement and super-resolution address both detail recovery and color correction, their designed networks lack the capability for global dependency modeling, making it challenging to capture critical information away from the regions of interest and potentially limiting their adaptability to various underwater environments.

2.3. Non-Local Block

Models based on CNNs or GANs typically use the depth stack of the convolutional layers to model long-range dependency, where each layer models pixel relationships within a local neighborhood. However, direct stacking of convolutional layers is inefficient

and difficult to optimize due to the difficulty of transferring information between distant positions. To address this issue, Wang et al. [37] designed a single-layer non-local block to directly model long-range dependency. Figure 1 illustrates the structure of a widely used instantiation of the non-local block, embedded Gaussian.

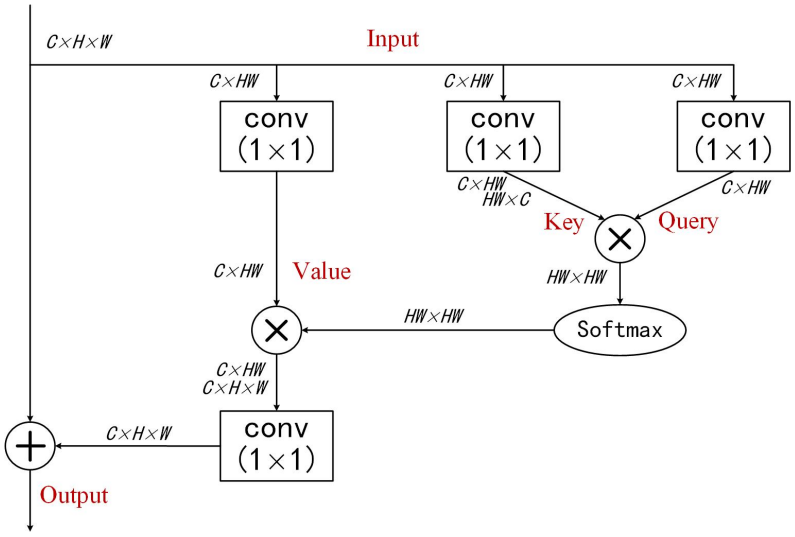

Figure 1. An instantiation of the non-local block: embedded Gaussian. C, H, and W represent the channel number, height, and width of the input maps, respectively. $\otimes$ denotes matrix multiplication, and $\oplus$ denotes elementwise addition.

The non-local block can be viewed as a query-specific global-context-modeling module, which enhances features at the query position by querying specific global context vectors. The process involves matrix multiplication of the query and key to calculate the similarity between the positions of each pixel and other pixel positions. Subsequently, the calculated similarity is used to assign weights to each position. Matrix multiplication is then performed between the weight map and the value to obtain the weighted sum and compute the query-specific global context vector. The non-local block is a type of Softmax attention with high expressive power [35], but with high computational complexity squared to the number of positions. In comparison to the typical non-local block, the agent-guided non-local attention module designed in this paper maintains the advantages of the high expressiveness of Softmax attention and lower complexity while preserving the ability to model long-range dependency.

3. Proposed Method

3.1. Network Architecture

The overall architecture of the proposed network is illustrated in Figure 2. The network can be divided into four parts: shallow feature extraction, deep feature extraction, adaptive fusion, and image reconstruction. Each of these parts will be explained in detail next.

Shallow feature extraction. The goal of this part is to obtain shallow features with diverse receptive fields and rich color information. The Lab color space separates color information into luminance and two chromatic channels, a and b, which are more aligned with human perception of color. The HSI color space divides color information into hue, saturation, and intensity, making color information more interpretable and manageable. Considering that using large convolutional kernels for processing the intensity component is more effective in capturing large-scale features in the image and since both HSI and Lab include a luminance attribute, but Lab simultaneously considers human color perception, using large convolutional kernels may not preserve luminance information as prominently as in the HSI color space. Therefore, we employed 3×3, 5×5, and 7×7 convolutional

kernels for feature extraction in the RGB, Lab, and HSI color spaces of underwater images, respectively, and obtained the corresponding contextual features:

$$\begin{aligned} D_{Lab} &= RGB2Lab(D_{RGB}) \\ D_{HSI} &= RGB2HSI(D_{RGB}) \\ f^1_{RGB} &= \delta(bn(conv^{3\times3}(D_{RGB}))) \\ f^1_{Lab} &= \delta(bn(conv^{5\times5}(D_{Lab}))) \\ f^1_{HSI} &= \delta(bn(conv^{7\times7}(D_{HSI}))) \end{aligned} \tag{1}$$

where $RGB2Lab(\cdot)$ and $RGB2HSI(\cdot)$ denote the conversion of the color space of an image from RGB to Lab and HSI, respectively. $conv^{3\times3}$, $conv^{5\times5}$, and $conv^{7\times7}$ denote the convolution operations with kernels 3×3, 5×5, and 7×7, respectively. bn and δ denote the batch normalization and PReLU activation functions, respectively.

The output of this part is:

$$M^1 = f^1_{RGB} \odot f^1_{Lab} \odot f^1_{HSI} \tag{2}$$

where $\odot$ denotes concatenation.

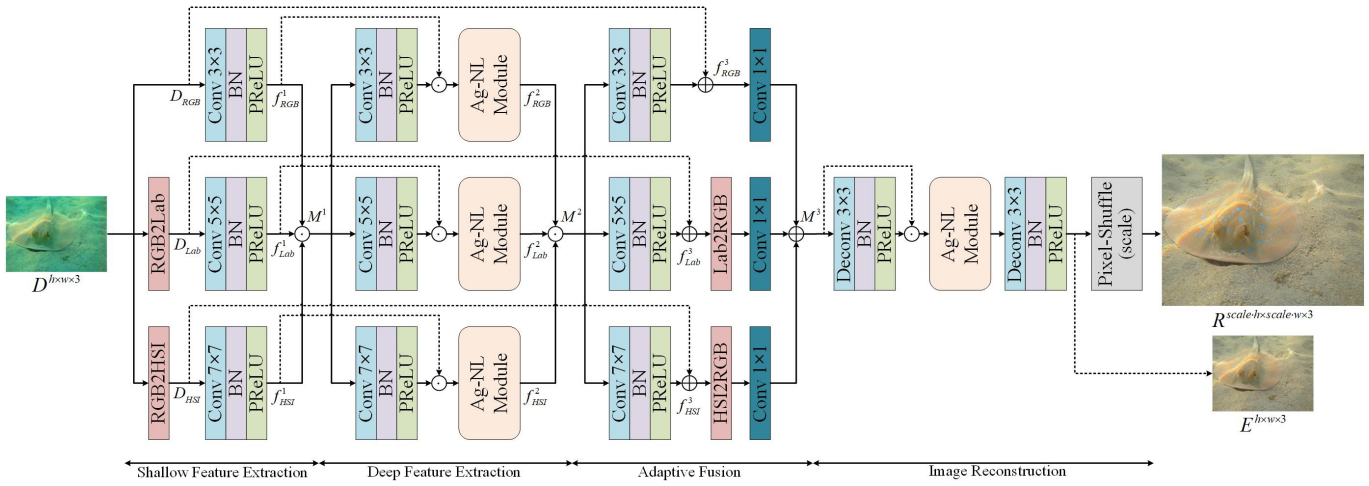

Figure 2. The overall network architecture of the proposed network.

Deep feature extraction. The goal of this stage is to efficiently model long-range dependency with lower computational complexity and obtain non-local contextual residuals in different color spaces. Firstly, similar to the shallow-feature-extraction part, 3×3, 5×5, and 7×7 convolutional kernels are used to extract deeper feature representations. Subsequently, the obtained feature maps are concatenated with the corresponding shallow feature maps and fed into the agent-guided non-local attention module to generate feature maps containing non-local contextual information:

$$\begin{aligned} f^2_{RGB} &= \delta(bn(conv^{3\times3}(M^1))) \odot f^1_{RGB} \\ f^2_{Lab} &= \delta(bn(conv^{5\times5}(M^1))) \odot f^1_{Lab} \\ f^2_{HSI} &= \delta(bn(conv^{7\times7}(M^1))) \odot f^1_{HSI} \end{aligned} \tag{3}$$

$$\begin{aligned} f^2_{RGB} &= Ag{-}NL(f^2_{RGB}) \\ f^2_{Lab} &= Ag{-}NL(f^2_{Lab}) \\ f^2_{HSI} &= Ag{-}NL(f^2_{HSI}) \end{aligned} \tag{4}$$

where $Ag\text{-}NL$ refers to the Ag-NL module.

The output of this part can be expressed as:

$$M^2 = f^2_{RGB} \odot f^2_{Lab} \odot f^2_{HSI} \tag{5}$$

Adaptive fusion. This part aims to integrate the results obtained through feature learning in different color spaces as comprehensively as possible, thereby generating a color-corrected map:

$$\begin{array}{l} f^3_{RGB} = \delta(bn(conv^{3\times3}(M^2))) \oplus D_{RGB} \\ f^3_{Lab} = \delta(bn(conv^{5\times5}(M^2))) \oplus D_{Lab} \\ f^3_{HSI} = \delta(bn(conv^{7\times7}(M^2))) \oplus D_{HSI} \end{array} \tag{6}$$

$$M^3 = conv^{1\times1}(f^3_{RGB}) \oplus conv^{1\times1}(Lab2RGB(f^3_{Lab})) \oplus conv^{1\times1}(HSI2RGB(f^3_{HSI})) \tag{7}$$

where $\oplus$ denotes elementwise addition, $conv^{1\times1}$ denotes convolution with kernel 1×1, and $Lab2RGB(\cdot)$ and $HSI2RGB(\cdot)$ denote the conversion of the color space of the image from Lab and HSI back to RGB.

Image reconstruction. The goal of this stage is to further capture long-range dependency in the corrected image and ultimately generate enhanced and super-resolution images. Initially, local features of the color-corrected map are extracted using a 3×3 deconvolution operation. Subsequently, these features are concatenated with the color-corrected map and fed into the agent-guided non-local attention module. Finally, a 3×3 deconvolution layer (followed by a pixel-shuffle layer for the SR task) is employed to generate the predicted image:

$$\begin{array}{l} f^4 = Ag{-}NL(\delta(bn(deconv^{3\times3}(M^3))) \odot M^3) \\ E^{h\times w\times3} = \delta(bn(deconv^{3\times3}(f^4))) \\ R^{scale\cdot h\times scale\cdot w\times3} = pixelshuffle(E^{h\times w\times3\cdot scale^2}) \end{array} \tag{8}$$

where $deconv^{3\times3}$ denotes the deconvolution layer with kernel 3×3, $pixelshuffle$ denotes the pixel-shuffle layer, and $scale$ refers to the scale factor of SR. $E^{h\times w\times3}$ denotes the enhanced underwater image, and $R^{scale\cdot h\times scale\cdot w\times3}$ denotes the super-resolution underwater image.

3.2. Agent-Guided Non-Local Attention Module

Capturing long-range dependency in underwater images is advantageous for enhancing the network's perception of the overall context. This enables the network to consider various parts of the image more comprehensively, leading to improved handling of color distortion and more-accurate restoration of detailed information. The typical non-local block, while effective in long-range modeling, comes with high computational complexity and is typically applied sparingly in the network architecture. To achieve the efficient capture of long-range dependency, we designed the agent-guided non-local attention module (Ag-NL module), as illustrated in Figure 3.

The Ag-NL module can be represented as:

$$\begin{array}{l} O = \sigma(QA^T)\sigma(AK^T)V + DWC(I) \\ A = Globalpooling(I) \in {}^{C\times1} \end{array} \tag{9}$$

where $Q, K, V \in \mathbb{R}^{C\times HW}$ denote $Query$, Key, and $Value$, respectively. σ denotes the Softmax activation function, and $DWC(\cdot)$ denotes the depthwise convolution. I denotes the input; O denotes the output; A denotes the agent, which is obtained by applying global average pooling to the input I.

First, A is used as the query, and attention computation is performed between A, K, and V to aggregate global information from K and V, resulting in agent features. Subsequently, A is used as the key, and the agent features are used as the value. By employing Q for the second attention computation, the global information of the agent features is broadcast to each query, thereby achieving efficient global context modeling. Simultaneously, the depthwise convolution is applied to the input feature maps to enhance feature diversity. Finally, the obtained global context is utilized to strengthen the input feature maps with diversity. The computational complexity of the typical non-local block is $\mathcal{O}(C(HW)^2)$, where $\mathcal{O}(\cdot)$ denotes complexity. In contrast, our Ag-NL module has a complexity of $\mathcal{O}(CHW)$, reducing the computational complexity while retaining the ability for long-range dependency modeling, making it applicable to multiple locations in the network architecture.

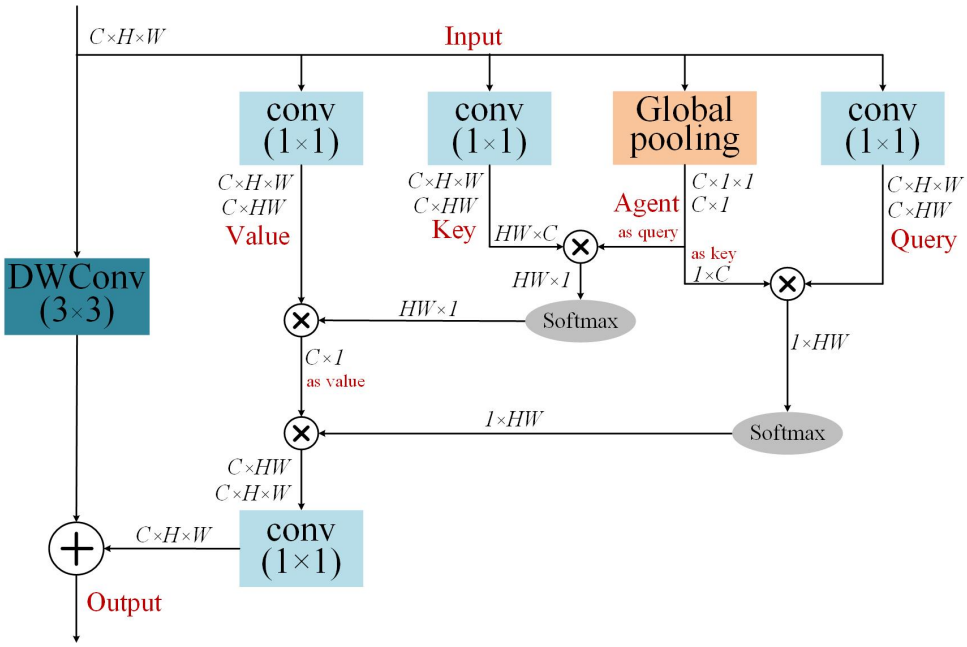

Figure 3. Agent-guided non-local attention module (Ag-NL Module). C, H, and W represent the channel number, height, and width of the input maps, respectively. $\otimes$ denotes matrix multiplication, and $\oplus$ denotes elementwise addition.

3.3. Loss Function

To train the proposed network, we designed a comprehensive loss function that includes the L_1 loss, SSIM loss, and perceptual loss. The L_1 loss is used to measure global similarity; the SSIM loss is employed for structural similarity; the perceptual loss assesses the consistency in image content. The L_1 loss, also known as the mean absolute error (MAE), calculates the mean of the absolute difference between all pixels of the predicted image $\hat{y}$ and the ground-truth image y, as follows:

$$L_1 = \sum_{m=1}^{W} \sum_{n=1}^{H} |\hat{y}(m,n) - y(m,n)| \tag{10}$$

The SSIM loss is used to enhance the structural and textural similarity between the predicted image $\hat{y}$ and the ground-truth image y, and the computation can be expressed as:

$$L_{SSIM} = 1 - \frac{(2u_{\hat{y}}u_y + \mu_1)(2\sigma_{\hat{y}y} + \mu_2)}{(u_{\hat{y}}^2 + u_y^2 + \mu_1)(\sigma_{\hat{y}}^2 + \sigma_y^2 + \mu_2)} \tag{11}$$

where $u_{\hat{y}}$ and $\sigma_{\hat{y}}$ represent the mean and standard deviation of the predicted image $\hat{y}$, u_y and σ_y represent the mean and standard deviation of the real image y, and $\sigma_{\hat{y}y}$ represents the covariance.

Perceptual loss is used to calculate the distance between the predicted image $\hat{y}$ and the feature representation of the ground-truth image y, which is based on the trained VGG network. Let $\phi(\cdot)$ be the last convolutional layer of the VGG-19 network [38]. The perceptual loss is calculated as the L1 loss between the predicted image $\hat{y}$ and the ground-truth image y; L_{perc} can be expressed as:

$$L_{perc} = \sum_{m=1}^{W} \sum_{n=1}^{H} |\phi(\hat{y})(m,n) - \phi(y)(m,n)| \tag{12}$$

Thus, our comprehensive loss function can be expressed as:

$$L = \omega_1 L_1 + \omega_2 L_{SSIM} + \omega_3 L_{perc} \tag{13}$$

where ω_1, ω_2, and ω_3 were empirically set to 1.0, 1.0, and 0.5, respectively, to balance different loss terms.

4. Experimental Results and Analysis

We first introduce the experimental datasets and evaluation metrics, along with the implementation details of the experiments, then present the underwater-image-enhancement results and the underwater-image-super-resolution results, as well as the analysis of the results, while the ablation study in the final vignette demonstrates the effectiveness of our method.

4.1. Datasets and Implementation Details

We utilized publicly available underwater image datasets to train and evaluate our network, namely UIEBD [13], SQUID [39], USR-248 [31], and UFO-120 [33]. UIEBD comprises 890 pairs of real-world underwater images and their corresponding reference images, along with 60 challenging unpaired degraded underwater images. It serves as an underwater-enhancement dataset, and the 60 unpaired images are available for testing the generalization performance. SQUID contains a large number of images captured at different locations with varying water characteristics, with 16 representative images often used for testing the performance of the enhancement methods. The USR-248 dataset consists of 1060 image pairs for training and 248 pairs for testing, serving as an underwater super-resolution dataset. The UFO-120 dataset includes 1500 training pairs and 120 testing pairs, serving as both an underwater super-resolution and enhancement dataset. In our experiments, UIEBD was randomly split into 800 pairs for training (Train-U) and the remaining 90 pairs for testing (Test-U90), while UFO-120 was split into 1500 pairs for training (Train-UFO) and 120 pairs for testing (Test-UFO120) for the underwater image enhancement task. USR-248 and UFO-120 were used for the underwater super-resolution task. The challenging 60 images from UIEBD (Test-C60) and SQUID were employed to test the generalization performance of competing enhancement methods.

The proposed network was trained using the Adam optimizer with the parameters set to $\beta_1 = 0.5$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. The initial learning rate was $5{\times}10^{-4}$, and we set $milestones = [150, 250, 400]$ with $gamma = 0.5$ for non-uniformly adjusting the learning rate. The total epoch was set to 500 with batch sizes of 10 (for SE and 2× SR) and 15 (for 4× SR). We implemented our network using the PyTorch (the version number: 1.7.1) framework and an NVIDIA RTX 2080TI GPU (NVIDIA, Santa Clara, CA, USA). Due to memory constraints, the UIEBD training set was adjusted to 256 × 256 and randomly cropped to a 128 × 128 patch. For the SE task on UFO-120, the images were randomly cropped to a 160 × 120 patch. The training set for the 2× SR task was randomly cropped to a 160 × 120 patch, while the 4× SR training set was randomly cropped to an 80 × 60 patch.

Our network has 1.36 MB parameters and can perform 2× SR on 18 images of size 320 × 240 in 1 s.

4.2. Evaluation Metrics

We employed both full-reference and non-reference image-quality metrics to comprehensively evaluate and compare the proposed method with competing methods. The full-reference metrics included the mean-squared error (MSE), the peak signal-to-noise ratio (PSNR), and the structural similarity index (SSIM), which assess the proximity and structural–textural similarity between the generated images and the reference images. For the PSNR and SSIM, higher values indicate better performance, while for the MSE, lower values are desirable. The non-reference metrics consist of the underwater image quality measure (UIQM) [40] and underwater color image quality evaluation (UCIQE) [41], which provide a comprehensive evaluation of underwater image quality based on factors such as color intensity, saturation, clarity, and contrast.

4.3. Underwater Image Enhancement Results

We compared our method with UDCP [4], ULAP [5], HE [8], UWCNN [12], Water-Net [13], FUnIE-GAN [19], Ucolor [15], URSCT-SESR [23], Deep SESR [33], and U-Trans [24].

Quantitative evaluation. In Table 1, we present the full-reference evaluation results on Test-U90 and the non-reference results on Test-C60 and SQUID. In Table 2, we provide the full-reference and non-reference evaluation results on Test-UFO120. Additionally, we present the model parameters of the competing methods in Table 1.

Table 1. The evaluation results of different methods on Test-U90, Test-C60, and SQUID. The best result is in red under each case, and the second-best is in blue.

Method	Params (M)	T-U90			Test-C60		SQUID	
		MSE	PSNR (dB)	SSIM	UIQM	UCIQE	UIQM	UCIQE
UDCP [4]	-	4542.44	12.09	0.59	1.45	0.53	0.94	0.55
ULAP [5]	-	2254.50	16.27	0.76	1.66	0.53	0.87	0.46
UWCNN [12]	0.04	3260.72	13.92	0.67	2.38	0.47	2.12	0.44
Water-Net [13]	1.09	848.16	20.05	0.85	2.67	0.56	2.34	0.54
FUnIE-GAN [19]	7.02	1448.76	17.75	0.75	2.64	0.52	2.12	0.48
Ucolor [15]	149	433.06	22.65	0.89	2.61	0.53	2.13	0.50
URSCT-SESR [23]	11.19	2097.75	16.18	0.78	1.51	0.54	1.37	0.49
U-Trans [24]	31.59	595.95	21.32	0.78	2.59	0.53	2.13	0.51
Ours	1.36	432.14	22.99	0.91	2.64	0.55	2.25	0.50

Table 2. The full-reference and non-reference evaluation results of different methods on Test-UFO120. The best result is in red under each case, and the second-best is in blue.

Method	MSE	PSNR (dB)	SSIM	UIQM
HE [8]	3433.87	13.52	0.55	2.44
UDCP [4]	3020.50	14.47	0.50	1.95
ULAP [5]	1017.88	18.77	0.65	2.22
UWCNN [12]	1984.30	16.41	0.59	2.88
Water-Net [13]	819.03	19.70	0.69	3.02
Ucolor [15]	821.32	19.56	0.69	3.10
Deep SESR [33]	332.26	23.76	0.75	3.12
Ours	268.20	24.54	0.78	2.90

As shown in Table 1, our method achieved the best performance on Test-U90. For the MSE, PSNR, and SSIM metrics, our method obtained improvements of at least 0.21%, 1.35%, and 2.25%, respectively. Additionally, our method ranked second in results on Test-C60, just behind the best-performing Water-Net [13]. On the SQUID testing dataset, our method achieved the second-highest UIQM. Table 2 indicates that the non-reference result (UIQM) of our method on Test-UFO120 was not as high as Deep SESR [33], Ucolor [15], and Water-Net [13], but our full-reference evaluation results were the best, where the PSNR improved by 3.28% compared to the second-best performer. Tables 1 and 2 demonstrate that our method had a significant advantage in improving the quality and restoring the structure and texture of the underwater images, along with good generalization performance.

Visual assessment. In Figures 4–6, we present the visual quality of the underwater images enhanced by competing methods and our method, sourced from Test-U90, Test-C60, SQUID, and Test-UFO120. As shown in Figures 4 and 6, noticeable color distortions and low sharpness are observed in the raw images. ULAP [5] and UWCNN [12] to some extent improved the quality of the degraded images, but the enhanced images exhibited some color casts. Water-Net [13] enhanced the image quality of the degraded images and alleviated the color distortion, but the overall enhanced images appear darker. FUnIE-GAN [19] improved the quality of bluish and greenish underwater images and enhanced the visual quality, but it tended to overly enhance low-illumination underwater images and introduced artifacts. Ucolor [15] effectively mitigated color casts in the degraded images and improved the image sharpness, but the enhanced images still exhibited little color bias. While URSCT-SESR [23] improved the quality of most underwater images, it exacerbated the color casts significantly and introduced over-enhanced artifacts. U-Trans [24] and Deep SESR [33] successfully removed the color deviations from the degraded images and enhanced the visual quality, but the quality of the enhanced images can still be improved. In contrast, our method not only effectively corrected the color deviations in the degraded underwater images, but also significantly improved the image quality, resulting in clear images with visually satisfying outcomes. In Figure 5, ULAP [4], FUnIE-GAN [19], and URSCT-SESR [23] exhibit severe color casts in the enhanced results of yellowish and greenish underwater images, with almost no improvement in brightness for the low-illumination underwater images. Water-Net [13] improved the visual effects of underwater images presenting different color distortions, but exhibited an overall dark appearance in the enhanced results of bluish underwater images. Ucolor [15] and U-Trans [24] effectively improved the visual quality of the underwater images with different color distortions and enhanced the brightness for low-illumination underwater images, but the enhanced images still had some color bias. Our method significantly improved the brightness and contrast of the low-illumination underwater images and enhanced the visual quality of the greenish, bluish, and yellowish underwater images.

4.4. Underwater Image Super-Resolution Results

We compared our method with several state-of-the-art methods, including SRCNN [25], DSRCNN [26], SRResNet [28], EDSRGAN [27], SRGAN [28], SRDRM [31], SRDRM-GAN [31], PAL [32], Deep SESR [33], and SwinIR-NG [30].

Quantitative evaluation. We present the quantitative evaluation results on the USR-248 and UFO-120 testing datasets, as shown in Tables 3 and 4, respectively. Additionally, we present the model parameters of the competing methods in Table 3. In Table 4, each assessment metric is a predicted value obtained by adding or subtracting the standard deviation from the mean. In experiments, metric values may vary due to factors such as noise or algorithm sensitivity, and the mean plus or minus the standard deviation quantifies this uncertainty to present more-robust data.

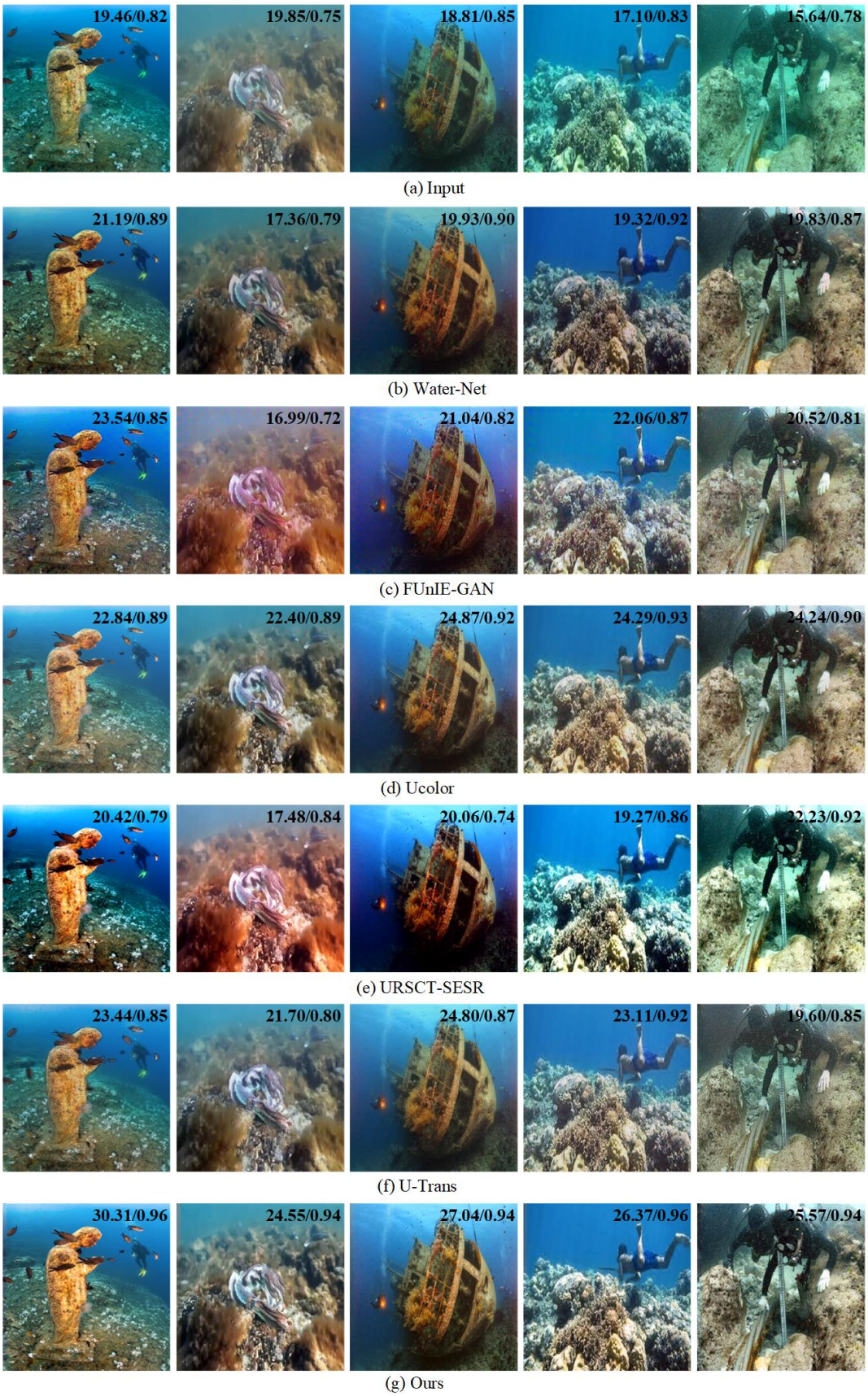

Figure 4. Visual comparisons on underwater images sampled from Test-U90. The number on the top-right corner of each image refers to its PSNR/SSIM (the larger, the better).

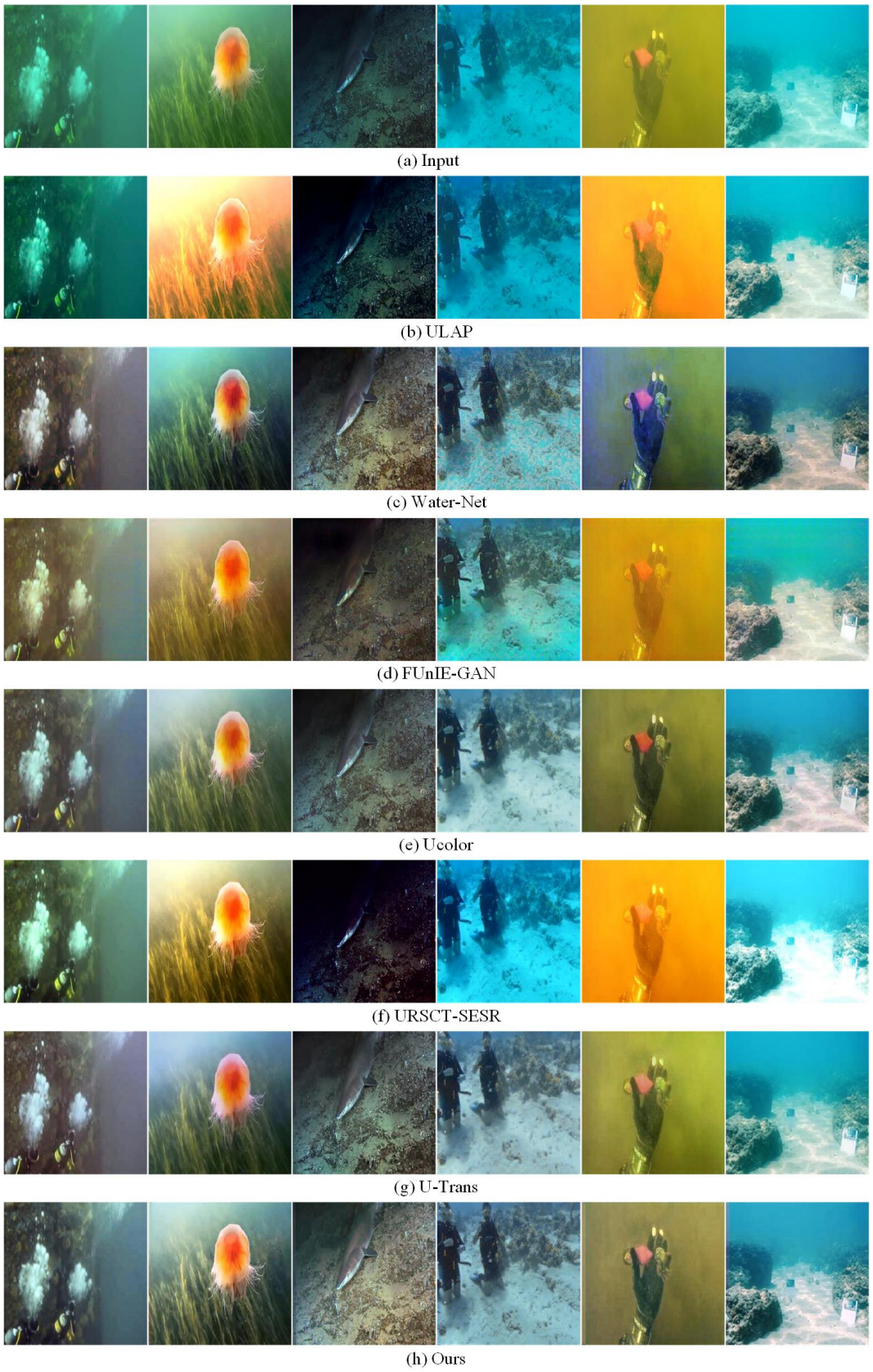

Figure 5. Visual comparisons on underwater images sampled from Test-C60 and SQUID. (**a**) Input raw image, (**b**) ULAP [4], (**c**) Water-Net [13], (**d**) FUnIE-GAN [19], (**e**) Ucolor [15], (**f**) URSCT-SESR [23], (**g**) U-Trans [24], and (**h**) ours.

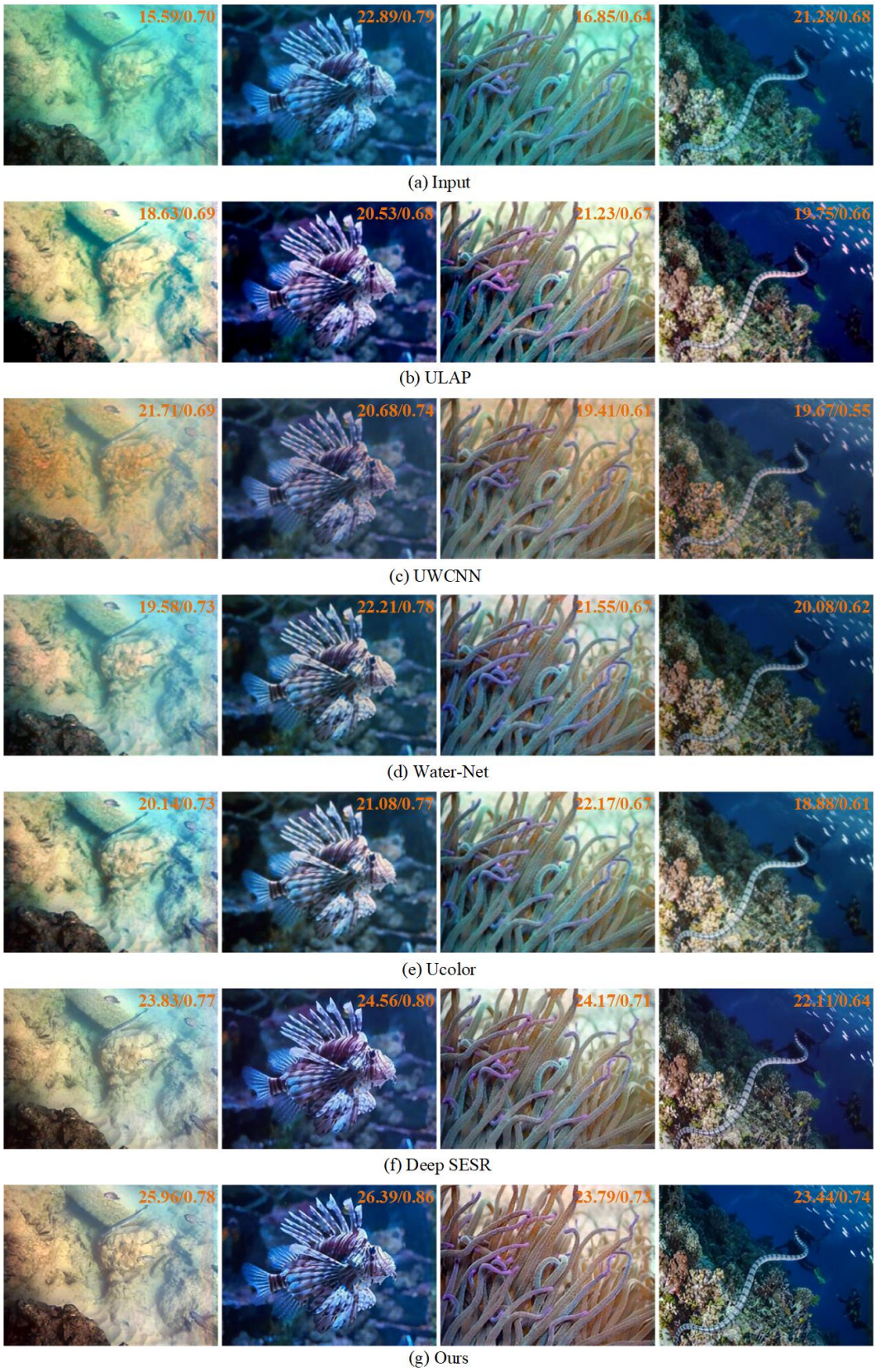

Figure 6. Visual comparisons of underwater images sampled from Test-UFO120. The number on the top-right corner of each image refers to its PSNR/SSIM (the larger, the better).

Table 3. Quantitative results on USR-248 with scale factors 2× and 4× for underwater image SR. The best result is in red under each case, and the second-best is in blue.

Method	Params (M)		PSNR		SSIM		UIQM	
	2×	4×	2×	4×	2×	4×	2×	4×
SRCNN [25]	0.06	0.06	26.81	23.38	0.76	0.67	2.74	2.38
DSRCNN [26]	1.11	1.11	27.14	23.61	0.77	0.67	2.71	2.36
EDSRGAN [27]	1.38	1.97	27.12	21.65	0.77	0.65	2.67	2.40
SRGAN [28]	5.95	5.95	28.05	24.76	0.78	0.69	2.74	2.42
SRDRM [31]	0.83	1.90	28.36	24.64	0.80	0.68	2.78	2.46
SRDRM-GAN [31]	11.31	12.38	28.55	24.62	0.81	0.69	2.77	2.48
PAL [32]	0.83	1.92	28.41	24.89	0.80	0.69	–	–
SwinIR-NG [30]	1.18	1.20	29.03	25.66	0.79	0.68	2.65	2.54
Ours	1.36	1.39	29.10	25.76	0.80	0.67	2.71	2.60

Table 4. Quantitative results on the UFO-120 with scale factors 2× and 4× for underwater image SR. The best result is in red under each case, and the second-best is in blue.

Method	PSNR		SSIM		UIQM	
	2×	4×	2×	4×	2×	4×
SRCNN [25]	24.75 ± 3.7	19.05 ± 2.3	0.72 ± 0.07	0.56 ± 0.12	2.39 ± 0.35	2.02 ± 0.47
SRResNet [28]	25.23 ± 4.1	19.13 ± 2.4	0.74 ± 0.08	0.56 ± 0.05	2.42 ± 0.37	2.09 ± 0.30
SRGAN [28]	26.11 ± 3.9	21.08 ± 2.3	0.75 ± 0.06	0.58 ± 0.09	2.44 ± 0.28	2.26 ± 0.17
SRDRM [31]	24.62 ± 2.8	22.26 ± 2.5	0.72 ± 0.17	0.59 ± 0.05	2.59 ± 0.64	2.28 ± 0.35
SRDRM-GAN [31]	24.61 ± 2.8	22.21 ± 2.4	0.72 ± 0.17	0.58 ± 0.13	2.59 ± 0.64	2.27 ± 0.44
Deep SESR [33]	25.49 ± 3.3	24.75 ± 2.8	0.71 ± 0.19	0.66 ± 0.05	2.82 ± 0.47	2.55 ± 0.35
Ours	25.79 ± 3.0	24.97 ± 2.8	0.72 ± 0.16	0.71 ± 0.17	2.66 ± 0.56	2.62 ± 0.54

As shown in Table 3, our method achieved the highest PSNR values on the USR-248 testing dataset for both scale factors 2× and 4×. For the 2× underwater SR, our method achieved SSIM values close to the top performer. At the 4× scale factor, our method achieved the best UIQM results, with an improvement of 2.36% compared to the second-best performer. In Table 4, for the 2× scale factor on the UFO-120 testing dataset, our method achieved the second-best PSNR and UIQM values. At the 4× scale factor, our method demonstrated superiority in all image-quality-evaluation metrics on the UFO-120 testing dataset, with improvements of 0.89%, 7.58%, and 2.75% in the PSNR, SSIM, and UIQM, respectively, compared to the second-best method.

Visual assessment. Figure 7 presents a visual comparison of the super-resolution results on the USR-248 testing dataset with scale factors 2× and 4× between our method and competing methods, and Figure 8 shows a visual comparison of the UFO-120 testing dataset with scale factors 2× and 4×. As shown in Figure 7, we observed that images generated by EDSRGAN [27], SRDRM-GAN [31], and SwinIR-NG [30] at a scale factor of four exhibited noticeable artifacts and distortion. Images generated by SRCNN [25], DSRCNN [26], and SRDRM [31] lacked clear texture details. SRGAN [28] and PAL [32] successfully recovered the texture details, but lacked color consistency in images generated at a scale factor of two. In contrast, our method generated images that closely resembled the HR images, with richer texture details and without unpleasant artifacts. In Figure 8, the images generated by SRCNN [25], SRResNet [28], SRGAN [28], and SRDRM [31] showed some color deviations and failed to recover the texture details at a scale factor of four. SRDRM-GAN [31] generated high-resolution images, but exhibited some color deviations. The images generated by Deep SESR [33] had color similarity with respect to the HR images, but needed improvement in terms of sharpness. In comparison, our method not only enhanced the image sharpness, but also effectively corrected the color deviations, resulting in improved visual quality.

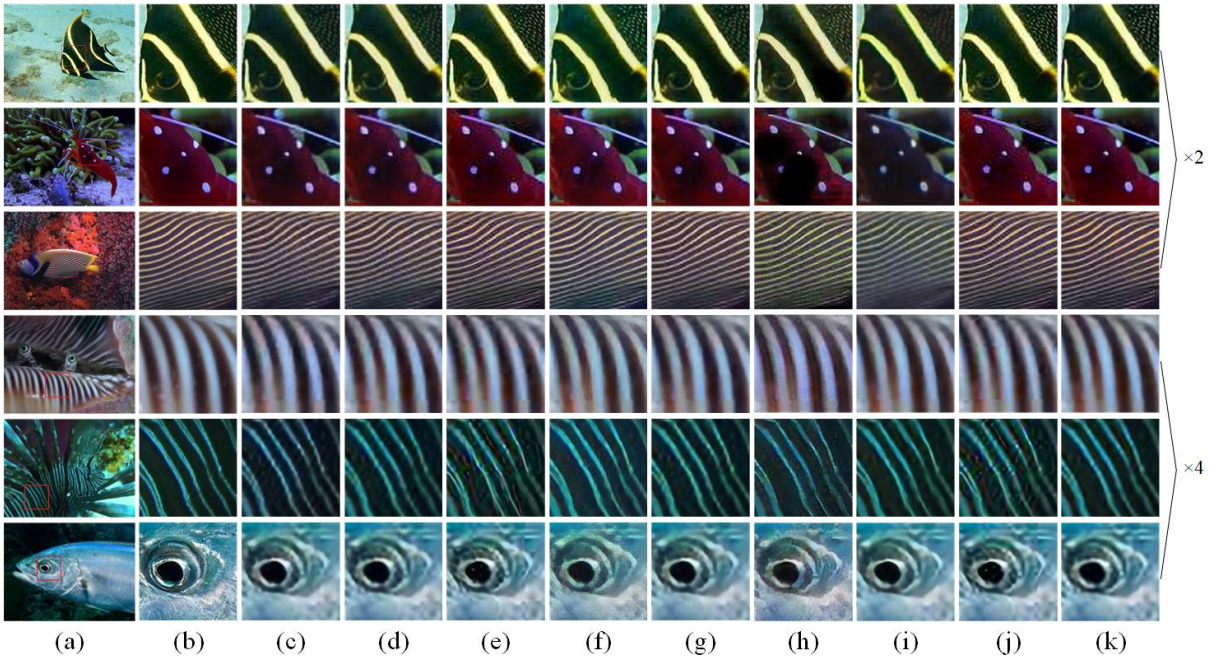

Figure 7. Visual comparisons of underwater images sampled from the USR-248 testing dataset with scale factors 2× and 4×. (**a**) Input raw image, (**b**) HR, (**c**) SRCNN [25], (**d**) DSRCNN [26], (**e**) EDSRGAN [27], (**f**) SRGAN [28], (**g**) SRDRM [31], (**h**) SRDRM-GAN [31], (**i**) PAL [32], (**j**) SwinIR-NG [30], and (**k**) ours.

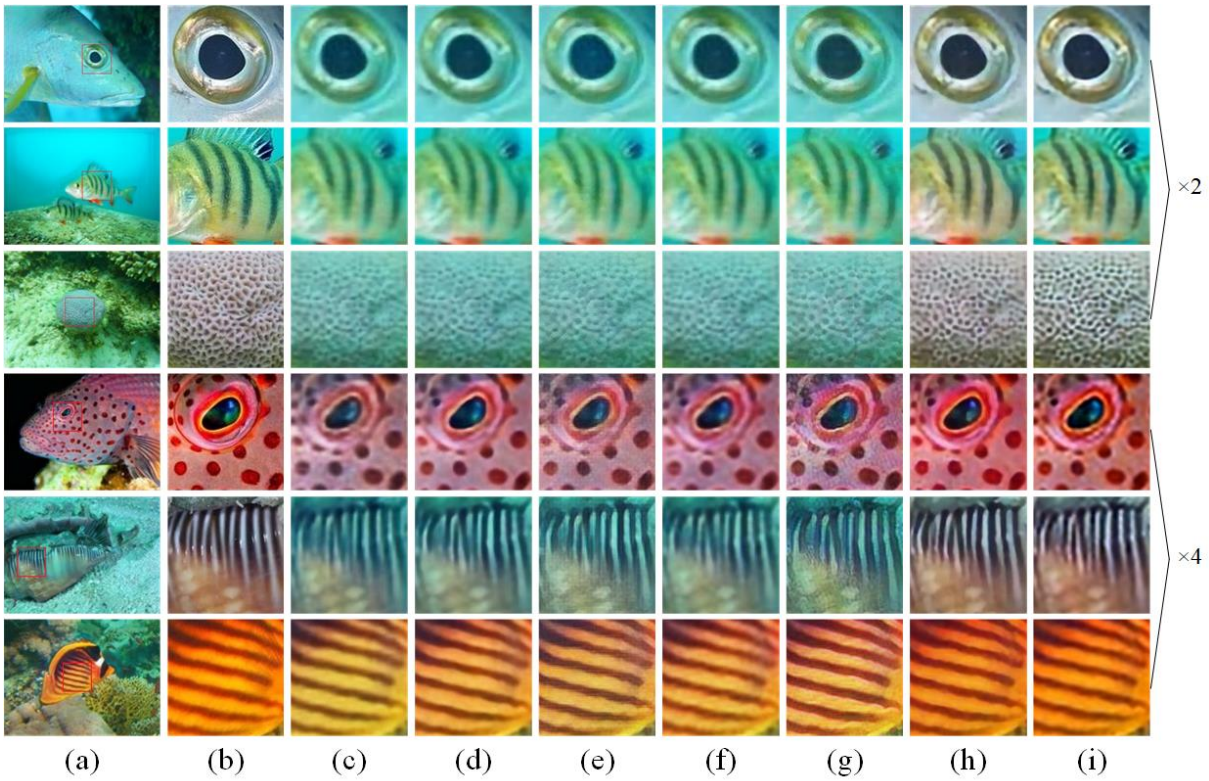

Figure 8. Visual comparisons of underwater images sampled from the UFO-120 testing dataset with scale factors 2× and 4×. (**a**) Input raw image, (**b**) HR, (**c**) SRCNN [25], (**d**) SRResNet [28], (**e**) SRGAN [28], (**f**) SRDRM [31], (**g**) SRDRM-GAN [31], (**h**) Deep SESR [33], and (**i**) ours.

4.5. Analysis and Discussion

The physical-model-based underwater-image-enhancement methods UDCP [4] and ULAP [5] rely on manual priors for estimating the parameters in an underwater imaging model. This makes their predictions susceptible to being dominated by prior information, leading to the risk of over-/under-recovered artifacts in scenes beyond the assumed conditions. The non-physical-model-based method HE [8] directly modifies the pixel values, overlooking the varying light attenuation characteristics under water, resulting in an introduction of excessive red components and exacerbating color bias. UWCNN [12], trained on a synthetically generated underwater dataset with prior knowledge of underwater scenes, may face challenges when adapting to real-world underwater image enhancement. Water-Net [13] introduces white balancing, which is not always reliable for underwater images and often leads to visually darker restored images. FUnIE-GAN [19], due to its simple model design, may easily reach bottlenecks in the complex-feature-learning process, potentially causing color casts in the recovery results. Ucolor [15], incorporating estimated transmission maps from a presumed underwater imaging model into a deep network, may struggle to adapt to all underwater scenes. URSCT-SESR [23], a U-Net-based network with multiple downsampling layers, might lose some image information, resulting in unsatisfactory outcomes. U-Trans [24], splitting the input image into fixed-size patches and independently processing each patch, could produce boundary artifacts. Given the distinct nature of underwater imaging compared to terrestrial scenes, super-resolution methods like SRCNN [25], SRGAN [28], and SwinIR-NG [30] may struggle to effectively handle degraded underwater images with color distortion. Residual-learning-based deep networks like SRDRM [31], SRDRM-GAN [31], and Deep SESR [33] might be constrained in performance due to insufficient attention to long-range dependencies. PAL [32], a progressive network based on channel attention, may fail to provide satisfactory results when dealing with underwater images exhibiting significant color distortion or low illumination, as the RGB color space is insensitive to saturation and brightness. Our method utilizes the designed non-local attention module with high expressiveness and low complexity to model long-range dependency in the multi-color space, enabling the generation of visually pleasing underwater images.

4.6. Ablation Study

We conducted ablation studies to demonstrate the effectiveness of multi-color space utilization, multi-receptive field feature extraction (i.e., using convolution operations with different kernels k), and the agent-guided non-local attention module (Ag-NL module). Specifically, we trained the models in different cases on the UIEBD dataset for underwater image enhancement, then quantitatively evaluated the results on Test-U90 using the full-reference metrics MSE, PSNR, and SSIM, and assessed the performance on Test-C60 using the non-reference metrics UIQM and UCIQE. Table 5 presents the model parameters, FLOPs, and evaluation results, and visual comparisons are shown in Figure 9. As shown in Table 5, using only the RGB color space and features extracted with the same receptive field size yielded decent results, but employing the multi-color space and different receptive fields further improved the performance. It is noteworthy that, compared to models without the Ag-NL module, our model, while having slightly higher complexity, exhibited significantly better performance. From Figure 9, we can observe that the images generated by our method had the highest image quality and the best visual effects, while the images generated in other cases may exhibit some color bias or unsatisfactory clarity. Table 5 and Figure 9 demonstrate that utilizing the multi-color space and multi-receptive field feature extraction is beneficial for capturing more-diverse features, thereby enhancing network performance. The agent-guided non-local attention module (Ag-NL module) efficiently modeled long-range dependency with lower computational complexity, significantly boosting network performance.

Table 5. Image quality assessment of models in different cases.

Model	Params (M)	FLOPs (G)	T-U90			Test-C60	
			MSE	PSNR (dB)	SSIM	UIQM	UCIQE
with only RGB color space	1.36	44.54	447.15	22.91	0.90	2.43	0.54
with same receptive field ($k = 5$)	1.25	40.96	453.41	22.84	0.91	2.45	0.55
without Ag-NL module	1.19	39.06	785.66	20.21	0.89	2.50	0.51
ours	1.36	44.54	432.14	22.99	0.91	2.64	0.55

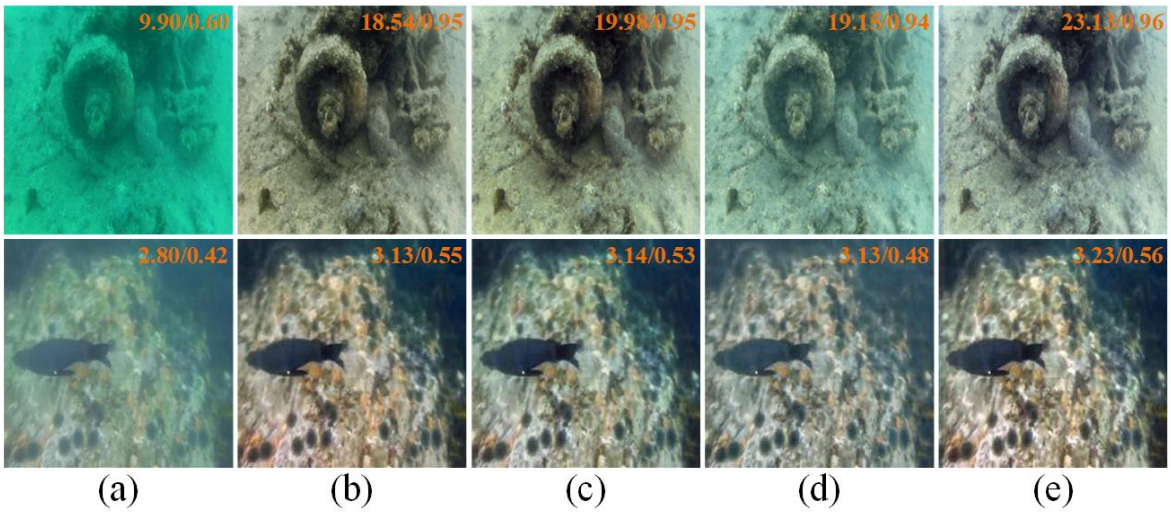

Figure 9. From left to right: (**a**) input raw image, (**b**) Results of only using RGB color space, (**c**) results with same receptive field, (**d**) results without Ag-NL module, and (**e**) ours. Top row: the sample is from Test-U90, and the number on the top-right corner of each image refers to its PSNR/SSIM (the larger, the better); bottom row: the sample is from Test-C60, and the number on the top-right corner of each image refers to its UIQM/UCIQE (the larger, the better).

5. Conclusions

In this paper, we proposed an agent-guided non-local attention-based network using the multi-color space for underwater image enhancement and super-resolution. The network consists of four parts: shallow feature extraction, deep feature extraction, adaptive fusion, and image reconstruction. Specifically, in the shallow-feature-extraction part, we simultaneously extracted features from the RGB color space and Lab/HSI color spaces, which are closer to human perception of color, enriching the color diversity of features for underwater images. Multiple designed agent-guided non-local attention modules were seamlessly applied in the deep-feature-extraction part, efficiently accomplishing long-range dependency modeling and aiding the network in better understanding the global structure of underwater scenes. Extensive experimental results demonstrated that the proposed method outperformed other state-of-the-art methods on multiple benchmark datasets. However, the proposed method has some weaknesses, such as training data dependency and computational complexity. The limited availability of diverse and representative training data may affect the robustness and effectiveness of the method, especially under challenging underwater conditions. The computational complexity of the method makes it less suitable for real-time applications. In the future, we will aim to extend the proposed approach to underwater video enhancement and super-resolution to meet the demands of practical applications and enhance the practicality and applicability of underwater-image-processing technology.

Author Contributions: Conceptualization, R.W. and Y.Z. (Yulu Zhang); methodology, R.W.; software, R.W.; validation, R.W. and Y.Z. (Yulu Zhang); formal analysis, R.W. and Y.Z. (Yulu Zhang); data curation, Y.Z. (Yonghui Zhang); writing—original draft preparation, R.W.; writing—review and editing, Y.Z. (Yonghui Zhang). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by 'Key Research and Development Project of Hainan Province of China' (Grant No. ZDYF2019024).

Data Availability Statement: Data are contained within the article.

Acknowledgments: We thank the anonymous reviewers for their comments and suggestions, which greatly improved the manuscript.

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

References

1. Anwar, S.; Li, C. Diving deeper into underwater image enhancement: A survey. *Signal Process. Image Commun.* **2020**, *89*, 115978. [CrossRef]
2. Zhou, J.; Yang, T.; Zhang, W. Underwater vision enhancement technologies: A comprehensive review, challenges, and recent trends. *Appl. Intell.* **2023**, *53*, 3594–3621. [CrossRef]
3. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 2341–2353.
4. Drews, P.L.J.; Nascimento, E.R.; Botelho, S.S.C.; Campos, M. Underwater depth estimation and image restoration based on single images. *IEEE Comput. Graph. Appl.* **2016**, *36*, 24–35. [CrossRef]
5. Song, W.; Wang, Y.; Huang, D.; Tjondronegoro, D. A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration. In Proceedings of the Advances in Multimedia Information Processing–PCM 2018: 19th Pacific-Rim Conference on Multimedia, Hefei, China, 21–22 September 2018; Springer International Publishing: Cham, Switzerland, 2018; pp. 678–688.
6. Song, W.; Wang, Y.; Huang, D.; Liotta, A.; Perra, C. Enhancement of underwater images with statistical model of background light and optimization of transmission map. *IEEE Trans. Broadcast.* **2020**, *66*, 153–169. [CrossRef]
7. Liang, Z.; Ding, X.; Wang, Y.; Yan, X.; Fu, X. GUDCP: Generalization of underwater dark channel prior for underwater image restoration. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 4879–4884. [CrossRef]
8. Singhai, J.; Rawat, P. Image enhancement method for underwater, ground and satellite images using brightness preserving histogram equalization with maximum entropy. In Proceedings of the International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007), Sivakasi, India, 13–15 December 2007; IEEE: Piscataway, NJ, USA, 2007; Volume 3, pp. 507–512.
9. Fu, X.; Zhuang, P.; Huang, Y.; Liao, Y.; Zhang, X.; 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; IEEE: Piscataway, NJ, USA, 2014; pp. 4572–4576.
10. 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; IEEE: Piscataway, NJ, USA, 2012; pp. 81–88.
11. Ancuti, C.O.; Ancuti, C.; Vleeschouwer, C.D.; Bekaert, P. Color balance and fusion for underwater image enhancement. *IEEE Trans. Image Process.* **2018**, *27*, 379–393. [CrossRef]
12. Li, C.; Anwar, S.; Porikli, F. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognit.* **2020**, *98*, 107038. [CrossRef]
13. 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]
14. Hu, J.; Jiang, Q.; Cong, R.; Gao, W.; Shao, F. Two-branch deep neural network for underwater image enhancement in HSV color space. *IEEE Signal Process. Lett.* **2021**, *28*, 2152–2156. [CrossRef]
15. Li, C.; Anwar, S.; Hou, J.; Cong, C.; Guo, C.; Ren, W. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Trans. Image Process.* **2021**, *30*, 4985–5000. [CrossRef]
16. Chen, Y.; Li, H.; Yuan, Q.; Wang, Z.; Hu, C.; Ke, W. Underwater Image Enhancement based on Improved Water-Net. In Proceedings of the 2022 IEEE International Conference on Cyborg and Bionic Systems (CBS), Wuhan, China, 24–26 March 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 450–454.
17. Yan, J.; Wang, Y.; Fan, H.; Huang, J.; Grau, A.; Wang, C. LEPF-Net: Light Enhancement Pixel Fusion Network for Underwater Image Enhancement. *J. Mar. Sci. Eng.* **2023**, *11*, 1195. [CrossRef]
18. 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; IEEE: Piscataway, NJ, USA, 2018; pp. 7159–7165.

19. Islam, M.J.; Xia, Y.; Sattar, J. Fast underwater image enhancement for improved visual perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [CrossRef]
20. Hambarde, P.; Murala, S.; Dhall, A. UW-GAN: Single-image depth estimation and image enhancement for underwater images. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [CrossRef]
21. Zhang, S.; Zhao, S.; An, D.; Li, D.; Zhao, R. MDNet: A Fusion Generative Adversarial Network for Underwater Image Enhancement. *J. Mar. Sci. Eng.* **2023**, *11*, 1183. [CrossRef]
22. Cong, R.; Yang, W.; Zhang, W.; Li, C.; Guo, C.; Huang, Q.; Kwong, S. PUGAN: Physical Model-Guided Underwater Image Enhancement Using GAN with Dual-Discriminators. *IEEE Trans. Image Process.* **2023**, *32*, 4472–4485. [CrossRef]
23. Ren, T.; Xu, H.; Jiang, G.; Yu, M.; Zhang, X.; Wang, B.; Luo, T. Reinforced swin-convs transformer for simultaneous underwater sensing scene image enhancement and super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [CrossRef]
24. Peng, L.; Zhu, C.; Bian, L. U-shape transformer for underwater image enhancement. *IEEE Trans. Image Process.* **2023**, *32*, 3066–3079. [CrossRef]
25. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [CrossRef]
26. Mao, X.J.; Shen, C.; Yang, Y.B. Image restoration using convolutional auto-encoders with symmetric skip connections. *arXiv* **2016**, arXiv:1606.08921.
27. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.
28. Ledig, C.; Theis, L.; Huszár, F.; Cunningham, J.C.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; Shi, W. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
29. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1833–1844.
30. Choi, H.; Lee, J.; Yang, J. N-gram in swin transformers for efficient lightweight image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 2071–2081.
31. Islam, M.J.; Enan, S.S.; Luo, P.; Sattar, J. Underwater image super-resolution using deep residual multipliers. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 900–906.
32. Chen, X.; Wei, S.; Yi, C.; Quan, L.; Lu, C. Progressive attentional learning for underwater image super-resolution. In *Intelligent Robotics and Applications: Proceedings of the 13th International Conference, ICIRA 2020, Kuala Lumpur, Malaysia, 5–7 November 2020*; Springer International Publishing: Cham, Switzerland, 2020; pp. 233–243.
33. Islam, M.J.; Luo, P.; Sattar, J. Simultaneous Enhancement and Super-Resolution of Underwater Imagery for Improved Visual Perception. In *16th Robotics: Science and Systems, RSS 2020*; MIT Press Journals: Cambridge, MA, USA, 2020.
34. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–4.
36. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
37. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
38. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
39. 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**, *4*, 2822–2837. [CrossRef]
40. Panetta, K.; Gao, C.; Agaian, S. Human-visual-system-inspired underwater image quality measures. *IEEE J. Ocean. Eng.* **2015**, *41*, 541–551. [CrossRef]
41. Yang, M.; Sowmya, A. An underwater color image quality evaluation metric. *IEEE Trans. Image Process.* **2015**, *24*, 6062–6071. [CrossRef]