Journal of Marine Science and Technology

Volume 32 | Issue 2 Article 5

Underwater Image Enhancement Algorithm for Dual Color Spaces

Xingsheng Shen
Henan University, Software College, Longting District, Kaifeng, China, 475000

Yalin Song
Henan University, Software College, Longting District, Kaifeng, China, 475000, syl@vip.henu.edu.cn

Shichang Li
Henan University, Software College, Longting District, Kaifeng, China, 475000

Xiaoshu Hu
Henan University, Software College, Longting District, Kaifeng, China, 475000

REVIEW

Underwater Image Enhancement Algorithm for Dual Color Spaces

Xingsheng Shen, Yalin Song*, Shichang Li, Xiaoshu Hu

Henan University, Software College, Longting District, Kaifeng, 475000, China

Abstract

Targeting issues related to low contrast, blurring, and loss of detail prevalent in underwater image enhancement algorithms, we propose a dual-color space multiscale residual network (DMR-SCNet) based on SCNet. First, we introduce the HSV color space feature extraction module, which aims to optimize the color representation and saturation of underwater images. Subsequently, we propose the RGB color space denoising module, which focuses on repairing the content and structure of underwater images to enhance their clarity and visual quality. Finally, by designing the residual attention (RAB) module, we aim to further refine the detailed representation and feature extraction of underwater images.

The results obtained from assessments conducted on the UIEBD and EUVP datasets indicate that the proposed method outperforms current prevalent deep learning methodologies, showing superior performance in terms of the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and underwater image quality measure (UIQM). Our findings moreover indicate that DMR-SCNet yields substantial improvements in underwater image enhancement within diverse underwater environments. This approach shows promise for implementation in underwater image processing to markedly enhance the overall quality and usefulness of underwater imagery, offering broad applications in underwater visual data enhancement.

Keywords: Underwater image enhancement, Image processing, Convolutional neural network, Residual learning

1. Introduction

The ocean covers approximately 71% of the Earth's surface. In light of the scarcity of resources on land, the development and use of marine resources have drawn significant attention from nations worldwide. Underwater images are essential for exploring and utilizing underwater resources. Frequently, such images degrade because of absorption and scattering, which are influenced by wavelength and distance. Water absorbs more red light than blue and green light, leading to an absorption effect and the characteristic blue–green color cast of underwater images. Scattering phenomena, such as forward and backward scattering, are caused by the suspension of sediment particles in water. These particles create uniform background noise and a foggy appearance, thus degrading image quality. Poor-quality underwater images can negatively affect subsequent underwater visual tasks, including underwater target detection and classification, as well as other visual processing tasks. The substandard quality of underwater images significantly hampers activities such as underwater resource exploration, underwater archaeology, and marine military operations. Hence, it is crucial to develop a more proficient approach to enhancing subaquatic images to obtain superior underwater image quality and to further investigate the marine world.

Early methods for enhancing underwater images can be divided into two main categories: physical model-based methods and nonphysical models. Nonphysical model image enhancement techniques

Received 27 November 2023; revised 24 April 2024; accepted 26 April 2024.
Available online 21 June 2024

* Corresponding author.
E-mail address: syl@vip.henu.edu.cn (Y. Song).

https://doi.org/10.51400/2709-6998.2737

aim to enhance image data at the pixel level without considering the physical processes involved in image generation. Thus, they adjust pixel intensity or color values to enhance image quality while disregarding the influence of optics, sensors, or other physical phenomena related to image formation. These methods include histogram stretching [1], histogram equalization [2], white balancing [3], grayscale edge assumption [4], and methods based on retinex theory [5]. Physical model-based approaches address underwater image enhancement as an image restoration challenge. This technique diminishes the complexity of coping with ambiguity through an inverse problem solution that utilizes prior information from natural images. Common algorithms in this category include the dark channel prior method [6], red channel prior method [7], adaptive attenuation curve prior method [8], and minimum information prior method [9]. While conventional algorithms consider the causes of degradation in underwater images and the impact on image pixel values, they possess certain limitations and frequently fail to yield exceptional outcomes when applied to underwater images.

However, the use of deep learning techniques has led to impressive improvements in underwater image quality. The algorithm proposed by WaterNet et al. [10] utilizes a gated fusion network architecture for underwater image enhancement. Although the quantitative analysis results are subpar, there is significant potential for improvement. Li et al. [11] introduced a UWCNN network model that leverages deep learning techniques to improve the clarity and detail of images captured in underwater environments, taking into account the unique characteristics and prior knowledge of underwater scenes. However, accurate predictions by a single model are unachievable, and there exists an issue of overcompensation for images that exhibit significant attenuation. Slam et al. [12] introduced FUnIE-GAN, a computational model that exhibits high processing efficiency and demonstrates effective color restoration capabilities for underwater images. However, the model is trained primarily on synthetic image pairs, thereby disregarding the inherent differences between virtual and actual scene images. Consequently, its efficiency tends to deteriorate when evaluated on real-world data. Gong et al. [13] proposed an underwater image enhancement method based on color feature fusion. This method takes advantage of the characteristics of underwater light propagation and adopts a multichannel feature extraction strategy. It uses convolution blocks of different sizes to extract features from the three channels of red, green and blue to effectively capture global and local information in underwater images. In addition, the researchers introduced an attention mechanism to design a residual enhancement module, thus enhancing feature representation. However, depending on water conditions, this method may be overenhanced or underenhanced. Li et al. [14] proposed a novel neural network structure named the multi-channel attention network (MCANet). This network effectively integrates features in different color spaces by introducing a multicolour space encoder and uses the multichannel attention path aggregation strategy to acquire deeper global and local features of an image from multiple dimensions. Finally, the network's ability to perceive and learn from image features is continuously improved by embedding and stacking multichannel attention modules multiple times.

In summary, while deep learning algorithms focus more on color correction in enhancing underwater images, thereby amplifying their effectiveness, most of them use only a single color space for image enhancement without considering the complex environment of the underwater image. A single color space cannot comprehensively process the chromatic details within the image or improve the color saturation, hue, or other aspects of the problem.

Given these issues, we propose a model for enhancing underwater images in dual-color space. The primary contributions of this study can be summarized as follows:

1. The present study introduces the structure of both the RGB color model and the HSV color model. Among the array of color models, the HSV color space can be globally adjusted using a neural curve after the deep learning convolution layer. This adjustment effectively enhances the background and tone of underwater images. Subsequently, the processed underwater image is further refined using an RGB color space denoising network.
2. The performance of the RGB denoising network is improved. Based on the U-Net architecture, this paper modifies each convolution module by employing multiscale expansion convolution instead of conventional convolution. This approach enables more effective extraction of multiscale features, expansion of the receptive field, and utilization of contextual information. A more comprehensive understanding of the visual information and composition within the image facilitates the precise elimination of noise in underwater images and the subsequent restoration of clear visuals.

3. A module called the residual attention block (RAB) module is proposed. We suggest incorporating channel attention and spatial attention mechanisms into the convolutional layer, along with the addition of skip connections for feature fusion. The objective is to improve the performance of underwater image restoration tasks by reducing noise propagation and improving the overall usability of the restored images. This study aims to provide information and enhance the quality of images.

2. Related work

2.1. HSV color model and RGB color model feature extraction

Wang et al. [15] proposed the UIEC^2-Net model, which combines the RGB color space and HSV color space into a unified convolutional neural network. The network consists of three modules, followed by RGB pixel-level blocks to achieve basic denoising and removal of the cast; the HSV color space global adjustment block, which employs a layer of neural curves to adjust the luminance, hue, and saturation of underwater images; and the attention module, which assigns weight to the output of the RGB and HSV blocks at the pixel level using an attention mechanism, aiming to achieve superior enhancement for underwater images. According to UIEC^2-Net, we introduce the HSV color space feature extraction module to improve the color deviation of underwater images and introduce the pixel weight of the HSV color space in the attention module to enhance image quality.

FU et al. [16] proposed the SCNet model, which performs normalization separately for spatial and channel dimensions at every scale of the U-Net architecture to obtain multiscale representations. Specifically, spatial normalization is conducted using instance whitening techniques to mitigate the impact of varying water conditions at the skip connections of the U-Net model [17]. PONO-MS [18] is added to the network to input the previous activations in the U-Net encoder. Both the mean and standard deviation are used to direct the models toward the matching decoder modules of the U-Net architecture, while using a 1×1 convolution to optimize the channel. PONO-MS is normalized across channels. This method enables the model to grasp the structural intricacies of a particular input image effectively by extracting relevant features and reinjecting/transmitting them to the subsequent layer, which can effectively improve network performance. However, during training, the network lacks crucial information regarding the image content. In this paper, we construct an RGB color space denoising module based on the SCNet model and enhance the restoration of underwater images by improving the convolutional layer and attention block of the SCNet.

2.2. Multiscale dilated convolution

In a traditional convolutional neural network (CNN), the convolution operation typically involves using a fixed-size convolutional kernel to conduct convolutions on the input data. This method may have some limitations when dealing with objects or features at different scales. Multiscale convolution solves the inability of traditional convolution operations to capture features of various scales by incorporating multiple convolution kernels of differing sizes.

The main idea of multiscale convolution is to apply convolutional operations to an image at different scales, thereby considering both local details and global structures to improve denoising effects. Multiscale convolution is prevalent in the domain of image processing. Zhang et al. [19] introduced the concept of a multiscale convolutional network (DnCNN) for image denoising. The DnCNN structure consists of multiple convolutional layers and residual connections, which reduce noise through the learning of residuals. Lehtinen et al. [20] proposed a denoising method that does not require clean image data. It employs multiscale convolution to train a denoising network by applying multiple random transformations and reconstructions to noisy images to generate training data. With the introduction of dilated convolution [21], many researchers have combined dilated convolution with multiscale convolution for various visual tasks. For instance, Chen et al. [22] proposed the DeepLab network, which combines multiscale convolution and dilated convolution for semantic image segmentation tasks. In the context of image denoising, Liu et al. [23] presented an image restoration method that utilizes a combination of multilevel wavelet transforms and convolutional neural networks, incorporating multiscale convolution and dilated convolution operations to address image noise. Wang et al. [24] proposed an image-denoising technique that leverages deep convolutional neural networks and incorporates multiscale convolution and dilated convolution to accomplish multiscale feature fusion.

The incorporation of the multiscale dilated convolution module in this study is predicated on the distinctive attributes of both multiscale

convolution and dilated convolution. The integration of the convolution layer into the RGB denoising module aims to enhance the model's field of view, thereby facilitating the effective extraction of images with varying scales and spatial ranges. The inclusion of characteristics that can capture both overall and specific features is paramount in preserving the structural integrity of an image and enhancing the effectiveness of image denoising.

2.3. Attention mechanism and residual connections

The integration of the attention mechanism has been demonstrated to augment the model's expressive and learning capacities, enabling it to effectively address intricate tasks and handle extensive datasets. By employing a dynamic focus on crucial elements of the input, the model can enhance its information processing capabilities, leading to improved performance and generalization. Park et al. [25] introduced a lightweight attention module called the channel and spatial attention mechanism (CBAM), as depicted in Fig. 1. This module enhances the representation ability of feature maps by incorporating channel attention and spatial attention mechanisms. Thus, it can effectively distinguish between useful information and noise in the image, thereby improving the denoising effect. He et al. [26] introduced the residual network (ResNet), which incorporates the novel concept of residual blocks. By incorporating residual connections within these blocks, the network can learn mapping of both identity and differences. This approach effectively addresses the issue of vanishing gradients and allows for increased network depth.

Based on the aforementioned concepts, the RAB module incorporated in the proposed model introduces the CBAM attention mechanism and residual connection. This integration allows the model to effectively capture residual information and assign appropriate weights to features, thereby emphasizing the importance of certain features through increased attention. The performance of the image enhancement model is thus enhanced.

3. Dual-color space multiscale residual network model (DMR-SCNet)

In this paper, we introduce a novel dual-color space multiscale residual network model. The model consists of three main modules: the HSV color space feature extraction module, the RGB color space denoising module, and the residual attention block (RAB) module. The overall architecture of our model is shown in Fig. 2.

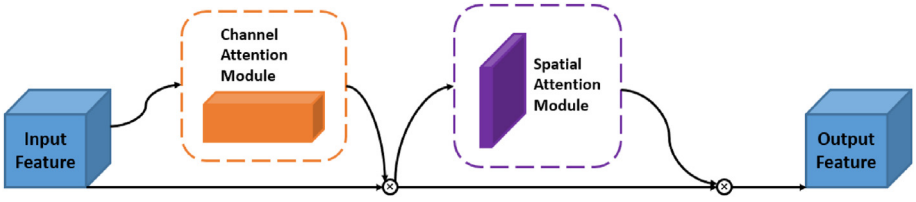

Fig. 1. Diagram of CBAM attention structure.

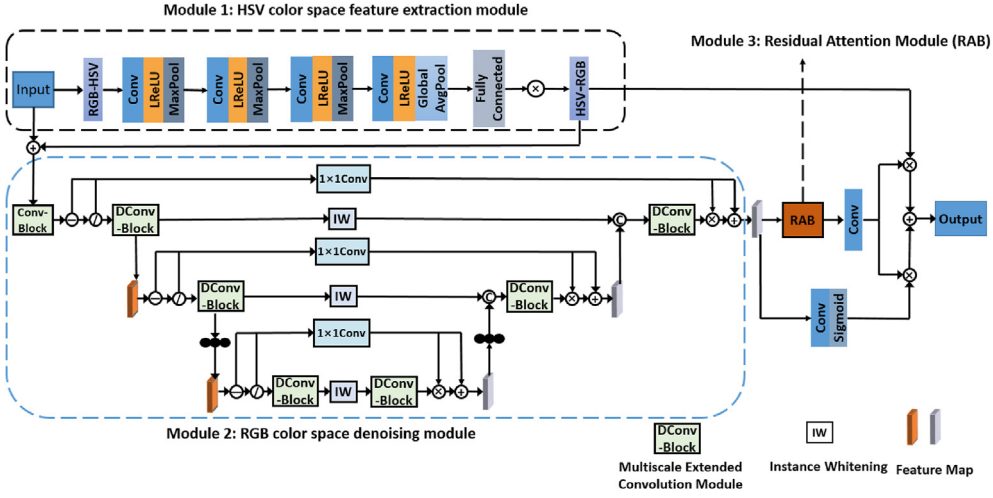

Fig. 2. DMR-SCNet model structure diagram.

3.1. HSV color space feature extraction module

The RGB color space is a prevalent color model employing three channels, red (R), green (G), and blue (B), to define the colors within an image. It is difficult to restore the image using only the RGB space; however, the chromatic details in underwater images can be seriously distorted and shifted as a result of light propagation and absorption within water. To overcome this problem, this paper employs an approach that enhances the effectiveness of underwater image restoration through the integration of the HSV (hue, saturation, value) color space. A straightforward neural network curve layer is introduced within the HSV color space to perform piecewise linear scaling of HSV color attributes, especially hue and saturation, thereby alleviating the color distortion problem of underwater images.

According to UIEC^2-Net [15], we propose the HSV color space feature extraction module. The original underwater image serves as the input for the feature extraction module within the HSV color space. The RGB image is first transformed into an HSV image, followed by three 3x3 convolutional layers with a stride of 1. After each convolutional layer, a leaky ReLU activation function and max pooling are applied. After applying a global average pooling layer that condenses the feature map into a 1x1xC shape, a fully connected layer is subsequently appended to it, which is used to return the piecewise linear curve node. This curve resizes the predicted image by adjusting the pixel sizes through the formula described in Equation (1).

$$S\left(\widehat{I}_i^{jl}\right) = k_0 + \sum_{m=0}^{M-1}(k_{m+1} - k_m)\delta\left(M\widehat{I}_i^{jl} - m\right) \quad (1)$$

where

$$\delta(x) = \begin{cases} 0\, x < 0 \\ x\, 0 \le x \le 1 \\ 1\, x > 1 \end{cases}$$

M represents the number of predicted knot points, $\widehat{I}_i^{jl}$ denotes the value of the jth pixel in the 1st color channel of the i-th image, and k_m represents the value associated with node M. The core of the HSV color space feature extraction module involves multiplying the pixel value by the curve's scale factor to obtain an enhanced global adjustment image.

The HSV space separates the values of the three channels of the image—hue, saturation and value. The image properties are optimized by the corresponding adjustment curves. The optimized HSV image is converted back to RGB space by differentiable HSV to RGB conversion, and then the resulting features are combined with the original underwater input image through a long jump join, passing the features to the RGB color space denoising module of the model. This process improves the stability of model training while combining the basic features from the underlying original image.

3.2. RGB color space denoising module

To address the issues of blurring or lack of detail in underwater images with complex textures or subtle structures, this study adopts a novel approach. Specifically, we introduce an RGB color space denoising module and incorporate a multiscale dilated convolution to replace the ordinary convolution in the U-Net encoder and decoder components. This modification aims to enhance the network performance in handling underwater images with complex textures or subtle structures. Ordinary convolution operations employ convolution kernels of fixed size, thereby lacking the ability to directly capture multiscale information. Therefore, the use of multiscale dilated convolution allows the network to effectively extract features at various scales, which enables the network to adapt to feature changes occurring at different scales while also capturing larger receptive field features. Consequently, this approach enhances the network's ability to comprehend the features present in the image.

The multiscale dilated convolution structure is based on the inception module. It consists of three parallel convolution branches, each utilizing convolutions that are 3×3, 5×5, and 7×7. To decrease the computational demands of the model, the 5×5 and 7×7 convolutions are decomposed, and smaller, more lightweight 3×3 convolutions are employed as substitutes. Different dilation rates are applied to each parallel convolution branch to perform dilated convolutions. By employing dilated convolutions with varying dilation rates, this technique facilitates feature extraction at multiple scales, thereby enhancing fine details in underwater images. This approach additionally aids in mitigating concerns associated with diminished feature map resolution and potential information loss that may occur when multiple convolution layers are stacked. Subsequently, the outputs from the three branches are combined to extract spatial features at various scales. Subsequently, a 1×1 convolution is employed to modify the channel dimensions,

thereby reducing the parameter count and computational load of the model for subsequent operations. Finally, to enable the addition of input and output features in an elementwise addition, a residual connection is introduced by incorporating a 1×1 convolution. The purpose of this addition is to enhance the network's ability to preserve and transmit essential image information during training, which facilitates the recovery of local image features. Fig. 3 depicts the architecture of the multiscale dilated convolution.

3.3. RAB residual attention module

While the encoder-decoder architecture of the RGB color space denoising module effectively captures multiscale features through multiple upsampling operations, it is nevertheless susceptible to the loss of pixel-level spatial detail features. To preserve the intricate texture characteristics present in the input image, the CBAM module and residual connections are employed within the RAB residual attention module. As depicted in Fig. 4, the module initially undergoes convolution and activation, followed by another convolution operation. Subsequently, the CBAM attention module is employed, and residual connections are utilized to address these challenges when training a deep network. The CBAM module aims to capture both spatial and channel attention in images. The system comprises two primary modules, namely, the channel attention module (CAM) and the spatial attention module (SAM).

The CAM submodule adaptively adjusts the importance of different channels by learning channel weights to enhance the network expressiveness, allowing the model to better capture the valuable characteristics present within the input data and thus improving underwater image enhancement.

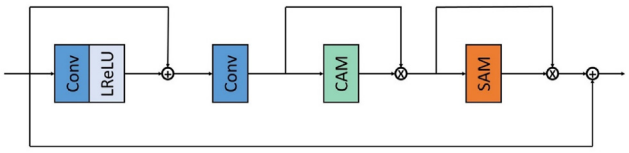

Fig. 4. Diagram of the RAB module structure.

The CAM submodule structure is shown in Fig. 5. The input features ($C \times H \times W$) are first subjected to global maxpooling and global average pooling to obtain shape features ($C \times 1 \times 1$). The features are subsequently input into a multilayer perceptron (MLP) to learn the channel weight. This MLP network consists of one or more fully connected layers, each of which introduces nonlinearities using a rectified linear unit (ReLU) activation function, and the outputs of the MLP represent the weight coefficient of each channel. The resulting features from the MLP undergo elementwise summation, followed by the application of a sigmoid activation function, thereby producing the ultimate channel attention weight.

The SAM submodule is used to adaptively adjust the importance of each spatial position in the feature map to more accurately grasp the spatial dimensions of the features and restore the primary content of the image so that the network can discriminate diverse localized regions and will pay more attention to more important and harder to augment regions. Fig. 6 illustrates the structure of the SAM submodule. First, the input feature map is subjected to max-pooling and average-pooling operations, resulting in two separate feature maps, each with dimensions of ($H \times W \times 1$). The two feature maps are then combined using the concatenation operation. Then, through a 7×7 convolution operation to change the feature map feature channel to 1, the spatial attention feature is generated by a sigmoid activation function.

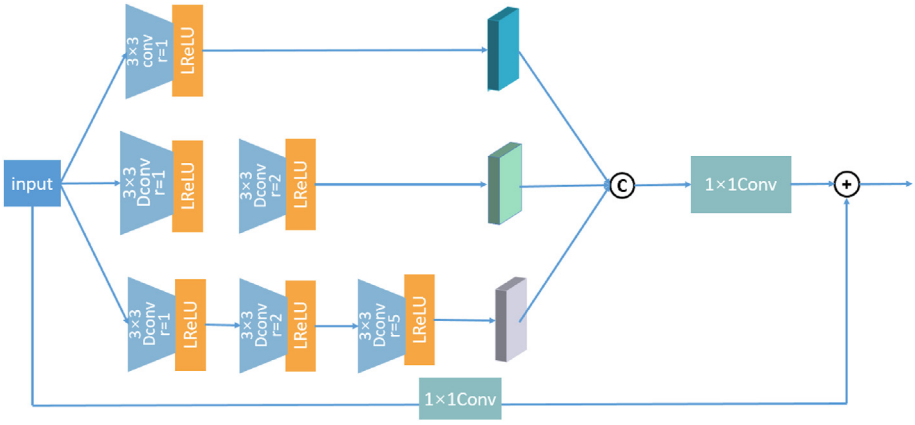

Fig. 3. Diagram of the multiscale dilated convolution structure.

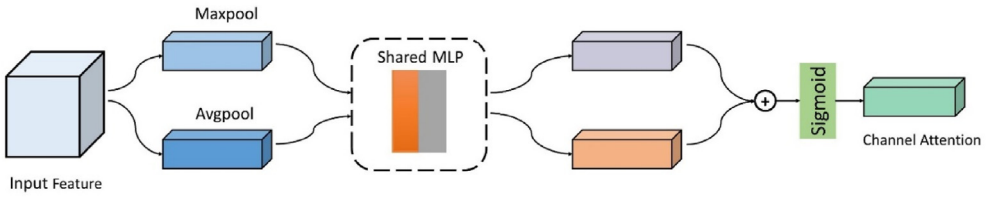

Fig. 5. Diagram of the channel attention submodule structure.

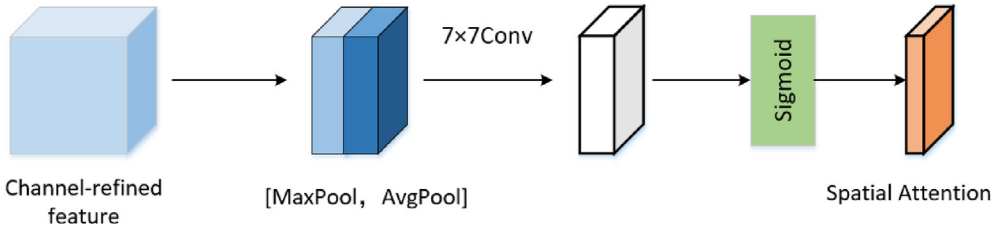

Fig. 6. Diagram of the spatial attention submodule structure.

The RAB module can enhance the model's overall performance to effectively address quality improvement and noise suppression in images. Consequently, this approach can significantly bolster underwater image enhancement. After the RAB module and the application of a convolutional layer, a tensor with 6 channels is generated. The first three channels are multiplied by the result obtained from the module that extracts features in the HSV color space, while the remaining three channels are multiplied by the output of the RGB color space denoising module that has undergone a convolution and a sigmoid activation function. Finally, the two aforementioned results are combined to achieve a high-quality enhanced underwater image.

3.4. Loss function

To maintain the edge information of the image and enhance the texture similarity of the image, in this paper, the mean square error (MSE) loss, content-aware loss and SSIM loss are used to train the model. The MSE loss is the mean value of the squared difference between the model output image and the reference image. The MSE loss is sensitive to outliers and better guides the model to learn sample mapping from the global similarity space. The MSE loss can be expressed as shown in Equation (2).

$$Loss_{mse} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \tag{2}$$

where N is the number of samples in each training batch, x_i denotes the augmented image output by the model, and y_i denotes the corresponding reference image.

Inspired by the literature [27], this paper introduces content-aware loss to enhance image details and correct image colors. Perceived similarity is defined as the Euclidean distance between the feature representations of the augmented image and the clean instances. The content-aware loss can be expressed as shown in Equation (3).

$$Loss_{vgg} = \frac{1}{m}\sum\left(\varphi_{i,j}(\hat{I}) - \varphi_{i,j}(I)\right)^2 \tag{3}$$

where $\varphi_{i,j}$ denotes the feature mapping obtained through the jth convolution (after activation) before the ith maximal pooling into within the pretrained VGG16 network and m is the number of pixels of the extracted feature map.

The SSIM loss is designed to improve the structural similarity between the feature representations of the output image and the reference image. It can be expressed as shown in Equation (4).

$$Loss_{ssim} = 1 - \frac{1}{N}\sum_{i=1}^{N} SSIM \tag{4}$$

The final loss is a linear combination of the mean square error loss, content perception loss and SSIM loss. It can be expressed as shown in Equation (5).

$$Loss_{final} = \lambda_1 Loss_{mse} + \lambda_2 Loss_{\mathrm{vgg}} + \lambda_3 Loss_{ssim} \tag{5}$$

where λ_1, λ_2 and λ_3 are empirically set to 0.6, 0.3 and 0.1, respectively, to balance the proportions of different losses.

4. Experiment

In this section, we initially present the dataset employed in the experiments, the evaluation index of underwater image enhancement, and the implementation details. Then, this paper makes a quantitative comparison with other deep learning methods, including a subjective comparison and objective evaluation index comparison. Finally, we perform ablation experiments to validate the contribution of each proposed module.

4.1. Dataset

We employ the publicly available underwater image datasets UIEBD [10] and EUVP [12] for training and testing our model, respectively. The UIEBD dataset comprises 890 underwater images paired with corresponding labels. The reference images within the dataset are chosen subjectively from 12 UIE results, including underwater images from diverse waters and depths. In total, 800 images were randomly selected and employed as the training set, while the remaining 90 images were designated for the test set. Due to the variability in image sizes, during model training, we resized the images to 128 × 128.

The EUVP underwater ImageNet dataset contains underwater image data for different scenes, water types, and lighting conditions. We randomly divided this dataset into training and testing sets maintaining an 8.5 to 1.5 ratio, where 3145 pairs of images were trained and the remaining 555 pairs of images were tested, including various real underwater images such as low contrast, missing details, and color distortion. Given that the EUVP dataset contains images of various resolutions, including sizes such as 800 × 600, 640 × 480, 256 × 256, and 224 × 224, we included a preprocessing step to standardize these images to a consistent size of 256 × 256 before model training. Ensuring uniform image dimensions during model training allowed for better adaptation to the model's learning and feature extraction procedures.

4.2. Experimental environment

Table 1 displays the precise experimental setup, detailing the hardware and software configurations essential for conducting the research. During the training phase, the network training process involved PyTorch framework implementation utilizing the Adam optimizer with an initial learning rate of 1e-4. A batch size of 1 is set, and the training comprises 100 rounds.

Table 1. Experimental environment.

Environment	properties
System	Ubuntu 18.04
CPU	Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60 GHz
GPU	RTX3090
CUDA	11.0
build environment	PyTorch = = 1.7.0+Python 3.8

4.3. Evaluation metrics

Objective evaluation metrics, specifically full-reference metrics such as the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), are employed. A higher PSNR indicates a closer content resemblance between the output and reference images. Conversely, a higher structural similarity index (SSIM) suggests greater structural similarity between two images. The evaluation metric for enhancing underwater images without reference is known as the underwater image quality metric (UIQM).

The PSNR, a metric utilizing the mean squared error (MSE), gauges signal distortion through a comparison of the original and enhanced images. A higher PSNR denotes diminished image distortion and elevated image quality. Its formulation is illustrated in Equation (6).

$$\mathrm{PSNR}(\mathrm{x},\mathrm{y}) = 10\, log_{10}\left[\frac{255^2}{MSE(x,y)}\right] \tag{6}$$

The SSIM, a metric for structural similarity, quantifies the resemblance between the output image (y) and the reference image (x). It accounts for the structural information within images, evaluating similarities in luminance, contrast, and structure. SSIM values range between 0 and 1, with values closer to 1 denoting heightened image similarity and enhanced image quality. Its computation is shown in Equation (7).

$$\mathrm{SSIM}(\mathrm{x},\mathrm{y}) = \frac{\left(2\mu_x\mu_y + c_1\right)}{\left(\mu_x{}^2 + \mu_y{}^2 + c_1\right)} \frac{\left(2\sigma_{xy} + c_2\right)}{\left(\sigma_x{}^2 + \sigma_y{}^2 + c_2\right)} \tag{7}$$

In the formula, μ_x and μ_y represent the brightness means of x and y, $\sigma_x{}^2$ and $\sigma_y{}^2$ represent the variances of x and y, σ_{xy} is their covariance, and c_1 and c_2 are constants with values $c_1 = (255 \times 0.01)^2$ and $c_2 = (255 \times 0.03)^2$.

μ_x and μ_x signify the mean brightness values of x and y, respectively, while $\sigma_x{}^2$ and $\sigma_y{}^2$ denote their respective variances. The variable σ_{xy} represents the covariance between x and y. Additionally, the constants c_1 and c_2 are defined as $c_1 = (255 \times 0.01)^2$ and $c_2 = (255 \times 0.03)^2$.

UIQM [27] is derived from a weighted sum comprising the underwater image colorfulness measure (UICM), underwater image sharpness measure (UISM), and underwater image contrast measure (UIConM). While a higher UIQM score typically indicates improved human visual perception, it may not consistently offer precise evaluations of image quality, especially when considering no-reference metrics. Therefore, in this paper, we use the UIQM score as a reference. The UIQM is calculated as shown in Equation (8).

$$\mathrm{UIQM} = C_1 \times UICM + C_2 \times UISM + C_3 \times UIConM \tag{8}$$

According to [27], this paper sets the following values: $c_1 = 0.0282$, $c_2 = 0.2953$, and $c_3 = 3.5753$.

4.4. Comparative analysis between different models

To evaluate the effectiveness of the proposed approach, two real underwater datasets, namely, the UIEBD and EUVP datasets, were selected. These datasets were then compared with the most widely used underwater image enhancement techniques, including the fusion-based method (Fusion), the unified dark channel prior method (UDCP), WaterNet, UWCNN, FunIE-GAN, and SCNet. We conducted a qualitative evaluation and quantitative analysis of the methodology employed in this paper.

4.4.1. Objective analysis

This paper endeavors to thoroughly examine the influences of modeling from diverse viewpoints, seeking a comprehensive understanding of the model's behavior across various scenarios and perspectives. To achieve this goal, we conducted objective analysis experiments on the UIEBD and EUVP datasets. The quantitative outcomes of diverse algorithms for improving underwater images on the underwater image enhancement benchmark dataset (UIEBD) are shown in Table 2. The data in Table 2 distinctly demonstrate the superior performance of the proposed model relative to existing approaches, showcasing its dominance across a broad spectrum of image quality metrics. Our method demonstrates exceptional performance on the limited dataset UIEBD.

Compared with SCNet, our method shows a substantial improvement in PSNR of 4.5% and in SSIM of 1.8%. Compared with the traditional methods Fusion and UDCP, our method greatly improves the metrics because Fusion is a relatively good traditional method and can achieve results close to those of deep learning methods. Compared to the UWCNN algorithm, our method exhibits a marginal increase in PSNR while maintaining similar SSIM scores, suggesting similarity in image structure and details between the two methods. Conversely, compared to the FunIE-GAN algorithm, our method demonstrates significant enhancements in both PSNR and SSIM, illustrating its ability to capture image details and realistic color perception. Additionally, when juxtaposed with the WaterNet algorithm, our method displays substantial improvements in both PSNR and SSIM metrics. Compared to the UIEC^2-Net algorithm, our method shows improvements in both the PSNR and SSIM metrics. Notably, our method records the lowest MSE values among the various algorithms, indicating superior quality and closer alignment with reference images in underwater image enhancement. The experimental results presented in Table 2 underscore the effectiveness of our proposed method in enhancing underwater images, showcasing its ability to closely replicate reference images in terms of contrast, texture, structure, and detail, thus ensuring higher-quality outcomes.

We evaluated the efficacy of our proposed method using the EUVP dataset, employing evaluation indices such as PSNR, SSIM, and UIQM, as depicted in Table 3. The data in Table 3 demonstrate the superior performance of our proposed model over existing methodologies across various image quality metrics. Specifically, compared with SCNet, our method exhibits a remarkable improvement in PSNR of 4.2% and in SSIM of 3.5%, demonstrating

Table 2. Quantitative comparison of different algorithms on the test set UIEBD.

Methods	PSNR	SSIM	MSE
Fusion	19.03	0.8021	859.61
UDCP	11.76	0.5101	5157.36
WaterNet	19.11	0.7964	797.60
FunIE-GAN	17.75	0.7893	1041.03
UWCNN	19.22	0.8706	742.58
UIEC^2-Net	20.33	0.8159	696.91
SCNet	20.48	0.8459	596.91
Ours	21.52	0.8612	407.53

Table 3. Quantitative comparison of different algorithms on the test set EUVP.

Methods	PSNR	SSIM	UIQM
Fusion	19.68	0.7014	2.79
UDCP	13.43	0.5300	2.31
WaterNet	21.53	0.7988	2.86
UWCNN	23.54	0.8263	2.65
FUnIE-GAN	20.14	0.6879	2.96
UIEC^2-Net	23.42	0.8315	2.95
SCNet	24.08	0.8232	3.07
Ours	25.10	0.8526	3.17

its superiority in diverse underwater environments. Compared with Fusion and UDCP, the proposed method significantly improved the evaluated metrics, indicating that the traditional methods still have limitations regarding complex underwater image enhancement. Furthermore, in contrast to the UWCNN, our method shows improvements in both the PSNR and SSIM, indicating its ability to generate higher-quality reconstructed images while effectively reducing image noise. Similarly, compared to WaterNet, our method demonstrates substantial advancements in both PSNR and SSIM, highlighting its ability to minimize color distortion and enhance image details. Additionally, our method achieves noteworthy improvements in PSNR and SSIM over FunIE-GAN, suggesting the limitations of the latter in handling real underwater images. Compared with the UIEC^2-Net algorithm, our proposed method shows improvements in all the metrics. These results emphasize that our proposed method effectively enhances the clarity, contrast, color, and overall visual appeal in underwater image enhancement (see Table 3).

4.4.2. Subjective analysis

To assess the effectiveness of the proposed method in enhancing visual quality, subjective analysis experiments were performed using the UIEBD and EUVP datasets. This approach involved a quantitative comparison with prominent deep learning methods that have emerged in recent years. The influence of different algorithms on the UIEBD and EUVP test sets is visually depicted in Figs. 7 and 8, highlighting their respective enhancement effects on underwater images.

Figures 7 and 8 clearly show that the utilization of WaterNet in image processing is a significant issue because it intensifies the impact of color bias. Fusion does not correct for the color bias, while the contrast is low, and the details are not well expressed. In addition, UDCP fails to restore the image effectively and does not address the green color bias, making the overall colors dull. When applying WaterNet for image enhancement, the resulting enhanced image exhibits a darker appearance while still preserving the predominant blue and green backgrounds present in the original image. Nevertheless, this procedure results in the loss of fine image details due to blurring.

In the case of UWCNN, the network attempts to compensate for color shifts in the underwater environment with an overemphasis on the red channel, resulting in a noticeable red color bias effect. Although this processing improves the sharpness contrast of the image, it adversely affects the overall color balance of the image and leads to oversmoothing or blurring of the image details during enhancement, which reduces the image's

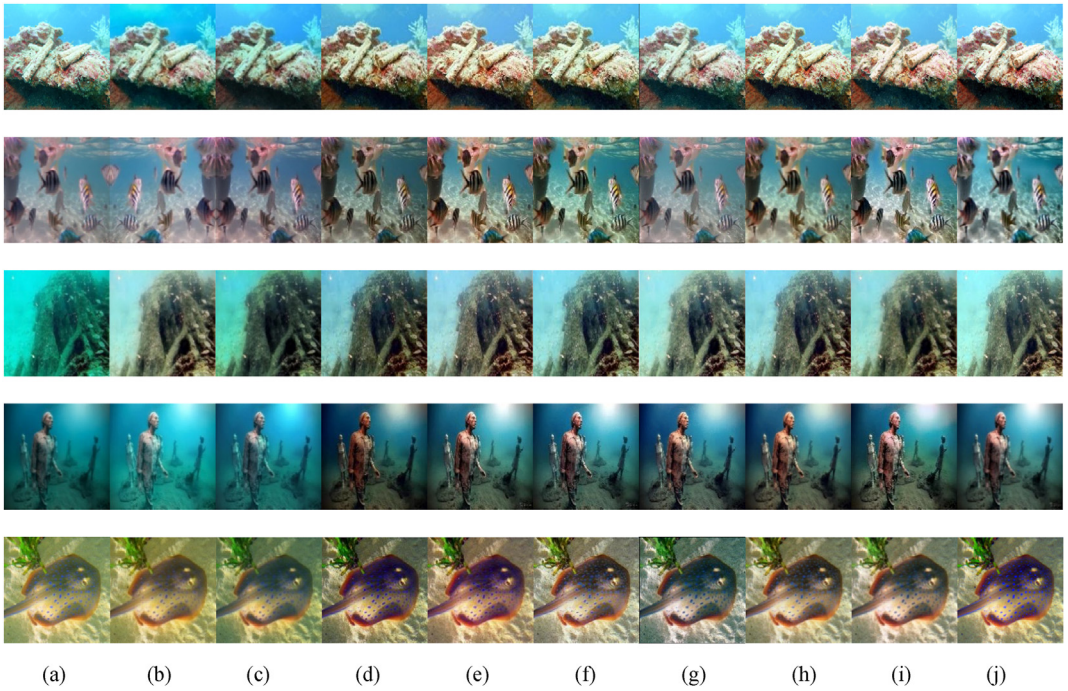

Fig. 7. Processing results of different algorithms on the UIEBD test set (a) Original image; (b) Fusion; (c) UDCP; (d) WaterNet; (e) UWCNN; (f) FunIE-GAN; (g) UIEC^2-Net; (h) SCNet; (i) Proposed method; (j) Reference image.

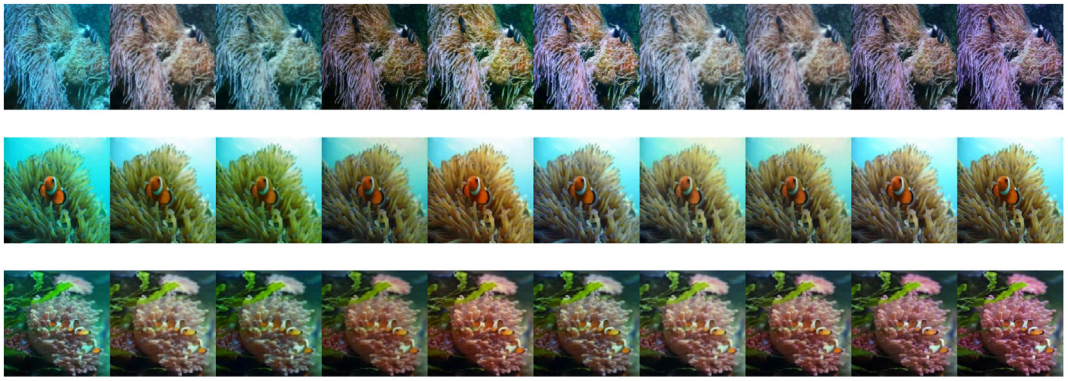

Fig. 8. Processing results of different algorithms on the EUVP test set (a) Original image; (b) Fusion; (c) UDCP; (d) WaterNet; (e) UWCNN; (f) FunIE-GAN; (g) UIEC^2-Net; (h) SCNet; (i) Proposed method; (j) Reference image.

ability to retain detail. FunIE-GAN effectively removes the blue–green color bias. However, this approach can lead to problems; for example, the enhanced image usually appears to be undersaturated, which means that the colors in the image appear to be softer, resulting in the loss of some color features. Although this method can reduce the blue–green bias, it also results in less expressive colors in the image, with some of the otherwise vibrant colors becoming flatter. The consequence is a reduction in the level of detail and vibrancy of colors in the image. The UIEC^2-Net algorithm effectively improves image clarity, but the color correction of the water body is not obvious enough, and a blue color bias is still present. The SCNet method demonstrates a noticeable improvement, but it fails to address blurring in certain parts of the image, resulting in the loss of detail and poor clarity in those areas. Although SCNet effectively restores the dehydration characteristics of the image, its enhancement effect was not as pronounced as that of our model. Additionally, it does not sufficiently address the problem of color casting, and it still retains some of the distortion present in the original image. In contrast, our method exhibits remarkable color correction abilities, effectively eliminating color biases, enhancing texture details, and refining overall image clarity. Furthermore, it accurately preserves the genuine color distribution present in the original underwater image while augmenting contrast and brightness to a certain extent. Consequently, the proposed method generates images enriched with detailed features, contributing to an overall improvement in the quality and lifelike representation of underwater images, offering viewers a more realistic visual experience.

When evaluating the performance of deep learning-based models, in addition to assessing whether they effectively enhance the image quality, the number of parameters and GFLOPS of the model are also pivotal. These metrics not only reflect the complexity and computational power of the model but are also relate directly to the feasibility and efficiency of real-time applications for underwater robots. In particular, in complex and changing underwater environments, the ability to respond in real time and process data efficiently is crucial for robots.

Table 4 shows that compared with the FunIE-GAN algorithm, our proposed model decreases the number of parameters; however, it slightly increases them relative to the other algorithms. In terms of the GFLOPs metrics, the proposed model shows better results than all the other models.

4.5. Ablation study

To demonstrate the effectiveness of the HSV color space feature extraction module, the multiscale dilated convolution module, and the RAB module, ablation studies were carried out using the UIEBD dataset. The resulting average PSNR and SSIM values obtained from these experiments are summarized in Table 5. In the ablation studies, each of the three modules, namely, the HSV color space feature extraction module, multiscale dilated convolution module, and RAB module, were individually

Table 4. Comparison of model performance metrics.

Model	Parameters	GFLOPs
WaterNet	1.07×10^6	140.91
UWCNN	4.00×10^4	1.63
FunIE-GAN	7.02×10^6	0.73
SCNet	8.17×10^5	5.88
UIEC^2-Net	3.14×10^6	7.06
Ours	5.24×10^6	5.27

Table 5. Ablation study results for each module.

HSV color space feature extraction module	Multiscale dilated convolution	RAB module	PSNR	SSIM
			20.48	0.8459
√			20.74	0.8524
	√		20.80	0.8535
		√	20.69	0.8516
√	√		21.13	0.8573
√		√	20.85	0.8526
	√	√	20.98	0.8565
√	√	√	21.52	0.8612

tested for experimentation. The findings clearly indicate that all three modules played a significant role in enhancing the outcomes. In terms of the PSNR and SSIM, the multiscale dilated convolution module achieved notable improvements of 4.5% and 1.8%, respectively. Subsequently, ablation experiments were conducted to test two random combinations of these three modules. The findings demonstrate that the model's performance was positively influenced by each pair of enhancements, as evidenced by improvements in the PSNR and SSIM. Notably, the combination of the HSV color space feature extraction module and the multiscale dilated convolution module resulted in substantial improvements in PSNR and SSIM of 4.2% and 3.5%, respectively. Furthermore, ablation studies were conducted to evaluate the impact of all three modules when used together. The findings of the ablation studies clearly show that this particular combination produced the most favorable outcomes, exhibiting a notable 4.5% enhancement in the peak signal-to-noise ratio (PSNR) and a 1.8% improvement in the structural similarity index (SSIM). The findings also provide evidence that the incorporation of the HSV color space feature extraction module, multiscale dilated convolution, and RAB module, as proposed in this research, enables the model to attain superior outcomes and acquire underwater images of higher quality.

5. Conclusion

In this paper, we introduce a new methodology referred to as the dual-color-space multiscale residual network (DMR-SCNet), which is an advancement built upon the foundation of the SCNet architecture. DMR-SCNet is developed to address the prevalent issues with underwater images, particularly low contrast and blurred details. By effectively enhancing image quality, this approach establishes a favorable foundation for subsequent advanced visual tasks, such as underwater target detection and recognition. First, the issue of image color asymmetry is mitigated through color space conversion, which is then followed by curve adjustment. The proposed module for denoising in the RGB color space incorporates multiscale expansion convolution instead of ordinary convolution. This modification enhances the model's performance by leveraging the multiscale and multiresolution characteristics of the feature map. Additionally, it refines the learning process in terms of channel and space. Finally, we propose the RAB module, which aims to enhance image detail restoration and better enhance important features by incorporating an attentional mechanism and residual connectivity. The experimental results confirm the ability of our method to alleviate color bias, improve image clarity, and augment contrast. Therefore, it represents a highly effective approach for enhancing underwater images. Our study thus holds significant value for future research on underwater vision tasks because it contributes advancing marine resource exploitation and the exploration of the marine environment.

Code, data, and materials availability

The analysis conducted in this study utilized publicly available datasets. The specific dataset information can be accessed through the following link: UIEBD, https://li-chongyi.github.io/proj%5Fbenchmark.html; EUVP, https://irvlab.cs.umn.edu/resources/euvp-dataset.

Conflict of interest

The research for this paper did not address conflicts of interest.

Acknowledgments

This work was financially supported by the Key Science and Technology R&D Project of Henan Province, China, under Grant No. 212102210078; the Major Science and Technology Project of Henan Province, China, under Grant No. 201300210400;

and the Key R&D and Promotion Project of Henan Province (Science and Technology Breakthrough), China, under Grant No. 202102210380.

References

[1] Iqbal K, Odetayo M, James A, Abdul Salam Rosalina, Talib Abdullah Zawawi Hj. Enhancing the low quality images using unsupervised colour correction method. In: 2010 IEEE international conference on systems, man and cybernetics, Istanbul, Turkey; 2010. p. 1703–9. https://doi.org/10.1109/ICSMC.2010.5642311.

[2] Abdullah-Al-Wadud M, Kabir MH, Dewan MAA, Chae O. A dynamic histogram equalization for image contrast enhancement. IEEE Trans Consum Electron 2007;53(2):593–600. https://doi.org/10.1109/TCE.2007.381734.

[3] Liu YC, Chan WH, Chen YQ. Automatic white balance for digital still camera. IEEE Trans Consum Electron 1995;41(3):460–6. https://doi.org/10.1109/30.468045.

[4] van de Weijer J, Gevers T, Gijsenij A. Edge-based color constancy. IEEE Trans Image Process Sept. 2007;16(9):2207–14. https://doi.org/10.1109/TIP.2007.901808.

[5] Land E. Lightness and retinex theory. J Opt Soc Am 1971;61(1):1–11.

[6] Drews P, Nascimento E, Moraes F, Botelho S, Campos M. Transmission estimation in underwater single images. In: Proceedings of the IEEE international conference on computer vision (ICCV) workshops; 2013. p. 825–30.

[7] Galdran A, Pardo D, Picón A, Alvarez-Gila A. Automatic red channel underwater image restoration. J Vis Commun Image Represent 2015;26:132–45. https://doi.org/10.1016/j.jvcir.2014.11.006.

[8] Liu K, Liang Y. Underwater image enhancement method based on adaptive attenuation-curve prior. Opt Express 2021;29(7):10321–45.

[9] Li C-Y, Guo J-C, Cong R-M, Pang Y-W, Wang B. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. IEEE Trans Image Process 2016;25(12):5664–77. https://doi.org/10.1109/TIP.2016.2612882.

[10] Li C, Guo C, Ren W, Cong R, Hou J, Kwong S, et al. An underwater image enhancement benchmark dataset and beyond. IEEE Trans Image Process 2020;29:4376–89. https://doi.org/10.1109/TIP.2019.2955241.

[11] Li C, Anwar S, Porikli F. Underwater scene prior inspired deep underwater image and video enhancement [J]. Pattern Recognit 2020;98:107038.

[12] Islam MJ, Xia YY, Sattar J. Fast underwater image enhancement for improved visual perception. In: IEEE robotics and automation letters. vol. 5; 2020. p. 3227–34. https://doi.org/10.1109/LRA.2020.2974710.

[13] Gong T, Zhang M, Zhou Y, Bai H. Underwater image enhancement based on color feature fusion [J]. Electronics 2023;12(24):4999.

[14] Li G, Wang F, Zhou L, Jin S, Xie X, Ding C, et al. MCANet: multi-channel attention network with multi-color space encoder for underwater image classification [J]. Comput Electr Eng 2023;108:108724.

[15] Wang Y, Guo J, Gao H, Yue H. UIEC^2-Net: CNN-based underwater image enhancement using two color space[J]. Signal Processing. Image Commun 2021;96:116250.

[16] Fu Z, Lin X, Wang W, Huang Y, Ding X. Underwater image enhancement via learning water type desensitized representations. In: ICASSP 2022 - 2022 IEEE international conference on acoustics, speech and signal processing (ICASSP), Singapore, Singapore; 2022. p. 2764–8. https://doi.org/10.1109/ICASSP43922.2022.9747758.

[17] Pan Xingang, Zhan Xiaohang, Shi Jianping, Tang Xiaoou, Luo Ping. Switchable whitening for deep representation learning. In: 2019 IEEE/CVF international conference on computer vision (ICCV); 2019. p. 1863–71.

[18] Li Boyi, Wu Felix, Weinberger Kilian Q, Belongie Serge. Positional normalization. 2019. p. 1622–34.

[19] Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans Image Process July 2017;26(7):3142–55. https://doi.org/10.1109/TIP.2017.2662206.

[20] Lehtinen J, Munkberg J, Hasselgren J, Laine S, Karras T, Aittala M. Noise2Noise: learning image restoration without clean data. ACM Trans Graph 2018;37(4) [arXiv:1803.04189].

[21] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. In: International conference on learning representations; 2016. p. 1–13.

[22] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: semantic image segmentation with deep convolutional Nets, Atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 2017;40(4):834–48.

[23] Liu P, Zhang H, Lian W, Zuo W. Multi-level wavelet convolutional neural networks. IEEE Access 2019;7:74973–85. https://doi.org/10.1109/ACCESS.2019.2921451.

[24] Wang Z, Liu Z, Yang J. DnCNN-based multi-scale deep feature fusion for image denoising. IEEE Trans Image Process 2020;29:4269–80.

[25] Woo S, Park J, Lee JY, Kweon IS. CBAM: convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV); 2018.

[26] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR); 2015.

[27] Panetta K, Gao C, Agaian S. Human-visual-system-Inspired underwater image quality measures. IEEE J Ocean Eng July 2016;41(3):541–51. https://doi.org/10.1109/JOE.2015.2469915.