frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 11 March 2025
DOI 10.3389/fmars.2025.1555128

OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Zheng Jianhua,
Zhongkai University of Agriculture and Engineering, China
Zifan Lin,
University of Western Australia, Australia

*CORRESPONDENCE
Weidong Zhang
zwd_wd@163.com

RECEIVED 03 January 2025
ACCEPTED 17 February 2025
PUBLISHED 11 March 2025

CITATION
Zhao G, Wu Y, Zhou L, Zhao W and Zhang W (2025) Multi-scale cascaded attention network for underwater image enhancement. *Front. Mar. Sci.* 12:1555128. doi: 10.3389/fmars.2025.1555128

Multi-scale cascaded attention network for underwater image enhancement

Gaoli Zhao[1], Yuheng Wu[1], Ling Zhou[2], Wenyi Zhao[3] and Weidong Zhang[2,4]*

[1]School of Computer and Technology, Henan Institute of Science and Technology, Xinxiang, China, [2]School of Information Engineering, Henan Institute of Science and Technology, Xinxiang, China, [3]School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China, [4]School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, China

The complexity of underwater environments combined with light attenuation and scattering in water often leads to quality degradation in underwater images, including color distortion and blurred details. To eliminate obstacles in underwater imaging, we propose an underwater image enhancement method based on a cascaded attention network called MSCA-Net. Specifically, this method designs an attention-guided module that connects channel and pixel attention in both serial and parallel ways to simultaneously achieve channel feature refinement and feature representation enhancement. Afterward, we propose a multi-scale feature integration module to capture information and details at different scales within the image. Meanwhile, residual connections are introduced to assist in deep feature learning via acquiring more detailed information from shallow features. We conducted extensive experiments on various underwater datasets, and the results demonstrate that our method still holds an advantage when compared to the latest underwater image enhancement methods.

KEYWORDS

underwater image enhancement, cascaded attention network, multi-scale feature integration, computer vision, deep learning

1 Introduction

With the development of ocean exploration missions, remote sensing technology has been widely applied in underwater scene analysis (Lin et al., 2021; Li et al., 2018), marine resource exploration (Shen et al., 2021), and marine archaeology (Zhang et al., 2022a). Unfortunately, due to the wavelength-dependent absorption and scattering of light as it travels through water, underwater images often suffer from low contrast, blurriness, and color distortion. Remote sensing-based underwater imaging equipment struggles to capture clear and accurate underwater images, severely affecting underwater visual tasks. To address these challenges, many scholars have conducted extensive and in-depth research on underwater image enhancement technology and have achieved significant results.

Over the past few decades, deep learning has emerged as a driving force in the development of artificial intelligence technologies. Deep learning has seen significant achievements across various areas of computer vision fields because of its powerful nonlinear modeling capabilities, such as object recognition (Zhang et al., 2024a), image fusion (Zhang et al., 2020) and image restoration (Zhang et al., 2024b). Compared with traditional physics-based methods, Convolutional Neural Networks (CNNs) that use deep learning technology have significantly improved image processing effects and time performance. Convolutional neural networks can model complex nonlinear systems end-to-end by learning from extensive paired data, thereby improving perceived image quality. In underwater visual tasks, several neural network models have achieved remarkable success. These models mainly involve generative adversarial networks, multi-scale dense networks (Liu et al., 2024a), self-attention networks, and lightweight image enhancement networks. Among them, the Comparative Learning Network (CLUIE-Net) (Li et al., 2023b) and the Semantic Guidance-based Network using Multi-scale Perception (SGUIE-Net) (Qi et al., 2022) represent the cutting edge methods for underwater image enhancement. Nonetheless, most existing methods based on CNN models are designed with a single attention mechanism (Woo et al., 2018; Fan et al., 2022), paying little attention to extracting global features at different scales. The size of the effective receptive field limits their ability to simultaneously capture global and local features (Liu et al., 2024b), resulting in deficiencies in hierarchical feature extraction and fusion in most current methods. Therefore, more and more image enhancement methods adopt multi-scale feature extraction branches to capture image details and global information at different scales, improving the contrast while enhancing image details, textures, and overall structure. MFMN (Zheng et al., 2024b) achieves diversified feature extraction using only a small number of 1×1 and 3×3 convolution combinations, avoiding complex operations such as large convolution kernels, frequent skip connections, and channel reordering, significantly reducing the parameter count and computational complexity. MCRNet (Zhang et al., 2024c) effectively integrates spatial global information across four different scales through convolution operations, enhancing the network's information representation ability while avoiding the problem of simple feature stacking. Compared to the aforementioned methods, this paper uses a larger number of multi-scale convolutions and adaptive weighted fusion, which not only extracts multi-level features more comprehensively but also selectively enhances the representation capability of different scale features through weight selection, improving the model's overall perception and representation performance. Additionally, considering issues such as uneven lighting distribution and imbalanced color information in underwater images, we propose a Dual-Path Attention Enhancement Module (DAEM), which cascades Channel Attention (CA) and Pixel Attention (PA) in both serial and parallel ways, alternating the transfer of deep semantic information and shallow feature information, thereby better eliminating color distortion and low illumination problems in underwater images. Moreover, we adopt residual connections in the dense convolution module used for image denoising to extract representative noise information at both local and global scales, introducing residual connections to pass the original noise information to subsequent layers, thus improving the network's denoising performance (Figure 1).

We emphasize the contributions of our work as follows:

- We propose a dual-path attention enhancement module, which gradually strengthens feature selectivity at different spatial frequencies through a serial structure while focusing on channel-spatial domain features via a parallel structure. This module enables adaptive weighting and regulation of multi-dimensional information, effectively enhancing multi-scale feature representation capabilities.
- We present a dense convolutional denoising module, which captures the noise information in the original image through a deep convolutional network and performs a difference operation between the original and noisy images to achieve denoising. Meanwhile, we optimize the residual connections to avoid the gradient propagation

FIGURE 1
Comparison of sample enhancement results with the raw images, including blue-green cast images, yellow cast images, and hazy images.

problem. This module effectively separates the noise in the image.

- We propose a novel hybrid loss function, which consists of Laplacian loss to enhance image edge details, perceptual reconstruction loss to capture higher-level semantic information, and SSIM loss to strengthen image structure and texture. This hybrid loss function optimizes image details more effectively, resulting in clearer and more natural enhanced images.

2 Related works

Presently, more and more researchers are conducting studies on underwater image enhancement. These works can be generally categorized into traditional methods and deep learning-based methods. This section lists diverse methods based on different principles and provides a brief overview.

2.1 Traditional methods

Generally, model-based methods as a common category of traditional underwater image enhancement methods, typically use physical models to reverse or mitigate the distortion caused by the underwater environment. In contrast, non-model-based methods which are also part of traditional methods, rely on heuristic techniques and focus on improving the visual quality of images through algorithms.

2.1.1 Model-based methods

Estimate transmission parameters through the underwater image formation model and reverse the process to recover clear images. These methods can construct models to simulate the image generation process in underwater environments and reverse distorted images into clear ones, primarily applied in scenarios that require accurate recovery of the physical properties and structures of the images. Hou et al. (Hou et al., 2024) proposed a Laplacian variational model that achieved significant results in image dehazing. Peng et al. (Peng and Cosman, 2017) focused on underwater image restoration and enhancement by estimating scene depth through analysis of image sharpness and light absorption. Liang et al. (Liang et al., 2022) proposed a method for estimating backscatter light using hierarchical search technology, integrated with the dark channel prior, to efficiently achieve underwater image dehazing. However, traditional physics-based underwater image enhancement methods exhibit poor robustness due to their dependence on precise imaging models and additional prior information. As a result, the outcomes are not always satisfactory, posing significant challenges for enhancement.

2.1.2 Non-Model-based methods

Generally focus on enhancing images by adjusting pixel intensity levels, making them more suitable for scenarios that require quick improvement of image visual quality. Some researchers employed histogram adjustment to enhance underwater images by stretching pixel intensities (Li et al., 2016), though they often perform poorly in correcting color bias. Drawing on the principles of minimum color loss and maximum attenuation map-guided fusion, Zhang et al. (Zhang et al., 2022b) applied distinct correction strategies to color channels based on their varying levels of attenuation. By also adjusting the contrast in local regions, they effectively improved the color accuracy of the corrected distorted images. Zhuang et al. (Zhuang and Ding, 2020) proposed an edge-preserving Retinex filtering algorithm that incorporates guided enhancement and combines light correction with guided image filtering to improve contrast and sharpness in underwater images. In recent past, Zhou et al. (Zhou et al., 2022) categorized color biasbased on the average intensity values of color channels, and simultaneously enhanced key image information using optical attenuation characteristics and multi-scene, block-based histogram stretching methods while calculating color information loss. Bi et al. (Bi et al., 2024) effectively enhanced the details of multi-degraded underwater images by applying a dehazing method based on multi-exposure image fusion, following color compensation and white balance for color correction. Overall, despite effectively enhancing images from the perspective of human perception, these algorithms do not completely resolve the complex distortions present in underwater images.

2.2 Deep learning-based methods

With the rapid advancements in deep learning technologies and their widespread application across various computer vision tasks, an increasing number of scholars are applying them to image processing in underwater environments due to their superior performance. Deep learning-based methods can generally be divided into data-driven methods, contrastive learning methods, and attention mechanism-based methods.

2.2.1 Data-driven methods

Utilize the powerful computational capabilities of models by training on large volumes of high-quality data. Through multi-layer nonlinear transformations, these models can automatically extract useful features to accomplish complex tasks, maintaining high versatility in underwater environments with diverse scenarios and varying image quality. Li et al. (Li et al., 2020a) designed a lightweight network model (UWCNN) for underwater scene enhancement, synthesizing ten underwater image datasets based on different water types, achieving satisfactory results in underwater video and image enhancement tasks. Wang et al. (Wang et al., 2024) combined CNN with transformers to extract depth information from images, and this method demonstrated excellent performance removing underwater target edge artifacts. Li et al. (Li et al., 2020b) combined a total of 950 images with various degradation to create a realistic underwater dataset, including 890 images paired with

reference images, and proposed the (WaterNet) enhancement network, delivering visually pleasing results. Guan et al. (Guan et al., 2024) proposed an underwater image enhancement method called DiffWater, based on a conditional denoising diffusion probabilistic model (DDPM) trained on a large number of underwater images. By employing a color compensation method tailored to different water conditions and lighting scenarios, it achieves high-quality enhancement of degraded underwater images. However, data-driven methods rely heavily on large datasets to learn the mapping relationships in the image feature space, which can easily overlook the diverse feature distributions in different domains.

2.2.2 Contrastive learning methods

Neural networks in distinguishing and classifying data by enabling input samples to learn similarity with positive samples while differentiating from negative samples, making them suitable for scenarios where underwater image data is scarce or difficult to obtain labeled data. Li et al. (Li et al., 2023b) introduced a comparative learning framework that learns from multiple enhanced reference candidates by designing a regional quality advantage discrim to generate paired data and then trained the UGAN network with it. Fabbri et al. (Fabbri et al., 2018) proposed using CycleGAN to generate transformation pairs between source and target images, thereby enriching the paired data used to train the UGAN network, and continuously optimizing the model through adversarial training between real and generated samples. Liang et al. (Liang et al., 2024) developed an image quality enhancement model using unsupervised learning, andeffectivelyaddressedimage distortionissuesthroughadataaugmentationmethod utilizing non-real-world transformations. Recently, Yu et al. (Yu et al., 2024) proposing a semantic-aware contrastive module based on disentangled representations that mitigates the impact of critical information loss required for machine vision tasks through contrastive learning strategies. Jia et al. (Jia et al., 2024) designed an unsupervised generative adversarial network based on multi-scale feature extraction, effectively enhancing the details and color information of underwater images by incorporating perceptual loss and edge detection modules. However, in the field of underwater image enhancement, challenges such as difficult positive-negative sample selection, lack of high-quality annotated data, and complex feature distributions severely limit the effectiveness of contrastive learning applications.

2.2.3 Attention mechanism-based methods

Dynamically allocate the model's focus on input data to promote information interaction and handle complex sequential data more effectively, primarily applied in the fine-grained processing of underwater image features and key information extraction. Hu et al. (Hu et al., 2018) started by capturing channel feature information and extensively modeled the interdependencies between channels. During training they dynamically adjusted the output of each channel, which substantially enhanced the model's ability to represent features. Qi et al. (Qi et al., 2022) introduced semantic information as shared guidance among different images within semantic regions, utilizing an attention-aware enhancement module to maximally preserve spatial details in images. Misra et al. (Misra et al., 2021) employed a triple-branch structure to compute attention weights and capture cross-dimensional interactions, ensuring efficient computation while capturing cross-dimensional features in tensors. To address scale degradation and non-uniform color bias in underwater images, Tolie et al. (Tolie et al., 2024) proposed a lightweight network based on multi-scale channel attention, which enhances color richness and refines color distributions satisfactorily. In the past few years, self-attention mechanisms have excelled in capturing global dependencies and flexibly adjusting attention weights, particularly effective in handling long-distance dependencies and multi-task learning. Liu et al. (Liu et al., 2022) highlighted important image features by leveraging parallel attention modules and adaptive learning modules, thereby enhancing the network's feature representation capabilities. In summary, attention-based methods significantly improve enhancement effects but often require higher computational resources and memory. To overcome these challenges, we propose an innovative attention module for enhancing underwater image quality, termed the Dual-Path Attention Enhancement Module (DAEM). Compared to previous methods, DAEM reduces computational complexity while effectively capturing high-priority local features in images.

3 Methodology

In this section, we introduce the proposed MSCA-Net in detail. We will first give an overview of the MSCA-Net model framework and then describe each component of the network architecture in detail. Finally, we will elaborate on the three loss functions used during the training phase.

3.1 Network architecture

We fully leverage the advantages of multi-scale feature extraction and cascaded attention mechanisms in the design of MSCA-Net. By introducing a multi-branch architecture, we enable parallel processing of features at different scales, improving the model's ability to capture both local details and global semantic information, thus enhancing its robustness in complex underwater scenarios. The cascaded attention mechanism further optimizes feature selection and aggregation by dynamically adjusting the network's focus across multiple levels, effectively increasing the efficiency and accuracy of key feature extraction and improving the

model's performance in handling complex image enhancement tasks. The general process of MSCA-Net is summarized in Figure 2. The network structure primarily utilizes three modules: (1) the Multi-Scale Feature Integration Module (MFIM), (2) the Dense Convolutional Denoising Module (DCDM), and (3) the Dual-PathAttention Enhancement Module (DAEM). Firstly, the Multi-Scale Feature Integration Module captures multi-level features from the raw input image. It then generates weights and performs weighted fusion on the captured feature information to enable the neural network to comprehensively consider local features together with global features when capturing an image. Next, we employ dense connections and introduce an attention mechanism in the dense convolutional layers to achieve image denoising. Finally, the dual-path attention enhancement module efficiently aggregates attention sample information, helping the neural network focus on key features from different dimensions.

3.2 Multi-scale feature integration module

The input parameters of the convolutional layers vary, resulting in different feature extraction capabilities. If features are extracted using only a single-scale convolutional layer, key information may be missed, leading to reduced model robustness and accuracy. Therefore, we construct feature extraction branches of different sizes to capture features at various scales and levels of abstraction, as

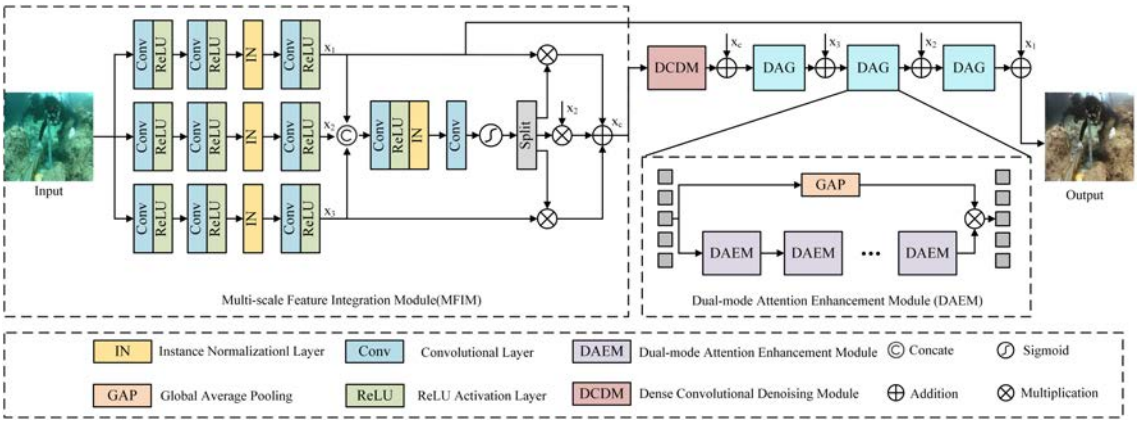

FIGURE 2
Flowchart of MSCA-Net. First, three different convolution kernels are used to perform multi-scale feature extraction on the given original underwater image. Subsequently, the three feature maps are fused and weighted with the original image to obtain a fused image. Next, the fused image is processed through a dense denoising module to remove noise, with residual connections introduced to preserve texture details within the image. Finally, the image is passed through a dual-path attention enhancement module to highlight important details, resulting in the final enhanced image.

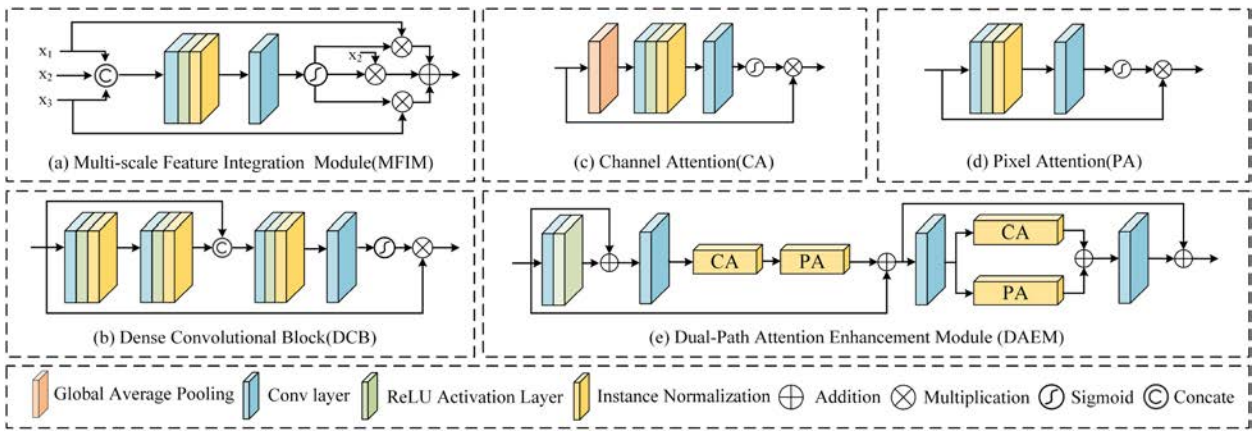

FIGURE 3
The structure of some modules in MSCA-Net. **(a)** The Multi-Scale Feature Integration Module takes multi-scale features extracted by different convolution kernels as input, which are then concatenated, fused, and weighted before output. **(b)** The Dense Convolution Block, where the Dense Convolution Denoising Module is composed of six cascaded Dense Convolution Blocks. **(c)** The structure of the Channel Attention Module. **(d)** The structure of the Pixel Attention Module. **(e)** The Dual-Path Attention Enhancement Module composed of a series and a parallel form, with multiple Dual-Path Attention Enhancement Modules cascaded to form the Dual-Path Attention Enhancement Group.

shown in Figure 3a. Since using a single receptive field alone is not effective or comprehensive enough for extracting feature information from underwater images, we employ 3 × 3, 5 × 5, and 7 × 7 convolutional kernels to capture local detail features and global semantic features respectively inspired by (Liu et al., 2022). Subsequently, we concatenate the features from different branches into a single feature map. Following dimensionality reduction and weight calculation, we multiply the feature map of each branch by its corresponding weight and then sum the weighted feature maps to obtain the final fused feature map X_m, as illustrated below

$$X_m = x_1\delta_1(x) + x_2\delta_2(x) + x_3\delta_3(x), \tag{1}$$

where x_1, x_2, x_3 respectively represent the results of feature extraction branches using 3 × 3, 5 × 5, and 7 × 7 convolutional kernel sizes applied to the input features. $\delta_i(x)$, $i \in (1,2,3)$ represents the weight calculation for each scale feature. In the end, the input features are selectively weighted according to these different weights to generate a new feature representation.

3.3 Dense convolutional denoising module

Typically, the process of image denoising can be decomposed into extracting the unrecovered noise map and eliminating erroneous high-frequency information. Therefore, we propose a dense convolutional module with residual connections for image denoising after multi-level feature fusion. In this module, we designed two different noise elimination layers (multi-convolutional layers and dense attention layers) and introduced dense residual connections. By increasing the number of convolutional layers and deepening the network structure, the module significantly enhances the network's information capture capability.

The specific implementation is shown in Figure 3b. The input image of the densely connected attention module is represented as X_{in}. The intermediate feature map X_{mid} is obtained through the first two convolutional layers of the dense connection attention block, and the process is

$$X_{\text{mid}} = \text{Conv}_{1,2}(X_{\text{in}}), \tag{2}$$

where $\text{Conv}_{1,2}(\cdot)$ represents the first two convolutional layers. Subsequently, by passing through the last two convolutional layers and applying the sigmoid activation function to weight the results and sum them up, we obtain an output feature map X_{out}. The calculation formula is represented as:

$$X_{\text{out}} = \text{Conv}_{3,4}(C(X_{\text{in}}, X_{\text{mid}})) \cdot X_{\text{in}}, \tag{3}$$

where $\text{Conv}_{3,4}(\cdot)$ represents the final two convolutional layers, and $C(\cdot)$ represents the feature map concatenation operation. We incorporated residual connections to ensure that the training of the neural network is not affected by gradient vanishing or gradient exploding issues. Through the dense attention residual block, we can extract erroneous high-frequency information from the X_{in}. Finally, by subtracting the erroneous high-frequency information from the unrecovered noise image, we obtain the denoised image.

3.4 Dual-path attention enhancement module

Although existing feature refinement methods based on channel and spatial attention modules are widely used to address issues such as resolution reduction and detail loss caused by hierarchical down sampling, it remains challenging for the feature maps generated by each module to coordinate effectively. Integrating information from different spatial regions often results in overlaps or conflicts. As a result, relying solely on serial or parallel generation modules may limit the model's performance in capturing complex scenes. To fully leverage the advantages of various attention mechanisms, we designed a Dual-Path Attention Group (DAG) as the foundational structure of the attention module. The DAG consists of ten cascaded Dual-Path Attention Enhancement Modules (DAEMs), with each DAEM's structure illustrated in Figure 3e. First, we connect the channel attention and pixel attention modules serially, allowing the feature map to be adjusted based on the output of the previous step, thereby flexibly controlling the feature extraction and weighting process. Afterward, we cascade the channel attention and pixel attention modules in parallel to enhance the model's computational efficiency. To prevent feature blurring and information loss due to excessive network depth, we incorporate skip connections. The logical process constructed by the dual-cascaded multi-attention mechanism can be represented as:

$$F_{\text{mid}} = F_{\text{in}} + A_p(A_c(\text{Conv}_2(F_{\text{in}} + \text{Conv}_1(F_{\text{in}})))), \tag{4}$$

$$F_{\text{out}} = F_{\text{mid}} + \text{Conv}_4(A_c(\text{Conv}_3(F_{\text{mid}})) + A_p(\text{Conv}_3(F_{\text{mid}}))), \tag{5}$$

where F_{in} represents input features, F_{mid} represents the input features processed through the serially cascaded multi-attention mechanism, and F_{mid} represents the output features F_{out} obtained after further processing by the parallel cascaded multi-attention mechanism. $A_p(\cdot)$ denotes the pixel attention module, $A_c(\cdot)$ denotes the channel attention module, and Conv$i$, $i \in (1,2,3,4)$ represents the four convolutional layers in the module from front to back.

The channel attention and pixel attention mechanisms used in the DAEM are shown in Figures 3c, d. Channel attention can balance the color features of the image and improve detail and texture information. The process is illustrated in Figure 3c. First, a global average pooling layer is applied to obtain the global feature information of each channel. Then, two convolutional layers and a sigmoid activation function are used to obtain the attention weights for each channel. Finally, the weights multiply the input features. In addition, we use pixel attention to focus on each pixel's brightness details or color components. Two convolutional layers reduce the feature map to a single channel. This allows us to calculate the weight for each pixel in the feature map, highlighting important pixels while suppressing unimportant ones, and then return the weighted feature map as the output.

Inspired by (Woo et al., 2018), the serial part of the Dual-Path Attention Enhancement Module places a Channel Attention

module (CA) and a Pixel Attention module (PA) sequentially. Ablation experiments demonstrate that embedding PA and CA sequentially within the serial-parallel framework outperforms using both serial and parallel modes simultaneously.

3.5 Loss function

3.5.1 Laplace loss

To help the model better capture details and textures in the image, thereby achieving more refined image enhancement, we introduced the Laplace loss function. Assuming the generated image is $\hat{\mathbf{I}}$ and the ground truth image is $\mathbf{I}$, the Laplace function value L_{Laplace} can be expressed as:

$$L_{\text{Laplace}} = \frac{1}{N}\sum_{i=1}^{N} || L_i(\hat{\mathbf{I}}) - L_i(I)||_1, \tag{6}$$

where N represents the total pixel count of the input image, $L_i(\hat{\mathbf{I}})$ represents the ith pixel of the generated image, $L_i(\mathbf{I})$ represents the $L_i(\mathbf{I})$ pixel of the reference images, and $\|\cdot\|_1$ denotes the L_i norm of between the two images.

3.5.2 Perceptual reconstruction loss

To maintain consistency between pixels, we use the L_1 loss as the content loss and the L_1 loss to measure the difference between the generated and the ground truth images. The formula can be expressed as:

$$L_1 = \frac{1}{N}\sum_{i=1}^{N}\left|\mathbf{I} - \hat{\mathbf{I}}\right|. \tag{7}$$

The perceptual reconstruction of underwater images is achieved using the VGG network (Simonyan and Zisserman, 2014). Its expression is defined as:

$$L_{\text{VGG}} = \frac{1}{C \times W \times H}\sum_{c=1}^{C}\sum_{w=1}^{W}\sum_{h=1}^{H}(\text{VGG(I)} - \text{VGG}(\hat{\text{I}}))^2, \tag{8}$$

where C, W, and H represent the channels, width, and height of the image, VGG($\mathbf{I}$) and VGG($\hat{\mathbf{I}}$) represent the nonlinear transformations of the generated and ground truth images performed by the VGG network, respectively. We linearly combine the L_1 loss with the weighted perceptual loss function to balance the retention of low-level details and the optimization of high-level features, enhancing both the reconstruction accuracy and perceptual quality of the image. The perceptual reconstruction loss L can be expressed as:

$$L = L_1 + \lambda L_{\text{VGG}}, \tag{9}$$

where λ is the adjustment weight. An excessively large λ would neglect detail accuracy, leading to the loss of edge and texture information, while an excessively small λ would cause color distortion and low contrast. To achieve a balance between detail preservation and high-level feature optimization, λ was set to 0.5 after multiple experimental adjustments.

3.5.3 Hybrid loss

To minimize the loss of detailed textures in the raw image during the enhancement process, we utilized the Structural Similarity Index Measure loss L_{SSIM}. The calculation formula is

$$L_{\text{SSIM}} = 1 - \frac{1}{N}\sum_{i=1}^{N} \text{SSIM}(\mathbf{I}, \hat{\mathbf{I}}). \tag{10}$$

Finally, we combine the SSIM loss with the L_1 loss. The final hybrid loss used during the training phase is expressed as (Equations 1-11):

$$L_{\text{Mix}} = \alpha \cdot L_{\text{SSIM}}(\mathbf{I}, \hat{\mathbf{I}}) + (1 - \alpha) \cdot L_1(\mathbf{I}, \hat{\mathbf{I}}), \tag{11}$$

among them, an excessively large weight α improves SSIM but significantly decreases PSNR and MSE. After multiple experimental adjustments, we determined the optimal weight value to be 0.86.

4 Experiments

In this portion of the text, we first introduce the experimental settings and the details of the implementation. Then, we compare MSCA-Net with eleven methods on the same datasets. These methods include CLUIE- Net (Li et al., 2023b), WaterNet (Li et al., 2020b), SGUIE-Net (Qi et al., 2022), DC-Net (Zheng et al., 2024b), FUnIE-GAN (Islam et al., 2020b), HFM (An and Xu, 2024), OGO-ULAP (Li et al., 2023a), TEBCF (Yuan et al., 2021), ICSP (Hou et al., 2023), WWPE (Zhang et al., 2023b) and PCDE (Zhang et al., 2023a). Finally, we have carefully analyzed the modules in our neural network by designing extensive ablation experiments.

4.1 Implementation details

For a fair comparison of all methods, we used two publicly available underwater image datasets, UIEB (Li et al., 2020b) and EUVP (Islam et al., 2020a), and maintained uniform experimental conditions throughout. The UIEB (Li et al., 2020b) dataset includes 890 underwater images with various types of distortions, while the EUVP (Islam et al., 2020a) dataset includes over 6000 underwater images from various categories such as ImageNet and Scenes. We selected 800 images from the UIEB (Li et al., 2020b) dataset for training and an additional 90 images for testing. From the EUVP (Islam et al., 2020a) dataset, we selected 1600 images from the ImageNet and Scenes categories for training and 400 images for testing. We made the input image size 256 × 256 and performed the training tasks using the PyTorch framework on an Intel(R) I5-12600KF CPU with 32 GB RAM and a NVIDIA RTX 3090 GPU. For the training of the MSCA-Net, the learning rate was set to 1×10^{-3}, we set the batch size to 2 and trained for 40 epochs.

4.2 Qualitative analysis

To validate the effectiveness of our experiment in complicated underwater environments, we selected various degraded

underwater images from the UIEB (Li et al., 2020b) and EUVP (Islam et al., 2020a) datasets for comparison, as shown in Figures 4–8.

The WaterNet (Li et al., 2020b) and DC-Net (Zheng et al., 2024b) methods can significantly reduce color cast in underwater images, but they leave a layer of shadow on the image surface. This makes them ineffective in improving low-light conditions and failing to meet human subjective visual needs. Due to the varying light conditions and water properties, collected underwater images often have severe color distortions that HFM (An and Xu, 2024) and SGUIE-Net (Qi et al., 2022) methods cannot fully eliminate. HFM (An and Xu, 2024) method uses a white balance correction

FIGURE 4
(a-n) Comparison of underwater image experiments on the UIEB (Li et al., 2020b) dataset. From left to right: CLUIE-Net (Li et al., 2023b), WaterNet (Li et al., 2020b), SGUIE-Net (Qi et al., 2022), DC-Net (Zheng et al., 2024b), FUnIE-GAN (Islam et al., 2020b), HFM (An and Xu, 2024), OGO-ULAP (Li et al., 2023a), TEBCF (Yuan et al., 2021), ICSP (Hou et al., 2023), WWPE (Zhang et al., 2023b) and PCDE (Zhang et al., 2023a).

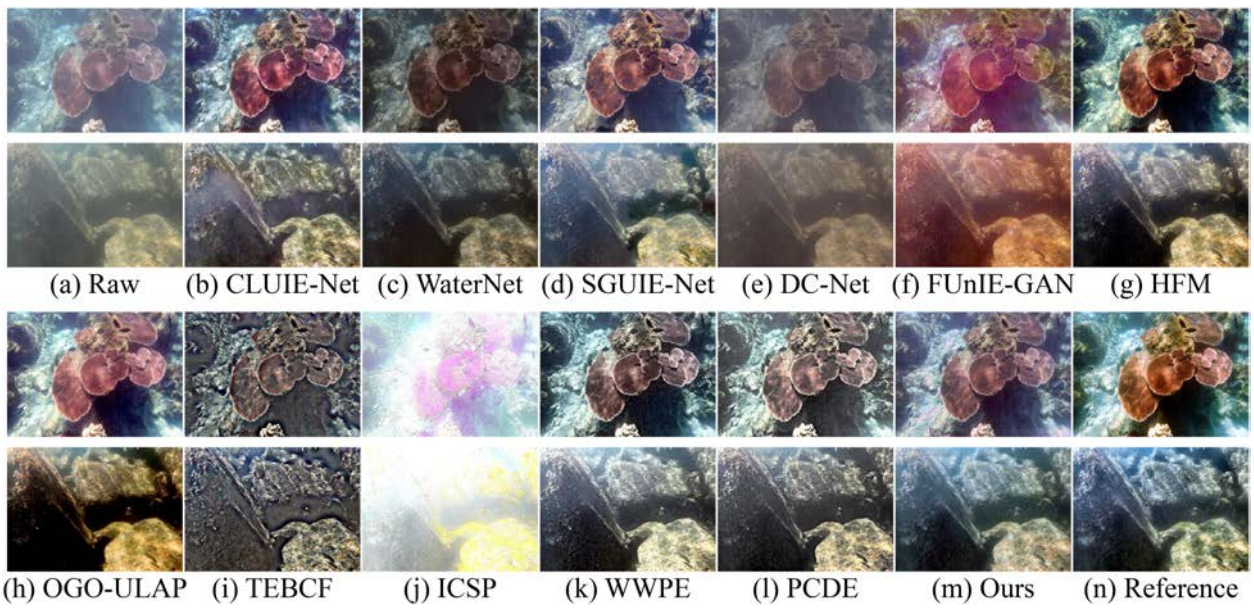

FIGURE 5
(a-n) Comparison of underwater image experiments on the UIEB (Li et al., 2020b) dataset. From left to right: CLUIE-Net (Li et al., 2023b), WaterNet (Li et al., 2020b), SGUIE-Net (Qi et al., 2022), DC-Net (Zheng et al., 2024b), FUnIE-GAN (Islam et al., 2020b), HFM (An and Xu, 2024), OGO-ULAP (Li et al., 2023a), TEBCF (Yuan et al., 2021), ICSP (Hou et al., 2023), WWPE (Zhang et al., 2023b) and PCDE (Zhang et al., 2023a).

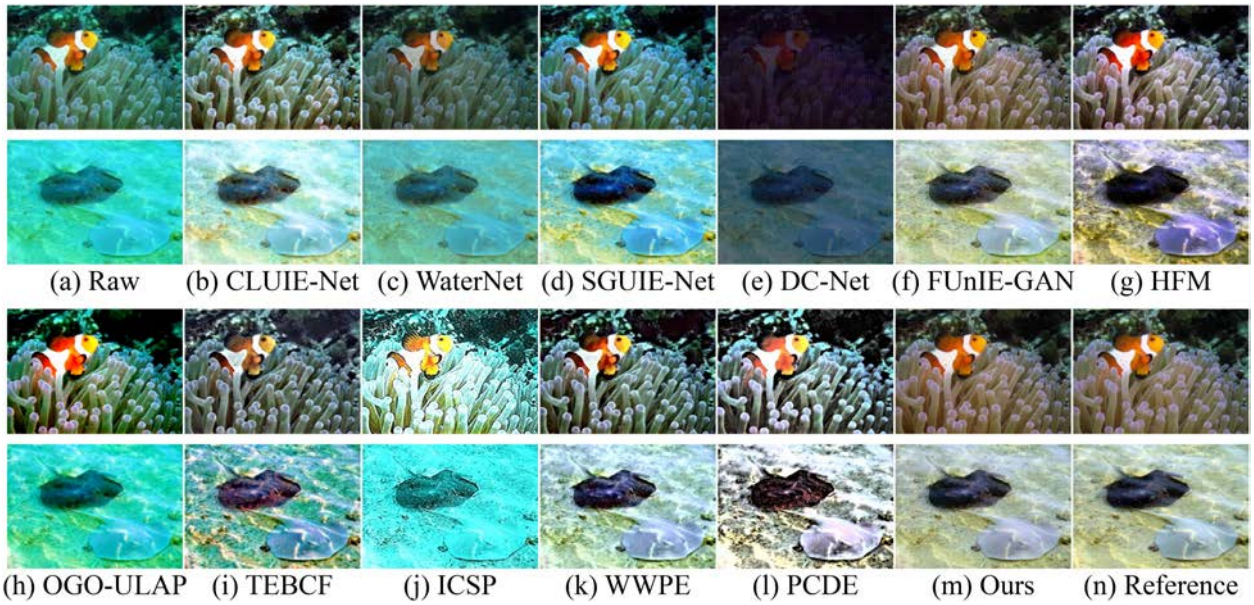

FIGURE 6
(a-n) Comparison of underwater image experiments on the EUVP (Islam et al., 2020a) dataset. From left to right: CLUIE-Net (Li et al., 2023b), WaterNet (Li et al., 2020b), SGUIE-Net (Qi et al., 2022), DC-Net (Zheng et al., 2024b), FUnIE-GAN (Islam et al., 2020b), HFM (An and Xu, 2024), OGO-ULAP (Li et al., 2023a), TEBCF (Yuan et al., 2021), ICSP (Hou et al., 2023), WWPE (Zhang et al., 2023b) and PCDE (Zhang et al., 2023a).

module to fix color bias in underwater images, but it results in an overall yellowish tone in the processed images. Meanwhile, the quality of the results is affected because the CLUIE-Net (Li et al., 2023b) method introduces a slight color cast in local areas while enhancing the images. The FUnIE-GAN (Islam et al., 2020b) method can effectively remove blue and green distortions, but the image quality, particularly in contrast, brightness, and texture detail, remains unsatisfactory. The TEBCF (Yuan et al., 2021) method sharpens images, greatly enhancing contrast and effectively improving low-light conditions. However, excessive

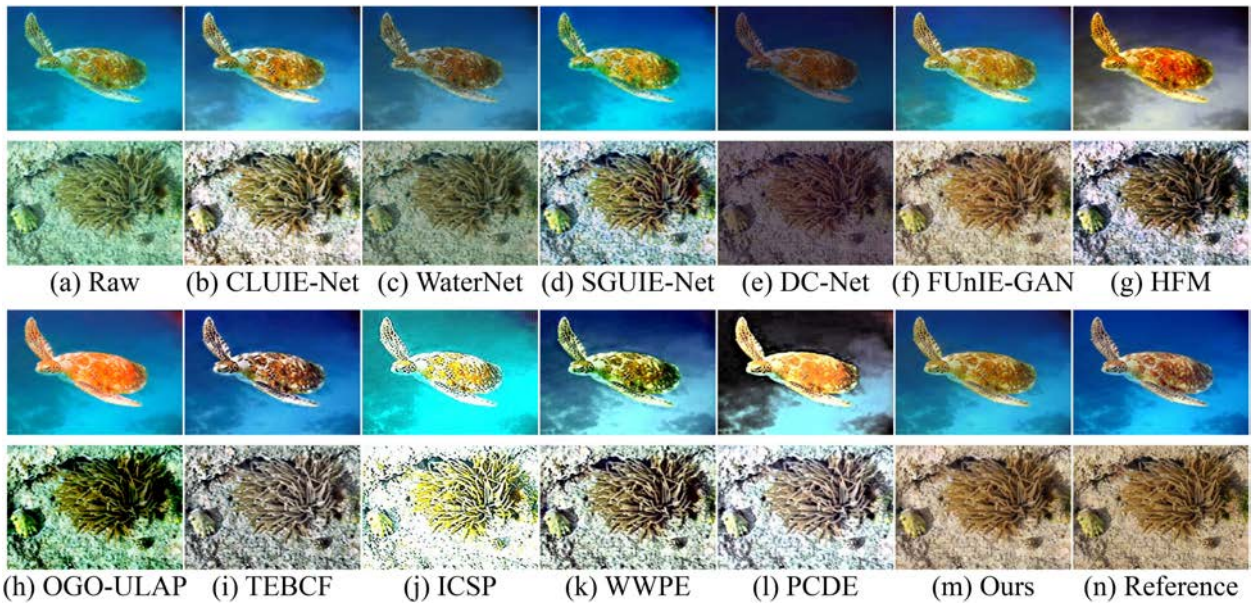

FIGURE 7
(a-n) Comparison of underwater image experiments on the EUVP (Islam et al., 2020a) dataset. From left to right: CLUIE-Net (Li et al., 2023b), WaterNet (Li et al., 2020b), SGUIE-Net (Qi et al., 2022), DC-Net (Zheng et al., 2024b), FUnIE-GAN (Islam et al., 2020b), HFM (An and Xu, 2024), OGO-ULAP (Li et al., 2023a), TEBCF (Yuan et al., 2021), ICSP (Hou et al., 2023), WWPE (Zhang et al., 2023b) and PCDE (Zhang et al., 2023a).

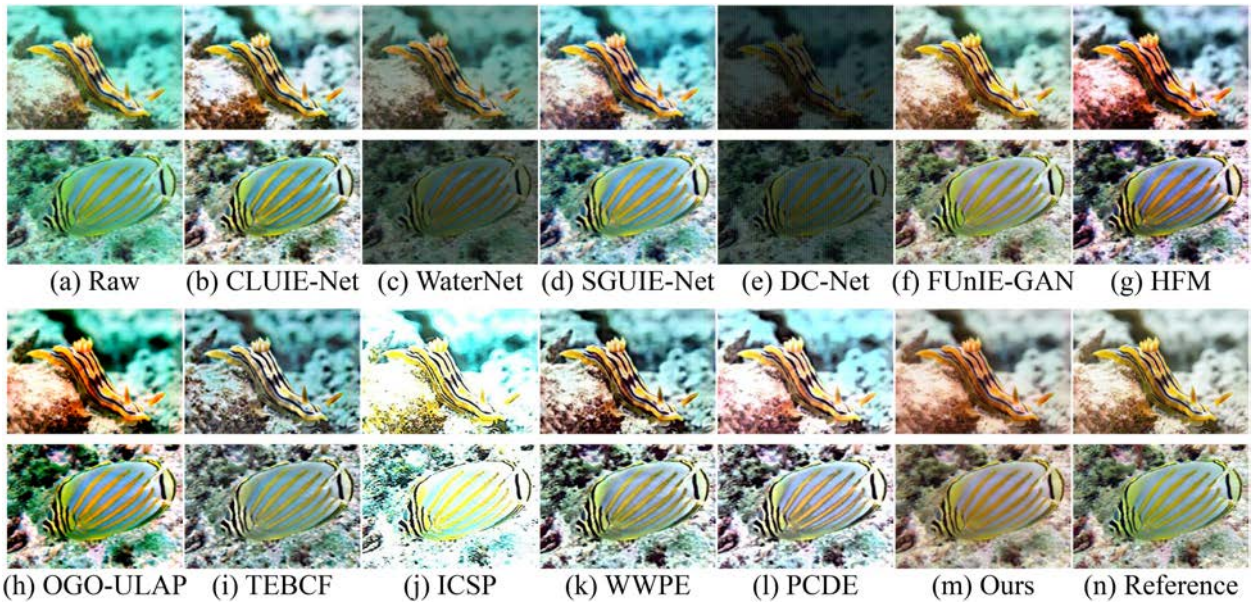

FIGURE 8
(a-n) Comparison of underwater image experiments on the EUVP (Islam et al., 2020a) dataset. From left to right: CLUIE-Net (Li et al., 2023b), WaterNet (Li et al., 2020b), SGUIE-Net (Qi et al., 2022), DC-Net (Zheng et al., 2024b), FUnIE-GAN (Islam et al., 2020b), HFM (An and Xu, 2024), OGO-ULAP (Li et al., 2023a), TEBCF (Yuan et al., 2021), ICSP (Hou et al., 2023), WWPE (Zhang et al., 2023b) and PCDE (Zhang et al., 2023a).

edge sharpening makes the images appear unnatural and leads to noticeable trailing shadows around the edges of objects. The OGO-ULAP (Li et al., 2023a) method employs a novel image gradient hypothesis, but images processed with OGO-ULAP (Li et al., 2023a) exhibit a significant red color cast and do not handle bluish-green biased images well. The WWPE (Zhang et al., 2023b) method uses weighted wavelet visual perception fusion technology, significantly improving image contrast and clarity, but still retains a slight bluish-green bias. The ICSP (Hou et al., 2023) method aims to address underwater non-uniform lighting using a variational framework, greatly enhancing image quality in terms of lighting improvement and detail preservation, but the overall color of the images is unnatural, and some areas still suffer from severe overexposure. Our proposed neural network effectively eliminates color casts and the effects of low-light environments, enhancing underwater image brightness and saturation.

4.3 Quantitative evaluation

To objectively evaluate degraded image restoration effectiveness, we further validate the effectiveness of our method by providing various evaluation metrics on samples from the UIEB (Li et al., 2020b) and EUVP (Islam et al., 2020a) datasets. In terms of full-reference metrics, we use the Structural Similarity Index Measure (SSIM), to measure the level of structural information loss during the enhancement process, the Peak Signal-to-Noise Ratio (PSNR) to assess the noise level in the image, and the Mean Squared Error (MSE) to reflect image distortion at the pixel level. Additionally, we introduce the Learned Perceptual Image Patch Similarity (LPIPS) and Fréchet Inception Distance (FID) to evaluate the perceptual quality of the image and the distribution difference between generated images and real images. Furthermore, to evaluate image quality from an overall visual perception perspective, we used the no-reference image quality assessment metric Underwater Image Quality Measure [UIQM (Panetta et al., 2015)] to quantitatively evaluate the result images in terms of color, edges and other aspects. Through a comprehensive comparison of all metrics, our method is observed to generally deliver superior performance. Table 1 presents the PSNR, MSE, SSIM, and UIQM (Panetta et al., 2015) metrics for different methods. The results show that our method outperforms the others on most metrics. The results show that our method outperforms all other compared methods on most metrics. Specifically, although the performance on the LPIPS and FID metrics is mediocre, compared to the second-best performing method, our network achieves percentage gains of 0.35 in PSNR on the UIEB (Li et al., 2020b) dataset. Additionally, our method obtains the lowest MSE score of 377, further demonstrating its superiority in preserving image detail and texture.

To further illustrate the our method's generalization ability, the ImageNet and scenes portions of the EUVP (Islam et al., 2020a) dataset are applied for testing. Tables 2, 3 presents the average scores of all metrics for the proposed MSCA-Net and ten UIE methods across two datasets. The data show that our method achieves the highest PSNR and SSIM on the Image dataset and the highest SSIM score on the Scenes dataset, with the PSNR score being only slightly lower than that of FUnIE-GAN (Islam et al., 2020b). Accordingly, our method demonstrates superior generalization ability compared to others. For the EUVP dataset, although the PSNR score of the FUnIE-GAN (Islam et al., 2020b) method is slightly higher than our method, visual inspection indicates that the images output by the FUnIE-GAN (Islam et al.,

TABLE 1 PSNR, SSIM, MSE, LPIPS, FID and UIQM (Panetta et al., 2015) are used for image quality assessment on the UIEB (Li et al., 2020b) dataset.

Method	PSNR↑	SSIM↑	MSE↓	UIQM↑	LPIPS↓	FID↓
HFM (An and Xu, 2024)	18.966	0.870	1104.454	4.790	0.216	66.021
OGO-ULAP (Li et al., 2023a)	16.681	0.810	1783.974	3.376	0.260	70.548
TCBEF (Yuan et al., 2021)	19.350	0.769	854.116	4.272	0.251	79.639
ICSP (Hou et al., 2023)	13.006	0.639	4426.687	2.166	0.502	106.377
WWPE (Zhang et al., 2023b)	19.445	0.775	897.869	4.002	0.227	57.250
PCDE (Zhang et al., 2023a)	16.829	0.660	1805.942	4.242	0.337	96.675
CLUIE-Net (Li et al., 2023b)	20.848	0.878	657.191	3.866	0.160	59.801
FUnIE-GAN (Islam et al., 2020b)	19.538	0.871	972.875	3.999	0.217	67.108
WaterNet (Li et al., 2020b)	16.333	0.818	1955.765	4.457	0.152	63.951
SGUIE-Net (Qi et al., 2022)	22.156	0.880	467.683	4.015	0.162	55.875
DC-Net (Zheng et al., 2024b)	15.983	0.659	1945.656	4.341	0.347	143.038
Ours	22.500	0.872	377.056	4.513	0.187	68.377

Deep learning-based methods and traditional methods are separated by a line, with deep learning methods listed in the upper half and traditional methods in the lower half. The top-performing method in each metric is highlighted in red, while the second-best is highlighted in blue.

2020b) method still exhibit background color bias, and the images generated by DC-Net (Zheng et al., 2024a) appear dark and contain some noise, as shown in Figure 8. Moreover, in terms of LPIPS and FID metrics, although our method did not achieve the best results on the UIEB dataset, only FUnIE-GAN can compare with our method on the EUVP dataset. However, FUnIE-GAN does not perform well on the UIEB dataset, which fully demonstrates the strong generalization ability of our method. Overall, our method not only improves color bias but also demonstrates superior performance with regard to brightness and clarity. Meanwhile, compared to all the methods, our method delivers the highest objective scores on most evaluation metrics.

We resize the test images to 256 × 256 and compare the runtimes of all methods on the UIEB dataset. The average runtime of each method is shown in Table 4, and our method achieves the second-best result, indicating that our method has an advantage in terms of runtime, but there is still room for improvement. In future work, we will investigate optimization strategies to cut computational complexity, including simplifying the model by reducing convolutional layers. Our aim is to decrease

TABLE 2 PSNR, SSIM, MSE and UIQM (Panetta et al., 2015) are used for image quality assessment on the on EUVP (Islam et al., 2020a) dataset.

Method	ImageNet				Scenes			
	PSNR↑	SSIM↑	MSE↓	UIQM↑	PSNR↑	SSIM↑	MSE↓	UIQM↑
HFM	18.978	0.795	1131.783	5.030	18.327	0.768	925.045	5.032
OGO-ULAP	16.190	0.684	1804.985	4.857	16.247	0.706	1829.933	7.290
TEBCF	17.335	0.682	1319.893	4.613	17.784	0.731	1289.933	4.789
ICSP	11.175	0.549	5327.004	2.852	11.405	0.599	5219.309	3.623
WWPE	16.328	0.651	1676.526	4.699	24.407	0.728	1543.944	4.898
PCDE	15.120	0.606	2154.250	5.852	15.814	0.658	1797.523	5.301
CLUIE-Net	19.193	0.827	925.297	4.567	18.814	0.782	825.461	4.680
FUnIE-GAN	23.840	0.813	300.239	5.100	26.568	0.872	160.810	5.288
WaterNet	16.060	0.529	1899.209	2.134	14.869	0.658	3590.675	2.345
SGUIE-Net	18.063	0.823	1148.591	3.762	18.532	0.846	1005.458	3.588
DC-Net	11.106	0.334	5437.929	5.337	11.357	0.411	5781.909	4.412
Ours	25.624	0.868	218.870	5.151	25.449	0.882	182.442	5.357

The top method in each metric is marked in red, and the second in blue.

TABLE 3 LPIPS and FID are used for image quality assessment on the EUVP (Islam et al., 2020a) dataset.

Method	ImageNet		Scenes	
	LPIPS↓	FID↓	LPIPS↓	FID↓
HFM	0.265	44.616	0.289	55.247
OGO-ULAP	0.313	48.826	0.351	64.204
TEBCF	0.293	48.957	0.315	62.468
ICSP	0.485	107.173	0.425	85.344
WWPE	0.296	43.271	0.302	53.674
PCDE	0.332	60.597	0.342	72.244
CLUIE-Net	0.248	39.103	0.267	44.032
FUnIE-GAN	0.173	27.245	0.131	29.302
WaterNet	0.264	90.151	0.291	93.682
SGUIE-Net	0.301	39.483	0.331	47.802
DC-Net	0.493	163.923	0.614	237.696
Ours	0.136	19.562	0.201	38.334

The top method in each metric is marked in red, and the second in blue.

complexity while maintaining accuracy, thus improving the practicality of our method.

4.4 Ablation study

To evaluate the effectiveness of the main modules in the network, we conducted extensive ablation experiments to analyze our method. We set up four groups of control experiments with the following specific details: (a) the original underwater image, (b) the network without the dense attention denoising module, (c) the network without residual connections in the dense convolution module, (d) The network with separate use of channel attention and pixel attention in the dual-path attention enhancement module, (e) the network with the dual-path attention enhancement module connected in parallel, (f) the network without the SSIM loss function, (g) the complete MSCA-Net, and (h) the reference image.

We compared all the ablation experiments, as shown in Figure 9. Visually, our complete model demonstrates the best overall enhancement effect in both objective metrics and visual quality. From Figure 9, it is evident that (b) and (c) exhibit low clarity, blurred details, and textures, with significant noise around the edges of objects, indicating that the dense attention denoising module did not achieve the desired effect. Comparing (f) and (g) reveals that the SSIM loss effectively eliminates the ghosting around the edges of objects in underwater images and improves image details. To evaluate the impact of using pixel attention and channel attention separately, we compared the two attention modules by placing them in different branches. As shown in (d), embedding PA and CA simultaneously within the serial-parallel framework results in severe green color bias throughout the image, and the excessive network depth significantly increases training time. The severe green color bias and low illumination hinder capturing image details, deviating from normal human perception. Moreover, when we stack the Dual-Path Attention Enhancement Module using parallel connections, the experimental results (e) show that the enhanced images contain slight artifacts, and most evaluation metrics significantly decline. The results indicate a cascaded connection outperforms a parallel one for attention modules, and a well-structured module arrangement yields better underwater images.

The metrics for each control group in the ablation experiments are presented in Table 5. We observe that excluding the dense attention denoising module and omitting the residual connections in the dense convolution module both result in a significant drop in the PSNR metric, highlighting the crucial effectiveness of the dense attention denoising module structure we employed. Additionally, due to the absence of the SSIM loss function for adjustment during training, Group (e) shows a significant decrease in the SSIM metric. Therefore, the results of the ablation experiments confirm that each module in MSCA-Net and loss function are essential for optimal performance.

4.5 Underwater vision applications

To further demonstrate the our method's effectiveness in underwater object detection, we tested the original and enhanced underwater images using the well-known YOLOv5 (Lei et al., 2022) and YOLOv7 (Liu et al., 2023) detection networks, as well as saliency detection techniques. By applying our proposed neural network model, all original underwater images in the dataset were enhanced to obtain their enhanced versions. The object detection results using YOLOv5 (Lei et al., 2022) and YOLOv7 (Liu et al., 2023) on original and enhanced underwater images are shown in Figures 10, 11, respectively. Figure 10 shows that noise and color bias in underwater images cause YOLOv5 (Lei et al., 2022) to produce errors and low confidence levels in object detection. There are even instances where the detector misidentifies objects, such as detecting a surfboard instead of a diver or failing to detect a diver

TABLE 4 Average runtime of different comparison methods.

Method	HFM	OGO-ULAP	TEBCF	ICSP	WWPE	PCDE
Time/s	0.463	0.080	1.017	0.148	0.342	0.219
Method	CLUIE-Net	FUnIE-GAN	WaterNet	SGUIE-Net	DC-Net	Ours
Time/s	0.231	0.002	0.324	0.247	0.077	0.075

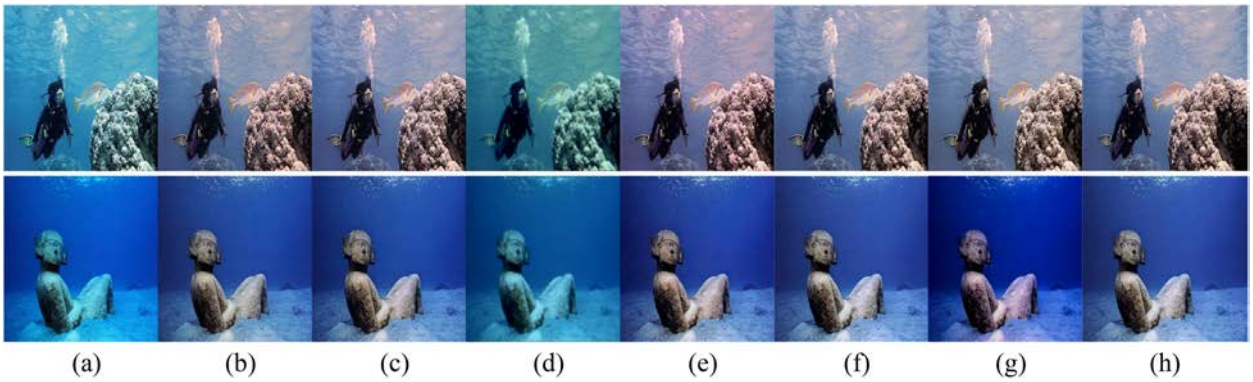

FIGURE 9
(a-h) The results of the ablation experiment.

TABLE 5 Quantitative results of ablation study on challenging images from the UIEB dataset.

Group	PSNR↑	SSIM↑	MSE↓	UIQM↑
b	17.225	0.780	453.560	4.138
c	17.499	0.785	567.878	4.204
d	19.493	0.881	493.637	4.373
e	15.662	0.839	1335.558	2.650
f	16.298	0.757	1983.130	4.556
g	20.116	0.872	377.056	4.513

The highest score is highlighted in red.

while detecting a backpack. The detection accuracy and confidence are significantly improved after enhancement with MSCA-Net.

The detection results using YOLOv7 (Liu et al., 2023) in Figure 11 are similar to those in Figure 10, where it is difficult to accurately detect fish and other aquatic organisms in the original underwater images. The edge details in the images are more prominent before and after enhancement, which significantly improves detection accuracy. This result confirms the practical value of our method in meeting the application needs for both human and machine-oriented tasks.

FIGURE 10
Underwater object detection using the YOLOv5 (Lei et al., 2022) model. Detection results for the raw images and enhanced images are respectively shown at the top and bottom.

FIGURE 11
Underwater object detection using the YOLOv7 (Liu et al., 2023) model. Detection results for the raw images and enhanced images are respectively shown at the top and bottom.

5 Conclusion

To more effectively fuse high-level semantic information and low-level detail information in underwater images for image enhancement, we propose an noval method based on a dual-path attention network. Our proposed method maximizes the seamless integration and effective fusion of both pixel-level information and detailed high-level features by employing dual-path connections between the channel attention mechanisms and pixel attention mechanisms, with the aim of minimizing computational complexity while maintaining high processing efficiency. Additionally, we utilize dense attention convolutional blocks to effectively extract and filter noise-related information from images, and we recommend the use of residual connections in the stacked network structure to achieve more robust and accurate image denoising. Experimental results clearly demonstrate that our method is highly capable of handling various complex image processing tasks, including significantly reducing unwanted color bias and substantially improving the overall image clarity and quality. Furthermore, extensive and detailed application studies confirm that our method achieves highly promising and competitive results in the fields of underwater target detection and recognition.

Although our method performs well in most underwater scenes, there are still some limitations. The deep convolutional layers in the dense convolutional denoising module have led to a significant increase in the model's training time. Therefore, in future work, we plan to reduce the model's training time by adopting deep convolutional networks to replace numerous convolutional layers, while also decreasing the number of stacked dense convolutional layers.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

GZ: Methodology, Resources, Supervision, Writing – original draft. YW: Formal analysis, Methodology, Resources, Software, Writing – original draft. LZ: Resources, Supervision, Validation, Writing – original draft. WYZ: Methodology, Supervision, Writing – original draft. WDZ: Funding acquisition, Investigation, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the China Postdoctoral Science Foundation project under Grant 2024M750747, the National Natural Science Foundation of China under Grant 62171252, the Key Specialized Research and Development Program of Science and Technology of Henan Province under Grants 242102211096, 242102210075, the Teacher Education Curriculum Reform Research of Henan Province under Grant 2024-JSJYYB-099, the Henan Provincial Science and Technology Research and Development Joint Foundation Project under Grant 235200810066, the Postdoctoral Fellowship Program of CPSF under Grant Number GZC20240163, and the National Natural Science Foundation of China under Grant 62403066.

Acknowledgments

This brief text acknowledges the contributions of specific colleagues, institutions, or agencies that assisted the authors' efforts.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

An, S., and Xu, L. (2024). Hfm: A hybrid fusion method for underwater image enhancement. *Eng. Appl. Artif. Intell.* 127, 107219. doi: 10.1016/j.engappai.2023.107219

Bi, X., Wang, P., Guo, W., Zha, F., and Sun, L. (2024). Rgb/event signal fusion framework for multi-degraded underwater image enhancement. *Front. Mar. Sci.* 11. doi: 10.3389/fmars.2024.1366815

Fabbri, C., Islam, M. J., and Sattar, J. (2018). "Enhancing underwater imagery using generative adversarial networks," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. (Brisbane, Queensland, Australia: 2018 IEEE International Conference on Robotics and Automation, ICRA 2018), 7159–7165.

Fan, Y., Qian, Y., Qin, Y., Wan, Y., Gong, W., Chu, Z., et al. (2022). Mslaenet: Multiscale learning and attention enhancement network for fusion classification of hyperspectral and lidar data. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 15, 10041–10054. doi: 10.1109/JSTARS.2022.3221098

Guan, M., Xu, H., Jiang, G., Yu, M., Chen, Y., Luo, T., et al. (2024). Diffwater: Underwater image enhancement based on conditional denoising diffusion probabilistic model. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 17, 2319–2335. doi: 10.1109/JSTARS.2023.3344453

Hou, G., Li, N., Zhuang, P., Li, K., Sun, H., Li, C., et al. (2023). Non-uniform illumination underwater image restoration via illumination channel sparsity prior. *IEEE Trans. Circuits Syst. Video Technol.* 34, 799–814. doi: 10.1109/TCSVT.2023.3290363

Hou, G., Li, N., Zhuang, P., Li, K., Sun, H., and Li, C. (2024). Non-uniform illumination underwater image restoration via illumination channel sparsity prior. *IEEE Trans. Circuits Syst. Video Technol.* 34, 799–814. doi: 10.1109/TCSVT.2023.3290363

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Piscataway, NJ, USA: IEEE), 7132–7141.

Islam, M. J., Xia, Y., and Sattar, J. (2020a). Fast underwater image enhancement for improved visual perception. *IEEE Robotics Automation Lett.* 5, 3227–3234. doi: 10.1109/LSP.2016.

Islam, M. J., Youya, A., and Sattar, J. (2020b). Fast underwater image enhancement for improved visual perception. *IEEE Robotics Automation Lett.* 5, 3227–3234. doi: 10.1109/LSP.2016.

Jia, Y., Wang, Z., and Zhao, L. (2024). An unsupervised underwater image enhancement method based on generative adversarial networks with edge extraction. *Front. Mar. Sci.* 11. doi: 10.3389/fmars.2024.1471014

Lei, F., Tang, F., and Li, S. (2022). Underwater target detection algorithm based on improved yolov5. *J. Mar. Sci. Eng.* 10, 310. doi: 10.3390/jmse10030310

Li, C., Anwar, S., and Porikli, F. (2020a). Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition* 98, 107038. doi: 10.1016/j.patcog.2019.107038

Li, C.-Y., Guo, J.-C., Cong, R.-M., Pang, Y.-W., and Wang, B. (2016). Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *IEEE Trans. Image Process.* 25, 5664–5677. doi: 10.1109/TIP.2016.2612882

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., et al. (2020b). An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.* 29, 4376–4389. doi: 10.1109/TIP.83

Li, J., Hou, G., and Wang, G. (2023a). Underwater image restoration using oblique gradient operator and light attenuation prior. *Multimedia Tools Appl.* 82, 6625–6645. doi: 10.1007/s11042-022-13605-5

Li, J., Skinner, K. A., Eustice, R. M., and Johnson-Roberson, M. (2018). Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robotics Automation Lett.* 3, 387–394. doi: 10.1109/LRA.2017.2730363

Li, K., Wu, L., Qi, Q., Liu, W., Gao, X., Zhou, L., et al. (2023b). Beyond single reference for training: Underwater image enhancement via comparative learning. *IEEE Trans. Circuits Syst. Video Technol.* 33, 2561–2576. doi: 10.1109/TCSVT.2022.3225376

Liang, D., Chu, J., Cui, Y., Zhai, Z., and Wang, D. (2024). Npt-ul: An underwater image enhancement framework based on nonphysical transformation and unsupervised learning. *IEEE Trans. Geosci. Remote Sens.* 62, 1–19. doi: 10.1109/TGRS.2024.3363037

Liang, Z., Ding, X., Wang, Y., Yan, X., and Fu, X. (2022). Gudcp: Generalization of underwater dark channel prior for underwater image restoration. *IEEE Trans. Circuits Syst. Video Technol.* 32, 4879–4884. doi: 10.1109/TCSVT.2021.3114230

Lin, Y., Zhou, J., Ren, W., and Zhang, W. (2021). Autonomous underwater robot for underwater image enhancement via multi-scale deformable convolution network with attention mechanism. *Comput. Electron. Agric.* 191, 106497. doi: 10.1016/j.compag.2021.106497

Liu, S., Fan, H., Lin, S., Wang, Q., Ding, N., and Tang, Y. (2022). Adaptive learning attention network for underwater image enhancement. *IEEE Robotics Automation Lett.* 7, 5326–5333. doi: 10.1109/LRA.2022.3156176

Liu, K., Sun, Q., Sun, D., Peng, L., Yang, M., and Wang, N. (2023). Underwater target detection based on improved yolov7. *J. Mar. Sci. Eng.* 11, 677. doi: 10.3390/jmse11030677

Liu, T., Zhu, K., Wang, X., Song, W., and Wang, H. (2024a). Lightweight underwater image adaptive enhancement based on zero-reference parameter estimation network. *Front. Mar. Sci.* 11. doi: 10.3389/fmars.2024.1378817

Liu, Y., Li, E., Liu, W., Li, X., and Zhu, Y. (2024b). Lfemap-net: Low-level feature enhancement and multiscale attention pyramid aggregation network for building extraction from high-resolution remote sensing images. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 17, 2718–2730. doi: 10.1109/JSTARS.2023.3346454

Misra, D., Namada, T., Arasanipalai, A. U., and Hou, Q. (2021). "Rotate to attend: Convolutional triplet attention module," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. (Piscataway, NJ, USA: IEEE), 3139–3148.

Panetta, K., Gao, C., and Agaian, S. (2015). Human-visual-system-inspired underwater image quality measures. *IEEE J. Oceanic Eng.* 41, 541–551. doi: 10.1109/JOE.2015.2469915

Peng, Y.-T., and Cosman, P. C. (2017). Underwater image restoration based on image blurriness and light absorption. *IEEE Trans. Image Process.* 26, 1579–1594. doi: 10.1109/TIP.2017.2663846

Qi, Q., Li, K., Zheng, H., Gao, X., Hou, G., and Sun, K. (2022). Sguie-net: Semantic attention guided underwater image enhancement with multi-scale perception. *IEEE Trans. Image Process.* 31, 6816–6830. doi: 10.1109/TIP.2022.3216208

Shen, Y., Zhao, C., Liu, Y., Wang, S., and Huang, F. (2021). Underwater optical imaging: Key technologies and applications review. *IEEE Access* 9, 85500–85514. doi: 10.1109/ACCESS.2021.3086820

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for rge-scale image recognition. *arXiv preprint arXiv* 1409, 1556. doi: 10.48550/arXiv.1409.1556

Tolie, H. F., Ren, J., and Elyan, E. (2024). Dicam: Deep inception and channel-wise attention modules for underwater image enhancement. *Neurocomputing* 584, 127585. doi: 10.1016/j.neucom.2024.127585

Wang, B., Xu, H., Jiang, G., Yu, M., Ren, T., Luo, T., et al. (2024). Uie-convformer: Underwater image enhancement based on convolution and feature fusion transformer.

IEEE Trans. Emerging Topics Comput. Intell. 8, 1952–1968. doi: 10.1109/TETCI.2024.3359061

Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. (2018). "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*. (Berlin, Heidelberg, Germany: Springer).

Yu, M., Shen, L., Wang, Z., and Hua, X. (2024). Task-friendly underwater image enhancement for machine vision applications. *IEEE Trans. Geosci. Remote Sens.* 62, 1–14. doi: 10.1109/TGRS.2024.3509985

Yuan, J., Cai, Z., and Cao, W. (2021). Tebcf: Real-world underwater image texture enhancement model based on blurriness and color fusion. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15. doi: 10.1109/TGRS.2021.3110575

Zhang, L., Ma, Z., Zhou, J., Li, K., Li, M., Wang, H., et al. (2024c). A multi-scale convolutional hybrid attention residual network for enhancing underwater image and identifying underwater multi-scene sea cucumber. *IEEE Robotics Automation Lett.* 9, 7397–7404. doi: 10.1109/LRA.2024.10595499

Zhang, S., Zhao, S., An, D., Liu, J., Wang, H., Feng, Y., et al. (2022a). Visual slam for underwater vehicles: A survey. *Comput. Sci. Rev.* 46, 100510. doi: 10.1016/j.cosrev.2022.100510

Zhang, W., Jin, S., Zhuang, P., Liang, Z., and Li, C. (2023a). Underwater image enhancement via piecewise color correction and dual prior optimized contrast enhancement. *IEEE Signal Process. Lett.* 30, 229–233. doi: 10.1109/LSP.2023.3255005

Zhang, W., Li, Z., Li, G., Zhuang, P., Hou, G., Zhang, Q., et al. (2024a). Gacnet: Generate adversarial-driven cross-aware network for hyperspectral wheat variety identification. *IEEE Trans. Geosci. Remote Sens.* 62, 1–14. doi: 10.1109/TGRS.2023.3347745

Zhang, W., Zhao, W., Li, J., Zhuang, P., Sun, H., Xu, Y., et al. (2024b). Cvanet: Cascaded visual attention network for single image super-resolution. *Neural Networks* 170, 622–634. doi: 10.1016/j.neunet.2023.11.049

Zhang, W., Zhou, L., Zhuang, P., Li, G., Pan, X., Zhao, W., et al. (2023b). Underwater image enhancement via weighted wavelet visual perception fusion. *IEEE Trans. Circuits Syst. Video Technol.* 34, 2469–2483. doi: 10.1109/TCSVT.2023.3299314

Zhang, W., Zhuang, P., Sun, H.-H., Li, G., Kwong, S., and Li, C. (2022b). Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Trans. Image Process.* 31, 3997–4010. doi: 10.1109/TIP.2022.3177129

Zhang, Y., Liu, Y., Sun, P., Yan, H., Zhao, X., and Zhang, L. (2020). Ifcnn: A general image fusion framework based on convolutional neural network. *Inf. Fusion* 54, 99–118. doi: 10.1016/j.inffus.2019.07.011

Zheng, S., Wang, R., and Chen, G. (2024a). Underwater image enhancement using divide-and-conquer network. *PloS One* 19, e0294609. doi: 10.1371/journal.pone.0294609

Zheng, S., Wang, R., Zheng, S., Wang, F., Wang, L., and Liu, Z. (2024b). A multi-scale feature modulation network for efficient underwater image enhancement. *J. King Saud Univ. - Comput. Inf. Sci.* 36, 101888. doi: 10.1016/j.jksuci.2024.101888

Zhou, J., Wei, X., Shi, J., Chu, W., and Zhang, W. (2022). Underwater image enhancement method with light scattering characteristics. *Comput. Electrical Eng.* 100, 107898. doi: 10.1016/j.compeleceng.2022.107898

Zhuang, P., and Ding, X. (2020). Underwater image enhancement using an edge-preserving filtering retinex algorithm. *Multimedia Tools Appl.* 79, 17257–17277. doi: 10.1007/s11042-019-08404-4