



OPEN ACCESS

EDITED BY
Xuebo Zhang,
Northwest Normal University, China

REVIEWED BY
Ke Qu,
Guangdong Ocean University, China
Hanhao Zhu,
Zhejiang Ocean University, China

*CORRESPONDENCE
Liquan Zhao
✉ zhaoliqun@neepu.edu.cn

RECEIVED 26 July 2024
ACCEPTED 20 November 2024
PUBLISHED 12 December 2024

CITATION
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. 11:1471014.
doi: 10.3389/fmars.2024.1471014

COPYRIGHT
© 2024 Jia, Wang and Zhao. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

An unsupervised underwater image enhancement method based on generative adversarial networks with edge extraction

Yanfei Jia¹, Ziyang Wang² and Liquan Zhao^{3*}

¹College of Electrical and Information Engineering, Beihua University, Jilin, China, ²Mobile Internet Service Center, Beijing Zhongdian Feihua Communication Company Limited, Beijing, China, ³College of Electrical Engineering, Northeast Electrical Power University, Jilin, China

Underwater environments pose significant challenges for image capture due to factors like light absorption, scattering, and the presence of particles in the water. These factors degrade the quality of underwater images, impacting tasks like target detection and recognition. The challenge with deep learning-based underwater image enhancement methods is their reliance on paired datasets, which consist of degraded and corresponding ground-truth images. Obtaining such paired datasets in natural conditions is challenging, leading to performance issues in these algorithms. To address this issue, we propose an unsupervised generative adversarial network with edge detection for enhancing underwater images without needing paired data. First, we introduce the perceptual loss function into the conventional loss function to better measure the performance of two generative networks. Second, we propose an edge extraction block based on the Laplacian operator, an attention module with an edge extraction block, a multi-scale feature module, a novel upsampling module, and a new downsampling module. We use these proposed modules to design a new generative network. Third, we use the proposed multi-scale feature and downsampling modules to design the adversarial network. We tested the algorithm's performance on both synthetic and authentic underwater images. Compared to existing state-of-the-art methods, our proposed approach better enhances image details and restores color information.

KEYWORDS

underwater image enhancement, generative adversarial networks, image quality, deep learning, edge extraction

1 Introduction

Underwater images are crucial for acquiring and documenting underwater information and accurately preserving and exploring underwater resources. However, underwater images often suffer from low brightness and color distortion caused by water's selective light absorption. Additionally, suspended particles in the water can scatter light, reducing

the contrast and clarity of underwater images. These distortions make extracting valid information from underwater images challenging, significantly impacting advanced visual tasks such as marine object detection (Hua et al., 2023; Xu et al., 2023) and visual tracking (Li et al., 2023; Cai et al., 2023a). Therefore, enhancing underwater images is essential to improve the performance of underwater visual perception.

With the continuous exploration of the marine world, numerous methods for enhancing underwater images have been proposed. Traditional underwater image enhancement methods can be classified into model-free and model-based methods. Model-free methods improve image quality by adjusting pixels (Zhang et al., 2022; Muniraj and Dhandapani, 2023). These methods are easy to implement but often overlook the unique characteristics of underwater images, resulting in enhanced images with noise and artifacts. On the other hand, model-based methods establish a degradation model for underwater images to restore their quality (Zhang et al., 2023a; Zhou et al., 2023a). While these methods perform well in specific water conditions due to the selection of priors, they suffer from poor robustness. Moreover, fixed physical models struggle to simulate complex and variable underwater environments, making them inadequate for applications under different environmental conditions.

With the development of deep learning, numerous learning-based methods for underwater image enhancement have emerged in recent years (Shen et al., 2023; Peng et al., 2023). These methods do not require the construction of physical models and exhibit good robustness by training networks with large datasets of degraded and high-quality images (Guan et al., 2024). The Generative Adversarial Networks (GANs) is a typical deep learning model comprising generative and adversarial networks. Thanks to the adversarial learning approach, underwater image enhancement methods based on GANs generally outperform those based on convolutional neural networks, making them widely used for this purpose (Han et al., 2023; Zheng et al., 2023).

However, most deep learning-based methods depend on paired datasets for training. In practical applications, acquiring many paired underwater images is challenging, which limits the performance of these methods. To address this issue and achieve underwater image enhancement without paired images, we propose an unsupervised underwater image enhancement method based on generative adversarial networks with edge extraction.

The main contributions of this paper are summarized as follows:

- We designed some modules that include a new edge extraction block based on the Laplacian operator, a new attention module with the proposed edge extraction block, a new multi-scale feature module, a new upsampling module, and a new downsampling module.
- We designed a novel generative adversarial network using the above proposed modules to enhance the underwater image. It contains two generative networks with the same architecture and two adversarial networks with the same architecture. It can be training using the unpaired images.

- We proposed an improved loss function by introducing the perceptual loss function into the conventional loss function to better measure the performance of two generative networks. It can improve the image enhancement performance of the generative network.

2 Related work

In recent years, With the advancement of marine science research, many methods for underwater image enhancement have been proposed. These methods are divided into three categories: model-free, model-based, and deep learning-based. Model-free methods mainly use the improved traditional image enhancement methods to achieve underwater image enhancement, such as histogram equalization (Ulutas and Ustubioglu, 2021; Xiang et al., 2023), white balance (An et al., 2024), and Retinex-based methods (Mishra et al., 2023; Zhang et al., 2023b). Model-free underwater image enhancement methods are usually effective in correcting image color and improving image contrast. However, the color style of the enhanced image is usually unnatural and prone to noise and artifacts, as well as over-enhancement.

In contrast, the model-based underwater image enhancement methods establish physical models for the degradation process of underwater images and use prior knowledge to estimate the parameters of the model. The model is finally reversed to obtain the enhanced underwater image (Zhou et al., 2022; Ding et al., 2022). While these methods demonstrate outstanding performance in specific environments, they exhibit poor robustness, making them challenging to adapt to the diverse light and visual conditions encountered underwater in practical applications. Furthermore, acquiring precise prior knowledge is challenging due to the underwater environment's complexity and variability. Consequently, the estimation of model parameters may not be sufficiently accurate, affecting the quality of the enhanced underwater images.

In recent years, deep learning techniques have achieved great success. Due to its great nonlinear mapping learning ability, the convolutional neural network is widely used in the fields of image dehazing (Ding et al., 2023; Chen et al., 2024), image segmentation (Wang et al., 2023; Zhang T. et al., 2023), and underwater image enhancement (Cai et al., 2023b; Zhang et al., 2024a). It learns the mapping from degraded images to high-quality images through a large amount of training data, which is more suitable for different underwater environments. For example, Zhou et al., (2023b) proposed CVE-Net. This network is based on a cross-domain learning method for underwater image enhancement, using a feature alignment module to utilize different temporal features. The method performs better in detail enhancement and color correction but is poorly robust for practical applications. Liu et al., (2023a) proposed MSDC-Net consisting of a color correction block and an asymmetric multi-scale encoder-decoder. This network extracts features in both RGB and HSV color spaces, thus extracting diverse features. The method can improve the

quality of underwater images, but its color correction effect is unsatisfactory. These methods significantly improve the quality of underwater images through convolutional neural networks' powerful feature extraction capability. However, they rely on many paired data and have poor generalization ability in practical applications.

Generative Adversarial Networks have advantages in detail texture generation and style transfer of images by adversarial training of two networks. Therefore, it is applied to underwater image enhancement. Panetta et al., (2022). proposed a cascaded residual network for underwater image enhancement method for generating high-quality underwater images. They designed a cascaded block that consists of three residual blocks and used the cascaded blocks with convolutions to design the generative network with U-Net architecture. It has better performance in reducing the effect of light refraction and attenuation, and color distortion in enhancing underwater images. However, the use of cascaded blocks leads to higher model complexity, and it requires pairs of images during training. Liu et al., (2023b) proposed a weak-strong dual supervised generative adversarial network for enhancing underwater images. It included two phases: the weak supervised learning phase and the strong supervised learning phase. In the weak supervised learning phase, it used unpaired underwater images to train the network. In the strongly supervised learning phase, it used a small amount of paired underwater images to train the network. Although it reduces the dependence of the training process on paired images, it still requires a few paired images and the details of enhanced images are not satisfactory. Besides, it requires training the model using unpaired and paired images separately, resulting in difficult model training. Cong et al. (Cong et al., 2023). proposed a physical model-guided GAN model for enhancing underwater images. They separately split the generative network into two sub-networks for physical model parameter estimation and model-guided image enhancement. The method also designed dual-discriminative networks for style-content adversarial constraints to improve the visualization of the image. The method can achieve effective underwater image enhancement when there is sufficient prior information. However, it requires estimation of the physical model parameters of underwater images, making it challenging to accurately estimate these parameters without prior information, which directly affects the image enhancement performance of the algorithm. Additionally, the physical model parameters vary across different underwater environments, making it difficult for the pre-trained model to adapt to underwater image enhancement in varying conditions. Besides, it also required the paired images to train the module. Wang et al. (Wang et al., 2024). proposed a novel Self-Adversarial Generative Adversarial Network. The network uses a self-adversarial mode to improve the quality of the generated image by two pairwise image quality discriminators. This method effectively corrects the image color, but the enhanced image still suffers from detail blurring. In addition, the method still needs to utilize pairs of images in the training process. Jiang et al., (2024). proposed an unsupervised perception-driven unsupervised generative adversarial network to enhance the underwater images. They trained a pairwise quality ranking network consistent with

human visual perception. After that, they used the pre-trained pairwise quality ranking network as a loss function to improve the performance of the enhancement network. This method can improve the visual quality of images. However, the pairwise quality ranking model still be trained using underwater image pair. Islam et al. (Islam et al., 2022). proposed a FUnIE-GAN method based on conditional generative adversarial networks for enhancing underwater images. They used L1 loss and perceptual loss as the loss function of the network and supported both paired and unpaired data training. The network improves the visual quality of underwater images on the EUVP underwater dataset and has faster inference rates. However, it is not very effective for enhancing some degraded and texture-less images and prone to training instability. Liu et al. (Liu et al., 2022). proposed TACL for task-oriented underwater image enhancement. The model constructed a bilateral constrained closed-loop adversarial enhancement module to realize unpaired training. During the training process, a task-aware feedback module is employed to enhance the visual quality of the detected targets in the image. Although this method can achieve the enhancement of underwater images without the need for training on pairs of images, the enhanced images still have some color and texture distortion. They only used VGG-19 as a perceptual feature extraction network, which is not sufficient for feature extraction. Yan et al., (2023). proposed UW-CycleGAN. They used a CycleGAN-based network to estimate the parameters of the underwater image degradation model to recover the color and details of underwater images. This method can effectively improve the accuracy of underwater object detection. It has better performance in restoring color-distorted and blurred underwater images and does not require training the model with the paired underwater images. However, the model requires the estimation of information such as underwater illumination and depth, making it difficult to adapt to different underwater environments. Additionally, the high complexity of the model impacts the speed of underwater image enhancement.

For the traditional model-free methods, the complexity of underwater environments makes it difficult to obtain satisfactory enhanced underwater images in dynamic underwater conditions. Although conventional deep learning-based underwater image enhancement methods can achieve better results than traditional model-free methods, they often suffer from issues of over-enhancement and require paired training images. The generative adversarial networks (GANs) are a specialized deep learning method consisting of generative network and adversarial network. Compared to conventional networks, GANs offer stronger underwater image enhancement capabilities, resulting in visually superior enhanced images. The supervised generative adversarial networks require paired underwater images during the training process. The effectiveness of supervised methods depends to some extent on the quality of the paired data sets, which are typically synthesized using *a priori* information. However, the synthesized images can hardly reflect the real conditions of underwater images, resulting in low generalizability in real-world applications. Although underwater image enhancement methods based on unsupervised generative adversarial networks do not rely on paired data and are more suitable for practical applications,

existing methods have not fully utilized edge feature information and lack adequate feature extraction. This results in recovered images still suffering from some detail and color distortion. To solve the problems, we proposed an unsupervised generative adversarial network for underwater image enhancement to improve the quality of underwater images. This model's training does not rely on paired underwater images, making it more suitable for practical applications. The network consists of our proposed multi-scale feature module, attention module, upsampling module, and downsampling module. Additionally, to better restore object edge details, we proposed first extracting image edges using the Laplacian operator, then fusing the extracted edge image with the original image, followed by feature extraction and recovery on the fused image.

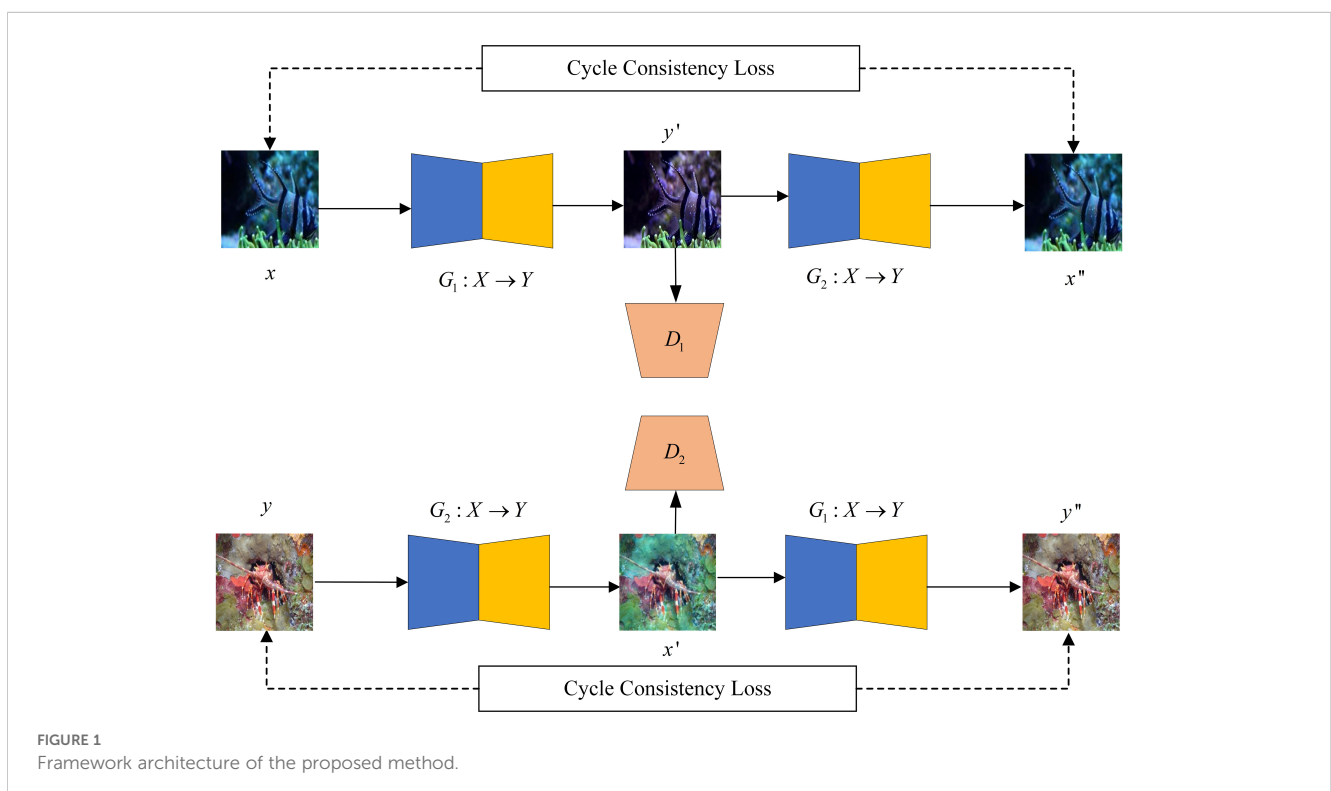
3 Proposed method

To enhance underwater images in the absence of paired underwater images, we proposed an unsupervised generative adversarial network with edge detection. The network is based on the CycleGAN framework with the structure shown in Figure 1, which contains two generative networks and two adversarial networks. The generative network G_1 enhances the degraded underwater image, and the adversarial network D_1 determines whether the input is a generative enhanced image or an original high-quality image. Similarly, the generative network G_2 simulates the degradation process of underwater images, and the adversarial network D_2 is responsible for determining whether the input is a generated degraded image or an original degraded image. During

training, the whole network consists of two branches: the enhancement-degradation branch and the degradation-enhancement branch. Each branch consists of two generative networks constituting a recurrent structure, which is convenient for constraining network training using the cycle consistency loss. Meanwhile, the two pairs of generative networks and adversarial networks have the same structure and share the network parameters to ensure the efficiency of the training process and the simplicity of the model. Next, we will introduce the structure of the generative network and adversarial network and the composition of the loss function.

3.1 Edge extraction based on Laplacian

Underwater imaging is affected by the absorption and scattering of water, often resulting in issues such as blurred object edges. Laplacian edge detection is a classical method that emphasizes details and edge information in an image through second-order differential operators, enhancing clarity and contrast. By applying the Laplacian operator to underwater images for edge detection, the edges of submerged objects can be highlighted, facilitating subsequent feature extraction modules to obtain more effective edge and detail information of underwater objects, ultimately improving the quality of the enhanced underwater images. Edge information in images is typically characterized by abrupt changes in gray values. Consequently, the Laplacian operator is frequently utilized in underwater image processing to capture the edge and texture details of an image. In our method, we employ the Laplacian operator as a convolutional kernel in the convolutional layer to



extract edge features from the images. We first apply the Laplacian operator to the degraded underwater image for edge detection, resulting in an edge map that exhibits more pronounced edge features compared to the original image. Next, we perform a channel-wise concatenation of the edge map with the degraded underwater image. This concatenated image effectively highlights object edge features, making it suitable as input for the feature extraction network. This approach facilitates the extraction of richer edge characteristics, ultimately enhancing the capabilities of underwater image enhancement. If f is a second-order differentiable function in two dimensions, the Laplacian operator is typically expressed as follows:

$$\nabla^2 f(x, y) = \frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2} \quad (1)$$

where $f(x, y)$ is the gray scale function of the image, ∇^2 is the Laplacian operator, $\frac{\partial^2 f(x, y)}{\partial x^2}$ and $\frac{\partial^2 f(x, y)}{\partial y^2}$ are the second-order partial derivatives of the image function in the horizontal and vertical directions, respectively. The image function $f(x, y)$ is typically discrete. Therefore, the second-order partial derivatives are approximated using the second-order central difference with the following equations:

$$\frac{\partial^2 f(x, y)}{\partial x^2} = f(x + 1, y) + f(x - 1, y) - 2f(x, y) \quad (2)$$

$$\frac{\partial^2 f(x, y)}{\partial y^2} = f(x, y + 1) + f(x, y - 1) - 2f(x, y) \quad (3)$$

At this point, the Laplacian operator of the image is:

$$\begin{aligned} \nabla^2 f(x, y) = & f(x + 1, y) + f(x - 1, y) + f(x, y + 1) + f(x, y - 1) \\ & - 4f(x, y) \end{aligned} \quad (4)$$

In the network, edge extraction based on the Laplacian operator can be achieved through convolution. The Laplacian convolutional kernel is illustrated in Figure 2. The convolution calculation result is zero when neighboring pixels have the same gray value. However, when the convolution kernel slides to the edge part where neighboring pixels have different gray values, the convolution calculation result is non-zero. This indicates the extraction of this edge feature.

3.2 Generative network

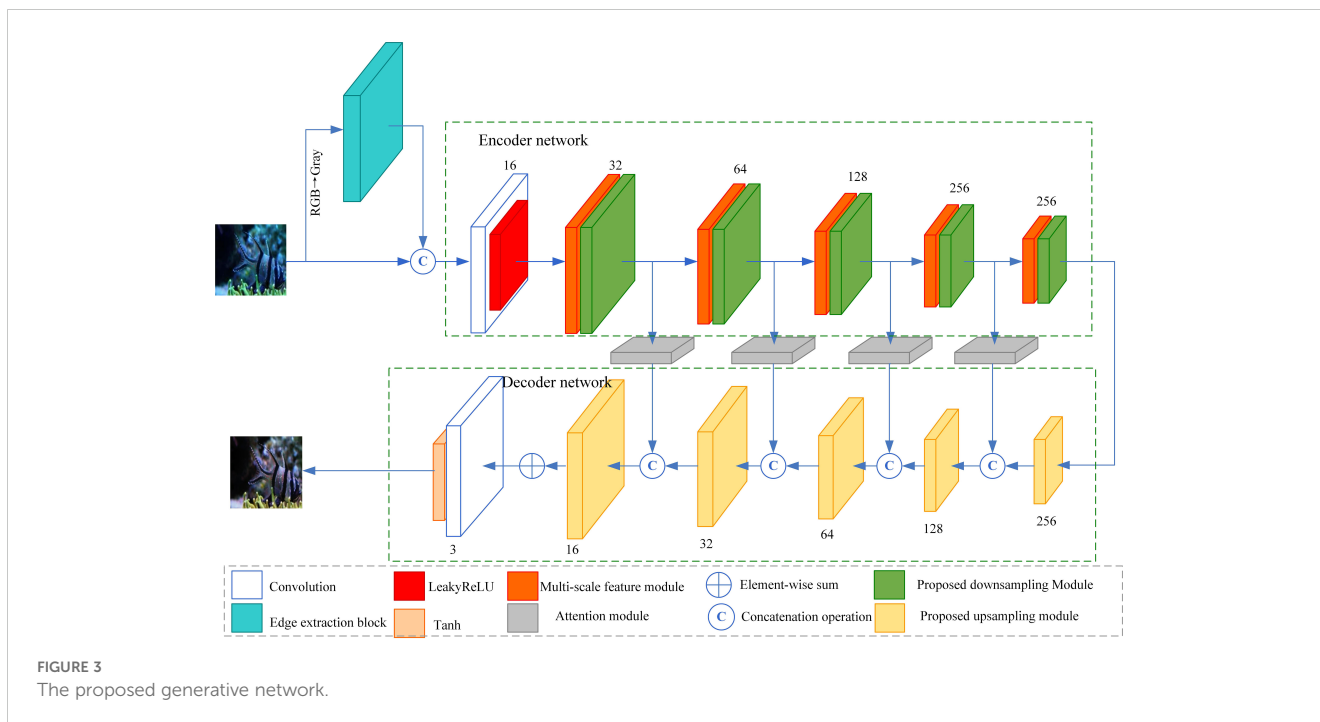
3.2.1 Overall introduction of the proposed generative network

Our proposed generative network is a deep convolutional neural network based on an encoder-decoder structure. The structure is shown in Figure 3. The underwater degraded image is first transformed into a grayscale image before it is imported into the generative network, and an edge image is generated after the edge extraction module using the Laplacian operator. This edge image is concatenated with the underwater degraded image in the channel dimension and later imported into the encoder network. The generative network consists of an encoder network and a decoder network. The encoder network extracts the deep feature information of the image through downsampling operations. The decoder network recovers the image size and reconstructs the image details through upsampling operations.

In the encoder network, we first use a convolutional layer and LeakyReLU activation function to extract local features and increase the number of input image channels from 4 to 16. Then, five multi-scale feature modules and downsampling modules are used to reduce the feature map size and extract multi-scale features. The multi-scale feature module extracts the multi-scale features without

0	1	0
1	-4	1
0	1	0

FIGURE 2
Laplacian convolution kernel.



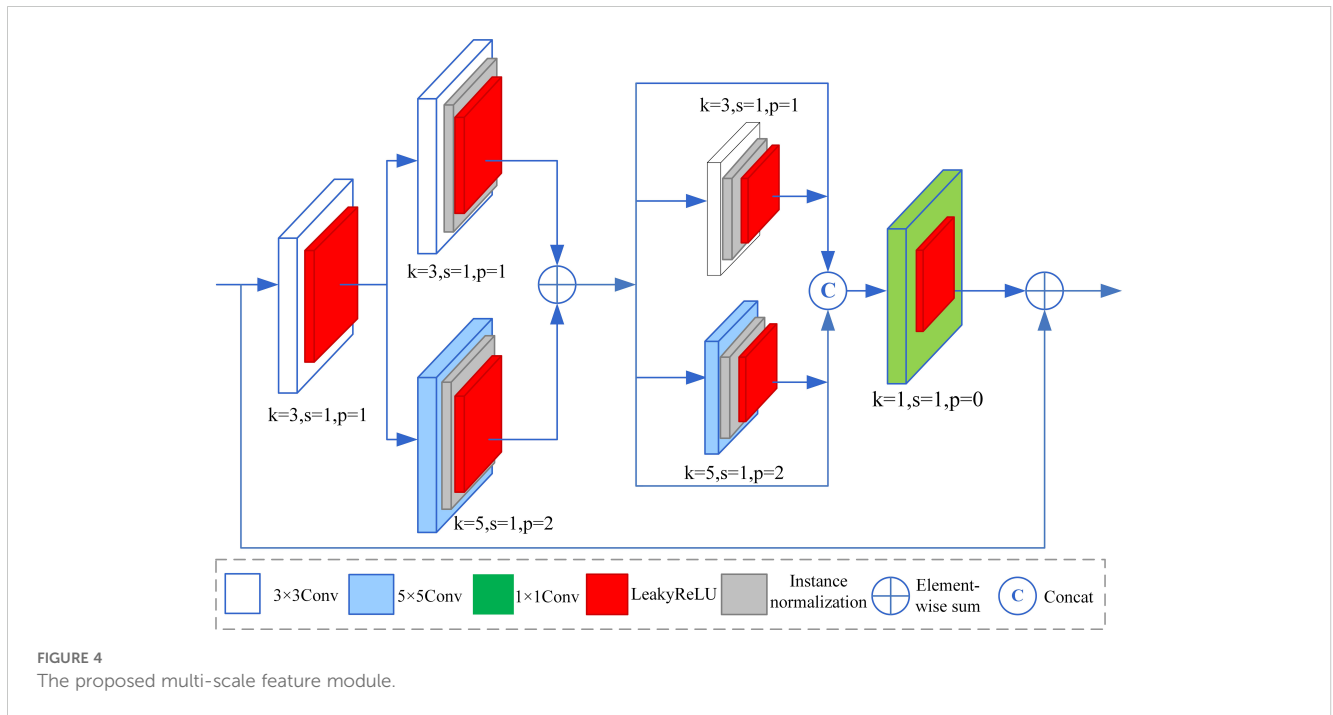
changing the feature map size and the number of channels, which is convenient for the following operations. Each proposed downsampling module downsamples the input feature maps to reduce the size of the output feature maps to half that of the input feature maps. The number of channels of the output feature maps for five downsampling operations are 32, 64, 128, 256, and 256, respectively. Finally, we use four skip connections to transmit the output feature maps of the proposed downsampling modules to the decoder network. Each skip connection contains an attention module to reduce the interference of redundant feature information. In the encoder network, a small number of downsampling modules can only extract shallow features from underwater images. By increasing the number of downsampling modules, the network can capture more complex features, thereby enhancing its representational and generalization capabilities. However, as the number of downsampling modules increases, the feature dimensions decrease, leading to a less comprehensive feature extraction and negatively impacting underwater image enhancement. Additionally, increased network complexity may cause issues like vanishing or exploding gradients, complicating the training process. Therefore, the number of downsampling modules directly affects the enhancement results. Based on our experimental results, we set the number of downsampling modules to five, which provides the best enhancement capability for underwater images. The number of upsampling modules in the decoder network matches that of the downsampling modules, we also set the upsampling modules to five.

In the decoder network, we use five proposed upsampling modules to gradually recover the feature map size and reconstruct the image content. Each upsampling module uses sub-pixel convolution to restore the size of the output feature maps to twice the size of the input feature maps. The number of channels of the output feature maps for the five proposed upsampling

modules is 256, 128, 64, 32, and 16, respectively. Finally, the feature maps are reconstructed as an enhanced image using a convolutional layer and a Tanh activation function. To reduce the negative impact of feature loss during downsampling on the reconstruction of image content during upsampling, a skip connection is used between each layer of downsampling and the corresponding upsampling, and an attention module with edge extraction is cascaded over the skip connection. This attention module improves the network's ability to enhance edge details.

3.2.2 Proposed multi-scale feature module

Extracting multi-scale features can enhance the capability of the network to perceive details and overall structure across various scales. Consequently, this enables the network to generate underwater images with higher quality. Therefore, we proposed a multi-scale feature module to extract multi-scale information in images. The structure of the multi-scale feature module is shown in Figure 4. It consists of residual mapping and identity mapping. The residual mapping consists of four parts. The first part consists of a 3×3 convolution and LeakyReLU function to extract features initially. The second part consists of two parallel branches to extract multi-scale features. The first branch contains a 3×3 convolution, an instance normalization, and a LeakyReLU function. The second branch contains a 5×5 convolution, an instance normalization, and a LeakyReLU function. The element-wise sum operation fuses the output features of the two branches. The fused features are used as the input feature of the third part. The third part also consists of two parallel branches to extract multi-scale features further. Each branch is a residual structure. The residual mappings are also 3×3 convolution and 5×5 convolution, respectively. An instance normalization and a LeakyReLU function follow each convolution. The concatenation operation fuses the output features of two branches. The fused feature is used as the input



feature of the final part. Finally, a 1×1 convolutional layer with a LeakyReLU activation function adjusts the number of feature map channels. This adjustment is made to facilitate the summation of subsequent elements. Additionally, residual connections are utilized throughout the module to facilitate efficient propagation of gradients in deep networks. This approach helps prevent gradient vanishing during training and enhances the stability of the network. In the proposed multi-scale feature extraction module, we primarily utilize two parallel convolutions: a 3×3 convolution and a 5×5 convolution, to extract multi-scale features. The 3×3 convolution has a smaller receptive field, while the 5×5 convolution has a larger receptive field. The smaller receptive field can capture local detail features in underwater images, such as the edges and textures of underwater objects. In contrast, the larger receptive field can capture features over a broader range, such as the overall shape and structure of underwater objects. Therefore, compared to a single-scale feature extraction module, the multi-scale feature extraction module employs convolutions with different receptive fields to extract both more local and global features of underwater images. This approach allows for the acquisition of high-quality underwater images that contain richer detail features.

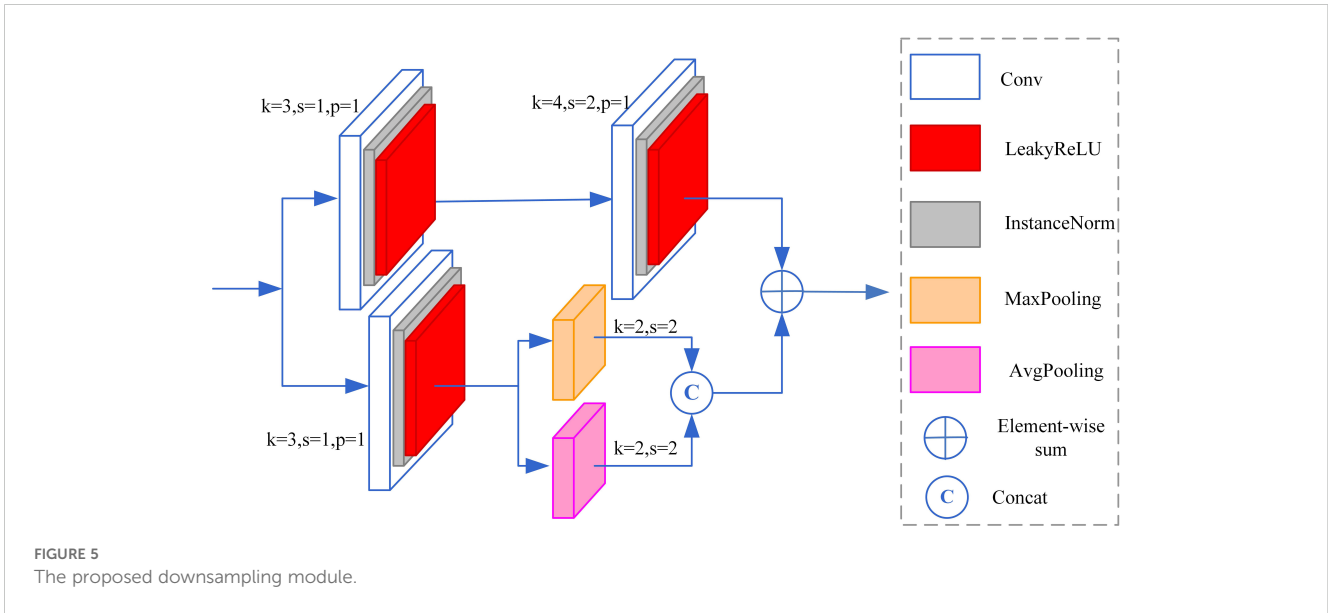
3.2.3 Proposed downsampling module

A downsampling module has been proposed to mitigate feature loss from downsampling operations and enhance the network's capability to extract diverse features. The structure of the module is shown in Figure 5. The module consists of two branches. The bottom branch consists of a basic convolution module and two pooling sub-branches. The basic convolution module contains a 3×3 convolution, instance normalization, and a LeakyReLU function. The feature maps extracted through average pooling are more sensitive to background information, whereas those obtained through max pooling are more sensitive to texture detail

information. The two sub-branches halve the feature map size using average pooling and maximum pooling, respectively. Subsequently, the feature maps are fused by channel-wise concatenation, aiding the network in recovering the image's color and details. In addition, the top branch uses two basic convolution blocks. The first convolution block increases the number of output feature map channels to twice that of input feature map channels. The second convolutional block reduces the output feature map size to half of the input feature maps. The final output feature maps of the two branches are fused by element-wise summation to compensate for features lost during downsampling. In the proposed downsampling module, employing convolution for downsampling enhances the extraction of more abstract features. The max pooling operation selects the maximum value within a local region as the output, preserving the most prominent and important features in the image. Therefore, downsampling with max pooling can capture local information in underwater images, effectively retaining textural features. In contrast, average pooling obtains the overall features of a local region by calculating the average value, allowing it to capture global information and highlight background details. By combining convolution, max pooling, and average pooling, we can harness their complementary advantages to extract richer feature information from underwater images during the downsampling process.

3.2.4 Proposed upsampling module

The upsampling operation in a convolutional neural network is usually realized by deconvolution. Deconvolution achieves upsampling by interpolating 0 between pixel values of the feature maps with learnable parameters. However, the more significant number of deconvolution parameters increases the model complexity, and the image reconstructed by the deconvolution may produce a checkerboard effect. Sub-pixel convolution is a



learnable upsampling method with fewer parameters than deconvolution. Instead of upsampling by interpolating 0 operations, this method utilizes the features of different channels to reconstruct the relationship between pixels and increase the size of the feature maps. As a result, sub-pixel convolution can better preserve the structure and details of the input feature maps, thereby assisting in producing a clearer output image (Shi et al., 2016). In sub-pixel convolution, a series of convolution operations are first applied to generate a feature map of size $H \times W$ with r^2 channels (where H and W are the size of the input feature map, and r is the upscaling factor). Next, the PixelShuffle operation is used to transform this into a feature map of size $(H \times r) \times (W \times r)$. Specifically, a low-resolution pixel is divided into $r \times r$ smaller grids and the values from the corresponding positions in the $r \times r$ feature maps are used to fill these small grids according to a set rule. By filling each of the smaller grids created from low-resolution pixels in the same manner, the upsampling process is completed. The core idea of sub-pixel convolution is to obtain a large number of feature map channels through the network's convolution layers, and then rearrange these channels to achieve a predetermined image size. Since sub-pixel convolution can capture global

features, the network can utilize more contextual information to restore more realistic details. This effectively enhances the representation of underwater image details, thereby improving the quality of the enhanced underwater images. Based on the above analysis, we designed an upsampling module based on sub-pixel convolution. The structure is shown in Figure 6. The module begins with a basic convolution block that adjusts the number of channels of the input feature maps to 4 times the target number of channels. The basic convolution module contains a 3×3 convolution, instance normalization, and a LeakyReLU function. Sub-pixel convolution then enlarges the size of the output feature maps to twice the size of the input feature maps while simultaneously reducing the number of channels of the feature maps to the target number. Finally, a residual block consisting of two basic convolution blocks is employed to further process the upsampled feature map and extract more features.

3.2.5 Proposed attention module

We designed an attention module to encourage the network to focus more on features conducive to underwater image enhancement and effectively enhance the quality of the output

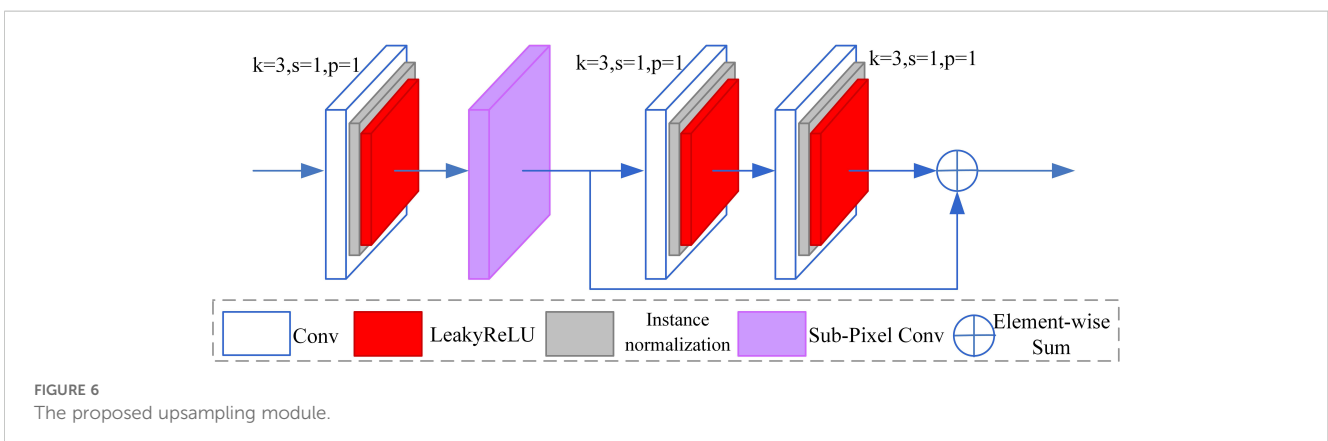


image. The structure of the module is shown in Figure 7. The module is divided into channel attention and spatial attention.

Channel attention enhances essential features of the channel dimension. In this section, we employ global maximum pooling and global average pooling on both branches to extract detailed texture and background features. We then utilize element-wise summation for feature fusion. Afterward, the number of channels is adjusted by a 1×1 convolutional layer, a LeakyReLU activation function, and another 1×1 convolutional layer. This step enhances the correlation between channels. Finally, the attention map is generated using a Sigmoid activation function. The attention map size is 1×1, which needs to be weighted using the element-wise multiplication operation after expanding it to the size of the input feature map. The weighted feature map continues to be used as the input for spatial attention. The output feature maps of channel attention can be represented as follows:

$$f_1(x) = x \cdot \sigma(\text{conv}_2(\text{LReLU}(\text{conv}_1(\text{avg}(x) \oplus \text{max}(x)))))) \quad (5)$$

where x is the input feature map of channel attention, σ is the Sigmoid activation function, conv is a 1×1 convolutional layer, LReLU is the LeakyReLU activation function, avg is global average pooling, max is global maximum pooling.

Spatial attention enhances important features of spatial dimensions. The spatial attention part of this module introduces edge extraction to improve the module’s ability to enhance edge details. In this part, we employ a 1×1 convolutional layer to reduce the number of feature map channels to 1. Subsequently, it is divided into two branches. The lower branch utilizes a convolutional layer to perform spatial feature extraction, while the upper branch employs a convolutional layer based on the Laplacian operator to extract image edge features. The outputs of the two branches are fused by element-wise summation, and the attention map is generated using the Sigmoid activation function. The number of channels in the attention map is 1, which needs to be expanded to the size of the input feature map and then weighted using the element-wise multiplication operation. The output feature maps of spatial attention can be represented as follows:

$$f_2(x) = x \cdot \sigma(\text{conv}_2(\text{conv}_1(x)) \oplus \text{Lconv}(\text{conv}_1(x))) \quad (6)$$

Where x is the input feature map of spatial attention, Lconv represents convolutional layers using the Laplacian operator. In the attention module, the channel attention and the spatial attention are cascaded to enhance essential features in the channel and spatial dimensions separately.

3.3 Adversarial network

The adversarial network is a crucial component of the generative adversarial network. It enhances the performance of the generative network and plays a significant role in underwater image enhancement. The structure of the adversarial network we used is depicted in Figure 8. This adversarial network is based on a Markovian discriminator and contains four proposed downsampling layers. First, features are extracted using a convolutional layer followed by a LeakyReLU activation function. Then multi-scale features are extracted using a designed multi-scale feature module shown in Figure 4 to enhance the multi-scale discriminative capability of the network. After that, four downsampling operations are performed. The downsampling operation is implemented using a designed downsampling module shown in Figure 5 to reduce the feature map size and extract various kinds of feature information. A small number of downsampling modules can only extract shallow features while increasing their number allows for the extraction of deeper, more complex features. However, as the number of downsampling modules increases, the dimensions of the deep feature maps may become too small to extract effective features, leading to redundant features that negatively impact underwater image enhancement. Additionally, increased network complexity can result in issues such as vanishing or exploding gradients, making training more difficult. Based on our experimental results, the network performs optimally with four downsampling modules in the adversarial network. Therefore, we set the number of downsampling modules in the adversarial network to four. Finally, a 1×1 convolutional layer is used to reduce the number of feature channels to 1 to generate a judgment score. Based on this score, the network determines

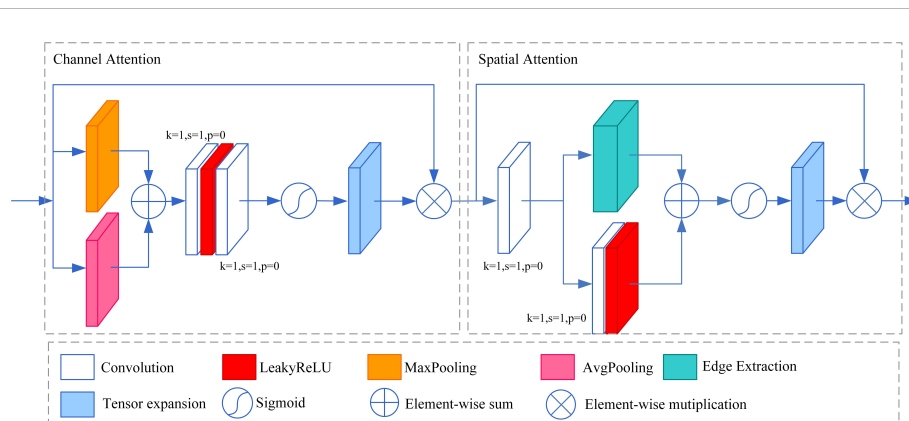
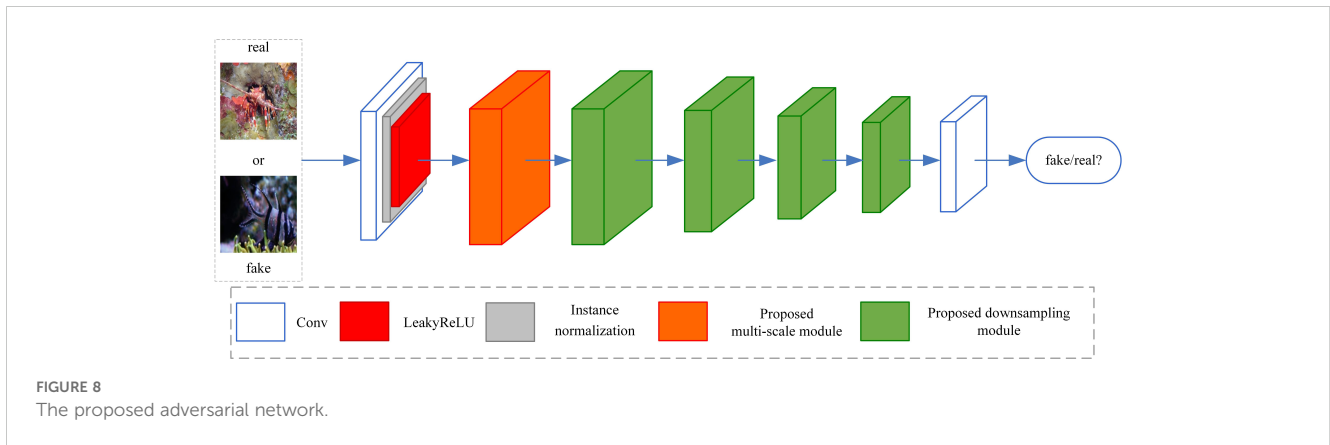


FIGURE 7 The proposed attention module.



whether the input is an original high-quality or enhanced image generated by the generative network.

3.4 Loss function

The loss function used in our proposed network consists of four parts: adversarial loss, cycle consistency loss, Identity loss, and perception loss. The total loss function can be expressed as follows:

$$L = L_{GAN}(G_1, D_Y, X, Y) + L_{GAN}(G_2, D_X, X, Y) + L_{cyc} + L_{idt} + L_{per} \quad (7)$$

where $L_{GAN}(G_1, D_Y, X, Y)$ and $L_{GAN}(G_2, D_X, X, Y)$ are adversarial loss, L_{cyc} is cycle consistency loss, L_{idt} is identity loss, L_{per} is perception loss. Cycle consistency loss is employed to ensure that the input image can be reconstructed to the original image after undergoing two transformations. This aids the network in learning the image transformation process effectively. Identity loss ensures that when an image from the target domain is fed into the network, the output should be identical to the input. This helps the network maintain the colors and details in the image and prevents unauthorized changes to the content of the image. Perceptual loss encourages the network to capture more complex features, leading to the generation of more realistic and natural images. $L_{GAN}()$ is the adversarial loss function, which can be expressed as follows:

$$L_{GAN}(G_1, D_Y, X, Y) = E_Y(\|D_Y(Y) - 1\|_2) + E_X(\|D_Y(G_1(X))\|_2) \quad (8)$$

$$L_{GAN}(G_2, D_X, X, Y) = E_X(\|D_X(X) - 1\|_2) + E_Y(\|D_X(G_2(Y))\|_2) \quad (9)$$

where X is degraded images, Y is high quality images, G_1 is used to enhance underwater images, G_2 is used to generated the degraded images, D_Y is used to determine whether the input image is an original high quality image or an enhanced image, D_X is used to determine whether the input image is an original degraded image or a generated degraded image.

The L_{cyc} can be expressed as follows:

$$L_{cyc} = E[\|G_2(G_1(X)) - X\|_1] + E[\|G_1(G_2(Y)) - Y\|_1] \quad (10)$$

where $G_2(G_1(X))$ is the generated degraded image. $G_1(G_2(Y))$ is the generated high quality image. L_{cyc} is used to calculate the difference between the original image and corresponding generated image.

The L_{idt} can be expressed as follows:

$$L_{idt} = E[\|G_1(Y) - Y\|_1] + E[\|G_2(X) - X\|_1] \quad (11)$$

It is used to calculate L1 loss between the input image and the output image of the generative network.

The L_{per} can be expressed as follows:

$$L_{per} = E[\|\phi(G_1(X)) - \phi(X)\|_2] + E[\|\phi(G_2(Y)) - \phi(Y)\|_2] \quad (12)$$

where $\phi(\cdot)$ are the high-level features extracted by a pre-trained VGG-19 network.

4 Simulation and discussion

In this section, we train our network using unpaired datasets and evaluate its effectiveness on synthetic images and real images. We compare our method with four underwater image enhancement methods: the CycleGAN method (Zhu et al., 2017), the FUNIEGAN method (Islam et al., 2022), the TAEL method (Liu et al., 2022), and the UW-CycleGAN (Yan et al., 2023) method, respectively. We use peak signal-to-noise ratio (PSNR), structural similarity (Wang et al., 2004) (SSIM), and underwater image quality measures (Panetta et al., 2015) (UIQM) to evaluate the enhancement effects of each method quantitatively. Finally, we performed an ablation study to validate the effectiveness of the modules of the method.

4.1 Dataset and metrics

We utilize Enhancing Underwater Visual Perception (EUVP) datasets (Islam et al., 2022) and the Underwater Image Enhancement Benchmark (UIEB) (Li et al., 2019), commonly used in existing underwater image enhancement methods (Yan et al., 2023; Zhou et al., 2023a; Zhang et al., 2024b), as the training and testing datasets for various methods in this paper. The EUVP dataset contains Paired dataset, Unpaired dataset, and Test_samples dataset. The Paired dataset

contains 11335 paired training images. The Unpaired dataset contains 6335 unpaired training images, of which 3195 are poor perceptual quality images and 3140 are good perceptual quality images. The Test_samples dataset contains 613 paired testing images. The images in EUVP are captured using seven different cameras over various visibility conditions during oceanic explorations and human-robot collaborative experiments in different locations under various visibility conditions. Additionally, images extracted from a few publicly available YouTube videos are included in the dataset. The images are carefully selected to accommodate a wide range of natural variability (e.g., scenes, waterbody types, lighting conditions, etc.) in the data (Islam et al., 2022). The UIEB dataset comprises 950 real-world underwater images collected from various scenes, each with a size of 256×256 pixels. The 400 underwater degraded images and 400 high-quality images, a total of 800 images constitute the training set named train800. The remaining 150 degraded images are used as the testing set named test150. These underwater images in the UIEB are taken under natural light, artificial light, or a mixture of natural light and artificial light. The underwater images in the UIEB have diverse color ranges and degrees of contrast decrease (Li et al., 2019). Both datasets are commonly used for training and testing underwater image enhancement algorithms and include underwater images from different scenes and lighting conditions in regular environments, making them quite representative. Therefore, this paper selects these two datasets as the training and testing datasets for the all methods.

All methods are implemented using the PyTorch framework with the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay = 0.0001). The batch size is 8 and the number of epochs is 100 for two training datasets. Firstly, we complete 100 epochs of training on the Unpaired dataset which is one part of the EUVP dataset with a learning rate of 0.0002. Secondly, we continue to train the network for another 100 epochs on the train800 which is one part of the UIEB dataset with a learning rate of 0.0002. The train800 consists of 400 underwater degraded images and 400 high-quality images.

The UIQM consists of three parts: the underwater image colorfulness measure (UICM), the underwater image sharpness measure (UISM), and the underwater image contrast measure (UIConM). It is expressed as follows:

$$UIQM = c_1 UICM + c_2 UISM + c_3 UIConM \quad (13)$$

where $c_1=0.0282$, $c_2=0.2953$, $c_3=3.5753$. The UIQM (Underwater Image Quality Measure) metric measures the quality of underwater images by assessing various factors such as color fidelity, contrast, brightness, and overall clarity. It is designed to evaluate the perceptual quality of images without requiring a reference image for comparison, making it particularly useful for underwater imaging scenarios where ideal reference images may not be available. The larger the UIQM value, the higher the overall visual quality of the enhanced image.

4.2 Underwater image enhancement on synthetic images

We evaluate the performance of our proposed method on the synthetic dataset using the Test_samples subset of the EUVP

dataset. We randomly selected three underwater degraded images and the corresponding reference images to show the enhancement results of the CycleGAN, the FUnIEGAN, the TACL, the UW-CycleGAN, and our method. The raw images (degraded underwater images), reference images (high quality images), and underwater images enhanced by different methods are shown in Figure 9. From left to right, each row of the figure shows the degraded underwater image, the high-quality reference image, and the images enhanced by the CycleGAN method, the FUnIEGAN method, the TACL method, the UW-CycleGAN method, and our method. Below each image, there is a partial enlargement with the enlarged area marked by a red box in the original image. In the first row, the CycleGAN method corrects the image color, but the enhanced image still suffers from blurred details. The FUnIEGAN method enhances the image contrast but fails to correct the color bias.

Additionally, the presence of noise and artifacts in the image adversely affects its quality. The TACL and UW-CycleGAN methods effectively enhance the image contrast, but the enhanced image suffers from overexposure issues. Our method enhances the image with natural colors and precise details; the visual effect is closest to the reference image. In the second row, The image enhanced by the CycleGAN method has blurred details and artifacts. The FUnIEGAN, TACL, and UW-CycleGAN methods enhanced images cannot correct the image color effectively, and the enhanced images still have color bias. In contrast, our method can effectively correct image color and sharpen image details. In the third row, The image enhanced by the CycleGAN method has blurred edge texture and artifacts. The image enhanced by the FUnIEGAN method still suffers from color bias and low quality. The TACL and UW-CycleGAN methods can improve the image contrast but cannot completely correct the image color bias. The images enhanced by our method have the most natural colors and sharpest details, and the overall quality is closest to the reference image.

The CycleGAN method can correct image color but performs poorly in image detail recovery and denoising. The FUnIEGAN method can improve the image contrast but cannot correct the image color effectively. The image details enhanced by the TACL method are better in quality, but there is still some color bias. The UW-CycleGAN method effectively enhances image contrast and sharpens image details, but the color of the enhanced image is not natural. In contrast, our method performs best in correcting image color and sharpening image details. Furthermore, The images enhanced by our method closely resemble the high-quality reference image.

To quantitatively test the enhancement performance of different methods, we calculated the PSNR, SSIM, and UIQM of the images enhanced by different methods using all the paired images in this test set. The average results are shown in Table 1. The average PSNR values for the CycleGAN method, the FUnIEGAN method, the TACL method, the UW-CycleGAN method, and our method are 21.98, 20.85, 24.65, 23.49 and 26.51, respectively. Our method had the highest PSNR value, followed by the TACL and UW-CycleGAN methods. The average SSIM values for the CycleGAN method, the FUnIEGAN method, the TACL method, the UW-CycleGAN



FIGURE 9 Visual comparison among the competing methods on the synthetic underwater images selected from the EUVP dataset. The degraded underwater images are listed in the first column, and the reference images are in the second column. Columns 3-6 are the enhanced images of comparison methods, and ours are in the last column.

method, and our method are 0.61, 0.59, 0.73, 0.68, and 0.76, respectively. Our method had the highest SSIM value, followed by the TACL and UW-CycleGAN methods. The average UIQM values for the CycleGAN method, the FUnIEGAN method, the TACL method, the UW-CycleGAN method, and our method are 2.94, 2.78, 3.07, 2.98, and 3.21, respectively. Our method had the highest UIQM value, followed by the TACL and UW-CycleGAN methods. In summary, our proposed method has the highest PSNR, SSIM, and UIQM on the test_samples subset of the EUVP dataset, which shows that the method can effectively remove the image noise, preserve the image content, and improve the visual quality of the image.

4.3 Underwater image enhancement on real images

We also evaluate the performance of our proposed method on the actual image dataset using the test150 test set. The test set contains 150 degraded underwater images taken from the UIEB dataset. We randomly selected three underwater degraded images to test the performance of the CycleGAN, the FUnIEGAN, the TACL, the UW-CycleGAN, and our method. The raw and underwater images enhanced by different methods are shown in Figure 10. From left to right, each row of the figure shows the degraded underwater image and the images enhanced by the CycleGAN method, the FUnIEGAN method, the TACL method, the

TABLE 1 Performances of underwater image enhancement methods for the synthetic images.

	CycleGAN	FUnIEGAN	TACL	UW-CycleGAN	Ours
PSNR	21.98	20.85	24.65	23.49	26.51
SSIM	0.61	0.59	0.73	0.68	0.76
UIQM	2.94	2.78	3.07	2.98	3.21

UW-CycleGAN method, and our method. Below each image, there is a partial enlargement with the enlarged area marked by a red box in the original image. In the first row, The image enhanced by the CycleGAN method exhibits blurrier texture details.

The FUnIEGAN method introduces artifacts in the image, severely impacting its quality. The TACL method enhances the image brightness, but the color correction is unsatisfactory. The UW-CycleGAN method improves the image contrast, but the color bias of the enhanced image is obvious. Our method effectively enhances image brightness and corrects image color. In the second row, the CycleGAN method fails to effectively correct the image color, and the enhanced image exhibits checkerboard artifacts. The image’s brightness and contrast after enhancement by the FUnIEGAN method are low. The TACL and UW-CycleGAN methods corrected the image color but failed to effectively improve the region’s brightness. In contrast, our method effectively enhances image brightness and corrects image color. In the third row, The

image enhanced by the CycleGAN method suffers from color bias and detail blurring. There are artifacts in the images enhanced by the FUnIEGAN method. The TACL method effectively corrects the image color, but the color saturation of the enhanced image was low. The image enhanced by the UW-CycleGAN method exhibits low brightness and blurred details. Our method effectively corrects image color and sharpens image details. The visual quality of the images obtained by our method is optimal.

In summary, the CycleGAN method is ineffective in correcting the image color, and the enhanced image still has a color bias. The images enhanced by the FUnIEGAN method have artifacts that severely degrade the quality of image details. The TACL method is not effective in improving image color saturation. The images enhanced by the UW-CycleGAN method still have a slight color bias and low brightness. In contrast, our proposed method can effectively correct image color, sharpen image edge details, and improve the visual quality of authentic underwater degraded images.

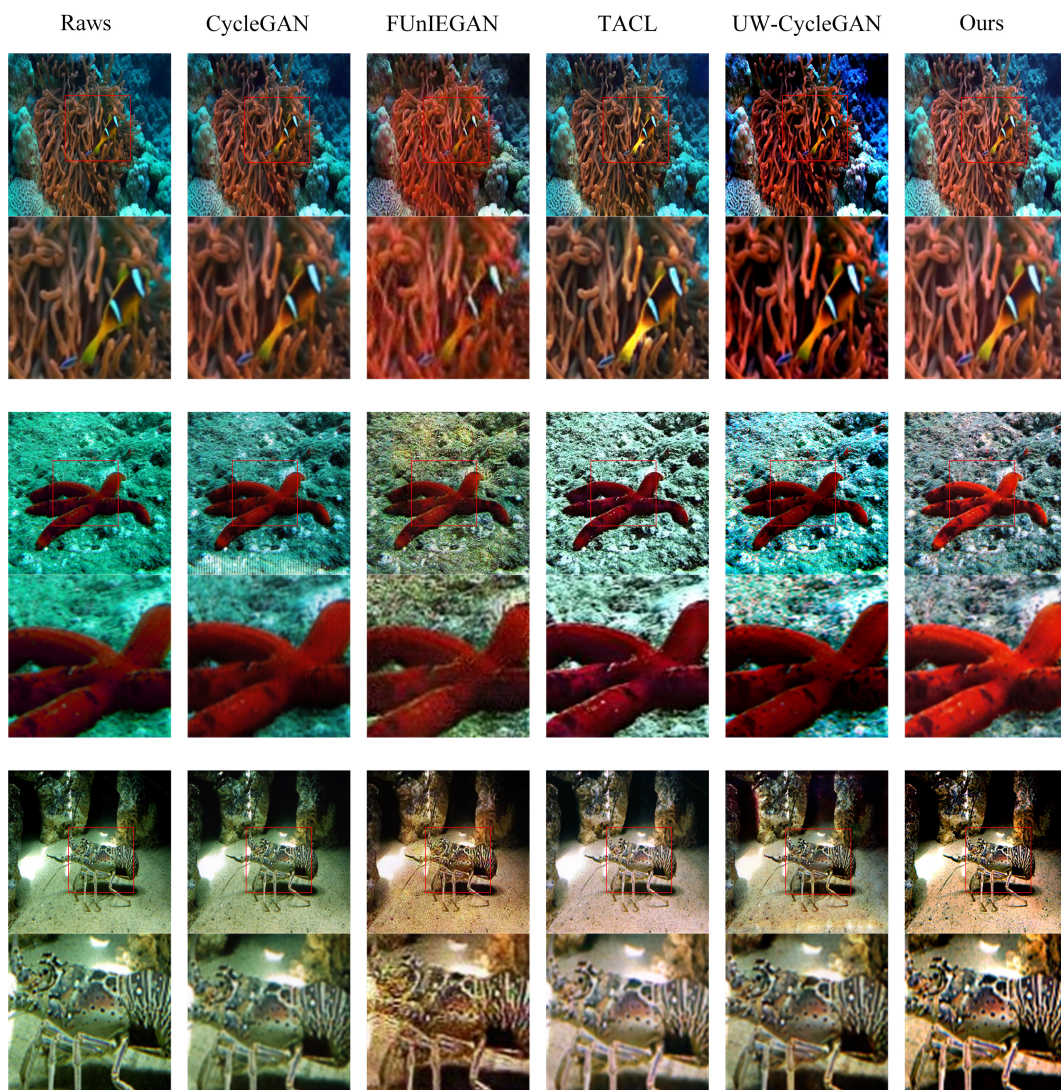


FIGURE 10 Visual comparison among the competing methods on the real underwater images selected from the UIEB dataset. The degraded underwater images are listed in the first column. Columns 2-5 are the enhanced images of comparison methods, and ours are in the last column.

To quantitatively test the enhancement performance of different methods, we calculated the UICM, UISM, UIConM, and UIQM of the images enhanced by different methods using all images in test150. The average results are shown in Table 2. The average UICM values for the CycleGAN method, the FUnIEGAN method, the TACL method, the UW-CycleGAN method, and our method are 4.95, 6.15, 6.31, 6.36, and 6.46, respectively. Our method had the highest UICM value, followed by the TACL and UW-CycleGAN methods. The average UICM values for the CycleGAN method, the FUnIEGAN method, the TACL method, the UW-CycleGAN method, and our method are 6.86, 5.78, 7.02, 6.65, and 7.34, respectively. Our method had the highest UICM value, followed by the TACL and CycleGAN methods. The average UIConM values for the CycleGAN, FUnIEGAN, TACL, and UW-CycleGAN and our method are 0.20, 0.22, 0.22, 0.21, and 0.23, respectively. Our method had the highest UIConM value, followed by the TACL and FUnIEGAN methods. The average UIQM values for the CycleGAN method, the FUnIEGAN method, the TACL method, the UW-CycleGAN method, and our method are 2.88, 2.68, 3.03, 2.89, and 3.16, respectively. Our method had the highest UIQM value, followed by the TACL and UW-CycleGAN methods. In summary, our proposed method obtained the highest UICM, UISM, and UIConM, representing that the method performs better than other comparison methods in terms of color correction, contrast enhancement, and sharpness enhancement of images. In addition, our method also obtained the highest UIQM, confirming the advanced performance of the method in improving the overall visual quality of underwater images.

To better analyze the complexity of underwater image enhancement methods and their processing speed, we measured the algorithms' FLOPs (Floating Point Operations), parameters, and FPS. The results are presented in Table 3. From Table 3, it is evident that the FLOPs and parameters of our proposed method are lower than those of other methods, except for the FUnIEGAN method. Additionally, our method achieves a higher underwater image enhancement speed compared to other methods, again except for FUnIEGAN. Although the proposed method is relatively slower in enhancing underwater images compared to the FUnIEGAN method, the quality of the enhanced underwater images obtained from our method is better.

4.4 Ablation study

To analyze the contribution of the edge extraction convolution, the multi-scale feature module, the downsampling Module, the

upsampling Module, the attention module, and the perceptual loss, we conducted the following ablation studies:

- w/o-EEC: without the edge extraction convolution;
- w/o-MFM: without the multi-scale feature module;
- w/o-DSM: without the downsampling module;
- w/o-USM: without the upsampling module;
- w/o-AM: without the attention module;
- w/o-Per Loss: without the perceptual loss;

We removed the edge extraction convolution module, attention module, and perceptual loss from our proposed complete method to obtain the w/o-EEC method, w/o-AM, and w/o-Per Loss, respectively. We use the 3×3 convolution instead of the multi-scale feature module to obtain the w/o-MFM method. We replace the downsampling and upsampling modules with convolutions and deconvolutions that have a kernel size of 4, a stride of 2, and a padding of 1, resulting in the w/o-DSM method and w/o-USM method, respectively. The parameters of the multi-scale feature module, downsampling module, upsampling module, and attention module are shown in Figures 4–7, respectively.

The PSNR, SSIM, and UIQM scores on the test_samples subset from the EUVP dataset are shown in Table 4. As can be seen from the table, the PSNR, SSIM, and UIQM obtained by our complete model are higher than those obtained by all the ablated models. This result validates the effectiveness of edge extraction convolution, multi-scale feature module, downsampling Module, upsampling Module, attention module, and perceptual loss.

4.4.1 Ablation study on EEC

Edge extraction convolution improves the detail sharpness of the enhanced image by extracting the image edge details. From Table 4, compared with our method without edge extraction convolution, our complete method improved the PSNR, SSIM, and UIQM scores by nearly 2.9%, 2.6%, and 3.1%, respectively. It is shown that the edge extraction convolution used in our method improves the detail quality of the images enhanced by the generative network.

4.4.2 Ablation study on MFM

The multi-scale feature module in the generative network is designed to enhance the network's multi-scale feature extraction ability. From Table 4, compared with our method without multi-scale feature module, our complete method improved the PSNR, SSIM, and UIQM scores by nearly 16.0%, 9.2%, and 7.5%, respectively. It is shown that the multi-scale feature module used

TABLE 2 Performances of underwater image enhancement for different methods on test150.

	CycleGAN	FUnIEGAN	TACL	UW-CycleGAN	Ours
UICM	4.95	6.15	6.31	6.36	6.46
UISM	6.86	5.78	7.02	6.65	7.34
UIConM	0.20	0.22	0.22	0.21	0.23
UIQM	2.88	2.68	3.03	2.89	3.16

TABLE 3 Computational complexity and speed of underwater image enhancement methods.

	CycleGAN	FUnIEGAN	TACL	UW-CycleGAN	Ours
FLOPs(G)	63.57	13.16	87.03	50.07	20.04
Parameters(M)	19.56	10.69	21.29	24.4	18.34
FPS	2.52	21.74	2.17	2.66	13.69

in our method enhances the ability of the network to extract multi-scale features, which is positive for color correction and detail enhancement of underwater images.

4.4.3 Ablation study on DSM

The downsampling module in the generative network is designed to extract different types of feature information and reduce feature loss, thus improving the quality of the enhanced images. From Table 4, compared with our method without the downsampling module, our complete method improved the PSNR, SSIM, and UIQM scores by nearly 9.1%, 3.9%, and 4.4%, respectively. It is shown that the downsampling module used in our method reduces noise and improves the detail quality of the enhanced image.

4.4.4 Ablation study on USM

The upsampling module in the generative network uses sub-pixel convolution to recover the image size, aiming to improve the detail quality of the reconstructed image by the network. From Table 4, compared with our method without the upsampling module, our complete method improved the PSNR, SSIM, and UIQM scores by nearly 7.0%, 3.9%, and 5.3%, respectively. It is shown that the upsampling module used in our method can keep more detailed features while recovering the image size and improving the image's detail quality.

4.4.5 Ablation study on AM

The attention module in the generative network is designed to enhance the network's focus on essential features that facilitate underwater image enhancement as well as edge features. From Table 4, compared with our method without the attention module, our complete method improved the PSNR, SSIM, and UIQM scores by nearly 5.5%, 2.6%, and 4.0%, respectively. It is shown that the attention module used in our method enhances the network's focus on details and improves the visual effect of the enhanced image.

4.4.6 Ablation study on per loss

The perceptual loss is added to the total loss function to improve the enhanced image's content quality. From Table 4,

compared with our method without perceptual loss, our complete method improved the PSNR, SSIM, and UIQM scores by nearly 2.0%, 5.3%, and 9.3%, respectively. It is shown that the perceptual loss prompts the network to generate images with greater clarity, detail, and texture visual quality.

5 Conclusion

This paper proposes a new unsupervised generative adversarial network to enhance underwater images. It can be trained without using the paired images. It proposes an edge extraction block based on the Laplacian operator, an attention module with the edge extraction block, a multi-scale feature module, a novel upsampling module, and a new downsampling module. It constructs the generative network and adversarial network using these proposed modules. Besides, it proposes an improved loss function by introducing the perceptual loss function into the conventional loss function to better the image generation capability of two generative networks. Compared to state-of-the-art methods, the proposed method achieves the highest PSNR, SSIM, and UIQM values on the synthetic underwater images dataset. The proposed method improves by 7.5%, 4.1%, and 4.6% over the second-best method in terms of PSNR, SSIM, and UIQM, respectively. The proposed method achieves the highest UICM, UISM, UICoM, and UIQM values on the authentic underwater images dataset. The proposed method improves by 1.6%, 4.6%, 4.5%, and 4.3% over the second-best method in terms of UICM, UISM, UICoM, and UIQM respectively. These prove that the proposed method effectively restores underwater image color, contrast, and detail information more than others. In addition, we compare the FLOPs, parameters, and image enhancement speeds of different methods. The FLOPs and parameters of our method are only higher than those of the FUnIEGAN method. The image enhancement speed is only slower than that of the FUnIEGAN method. An ablation study incorporating several proposed modules is conducted. Compared with our method without edge extraction convolution, our complete method improve the PSNR, SSIM, and UIQM scores by nearly 2.9%, 2.6%, and 3.1%, respectively. Compared with our method without multi-scale feature module, our complete method improve the PSNR, SSIM, and UIQM

TABLE 4 Experimental results of the ablation study.

	w/o- EEC	w/o- MFM	w/o- DSM	w/o- USM	w/o- AM	w/o- Per Loss	Ours
PSNR	25.74	22.26	24.11	24.65	25.05	25.98	26.51
SSIM	0.74	0.69	0.73	0.73	0.74	0.72	0.76
UIQM	3.11	2.97	3.07	3.04	3.08	2.91	3.21

scores by nearly 16.0%, 9.2%, and 7.5%, respectively. Compared with our method without the downsampling module, our complete method improve the PSNR, SSIM, and UIQM scores by nearly 9.1%, 3.9%, and 4.4%, respectively. Compared with our method without the upsampling module, our complete method improve the PSNR, SSIM, and UIQM scores by nearly 7.0%, 3.9%, and 5.3%, respectively. Compared with our method without the attention module, our complete method improve the PSNR, SSIM, and UIQM scores by nearly 5.5%, 2.6%, and 4.0%, respectively. Compared with our method without perceptual loss, our complete method improve the PSNR, SSIM, and UIQM scores by nearly 2.0%, 5.3%, and 9.3%, respectively. These results validated the effectiveness of each module.

This paper focuses on underwater image enhancement in conventional environments. It does not address special conditions, such as deep water with low light or non-uniform lighting. In future research, we will collect more underwater images from unconventional environments. We aim to design a separate illumination feature extraction module within the network architecture. Additionally, we will incorporate available prior information to improve the network's ability to enhance underwater images under these conditions, thereby expanding the algorithm's applicability.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Author contributions

YJ: Funding acquisition, Writing – original draft, Writing – review & editing, Conceptualization. ZW: Formal Analysis,

Software, Writing – original draft, Conceptualization, Data curation. LZ: Data curation, Funding acquisition, Methodology, 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 by the Foundation for Natural Science Foundation of Jilin Province under Grants (NO. 61271115), Jilin Provincial Department of Education Science and Technology Research Project (No. JJKH20240085KJ).

Conflict of interest

ZW is employed by Beijing Zhongdian Feihua Communication Company Limited, Beijing, China.

The remaining 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.

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., Xu, L., Deng, Z., and Zhang, H. (2024). HFM: A hybrid fusion method for underwater image enhancement. *Eng. Appl. Artif. Intell.* 127, 107219. doi: 10.1016/j.engappai.2023.107219
- Cai, L., McGuire, N. E., Hanlon, R., Mooney, T. A., and Gridhar, Y. (2023a). Semi-supervised visual tracking of marine animals using autonomous underwater vehicles. *Int. J. Comput. Vision* 131, 1406–1427. doi: 10.1007/s11263-023-01762-5
- Cai, X., Jiang, N., Chen, W., Hu, J., and Zhao, T. (2023b). Cure-net: A cascaded deep network for underwater image enhancement. *IEEE J. Oceanic Eng.* 49, 226–236. doi: 10.1109/JOE.2023.3245760
- Chen, Z., He, Z., and Lu, Z. M. (2024). DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention. *IEEE Trans. Image Process.* 33, 1002–1015. doi: 10.1109/TIP.2024.3354108
- Cong, R., Yang, W., Zhang, W., Li, C., Guo, C., Huang, Q., et al. (2023). Pugan: Physical model-guided underwater image enhancement using gan with dual-discriminators. *IEEE Trans. Image Process.* 32, 4472–4485. doi: 10.1109/TIP.2023.3286263
- Ding, X., Wang, Y., Liang, Z., and Fu, X. (2022). A unified total variation method for underwater image enhancement. *Knowledge-Based Syst.* 255, 109751. doi: 10.1016/j.knsys.2022.109751
- Ding, B., Zhang, R., Xu, L., Liu, G., Yang, S., Liu, Y., et al. (2023). U2D2Net: Unsupervised unified image dehazing and denoising network for single hazy image enhancement. *IEEE Trans. Multimedia* 26, 202–217. doi: 10.1109/TMM.2023.3263078
- Guo, 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
- Han, G., Wang, M., Zhu, H., and Lin, C. (2023). UIEGAN: adversarial learning-based photo-realistic image enhancement for intelligent underwater environment perception. *IEEE Trans. Geosci. Remote Sens.* 61, 5611514. doi: 10.1109/TGRS.2023.3281741
- Hua, X., Cui, X., Xu, X., Qiu, S., Liang, Y., Bao, X., et al. (2023). Underwater object detection algorithm based on feature enhancement and progressive dynamic aggregation strategy. *Pattern Recognition* 139, 109511. doi: 10.1016/j.patcog.2023.109511
- Islam, M. J., Xia, Y., and Sattar, J. (2022). Fast underwater image enhancement for improved visual perception. *IEEE Robotics Automation Lett.* 5, 3227–3234. doi: 10.1109/LRA.2020.2974710
- Jiang, Q., Kang, Y., Wang, Z., Ren, W., and Li, C. (2024). Perception-driven deep underwater image enhancement without paired supervision. *IEEE Trans. Multimedia* 26, 4884–4897. doi: 10.1109/TMM.2023.3327613
- Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., et al. (2019). An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.* 29, 4376–4389. doi: 10.1109/TIP.2019.2955241
- Li, Y., Wang, B., Li, Y., Liu, Z., Huo, W., Li, Y., et al. (2023). Underwater object tracker: UOSTrack for marine organism grasping of underwater vehicles. *Ocean Eng.* 285, 115449. doi: 10.1016/j.oceaneng.2023.115449

- Liu, R., Jiang, Z., Yang, S., Yang, S., and Fan, X. (2022). Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Trans. Image Process.* 31, 4922–4936. doi: 10.1109/tip.2022.3190209/2022twinn
- Liu, C., Shu, X., Pan, L., Shi, J., and Han, B. (2023a). Multi-scale underwater image enhancement in RGB and HSV color spaces. *IEEE Trans. Instrumentation Measurement* 72, 1–14. doi: 10.1109/TIM.2023.3298395
- Liu, Q., Zhang, Q., Chen, W., Liu, X., and Wang, X. (2023b). WSDS-GAN: A weak-strong dual supervised learning method for underwater image enhancement. *Pattern Recognition* 143, 109774. doi: 10.1016/j.patcog.2023.109774
- Mishra, A. K., Choudhry, M. S., and Kumar, M. (2023). Underwater image enhancement using multi-scale decomposition and gamma correction. *Multimedia Tools Appl.* 82, 15715–15733. doi: 10.1007/s11042-022-14008-2
- Muniraj, M., and Dhandapani, V. (2023). Underwater image enhancement by modified color correction and adaptive look-up-table with edge-preserving filter. *Signal Processing: Image Communication* 113, 116939. doi: 10.1016/j.image.2023.116939
- 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
- Panetta, K., Kezebou, L., Oludare, V., and Agaian, S. (2022). Comprehensive underwater object tracking benchmark dataset and underwater image enhancement with GAN. *IEEE J. Oceanic Eng.* 47, 59–75. doi: 10.1109/JOE.2021.3086907
- Peng, L., Zhu, C., and Bian, L. (2023). U-shape transformer for underwater image enhancement. *IEEE Trans. Image Process.* 32, 3062–3079. doi: 10.1109/TIP.2023.3276332
- Shen, Z., Xu, H., Luo, T., Song, Y., and He, Z. (2023). UDAformer: underwater image enhancement based on dual attention transformer. *Comput. Graphics* 111, 77–88. doi: 10.1016/j.cag.2023.01.009
- Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., et al. (2016). “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Institute of Electrical and Electronics Engineers, 1874–1883. doi: 10.1109/CVPR.2016.207
- Ulutas, G., and Ustubioglu, B. (2021). Underwater image enhancement using contrast limited adaptive histogram equalization and layered difference representation. *Multimedia Tools Appl.* 80, 15067–15091. doi: 10.1007/s11042-020-10426-2
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612. doi: 10.1109/TIP.2003.819861
- Wang, Z., Wu, R., Xu, Y., Liu, Y., Chai, R., and Ma, H. (2023). A two-stage CNN method for MRI image segmentation of prostate with lesion. *Biomed. Signal Process. Control* 82, 104610. doi: 10.1016/j.bspc.2023.104610
- Wang, H., Yang, M., Yin, G., and Dong, J. (2024). Self-adversarial generative adversarial network for underwater image enhancement. *IEEE J. Oceanic Eng.* 49, 237–248. doi: 10.1109/JOE.2023.3297731
- Xiang, D., Wang, H., He, D., and Zhai, C. (2023). Research on histogram equalization algorithm based on optimized adaptive quadruple segmentation and cropping of underwater image. *IEEE Access* 11, 69356–69365. doi: 10.1109/ACCESS.2023.3290201
- Xu, S., Zhang, M., Song, W., Mei, H., He, Q., and Liotta, A. (2023). A systematic review and analysis of deep learning-based underwater object detection. *Neurocomputing* 527, 204–232. doi: 10.1016/j.neucom.2023.01.056
- Yan, H., Zhang, Z., Xu, J., Wang, T., An, P., Wang, A., et al. (2023). UW-cycleGAN: model-driven cycleGAN for underwater image restoration. *IEEE Trans. Geosci. Remote Sens.* 61, 1–17. doi: 10.1109/TGRS.2023.3315772
- 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, T., Wang, D., and Lu, Y. (2023). ECSNet: An accelerated real-time image segmentation CNN architecture for pavement crack detection. *IEEE Trans. Intelligent Transportation Syst.* 24, 15105–15112. doi: 10.1109/TITS.2023.3300312
- Zhang, W., Wang, H., Ren, P., and Zhang, W. (2023b). Underwater image color correction via color channel transfer. *IEEE Geosci. Remote Sens. Lett.* 21, 1–5. doi: 10.1109/LGRS.2023.3344630
- Zhang, S., Wang, R., Zheng, S., Wang, L., and Liu, Z. (2024b). A learnable full-frequency transformer dual generative adversarial network for underwater image enhancement. *Front. Mar. Science.* 11. doi: 10.3389/fmars.2024.1321549
- Zhang, D., Wu, C., Zhou, J., Zhang, W., Lin, Z., and Palat, K. (2024a). Robust underwater image enhancement with cascaded multi-level sub-networks and triple attention mechanism. *Neural Networks* 169, 685–697. doi: 10.1016/j.neunet.2023.11.008
- Zhang, W., Zhuang, P., Sun, H. H., Li, G., Kwong, S., and Li, C. (2022). 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
- Zheng, Z., Xin, Z., Yu, Z., and Yeung, S. K. (2023). Real-time GAN-based image enhancement for robust underwater monocular SLAM. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1161399
- Zhou, J., Liu, Q., Jiang, Q., Ren, W., Lam, K. M., and Zhang, W. (2023a). Underwater camera: Improving visual perception via adaptive dark pixel prior and color correction. *Int. J. Comput. Vision*, 1–19. doi: 10.1007/s11263-023-01853-3
- Zhou, J., Yang, T., Chu, W., and Zhang, W. (2022). Underwater image restoration via backscatter pixel prior and color compensation. *Eng. Appl. Artif. Intell.* 111, 104785. doi: 10.1016/j.engappai.2022.104785
- Zhou, J., Zhang, D., and Zhang, W. (2023b). Cross-view enhancement network for underwater images. *Eng. Appl. Artif. Intell.* 121, 105952. doi: 10.1016/j.engappai.2023.105952
- Zhu, J. Y., Park, T., Isola, P., and Efros, A. A. (2017). “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *IEEE International Conference on Computer Vision Venice*. Venice, Italy: Institute of Electrical and Electronics Engineers. 2223–2232. doi: 10.1109/ICCV.2017.244