Check for updates

Generative adversarial networks with multi-scale and attention mechanisms for underwater image enhancement

Ziyang Wang[1], Liquan Zhao[1], Tie Zhong[1]*, Yanfei Jia[2] and Ying Cui[3]

[1]Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education (Northeast Electric Power University), Jilin, China, [2]College of Electrical and Information Engineering, Beihua University, Jilin, China, [3]Communication Network Operations Team of System Operations Department, Zhuhai Power Supply Bureau, Guangdong Power Grid Co., Ltd., Zhuhai, China

The images captured underwater are usually degraded due to the effects of light absorption and scattering. Degraded underwater images exhibit color distortion, low contrast, and blurred details, which in turn reduce the accuracy of marine biological monitoring and underwater object detection. To address this issue, a generative adversarial network with multi-scale and an attention mechanism is proposed to improve the quality of underwater images. To extract more effective features within the generative network, several modules are introduced: a multi-scale dilated convolution module, a novel attention module, and a residual module. These modules are utilized to design a generative network with a U-shaped structure. The multi-scale dilated convolution module is designed to extract features at multiple scales and expand the receptive field to capture more global information. The attention module directs the network's focus towards important features, thereby reducing the interference from redundant feature information. To improve the discriminative power of the adversarial network, a multi-scale discriminator is designed. It has two output feature maps with different scales. Additionally, an improved loss function for the generative adversarial network is proposed. This improvement involves incorporating the total variation loss into the traditional loss function. The performance of different methods for enhancing underwater images is evaluated using the EUVP dataset and UIEB dataset. The experimental results demonstrate that the enhanced underwater images exhibit better quality and visual effects compared to other methods.

KEYWORDS

underwater image enhancement, generative adversarial network, image quality, image visual effect, deep learning

1 Introduction

In recent years, computer vision technology has played an important role in marine engineering fields such as ocean data collection (Gavrilov and Parnum, 2010; Zhang et al., 2022), deep ocean resources exploration, and ocean environmental protection (Bell et al., 2022; Townhill et al., 2022). The quality of underwater images directly influences the performance of computer vision technologies. However, the underwater images are usually not clear enough with color distortion, uneven illumination, and lower contrast, which are caused by forward scattering, backward scattering, and absorption of light in the water medium. Degraded underwater images directly reduce the performance of underwater object detection, underwater image segmentation, and underwater object tracking (Zou et al., 2021; Yu et al., 2023). Most underwater vision systems cannot guarantee satisfactory performance under poor water conditions. Therefore, it is necessary to enhance the underwater images to improve the performance of underwater computer vision.

The traditional methods of image enhancement are mainly based on methods such as histogram equalization (Huang et al., 2021; Ulutas and Ustubioglu, 2021), Gamma correction (Huang et al., 2016; Huang et al., 2018a) and Retinex-based methods (Huang et al., 2018b). Although these methods are simple, they are not fully suitable for underwater image enhancement. There are still significant distortions in underwater images enhanced by these methods. Besides, underwater image enhancement methods based on the degradation model have been proposed (Cui et al., 2022; Luo et al., 2022). These methods necessitate the construction of an appropriate degradation model for underwater images, followed by the restoration of the original, undegraded underwater image by simulating the degradation process. While they exhibit improved performance for certain underwater images, they lack robustness. The complexities and variations within underwater environments make constructing a suitable degradation model challenging. Additionally, adapting a static degradation model to different underwater environments has proven difficult. As a result, underwater image enhancement methods relying on degradation models often yield subpar results for images captured in diverse underwater settings.

With the development of deep convolutional networks, many underwater image enhancement methods based on deep learning have been proposed in the last few years (Yan et al., 2022; Zheng and Luo, 2022). They use a large number of degraded images and high-quality images to train the network model without constructing the degradation model. Therefore, it is more suitable for different underwater environments. A generative adversarial network can be seen as a special deep learning method. Compared with the conventional deep learning network, it consists of a generative network and an adversarial network. In underwater image enhancement, the generative network is used to enhance the image, and the adversarial network is used to determine whether the input high-quality image is the generated image by the generative network or the original high-quality image. The adversarial network aids the generative network in improving the performance of underwater image enhancement. Therefore,

compared with the underwater image enhancement methods based on conventional deep learning networks, the methods based on generative adversarial networks have better performance. Although many generative adversarial networks have been used for underwater image enhancement and in attempts to improve the quality of underwater images(Estrada et al., 2022; Xu et al., 2023), the enhanced underwater images still contain much color distortion and detail loss, which affect underwater object detection. To further improve the quality of underwater images, an efficient generative adversarial network is proposed.

The main contributions of this paper are summarized as follows:

- A generative adversarial network with a U-shaped network structure is proposed for enhancing the underwater images. The network incorporates a new multi-scale dilated convolution module, a novel attention module, and a residual module. The generative adversarial network is capable of extracting more effective features at different scales. As a result, the enhanced underwater images produced by the proposed generative adversarial network preserve more details and color information.

- A multi-scale adversarial network is also proposed to improve the discriminative power, which is helpful in improving the optimization of the generative network. It contains two judgment score matrices at different scales to determine whether the input image is the image generated by the generative network or the original high-quality image at multiple scales.

- An improved loss function of the generative adversarial network is also proposed to provide a more accurate measurement of the difference between the enhanced image and the high-quality image. This improvement involves incorporating the total variation loss into the traditional loss function. By considering the difference between adjacent image pixels, the total variation loss promotes image smoothness and enhances the overall quality of the generated image.

2 Related works

In recent years, a number of underwater image enhancement methods have been proposed. These methods can be broadly classified into three categories: model-free methods, model-based methods, and deep learning-based methods. The model-free methods do not require a degradation model for underwater images. They correct the color and improve the contrast of underwater images by directly adjusting the image pixel values (Zhang et al., 2021; Zhuang et al., 2021). While the model-free method can effectively correct image color, it struggles to address issues such as blurred details and noise interference. Furthermore, these methods typically involve multiple steps to process images, making practical applications challenging. In contrast, model-based

methods establish a degradation model for underwater images and estimate model parameters using prior knowledge. These methods obtain enhanced images by inversion of the model (Ding et al., 2022; Zhou et al., 2022). These model-based methods exhibit superior performance in terms of color correction. Nevertheless, a majority of these methods rely on scene-specific depth and illumination information as *a priori* knowledge to estimate the model parameters. This dependence on such information is less stable and cannot be easily adapted to varying underwater scenes.

Recently, deep learning technology has developed rapidly. It has been widely used in image dehazing (Liu W. et al., 2019; Wang et al., 2022), super-resolution (Srivastava et al., 2022; Tian et al., 2022), remotely sensed image enhancement (Huang et al., 2022), and underwater image enhancement (Li et al., 2022; Zhou et al., 2023). It uses a large number of degraded images and high-quality images to train the network model without constructing the degradation model. Therefore, it is more suitable for different underwater environments. Compared with traditional deep learning with a single network, generative adversarial networks consist of two networks, the generative network and the adversarial network. The adversarial network can indirectly improve the performance of the generative network. Due to the advantages of a generative adversarial network, it is widely used in underwater image enhancement. Fabbri et al. proposed UGAN (Fabbri et al., 2018). They used CycleGAN to generate paired underwater image datasets for the following network training. Meanwhile, they added L1 loss and gradient loss to the original loss of Wasserstein GAN to better restore the degraded underwater images. This method can well correct the color of underwater images, but the model has some limitations in the types of images. Liu et al. proposed MLFcGAN (Liu X. et al., 2019). The network extracted multi-scale features and enhanced local features by using global features. The network was effective in restoring underwater image colors and eliminating unwanted artifacts but performed poorly in restoring image texture details. Guo et al. proposed a multi-scale dense generative adversarial network (Guo et al., 2019). This network improved its performance through the design of a multi-scale dense module, alongside the incorporation of L1 loss and gradient loss. These additions enabled the network to generate images of superior quality. While this method demonstrated strong results with both synthetic and real underwater images, it falls short in producing aesthetically pleasing underwater images.

Yang et al. proposed an underwater image enhancement method based on conditional generative adversarial networks (Yang et al., 2020). The network employed a dual discriminator that guided the generator to generate higher-quality images in terms of both global semantics and local features. Islam et al. proposed a fast underwater enhancement method (Islam et al., 2020). The method used a CGAN-based network model with L1 loss and Perceptual loss as loss functions. It showed good enhancement on the EUVP dataset but was less effective for enhancing images with severe degradation. Wu et al. proposed a multi-scale fusion generative adversarial network (Wu et al., 2022). To help the model achieve better performance, the network used three prior images as input to refine the prior features and then these features are fused into the encoding and decoding process. Jiang et al.

proposed a target-oriented perceptual fusion generative adversarial network (Jiang et al., 2022). The network used a Multi-scale Dense Boosted module and Deep Aesthetic Render module for contrast enhancement and color correction. A dual discriminator for global-local discrimination was used, too. Liu R. et al. proposed an unsupervised model to enhance underwater images (Liu et al., 2022). They constructed a twin adversarial constrained enhancement module to eliminate the reliance on paired images. It showed good performance in improving image quality.
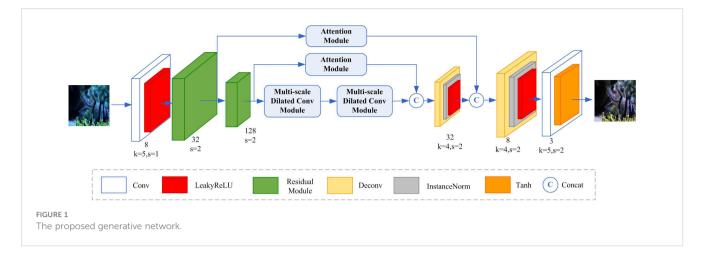
Although underwater image enhancement based on generative adversarial networks attempts to improve the quality of underwater images, the enhanced underwater images still contain much color distortion and detail loss, which affect the performance of high-level computer vision for analysis of underwater images.

3 Proposed method

To improve the quality of the enhanced underwater images, a generative adversarial network for enhancing underwater images is proposed. A generative adversarial network consists of a generative network and an adversarial network. The generative network is responsible for enhancing degraded underwater images, aiming to produce visually improved outputs. It takes degraded underwater images as input and generates enhanced versions of those images. The adversarial network plays a crucial role in evaluating the performance of the generative network. It serves as a discriminator and is responsible for determining whether an input image belongs to the set of generated images produced by the generative network or the original high-quality images. The adversarial component guides and offers feedback to the generative network, promoting its improvement by differentiating between real and generated images. Through the interaction of the generative and adversarial networks, the proposed GAN framework strives to augment the quality of underwater images. This is achieved by training the generative network to generate visually appealing and realistic outputs, with the adversarial network providing feedback to refine its performance.
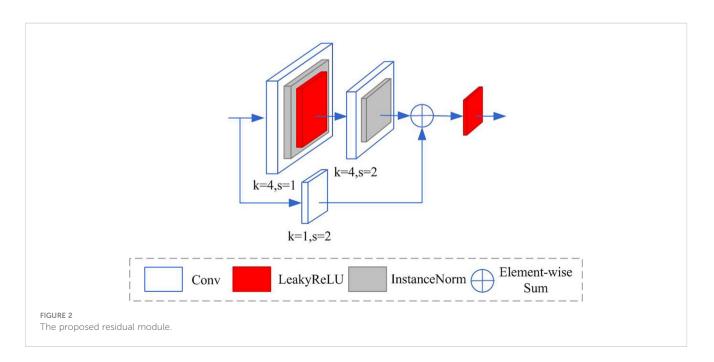
3.1 Generative network

The proposed generative network consists of three primary modules: a residual module, a multi-scale dilated convolution module, and an attention module. The structure of the generative network is illustrated in Figure 1. Firstly, a convolutional layer coupled with the LeakyReLU activation function is employed to extract local features. The number of channels also increases from 3 to 8. Secondly, two designed residual modules are used to reduce the size of feature maps and extract feature maps with different scales. The residual module reduces the output feature map size to one-half of the input feature map size. The number of channels of the output feature maps for the two residual modules are 32 and 128, respectively. Thirdly, two designed multi-scale dilated convolution modules are used to expand the receptive field and extract multi-scale features without changing the number of channels and the size

FIGURE 1
The proposed generative network.

of the feature map. Fourthly, we use deconvolution to increase the size of the feature map. In contrast to the conventional upsampling operation, deconvolution allows for the learning of parameters, which aids in recovering feature information. Besides, a new attention module is proposed and used to extract important features that are beneficial for enhancing underwater images. In the end, the extracted feature maps are restored to an RGB image using a convolution layer and a Tanh activation function. This restored underwater image is the enhanced underwater image.

In conventional generative networks, multiple downsampling is employed to broaden the receptive field and extract feature information on different scales. The downsampling will lead to the loss of information, which reduces the quality of the enhanced image. To reduce the loss of information, we design a residual module and use the designed residual module to realize the downsampling operation. In the designed generative network, two designed residual modules are used to expand the receptive field and extract multi-scale feature information. The structure of the proposed designed residual module is illustrated in Figure 2. It

consists of two branches. The upper branch consists of a 4×4 convolution layer with stride 1, an instance normalization layer, a LeakyReLU activation function, a 4×4 convolution layer with stride 2, and an instance normalization layer. The first 4×4 convolution layer is used to increase the number of channels without changing the size of the feature map. The instance normalization layer is used to improve network stability and speed up network convergence. The second 4×4 convolution layer is used to decrease the size of feature maps, instead of pooling operation for downsampling. In the lower branch, a 1×1 convolution layer with stride 2 is used to directly decrease the size of feature maps. The output feature maps of two branches are fused by the element addition operation.

A larger receptive field proves advantageous for extracting global features, whereas a smaller receptive field is well-suited for extracting local texture features. To extract more comprehensive information, we have designed a multi-scale dilated convolution module. This module integrates five distinct receptive fields, allowing it to extract features across various scales simultaneously. The architecture of the proposed multi-scale dilated convolution

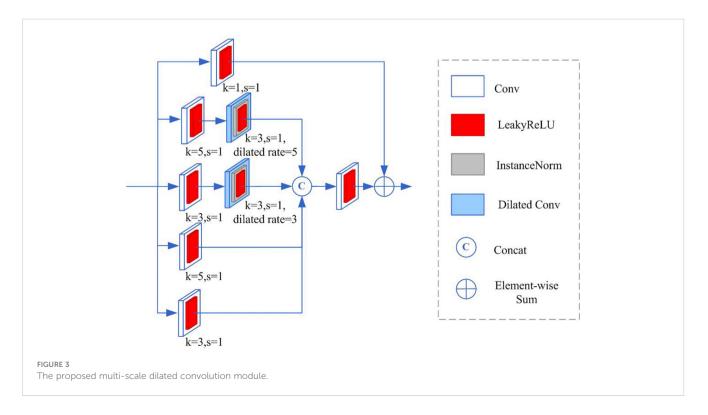**FIGURE 2**
The proposed residual module.

module is illustrated in Figure 3. It consists of five branches. The first branch only contains a 1×1 convolution with a LeakyReLU activation function. The second branch contains a 5×5 convolution with a LeakyReLU activation function, a 3×3 dilated convolution with dilated rate 5, an instance normalization, and a LeakyReLU activation function. The third branch contains a 3×3 convolution with a LeakyReLU activation function, a 3×3dilated convolution with dilated rate 3, an instance normalization, and a LeakyReLU activation function. The fourth branch contains a 5×5 convolution with a LeakyReLU activation function. The fifth branch contains a 3×3 convolution with a LeakyReLU activation function. The receptive fields for the five branches are 1×1, 15×15, 9×9, 5×5 and 3×3, respectively. The output feature maps of the second branch to the fifth branch are fused by the concatenation operation. The last 1×1 convolution layer is used to adjust the number of channels of the fused feature map to be equal to the number of channels in the first branch. In the end, the output feature maps of the last 1×1 convolution and the first branch are fused by an element-wise addition operation.
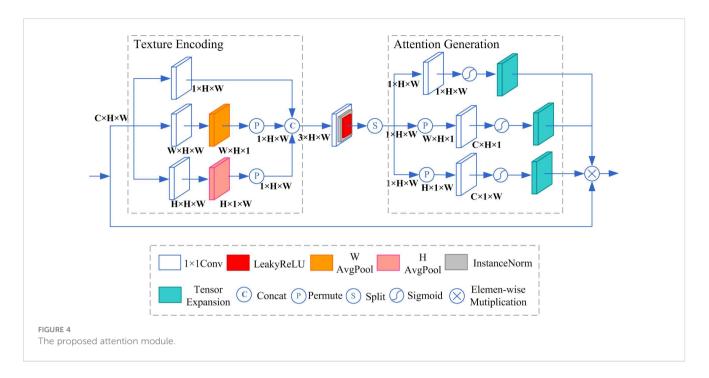
For the extraction of crucial features that contribute to the enhancement of underwater images, an attention module has been introduced, as depicted in Figure 4. This module comprises three key components: a texture encoding module, a feature extraction module, and an attention generation module. The texture encoding module consists of three branches. The first branch only contains a 1×1 convolution layer that is used to encode features in channel dimension. The second branch contains a 1×1 convolution layer, an average pooling in the horizontal direction, and a permute operation. The convolution layer is used to adjust the number of channels. The average pooling is used to encode the features along the horizontal direction. The permute operation is used to rearrange the dimensions of the feature maps. The third branch contains a

1×1 convolution layer, an average pooling in the vertical direction, and a permute operation. The average pooling is used to encode the features along the vertical direction. In the end, the output feature maps of three branches are concatenated by the concatenate operation. The size of the feature map of the texture encoding module is 3×H×W. The feature extraction module contains a 1×1 convolution with a LeakyReLU activation function to extract features. The attention generation module contains three branches that contain a permute operation, a 1×1convolution, and a sigmoid function. The permute operation is to rearrange the dimensions of the feature maps. The 1×1 convolution is used to adjust the number of channels. The sigmoid function is used to generate weight. Utilizing the attention generation module, three weights are derived for the channel dimension, vertical dimension, and horizontal dimension. Eventually, these three weights are applied as multipliers to the input feature map, resulting in the production of the output feature map.

3.2 Discriminative network

A multi-scale adversarial network based on the Markovian adversarial network is proposed and shown in Figure 5. It consists of 5 layers of downsampling. The downsampling operation is implemented by the residual module to reduce the loss of feature information. To improve the multi-scale feature extraction capability of the adversarial network, the concatenation operation is employed to fuse the features obtained from three residual modules at different scales. Furthermore, we use the convolution operation to adjust the number of channels in the output feature maps of the 4th and 5th residual modules to 1, respectively. In the end, we get the two judgment scores from two
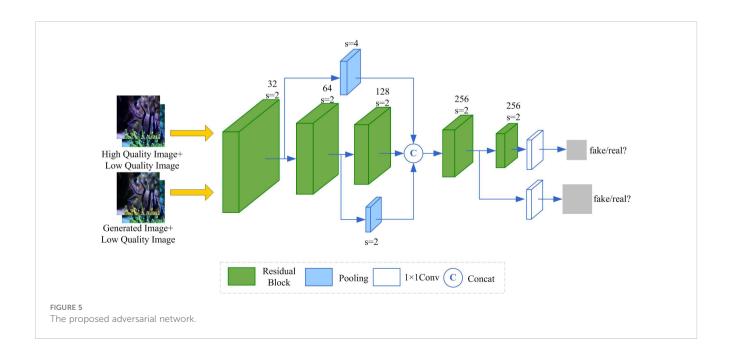

FIGURE 3
The proposed multi-scale dilated convolution module.

FIGURE 4
The proposed attention module.

adjusted output feature maps. The weights of the two scores are 0.6 and 0.4, respectively.

3.3 Loss function

An improved loss function is proposed by introducing the total variation loss into the conventional loss function. The improved loss function is expressed as follows.

$$L = \lambda_1(0.6 \times L_{GAN}(G, D_1) + 0.4 \times L_{GAN}(G, D_2)) + \lambda_2$$
$$\times L_{L1}(G) + \lambda_3 \times L_{per}(G) + \lambda_4 \times L_{TV}(G) \quad (1)$$

where λ_1, λ_2, λ_3 and λ_4 are scaling coefficients to adjust the importance of each loss function. We initially set the coefficient initial values based on the importance of the function, where the higher importance of the function corresponds to a larger coefficient value. We randomly set the coefficient initial values to $\lambda_1 = 0.2$; $\lambda_2 = 0.4, 0.42, 0.44, 0.46, 0.48$, or 0.5; and $\lambda_3 = 0.2, 0.19, 0.18, 0.17, 0.16$, or 0.15; and $\lambda_4 = 0.2, 0.19, 0.18, 0.17, 0.16$, or 0.15. In our experiments, when $\lambda_1 = 0.2$, $\lambda_2 = 0.48$, $\lambda_3 = 0.16$, and $\lambda_4 = 0.16$, our proposed method has better performance in enhancing underwater images. Therefore, we set $\lambda_1 = 0.2$, $\lambda_2 = 0.48$, $\lambda_3 = 0.16$, and $\lambda_4 = 0.16$. $L_{GAN}(\cdot)$ is the adversarial loss function, which is calculated as follows.

FIGURE 5
The proposed adversarial network.

$$L_{GAN}(G, D_1) = E_{X,Y}[\log D_1(Y, X)] + E_{X,Y}[\log(1 - D_1(G(X), X))] \tag{2}$$

$$L_{GAN}(G, D_2) = E_{X,Y}[\log D_2(Y, X)] + E_{X,Y}[\log(1 - D_2(G(X), X))] \tag{3}$$

where X is the low-quality underwater image, Y is the high-quality underwater image, G is the generator, $G(X)$ is the output of the generator, $D_1(\cdot)$ is the first output score of the discriminator, and $D_2(\cdot)$ is the second output score of the discriminator. $L_{L1}(G)$ is the L1 loss function, which is expressed as follows.

$$L_{L1}(G) = E_{X,Y}[||Y - G(X)||_1] \tag{4}$$

$L_{per}(G)$ is the perceptual loss function, which is expressed as follows.

$$L_{per}(G) = E_{X,Y}[||\phi(Y) - \phi(G(X))||_2] \tag{5}$$

where $\phi(\cdot)$ are the high-level features extracted by a pre-trained VGG-19 network. The total variation loss is expressed as follows.

$$L_{TV}(G(X)) = \sum_{i,j}((x_{i,j+1} - x_{i,j})^2 + (x_{i+1,j} - x_{i,j})^2)^{\frac{\beta}{2}} \tag{6}$$

where $x_{i,j}$ is the pixel in the ith row and the jth column of the enhanced image that is the $G(X)$.

4 Simulation and discussion

In this section, we use a synthetic paired underwater image dataset and a real underwater image dataset to test the effectiveness of our proposed method through qualitative and quantitative experiments. We compared our method with four underwater image enhancement methods: the FUnIEGAN (Islam et al., 2020), the FWGAN (Wu et al., 2022), the TOPAL (Jiang et al., 2022) and the TACL (Liu et al., 2022). We evaluated each method quantitatively using two full-reference evaluation metrics which were peak signal-to-noise ratio (PSNR), structural similarity (SSIM, Wang et al., 2004), and one no-reference evaluation metric which was underwater image quality measures (UIQM, Panetta et al., 2015).

4.1 Datasets and metrics

The underwater ImageNet dataset (Fabbri et al., 2018) which contains 6,128 pairs of underwater images is used as the training set. Fabbri et al., first selected underwater images from the ImageNet dataset, and secondly classified the selected underwater images into degraded underwater images and high-quality underwater images. Thirdly, CycleGAN was trained on these images and used to generate 6,128 pairs of high-quality underwater images and their corresponding degraded underwater images. The EUVP dataset (Islam et al., 2020) which contains paired images, and the UIEB dataset (Li et al., 2019), which contains real-world underwater

images, were used to test the performance of different methods, respectively. The PSNR, SSIM, and UIQM were evaluation metrics. The PSNR is expressed as follows.

$$PSNR(x, y) = 10\log_{10}[\frac{255^2}{MSE(x, y)}] \tag{7}$$

where x and y are the enhanced image and the corresponding ground truth image. MSE is the mean square error. It is expressed as follows.

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(x_{i,j} - y_{i,j})^2 \tag{8}$$

where $x_{i,j}$ is a pixel of enhanced image x, $y_{i,j}$ is a pixel of the ground truth image y. The larger the value of PSNR is, the closer the enhanced image is to the ground truth image. The SSIM is expressed as follows.

$$SSIM(x, y) = \left(\frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}\right)\left(\frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}\right) \tag{9}$$

where μ_x and μ_y are the mean of x and y, respectively. The σ_x^2 and σ_y^2 are the variance of x and y, respectively. The σ_{xy} is the covariance of x and y. The closer the SSIM value is to 1, the closer the enhanced image is to the ground truth image.
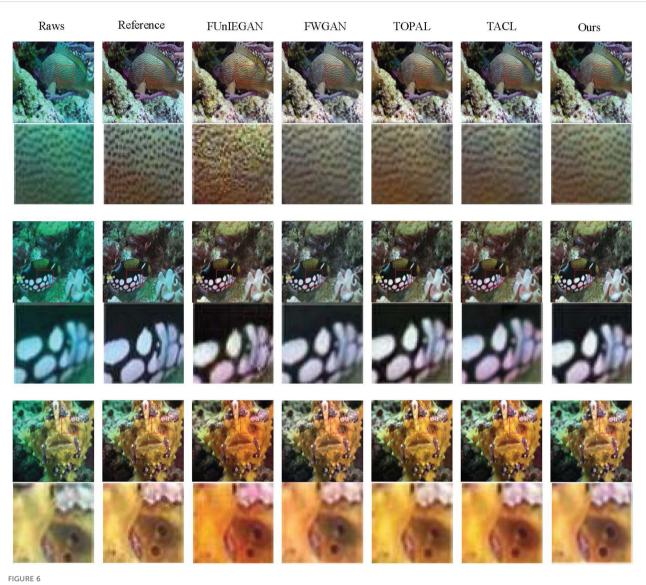
The UIQM consists of three parts, which are UICM (underwater image colorfulness measure), UISM (underwater image sharpness measure), and UIConM (underwater image contrast measure). It is expressed as follows.

$$UIQM = c_1 UICM + c_2 UISM + c_3 UIConM \tag{10}$$

where $c_1 = 0.0282$, $c_2 = 0.2953$, and $c_3 = 3.5753$. The higher the UIQM value is, the better the color balance, sharpness, and contrast of the enhanced image are.

4.2 Underwater image enhancement on synthetic images

We evaluated the performance of our proposed method on the synthetic images using the test_samples subset from the EUVP dataset. This subset consisted of 515 pairs of underwater degraded images along with their corresponding high-quality images. We randomly selected three paired images that consisted of reference images (high-quality images) and corresponding degraded images that were synthesized to test the performances of FUnIEGAN, FWGAN, TOPAL, TACL, and our method. The reference underwater images (high-quality underwater images), raw images (degraded underwater images), and underwater images enhanced by different methods are shown in Figure 6. In the first row, the FUnIEGAN method introduces noise interference in the enhanced images, which negatively affects their quality. The FWGAN cannot effectively enhance the contrast. The TOPAL method fails to adequately restore the details in the enhanced images, leading to a loss of important visual information. The TACL method produces reddish color biases in the enhanced images, which deviates from

FIGURE 6
The images enhanced by different methods from the EUVP dataset.

the desired color accuracy and may impact the overall visual quality. The image enhanced by our method stands out due to its exceptional color correction and contrast enhancement. In the second row, the FUnIEGAN method introduces artifacts in the enhanced image, negatively impacting its visual quality. The FWGAN method, although improving the image, still leaves a residual haze effect. However, the TACL, TOPAL, and our proposed method effectively enhance the image details, resulting in sharper visual features. Nonetheless, our approach excels in terms of color fidelity, yielding a more authentic and natural color rendition in the enhanced image. In the third row, the FUnIEGAN method introduces reddish color biases to the enhanced image, which diverges from the original color representation. The FWGAN method fails to provide clear details in the enhanced image, resulting in reduced clarity. Both the TOPAL and TACL methods leave a residual haze effect on the enhanced images, diminishing their overall quality. Conversely, our proposed method produces enhanced images that closely resemble

the reference image, demonstrating better color accuracy and preserving finer details.

In summary, the FUnIEGAN method falls short in effectively enhancing degraded underwater images and addressing issues related to color bias and image details. The FWGAN, TOPAL, and TACL methods struggle to adequately sharpen image details. In contrast, our proposed method successfully generates enhanced images that demonstrate satisfactory color correction, contrast enhancement, and detail sharpening. Furthermore, the images enhanced by our method closely resemble the corresponding reference images, indicating its superior performance.

To quantitatively analyze the performance of different methods, we used all paired images in the EUVP test set as test images and computed the PSNR, SSIM, and UIQM of the images enhanced by different methods. The average results are shown in Table 1 The PSNR values for the FUnIEGAN method, FWGAN method, TOPAL method, TACL method, and our proposed method were 23.23, 26.87, 25.41, 26.65, and 27.28, respectively. Our proposed

TABLE 1 Performances of underwater image enhancement for different methods on the EUVP dataset.

	FUnIEGAN	FWGAN	TOPAL	TACL	Ours
PSNR	23.23	26.87	25.41	26.65	28.28
SSIM	0.70	0.80	0.76	0.79	0.85
UIQM	3.04	2.97	3.06	2.95	3.25

method had the highest PSNR, followed by the FWGAN method and TACL method. The SSIM values for the FUnIEGAN method, FWGAN method, TOPAL method, TACL method, and our proposed method were 0.70, 0.80, 0.76, 0.79 and 0.81, respectively. Our proposed method had the highest SSIM, followed by the FWGAN method and TACL method. The UIQM values for the FUnIEGAN method, FWGAN method, TOPAL method, TACL method, and our proposed method were 3.04, 2.97, 3.06, 2.95, and 3.14, respectively. Our proposed method had the highest UIQM, followed by the FUnIEGAN method and TOPAL method. As shown in Table 1, our proposed method had the highest PSNR, SSIM, and UIQM, which shows that our proposed method had the best performance in enhancing underwater images on the EUVP dataset.

4.3 Underwater image enhancement on real-world images

We also assessed the effectiveness of our proposed method using 950 real-world images sourced from the UIEB dataset. This dataset encompasses a total of 950 underwater images in degraded conditions, consisting of 890 images from the raw-890 subset and an additional 60 images from the challenging-60 subset. We randomly selected three real-world degraded underwater images from the raw-890 subset to evaluate the performances of FUnIEGAN, FWGAN, TOPAL, TACL, and our proposed method. The images included both the original degraded underwater images and the corresponding enhanced underwater images generated by each method. The results are visualized in Figure 7. In the first row, the images enhanced by FUnIEGAN, FWGAN, and TACL exhibit reddish color biases, while the image enhanced by TOPAL shows greenish color biases. In contrast, the image enhanced by our proposed method achieves satisfactory contrast and saturation. In the second row, the image enhanced by FUnIEGAN exhibits checkerboard artifacts. The image enhanced by the FWGAN suffers from bluish color bias. The image enhanced by the TOPAL suffers from greenish color bias. TACL generates a darkish image. In contrast, our proposed method effectively removes the bluish hue, resulting in a more accurate and visually pleasing enhanced image. In the third row, the image enhanced by FUnIEGAN exhibits artifacts. The images enhanced by FWGAN and TACL suffer from reddish color bias. TOPAL fails to effectively sharpen details in the enhanced image. In contrast, our proposed method successfully enhances the images in terms of color correction and detail sharpening.

In summary, the FUnIEGAN usually introduces color bias and artifacts in the enhanced image. The FWGAN, TOPAL, and TACL cannot effectively correct color and sharpen details. The results

show that our method has the best generalization capability for real-world underwater image enhancement.
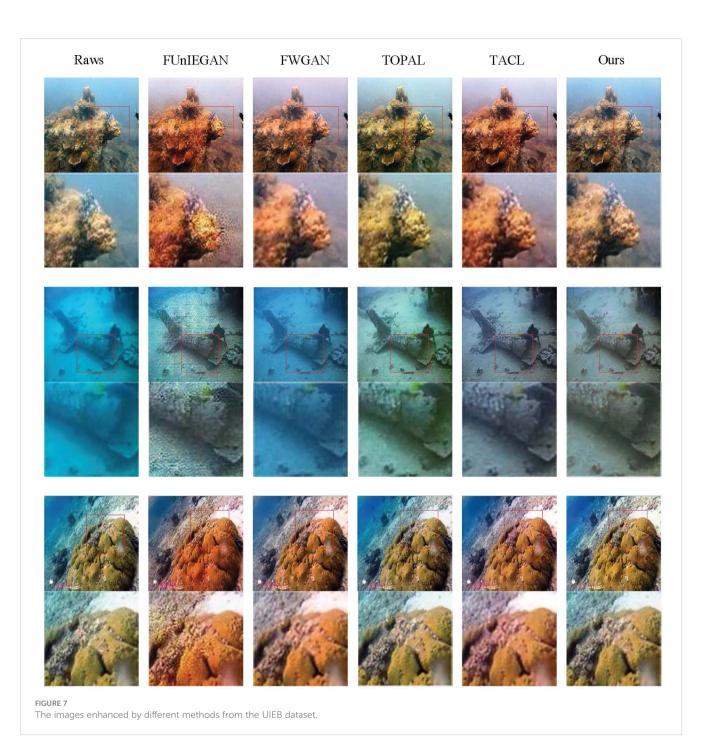
To quantitatively analyze the performance of different methods, we used all real-world degraded underwater images on the UIEB dataset as test images and computed the UICM, UISM, UIConM, and UIQM. The UICM, UISM, and UIConM are used to measure the colorfulness, sharpness, and contrast of enhanced underwater images, respectively. The UIQM which consists of UICM, UISM, and UIConM is used to measure the complete visual effect of enhanced underwater images. The average results are shown in Table 2. The UICM values for the FUnIEGAN method, FWGAN method, TOPAL method, TACL method, and our proposed method were 4.41, 4.43, 4.93, 4.68, and 5.02, respectively. Our proposed method had the highest UICM, followed by the TOPAL method and TACL method, which shows that our method had the best performance in restoring colorfulness. The UISM values for the FUnIEGAN method, FWGAN method, TOPAL method, TACL method, and our proposed method were 5.82, 5.99, 6.29, 6.16, and 6.68, respectively. Our proposed method had the highest UISM, followed by the TOPAL method and TACL method, which shows that our method had the best performance in restoring sharpness. The UIConM values for the FUnIEGAN method, FWGAN method, TOPAL method, TACL method, and our proposed method were 0.24, 0.24, 0.26, 0.27, and 0.28, respectively. Our proposed method had the largest UIConM, followed by the TOPAL method and TACL method, which shows that our method had the best performance in restoring contrast. The UIQM values for the FUnIEGAN method, FWGAN method, TOPAL method, TACL method, and our proposed method were 2.70, 2.78, 2.93, 2.91, and 3.12, respectively.

As shown in Table 3, our proposed method had the best performance in restoring colorfulness, sharpness, and contrast. Furthermore, our proposed method also had the largest UIQM, which shows that our method had the best performance in improving the visual effect of underwater images.

4.4 Ablation study

To analyze the contributions of the multi-scale adversarial network, the multi-scale dilated convolution module, the attention module, and the TV loss, we conducted the following ablation studies:

- w/o MAN: without multi-scale adversarial network
- w/o MDCM: without multi-scale dilated convolution module;
- w/o AM: without attention module;
- w/o TV loss: without TV loss;

FIGURE 7
The images enhanced by different methods from the UIEB dataset.

The PSNR, SSIM, and UIQM scores on the test_samples subset from the EUVP dataset are shown in Table 3. As is shown in the table, our complete model achieved the best performance compared to all the ablated models, which proves the effectiveness of the multi-scale adversarial network, the multi-scale dilated convolution module, the attention module, and the TV loss.

4.4.1 Ablation study on MAN

The multi-scale adversarial network is designed to improve the discriminative power, which is helpful in improving the quality of images generated by the generative network. From Table 3, compared with our method without a multi-scale adversarial network, our complete method improved the PSNR, SSIM, and UIQM scores by nearly 5%, 7%, and 9%, respectively. It is shown that the multi-scale adversarial network is effective in improving the performance of the network model.

4.4.2 Ablation study on MDCM

The multi-scale dilated convolution module in the generative network is designed to increase the receptive field and extract more feature information. From Table 3, compared with our method without multi-scale dilated convolution modules, our complete method improved the PSNR, SSIM, and UIQM scores by nearly 10%, 9%, and 9%, respectively. The result demonstrates the important role of the

TABLE 2 Performances of underwater image enhancement for different methods on the UIEB dataset.

	FUnIEGAN	FWGAN	TOPAL	TACL	Ours
UICM	4.41	4.43	4.93	4.68	5.52
UISM	5.82	5.99	6.29	6.16	6.68
UIConM	0.24	0.24	0.26	0.27	0.31
UIQM	2.70	2.78	2.93	2.91	3.24

multi-scale dilated convolution module in recovering Image color and enhancing detailed information.

4.4.3 Ablation study on AM

The attention module is designed to extract important features and reduce the influence of unimportant features. From Table 3, compared with our method without the attention module, our complete method improved the PSNR, SSIM, and UIQM scores by nearly 4%, 5%, and 4%, respectively. It is shown that the attention module is effective in improving the quality of images generated by our method.

4.4.4 Ablation study on TV loss

The TV loss is added to the traditional loss function to improve the visual effect of the generated images. From Table 3, compared with our method without TV loss, our complete method improved the PSNR, SSIM, and UIQM scores by nearly 2%, 4%, and 7%, respectively. The TV loss improved the image visual effect by smoothing image details, so the PSNR and SSIM scores were less increased, and the UIQM score, which is more suitable for the quality of human eye vision, was increased.

4.5 Performance of saliency detection based on enhanced underwater images

We also randomly selected an image from the UIEB dataset to test the performance of saliency detection (Liu and Yu, 2022) based on the enhanced underwater image. In Figure 8, the images in the first row are the real-world degraded underwater image (raw image) and enhanced images by FUnIEGAN, FWGAN, TOPAL, TACL, and our method. The images in the second row are the saliency probability maps by saliency detection from the images in the first row in Figure 8. All shark boundaries are not clear for the first image in the second row. The saliency detection method only detected one shark boundary from the enhanced images by TOPAL and TACL methods, respectively. While saliency detection effectively detects more shark boundaries in the enhanced images generated by FUnIEGAN, FWGAN, and our method, it is worth noting that the

boundary of the shark in the upper right part of the detected image appears less distinct in the enhanced images created by the FUnIEGAN and FWGAN methods. Compared with other methods, saliency detection detected more shark boundaries and the boundaries are clearer in the enhanced image from our method. It shows that our underwater image enhancement method is more useful in improving the performance of saliency detection.

4.6 Model complexity analysis

To further prove the performance of our model, we measured its complexity using two metrics: floating point operations (FLOPs) and parameters. FLOPs quantify the amount of computation required by the model and can be used to gauge its complexity. In our analysis, we focused on the convolution layer as it contributes significantly to the computational load. The FLOPs of the convolution layer can be calculated as follows.

$$FLOPs = (2 \times C_i \times K^2 - 1) \times H \times W \times C_o \qquad (11)$$

where C_i is the number of input feature map channels, K is the size of the convolution kernel, H, W, C_o are the height, width, and channels of the output feature map. The parameters of the convolution layer can be calculated as follows.

$$parameters = C_i \times C_o \times K^2 \qquad (12)$$

where C_i, C_o are the number of input and output channels, K is the size of the convolution kernel. Among them, the smaller the FLOPs, the smaller the computational complexity of the model. The fewer the parameters, the less memory the model occupies.

We compared the model complexity of FUnIEGAN, FWGAN, TOPAL, and TACL and our method. The measurement results are shown in Table 4. While our method had a larger number of FLOPs compared to FUnIEGAN and FWGAN, it was still smaller than TOPAL and TACL. Additionally, although our method had a larger number of parameters compared to FUnIEGAN, it was smaller than FWGAN, TOPAL, and TACL. When compared to TACL and TOPAL, which have demonstrated better performance in enhancing underwater

TABLE 3 Experimental results of the ablation study.

	w/o MAN	w/o MDCM	w/o AM	w/o TV loss	Ours
PSNR	26.97	25.54	27.28	27.61	28.28
SSIM	0.79	0.77	0.81	0.82	0.85
UIQM	2.95	2.97	3.12	3.01	3.25

FIGURE 8
Saliency detection based on enhanced underwater images.

TABLE 4 Measurement of model complexity.

	FUnIEGAN	FWGAN	TOPAL	TACL	Ours
FLOPs(G)	10.75	19.44	85.50	97.03	21.18
parameters(M)	7.72	10.51	30.04	21.29	7.81

images, our proposed method stands out with smaller values for FLOPs and parameters.

5 Conclusion

In this article, a new generative adversarial network is proposed for enhancing underwater images. First, we designed a multi-scale dilated convolution module, attention module, and residual module, and employed these modules to construct a generative network with a U-shaped network structure. The designed generative network was utilized to enhance the underwater images. Second, we designed a multi-scale adversarial network to indirectly improve the performance of the generative network. Finally, we proposed an improved loss function for the designed generative adversarial networks by incorporating the total variation loss into the traditional loss function. We used the paired underwater images consisting of high-quality images, degraded images, and real-world underwater images to test the performance of our proposed method. Compared with the other methods, the images enhanced by our proposed method from the synthesized degraded underwater images had the highest PSNR, SSIM, and UIQM. The images enhanced by our proposed method from the real-world degraded underwater images also had the highest UICM, UISM, UIConM, and UIQM. These show that the quality and visualization of enhanced underwater images from our method are better than other methods. Furthermore, ablation

studies on the network structure and loss function demonstrated the effectiveness of our method. Moreover, we also used the underwater image enhancement method as a pre-processing method for saliency detection. The saliency detection results indicated that our proposed method outperforms other methods in terms of saliency detection performance. Our proposed method can improve the performance of high-level image processing. In the final analysis, we performed a thorough complexity assessment of the proposed network model. The results conclusively show that our method demonstrated lower complexity and required less memory compared to alternative approaches.

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

ZW performed the experiments and wrote the manuscript. LZ and TZ revised the manuscript. YJ and YC provided the ideas and revised the article. LZ provided advice and GPU devices for parallel computing. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

Author YC was employed by the company Guangdong Power Grid Co., Ltd.

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

Bell, K. L. C., Chow, J. S., Hope, A., Quinzin, M. C., Cantner, K. A., Amon, D. J., et al. (2022). Low-cost, deep-sea imaging and analysis tools for deep-sea exploration: a collaborative design study. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.873700

Cui, Y., Sun, Y., Jian, M., Zhang, X., Yao, T., Gao, X., et al. (2022). A novel underwater image restoration method based on decomposition network and physical imaging model. *Int. J. Intelligent Syst.* 37 (9), 5672–5690. doi: 10.1002/int.22806

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.knosys.2022.109751

Estrada, D. C., Dalgleish, F. R., Den Ouden, C. J., Ramos, B., Li, Y., and Ouyang, B. (2022). Underwater LiDAR image enhancement using a GAN based machine learning technique. *IEEE Sensors J.* 22 (5), 4438–4451. doi: 10.1109/JSEN.2022.3146133

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, QLD, Australia: IEEE) 7159–7165.

Gavrilov, A. N., and Parnum, I. M. (2010). Fluctuations of seafloor backscatter data from multibeam sonar systems. *IEEE J. Oceanic Eng.* 35 (2), 209–219. doi: 10.1109/JOE.2010.2041262

Guo, Y., Li, H., and Zhuang, P. (2019). Underwater image enhancement using a multiscale dense generative adversarial network. *IEEE J. Oceanic Eng.* 45 (3), 862–870. doi: 10.1109/JOE.2019.2911447

Huang, Z., Fang, H., Li, Q., Li, Z., Zhang, T., Sang, N., et al. (2018a). Optical remote sensing image enhancement with weak structure preservation via spatially adaptive gamma correction. *Infrared Phys. Technol.* 94, 38–47. doi: 10.1016/j.infrared.2018.08.019

Huang, Z., Huang, L., Li, Q., Zhang, T., and Sang, N. (2018b). Framelet regularization for uneven intensity correction of color images with illumination and reflectance estimation. *Neurocomputing* 314, 154–168. doi: 10.1016/j.neucom.2018.06.063

Huang, Z., Wang, L., An, Q., Zhou, Q., and Hong, H. (2022). Learning a contrast enhancer for intensity correction of remotely sensed images. *IEEE Signal Process. Lett.* 29, 394–398. doi: 10.1109/LSP.2021.3138351

Huang, Z., Wang, Z., Zhang, J., Li, Q., and Shi, Y. (2021a). Image enhancement with the preservation of brightness and structures by employing contrast limited dynamic quadri-histogram equalization. *Optik* 226, 165887. doi: 10.1016/j.ijleo.2020.165877

Huang, Z., Zhang, T., Li, Q., and Fang, H. (2016). Adaptive gamma correction based on cumulative histogram for enhancing near-infrared images. *Infrared Phys. Technol.* 79, 205–215. doi: 10.1016/j.infrared.2016.11.001

Islam, M. J., Xia, Y., and Sattar, J. (2020). Fast underwater image enhancement for improved visual perception. *IEEE Robotics Automation Lett.* 5 (2), 3227–3234. doi: 10.1109/LRA.2020.2974710

Jiang, Z., Li, Z., Yang, S., Fan, X., and Liu, R. (2022). Target oriented perceptual adversarial fusion network for underwater image enhancement. *IEEE Trans. Circuits Syst. Video Technol.* 32 (10), 6584–6598. doi: 10.1109/TCSVT.2022.3174817

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, X., Hu, H., Huang, Y., Jiang, L., Che, L., Liu, T., et al. (2022). UCRNet: Underwater color image restoration via a polarization-guided convolutional neural network. *Front. Mar. Sci.* 2441. doi: 10.3389/fmars.2022.1031549

Liu, X., Gao, Z., and Chen, B. M. (2019). MLFcGAN: Multilevel feature fusion-based conditional GAN for underwater image color correction. *IEEE Geosci. Remote Sens. Lett.* 17 (9), 1488–1492. doi: 10.1109/LGRS.2019.2950056

Liu, R., Jiang, Z., 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

Liu, W., Yao, R., and Qiu, G. (2019). A physics based generative adversarial network for single image defogging. *Image Vision Computing* 92, 103815. doi: 10.1016/j.imavis.2019.10.001

Liu, L., and Yu, W. (2022). Underwater image saliency detection via attention-based mechanism. *J. Physics: Conf. Ser.* 2189 (1), 12012. doi: 10.1088/1742-6596/2189/1/012012

Luo, Z., Tang, Z., Jiang, L., and Ma, G. (2022). A referenceless image degradation perception method based on the underwater imaging model. *Appl. Intell.* 52, 1–17. doi: 10.1007/s10489-021-02815-3

Panetta, K., Gao, C., and Agaian, S. (2015). Human-visual-system-inspired underwater image quality measures. *IEEE J. Oceanic Eng.* 41 (3), 541–551. doi: 10.1109/JOE.2015.2469915

Srivastava, A., Chanda, S., and Pal, U. (2022). AGA-GAN: Attribute Guided Attention Generative Adversarial Network with U-Net for face hallucination. *Image Vision Computing* 126, 104534. doi: 10.1016/j.imavis.2022.104534

Tian, C., Yuan, Y., Zhang, S., Lin, C. W., Zuo, W., and Zhang, D. (2022). Image super-resolution with an enhanced group convolutional neural network. *Neural Networks* 153, 373–385. doi: 10.1016/j.neunet.2022.06.009

Townhill, B. L., Reppas-Chrysovitsinos, E., Sühring, R., Halsall, C. J., Mengo, E., Sanders, T., et al. (2022). Pollution in the Arctic Ocean: An overview of multiple pressures and implications for ecosystem services. *Ambio* 51, 471–483. doi: 10.1007/s13280-021-01657-0

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 (4), 600–612. doi: 10.1109/TIP.2003.819861

Wang, Y., Yan, X., Guan, D., Wei, M., Chen, Y., Zhang, X. P., et al. (2022). Cycle-snspgan: Towards real-world image dehazing via cycle spectral normalized soft likelihood estimation patch gan. *IEEE Trans. Intelligent Transportation Syst.* 23 (11), 20368–20382. doi: 10.1109/TITS.2022.3170328

Wu, J., Liu, X., Lu, Q., Lin, Z., Qin, N., and Shi, Q. (2022). FW-GAN: Underwater image enhancement using generative adversarial network with multi-scale fusion. *Signal Processing: Image Communication* 109, 116855. doi: 10.1016/j.image.2022.116855

Xu, H., Long, X., and Wang, M. (2023). UUGAN: a GAN-based approach towards underwater image enhancement using non-pairwise supervision. *Int. J. Mach. Learn. Cybernetics* 14 (3), 725–738. doi: 10.1007/s13042-022-01659-8

Yan, X., Qin, W., Wang, Y., Wang, G., and Fu, X. (2022). Attention-guided dynamic multi-branch neural network for underwater image enhancement. *Knowledge-Based Syst.* 258, 110041. doi: 10.1016/j.knosys.2022.110041

Yang, M., Hu, K., Du, Y., Wei, Z., Sheng, Z., and Hu, J. (2020). Underwater image enhancement based on conditional generative adversarial network. *Signal Processing: Image Communication* 81, 115723. doi: 10.1016/j.image.2019.115723

Yu, H., Li, X., Feng, Y., and Han, S. (2023). Multiple attentional path aggregation network for marine object detection. *Appl. Intell.* 53 (2), 2434–2451. doi: 10.1007/s10489-022-03622-0

Zhang, W., Dong, L., Zhang, T., and Xu, W. (2021). Enhancing underwater image via color correction and bi-interval contrast enhancement. *Signal Processing: Image Communication* 90, 116030. doi: 10.1016/j.image.2020.116030

Zhang, X., Yang, P., Huang, P., Sun, H., and Ying, W. (2022). Wide-bandwidth signal-based multi receiver SAS imagery using extended chirp scaling algorithm. *IET Radar Sonar Navigation* 16 (3), 531–541. doi: 10.1049/rsn2.12200

Zheng, M., and Luo, W. (2022). Underwater image enhancement using improved CNN based defogging. *Electronics* 11 (1), 150. doi: 10.3390/electronics11010150

Zhou, J., Sun, J., Zhang, W., and Lin, Z. (2023). Multi-view underwater image enhancement method via embedded fusion mechanism. *Eng. Appl. Artif. Intell.* 121, 105946. doi: 10.1016/j.engappai.2023.105946

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

Zhuang, P., Li, C., and Wu, J. (2021). Bayesian retinex underwater image enhancement. *Eng. Appl. Artif. Intell.* 101, 104171. doi: 10.1016/j.engappai.2021.104171

Zou, L., Zhao, M., Cao, F., Zan, S., Cheng, X., and Liu, X. (2021). Fish tracking based on feature fusion and scale adaptation in a real-world underwater environment. *Mar. Technol. Soc. J.* 55 (2), 45–53. doi: 10.4031/MTSJ.55.2.12