



OPEN ACCESS

EDITED BY

Chao Chen,
Suzhou University of Science and
Technology, China

REVIEWED BY

Xiaobo Li,
The Chinese University of Hong Kong, China
Hao Wang,
China University of Petroleum, China

*CORRESPONDENCE

Weitao Chen

✉ wtchen@cug.edu.cn

Xuwen Qin

✉ qinxuwen@163.com

RECEIVED 04 January 2025

ACCEPTED 17 March 2025

PUBLISHED 04 April 2025

CITATION

Shen S, Wang H, Chen W, Wang P, Liang Q
and Qin X (2025) A novel edge-feature
attention fusion framework for underwater
image enhancement.

Front. Mar. Sci. 12:1555286.

doi: 10.3389/fmars.2025.1555286

COPYRIGHT

© 2025 Shen, Wang, Chen, Wang, Liang and
Qin. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). 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.

A novel edge-feature attention fusion framework for underwater image enhancement

Shuai Shen¹, Haoyi Wang¹, Weitao Chen^{1*}, Pingkang Wang²,
Qianyong Liang³ and Xuwen Qin^{4,5*}

¹Faculty of Computer-Science, China University of Geosciences, Wuhan, Hubei, China, ²Department of Fundamental Investigations, China Geological Survey, Beijing, China, ³Guangzhou Marine Geological Survey, China Geological Survey, Guangzhou, Guangdong, China, ⁴Key Laboratory of Geological Survey and Evaluation of Ministry of Education, China University of Geosciences, Wuhan, Hubei, China, ⁵China Aero Geophysical Survey and Remote Sensing Centre for Natural Resources, China Geological Survey, Beijing, China

Underwater images captured by Remotely Operated Vehicles are critical for marine research, ocean engineering, and national defense, but challenges such as blurriness and color distortion necessitate advanced enhancement techniques. To address these issues, this paper presents the CUG-UIEF algorithm, an underwater image enhancement framework leveraging edge feature attention fusion. The method comprises three modules: 1) an Attention-Guided Edge Feature Fusion Module that extracts edge information via edge operators and enhances object detail through multi-scale feature integration with channel-cross attention to resolve edge blurring; 2) a Spatial Information Enhancement Module that employs spatial-cross attention to capture spatial interrelationships and improve semantic representation, mitigating low signal-to-noise ratio; and 3) Multi-Dimensional Perception Optimization integrating perceptual, structural, and anomaly optimizations to address detail blurring and low contrast. Experimental results demonstrate that CUG-UIEF achieves an average peak signal-to-noise ratio of 24.49 dB, an 8.41% improvement over six mainstream algorithms, and a structural similarity index of 0.92, a 1.09% increase. These findings highlight the model's effectiveness in balancing edge preservation, spatial semantics, and perceptual quality, offering promising applications in marine science and related fields.

KEYWORDS

underwater image enhancement, edge feature attention fusion, spatial crossattention, multidimensional perception optimization, attention-guided edge feature fusion

1 Introduction

Underwater images, captured in aquatic environments using remotely operated vehicles (ROVs), are crucial for marine exploration, underwater archaeology, and fishery monitoring, providing visual representations of underwater scenes and objects. However, underwater imaging environments are complex. The images obtained by ROVs are limited

by aggravated color distortion, objects with the same color in the background, and difficulty in edge distinction.

Underwater image enhancement (UIE) improves the quality of underwater images by mitigating their characteristic degradation features and bringing images closer to their true color and clarity, as observed in normal lighting environments. This enables more effective extraction and utilization of valuable features (Alsakar et al., 2024). High-quality underwater image data help reveal unknown marine life and geological features in the deep sea and provide critical information for biodiversity protection (Nazir and Kaleem, 2021), marine environmental monitoring (Wang et al., 2007), and resource sample collection (Mazzeo et al., 2022).

UIE techniques can be divided into two categories: traditional and deep learning-based methods. Traditional UIE techniques include color correction and image restoration methods. Color correction methods such as color balancing can improve color distortion but cannot address blurring and detail loss. Image restoration methods that incorporate physical models, such as light transmission or dehazing models, improve image clarity and optical effects more effectively (Hu et al., 2022). Common color correction methods often perform pixel-level restoration of image colors. For instance, Banik et al. (2018) used gamma correction in the value channel of the hue, saturation, value space to enhance low-light image contrast but introduced problems such as over-enhancement and halos. Garg et al. (2018) applied CLAHE and percentile methods to enhance underwater images and obtained good results in specific scenes but limited improvement in certain water environments. Image-restoration methods typically integrate physical models. Zhu (2023) proposed an enhancement algorithm based on graph theory that improves contrast and color using CIELab and red, green, blue (RGB) spaces combined with CLAHE. However, owing to the independent operations in each color space, the method lacks robustness in complex scenes. Drews et al. (2016) enhanced blue-green channels using a light propagation model but introduced red color distortion. Xiong et al. (2020) applied a linear model and nonlinear adaptive weighting strategy based on the Beer–Lambert law (Swinehart, 1962) to adjust underwater image colors. Recent studies have developed enhanced methods based on conventional algorithmic frameworks to address imaging degradation in specific scenarios. Zhang et al. (2025) proposes a cascaded restoration algorithm grounded in quadtree search-guided background region classification and cross-domain synergy, which integrates dynamic channel discrepancy compensation, S-curve-optimized homomorphic filtering, and chromatic space fusion, thereby significantly improving underwater image fidelity and object recognition robustness. Li et al. (2025) proposes a cascaded restoration algorithm integrating quadtree search-guided background region classification and a cross-domain collaboration mechanism, which effectively addresses color distortion and detail blurring in underwater optical imaging through dynamic channel discrepancy compensation and S-curve-optimized homomorphic filtering, thereby significantly enhancing object detection robustness and visual task performance. However, the methods do not perform well with foggy and low-light underwater images. In general, traditional methods based on fixed

underwater priors perform well in specific scenes but are limited by the unpredictability of underwater environments and thus lack general applicability.

Deep learning-based UIE methods use large datasets to train models that adaptively handle various problems, such as color distortion, blurring, and low contrast. These methods can restore image details more accurately and adapt to diverse underwater scenarios. Among the deep learning methods, generative adversarial networks (GANs) have gained prominence in the early stages of UIE for their ability to address limited data availability (Goodfellow et al., 2014). Li et al. (2017) proposed WaterGAN, which corrects underwater image colors by training on both aerial and underwater real images. However, the aerial image model introduces unrealistic background colors. Fabbri et al. (2018) proposed UGAN, which uses CycleGAN generated paired datasets and a Pix2Pix-like structure for UIE. However, CycleGAN generates artifacts under certain scenarios. Despite the requirement of high-quality training data, their proposed method struggles with low-quality underwater images. These methods effectively restore color but often face challenges such as over-enhanced contrast, information loss, instability, and convergence difficulties.

Convolutional neural network (CNN)-based methods (Wang et al., 2021; Lyu et al., 2022; Yang et al., 2023) are particularly effective for UIE tasks owing to their strong feature extraction capabilities and nonlinear feature mapping, which enable them to adapt to various underwater scenes. Wang et al. (2017) designed an end-to-end CNN-based network for color correction and deblurring by employing a pixel disturbance strategy to improve model convergence speed and accuracy. However, their method overfocuses on local features while neglecting the overall semantic information, global color, and light–shadow relationships in the image. Li et al. (2019) developed a paired underwater image enhancement benchmark (UIEB) dataset and proposed Water-Net, a CNN-based model that serves as a benchmark for CNN applications in UIE. Li et al. (2020) trained their proposed UWCNN on synthetic underwater images of various scenes, which resulted in different model parameters. However, owing to the singularity of the training data scenes, the model is overly sensitive to subtle changes in underwater environments and thus, performs poorly. Islam et al. (2020) proposed the UFO-120 dataset and a residual nested CNN called Deep SESR, which has a multimodal objective function for both enhancement and super-resolution of images. However, the shared feature space in this model can cause significant features from the super-resolution task to interfere with the color performance of image enhancement. The aforementioned CNN-based models have powerful feature-learning capabilities and can adapt to complex underwater environments; however, CNNs primarily extract features through local receptive fields, which renders fully capturing global information challenging. Consequently, enhanced images often show a marked locality with coordination problems among objects in complex underwater environments.

To address this limitation, several studies have used Swin Transformers (Liu et al., 2021) for UIE. Sun et al. (2022) enhanced the underwater image contrast by inputting images into

a Swin Transformer following gamma and white balance corrections. However, white balance and gamma correction cannot fully resolve the complex problems of underwater images, particularly in foggy and blurry scenes. Peng et al. (2023) constructed a large-scale underwater image dataset and proposed a channel multi-scale fusion transformer and spatial global feature transformer to enhance severely attenuated color channels and spatial regions. However, the sensitivity of different color spaces to various colors varies, which degrades model stability in scenarios with strong color contrast. Transformer architecture, with its unique mechanisms and processing methods, has tremendous potential and value as a primary framework for UIE. Zhu et al. (2024) proposed an adaptive multi-scale image fusion cascaded neural network that integrates polarization-based multi-dimensional features to improve image enhancement quality under low-quality imaging conditions. Concurrently, the team establishing a standardized evaluation framework for polarization-aware visual restoration algorithms. Zhu et al. (2025) proposed a Fourier-guided dual-channel diffusion network, enhances underwater images via phase-based edge refinement and amplitude mapping, coupled with a lightweight transformer denoiser, outperforming leading methods in generalization and visual quality on real underwater datasets. Wang et al. (2025) proposed a SAM-powered framework for underwater image enhancement, integrating precise foreground-background segmentation, region-specific color correction, adaptive contrast enhancement, and high-frequency detail reconstruction to mitigate crosstalk and blurring, thereby significantly improving restoration fidelity and visual quality. Considering that underwater images exhibit inconsistent attenuation characteristics across different color channels and spatial regions and that the object edges in these images degrade, the proposed network focuses on these characteristics to restore underwater image information and achieve high-quality underwater image data.

The main contributions of this paper are as follows:

1. We propose a network model, CUG-UIEF, based on U-Net and a multi-feature cross-fusion module, which greatly improves the quality of underwater images.
2. We introduce a multi-feature cross-fusion module that enhances the feature representation of images at different scales, thereby improving the overall quality and accuracy of the final output.
3. We evaluate the proposed CUG-UIEF model on the UIEB, low-light and super-resolution underwater image (LSUI), and U45 datasets and compared its performance with that of six other mainstream models. The experimental results show that CUG-UIEF achieves substantial improvements in the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). The results also demonstrated excellent performance in both underwater image quality metrics and underwater color image quality assessments, indicating that the CUG-UIEF effectively overcomes underwater environmental interference and can be applied in related fields.

2 Proposed method

2.1 Network structure

The overall structure of the CUG-UIEF is shown in Figure 1; it can be divided into three parts: an encoder, a multi-feature cross-fusion module (DDEM), and decoder. The encoder converts the input image into a deep feature representation. The decoder gradually fuses the features and performs upsampling to reconstruct an underwater image. In this study, the multi-scale features extracted by the encoder were input into the DDEM, and its output was fused with the upsampling results at each stage of the decoder. An enhanced underwater image was obtained after the final upsampling step.

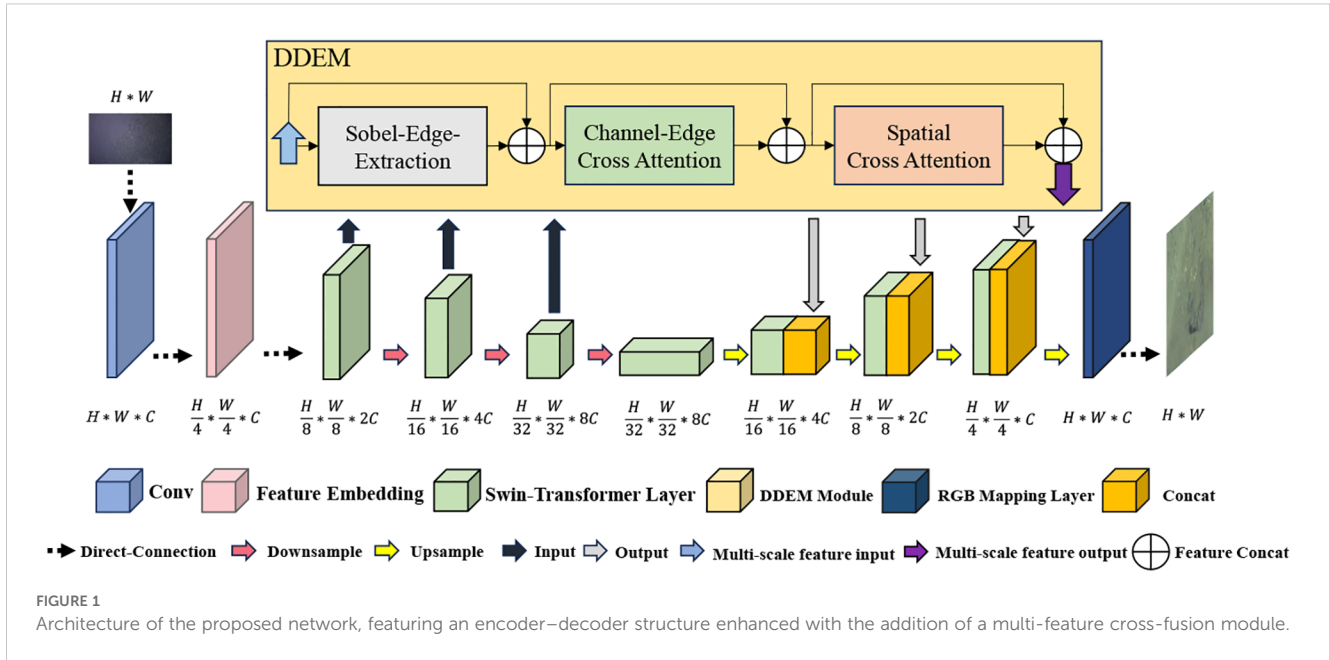
Encoder stage: This module extracts multi-scale features through the Swin Transformer layer and performs downsampling to capture the details and global information in the image. The deep feature representation provides rich semantic information for the subsequent DDEM module and decoder, which facilitates the final image reconstruction and enhancement.

DDEM module: Uses the Sobel operator to extract edge information from the multi-scale features extracted in the four stages of the encoder and inputs the edge and multi-scale features together into the channel cross-attention (CCA) module to fuse the feature information across channels. Subsequently, the output of the CCA is passed to the spatial cross-attention module to capture the long-distance dependencies among the multi-scale features. Following layer normalization and GeLU activation, the final features are sent to the decoder to gradually restore the spatial resolution and reconstruct the enhanced image.

Decoder stage: The decoder first upsamples the output of the final stage of the encoder and inputs it into the Swin Transformer block. Subsequently, the output of the DDEM is fused with the upsampled results of each decoder stage. The decoder restores the spatial resolution through gradual upsampling to reconstruct an enhanced image. The parameters of the Swin Transformer layer are adjusted at this stage to maintain the integrity of the features, whereas the upsampling layer is used to restore the size of the feature maps. The final upsampling restores the features to the resolution of the original input and projects them onto the RGB channels through the convolutional layer to generate an enhanced underwater image.

2.2 Multi-feature cross-fusion module (DDEM)

The proposed multi-feature cross-fusion module fuses the features extracted from the four multi-scale encoder stages (Figure 2). It generates enhanced feature representations and connects these enhanced features to the corresponding decoder stages. The module can be further divided into attention-guided edge fusion and spatial information enhancement modules.

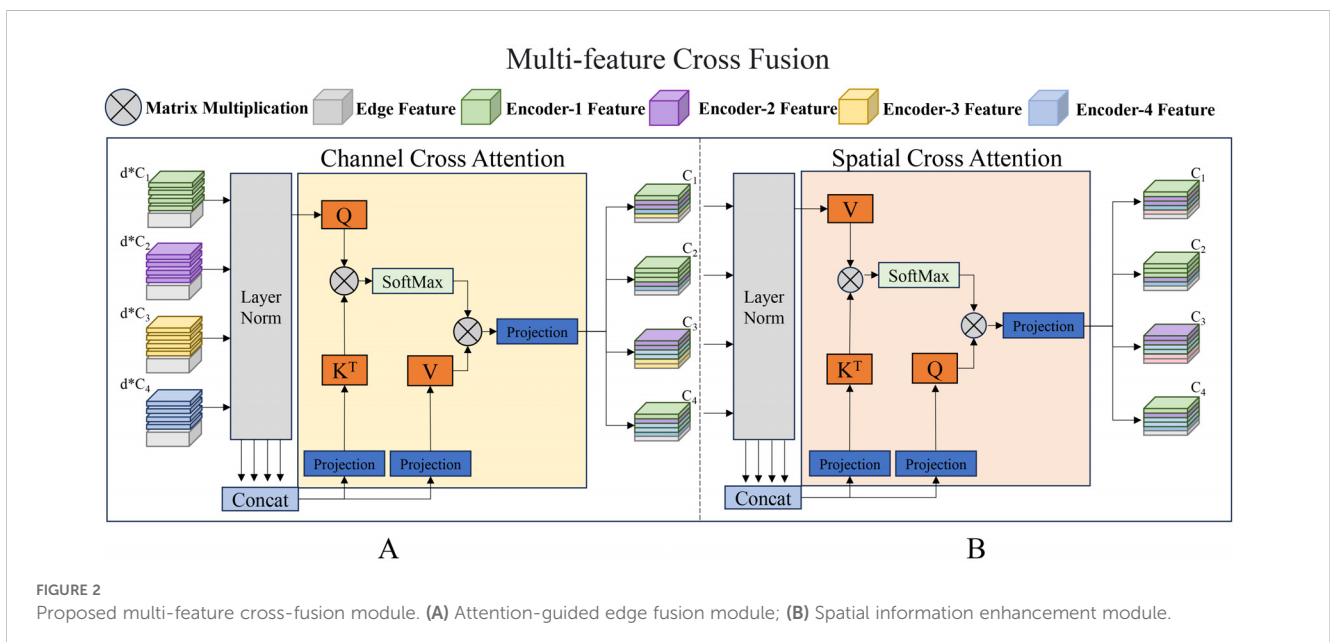


The specific operations performed by the module are as follows: The edge features are extracted from the multi-scale features output by the encoder and then input layer by layer together with the multi-scale features into the edge fusion and spatial information enhancement modules. Attention maps are constructed by fusing the features of the multi-scale encoder, enabling them to capture long-distance dependencies across different stages to achieve more accurate and comprehensive modeling of complex scenes and dynamic changes.

Through this series of operations, the output results are processed by layer normalization and subjected to nonlinear mapping via the GeLU activation function to establish dynamic correlations between the feature maps at different levels and edge feature maps.

2.2.1 Attention-guided edge fusion module

This module promotes information interactions between features at different levels in the channel dimension. In this study, the edge features are gradually fused in multiple stages. The weights of the weighted edge features are adjusted using CCA to ensure that detailed information, such as colors and textures, can be accurately transmitted. Upon being output to the decoder stage, as the decoding process proceeds, the weighting coefficients are dynamically adjusted based on the local information of the image; thus, the edge features are enhanced in detailed areas while minimizing interference in the background or smooth regions. In this manner, the edge information is strengthened in key areas (such as object boundaries and detailed parts), while the global consistency and natural appearance of the image are preserved.



The Sobel operator is applied for edge feature extraction. It identifies edge information by calculating the gray gradient of the area around each pixel in the image. The core of this algorithm lies in its elaborately designed convolution kernels, which perform convolution operations on the horizontal and vertical features of the image, thereby effectively capturing the edge changes in the image in different directions.

The change in the x-axis direction in the Sobel operator is

$$\mathbf{G}_x = \begin{pmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{pmatrix}$$

The change in the y-axis direction is

$$\mathbf{G}_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{pmatrix}$$

Approximate gradient values of the image in the horizontal and vertical directions can be obtained by performing convolution operations on the image using these two sets of convolution kernels. The gradient magnitude of each pixel point can then be obtained by calculating the square root of the sum of the squares of these two gradient values (or the sum of their absolute values) to determine the intensity of the edge.

$$\mathbf{G} = \sqrt{\mathbf{G}_x^2 + \mathbf{G}_y^2}$$

The Sobel operator extracts edge information across different scales, performs layer normalization along the channel dimension, and conducts weighted fusion with the multi-scale feature output from the encoder stage.

$$\mathbf{fused}_{\text{feature}} = \alpha \cdot \mathbf{edge}_{\text{feature}} + (1 - \alpha) \cdot \mathbf{encoder}_{\text{feature}}$$

Subsequently, the fused features automatically adjust the attention distribution by calculating the similarity between channels to strengthen key features in the image. At this stage, layer normalization is first performed on each token to stabilize the training process. Subsequently, all tokens are concatenated along the channel dimension to create unified keys and values while retaining each token as an independent query. The linear projection in the self-attention mechanism is replaced with a 1×1 depthwise convolutional projection. This enables cross-channel information integration and interaction and enhances its nonlinear characteristics. The process formula is as follows:

$$\mathbf{K}, \mathbf{V} = \mathbf{conv1D}(\mathbf{concat}(\mathbf{T}_1, \mathbf{T}_2 \dots \mathbf{T}_i))$$

$$\mathbf{Q}_i = \mathbf{conv1D}(\mathbf{T}_1, \mathbf{T}_2 \dots \mathbf{T}_i)$$

$$\mathbf{CCA}(\mathbf{Q}_i, \mathbf{K}, \mathbf{V}) = \mathbf{Softmax}(\mathbf{Q}_i^T \mathbf{K} \mathbf{S}) \mathbf{V}^T$$

\mathbf{Q}_i, \mathbf{K} , and \mathbf{V} are matrices that represent the queries, keys, and values, respectively, which are obtained by concatenating the tokens along the channel dimension. \mathbf{S} is the scaling factor. Once the output of the CCA is connected to the original tokens, the enhanced features are input into the spatial information enhancement module.

2.2.2 Spatial Information Enhancement Module

This module dynamically adjusts the feature weights of different regions of the image by calculating the correlations among different spatial positions, thereby enhancing the key features in the image. Edge features provide essential structural cues for the image, enabling the model to focus more on the detailed areas of the image while reducing attention allocation to smooth areas, thus avoiding excessive enhancement. Combined with enhanced multi-scale features, the module can capture the details of the image at different levels, effectively restoring the detail loss in underwater images caused by light attenuation and blurring. At this stage, all the tokens are first subjected to layer normalization along the channel dimensions and then concatenated. In contrast to the edge-fusion module, this module uses concatenated tokens as queries and keys; each token is used as a value. Moreover, 1×1 depthwise convolutions are also used for projection onto the queries, keys, and values. This design enables the spatial information enhancement module to focus on information integration in the spatial dimension, thereby complementing the edge fusion module to collaboratively establish a comprehensive and enhanced feature representation. The process is defined by the following formulas:

$$\mathbf{K}, \mathbf{Q} = \mathbf{conv1D}(\mathbf{concat}(\mathbf{T}_1, \mathbf{T}_2 \dots \mathbf{T}_i))$$

$$\mathbf{V}_i = \mathbf{conv1D}(\mathbf{T}_1, \mathbf{T}_2 \dots \mathbf{T}_i)$$

$$\mathbf{SCA}(\mathbf{Q}_i, \mathbf{K}, \mathbf{V}) = \mathbf{Softmax}(\mathbf{Q}_i^T \mathbf{K} \mathbf{S}) \mathbf{V}_i^T$$

\mathbf{Q}, \mathbf{K} , and \mathbf{V}_i are matrices that represent queries, keys, and values, respectively. \mathbf{S} is the scaling factor.

To ensure that the generated enhanced features can effectively serve the decoder, the following processing steps are adopted. First, layer normalization and the GeLU activation function are applied to the output to stabilize the features and introduce nonlinear transformations. Subsequently, through a combined sequence of an upsampling layer, a 1×1 convolution, batch normalization, and GeLU activation function, necessary size adjustments and enhancements are made to the features, which are fused with the features in the decoder stage. The upsampling layer is used to restore the spatial resolution of the feature maps, ensuring that the details of the image can be better reconstructed in the decoding stage. The 1×1 convolution is used for channel compression and feature fusion, enhancing the expressive ability of the model, while batch normalization ensures the consistency of features among different layers. The GeLU activation function introduces nonlinear transformations that aid in handling complex feature relationships. This method ensures the continuity and consistency of information and greatly improves the decoding efficiency and performance of the entire network.

2.3 Loss function

We propose a multi-dimensional perceptual loss function for training the CUG-UIEF to align the enhanced images with human visual perception and improve detail reconstruction.

1) Perceptual Loss.

Deep features capture high-level semantic information from images. By comparing the feature maps of the two images in a pretrained network, their perceptual similarity can be evaluated.

$$I_{\text{per}} = |x - y|,$$

x represents the predicted image, and y represents the real image.

2) Multi-scale Structural Similarity Loss.

Multi-Scale Structural Similarity (MSSSIM) is an image quality assessment metric that evaluates brightness, contrast, and structural features across multiple scales, providing a measure more aligned with human visual perception.

$$I_{\text{ms-ssim}} = 1 - \prod_{m=1}^M \left(\frac{2u_p u_g + c_1}{u_g^2 + u_p^2 + c_1} \right)^{\beta_m} \left(\frac{2\sigma_{pg} + c_2}{\sigma_p^2 + \sigma_g^2 + c_2} \right)^{\gamma_m}$$

Here, M represents different scales. u_g and u_p represent the means of the predicted image and ground truth, respectively. σ_p and σ_g represent the standard deviations between the predicted and real images. σ_{pg} represents the covariance between the predicted and real images. β_m and γ_m represent the relative importance constants between the two items. c_1 and c_2 are constants.

3) Charbonnier Loss.

The Charbonnier loss function is a variant of the L1 loss function. It prevents the denominator from reducing to zero by introducing a small positive number ϵ and ensures smoother changes when the gradient is large. It maintains the sharpness of an image while reducing noise.

$$I_{\text{charbonnier}} = \sqrt{x^2 + \epsilon^2}$$

X represents the difference between the predicted image and ground truth. ϵ is a small positive number used for numerical stability.

Finally, the loss function is expressed as

$$I = \lambda_1 I_{\text{per}} + \lambda_2 I_{\text{ms-ssim}} + \lambda_3 I_{\text{charbonnier}}.$$

Hyperparameters λ_1 , λ_2 , and λ_3 determine the balance between the overall performance and the local texture details. Following experimental analysis the parameters were set to 1, 2, and 1, respectively.

3 Experiments and analyses

3.1 Experimental environment and parameter settings

The proposed model was implemented using PyTorch 2.4.0. It was trained on an NVIDIA RTX 2080Ti GPU without a pretrained network. During the training process, the Adam optimizer was adopted, and the initial learning rate was set to 0.0005, with the β parameter pair being (0.9, 0.999). Training was performed for 700 epochs, and the number of samples in each batch was four.

3.2 Datasets

This study uses three datasets.

1. UIEB Dataset (Li et al., 2019): This dataset included 950 real underwater images. Among them, 890 images had corresponding reference images, and another 60 underwater images without reference images were used as the challenging data. In this study, 90 pairs of challenging images in multiple scenes with corresponding reference images from the UIEB were selected as the test set Test-U90, and 60 images without reference images were used as the test set Test-C60. The remaining images were divided into training and validation sets at a 8:2 ratio. The training data were enhanced using random cropping, size adjustment, and random rotation.
2. LSUI Dataset (Peng et al., 2023): This is a large-scale underwater image dataset that contains 5,004 underwater images with reference images. It contains richer underwater scenes. Forty-five images were selected from this dataset as the test set, Test-L45. The remaining images were divided into training and validation sets at a 8:2 ratio. The training data were enhanced through random cropping, size adjustment, and random rotation.
3. U45 Dataset: The U45 dataset is a publicly available underwater image test dataset that contains 45 underwater images in different scenes and involves underwater degradation phenomena, such as color shift, low contrast, and foginess. Forty-five images were used as the test set, Test-U45.

3.3 Evaluation metrics and comparative algorithms

Reference Evaluation Metrics: To quantify the performance of each model on the dataset with reference images, this study adopted two measurement standards: PSNR and SSIM. These two indicators help measure the similarity between the restored and reference images. PSNR is an objective quality metric calculated based on the mean squared error between the original image and the enhanced image, with the unit of decibel (dB). In UIE, a higher PSNR value indicates that the enhanced image has a smaller error than the original image and, thus, better quality. SSIM is an index used to measure the similarity between two images. It considers luminance, contrast, and structural information, and its value ranges from -1 to 1 . In UIE, the closer the SSIM value is to one, the more similar the enhanced image to the original image in terms of structure, luminance, and contrast, suggesting a higher image quality.

No-reference Evaluation Metrics: For the test sets of images without reference images, we adopted three evaluation methods: underwater color image quality evaluation (UCIQE), underwater image quality measure (UIQM) and Underwater Ranker (URanker)

(Guo et al., 2023). UCIQE focuses on the color density, saturation, and contrast of images and uses a linear combination of these three aspects as the quantitative form of color cast, blurring, and low contrast. UIQM includes color (UICM), sharpness (UISM), and contrast measurements (UIConM). As the scores of these methods increase, the image processing results become more aligned with the visual perception preferences of human beings.

Comparative Algorithms: The comparative algorithms adopted in this experiment are representative algorithms among traditional UIE methods and deep-learning-based UIE methods, which can verify the effectiveness and advancement of the proposed method, including the UIE algorithm based on color correction: Fusion (Ancuti et al., 2012); UIE algorithms based on image restoration: IBLA (Wang et al., 2013); HL (Berman et al., 2021); WWPF (Zhang et al., 2023); CBLA (Jha and Bhandari, 2024); UIE algorithms based on deep learning: UWCNN (Li et al., 2020), Shallow-UWnet (Naik et al., 2021), USUIR (Fu et al., 2022), URSCT (Ren et al., 2022), DiffWater (Guan et al., 2023).

3.4 Experimental results

All experimental results are presented with the best outcomes bolded and the second-best outcomes highlighted in blue font. This section first presents the test results of the model based on the UIEB training set on the Test-U90 dataset. As indicated in Table 1, CUG-UIEF outperformed the other algorithms in terms of the PSNR and SSIM. Moreover, compared to the second-best performance, CUG-UIEF achieved percentage gains of 8.41% and 0.1% in PSNR and SSIM, respectively. This study also conducted a no-reference evaluation comparison of Test-C60 and Test-U45. Table 2 presents all the statistical results. Both UIQM and UCIQE have specific feature biases and are relatively sensitive to the contrast of images. Therefore, results based on visual priors and physical models can yield higher scores. Our experimental results align with this conclusion. And the proposed method achieves the best performance on the URanker evaluation metric, with an average improvement of 12.21% over the second-best model. Therefore, the results cannot indicate whether the processed images are the best in all aspects. However, by combining the results of the two parameters, the images performed well in terms of contrast and color. CUG-UIEF obtained the second-best result among the models that were used in the experiment, only lower than that of the fusion method. Combined with the previous results, this shows that the generalization ability and actual performance of the CUG-UIEF are the best.

3.5 Comparative mechanism analysis of algorithms

The fusion algorithm addresses underwater color cast and low-contrast degradation through adaptive weight mapping, yet exhibits critical limitations when confronting specific technical challenges. Its

TABLE 1 PSNR and SSIM scores of different methods on the test set Test-U90.

	Method	TEST-U90	
		PSNR	SSIM
Traditional Method	HL	14.8429	0.6497
	IBLA	14.9395	0.6742
	Fusion	21.1843	0.8639
	CBLA	15.2359	0.6614
	WWPF	18.5371	0.7062
Deep-Learning Method	Sha-UWnet	17.4575	0.7174
	UWCNN	15.4532	0.7560
	USUIR	20.5514	0.8544
	URSCT	22.5976	0.9171
	Diff-Water	20.1567	0.8391
	Ours	24.4952	0.9262

In the results, boldface indicates the best data and blue denotes the suboptimal data.

edge restoration capability deteriorates in low signal-to-noise ratio (SNR) regions, producing blurred textures and artificial transitions around fine structural details, while contrast optimization remains suboptimal under non-uniform illumination caused by suspended particulates. Furthermore, the water-quality-dependent input generation mechanism demonstrates unstable color correction performance across chromatic water types, particularly failing to compensate for wavelength-specific absorption in turbid greenish waters where waterborne noise amplifies color inconsistency along depth gradients. These limitations stem from the algorithm's inherent constraints in decoupling overlapping degradation patterns and adapting to spatially variant underwater optical conditions.

The IBLA algorithm decomposes images via luminance-ordering error metrics and bright-pass filtering to separately regulate reflectance and illumination, dynamically adjusting their weights through dual-logarithmic transformations. While effective for uniform scenes, the framework suffers from edge-texture mismatches in areas with overlapping illumination-reflectance gradients, where low-SNR conditions exacerbate erroneous boundary segmentation and nonlinear illumination transitions degrade fine details. The logarithmic weight adaptation further struggles to resolve high dynamic range conflicts, causing halo artifacts near specular highlights and contextual inconsistency in shadowed low-contrast regions. These limitations arise from inadequate noise-robust disentanglement of radiometrically coupled components under complex degradation patterns.

The HL algorithm frames color restoration as a single-image dehazing task by estimating attenuation ratios for the blue-red and blue-green color channels, with a color distribution screening mechanism to identify optimal parameter combinations. However, this approach faces three critical limitations in addressing underwater-specific degradation: Its unified attenuation coefficient oversimplifies spectral interactions, failing to resolve edge blurriness

TABLE 2 UIQM and UCIQE scores of different methods on test sets C60 and U45.

	Method	Test-C60			Test-U45		
		UCIQE	UIQM	URanker	UCIQE	UIQM	URanker
Traditional Method	HL	0.5311	2.8774	0.094	0.5126	1.9423	0.751
	IBLA	0.5642	3.3236	0.815	0.4612	1.2768	0.945
	Fusion	0.5848	2.8092	0.745	0.6473	1.6984	0.726
	CBLA	0.4781	2.4273	1.285	0.5139	1.7141	1.392
	WWPF	0.5135	2.4861	1.348	0.5641	1.7311	1.351
Deep-Learning Method	Sha-UWnet	0.4198	2.2751	0.921	0.4595	1.6893	1.257
	UWCNN	0.4894	2.4523	1.687	0.4524	1.4338	1.582
	USUIR	0.5673	2.3234	1.618	0.5131	1.8952	1.685
	URSCT	0.5529	2.7453	1.713	0.5729	2.1861	1.724
	Diff-Water	0.5372	2.5894	1.632	0.5338	2.0142	1.583
	Ours	0.5737	2.8168	1.982	0.5937	2.3247	1.859

In the results, boldface indicates the best data and blue denotes the suboptimal data.

caused by wavelength-dependent scattering anisotropy. The channel-agnostic model amplifies noise in low-SNR scenarios, particularly in red-dominated deep-water regions where backscatter varies disproportionately. Linear color compensation ignores depth-related contrast attenuation gradients, leading to inaccurate recovery in shaded seabed areas with nonlinear illumination decay. These simplifications fundamentally disregard the photometric complexity of real underwater environments, where multi-band light refraction and particulate scattering create spatially varying attenuation patterns.

The color-balanced locally adjustable (CBLA) algorithm targets underwater color distortion and contrast degradation through dual-space hierarchical enhancement, yet reveals critical vulnerabilities when addressing complex photometric interactions. Its RGB-space color restoration mechanism struggles to decouple chromatic shifts from suspended particulate backscattering in high-turbidity environments, occasionally overcompensating blue-green dominance while neglecting wavelength-specific absorption residuals. The CIELAB-space contrast optimization demonstrates limited adaptivity to illumination gradients across depth-varying scenes, where aggressive luminosity adjustments in localized regions may amplify noise patterns and induce halo artifacts near high-frequency textures. Furthermore, the separate processing pipelines for color correction and contrast enhancement fail to maintain spectral consistency in transitional zones between adjusted and unprocessed areas, particularly under abrupt optical density changes caused by marine snow or biological layers. These deficiencies originate from the method's sequential processing framework that insufficiently models the nonlinear coupling between wavelength attenuation and turbidity-induced light diffusion.

The weighted wavelet visual perception fusion (WWPF) method tackles underwater color distortion and contrast degradation through multi-strategy hierarchical optimization, yet reveals critical constraints when handling complex photonic

interactions. Its attenuation-map-guided color correction exhibits incomplete spectral separation in high-turbidity greenish waters, where particulate backscattering interferes with wavelength-specific absorption estimation, occasionally preserving residual cyan dominance while overcompensating red-channel artifacts. The maximum entropy optimized global contrast enhancement demonstrates limited dynamic range adaptation across depth-varying illumination fields, where uniform intensity stretching may amplify noise in low-transmission regions while compressing texture details in high-clarity zones. Furthermore, the wavelet-based multi-scale fusion mechanism shows inadequate edge preservation at high-frequency subbands when processing particulate-laden scenes, as directional filter banks struggle to differentiate between authentic structural contours and suspended particle clusters, resulting in oversmoothed textures near marine snow interfaces. These limitations stem from the method's implicit assumption of linear degradations and insufficient modeling of nonlinear light-particle-camera interactions in turbid aquatic environments.

The UWCNN algorithm constructs a synthetic degradation dataset using spectral-attenuation priors to train a lightweight CNN for direct underwater image restoration, thereby reducing error propagation. While effective for general color cast correction, its wavelength-agnostic framework introduces spectral bias by oversimplifying depth-dependent chromatic shifts and angular illumination variations inherent in real underwater environments. Specifically, the model fails to address nonlinear wavelength absorption caused by suspended particulates and depth-varying water types, leading to color channel imbalance in scenes with multi-spectral artificial lighting or bioluminescent interference. Furthermore, its static prior integration neglects photometric divergence between shallow and deep-water zones, resulting in inconsistent color constancy when reconstructing red-depleted regions or high-turbidity sediments. These limitations stem from inadequate modeling of spectrally asymmetric degradation and

cross-domain generalization across heterogeneous underwater optical conditions.

The Sha-UWnet employs a parameter-efficient architecture to optimize underwater image enhancement, leveraging prioritized feature extraction to balance computational cost and restoration quality. While its streamlined design effectively addresses global color shifts, the constrained network depth impedes hierarchical abstraction of multi-scale edge contexts, resulting in blurred boundary delineation and textural discontinuities in low-contrast turbid waters. Specifically, the shallow structure fails to resolve edge-texture conflicts caused by suspended particle scattering, often miscalculating gradient magnitudes in regions with overlapping foreground-background chromaticity. Furthermore, its limited receptive field struggles to suppress waterborne noise while preserving high-frequency details, leading to artificial sharpening artifacts near bioluminescent features or sediment-rich zones. These limitations highlight the inherent trade-off between model efficiency and multi-scale degradation disentanglement in underwater optical environments.

The USUIR algorithm reformulates unsupervised restoration through homology-driven cycle consistency between original and synthetically re-degraded images, theoretically circumventing the need for paired training data. While effective for global error minimization, the framework exhibits edge gradient confusion in low signal-to-noise ratio regions, failing to resolve sub-pixel boundary discontinuities caused by suspended particle scattering or nonlinear light attenuation. This manifests as blurred bio-structural contours and textural oversmoothing in turbid waters where foreground-background chromatic similarity exacerbates edge ambiguity. Furthermore, its spectrally insensitive homology constraints inadequately model wavelength-dependent absorption, inducing color channel crosstalk that amplifies greenish hue bias in deep pelagic zones and artificial saturation spikes under multi-spectral artificial lighting. These limitations stem from insufficient physical priors to disentangle spatially coupled degradation patterns across heterogeneous underwater optical domains.

The URST algorithm integrates Swin Transformer into a U-Net framework to enhance global context modeling for structural and chromatic restoration, while its RSCTB module employs convolutional layers to refine local features. Although this hybrid design improves cross-scale feature aggregation in uniform underwater scenes, the global attention mechanism in Swin Transformer induces boundary erosion when processing low-contrast edges or suspended particle-induced textures, where multi-scale edge ambiguity arises from nonlinear light scattering. Concurrently, the convolutional RSCTB module exhibits limited texture-edge decoupling capacity, failing to recover high-frequency boundary cues lost during transformer-based global smoothing, particularly in high-turbidity regions with overlapping bio-optical signals. This synergistic deficiency manifests as gradient reversal artifacts along complex seabed contours and chromatic offsets in shadowed areas, highlighting the algorithm's inadequate fusion of spectral-spatial priors to address depth-variant degradation patterns in dynamic underwater environments.

The DiffWater method addresses underwater color distortion and quality degradation through conditional diffusion modeling, yet demonstrates critical vulnerabilities when confronting nonlinear photometric interactions in complex aquatic scenarios. Its channelwise color compensation mechanism in RGB space shows incomplete chromatic separation in turbid greenish waters, where wavelength-dependent scattering interferes with particulate density estimation, occasionally preserving blue-green dominance while introducing artificial magenta casts in shadow regions. The conditional DDPM framework exhibits unstable noise prediction capabilities under dynamic illumination fields, where conditional guidance from color-compensated inputs may misdirect the denoising trajectory, generating texture-inconsistent hallucinated details near high-particle-concentration zones. Furthermore, the sequential integration of color correction and diffusion processes demonstrates spectral incoherence in transitional depth layers, particularly failing to preserve wavelength attenuation gradients when processing scenes with abrupt optical density changes caused by algal blooms or sediment plumes. These limitations stem from the method's simplified assumption of additive degradation patterns and insufficient physical modeling of the nonlinear correlation between waterborne light scattering and depth-dependent chromatic absorption.

The proposed UIE algorithm in this study employs edge feature attention fusion to address critical problems, such as edge blurriness, low SNR, and low contrast in underwater images. It integrates three innovative modules: (1) Edge operators extract edge information through gradient-sensitive feature learning, while CCA fuses multi-scale features using cross-channel coherence analysis, restoring object edge details by jointly optimizing high-frequency components and improving visual performance. (2) A spatial cross-attention mechanism strengthens spatial structure information via edge-guided attention propagation, preserving details under low signal-to-noise ratio conditions through noise-adaptive feature reinforcement. (3) A multi-dimensional perception optimization method enhances semantic understanding, structural integrity, and local contrast using frequency-aware adversarial learning, while mitigating the effects of outliers through multi-scale degradation disentanglement. Collectively, these modules establish hierarchical edge-texture synchronization, where edge restoration and feature fusion are systematically coordinated to resolve cross-scale degradation conflicts in turbid underwater environments.

3.6 Component ablation and fusion validation

The excellent performance of the CUG-UIEF proposed in this study for UIE mainly benefits from the multi-feature cross-fusion module and the redesigned loss function. To verify the effectiveness of the modules proposed in this study, we conducted ablation studies using the UIEB dataset as the training set on Test-U90 and by selecting 45 challenging images from the LSUI dataset as Test-L45.

The original model selected in this study has been described in the literature (Ren et al., 2022). The specific experimental settings were consistent with those used in a previous experiment. Table 3 presents the results of this index. DDEM-1 represents CUG-UIEF with the edge fusion module removed, and DDEM-2 represents CUG-UIEF with the spatial information enhancement module removed.

As indicated in Table 3, the proposed model achieved the best quantitative performance on the two test datasets, reflecting the effectiveness of the combination of multi-feature cross-fusion and multidimensional perceptual loss function modules.

As shown in Figure 3, compared with the original model, the added multi-feature cross-fusion module better address the problem of cyan-green color casts. When processing images, the cross-attention mechanism can adaptively focus on the interactions between cyan-green channels and other channels, avoiding excessive or insufficient utilization of cyan-green channel information. It emphasizes local details than on the convolution in the original model. The color distribution in real scenes was better matched adjusting the weights and contributions of the cyan-green channels in the image to a more reasonable level, thus effectively correcting the cyan-green color cast and improving the accuracy and naturalness of the image colors. Moreover, after adding edge features, the attention mechanism can focus on the structural information in the image and avoid wasting resources in unimportant areas. Following the addition of edge features, the details of the stones and creatures in the two comparison images became clearer. As we can observe in Figure 4, owing to the multidimensional perceptual loss function, the obtained images exhibit enhanced details, improved color restoration, vivid object edges, high contrast, and clear boundaries.

3.7 Qualitative comparison through visualization

First, the image comparison results of the UIEB dataset are presented. A comprehensive training was conducted using the UIEB dataset. The test data selected for this study were sampled according to six scenes with distinct characteristics: shadow, texture, blur, blue, yellow, and green. The images that best

TABLE 3 Statistical results of the ablation study on the modules and loss functions.

Module	Test-U90		Test-L45	
	PSNR	SSIM	PSNR	SSIM
Origin	23.2074	0.9178	21.9878	0.9164
DDEM-1	23.4076	0.9183	22.3455	0.9142
DDEM-2	25.5647	0.9195	23.2346	0.9234
CUG-UIEF	26.4693	0.9286	24.5212	0.9276
Loss	PSNR	SSIM	PSNR	SSIM
I_p	24.3572	0.9142	22.0478	0.9123
$I_p I_m$	25.7823	0.9212	23.5689	0.9201
$I_p I_m I_c$	26.4693	0.9286	24.5212	0.9276

In the results, boldface indicates the best data.

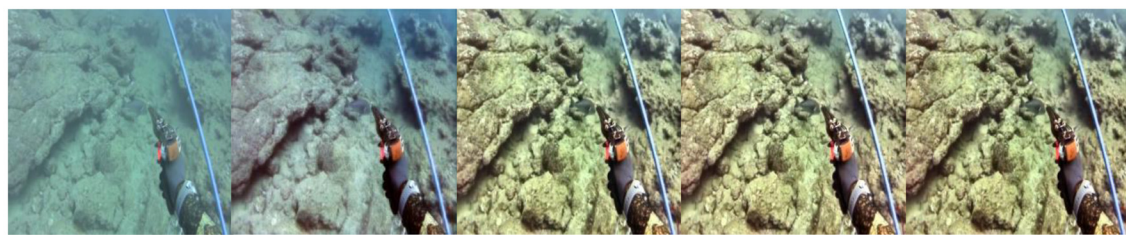
represented each type of scene and were to some extent challenging were chosen. Figure 5 presents the enhancement results of the different methods.

In the green scene, color casts occurred in the results of HL, Sha-UWnet, and UWCNN. Only Fusion, USUIR, URSCCT, and CUG-UIEF showed color restoration similar to that of the real image. However, CUG-UIEF achieved the closest color restoration effect to the real image, and the details in the shadowed parts were the clearest and most distinguishable. In the blue scene, except for USUIR, URSCCT, and CUG-UIEF, none of the methods restored the real illumination effect, and only CUG-UIEF truly restored the color texture of the fish in the upper left corner. In foggy and textured scenes, only the proposed URSCCT and CUG-UIEF achieved good effects in defogging and enhancing textures and object edges. However, compared to real situations, both methods had some deficiencies. In the final yellow scene, only CUG-UIEF retained delicate edge information during defogging. Overall, CUG-UIEF was visually superior to the other methods.

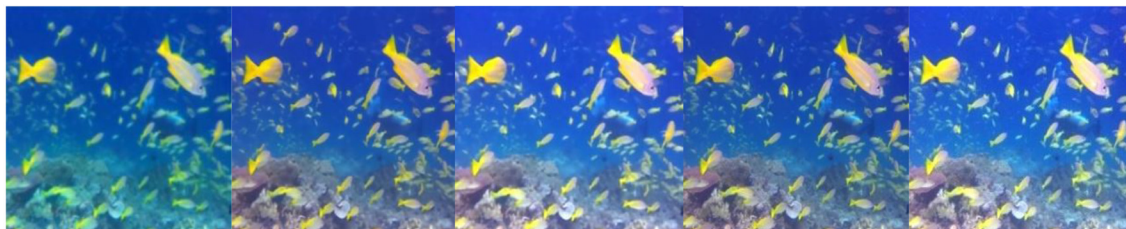
This section also presents the image results of CUG-UIEF on no-reference datasets. The tests were performed on two test sets, Test-C60 and Test-U45.

Test-C60 includes five underwater environments—red, yellow, green, blue, and foggy scenes—all of which were affected by high backscattering and color deviation. The most representative images of each type were selected for visual comparison. As shown in Figure 6, HL, CBLA, WWPF, Sha-UWnet, UWCNN, and URSCCT exhibited obvious color deviations in some cases. In the yellow scene, HL, IBLA, and USUIR restored the paddle blade to purple, whereas UWCNN and CUG-UIEF restored it to yellow, which is closer to the normal visual perception of humans. Moreover, CUG-UIEF can better restore blurred details in the original image. In the green and blue scenes, only URSCCT and CUG-UIEF achieved good restoration of the background and surfaces of the creatures. In the foggy scene, URSCCT had a significant defogging effect but overly enhanced the red color in the original image. CUG-UIEF attempted to retain the information of the original image while defogging, and the color restoration at the bottom background was more in line with normal perception. In the shadow and texture scenes, all the methods except USUIR, URSCCT, and CUG-UIEF, exhibited color restoration deviations. These three methods could restore the details in the shaded parts while retaining the natural illumination, but only CUG-UIEF could retain sufficient light–dark contrast and object details while providing improved color for the seawater background.

TEST-U45 contains multiple scenes, such as color deviation and foggy scenes. Multiple scenes were selected for the experiments, and representative scenes were selected for display. As is shown Figure 7, except for HL, UWCNN, CBLA, WWPF and Sha-UWnet, the methods exhibited a lower degree of color deviation. In the shadow and texture scenes, Fusion, USUIR, URSCCT, and CUG-UIEF performed well in color restoration and texture information preservation. However, in the blue scene, only the URSCCT and CUG-UIEF restored colors that were more in line with normal visual perception and preserved the texture information of the objects well. In the green scene, only CUG-UIEF could better reflect the natural illumination environment and delicate details.



A



B

FIGURE 3

Ablation study on the contribution of cross-fusion. Each panel includes the original image (RAW), the results of the original method, the results of DDEM-1, the results of DDEM-2, and the results of CUG-UIEF. (A) Test-U90, (B) Test-L45.

In summary, HL, UWCNN, WWPF and Sha-UWnet were prone to color-cast phenomena. The IBLA improved the quality of underwater images using local adaptive methods but performed poorly in yellow, foggy, and some blue scenes. Fusion greatly increased artificial colors to enhance contrast but could not adapt

well to the changes caused by foggy scenes. USUIR performed well in most scenes but often exhibited a red-shift phenomenon in blue scenes. URSCCT had good robustness and strong defogging ability; however, when restoring objects in foggy scenes, it was prone to red-shift phenomena. CUG-UIEF had good robustness and



A



B

FIGURE 4

Multi-dimensional ablation study. Each panel includes the original image (RAW) and the results of the original method. The results of using only L_p , the results of using both L_p and L_m , and the results of using L_p , L_m , and L_c simultaneously. (A) Test-U90, (B) Test-L45.

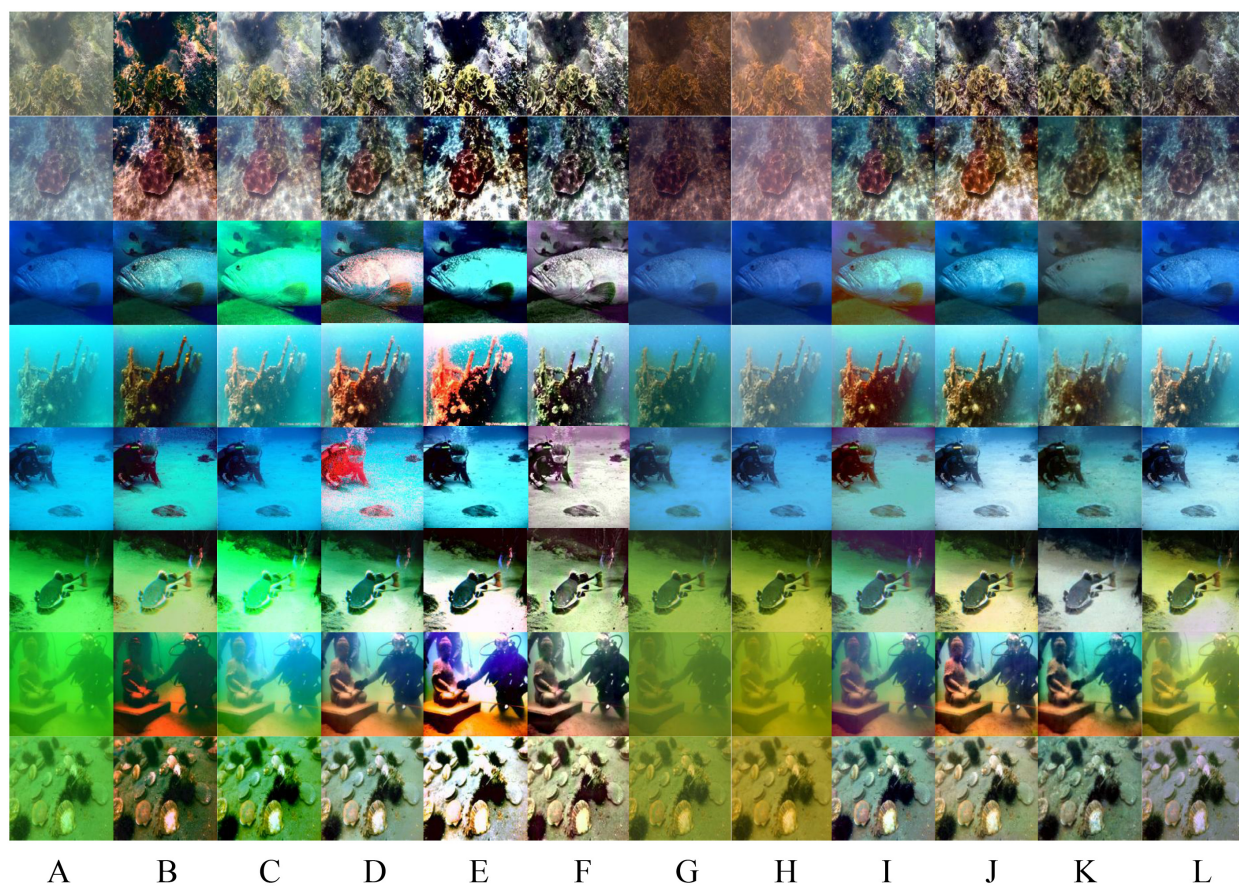


FIGURE 5
 Comparison of the underwater images sampled from the Test-U90 dataset. (A) RAW (B) HL (C) IBLA (D) Fusion (E) CBLA (F) WWPE (G) Sha-UWnet (H) UWCNN (I) USUIR (J) URSCT (K) Diff-Water (L) Ours (M) Ground truth.

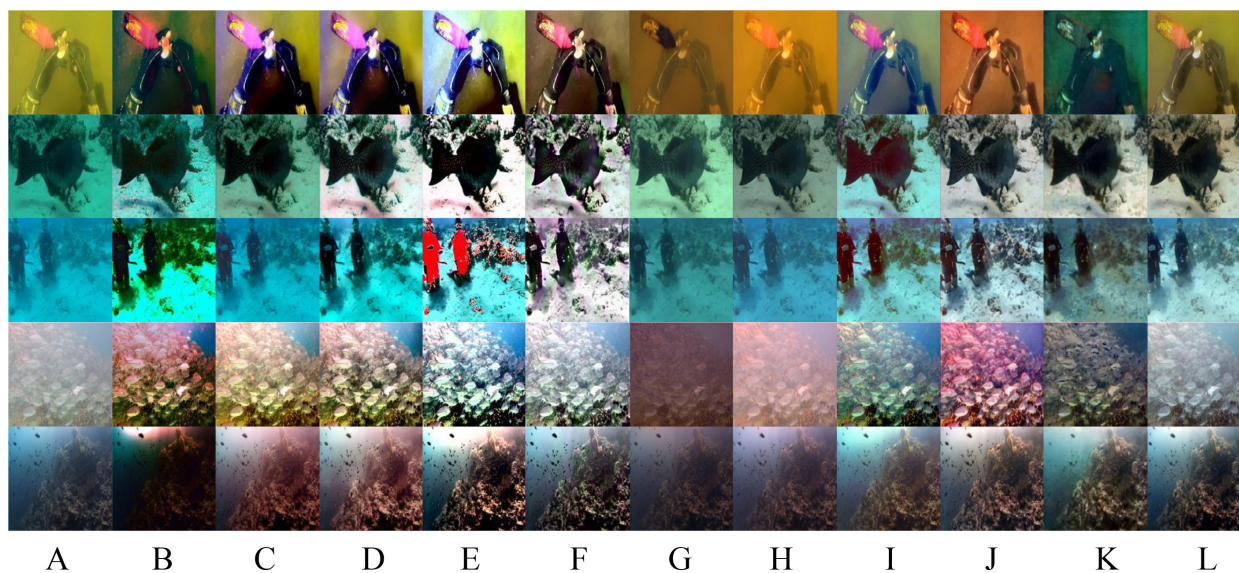


FIGURE 6
 Visual comparison of the underwater images sampled from the Test-C60 dataset. (A) RAW (B) HL (C) IBLA (D) Fusion (E) CBLA (F) WWPE (G) Sha-UWnet (H) UWCNN (I) USUIR (J) URSCT (K) Diff-Water (L) Ours.

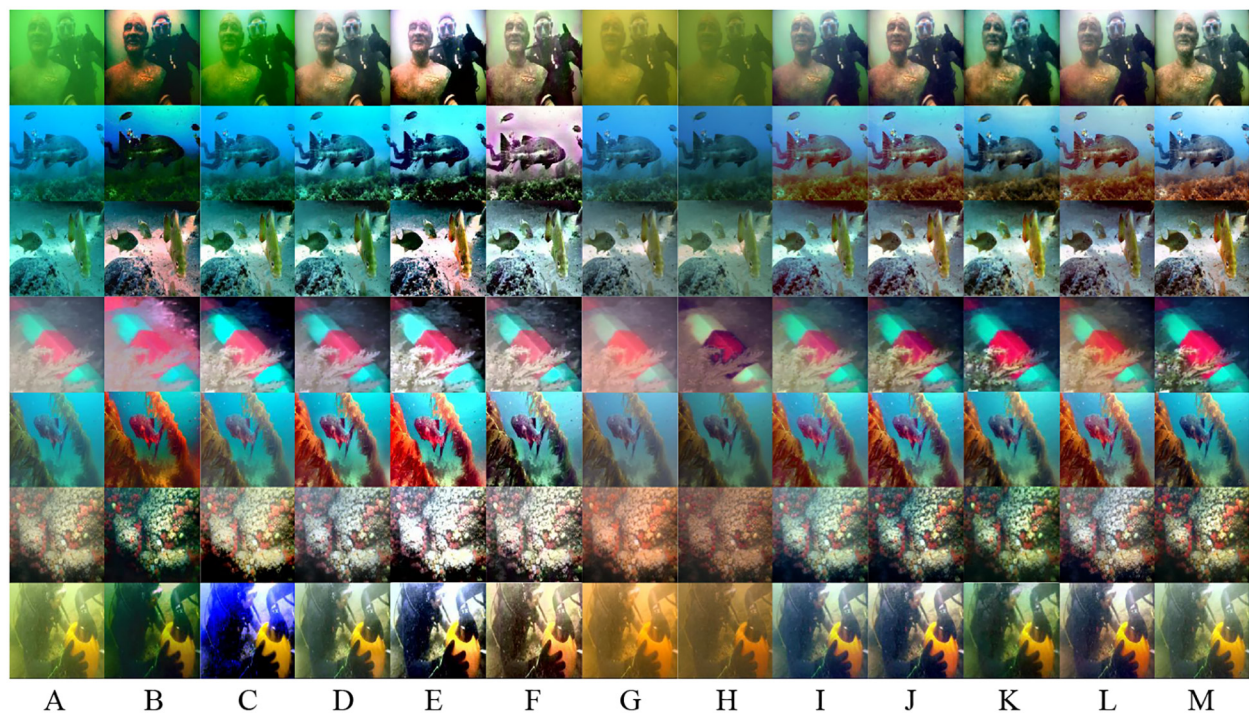


FIGURE 7

Visual comparison of the underwater images sampled from the Test-U45 dataset. (A) RAW (B) HL (C) IBLA (D) Fusion (E) CBLA (F) WWPE (G) Sha-UWnet (H) UWCNN (I) USUIR (J) URSCT (K) Diff-Water (L) Ours.

performed well in yellow, blue, green, foggy, shadowed, and textured scenes.

3.8 Application and inference efficiency

As feature extraction-matching and edge detection constitute core technical pillars in underwater image analysis, this study systematically validated the necessity of the proposed method as a preprocessing module for feature matching and edge detection. In feature matching tasks, the SIFT algorithm was utilized to extract feature points, complemented by the RANSAC algorithm for false match elimination. Feature matching was performed on preprocessed 256×256 pixels underwater stereo image pairs from the SQUID dataset. Figure 8 revealed that the proposed approach significantly optimized matching performance while concurrently improving visual quality. Table 4 demonstrates that compared to baseline methods, our scheme ranked second in both initial and valid matches, yet achieved the highest matching precision. Integrative qualitative-quantitative analyses corroborated the critical utility of this method for underwater feature matching tasks.

Regarding edge detection tasks, all images in the Test-C60 and Test-U45 datasets underwent enhancement prior to edge extraction and evaluation via the Canny operator. Detection performance was quantified using three metrics: Precision, F1 (harmonic mean of precision and recall), and Edge Pixel Ratio (EPR). Table 5 indicated that our method ranked first in accuracy and second in EPR relative to state-of-the-art approaches. Figure 9 demonstrates the

experimental findings: In Test-U45 fish samples, the enhanced edge detection preserves intact morphological contours while precisely discriminating target-background depth disparities, revealing underwater spatial hierarchy. In Test-C60 columnar targets, the algorithm achieves complete extraction of artificial structures' geometric edges with enhanced low-light gradient responses, where continuous seagrass blade edges further validate optical attenuation compensation. Convergent qualitative and quantitative evidence validated the significant contribution of this method to underwater edge detection tasks.

To evaluate the practical applicability of underwater image enhancement algorithms, we conducted a systematic comparison of inference efficiency among competing methods. The experiments were performed using the UIEB dataset as the benchmark, with average inference times calculated across all test samples. Traditional algorithms were executed in batch processing mode, while deep learning approaches employed pre-trained models on the UIEB training set for inference. Owing to significant architectural variations among deep learning algorithms, substantial discrepancies in inference times were observed across different models. As demonstrated in Table 6, conventional algorithms maintain absolute superiority in computational speed, whereas the proposed framework achieves the second-highest efficiency among deep learning methods while demonstrating a competitive advantage over structurally complex traditional approaches. These findings validate the proposed method's significant advantages in balancing computational complexity with practical deployment feasibility.

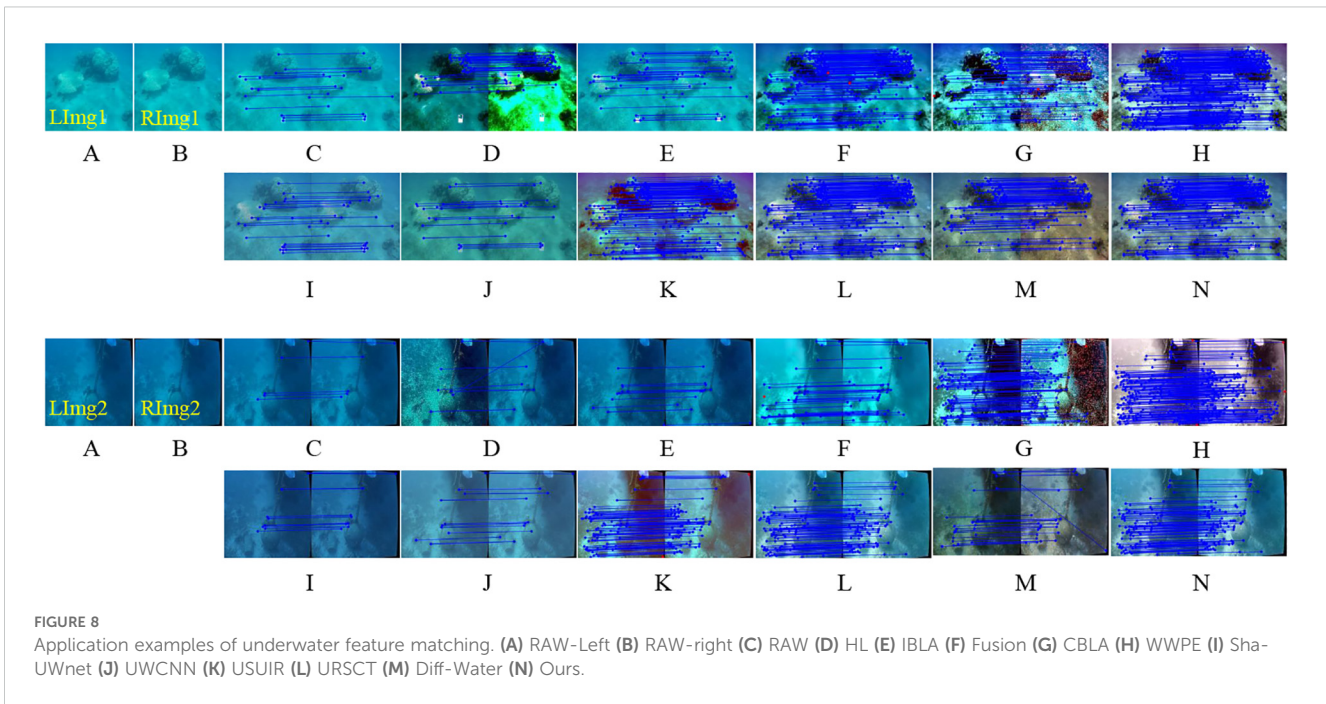


TABLE 4 Mean evaluation results of underwater feature matching on the SQUID dataset.

	Method	Initial matches	Valid matches	Precision
Traditional Method	HL	44.32	38.67	87.25%
	IBLA	37.42	31.73	84.79%
	Fusion	46.56	39.14	84.06%
	CBLA	87.35	78.15	89.46%
	WWPF	198.58	166.21	83.70%
Deep-Learning Method	Sha-UWnet	36.74	31.27	85.12%
	UWCNN	39.55	32.52	82.29%
	USUIR	152.46	135.14	88.64%
	URST	163.24	145.81	89.24%
	Diff-Water	108.47	97.09	89.51%
	Ours	172.68	155.81	90.23%

In the results, boldface indicates the best data and blue denotes the suboptimal data.

TABLE 5 Mean evaluation results of underwater feature matching on the Test-C60 and Test-U45 dataset.

	Method	Precision	F1	EPR
Traditional Method	HL	0.6011	0.1406	0.0303
	IBLA	0.6208	0.4789	0.1492
	Fusion	0.6149	0.3311	0.0926
	CBLA	0.6215	0.5435	0.2021
	WWPF	0.6571	0.5252	0.1888

(Continued)

TABLE 5 Continued

	Method	Precision	F1	EPR
Deep-Learning Method	Sha-UWnet	0.6666	0.0295	0.0057
	UWCNN	0.6523	0.3145	0.0649
	USUIR	0.6314	0.2227	0.0542
	URSCT	0.6289	0.4445	0.1436
	Diff-Water	0.6403	0.4474	0.1381
	Ours	0.6719	0.4761	0.1989

In the results, boldface indicates the best data and blue denotes the suboptimal data.

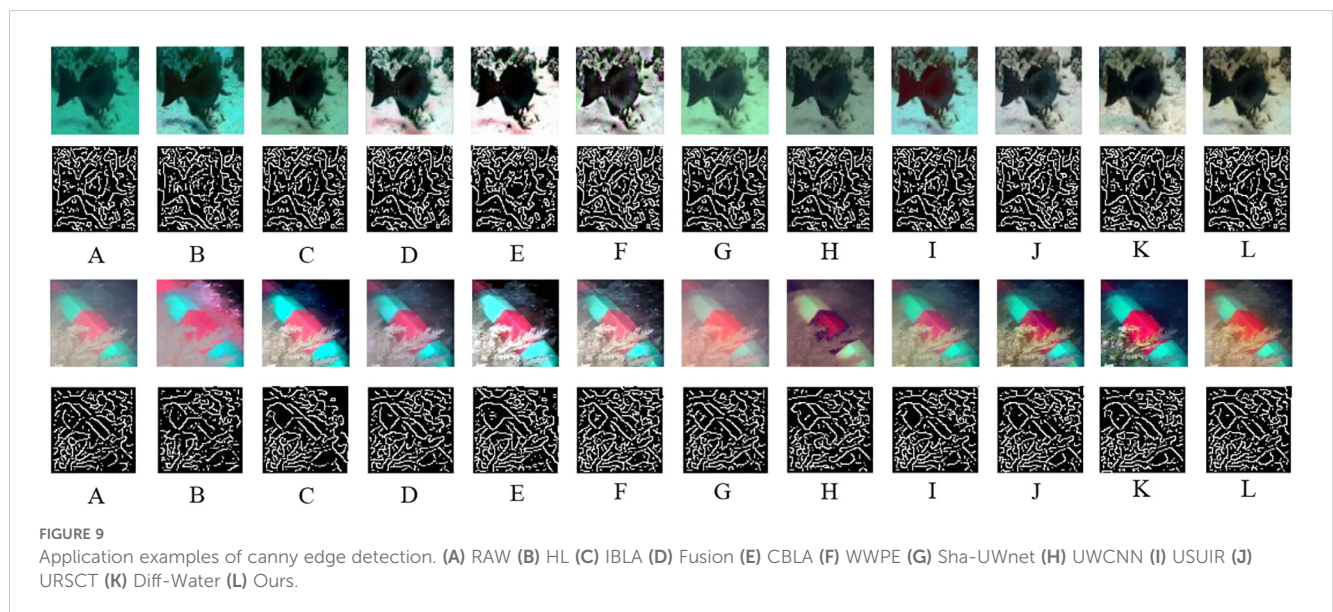


FIGURE 9 Application examples of canny edge detection. (A) RAW (B) HL (C) IBLA (D) Fusion (E) CBLA (F) WWPE (G) Sha-UWnet (H) UWCNN (I) USUIR (J) URSCT (K) Diff-Water (L) Ours.

TABLE 6 Inference Efficiency Comparison.

	Method	Per-image inference time
Traditional Method	HL	0.284s
	IBLA	0.622s
	Fusion	1.756s
	CBLA	0.199s
	WWPF	0.652s
Deep-Learning Method	Sha-UWnet	3.643s
	UWCNN	2.911s
	USUIR	1.876s
	URSCT	1.061s
	Diff-Water	44.322s
	Ours	1.083s

In the results, boldface indicates the best data and blue denotes the suboptimal data.

4 Conclusion

This study presented a deep learning model for UIE that improves blurring and color distortion caused by light scattering and attenuation. The proposed model integrates a multi-feature cross-fusion module, which combines edge features with encoder features and utilizes a channel-cross attention mechanism, effectively enhancing the clarity of blurred areas and improving edge detail capture. Additionally, the spatial information enhancement module strengthens feature interactions across different locations, enabling more natural restoration of color-distorted regions, thereby bringing the image closer to true colors and clarity. Through multi-dimensional perception optimization, the model further improves clarity, color accuracy, and edge details. Experimental results confirm the superior ability of the model to restore image details and correct color distortion. Ablation studies highlight the effectiveness of both the multi-feature cross-fusion module and multi-dimensional perception optimization in enhancing detail and overall color consistency. However, the

dehazing performance of the model in large-scale foggy underwater images requires further improvement. Future work will incorporate multispectral data to address the limitations, enhance dehazing performance, and improve the overall robustness and generalizability of the model in complex scenarios, ultimately providing more reliable image enhancement solutions for practical underwater operations and deep-sea exploration.

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 authors.

Author contributions

SS: Data curation, Methodology, Writing – original draft. WC: Supervision, Writing – review & editing, Methodology. HW: Data curation, Validation, Writing – original draft. PW: Data curation, Validation, Writing – review & editing. QL: Validation, Visualization, Writing – review & editing. XQ: Funding acquisition, Methodology, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was jointly

supported by the China Geology Survey Project (NO. 20241910244) and the Opening Fund of Key Laboratory of Geological Survey and Evaluation of Ministry of Education (No. GLAB2024ZR01).

Conflict of interest

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

Generative AI statement

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

Publisher's note

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

References

- Alsakar, Y. M., Sakr, N. A., El-Sappagh, S., Abuhmed, T., and Elmogy, M. (2024). Underwater image restoration and enhancement: A comprehensive review of recent trends, challenges, and applications. *Vis. Comput.* 1–49. doi: 10.1007/s00371-024-03630-w
- Ancuti, C., Ancuti, C. O., Haber, T., and Bekaert, P. (2012). "Enhancing underwater images and videos by fusion," in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Providence, RI, USA: IEEE), p. 81–88. doi: 10.1109/CVPR.2012.6247661
- Banik, P. P., Saha, R., and Kim, K.-D. (2018). "Contrast enhancement of low-light image using histogram equalization and illumination adjustment," in *2018 international conference on electronics, information, and communication (ICEIC)*, Vol. 2018. (Hawaii, USA: IEEE), p. 1–4. doi: 10.23919/ELINFOCOM.2018.8330564
- Berman, D., Levy, D., Avidan, S., and Treibitz, T. (2021). Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 2822–2837. doi: 10.1109/TPAMI.2020.2977624
- Dreus, P. L. J., Nascimento, E. R., Botelho, S. S. C., and Campos, M. F. M. (2016). Underwater depth estimation and image restoration based on single images. *IEEE Comput. Graph. Appl.* 36, 24–35. doi: 10.1109/MCG.2016.26
- 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, Australia: IEEE), p. 7159–7165. doi: 10.1109/ICRA.2018.8460552
- Fu, Z., Lin, H., Yang, Y., Chai, S., Sun, L., Huang, Y., et al. (2022). "Unsupervised underwater image restoration: From a homology perspective," in *Proceedings of the AAAI conference on artificial intelligence (AAAI)*, Vol. 36. (AAAI), 643–651. doi: 10.1609/aaai.v36i1.19944
- Garg, D., Garg, N. K., and Kumar, M. (2018). Underwater image enhancement using blending of CLAHE and percentile methodologies. *Multimedia Tool. Appl.* 77, 26545–26561. doi: 10.1007/s11042-018-5878-8
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. in *Advances in neural information processing systems*. Montreal, Quebec, Canada: 27(NIPS). doi: 10.5555/2969033.2969125.
- Guan, M., Xu, H., Jiang, G., Yu, M., Chen, Y., Luo, T., et al. (2023). DiffWater: Underwater image enhancement based on conditional denoising diffusion probabilistic model. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 17, 2319–2335. doi: 10.1109/JSTARS.2023.3344453
- Guo, C., Wu, R., Jin, X., Han, L., Zhang, W., Chai, Z., et al. (2023). "Underwater ranker: Learn which is better and how to be better," in *Proceedings of the AAAI conference on artificial intelligence (AAAI)*, Vol. 1. (Washington, DC, USA: AAAI), 702–709. doi: 10.1609/aaai.v37i1.25147
- Hu, K., Weng, C., Zhang, Y., Jin, J., and Xia, Q. (2022). An overview of underwater vision enhancement: From traditional methods to recent deep learning. *J. Mar. Sci. Eng.* 10, 241. doi: 10.3390/jmse10020241
- Islam, M. J., Luo, P., and Sattar, J. (2020). Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. In *16th Robotics: Science and Systems* (Corvallis, OR, USA: MIT Press Journals), 18–28. doi: 10.15607/RSS.2020.XVI.018
- Jha, M., and Bhandari, A. K. (2024). CBLA: color-balanced locally adjustable underwater image enhancement. *IEEE Trans. Instrum. Meas.* 73, 3396850. doi: 10.1109/TIM.2024.3396850
- Li, B., Chen, Z., Lu, L., Qi, P., Zhang, L., Ma, Q., et al. (2025). Cascaded frameworks in underwater optical image restoration. *Inform. Fusion.* 117, 102809. doi: 10.1016/j.inffus.2024.102809
- Li, C., Anwar, S., and Porikli, F. J. P. R. (2020). Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognit.* 98, 107038. doi: 10.1016/j.patcog.2019.107038
- 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, J., Skinner, K. A., Eustice, R. M., and Johnson-Roberson, M. (2017). WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robot. Autom. Lett.* 3, 1–1. doi: 10.1109/LRA.2017.2730363
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021). "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. (New York: IEEE), 9992–10002. doi: 10.1109/ICCV48922.2021.00986
- Lyu, Z., Peng, A., Wang, Q., and Ding, D. (2022). An efficient learning-based method for underwater image enhancement. *Displays* 74. doi: 10.1016/j.displa.2022.102174
- Mazzeo, A., Aguzzi, J., Calisti, M., Canese, S., Vecchi, F., Stefanni, S., et al. (2022). Marine robotics for deep-sea specimen collection: A systematic review of underwater grippers. *Sensors (Basel)*. 22, 648. doi: 10.3390/s22020648
- Naik, A., Swarnakar, A., and Mittal, K. (2021). Shallow-uwnet: Compressed model for underwater image enhancement (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence Online* Vol. 35. (AAAI), 15853–15854. doi: 10.1609/aaai.v35i18.17923
- Nazir, S., and Kaleem, M. (2021). Advances in image acquisition and processing technologies transforming animal ecological studies. *Ecol. Inf.* 61, 101212. doi: 10.1016/J.ECOINF.2021.101212
- Peng, L., Zhu, C., and Bian, L. (2023). U-shape transformer for underwater image enhancement. *IEEE Trans. Image Process.* 32, 3066–3079. doi: 10.1109/TIP.2023.3276332
- Ren, T., Xu, H., Jiang, G., Yu, M., and Luo, T. (2022). Reinforced swin-convts transformer for underwater image enhancement. *IEEE Trans. Geosci. Remote. Sens.* 60, 1–16. doi: 10.1109/TGRS.2022.3205061
- Sun, J., Dong, J., and Lv, Q. (2022). "2022. Swin transformer and fusion for underwater image enhancement," in *International Workshop on Advanced Imaging Technology (IWAIT)*. (Hong Kong, China: SPIE) Vol. 12177, pp. 627–631. doi: 10.1117/12.2626075
- Swinehart, D. F. (1962). The beer-lambert law. *J. Chem. Educ.* 39, 333. doi: 10.1021/ed039p333
- Wang, Y. D., Guo, J., Gao, H., and Yue, H. (2021). UIEC²-Net: CNN-based underwater image enhancement using two color space. *Signal Process. Image Commun.* 96. doi: 10.1016/j.image.2021.116250
- Wang, C. B., Ji, K., Huang, Q., and Xia, Y. (2007). "Generation of multi-spectral scene images for ocean environment," in *MIPPR 2007: Multispectral Image Processing*, Vol. 6787. (Wuhan, China: SPIE), p. 596–603. doi: 10.1117/12.751757
- Wang, H., Köser, K., and Ren, P. (2025). Large foundation model empowered discriminative underwater image enhancement. *IEEE Trans. Geosci. Remote. Sens.* 63, 1–17. doi: 10.1109/TGRS.2025.3525962
- Wang, Y., Zhang, J., Cao, Y., and Wang, Z. (2017). "A deep CNN method for underwater image enhancement," in *24th IEEE International Conference on Image Processing (ICIP)*. (IEEE), p. 1382–1386. doi: 10.1109/ICIP.2017.8296508
- Wang, S., Zheng, J., Hu, H. M., and Li, B. (2013). Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Process.* 22, 3538–3548. doi: 10.1109/TIP.2013.2261309
- Xiong, J., Zhuang, P., and Zhang, Y. (2020). "An efficient underwater image enhancement model with extensive Beer-Lambert law," in *IEEE International Conference on Image Processing (ICIP)*. (Abu Dhabi, United Arab Emirates: IEEE), p. 893–897. doi: 10.1109/ICIP40778.2020.9191131
- Yang, T., Yang, T., Gao, W., Wang, P., Li, X., Lv, Z., et al. (2023). Underwater target detection algorithm based on automatic color level and bidirectional feature fusion. *Laser Optoelectron. Prog.* 60, 11–18. doi: 10.3788/LOP213139
- Zhang, W., Li, X., Huang, Y., Xu, S., Tang, J., and Hu, H. (2025). Underwater image enhancement via frequency and spatial domains fusion. *Opt. Laser. Eng.* 186, 108826. doi: 10.1016/j.optlaseng.2025.108826
- Zhang, W., Zhou, L., Zhuang, P., Li, G., Pan, X., Zhao, W., et al. (2023). Underwater image enhancement via weighted wavelet visual perception fusion. *IEEE Trans. Circuits Syst. Video Technol.* 34.4, 2469–2483. doi: 10.1109/TCSVT.2023.3299314
- Zhu, D. (2023). Underwater image enhancement based on the improved algorithm of dark channel. *Mathematics*. 11, 1382. doi: 10.3390/math11061382
- Zhu, Z., Li, X., Ma, Q., Zhai, J., and Hu, H. (2025). FDNet: Fourier transform guided dual-channel underwater image enhancement diffusion network. *Sci. China. Technol. Sc.* 68.1, 1100403. doi: 10.1007/s11431-024-2824-x
- Zhu, Z., Li, X., Zhai, J., and Hu, H. (2024). POBB: A learning-based polarimetric object detection benchmark for road scenes in adverse weather conditions. *Inform. Fusion*. 108, 102385. doi: 10.1016/j.inffus.2024.102385