Dual Color Space Underwater Image Enhancement Network

Jiaying He JIAYING.HE@MPU.EDU.MO
K. L. Eddie Law* EDDIELAW@MPU.EDU.MO
Faculty of Applied Sciences, Macao Polytechnic University, Macao, China

Editors: Hung-yi Lee and Tongliang Liu

Abstract

In the field of underwater image enhancement, existing methods generally rely heavily on the RGB color space and ignore the potential advantages of the perceptually uniform XYZ color space in color correction. Additionally, CNN-based methods are prone to losing long-distance dependency relationships during local feature extraction process, thus affecting the image restoration quality. To address the above issues, we propose DSCNet, an underwater image enhancement framework based on dual color spaces. The framework aims to break through the limitations of traditional methods, by innovatively introducing a parallel processing mechanism for both RGB and XYZ color spaces. Upon fully taking advantages of the XYZ space in terms of perceptual linearity, the model can improve the color correction and brightness enhancement processes. Furthermore, we design a hybrid computing architecture which combines convolutional operations with a novel lightweight Transformer module. Through channel splitting and dimensionality reduction strategies, the computational complexity is reduced significantly while maintaining the ability to effectively model global contextual information. Experimental results show that DCSNet exhibits excellent enhancement performance in various underwater scenarios and delivers superior visual effects. Moreover, with its small number of model parameters, DCSNet can be deployed on embedded or edge devices for practical underwater visualization applications.

Keywords: Underwater image enhancement, dual color space processing, lightweight transformer, computational efficiency, color correction, global context modeling.

1. Introduction

Underwater imaging is crucial for marine exploration, ecological monitoring, and infrastructure inspection. However, images captured underwater suffer from severe quality degradation due to the absorption and attenuation characteristics of light when it propagates in water. This results in imagery color cast, presenting typically a blue-green tint, and a significant decrease in contrast (Yu et al., 2024). This phenomenon affects human eye observation and the applications of underwater robots, such as navigation, object detection, and recognition (Wang et al., 2024; Bogue, 2015). Hence, improving underwater image quality is of great significance in improving the distinguishability of target features. With the development of fields such as ocean engineering and underwater rescue, underwater image enhancement has become one of the key technologies (Anwar and Li, 2019; Wu et al., 2024).

Traditional underwater image enhancement methods are mainly divided into physical model-based and non-physical enhancement (Raveendran et al., 2021). Physical model

* Corresponding author. The MPU internal document identity number is fca.9915.fbc1.c

methods estimate optical parameters to reverse the degradation process, but they rely on accurate depth information and are limited in practical applications. Non-physical methods, including histogram techniques and the Retinex model, directly operate on pixel values without the need for physical modeling (Yuan et al., 2022). Fusion-based methods combine multiple enhancement results to improve the effect. However, these methods face problems such as color cast and noise amplification under complex underwater conditions, and their adaptability is limited (Cheng et al., 2025).

Deep learning, especially convolutional neural networks (CNNs), has made remarkable progress in underwater image enhancement (Yuan et al., 2024; Zhou et al., 2025; Zhang et al., 2024b). However, limited by the local receptive field, it is difficult to handle the global scattering effect. The emerging Transformer architecture effectively models global relationships through the self-attention mechanism, overcoming this limitation (Li et al., 2025). Some of the latest methods combine Transformer and CNN to take into account both local features and global context. However, pure Transformer has a relatively high computational complexity in high-resolution image processing, which affects its real-time application potential (Zhang et al., 2025c).

Current underwater image enhancement methods suffer from three major limitations. First, although a limited number of studies have attempted to utilize alternative color spaces, the vast majority of approaches still operate predominantly in the RGB color space, which often overlooks the benefits of perceptually uniform color spaces that better align with human vision and can facilitate more effective color correction. Second, existing architectures struggle to balance local feature extraction and global context modeling—CNNs excel at local patterns but ignore long-range dependencies, while transformers capture global context but are inefficient and may overlook fine details. Third, computational inefficiency remains a significant barrier, especially for transformer-based approaches that require quadratic computation relative to image size. Too complex network are difficult to deploy on some lightweight underwater equipment.

To address these limitations, we propose a novel Dual Color Space Network (DCSNet). Our approach introduces parallel processing in both RGB and XYZ color spaces, leveraging the complementary advantages of each representation. The XYZ color space, being device-independent and perceptually linear, provides a more effective basis for color correction and luminance enhancement. To overcome the local-global modeling trade-off, we design hybrid blocks that integrate convolutional operations for local feature extraction with efficient transformer components for global context modeling. Crucially, we introduce a lightweight attention mechanism that maintains the benefits of global modeling while dramatically reducing computational overhead through channel splitting and dimensionality reduction.

Our main contributions are summarized as follow:

1. **A Dual Color Space Network (DCSNet) is proposed for underwater image enhancement.** By leveraging the complementarity between color spaces, performing feature extraction and processing synchronously in RGB and XYZ color spaces. The network has relatively low requirements for computing resources and can be effectively deployed on low-computing-power devices.

2. **A Generalized Enhancement block (GenE Block) design is proposed optimized for underwater optical properties, balancing performance and effi-**

ciency: In this branch, local and global features of XYZ and RGB color spaces can be extracted through CNN and lightweight Transformer. This method can effectively remove image degradation while retaining the overall structural information of the image and maintaining well consistency.

3. **Comprehensive experimental verification demonstrates that our method can achieve good results in various underwater scenarios.** Moreover, the model parameters and size are much lower than those of current mainstream deep learning algorithms.

2. Related Works

Early methods for underwater image enhancement were mostly based on physical models and empirical assumptions. For example, the Underwater Dark Channel Prior (UDCP) method (Drews Jr et al., 2013) and histogram equalization. The method based on red channel correction and multi-scale fusion (Zhang et al., 2024a) enhances the image by physically correcting the red light attenuation and fusing multi-scale information, and can achieve a more natural visual effect in some scenarios. Other examples are the Bayesian Retinex method (Zhuang et al., 2021), and the multi-branch aggregation network MBANet (Yang et al., 2023). These methods focus on the separation and compensation of inherent degradation factors, and achieve good quality improvement on real underwater data.

In recent years, the advancement of deep learning has led to the widespread application of numerous data-driven algorithms in underwater image enhancement. For example, Li et al. constructed a large-scale real underwater image dataset UIEB and proposed a baseline convolutional neural network model named Water-Net (Li et al., 2020). Subsequently, various network architectures and learning strategies have been introduced to further improve enhancement performance. Among these developments, dual-branch networks have emerged as effective solutions for underwater image enhancement. By processing features in parallel – such as color and structure or global and local information – these models enhance clarity and detail while mitigating typical underwater distortions. For instance, UIEC^2-Net (Wang et al., 2021) and DBFNet (Sun and Tian, 2023) employ dual pathways to separately model complementary types of features, which are then fused to significantly improve visual quality. Such designs not only sharpen fine details but also address common issues including color casts, low contrast, and blur. Furthermore, recent approaches like UWMambaNet (Zhang et al., 2025a) and USLN (Xiao et al., 2022) incorporate lightweight architectures to enhance computational efficiency, facilitating practical deployment on resource-constrained devices.

In addition, several other directions have also been explored. Some studies integrate underwater imaging physical models and incorporate physical priors into network training (Li et al., 2024), while others employ Generative Adversarial Networks (GANs) (Liang et al., 2024) or diffusion models to produce more realistic enhancement results. For example, DCGF utilizes a diffusion probabilistic model to progressively refine image quality (Zhang et al., 2025b). There are also task-oriented enhancement methods, such as the approach proposed by Yu et al. (2023), which incorporates object detection branches or losses into the enhancement network to guide the generation of images that improve machine recognition performance. Experiments demonstrate that such task-friendly methods can enhance object

detection accuracy compared to enhancements optimized solely for human visual perception. Meanwhile, Transformer-based architectures like U-Transformer (Peng et al., 2023) and WaterFormer (Kang et al., 2023) have been introduced to underwater image enhancement, leveraging self-attention mechanisms for global information modeling and further improving color restoration and clarity.

3. Method

3.1. Motivation

Traditional image enhancement methods operate predominantly in the RGB color space. However, it does not always align with human visual perception or the physical properties of light. The XYZ color space offers a perceptually uniform representation and decouples luminance from chromaticity, making it more robust for color-related tasks. Motivated by this, we propose a dual-branch network, DCSNet, that processes both RGB and XYZ representations in parallel, allowing the model to exploit complementary information from both color spaces.

Furthermore, while convolutional neural networks (CNNs) excel at capturing local spatial features, they are inherently limited in modeling long-range dependencies. Recent advances have demonstrated the effectiveness of transformer architectures in capturing global context via self-attention mechanisms. However, standard transformers are computationally heavy for high-resolution images. To address this problem, we employ a lightweight transformer within each GenE block, enabling efficient global feature modeling without significant computational overhead.

3.2. Overview

DCSNet is a dual-branch structured network. The core idea is to break through the limitations of traditional single-color space processing by processing the complementary representations of the input image in different color spaces in parallel. The overall framework adopts a three-stage process of "color space transform - parallel processing - adaptive fusion," as shown in Figure 1. Firstly, the input RGB image is converted into the CIE-XYZ color space, then two identical but independent branches are used to process the RGB and XYZ representations, respectively. Finally, a learnable fusion convolution is used to integrate the dual-path features to generate the final output. This design enables the network to simultaneously utilize the original information in the RGB and the XYZ color space and to align. This framework maintains end-to-end differentiability and achieves collaborative optimization of multi-space features.

3.3. Dual-Color-Space Architecture

The dual-branch design is motivated by the fact that different color spaces emphasize distinct image features and degradation patterns. For instance, RGB captures rich color information, while XYZ space offers a perceptually uniform luminance and color representation. By processing both simultaneously, the network can extract and integrate complementary features, thereby improving restoration accuracy and visual quality. This complementary learning strategy enhances the model's ability to handle tasks such as denoising, deblurring,

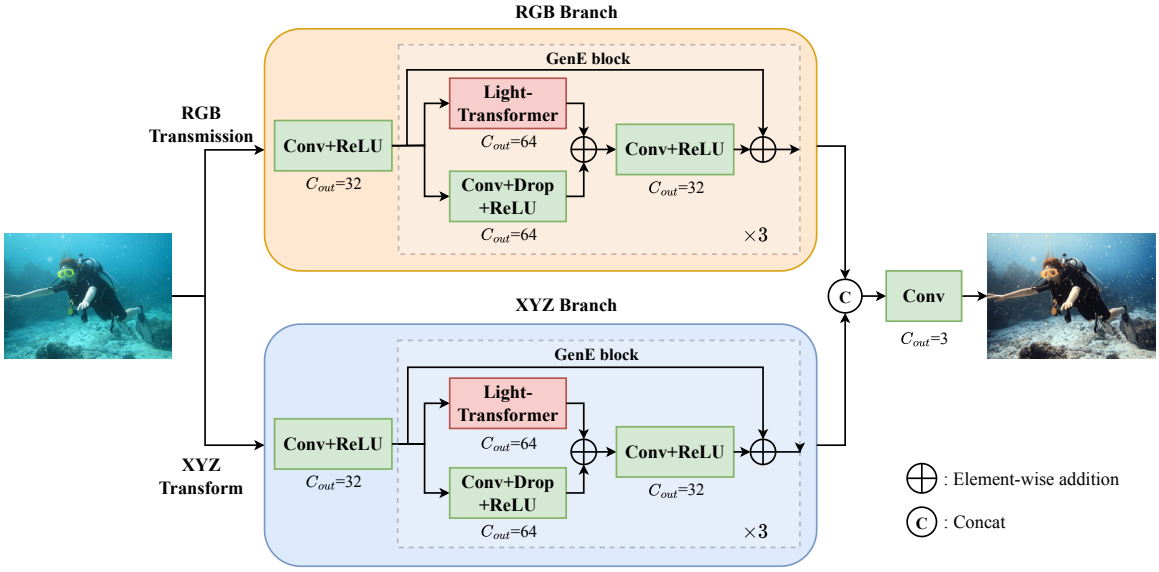

Figure 1: The overall structure of DCSNet.

and contrast adjustment. Given an input image I_{RGB}, we construct a parallel representation in the XYZ color space via:

$$I_{XYZ}(h, w) = \mathbf{M} \cdot I_{RGB}(h, w), \tag{1}$$

where h and w denote the height and width of the image, and $\mathbf{M}$ is the standard RGB-to-XYZ conversion matrix.

$$\mathbf{M} = \begin{bmatrix} 0.4124564 & 0.3575761 & 0.1804375 \\ 0.2126729 & 0.7151522 & 0.0721750 \\ 0.0193339 & 0.1191920 & 0.9503041 \end{bmatrix}. \tag{2}$$

This dual-path processing leverages the distinctive properties of each color space. Certain degradations or visual features are more pronounced and easier to process in one space than the other. RGB space, based on red, green, and blue primaries, aligns with human visual perception. In contrast, XYZ space offers a linear, standardized color representation conforming to optical measurement principles. Processing both spaces in parallel allows the model to adaptively exploit these advantages, leading to more comprehensive feature learning and improved enhancement performance.

DCSNet consists of two parallel branches, each built with identical GenE blocks designed to capture both local details and global semantics. One branch processes the input in RGB, while the other operates on the XYZ-transformed version. Each branch independently extracts and enhances features, producing two intermediate outputs: O_{RGB} and O_{XYZ}. These outputs retain enhanced features specific to their respective color spaces.

To fully leverage this complementary information, the two feature sets are concatenated and fused. The fusion is performed via element-wise addition, combining the strengths of

both representations. The resulting features are then processed through a 3×3 convolution layer to generate the final output. This procedure can be expressed as:

$$O_c = \mathrm{C}(F_{RGB}, F_{XYZ}), \tag{3}$$

$$O_{final} = \mathrm{Conv2d}_{3\times3}(O_c), \tag{4}$$

where F_{RGB} and F_{XYZ} denote the feature maps from the RGB and XYZ branches, respectively, O_c is the concatenated feature tensor, and O_{final} is the enhanced output image. $\mathrm{C}(\cdot)$ is the concat function, $\mathrm{Conv2d}_{3\times3}(\cdot)$ is the convolution function. The result exhibits improved color fidelity, contrast, noise suppression, and overall visual naturalness—making the method particularly suitable for complex underwater imaging conditions.

3.4. Generalized Enhancement Block

The backbone of each branch comprises of an initial convolutional layer, a series of GenE blocks and a final convolutional layer. These GenE blocks employ a lightweight transformer module to improve the ability to capture long-range dependencies without significantly increasing computational complexity.

For an input feature map F_{in}, the GenE block performs a series of well-defined transformations. First, the input is processed by a standard convolutional operation, which helps to effectively extract local spatial features. Then, the feature map is fed into the lightweight Transformer unit, where global context information is captured by modeling the interactions between different regions of the feature map. After the Transformer block, additional convolutional layers are used to further refine the resulting feature representation. These layers help to integrate rich features while maintaining spatial coherence. The entire process ensures that both local and global features of the input are fully captured, thus producing a more powerful feature representation. This hybrid design enhances the network's representational power while maintaining memory and computational efficiency, making it ideal for resource-constrained environments.

For an input feature map F_{in}, the block operates as follows:

$$\begin{aligned} F_{trans} &= LT(F_{in}), \\ F_{conv} &= \mathrm{ReLU}(\mathrm{Dropout}(\mathrm{Conv2d}_{3\times3}(F_{in}))), \\ F_{fused} &= F_{conv} + F_{trans}, \\ F_{out} &= \mathrm{Conv2d}_{3\times3}(\mathrm{ReLU}(\mathrm{Dropout}(F_{fused}))), \end{aligned} \tag{5}$$

where F_{trans} is the output of the lightweight Transformer function $LT(\cdot)$, F_{conv} is the output of the convolutional branch, and $\mathrm{Dropout}(\cdot)$ is the random dropout function. F_{fused} and F_{out} are the fused features and the final output features, respectively. The output is then concatenated with the original input I_{in} to preserve low-level details:

$$F_{add} = \mathrm{Concat}(F_{out}, I_{in}). \tag{6}$$

The lightweight Transformer is adapted for 2D feature maps, making it particularly effective in modeling long-range dependencies within spatial data. This adaptation enables the model to better capture and restore details that may be obscured by light absorption

and scattering in aquatic environments. By establishing interactions between distant spatial locations, the lightweight Transformer can effectively recover structural information and color fidelity from degraded underwater images. This approach improves the model's ability to understand contextual relationships while maintaining high computational efficiency. For input x:

$$[\mathbf{Q}, \mathbf{K}, \mathbf{V}] = \text{Conv2d}_{1\times1}(x), \text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\ (\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}})\mathbf{V}. \tag{7}$$

Among them, $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ represent the query vector (Query), the key vector (Key) and the value vector (Value) in the attention mechanism, respectively. Attention($\mathbf{Q}, \mathbf{K}, \mathbf{V}$) represents the attention mechanism function. $\sqrt{d_k}$ is called the scaling factor.

GenE effectively captures fine-grained local patterns through convolutional operations and acquires global context information via a self-attention mechanism. In complex image enhancement tasks, this dual-branch structure enhances overall visual coherence while preserving detailed textures. The model's ability to process both local and global information simultaneously makes the enhanced results structurally coherent and visually natural. In addition, the lightweight Transformer in this module helps overcome the locality constraints of traditional convolutional neural networks (CNNs). The model can understand the global structure and relationships within the entire image. This global perception is crucial when information from distant regions significantly influences local enhancement decisions. For example, when restoring missing or degraded textures, global information helps improve overall consistency of the image.

3.5. Loss Function

In the training process, we employ a combined loss function designed to promote both pixel-level accuracy and perceptual realism. The overall loss is formulated as follows:

$$\mathcal{L}_{\text{total}} = \underbrace{\alpha L_{\text{mse}}}_{\text{pixel-level matching}} + \underbrace{\beta L_{\text{Semantics}}}_{\text{semantic features}} + \underbrace{\gamma L_{\text{pixel}}}_{\text{pixel features}}. \tag{8}$$

The design of such a loss function aims to optimize the restoration effect of underwater images. It consists of three parts, which enhance the image quality from different aspects.

The first part is the pixel-level Mean Squared Error (MSE), which is used to measure the difference in pixel values between the reconstructed image and the original image. This part helps to preserve the basic structure and detail clarity of the image, making the restored image visually closer to the real scene. The weight of this term is controlled by the hyperparameter α to balance the need for pixel accuracy in the overall loss.

The second part extracts high-level semantic features based on a pre-trained VGG network. By comparing the differences between deep features, it ensures the semantic consistency of the image. This method is particularly suitable for solving common problems in underwater images such as color distortion and contrast reduction, and can maintain the true appearance of objects under complex lighting conditions. The weight of this part is adjusted by the hyperparameter β, enabling the model to pay more attention to the accuracy of semantic information during the training process.

The third part is the pixel-level semantic loss. By incorporating additional prior constraints, such as color correction and illumination compensation, it further optimizes the specific attributes of underwater images. The design of this loss function takes into account the special characteristics of the underwater imaging environment, such as the absorption and scattering effects of light, thus more effectively restoring the color and details of the image. The contribution degree of this part is controlled by the hyperparameter γ.

The adjustment of loss function weights (α, β, γ) remains a direction for future work. While currently set as equally important, task-specific adjustments—such as emphasizing pixel-level features for detail enhancement or prioritizing semantic features for recognition tasks—could further improve performance. This flexibility would enhance the model's adaptability to diverse underwater vision scenarios.

4. Experimental analysis

4.1. Datasets, Evaluation Metrics and Experimental Setup

We evaluated the proposed method on the UIEB (Li et al., 2020) and EUVP (Islam et al., 2020) datasets. The UIEB dataset comprises 950 underwater images, of which 890 have corresponding reference images and 60 belong to the challenge set. For training, we employed 800 pairs of real-world images, and testing was conducted on the remaining 90 UIEB images. The EUVP dataset contains about 20K underwater images organized into paired and unpaired subsets. The paired data provide aligned low- and high-quality images for supervised training, while the unpaired data include independent collections of poor- and good-quality images for adversarial training. With diverse underwater scenes and varying conditions, EUVP is widely used for training and evaluating image enhancement models.

In the evaluation of underwater image quality, UCIQE (Yang and Sowmya, 2015), UIQM (Panetta et al., 2016), CCF (Wang et al., 2018), and FDUM (Yang et al., 2021) are commonly used no-reference evaluation metrics, which quantify image quality from multiple dimensions. UCIQE is calculated through the weighted sum of color, saturation, and contrast. A larger value indicates less color distortion, higher contrast, and better overall visual quality. UIQM comprehensively considers the authenticity of color, image clarity, and contrast. A higher value indicates more natural colors and clearer images. CCF is composed of a linear combination of color richness, contrast, and haze, and is used to measure the degradation of image quality caused by medium scattering. A higher value means richer colors, better contrast, and weaker scattering effects. FDUM reflects image quality by estimating the degree of medium scattering in the image. A larger value indicates higher image transparency.

The DCSNet is implemented in PyTorch 1.13 on a system equipped with a NVIDIA 4090 GPU boasting 24 GB of RAM and an Intel (R) Xeon (R) W-2255 CPU. We use Adam Optimizer for training. The initial learning rate is set to 0.01, and the learning rate decays by 30% every 30 epochs. The training uses a batch size of 4 and is carried out for a total of 100 epochs. The input images are uniformly scaled to a resolution of 512×512. During the training process, the model checkpoint is saved every 10 epochs.

Table 1: Quantitative comparison on the **UIEB** dataset, the **red** is the highest, the **blue** is the second highest, and the **black** font is the third highest.

Method	UCIQE(↑)	UIQM(↑)	CCF(↑)	FDUM(↑)
Raw	0.519	1.157	20.514	0.447
Fusion (Ancuti et al., 2018)	0.592	1.345	20.053	0.538
Shallow-UWnet (Naik et al., 2021)	0.508	1.084	18.955	0.382
U-Shape (Peng et al., 2023)	0.569	1.343	23.732	0.495
DMWater (Tang et al., 2023)	**0.603**	**1.361**	**27.062**	**0.553**
NU^{2}Net (Guo et al., 2023)	0.598	1.258	20.694	0.509
HCLR-Net (Zhou et al., 2024)	**0.613**	**1.350**	**26.319**	**0.610**
Ours	**0.602**	**1.391**	**31.738**	**0.658**

Table 2: Quantitative comparison on the **EUVP** dataset, the **red** is the highest, the **blue** is the second highest, and the **black** font is the third highest.

Method	UCIQE(↑)	UIQM(↑)	CCF(↑)	FDUM(↑)
Raw	0.549	1.302	**30.267**	0.445
Fusion	0.588	**1.404**	28.310	0.486
Shallow-UWnet	0.562	1.296	25.972	0.440
U-Shape	0.408	0.845	12.434	0.423
DMWater	0.510	1.129	19.359	0.340
NU^{2}Net	**0.606**	1.389	29.123	**0.523**
HCLR-Net	**0.618**	**1.411**	**36.929**	**0.555**
Ours	**0.607**	**1.467**	**46.857**	**0.629**

4.2. Experiment on Datasets

In the quantitative comparison, our method ranked first in the three key metrics - UIQM, CCF and FDUM - on both the UIEB and EUVP datasets, exceeding the second-place method by 0.03, 4.676, and 0.048 respectively on UIEB, and by 0.056, 9.928, and 0.074 on EUVP. This indicates that our method demonstrates strong performance in color restoration, detail preservation, and structural fidelity. For the UCIQE metric, although our method ranked second on the EUVP dataset and third on the UIEB dataset, it trailed the top-performing HCLR-Net by only 0.011 in both cases.

The qualitative visual results highlight the advantages of our method under the UIQM and FDUM metrics, which assess color naturalness, contrast, and fine detail preservation, directly reflecting its perceived superiority. Figure 2 and Figure 3 present visual comparisons between our method and various underwater image enhancement models across different scenes from the UIEB and EUVP datasets. As shown in Figure 2, Shallow-UWnet

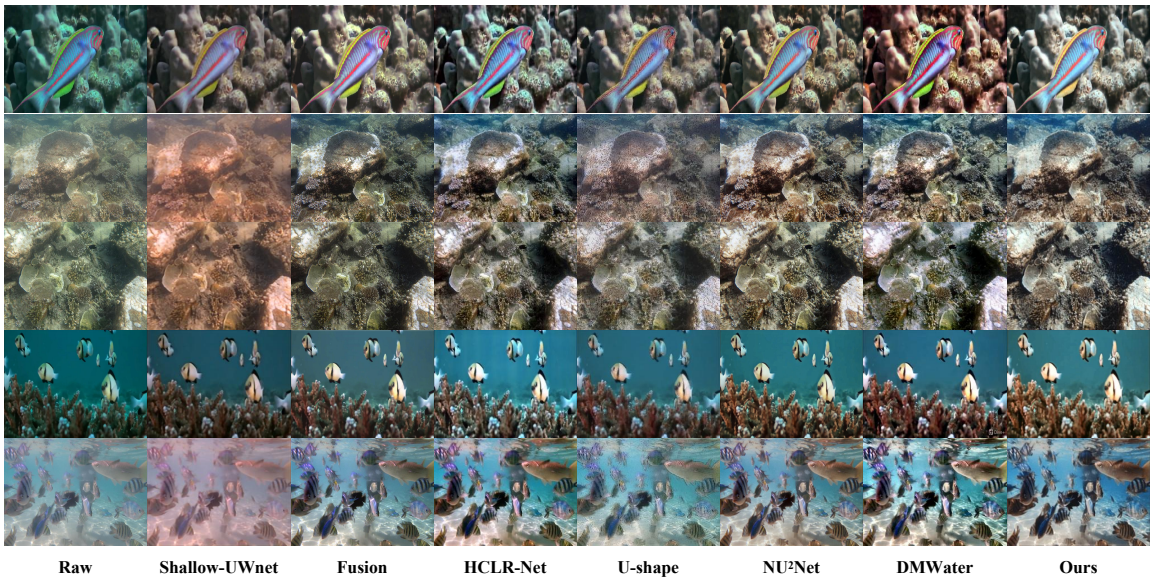

Figure 2: Qualitative evaluation of the proposed method compared to other techniques using the UIEB dataset.

Table 3: Comparison of Different UIE Models in Terms of GFLOPs (G) and Parameters (M), the **red** is the lowest, the **blue** is the second lowest, and the **black** font is the third lowest.

Method	Param(M)↓	GFLOPS(G)↓
NU^{2}Net	**3.15M**	**46.33G**
Shallow-UWnet	**0.22M**	**43.26G**
PUIE-Net (Fu et al., 2022)	1.40M	423.04G
Semi-UIR (Huang et al., 2023)	1.68M	550.76G
HCLR-Net (Zhou et al., 2024)	4.87M	5651.99G
Ours	**0.44M**	**86.54G**

demonstrates relatively good performance in restoring fish images but introduces additional red artifacts when processing turbid images. U-shape improves contrast, yet fails to recover fine structural details, resulting in blurred edges. Both NU^{2}Net and DMWater achieve a better balance between contrast and color fidelity, but still exhibit deficiencies such as loss of details in dark areas or red artifacts in certain regions. The visual results of HCLR-Net are comparable to those of ours, though some of its outputs suffer from incomplete color cast removal—for example, the fish or backgrounds in the first and fifth rows still retain a bluish tint.

Figure 3: Qualitative evaluation of the proposed method compared to other techniques using the EUVP dataset.

As observed in Figure 3, Shallow-UWNet tends to exhibit over-saturation or color distortion, especially in challenging regions like the yellow fish and sea turtle. Although U-Shape and DMWater improve color fidelity to some extent, they still display residual haze and reduced local contrast. Other methods generally perform well on the EUVP dataset, but often fail to fully eliminate the blue color cast in the background.

In contrast, our method restores natural colors and preserves textural details across all examples, while maintaining overall visual coherence without over-saturation. The enhanced results align more closely with actual scenes. These outcomes confirm that our method excels not only in quantitative metrics but also in delivering consistent and high-quality visual enhancements. The effectiveness of the dual-color space design is thereby underscored, demonstrating robust performance across complex underwater scenarios.

Furthermore, to comprehensively evaluate the real-time processing capability and deployment efficiency of underwater vision models, we systematically analyzed the model parameter counts and computational complexities of various architectures. The quantitative results (Table 3) indicate that the proposed method demonstrates excellent efficiency among deep learning-based solutions. Specifically, in terms of the key parameter scale (Param), our model remains highly compact. Although slightly higher than that of the lightest Shallow-UWNet, it achieves an 86% reduction compared to NU²Net. In terms of computational load (GFLOPS), our method reduces it by 79.5% compared to PUIE-Net. It is noteworthy that although Shallow-UWNet has the lowest computational cost, its restoration accuracy is significantly inferior to our method. Thus, the proposed method achieves

Table 4: Ablation study of various modules and loss functions on the UIEB dataset.

Method	Evaluation Metric			
	UCIQE	UIQM	CCF	FDUM
-w/o RGB branch	0.497	1.033	16.025	0.346
-w/o XYZ branch	0.517	1.080	16.338	0.411
-w/o GenE block	0.509	1.140	16.930	0.394
Full Model	**0.602**	**1.391**	**31.738**	**0.658**

an optimal balance between performance and efficiency, offering a practical solution for real-time deployment on resource-constrained platforms such as autonomous underwater vehicles and unmanned aerial vehicles.

4.3. Ablation Study

Ablation studies were conducted on the UIEB dataset to assess the contributions of individual components to model performance. Quantitative and visual comparisons under different architectural configurations are summarized in Table 4 and Figure 4.

The full model consistently outperformed all ablated variants: -w/o RGB branch, -w/o XYZ branch, and -w/o GenE block. Specifically, the UCIQE score increased from 0.497, 0.517, and 0.509 to 0.602; UIQM rose from 1.033, 1.080, and 1.140 to 1.391; CCF improved markedly from approximately 16 to 32; and FDUM increased from 0.346, 0.411, and 0.394 to 0.658. These results confirm that both color-space branches and the GenE block are critical to model performance, and their joint operation yields the best outcomes. The varying degree of performance degradation upon the removal of each component suggests their distinct contributions across different metrics.

Visual results further demonstrate the component-specific contributions: the -w/o XYZ output shows a bluish-white tint with detail loss, highlighting its role in color correction and detail retention. The -w/o RGB variant exhibits reduced color richness and clarity, confirming its importance for chromatic accuracy. Removing the GenE module introduces a yellowish cast and impairs color consistency, underscoring its global consistency effect. In contrast, the complete model produces the most natural colors and sharpest details, validating the complementary advantages of the dual-branch design and the GenE module.

These ablation studies confirm that the RGB and XYZ branches offer complementary feature representations, while the GenE module ensures global color consistency. Their integration enables optimal performance across metrics, providing an effective solution for underwater image enhancement.

5. Conclusion

In underwater imaging tasks, due to the absorption and scattering effects of light, the image quality deteriorates significantly, which affects the effective execution of subsequent

Figure 4: The visualization results of the ablation experiments conducted on the introduced framework.

tasks. To address this issue, this paper proposed an underwater image enhancement network based on a dual color space. The network consists of an RGB branch and an XYZ branch. In each branch, through the GenE block that integrates the CNN and Transformer structures, the network realizes the mixed extraction of global and local information of the image, and further fuses multi-scale features for image restoration. In this process, the network can maintain excellent color restoration ability in complex underwater environments. Experimental results show that this method outperforms existing algorithms on multiple underwater image enhancement datasets, demonstrating better visual effects and performance.

References

Codruta O. Ancuti, Cosmin Ancuti, Christophe De Vleeschouwer, and Philippe Bekaert. Color balance and fusion for underwater image enhancement. *IEEE Trans. Image Process.*, 27(1):379–393, 2018. doi: 10.1109/TIP.2017.2759252.

Saeed Anwar and Chongyi Li. Diving deeper into underwater image enhancement: A survey. *arXiv preprint arXiv:1907.07863*, 2019.

Robert Bogue. Underwater robots: a review of technologies and applications. *Industrial Robot: An International Journal*, 42(3):186–191, 2015.

Zheng Cheng, Guodong Fan, Jingchun Zhou, Min Gan, and C. L. Philip Chen. Fdce-net: Underwater image enhancement with embedding frequency and dual color encoder. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(2):1728–1744, 2025. doi: 10.1109/TCSVT.2024.3482548.

P. Drews Jr, E. do Nascimento, F. Moraes, S. Botelho, and M. Campos. Transmission estimation in underwater single images. In *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, pages 825–830, 2013. doi: 10.1109/ICCVW.2013.113.

Zhenqi Fu, Wu Wang, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Uncertainty inspired underwater image enhancement. In *European conference on computer vision*, pages 465–482. Springer, 2022.

Chunle Guo, Ruiqi Wu, Xin Jin, Linghao Han, Weidong Zhang, Zhi Chai, and Chongyi Li. Underwater ranker: Learn which is better and how to be better. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(1):702–709, Jun. 2023. doi: 10.1609/aaai.v37i1.25147. URL https://ojs.aaai.org/index.php/AAAI/article/view/25147.

Shirui Huang, Keyan Wang, Huan Liu, Jun Chen, and Yunsong Li. Contrastive semi-supervised learning for underwater image restoration via reliable bank. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18145–18155, 2023.

Md Jahidul Islam, Youya Xia, and Junaed Sattar. Fast underwater image enhancement for improved visual perception. *IEEE Robot. Autom. Lett.*, 5(2):3227–3234, 2020. doi: 10.1109/LRA.2020.2974710.

Jian Kang, Haiyan Guan, Lingfei Ma, Lanying Wang, Zhengsen Xu, and Jonathan Li. Waterformer: A coupled transformer and cnn network for waterbody detection in optical remotely-sensed imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 206: 222–241, 2023.

Bin Li, Li Li, Zhenwei Zhang, and Yuping Duan. Luieo: A lightweight model for integrating underwater image enhancement and object detection. *arXiv preprint arXiv:2412.07009*, 2024.

Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.*, 29:4376–4389, 2020. doi: 10.1109/TIP.2019.2955241.

Yuanyuan Li, Zetian Mi, Yulin Wang, Shuaiyong Jiang, and Xianping Fu. Taformer: A transmission-aware transformer for underwater image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(1):601–616, 2025. doi: 10.1109/TCSVT.2024.3455353.

Dan Liang, Jiale Chu, Yuguo Cui, Zhanhu Zhai, and Dingcai Wang. Npt-ul: An underwater image enhancement framework based on nonphysical transformation and unsupervised learning. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–19, 2024. doi: 10.1109/TGRS.2024.3363037.

Ankita Naik, Apurva Swarnakar, and Kartik Mittal. Shallow-uwnet: Compressed model for underwater image enhancement (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 15853–15854, 2021.

Karen Panetta, Chen Gao, and Sos Agaian. Human-visual-system-inspired underwater image quality measures. *IEEE J. Oceanic Eng.*, 41(3):541–551, 2016. doi: 10.1109/JOE.2015.2469915.

Lintao Peng, Chunli Zhu, and Liheng Bian. U-shape transformer for underwater image enhancement. *IEEE Transactions on Image Processing*, 32:3066–3079, 2023. doi: 10.1109/TIP.2023.3276332.

Smitha Raveendran, Mukesh D Patil, and Gajanan K Birajdar. Underwater image enhancement: a comprehensive review, recent trends, challenges and applications. *Artificial Intelligence Review*, 54:5413–5467, 2021.

Kaichuan Sun and Yubo Tian. Dbfnet: A dual-branch fusion network for underwater image enhancement. *Remote Sensing*, 15(5), 2023. ISSN 2072-4292. doi: 10.3390/rs15051195. URL https://www.mdpi.com/2072-4292/15/5/1195.

Yi Tang, Hiroshi Kawasaki, and Takafumi Iwaguchi. Underwater image enhancement by transformer-based diffusion model with non-uniform sampling for skip strategy. In *Proceedings of the 31st ACM International Conference on Multimedia*, MM '23, page 5419–5427, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701085. doi: 10.1145/3581783.3612378. URL https://doi.org/10.1145/3581783.3612378.

Hao Wang, Weibo Zhang, Lu Bai, and Peng Ren. Metalantis: A comprehensive underwater image enhancement framework. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–19, 2024. doi: 10.1109/TGRS.2024.3387722.

Yan Wang, Na Li, Zongying Li, Zhaorui Gu, Haiyong Zheng, Bing Zheng, and Mengnan Sun. An imaging-inspired no-reference underwater color image quality assessment metric. *Comput. Electr. Eng.*, 70:904–913, 2018. ISSN 0045-7906. doi: https://doi.org/10.1016/j.compeleceng.2017.12.006. URL https://www.sciencedirect.com/science/article/pii/S0045790617324953.

Yudong Wang, Jichang Guo, Huan Gao, and Huihui Yue. UIEC2-Net: Cnn-based underwater image enhancement using two color space. *Signal Process Image Commun*, 96:116250, 2021. ISSN 0923-5965. doi: https://doi.org/10.1016/j.image.2021.116250.

Siyuan Wu, Bangyong Sun, Xiao Yang, Wenjia Han, Jiahai Tan, and Xiaomei Gao. Reconstructing the colors of underwater images based on the color mapping strategy. *Mathematics*, 12(13):1933, 2024.

Ziyuan Xiao, Yina Han, Susanto Rahardja, and Yuanliang Ma. Usln: A statistically guided lightweight network for underwater image enhancement via dual-statistic white balance and multi-color space stretch. *ArXiv*, abs/2209.02221, 2022. URL https://api.semanticscholar.org/CorpusID:252090049.

Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. *IEEE Trans. Image Process.*, 24(12):6062–6071, 2015. doi: 10.1109/TIP.2015.2491020.

Ning Yang, Qihang Zhong, Kun Li, Runmin Cong, Yao Zhao, and Sam Kwong. A reference-free underwater image quality assessment metric in frequency domain. *Signal Processing: Image Communication*, 94:116218, 2021. ISSN 0923-5965. doi: https://doi.org/10.1016/j.image.2021.116218. URL https://www.sciencedirect.com/science/article/pii/S0923596521000503.

Peng Yang, Chunhua He, Shaojuan Luo, Tao Wang, and Heng Wu. Underwater image enhancement via triple-branch dense block and generative adversarial network. *Journal of Marine Science and Engineering*, 11(6):1124, 2023.

Meng Yu, Liquan Shen, Zhengyong Wang, and Xia Hua. Task-friendly underwater image enhancement for machine vision applications. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2023.

Meng Yu, Liquan Shen, Zhengyong Wang, and Xia Hua. Task-friendly underwater image enhancement for machine vision applications. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024. doi: 10.1109/TGRS.2023.3340244.

Jieyu Yuan, Zhanchuan Cai, and Wei Cao. Tebcf: Real-world underwater image texture enhancement model based on blurriness and color fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022. doi: 10.1109/TGRS.2021.3110575.

Jieyu Yuan, Zhanchuan Cai, and Wei Cao. A novel underwater detection method for ambiguous object finding via distraction mining. *IEEE Transactions on Industrial Informatics*, 20(7):9215–9224, 2024. doi: 10.1109/TII.2024.3383537.

Tao Zhang, Haibing Su, Bin Fan, Ning Yang, Shuo Zhong, and Jiajia Yin. Underwater image enhancement based on red channel correction and improved multiscale fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 2024a.

Weidong Zhang, Gongchao Chen, Peixian Zhuang, Wenyi Zhao, and Ling Zhou. Catnet: Cascaded attention transformer network for marine species image classification. *Expert Systems with Applications*, 256:124932, 2024b.

Yuhan Zhang, Xinyang Yu, and Zhanchuan Cai. Uwmambanet: Dual-branch underwater image reconstruction based on w-shaped mamba. *Mathematics*, 13(13), 2025a. ISSN 2227-7390. doi: 10.3390/math13132153. URL https://www.mdpi.com/2227-7390/13/13/2153.

Yuhan Zhang, Jieyu Yuan, and Zhanchuan Cai. DCGF: Diffusion-color-guided framework for underwater image enhancement. *IEEE Trans. Geosci. Remote Sens.*, 63:1–12, 2025b. doi: 10.1109/TGRS.2024.3522685.

Zengxi Zhang, Zhiying Jiang, Long Ma, Jinyuan Liu, Xin Fan, and Risheng Liu. Hupe: Heuristic underwater perceptual enhancement with semantic collaborative learning. *International Journal of Computer Vision*, pages 1–19, 2025c.

Jingchun Zhou, Jiaming Sun, Chongyi Li, Qiuping Jiang, Man Zhou, Kin-Man Lam, Weishi Zhang, and Xianping Fu. Hclr-net: hybrid contrastive learning regularization with locally randomized perturbation for underwater image enhancement. *International Journal of Computer Vision*, 132(10):4132–4156, 2024.

Jingchun Zhou, Zongxin He, Dehuan Zhang, Siyuan Liu, Xianping Fu, and Xuelong Li. Spatial residual for underwater object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):4996–5013, 2025. doi: 10.1109/TPAMI.2025.3548652.

Peixian Zhuang, Chongyi Li, and Jiamin Wu. Bayesian retinex underwater image enhancement. *Engineering Applications of Artificial Intelligence*, 101:104171, 2021.