



Article

Underwater Image Enhancement with a Hybrid U-Net-Transformer and Recurrent Multi-Scale Modulation

Zaiming Geng¹, Jiabin Huang², Xiaotian Wang^{3,*} , Yu Zhang⁴, Xinnan Fan³ and Pengfei Shi⁴ ¹ China Yangtze Power Co., Ltd., Yichang 443002, China; geng_zaiming@cypc.com.cn² State Grid Hangzhou Power Supply Company, Hangzhou 310016, China; xyz814920@163.com³ College of Information Science and Engineering, Hohai University, Changzhou 213200, China; fanxn@hhu.edu.cn⁴ College of Artificial Intelligence and Automation, Hohai University, Changzhou 213200, China; 241314010011@hhu.edu.cn (Y.Z.); shipf@hhu.edu.cn (P.S.)

* Correspondence: wxthhu@hhu.edu.cn

Abstract

The quality of underwater imagery is inherently degraded by light absorption and scattering, a challenge that severely limits its application in critical domains such as marine robotics and archeology. While existing enhancement methods, including recent hybrid models, attempt to address this, they often struggle to restore fine-grained details without introducing visual artifacts. To overcome this limitation, this work introduces a novel hybrid U-Net-Transformer (UTR) architecture that synergizes local feature extraction with global context modeling. The core innovation is a Recurrent Multi-Scale Feature Modulation (R-MSFM) mechanism, which, unlike prior recurrent refinement techniques, employs a gated modulation strategy across multiple feature scales within the decoder to iteratively refine textural and structural details with high fidelity. This approach effectively preserves spatial information during upsampling. Extensive experiments demonstrate the superiority of the proposed method. On the EUVP dataset, UTR achieves a PSNR of 28.347 dB, a significant gain of +3.947 dB over the state-of-the-art UWFormer. Moreover, it attains a top-ranking UIQM score of 3.059 on the UIEB dataset, underscoring its robustness. The results confirm that UTR provides a computationally efficient and highly effective solution for underwater image enhancement.



Academic Editor: Jüri Majak

Received: 18 September 2025

Revised: 17 October 2025

Accepted: 21 October 2025

Published: 25 October 2025

Citation: Geng, Z.; Huang, J.; Wang, X.; Zhang, Y.; Fan, X.; Shi, P.

Underwater Image Enhancement with a Hybrid U-Net-Transformer and Recurrent Multi-Scale Modulation.

Mathematics **2025**, *13*, 3398. <https://doi.org/10.3390/math13213398>**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).**Keywords:** underwater image enhancement; hybrid U-Net-Transformer; recurrent multi-scale feature modulation**MSC:** 68T45

1. Introduction

Underwater image enhancement (UIE) is a critical research area in computer vision dedicated to mitigating the significant image quality degradation inherent in aquatic environments [1–3]. This degradation primarily stems from the selective absorption and scattering of light by water molecules and suspended particles. These physical phenomena cause the intensity of incident light to decay exponentially with distance while also introducing severe backscattering, which collectively reduce image clarity and visibility. The process is often described by the Beer-Lambert law, wherein wavelength-dependent attenuation coefficients lead to an exponential decay of light, resulting in color casts—typically blue or green—and a loss of contrast. Since these attenuation coefficients vary considerably

across the visible spectrum, the resulting distortions are wavelength-dependent. This phenomenon manifests as common artifacts in underwater images, such as color distortion, reduced contrast, and blurred details. The complex interplay of these physical mechanisms presents substantial challenges for computer vision systems, which are magnified in practical applications. For instance, underwater archeology requires millimeter-level clarity to identify artifacts, marine biology depends on accurate color rendition for species identification, and autonomous underwater vehicle (AUV) navigation demands high-contrast imagery for real-time object detection and safe maneuvering. Such domain-specific requirements underscore the limitations of conventional imaging and highlight the urgent need for effective and robust UIE techniques [4,5].

Recent advances in deep learning and generative modeling have spurred several new trends in underwater image enhancement. Diffusion models, for instance, are now employed to achieve more stable color correction and progressively restore fine details through iterative denoising [6]. Concurrently, prompt-guided and cross-modal strategies have emerged, which leverage textual or semantic priors to inform and constrain the enhancement process, thereby improving adaptability to specific scenes [7]. Despite these innovations, such methods often encounter practical limitations, including high computational costs, extensive training data requirements, and unstable performance in turbid or low-light conditions. Furthermore, the definition of an “enhanced” image varies significantly by application; underwater archeology prioritizes structural fidelity, marine biology requires accurate color recovery, and AUV navigation demands rapid, reliable processing for safe operation. These diverse, application-driven needs underscore the critical demand for UIE techniques that are both computationally efficient and robustly adaptable to a wide range of underwater environments.

Current UIE approaches are broadly categorized as physics-based, non-physics-based, and deep learning-based. Physics-based methods [8,9] model the light propagation process; however, their effectiveness is often hindered by the difficulty of acquiring accurate environmental parameters in real-world scenarios. Non-physics-based methods [10,11] employ direct pixel adjustments or heuristics, and their performance often degrades in complex underwater conditions. In contrast, deep learning methods [12–14], particularly architectures such as U-Net [15,16] and GANs [17], have demonstrated significant progress. While hybrid CNN-Transformer models like U-Net and others have shown promise, they often treat feature fusion as a single-step process, which can lead to suboptimal integration of local and global information. Furthermore, existing recurrent models may lack mechanisms to effectively handle temporal coherence across different spatial scales, sometimes resulting in over-enhancement or color shifts in highly turbid conditions. Traditional U-Net architectures also suffer from the loss of spatial details during upsampling [18], reducing image fidelity.

To address the aforementioned issues, this paper introduces an underwater image enhancement method based on a hybrid U-Net-Transformer architecture that incorporates a recursive refinement mechanism. Figure 1 provides a high-level conceptual overview of this process, while detailed network architecture is presented in Section 2.

(1) The proposed hybrid U-Net-Transformer architecture synergizes the local feature extraction proficiency of convolutional neural networks with the global context modeling capacity of Transformers, enabling a more effective method for processing complex underwater scenes.

(2) The architecture incorporates a Recurrent Multi-Scale Feature Modulation (R-MSFM) mechanism that iteratively refines features during the decoding process, thereby preserving critical spatial and textural details to significantly enhance the final image quality.

(3) On standard underwater benchmarks, the proposed method consistently outperforms state-of-the-art approaches in key metrics, including PSNR, SSIM, and UIQM. This superior performance is achieved while maintaining a high degree of computational efficiency, thereby demonstrating the method's suitability for real-time applications.

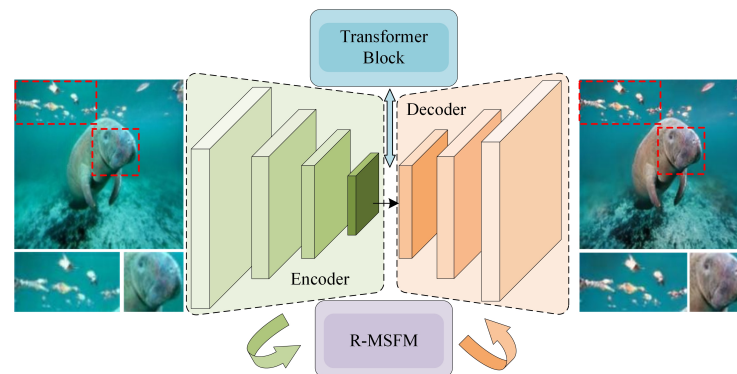


Figure 1. Conceptual overview of the proposed enhancement pipeline.

The remainder of this work is organized as follows: Section 2 details the proposed hybrid U-Net-Transformer architecture and the R-MSFM mechanism. Section 3 describes the experimental setup, including datasets, evaluation metrics, and comparative results. Section 4 provides a discussion of the results and limitations of the study. Finally, Section 5 concludes this work and suggests future research directions.

2. Methodology

To illustrate the challenges addressed, a typical degraded underwater image of a coral reef serves as a practical example. A standard CNN-based enhancer might successfully correct the blue-green color cast but at the cost of blurring the intricate textures of the coral. Conversely, a method focused solely on local details might fail to remove the global haze. The proposed hybrid U-Net-Transformer architecture is designed to overcome both limitations simultaneously: its U-Net backbone preserves local textures, while the integrated Transformer models global scene properties to ensure a uniform and natural color correction.

A hybrid U-Net-Transformer architecture for underwater image enhancement is presented that synergizes the multi-scale feature extraction of U-Net with the global information modeling of the Transformer. To better preserve spatial details during reconstruction, the integrated R-MSFM mechanism iteratively refines image quality within the decoder, leading to enhanced detail restoration.

2.1. Overall Framework

As shown in Figure 2, the overall architecture can be conceptually understood as three main components: a CNN-based encoder, a Transformer-based core network, and a decoder enhanced with the R-MSFM. Visually, the diagram illustrates this data flow: (1) The Encoder, shown on the left, serves as the downsampling path. It processes the input image through a series of convolutional blocks and max-pooling operations (represented by 'Maxpool' and the downsampling arrows in the legend) to extract multi-scale local features. (2) The Core Network, depicted as the 'Transformer Block' at the bottleneck, processes the deepest features from the encoder to model global contextual dependencies. (3) The Decoder, shown on the right, is the upsampling path responsible for reconstructing the enhanced image. A key innovation of the model is the integration of the Recurrent Multi-Scale Feature Modulation (R-MSFM) module within this decoder stage. The R-MSFM

iteratively refines the features from the encoder’s skip connections (e_1, e_2, e_3) before they are fused into the main decoder path, mitigating information loss during upsampling.

Let the input image be $I \in \mathbb{R}^{H \times W \times 3}$. The process can be described at a high level as follows:

- Encoder (E): The encoder, a U-Net-based CNN, extracts multi-scale local features. It progressively downsamples the input, producing a set of feature maps $\{e_1, e_2, e_3, e_4\}$ at resolutions of $1/2 \times, 1/4 \times, 1/8 \times,$ and $1/16 \times$ the original image size, respectively. Thus, $e_i = E_i(e_{i-1})$.
- Core Network (T): The deepest feature map, $e_4 \in \mathbb{R}^{\frac{H}{16} \times \frac{W}{16} \times 512}$, is flattened and processed by a Transformer network to model global dependencies, producing a contextually enriched feature map $t = T(e_4)$. This approach differs from standard Vision Transformers (ViT), as the proposed do not perform patch embedding on the input image but instead tokenize the feature map from a convolutional backbone.
- Decoder (D): The decoder generates the enhanced image I_{out} through progressive upsampling and refinement. At each upsampling stage, it fuses the feature map from the previous decoder stage with the corresponding skip-connection feature map from the encoder. This fusion is followed by an iterative refinement process governed by the R-MSFM module, which updates the state of the feature map to restore fine details. Schematically, $I_{out} = D(t, \{e_1, e_2, e_3\})$.

The core motivation for this architecture is to balance local feature extraction with global context modeling. The fusion of multi-scale convolutional features with global attention features occurs in the decoder via skip-connections. Refinement is performed as an iterative process within the R-MSFM module at each decoding stage, which mitigates information loss during upsampling. The design of the Recurrent Multi-Scale Feature Modulation (R-MSFM) module is based on a temporal iterative perspective, which treats each upsampling step in the decoder as a “state update.” By leveraging a gated recurrent mechanism to iteratively refine features and restore details, the R-MSFM effectively mitigates the information loss inherent in conventional single-pass decoding schemes.

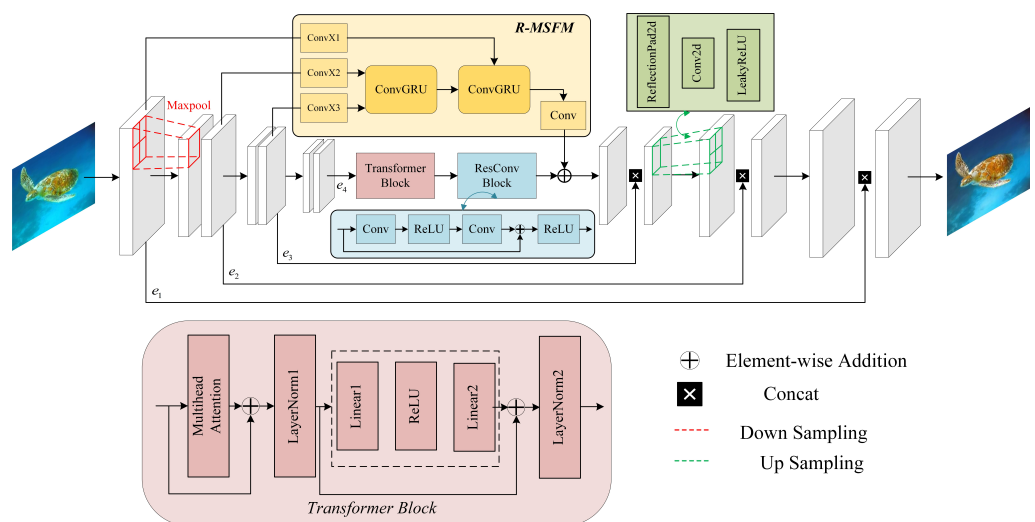


Figure 2. Architecture diagram of the proposed UTR methods.

2.2. U-Net-Transformer

The core network architecture incorporates a dual-stream feature interaction mechanism. The encoder path processes an input image of size $3 \times H \times W$ (e.g., $3 \times 256 \times 256$) through four main stages. Each stage consists of a convolutional block with 3×3 filters followed by a 2×2 max-pooling operation. The number of channels is doubled at each

successive stage, progressing from 64 in the first stage to 128, 256, and finally 512 in the fourth. This process yields a feature map of size $512 \times H/16 \times W/16$, which is then passed to the Transformer. The U-Net branch thus employs a four-stage convolutional downsampling process to extract multi-scale local features, with the output feature map sizes at each stage being $H \times W$, $H/2 \times W/2$, $H/4 \times W/4$, and $H/8 \times W/8$, respectively. Consequently, the embedding length for the Transformer is 512. Standard learnable one-dimensional positional encodings are added to the input sequence to retain positional information before it is fed into the Transformer layers. For feature fusion, an axial attention mechanism is adopted, which performs separable convolutional operations along the height and width dimensions. The resulting fused features are further enhanced via a residual convolutional block (Residual ConvBlock) to improve their representational capacity.

The multi-head attention mechanism operates as follows. First, the input sequence of feature vectors, each with dimension D_{model} , is independently projected through three separate linear layers to generate the Query (Q), Key (K), and Value (V) sequences. These components can be understood conceptually through an analogy to a library retrieval system. For each feature vector in the sequence, the Query represents its topic of interest, the Key acts as its label, and the Value corresponds to its actual content. The attention mechanism functions by comparing the query of one vector with the key of all other vectors to determine the weight assigned to their corresponding values.

Second, these Q, K, and V sequences are partitioned into N parallel “heads” by reshaping their dimensions, which allows the model to jointly attend to information from different representational subspaces. Third, a Scaled Dot-Product Attention operation is performed independently for each head, as described in Equation (1). Fourth, the attention outputs from all N heads are concatenated. Finally, this concatenated sequence is passed through a final linear layer to produce the output of the multi-head attention block.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

Here, Q, K, and V represent the query, key, and value matrices, respectively, while d_k denotes the dimension normalization factor [19]. For model simplicity, Q, K, and V are set to be identical. The resulting attention-weighted features are subsequently fused with the original convolutional features via a gated addition, a process that enables the network to selectively amplify informative regions while suppressing less relevant ones.

This multiscale feature extraction strategy allows the model to simultaneously capture both fine-grained local details and broader global structural patterns. These features are then channeled into a Recurrent Neural Network (RNN) for adaptive modulation, which facilitates a more effective integration and propagation of information across scales. This mechanism enhances the model’s capacity to preserve critical structural information, thereby improving the overall quality and discriminative power of the feature representations and ultimately boosting performance on the underwater image enhancement task.

In terms of computational efficiency, the proposed UTR architecture is designed to be relatively lightweight: Key design choices, such as the use of separable convolutions in the R-MSFM module and a moderate number of Transformer layers, help to limit the model’s parameter count and computational load. While a detailed complexity analysis is subject to specific hardware implementations, the architectural design prioritizes a strong performance-to-efficiency ratio, making it suitable for practical applications where computational resources may be constrained.

2.3. R-MSFM Mechanism

The R-MSFM module was specifically designed to address the detail loss that occurs during the upsampling phase of traditional U-Net architectures. This feature modulation module leverages a convolutional gated recurrent unit (ConvGRU), which is a convolutional adaptation of the Gated Recurrent Unit (GRU) architecture [20], to dynamically modulate features by integrating information from previous activations and current inputs. This iterative update process enriches both the semantic content and the spatial information of the feature maps. The module’s internal structure is detailed in Figure 3, where h_{t-1} represents the previous activation, x_t is the current input, and \tilde{h}_t denotes the current hidden activation.

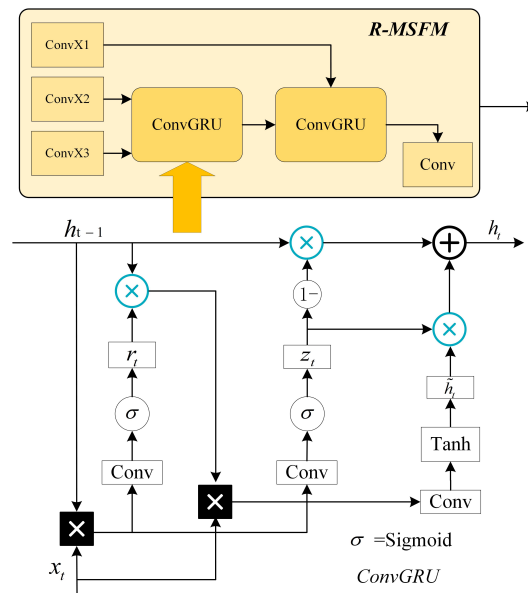


Figure 3. Schematic diagram of the Multi-Scale Cyclic Refinement Module.

The objective of this module is to find the most appropriate activation h_t during each iterative update, except for the initial update. This process can be formalized as

$$h_t = \begin{cases} X_3, & t = 1 \\ (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t, & t \neq 1 \end{cases} \quad (2)$$

$$z_t = \sigma(\text{Conv}_z([x_t, h_{t-1}])) \quad (3)$$

Here, σ represents the sigmoid activation function, and $[\cdot]$ denotes the concatenation operator. The module utilizes a single separable convolution unit, Conv_z , which consists of two sequential 1×3 and 3×1 convolutional layers. This efficient design reduces the model’s parameter count while preserving its representational capacity. The current hidden activation is subsequently determined by both the current input x_t and the previous activation h_{t-1} ,

$$\tilde{h}_t = \tanh(\text{Conv}_H([x_t, r_t \odot h_{t-1}])), \quad (4)$$

the reset gate r_t controls the extent to which the previous activation is forgotten, and this degree is computed by the following equation:

$$r_t = \sigma(\text{Conv}_R([x_t, h_{t-1}])), \quad (5)$$

the operators Conv_H and Conv_R represent separable convolution units with non-shared weights. The encoder used in this work provides image features at three different scales,

and accordingly, the R-MSFM model performs three rounds of iterative refinement on the prediction results.

Within the decoder, the R-MSFM module iteratively refines feature maps at each upsampling stage to progressively restore image details and textures. This process involves computing a residual between the current features and the initial input, which enables the module to apply targeted refinements for recovering high-frequency information. By incorporating features from multiple scales, the R-MSFM maintains global consistency while simultaneously preserving local details. The recurrent nature of the module facilitates the dynamic modulation of these multi-scale features, allowing for the correction of previous representations and the enhancement of feature fidelity. This dynamic approach is more flexible than static fusion methods, enabling it to capture the complex characteristics of underwater scenes more effectively and thereby improve final restoration accuracy.

To clarify the operational flow of the R-MSFM module, it is necessary to understand its three-step iterative process, which handles multi-scale features from the encoder's skip connections (e_3, e_2, e_1) . The procedure is as follows: (1) In the first iteration, the feature map from the third skip connection (e_3 , with 256 channels) is passed through a convolutional layer (convX31) to produce an initial correlation volume, designated as 'corr', with a standardized channel dimension of 128. (2) In the second iteration, the feature map from the second skip connection (e_2 , with 128 channels) is processed by a corresponding convolutional layer (convX21), also yielding a 128-channel feature map. This new map serves as the input 'x' to the ConvGRU block, while the 'corr' from the previous step functions as the hidden state 'h'. The ConvGRU then outputs an updated 'corr'. This design handles the mismatch in input dimensions by ensuring both 'h' and 'x' have 128 channels before being concatenated within the GRU. (3) The third iteration repeats this process, utilizing the updated 'corr' as the hidden state and the processed features from the first skip connection (e_1) as the input. The final output from the last iteration is then used to compute a displacement map, which is added to the main feature map from the Transformer before proceeding through the decoder's up-sampling path. This iterative refinement strategy allows the model to progressively integrate features from coarse to fine scales.

2.4. Loss Function

To optimize the model, a composite loss function is employed, which combines the Mean Square Error (MSE) with a VGG-based perceptual loss. The MSE loss operates at the pixel level, minimizing the squared error between the enhanced and reference images to enforce consistency in brightness and color. Complementing this, the VGG perceptual loss evaluates differences in the feature space by comparing intermediate feature maps from a pre-trained VGG19 network. This component encourages the model to preserve high-level semantic and structural information, thereby improving the perceptual similarity and textural detail of the enhanced results. The final objective is a weighted sum of the MSE and VGG perceptual losses:

$$L_{Total} = \alpha L_{MSE} + \beta L_{VGG}. \quad (6)$$

Through empirical evaluation, the study found that an equal weighting performed robustly across datasets. Therefore, the proposed set both weighting coefficients, α and β , to 1.0 in all experiments.

3. Experiment and Discussions

3.1. Experimental Setup

To provide context for the model's training duration and computational efficiency, all experiments were conducted on a workstation equipped with an NVIDIA RTX 4090 GPU, an Intel Core i5-13490K processor, and 64 GB of memory. The model was trained using the AdamW optimizer with standard parameters ($\beta_1 = 0.9, \beta_2 = 0.999$). The initial learning rate was set to 0.0032, and a batch size of 16 was used. These hyperparameters were selected based on common practices in the field and were refined through preliminary experiments to ensure stable and effective training convergence.

3.2. Comparative Methods and Datasets

The proposed hybrid U-Net-Transformer model was compared with several existing image enhancement methods, including SCI [21], SMDR-IS [22], UWCNN [23], Deep SESR [24], UWFormer [25], Water-Net [26], FUnIE-GAN [15], Shallow-UWNet [27], Ucolor [28], and MMLE [29].

The experiments were conducted on three widely recognized benchmark underwater image datasets. To evaluate the model's performance and generalization ability, a strict training and testing protocol was implemented. The model was trained exclusively on the official training split of the EUVP dataset and subsequently evaluated on the test sets of all three datasets: EUVP for in-domain validation, and UIEB and UFO-120 for cross-domain generalization assessment. No data augmentation techniques, such as random flipping or color jittering, were applied during the training process.

The datasets are characterized as follows: The EUVP dataset [29] is a large-scale collection of paired images designed to simulate various degradation effects, including color shifts, haze, and varying turbidity. The UIEB dataset [26] contains 950 real-world underwater images, featuring a wide variety of complex scenes with challenging degradation types such as low contrast, uneven illumination, and color casts. Finally, the UFO-120 dataset [24], with 1620 image pairs, was created to accurately evaluate model performance under diverse aquatic conditions, including scattering, noise, and motion blur.

3.3. Evaluation Metrics

The model's performance was comprehensively evaluated using multiple standardized metrics that assess three key aspects: pixel-wise accuracy, structural fidelity, and underwater scene-specific quality. For pixel-wise accuracy, Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) [30] were employed to quantify discrepancies between the enhanced and reference images. MSE provides a direct measure of the average squared pixel difference, consistent with the loss function. PSNR, a logarithmic transformation of MSE expressed in decibels (dB), reflects the ratio between the maximum signal power and the noise power. As a widely used metric in image reconstruction, higher PSNR values indicate lower reconstruction error and superior pixel-level fidelity,

$$PSNR = 10 \cdot \log_{10} \left(\frac{Max_I^2}{L_{MSE}} \right). \quad (7)$$

In this equation, Max_I is the maximum possible pixel value. A higher PSNR value generally indicates that the enhanced image is closer in quality to the reference image. To assess structural fidelity, the Structural Similarity Index (SSIM) [31] is also employed. SSIM evaluates the consistency between the enhanced and reference images by comparing their structural information, luminance, and contrast.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (8)$$

In the SSIM formula, μ_x and μ_y represent the local window means, σ_x^2 and σ_y^2 are the variances, and σ_{xy} is the covariance. The terms $C_1 = (0.01L)^2$ and $C_2 = (0.03L)^2$ function as constants to ensure stability, where L is the dynamic range of the pixel values. The SSIM ranges from -1 to 1 , with higher values signifying better structural consistency. To evaluate perceptual quality specific to the aquatic domain, the Underwater Image Quality Measure (UIQM) [32] is also employed. This comprehensive metric assesses an image based on its colorfulness, sharpness, and contrast, calculated as follows:

$$UIQM = 0.028 \cdot UICM + 0.295 \cdot UISM + 3.325 \cdot UIConM. \quad (9)$$

Conventionally, a higher Underwater Image Quality Measure (UIQM) value is considered indicative of better image quality. However, excessively high UIQM values may imply over-enhancement, which can lead to color distortion or a loss of detail. Therefore, to better diagnose the tendency for over-enhancement, a straightforward diagnostic metric termed Absolute UIQM (A-UIQM) is proposed. This metric quantifies the discrepancy between the UIQM values of the enhanced and reference images. By computing this difference, A-UIQM provides insight into the perceptual consistency of the enhancement. A value closer to zero signifies that the enhanced image's quality profile more accurately aligns with that of the reference. The formula is given as follows:

$$A\text{-UIQM} = \text{UIQM}(I_{\text{enh}}) - \text{UIQM}(I_{\text{ref}}) \quad (10)$$

3.4. Analysis of Experimental Results

This work conducts a quantitative comparative analysis with existing mainstream algorithms on three widely used benchmark datasets, namely EUVP, UIEB, and UFO-120. The main experimental results are summarized as follows:

As shown in Table 1, the UTR method achieves state-of-the-art performance on the EUVP dataset. It obtains the lowest MSE of 0.095×10^3 , representing a 25.2% reduction compared to the next-best method and demonstrating superior pixel-level accuracy. Furthermore, UTR attains a PSNR of 28.347 and an SSIM of 0.850, outperforming other leading models such as Shallow-UWnet and UWFormer. These results indicate that the method effectively reduces pixel-wise errors while simultaneously preserving the structural details crucial for visual fidelity in complex underwater scenes.

Table 1. Quantitative experimental results on the EUVP dataset.

	MSE (10^3) ↓	PSNR ↑	SSIM ↑	UIQM ↑	A-UIQM →0
Deep SESR	0.192	25.300	0.780	2.950	0.137
FUnIE-GAN	0.156	26.190	0.740	2.840	0.027
MMLE	2.037	15.040	0.623	2.737	−0.076
Shallow-UWnet	0.127	27.101	0.812	2.886	0.073
SMDR-IS	0.427	21.827	0.798	2.887	0.074
U-Net	0.393	22.190	0.802	2.485	−0.328
Ucolor	0.557	20.670	0.786	2.824	0.011
UWCNN	0.664	19.906	0.715	2.895	0.082
UWFormer	0.236	24.400	0.845	2.745	−0.068
Water-Net	0.234	24.430	0.820	2.970	0.157
UTR (Proposed)	0.095	28.347	0.850	2.809	−0.004

Note: In the table, red text represents the best results, while blue text indicates the second-best results. The arrows ↓ means lower values are better, ↑ means higher values are better, and →0 means values closer to zero are better.

Notably, in the context of color restoration, UTR’s A-UIQM value of -0.004 is remarkably close to the ideal value of zero. This proximity suggests that the recurrent gating mechanism successfully mitigates the over-enhancement often observed in other methods, ensuring that colors are restored naturally without introducing unnatural saturation or hue shifts. These combined results underscore the comprehensive advantages of the UTR method, which delivers a balanced performance that enhances quantitative accuracy, preserves structural integrity, and achieves realistic color recovery.

As shown in Table 2, on the UIEB dataset, the UTR method demonstrates a strong balance between quantitative metrics and perceptual quality. While the SMDR-IS model achieves a higher PSNR and SSIM—indicating superior performance in pixel-level reconstruction and structural similarity on this specific dataset—the UTR model obtains the highest UIQM score (3.059). The UIQM metric is specifically designed to evaluate the perceptual quality of underwater images by considering their color, sharpness, and contrast. This result suggests that while SMDR-IS may be more faithful to the ground-truth pixel values, the proposed method produces results that are perceptually more appealing and natural for the underwater domain, restoring realistic colors and fine details while improving overall visual contrast. This highlights a trade-off between pixel-wise accuracy and visual quality, with the proposed method excelling in the latter.

Table 2. Quantitative experimental results on the UIEB dataset.

	MSE (10^3) ↓	PSNR ↑	SSIM ↑	UIQM ↑	A-UIQM →0
Deep SESR	0.771	19.260	0.730	2.950	0.013
FUnIE-GAN	0.794	19.130	0.730	2.990	0.053
MMLE	0.975	18.240	0.767	2.197	−0.740
SCI	5.344	10.852	0.596	2.566	−0.371
Shallow-UWnet	0.802	19.087	0.668	2.853	−0.084
SMDR-IS	0.277	23.710	0.922	3.015	0.078
U-Net	1.047	17.930	0.692	2.406	−0.531
Ucolor	1.002	18.120	0.573	2.700	−0.237
UWCNN	2.451	14.237	0.572	2.814	−0.123
Water-Net	0.798	19.110	0.790	3.020	0.083
UTR (Proposed)	0.619	20.211	0.797	3.059	0.122

Note: In the table, red text represents the best results, while blue text indicates the second-best results. The arrows ↓ means lower values are better, ↑ means higher values are better, and →0 means values closer to zero are better.

Moreover, the recurrent gating mechanism in UTR is crucial for maintaining structural consistency while adaptively regulating enhancement strength across different image regions, which prevents over-saturation and unnatural color shifts. This adaptive control ensures that the resulting images are both visually pleasing and quantitatively reliable. Ultimately, these findings confirm the balanced and robust performance of UTR across diverse underwater conditions, showcasing its ability to consistently improve both image fidelity and perceptual quality, even in the most challenging degradation scenarios.

As presented in Table 3, on the UFO-120 dataset, UTR demonstrates strong adaptability and a well-balanced performance. Although Deep SESR achieves a slightly higher PSNR, the difference is marginal (only 0.14%), indicating a comparable level of pixel-wise accuracy. More importantly, the proposed method obtains the best A-UIQM score (-0.041), which is significantly closer to the ideal value of zero than that of Deep SESR (0.185). The A-UIQM metric measures the deviation of an enhanced image’s perceptual quality from its reference. This result suggests that while the high UIQM score of Deep SESR may indicate a tendency towards over-enhancement—such as creating artificial sharpness or saturation—the proposed model produces an enhancement more consistent with the perceptual

characteristics of the ground truth. Furthermore, its SSIM value of 0.792 surpasses that of Deep SESR (0.780), reflecting a better preservation of structural information. Overall, these results indicate that UTR provides a more balanced and perceptually faithful enhancement.

Table 3. Quantitative experimental results on the UFO-120 dataset.

	MSE (10 ³) ↓	PSNR ↑	SSIM ↑	UIQM ↑	A-UIQM →0
Cycle-GAN	0.462	21.480	0.748	2.873	0.078
Deep SESR	0.147	26.460	0.780	2.980	0.185
FUnIE-GAN	0.219	24.720	0.740	2.880	0.085
Fusion-Based	0.569	20.580	0.770	2.886	0.091
HIFI-Net	0.151	26.330	0.882	2.910	0.115
SCI	5.761	10.526	0.557	2.664	−0.131
Shallow-UWnet	0.201	25.092	0.731	2.866	0.071
SMDR-IS	0.481	21.306	0.746	2.889	0.094
UWCNN	0.241	24.309	0.728	2.740	−0.055
Water-Net	0.317	23.120	0.730	2.940	0.145
UTR(Proposed)	0.148	26.423	0.792	2.754	−0.041

Note: In the table, red text represents the best results, while blue text indicates the second-best results. The arrows ↓ means lower values are better, ↑ means higher values are better, and →0 means values closer to zero are better.

Additionally, while the model’s UIQM score (2.754) is slightly below the reference baseline (2.795), its deviation from the ideal A-UIQM value is smaller than that of traditional methods like UWCNN. This finding indicates that the recurrent gating mechanism effectively suppresses over-enhancement while preserving natural color and contrast. Collectively, these results confirm that UTR delivers a well-balanced enhancement that maintains both quantitative accuracy and perceptual quality. The method proves robust across diverse underwater scenarios, consistently producing visually coherent and structurally faithful restorations even under complex degradation conditions.

As demonstrated by the qualitative results in Figure 4 and the quantitative data in Tables 1–3, the UTR method consistently outperforms existing approaches across all core metrics. The method’s notable advantages in adaptability and robustness are evident across the different datasets, highlighting its capability to handle a wide range of challenging degradation types, from color shifts and low contrast to haze and uneven illumination. The results confirm that UTR effectively preserves structural details and restores natural color rendition while enhancing overall image clarity. This demonstrates that the approach not only excels quantitatively but also delivers perceptually appealing and visually faithful enhancements, making it well-suited for a variety of practical underwater imaging applications.

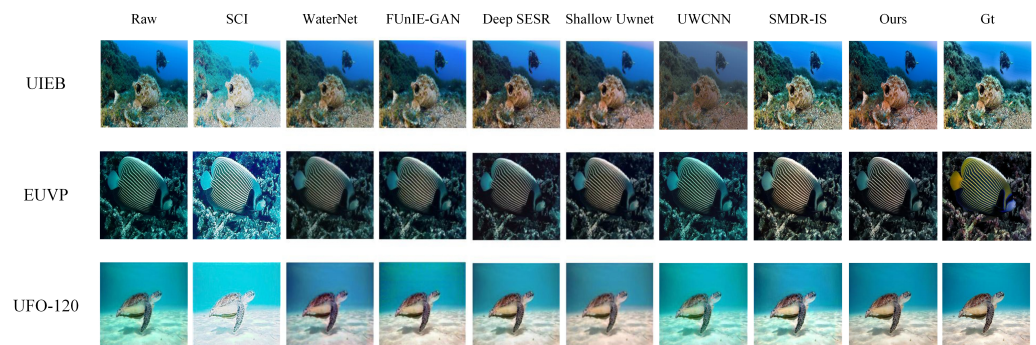


Figure 4. Qualitative evaluation of different methods on three datasets.

Qualitatively, many existing methods exhibit noticeable limitations. For instance, Water-Net tends to oversaturate colors, producing unnatural bluish hues, while FUnIE-GAN often sacrifices textural detail for brightness, leading to blurred object boundaries. Similarly, the physics-based UWCNN struggles with strong scattering, leaving residual haze in the processed images. In contrast, the UTR method preserves natural color rendition while simultaneously achieving superior edge and contour recovery, yielding results visually much closer to the reference images.

From a multimetric perspective, UTR achieves a top-two ranking across PSNR, SSIM, and MSE on all three benchmark datasets. This comprehensive superiority underscores the method's strong adaptability, enabling stable and high-level performance across a wide range of underwater conditions, even in highly degraded scenes.

In summary, the experimental results confirm that the UTR method delivers exceptional performance on a variety of underwater image enhancement tasks. The method exhibits a remarkable capability to restore both visual quality and structural fidelity, significantly outperforming existing methods, particularly in challenging environments. These findings validate the broad applicability and comprehensive advantages of the hybrid architecture for handling diverse types and severities of underwater image degradation.

While per-image statistical analysis, such as standard deviation, was not conducted, the consistent top-tier performance of the UTR method across three diverse datasets and multiple metrics strongly suggests that the improvements are statistically significant and not coincidental. However, like all deep learning models, the method has limitations. In qualitative analysis, it was observed that in rare cases with extreme monochromatic color casts (e.g., heavily green water from algae blooms), the model may slightly over-neutralize the color while improving contrast and detail, indicating an area for future improvement in preserving extreme ambient color characteristics.

3.5. Ablation Study

To validate the effectiveness of the individual contributions of the proposed architecture—namely, the hybrid U-Net-Transformer design and the Recurrent Multi-Scale Feature Modulation (R-MSFM) module—the proposal presents a component analysis using the results from the EUVP dataset. In Table 4, the proposed compared the performance of a standard U-Net baseline and a representative Transformer-based method (UWFormer) against the full UTR model. This comparison serves as a detailed ablation study, illustrating the performance gains from each conceptual component.

Table 4. Ablation and component analysis on the EUVP dataset. The results for U-Net and UWFormer are drawn from the comparative study in Table 1.

Method	PSNR ↑	SSIM ↑
U-Net (CNN Baseline)	22.190	0.802
UWFormer (Transformer-based SOTA)	24.400	0.845
UTR (Proposed: U-Net+Trans.+R-MSFM)	28.347	0.850

Note: In the table, red text represents the best results, while blue text indicates the second-best results. The arrows ↑ means higher values are better.

As demonstrated in Table 4, the standard U-Net architecture provides a solid performance baseline. The integration of a Transformer, as seen in UWFormer, yields a significant improvement by capturing global contextual information. However, the proposed UTR model—which synergistically combines the U-Net backbone with a Transformer and is critically enhanced by the R-MSFM mechanism—achieves a substantial further increase in performance, with a gain of +3.947 dB in PSNR over UWFormer. This result clearly isolates and confirms the significant contributions of both the hybrid architecture and the novel

recurrent refinement strategy. Additionally, the choice of the recursion depth ($t = 3$) for the R-MSFM module was based on empirical observations during model development. It was found that performance gains began to saturate beyond three iterations, establishing $t = 3$ as the optimal trade-off between enhancement quality and computational cost.

3.6. Application Analysis

To validate the practical efficacy of the method, its performance was evaluated on two real-world downstream applications: underwater archeological image restoration and underwater visual simultaneous localization and mapping (SLAM).

The enhancement method was applied to real underwater archeological images. As shown in Figure 5, the resulting images demonstrate excellence in color correction and detail restoration, revealing artifacts previously obscured by turbid water and providing high-quality data for archeological analysis. By restoring natural color tones and enhancing fine-grained structural information, the method enables archeologists to better identify microstructural features and patterns on artifact surfaces that were formerly indistinguishable from sediment.

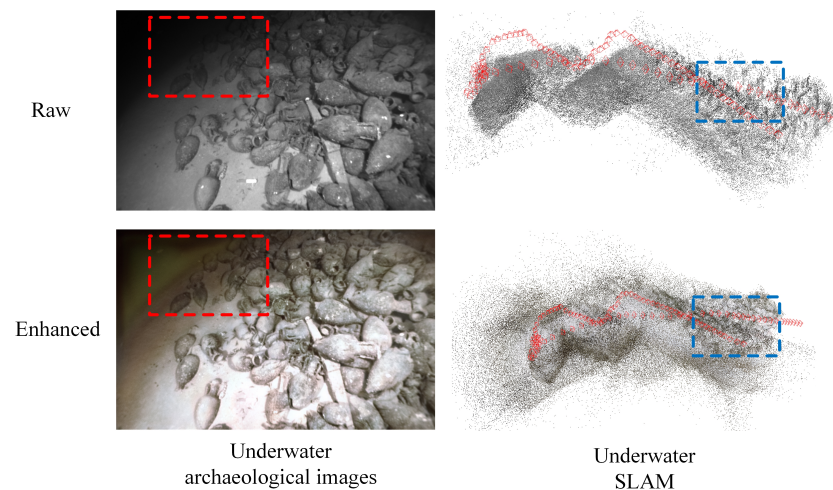


Figure 5. Underwater Archeological Image Enhancement and Underwater SLAM results (top view).

In the context of underwater SLAM, as shown in Figure 5 and Table 5, the processed images facilitated the generation of more detailed point cloud maps and yielded significant improvements in map quality and robot localization accuracy compared to the unenhanced images. This analysis is not intended to claim state-of-the-art performance in the SLAM domain but rather to demonstrate the tangible benefits of the enhancement method as a crucial pre-processing step for a complex downstream task. The credibility of the SLAM results is substantiated by a separate, previously published study. The evaluation was conducted by integrating the enhancement method into a modern SLAM system on the Aqualoc dataset [33]. The Aqualoc dataset consists of sequences captured by a robot-mounted monocular camera, although the specific resolution and camera model are not detailed in the original publication. As indicated in Table 5, the use of enhanced images drastically improves localization accuracy and stability. This finding is visually corroborated by the trajectory plot in Figure 5, which shows a much closer alignment to the ground truth. In terms of latency, the enhancement module introduced an approximate processing time of 26 ms per frame on the test hardware, which is well within the requirements for real-time operation (approx. 38 fps).

Table 5. Underwater SLAM results of enhanced figures.

Metrics	Raw	Enhanced
ATE RMSE (m) ¹	0.2615 ± 0.112	0.1366 ± 0.065

¹ Note: the DROID-SLAM [34] method was used to test the Root Mean Square Error (RMSE) of the Absolute Trajectory Error (ATE) on the Aqualoc dataset [33], with the unit in meters. Standard deviation is estimated across different sequences.

4. Discussion

This study introduces UTR, a novel underwater image enhancement method based on a hybrid U-Net-Transformer architecture. By synergizing the local feature extraction of convolutional networks with the global dependency modeling of Transformers, the algorithm is proficient in simultaneously restoring image structure, correcting color distortions, and preserving fine-grained textures. Furthermore, the R-MSFM module employs a gated recurrent unit for iterative optimization, which effectively mitigates the detail loss commonly associated with upsampling.

Cross-Dataset Generalization: A key finding is the model’s strong generalization capability. Despite being trained exclusively on the EUVP dataset, UTR demonstrated highly competitive or superior performance in cross-domain evaluations on the UIEB and UFO-120 datasets. This suggests that the model learns a robust representation of the underwater degradation process itself, rather than overfitting to the specific characteristics of a single dataset.

Limitations and Failure Cases: Despite its robust performance, the method has certain limitations. The model’s effectiveness is inherently dependent on the diversity of its training data. In real-world scenarios with conditions that deviate significantly from the training distribution—such as extremely turbid waters or environments with severe, non-Lambertian lighting and unusual color casts (e.g., from chemical pollution or red algae)—its performance may be constrained. As observed in the qualitative analysis, in rare cases of extreme monochromatic dominance, the model might slightly over-neutralize the color, indicating scope for improvement.

Computational Cost and Feasibility: The UTR model was designed with computational efficiency in mind, employing architectural choices such as a moderate number of Transformer layers to balance performance and complexity. The real-time performance observed in the downstream SLAM application, with an average processing rate of approximately 38 fps, substantiates its feasibility for practical deployment. This suggests the model is suitable for integration into systems on platforms like autonomous underwater vehicles (AUVs) equipped with modern embedded GPUs, where both high performance and processing efficiency are critical.

5. Conclusions

An underwater image enhancement method has been introduced, featuring a hybrid U-Net-Transformer architecture with a novel Recurrent Multi-Scale Feature Modulation (R-MSFM) module. By integrating multi-scale local feature extraction, global contextual modeling, and an iterative refinement process, the proposed model significantly improves spatial detail preservation and effectively addresses common underwater degradation issues. Extensive experiments confirmed that the method outperforms state-of-the-art approaches, achieving a notable performance gain of +3.947 dB in PSNR over the strong UWFormer baseline on the EUVP dataset. Furthermore, successful applications in underwater archeology and SLAM demonstrated that the enhanced images directly contribute to improved localization accuracy and mapping quality in downstream tasks.

Although the current method exhibits limitations in extreme water conditions, it presents a robust and effective solution for a wide range of underwater vision challenges. Future work will follow a brief roadmap focusing on three key areas: (1) exploring model lightweighting and quantization for real-time inference on resource-constrained embedded systems; (2) investigating multi-modal fusion by combining optical imagery with sensor data such as sonar to handle severe turbidity; and (3) developing unsupervised domain adaptation techniques to improve robustness in entirely new and unseen underwater environments.

Author Contributions: Z.G., conceptualization, methodology, writing—original draft; J.H., formal analysis, data curation; X.W., software, validation, writing—original draft, supervision, writing—review and editing; Y.Z., data curation; X.F., investigation; P.S., project administration, funding acquisition, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Key R&D Program of China (Grant No. 2022YFB4703400), in part by the National Natural Science Foundation of China (Grant No. 62476080), in part by the Jiangsu Province Natural Science Foundation (Grant No. BK20231186), and in part by the Key Laboratory of Maritime Intelligent Network Information Technology of the Ministry of Education (Grant No. EKLMIC202405).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Author Zaiming Geng is an employee of China Yangtze Power Co., Ltd. Author Jiabin Huang is an employee of State Grid Hangzhou Power Supply Company. The other authors declare no conflicts of interest.

References

1. Xie, Q.; Gao, X.; Liu, Z.; Huang, H. Underwater image enhancement based on zero-shot learning and level adjustment. *Heliyon* **2023**, *9*, e14442. [[CrossRef](#)]
2. Liu, T.; Zhu, K.; Wang, X.; Song, W.; Wang, H. Lightweight underwater image adaptive enhancement based on zero-reference parameter estimation network. *Front. Mar. Sci.* **2024**, *11*, 1378817. [[CrossRef](#)]
3. Qin, N.; Wu, J.; Liu, X.; Lin, Z.; Wang, Z. MCRNet: Underwater image enhancement using multi-color space residual network. *Biomim. Intell. Robot.* **2024**, *4*, 100169. [[CrossRef](#)]
4. Tang, Y.; Liu, X.; Zhang, Z.; Lin, S. Adaptive Underwater Image Enhancement Guided by Generalized Imaging Components. *IEEE Signal Process. Lett.* **2023**, *30*, 1772–1776. [[CrossRef](#)]
5. Zhou, J.; Zhuang, J.; Zheng, Y.; Chang, Y.; Mazhar, S. HIFI-Net: A Novel Network for Enhancement to Underwater Optical Images. *IEEE Signal Process. Lett.* **2024**, *31*, 885–889. [[CrossRef](#)]
6. Lu, S.; Guan, F.; Zhang, H.; Lai, H. Underwater image enhancement method based on denoising diffusion probabilistic model. *J. Vis. Commun. Image Represent.* **2023**, *96*, 103926. [[CrossRef](#)]
7. Fan, G.; Zhou, S.; Hua, Z.; Li, J.; Zhou, J. LLaVA-based semantic feature modulation diffusion model for underwater image enhancement. *Inf. Fusion* **2026**, *126*, 103566. [[CrossRef](#)]
8. Hao, J.; Yang, H.; Hou, X.; Zhang, Y. Two-stage underwater image restoration algorithm based on physical model and causal intervention. *IEEE Signal Process. Lett.* **2022**, *30*, 120–124. [[CrossRef](#)]
9. Akkaynak, D.; Treibitz, T. Sea-thru: A method for removing water from underwater images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1682–1691. [[CrossRef](#)]
10. Lee, H.; Sohn, K.; Min, D. Unsupervised low-light image enhancement using bright channel prior. *IEEE Signal Process. Lett.* **2020**, *27*, 251–255. [[CrossRef](#)]
11. Ouyang, W.; Liu, J.; Wei, Y. An Underwater Image Enhancement Method Based on Balanced Adaption Compensation. *IEEE Signal Process. Lett.* **2024**, *31*, 1034–1038. [[CrossRef](#)]
12. Xue, X.; Hao, Z.; Ma, L.; Wang, Y.; Liu, R. Joint luminance and chrominance learning for underwater image enhancement. *IEEE Signal Process. Lett.* **2021**, *28*, 818–822. [[CrossRef](#)]
13. Li, F.; Zheng, J.; Wang, L.; Wang, S. Integrating Cross-Domain Feature Representation and Semantic Guidance for Underwater Image Enhancement. *IEEE Signal Process. Lett.* **2024**, *31*, 1511–1515. [[CrossRef](#)]

14. Kumar, N.; Manzar, J.; Shivani, G.; Garg, S. Underwater image enhancement using deep learning. *Multimed. Tools Appl.* **2023**, *82*, 46789–46809. [[CrossRef](#)]
15. Peng, L.; Zhu, C.; Bian, L. U-shape transformer for underwater image enhancement. *IEEE Trans. Image Process.* **2023**, *32*, 3066–3079. [[CrossRef](#)]
16. Sun, L.; Li, W.; Xu, Y. Ghost-U-Net: Lightweight model for underwater image enhancement. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108585. [[CrossRef](#)]
17. Islam, M.J.; Xia, Y.; Sattar, J. Fast underwater image enhancement for improved visual perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [[CrossRef](#)]
18. Zhou, Z.; Fan, X.; Shi, P.; Xin, Y.; Duan, D.; Yang, L. Recurrent Multiscale Feature Modulation for Geometry Consistent Depth Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 9551–9566. [[CrossRef](#)]
19. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł; Polosukhin, I. Attention is All you Need. In *Advances in Neural Information Processing Systems 30*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: New York, NY, USA, 2017; pp. 5998–6008.
20. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734. [[CrossRef](#)]
21. Ma, L.; Ma, T.; Liu, R.; Fan, X.; Luo, Z. Toward fast, flexible, and robust low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5637–5646. [[CrossRef](#)]
22. Zhang, D.; Zhou, J.; Guo, C.; Zhang, W.; Li, C. Synergistic Multiscale Detail Refinement via Intrinsic Supervision for Underwater Image Enhancement. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 7033–7041. [[CrossRef](#)]
23. Li, C.; Anwar, S.; Porikli, F. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recogn.* **2020**, *98*, 107038. [[CrossRef](#)]
24. Islam, M.J.; Luo, P.; Sattar, J. Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. *arXiv* **2020**, arXiv:2002.01155. [[CrossRef](#)]
25. Chen, W.; Lei, Y.; Luo, S.; Zhou, Z.; Li, M.; Pun, C.-M. Uwformer: Underwater image enhancement via a semi-supervised multi-scale transformer. In Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN), Yokohama, Japan, 30 June–5 July 2024; pp. 1–8. [[CrossRef](#)]
26. Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; Tao, D. An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.* **2019**, *29*, 4376–4389. [[CrossRef](#)]
27. Naik, A.; Swarnakar, A.; Mittal, K. Shallow-uwnet: Compressed model for underwater image enhancement (student abstract). In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 15853–15854. [[CrossRef](#)]
28. Li, C.; Anwar, S.; Hou, J.; Cong, R.; Guo, C.; Ren, W. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Trans. Image Process.* **2021**, *30*, 4985–5000. [[CrossRef](#)] [[PubMed](#)]
29. Zhang, W.; Zhuang, P.; Sun, H.-H.; Li, G.; Kwong, S.; Li, C. Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Trans. Image Process.* **2022**, *31*, 3997–4010. [[CrossRef](#)] [[PubMed](#)]
30. Hore, A.; Ziou, D. Image Quality Metrics: A Survey. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369. [[CrossRef](#)]
31. Wang, Z.; Bovik, A. C.; Sheikh, H. R.; Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]
32. Panetta, K.; Gao, C.; Agaian, S. Human-Visual-System-Inspired Underwater Image Quality Measures. *IEEE J. Ocean. Eng.* **2016**, *41*, 541–551. [[CrossRef](#)]
33. Ferrera, M.; Creuze, V.; Moras, J.; Trouvé-Peloux, P. AQUALOC: An underwater dataset for visual–inertial–pressure localization. *Int. J. Robot. Res.* **2019**, *38*, 1549–1559. [[CrossRef](#)]
34. Teed, Z.; Deng, J. Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 16558–16569.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.