

Article

TFCNet: A Hybrid Architecture for Multi-Task Restoration of Complex Underwater Optical Images

Shengya Zhao ^{1,2} , Xiufen Ye ^{1,*} , Xinkui Mei ^{1,3} , Shuxiang Guo ⁴ and Haibin Qi ^{2,*}

¹ College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China; zhaosy@hrbeu.edu.cn (S.Z.); meixinkui@hrbeu.edu.cn (X.M.)

² Deep Sea Technology Department, National Deep Sea Center, Qingdao 266037, China

³ Harbin Electric Science and Technology Company Limited, Harbin 150001, China

⁴ Department of Electronic and Electrical Engineering, Southern University of Science and Technology, Shenzhen 518055, China; guo.shuxiang@sustech.edu.cn

* Correspondence: yexiufen@hrbeu.edu.cn (X.Y.); qhb@ndsc.org.cn (H.Q.)

Abstract: Underwater optical images are crucial in marine exploration. However, capturing these images directly often results in color distortion, noise, blurring, and other undesirable effects, all of which originate from the unique physical and chemical properties of underwater environments. Hence, various factors need to be comprehensively considered when processing underwater optical images that are severely degraded under complex lighting conditions. Most existing methods resolve one issue at a time, making it challenging for these isolated techniques to maintain consistency when addressing multiple degradation factors simultaneously, often leading to unsatisfactory visual outcomes. Motivated by the global modeling capability of the Transformer, this paper introduces TFCNet, a complex hybrid-architecture network designed for underwater optical image enhancement and restoration. TFCNet combines the benefits of the Transformer in capturing long-range dependencies with the local feature extraction potential of convolutional neural networks, resulting in enhanced restoration results. Compared with baseline methods, the proposed approach demonstrated consistent improvements, where it achieved minimum gains of 0.3 dB in the PSNR and 0.01 in the SSIM and a 0.8 reduction in the RMSE. TFCNet exhibited a commendable performance in complex underwater optical image enhancement and restoration tasks by effectively rectifying color distortion, eliminating marine snow noise to a certain degree, and restoring blur.

Keywords: multi-task underwater image restoration; hybrid architecture; Transformer encoding; convolutional neural network decoding; denoising; deblurring



Academic Editors: Weicheng Cui, Jian Zhang and Daqi Zhu

Received: 25 April 2025

Revised: 26 May 2025

Accepted: 27 May 2025

Published: 29 May 2025

Citation: Zhao, S.; Ye, X.; Mei, X.; Guo, S.; Qi, H. TFCNet: A Hybrid Architecture for Multi-Task Restoration of Complex Underwater Optical Images. *J. Mar. Sci. Eng.* **2025**, *13*, 1090. <https://doi.org/10.3390/jmse13061090>

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/>).

1. Introduction

Terrestrial resources are insufficient to meet the demands of human development. Thus, the development of marine resources, which cover 71% of the Earth's surface [1], is crucial. Underwater robots are vital for exploring these resources, where underwater optical images serve as the “eyes” of these robots, providing essential data and information for researchers [2].

The underwater imaging environment is influenced by factors such as marine organisms, suspended particles, and poor lighting conditions, which makes it more complex than imaging on land. Underwater images often exhibit a low contrast, color distortion, and blurred details [3]. The movement of underwater robots can further exacerbate these issues, particularly due to particles, including marine organisms, floating feces, suspended

sediments, and other inorganic matter. These particles vary in size, shape, and transparency. Scattering occurs when underwater light encounters small particles and is reflected back to the camera, resulting in a low contrast and blurriness in the captured images, as shown in Figure 1. Underwater optical images exhibit color distortion and are influenced by many impurities and motion blur. Consequently, these images cannot be directly used for target detection and require additional restoration.

The underwater environment is complex and dynamic, with suspended particles, light scattering, and light absorption, leading to color distortion and low contrast. The environment is also often accompanied by marine snow noise caused by plankton or sediment particles. Therefore, multiple issues must be simultaneously considered when dealing with severely degraded underwater optical images under complex lighting conditions, including the removal of marine snow noise, the restoration of blurring, and the basic color correction task. However, most of the existing methods focus on solving one problem at a time. Therefore, these isolated techniques struggle to maintain consistency when addressing multiple degradation factors simultaneously, and consequently, they do not achieve the desired visual results [4].

Furthermore, datasets designed for the multi-task restoration of complex underwater optical images are lacking. In particular, when dealing with varying lighting conditions, marine snow noise levels, and water composition, existing datasets lack the diversity required to represent diverse, complex scenarios, limiting the generalizability of current research [5].

Existing methods for restoring underwater optical images typically focus on color correction and enhancement rather than addressing impurity occlusion and blurriness in real underwater images. Moreover, the lack of publicly available datasets further complicates the restoration process [6]. To address these challenges, Mei et al. [7] proposed the Underwater Image Enhancement Benchmark Dataset (UIEBD-Snow dataset) to explore underwater optical image restoration involving impurities. However, datasets for restoring underwater images affected by motion blur are lacking.

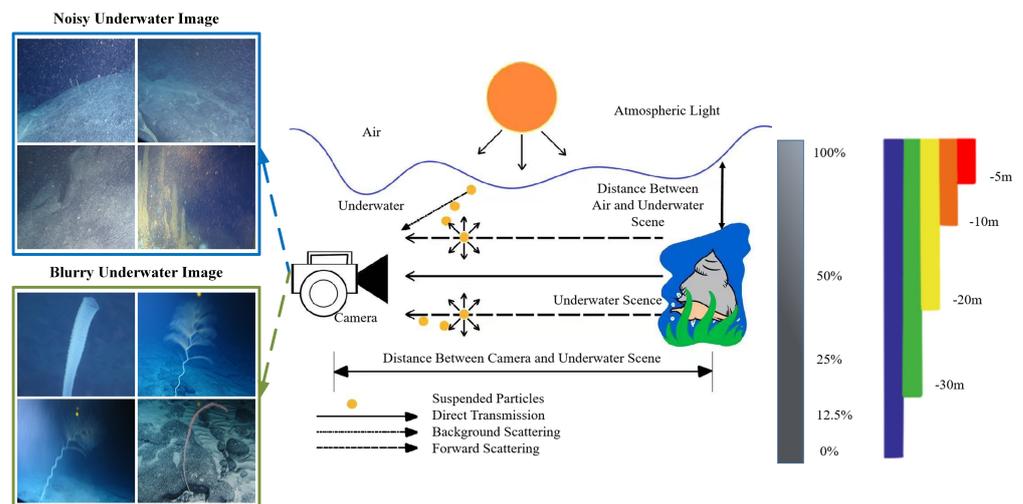


Figure 1. Schematic diagram illustrating the principle of underwater optical imaging.

Underwater image restoration is challenging, requiring color correction, impurity removal, and compensation for motion blur [8]. From a global perspective, restoration can be divided into two main parts: impurity removal and image enhancement. With the widespread application of deep learning technology, most underwater image enhancement and restoration methods often rely on pure convolutional neural network (CNN) structures or U-shape Transformer structures. A CNN offers several advantages. However, its

small receptive field makes it challenging to capture global features. Conversely, visual Transformers, such as U-shape Transformers, have been used for restoration tasks in underwater imaging tasks based on their ability to capture global information [9]. However, Transformers often lack the translation invariance and local correlation characteristics of a CNN, consequently requiring a large volume of training data to surpass CNN's performance. Therefore, combining the Transformer with a CNN to optimize their advantages and preserve global and local features is promising. Ordinary Transformer visual entities vary considerably because different shooting angles of the same object can lead to notable differences in binary images. Additionally, the performance of visual Transformers may vary across different scenarios. When dealing with high-resolution, pixel-dense images, the Vision Transformer model incurs a computational cost proportional to the square of the pixels due to self-attention [10].

Given the above background, this article introduces TFCNet, a network that integrates CNN-based local features and Transformer-based global representations. TFCNet incorporates the Transformer for encoding and CNN for decoding. TFCNet considerably enhances the quality of underwater image restoration by combining the Transformer's focus on global information with the CNN's ability to use underlying image features. Simultaneously, we propose an additional underwater image restoration dataset, UIEBD-Blur, by building on the UIEBD-Snow dataset to account for blurring. TFCNet has demonstrated promising results using the UIEBD-Snow and UIEBD-Blur datasets and can concurrently perform color correction, denoising, and deblurring for underwater images. In summary, the contributions of this article include the following:

1. This paper proposes TFCNet, a multi-task restoration method for underwater optical images, which is based on a hybrid architecture. It integrates a Swin Transformer-based encoder module with spatial adaptivity to efficiently capture global image features. The decoder module, comprising a CNN without activation functions, reduces computational complexity.
2. This paper introduces the UIEBD-Blur dataset, which was specifically designed for motion blur recovery in underwater optical images. Experiments using the UIEBD-Blur and UIEBD-Snow datasets demonstrated that TFCNet achieved superior visual outcomes in marine snow noise removal and color correction.
3. This study validated the feasibility of TFCNet in enhancing and restoring complex underwater optical images and its superiority to other methods. Ablation experiments further investigated the effectiveness of the hybrid architecture by evaluating the contributions of the Transformer-based encoder and CNN-based decoder within the framework.

2. Related Works

In underwater image restoration research, scholars focus on image enhancement and color correction techniques. These techniques can be categorized into three main types: model-free, model-based, and deep learning-based methods. K. Iqbal et al. [11] proposed an unsupervised color correction method for enhancing underwater images. This method is based on color balancing and contrast correction of the RGB and HSI color models. Hitam et al. [12] introduced the CLAHE method, which combines results from the RGB and HSV color models using Euclidean norms. Fu et al. [13] developed a widely used classical method based on retinex. Drews et al. [14] proposed Underwater DCP (UDCP), which applies adaptive DCP to estimate underwater scene transmission. Peng et al. [15] introduced Generalized DCP (GDCP) for image restoration, incorporating adaptive color correction into the image formation model. In addition, Mei et al. [16] proposed a method based on the optical geometric properties. These methods enhance underwater image details;

however, they often excessively amplify noise and distort colors, occasionally resulting in over-enhancement. Fu et al. [17] proposed a network based on probabilistic methods to obtain the enhancement distribution for degraded underwater images. Li et al. [18] developed WaterGAN, a network based on the generative adversarial network (GAN), to generate datasets of underwater images using air and depth pairings. These datasets are used for the unsupervised pipeline-based color correction of underwater images. Wang et al. [19] introduced UWdepth, a self-supervised model that obtains depth information from underwater scenes using monocular sequences. This depth information is subsequently used to enhance underwater images. Fabbri et al. [20] developed UGAN, a GAN-based model, to improve underwater image quality. Han et al. [21] designed their method by leveraging contrastive learning and generative adversarial networks to maximize the mutual information between raw and restored images. Islam et al. [22] presented a model based on conditional generative adversarial networks to enhance underwater images in real time. In addition, Zhou et al. [23] researched underwater image enhancement using deep learning techniques.

However, existing methods for enhancing underwater images only address color correction and do not fully satisfactorily restore underwater images degraded by visual impurities. This remains a challenge in the field. A few methods are available for removing impurities and blurring from underwater images, and even fewer methods can simultaneously enhance the image.

Jiang et al. [24] proposed a UDnNet network based on a GAN with skip connections to model the mapping relationship in underwater optical images contaminated by marine snow noise. This method generates marine-snow-free optical images from noisy inputs, partially suppressing noise. However, its limitations include high computational costs and lengthy training cycles. Sun et al. [25] introduced a CNN called NR-CCNet, which incorporates a recurrent learning strategy and an attention mechanism to address marine snow noise. This approach reduces noise to some extent; however, its restoration performance remains unsatisfactory. Sun et al. [26] developed a progressive multi-branch embedded fusion network to further improve the performance. This framework uses a dual-branch hybrid encoder–decoder module equipped with a triple attention mechanism to fuse distorted images and their sharpened versions, focusing on noisy regions and learning contextual features. It progressively learns a nonlinear mapping from degraded inputs, and the final output is refined and enhanced using a three-branch hybrid encoder–decoder module at each stage. Nevertheless, the multi-branch architecture increases the model complexity, and its effectiveness in marine snow noise removal remains limited because the network primarily targets underwater optical image enhancement.

Furthermore, image enhancement and restoration methods developed for atmospheric conditions also provide some valuable insights for marine snow noise removal.

The authors of [27] advocated for a revised median filter as a potent approach to mitigate the effects of underwater contaminants on these images. DB-ResNet [28] is a specialized structure, termed a “deep detail network”, which was specifically designed to remove natural raindrop patterns from captured images. A deep residual network (ResNet) is a parameter layer that encapsulates more complex image features and streamlines the network’s structure by reducing the mapping distance between the input and output features. Other authors [29] proposed a progressive optimization residual network (Progressive ResNet (PRN)) and a progressive recurrent network (PReNet) for image de-raining. Ren et al. [29] introduced PRN and a progressive recurrent network (PReNet) for image deraining. Maxim [30] is the latest MLP-based U-Net backbone network that combines global and local perceptual fields and can be directly applied to high-resolution images. Restormer [31] is an efficient Transformer that incorporates several pivotal enhancements

in the design of its improved multi-head attention and feed-forward networks. Multi-stage progressive restoration network (MPRNet) [32] is an innovative, collaborative design with a multi-stage structure aimed at learning the recovery features of degraded inputs while decomposing the entire recovery process. MPRNet learns context-dependent features using an encoder–decoder architecture and subsequently combines them with high-resolution branches that better preserve local information.

These methods are highly competitive in natural image restoration in the air; however, land-based and underwater imaging models cannot be used interchangeably. Sato et al. [33] proposed an underwater image restoration dataset multi-scale residual block (MSRB), which only performs underwater image denoising and cannot perform color correction. Mei et al. [7] proposed a lightweight baseline named UIR-Net, which simultaneously recovers and enhances underwater images while achieving notable recovery results. However, it still encounters limitations in color correction.

3. Method

Achieving satisfactory results for complex underwater optical image enhancement and restoration tasks using either a single Transformer architecture or a single convolutional neural network (CNN) structure is challenging.

To address these challenges, a hybrid architecture model is proposed in this study. This approach aims to maximize the local features and global representations by leveraging the strengths of the Transformer and CNN structures. The proposed multi-task restoration network, TFCNet, which is based on this hybrid architecture, is illustrated in Figure 2.

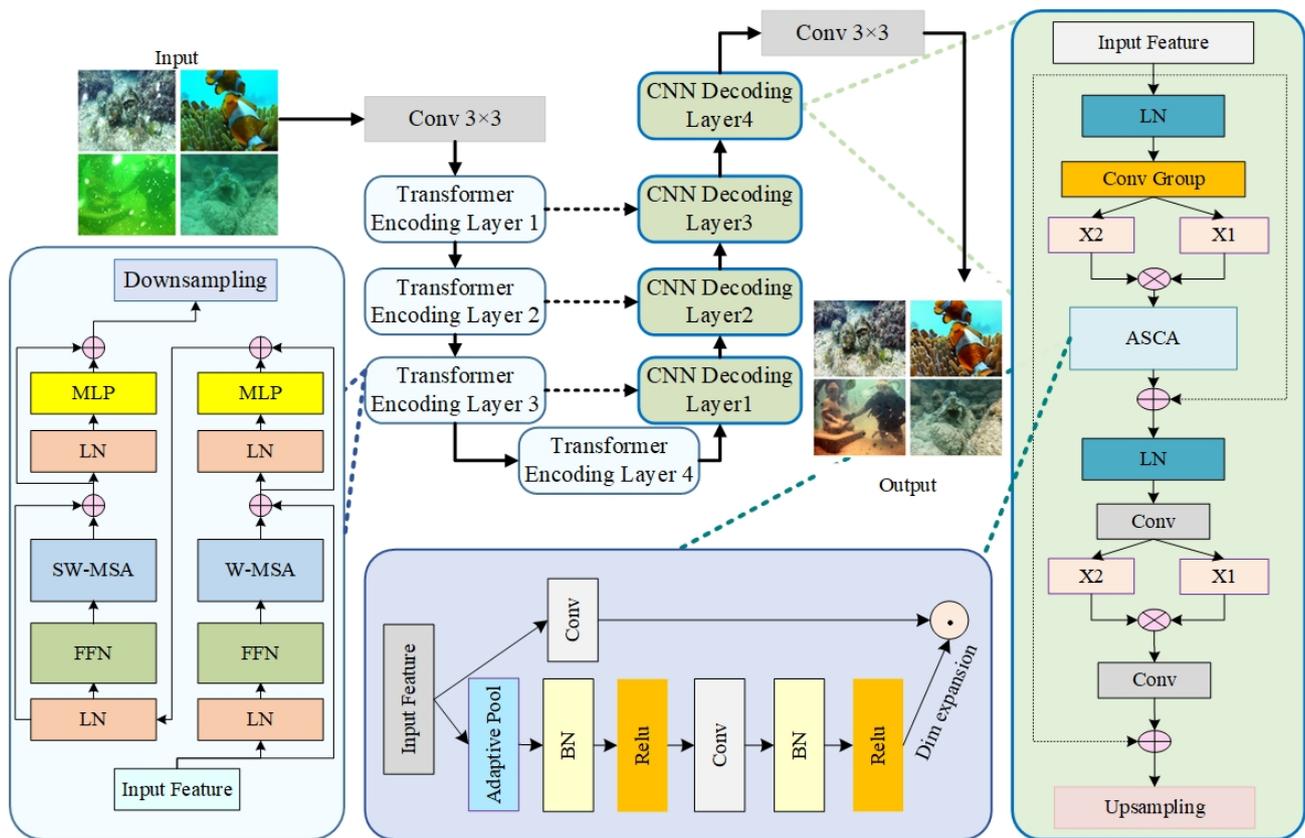


Figure 2. The structure of TFCNet; it is divided into a Transformer encoder and CNN decoder.

3.1. The Overall Structure of TFCNet

TFCNet integrates the strengths of the Transformer and CNN models. Transformers focus on global information but often overlook details at low resolutions, which impedes

the decoder’s ability to restore the pixel size, frequently leading to coarse results. Conversely, CNN models can effectively compensate for this limitation of Transformers. Thus, combining these two models presents greater advantages.

The entire TFCNet network adopts a hierarchical design encompassing eight stages. Initially, the encoder section serves as the core of TFCNet. Each stage begins by reducing the resolution of the input feature map through the Swin Transformer and downsampling layers, executing downsampling and progressively expanding the receptive field to capture global information. The encoder initially applies four stages of Swin Transformer structures to the input image, embedding image blocks obtained through the CNN into the feature map, necessitating positional encoding. Subsequently, the features extracted by the Transformer are passed on to the decoder, which uses conventional transposed convolution upsampling to restore the image pixels.

The overall processing pipeline is as follows: the underwater optical image requiring enhancement and restoration is the TFCNet network input. Considering the UIEBD-Blur dataset as an example, where I_{blur} represents the complex underwater optical image to be restored, the processing steps of the TFCNet encoder module can be simplified as follows:

The first-layer encoder module, based on the Swin Transformer architecture, is computed as follows:

$$FM_1 = f_1(W_1 * I_{blur} + B_1). \tag{1}$$

The encoder modules in the second to fourth layers, based on the Swin Transformer architecture, are computed as follows:

$$FM_i = f_i(W_i * FM_{(i-1)} + B_i), \tag{2}$$

where W_i represents the weight matrix of the i -th layer encoder module based on the Swin Transformer architecture, B_i denotes the bias term of the i -th layer encoder module, and f_i corresponds to the activation function of the i -th layer encoder module.

Subsequently, the processing steps of the decoder module, which is based on the CNN architecture, can be simplified as follows:

The first-layer decoder module, based on the CNN architecture, is computed as follows:

$$FM_5 = g_1(V_1 * FM_4 + C_1). \tag{3}$$

The decoder modules in the second to fourth layers, based on the CNN architecture, are computed as follows:

$$FM_{(i+4)} = g_i(V_i * FM_{(i+3)} + C_i), \tag{4}$$

where V_i represents the weight matrix of the i -th layer decoder module based on the CNN architecture, C_i denotes the bias term of the i -th layer decoder module, and g_i corresponds to the activation function of the i -th layer decoder module.

The output layer of TFCNet is as follows:

$$I_{deblur} = W_o * FM_8 + B_o, \tag{5}$$

where W_o is the weight matrix, and B_o is the bias of the output layer.

3.2. Encoding Network Design

The TFCNet network comprises four encoder modules based on the Swin Transformer structure and adopts a five-stage hierarchical design. Each stage progressively reduces the resolution of the input feature map and gradually expands the receptive field, akin to a CNN. The structure of the module, as Figure 3 illustrates, includes layer normalization, feature fusion normalization, a multi-layer perceptron, and an attention mechanism module.

In addition to the fundamental Swin Transformer module, a feature fusion normalization module is incorporated to further refine the layer-normalized features. Specifically, the image features processed via layer normalization undergo max and average pooling to extract high-frequency and global features, respectively. These features are then fused through stacking and convolutional operations. Subsequently, a sigmoid function is applied to obtain feature weight information, which is used to derive the feature information after suppressing the marine snow noise.

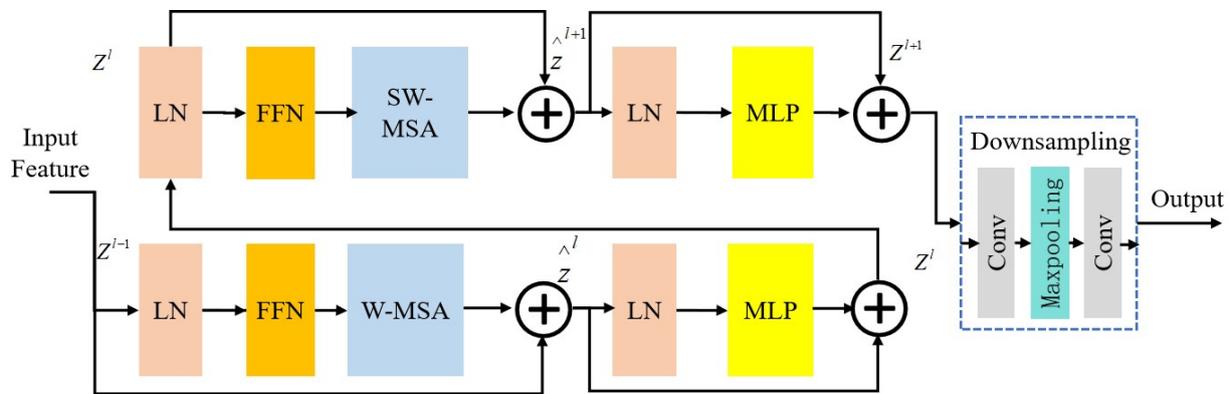


Figure 3. Schematic diagram of the TFCNet encoder module based on the Swin Transformer architecture.

The computational workflow of each Swin Transformer-based encoder module in the TFCNet network is detailed as follows:

$$\hat{z}^l = W - MSA \left(MaxAvg \left(LN \left(z^{l-1} \right) \right) \right) + z^{l-1}, \tag{6}$$

where $W - MSA()$ represents the window-based multi-head self-attention operation, $MaxAvg()$ denotes the min-max normalization, and $LN()$ corresponds to the layer normalization.

$$z^l = MLP \left(LN \left(\hat{z}^l \right) \right) + \hat{z}^l, \tag{7}$$

where $MLP()$ represents the multi-layer perceptron operation.

$$\hat{z}^{l+1} = SW - MSA \left(MaxAvg \left(LN \left(z^l \right) \right) \right) + z^l, \tag{8}$$

where $SW - MSA()$ represents the shifted window-based multi-head self-attention operation.

$$z^{l+1} = MLP \left(LN \left(\hat{z}^{l+1} \right) \right) + \hat{z}^{l+1}, \tag{9}$$

where \hat{z}^l and z^l represent the output features of the attention mechanism modules in each module. Finally, after further downsampling operations, the corresponding feature information is the output.

3.3. Decoding Network Design

The system becomes correspondingly complex because the complex multi-task restoration of underwater optical images necessitates simultaneous color correction, marine snow noise elimination, and blur restoration. Under such circumstances, nonlinear activation functions, such as Sigmoid, ReLU, GELU, and Softmax, are not essential [34]. Moreover, these nonlinear activation functions can be replaced by multiplication or directly removed. This ensures effective image enhancement and restoration and reduces the computational cost of the network. Guided by this concept, the TFCNet network constructs the decoder module based on a baseline network that does not require activation functions [34].

Figure 4 reveals that in this decoder module, the structure before upsampling has eliminated the nonlinear functions. The Gated Linear Units, which can introduce nonlinear computations, are replaced by the product of two feature maps, denoted as ϕ and θ . This reduces the computational complexity of the decoder module to a certain degree, thereby enhancing the efficiency of TFCNet.

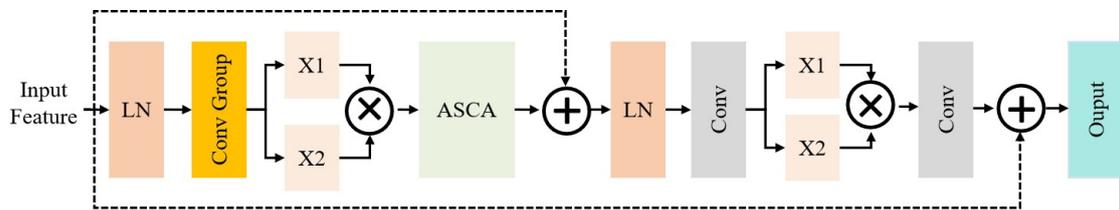


Figure 4. Schematic diagram of the CNN decoder in TFCNet based on the activation function-free network baseline.

The Adaptive Spatial Channel Attention (ASCA) mechanism was introduced to accelerate the convergence speed of the model, mitigate the risk of overfitting, and stabilize the model. Building on the SCA attention mechanism, a batch normalization layer was added between two convolutional layers, speeding up the convergence of the model and reducing the overfitting risk. Additionally, extra convolutional layers and ReLU activation functions were incorporated to enhance the model’s nonlinear expressive capacity and aid in extracting more complex features. Figure 5 illustrates the ASCA mechanism.

X represents the input feature map, and $dw_channel$ denotes the number of channels. Divide the number of channels of the input feature map by 2 for $in_channels$ and by 4 for $mid_channels$. W_i and B_i represent the weight and bias of the i -th conventional layer, respectively, and BN_i represents the i -th batch normalization layer. Finally, ReLU stands for the ReLU activation function.

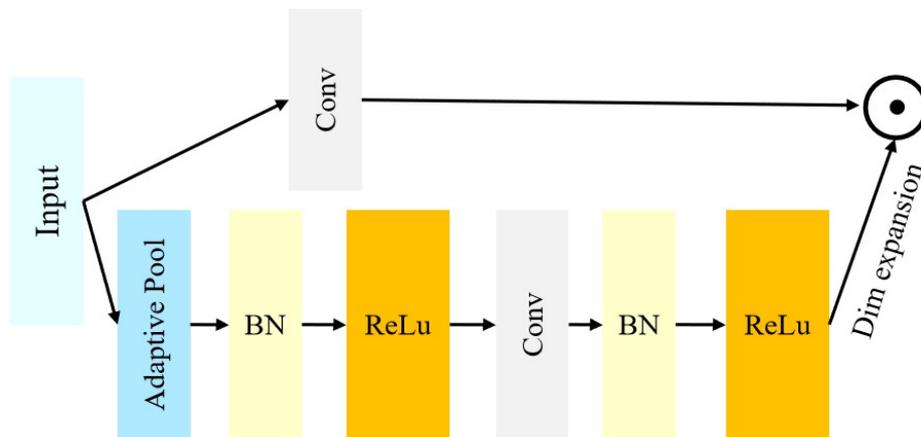


Figure 5. Schematic diagram of the Adaptive Spatial Channel Attention (ASCA) in TFCNet.

Adaptive average pooling:

$$x_{pooled} = AdaptiveAvgPool2d(x). \tag{10}$$

The computation formula for the i -th convolutional layer is as follows:

$$x_{conv_i} = Conv2d(x_{pooled}, W_i, B_i). \tag{11}$$

The i -th batch normalization layer:

$$x_{bn_i} = BN1(x_{conv_i}). \tag{12}$$

The i -th ReLU activation function:

$$x_{relu_i} = ReLU(x_{bn_i}). \tag{13}$$

Output: x_{relu_i} .

3.4. Loss Function

Regarding the loss function, this article adopts the commonly used loss functions in underwater image restoration, L_1 and L_{SSIM} , as follows:

$$L_{ALL} = k_1 L_1 + k_2 L_{SSIM}, \tag{14}$$

where $k_1 = 0.8$ and $k_2 = 0.2$.

The L_1 loss (Mean Absolute Error (MAE)) is the mean distance between the predicted value x of the model and the true value y . We used the L_1 loss to measure the pixel level loss between the reference network and the training results as follows:

$$L_1 = \frac{1}{N} \sum |x - y|. \tag{15}$$

The SSIM (structural similarity) loss is considered an indicator for luminance, contrast, and structure, incorporating human visual perception. We can obtain it as follows:

$$\begin{aligned} SSIM(x, y) &= [l(x, y)^\alpha * c(x, y)^\beta * s(x, y)^\lambda] \\ &= \frac{(2\mu_x\mu_y + c_1)(2\delta_x\delta_y + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\delta_x^2 + \delta_y^2 + c_2)}, \end{aligned} \tag{16}$$

where $l(x, y)$, $c(x, y)$, and $s(x, y)$ are as follows:

$$\begin{aligned} l(x, y) &= \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \\ c(x, y) &= \frac{2\delta_x\delta_y + c_2}{\delta_x^2 + \delta_y^2 + c_2} \\ s(x, y) &= \frac{\delta_{xy} + c_3}{\delta_x + \delta_y + c_3}. \end{aligned} \tag{17}$$

Therefore, we can determine L_{SSIM} as follows:

$$L_{SSIM} = 1 - SSIM(x, y). \tag{18}$$

4. Experimental Datasets and Discussion

4.1. Introduction to the Experimental Datasets

Experimental comparative analysis was conducted on two datasets, UIEBD-Snow and UIEBD-Blur, to verify the effectiveness and generalizability of the proposed method, which a focus on the impurities and motion blur in the underwater image restoration process. Both datasets are based on the commonly used public dataset UIEBD [35]. Figure 6 presents a schematic diagram of the UIEBD-Snow dataset construction.

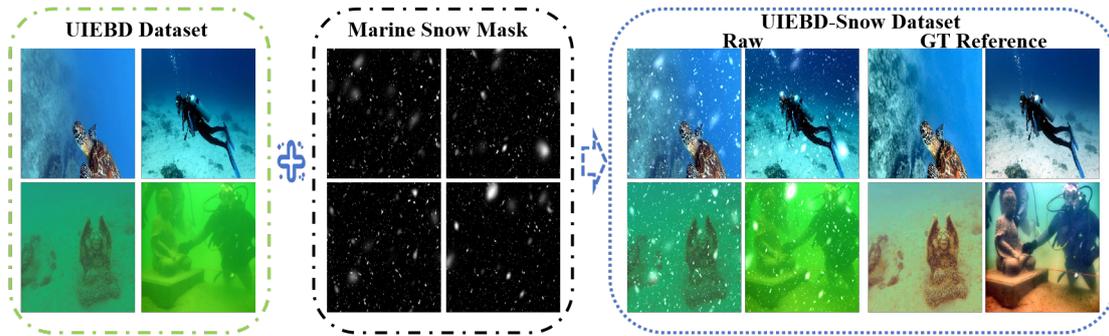


Figure 6. Schematic diagram of the UIEBD-Snow dataset construction.

Detail blurring in underwater optical images is expressed as a blur effect along a direction, with the extent of blurring being related to underwater imaging conditions and the speed of the underwater vehicle. Gaussian blur is a uniform blur treatment without direction. However, a directional kernel (motion blur kernel) was created in this study and combined with Gaussian blur to achieve a directional Gaussian blur effect, thereby simulating motion blur. The Gaussian function is the core concept of Gaussian blur. Figure 7 illustrates the entire dataset construction process.

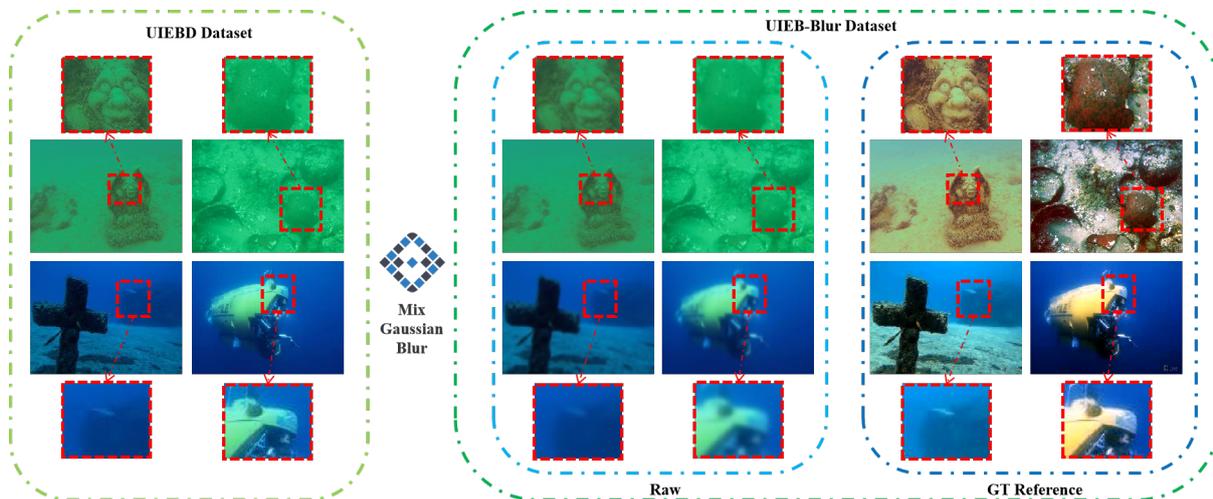


Figure 7. Schematic diagram of UIEBD-Blur dataset construction.

4.2. Training Parameter Settings

During the experimental process, the dataset was partitioned into 80% for training, 10% for testing, and 10% for validation. All other deep learning-based comparison methods were trained and tested using this identical data-splitting ratio.

The TFCNet model is an end-to-end trained model configured to perform underwater optical image enhancement and restoration tasks, encompassing color correction, denoising, and deblurring. Specifically, the Adam optimizer with an initial learning rate of 1×10^{-4} was used. Considering the model depth, the warm-up strategy gradually improved the learning

efficiency. The network was trained on a 256×256 image patch that was randomly cropped from training images. The batch size was 4. The learning rate was 1×10^{-4} . The number of epochs was 200. The model was trained on one 4090 GPU, which could be completed in 4 h.

4.3. Comparison with State-of-the-Art Methods for Underwater Image Enhancement

4.3.1. Qualitative Evaluations

The proposed method enhances and restores complex underwater optical images, including color correction. The experimental comparison includes four methods designed for underwater optical image enhancement with demonstrated effectiveness: Deepwave [36], PUIENet [17], Shallow [37], and FUnIE-GAN [22]. Additionally, three state-of-the-art methods for general optical image enhancement and restoration, DGUNet [38], MPRNet [32], and UIRNet [7], were included.

The comparative results of TFCNet and other methods on the UIEBD-Snow dataset in Figure 8 reveal that the methods solely designed for underwater optical image enhancement were ineffective at eliminating marine snow noise in complex underwater optical images. These methods focus on color correction and achieved suboptimal results owing to the interference of the marine snow noise. In contrast, DGUNet [38], MPRNet [32], and UIRNet [7] demonstrated some success in mitigating the marine snow noise while achieving satisfactory results in color correction. However, when examining the details compared with the ground truth data of UIEBD-Snow, the results processed by the TFCNet network aligned more closely with the reference ground truth, where it outperformed other comparative methods overall. Furthermore, the experimental results on the UIEBD-Snow dataset indicate that the methods that exclusively target underwater optical image enhancement were insufficient for restoring the complex underwater scenes.

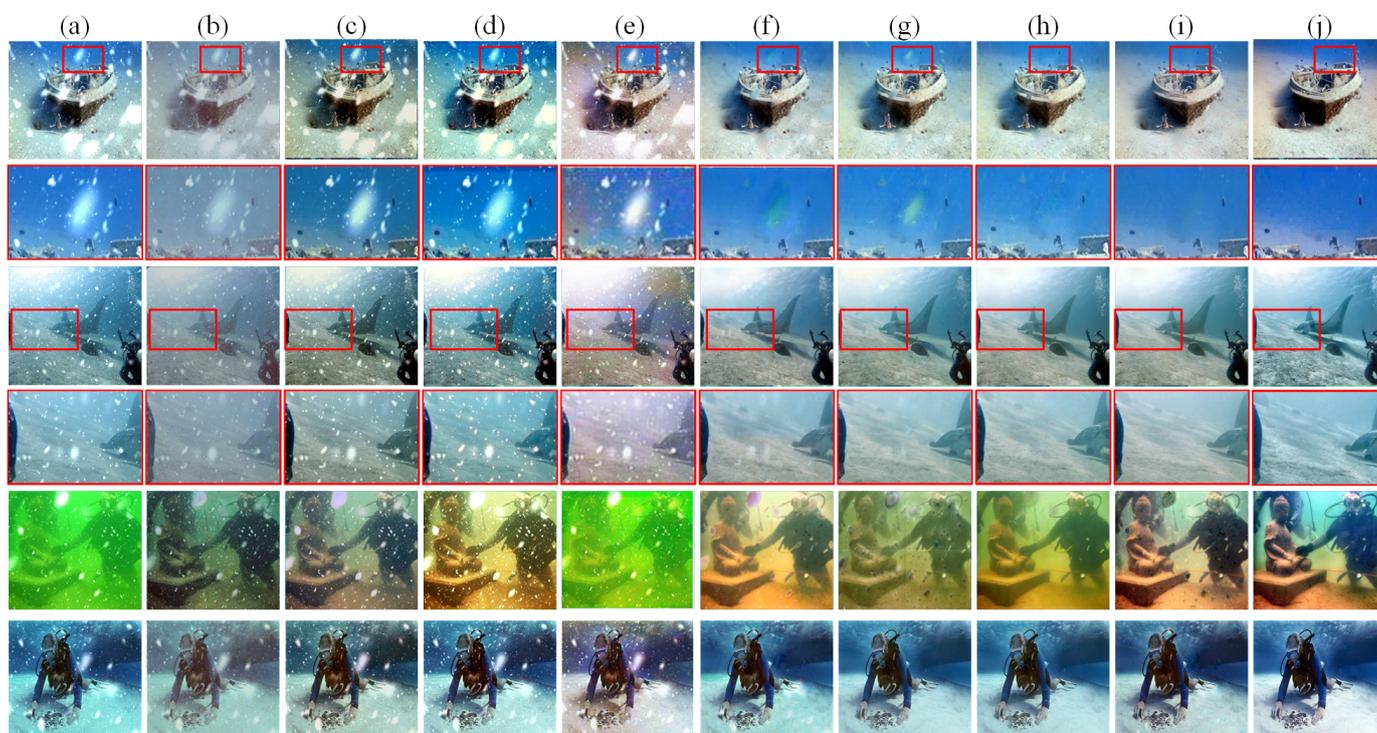


Figure 8. Comparison of TFCNet and the contrastive methods on the UIEBD-Snow dataset: (a) raw; (b) Deepwave [36]; (c) PUIENet [17]; (d) Shallow [37]; (e) FUnIE-GAN [22]; (f) DGUNet [38]; (g) MPRNet [32]; (h) UIRNet [7]; (i) TFCNet; (j) GT.

Figure 9 illustrates that Deepwave [36], PUIENet [17], and Shallow [37] demonstrated strong color correction capabilities while training the UIEBD-Blur dataset. However, these

methods are limited when removing blur. MPRNet [32], DGUNet [38], and UIRNet [7] improved the color correction and blur removal. Nevertheless, when considering the overall results, the output processed by the TFCNet network aligned more closely with the reference ground truth data provided by the UIEBD-Blur dataset, where it demonstrated superior performance.

A comprehensive comparative analysis of Figures 8 and 9 reveals that TFCNet, leveraging its hybrid architecture, effectively enhanced and restored the complex underwater optical images on the UIEBD-Snow and UIEBD-Blur datasets. This included fundamental color correction, more challenging marine snow noise elimination, and blur restoration tasks.

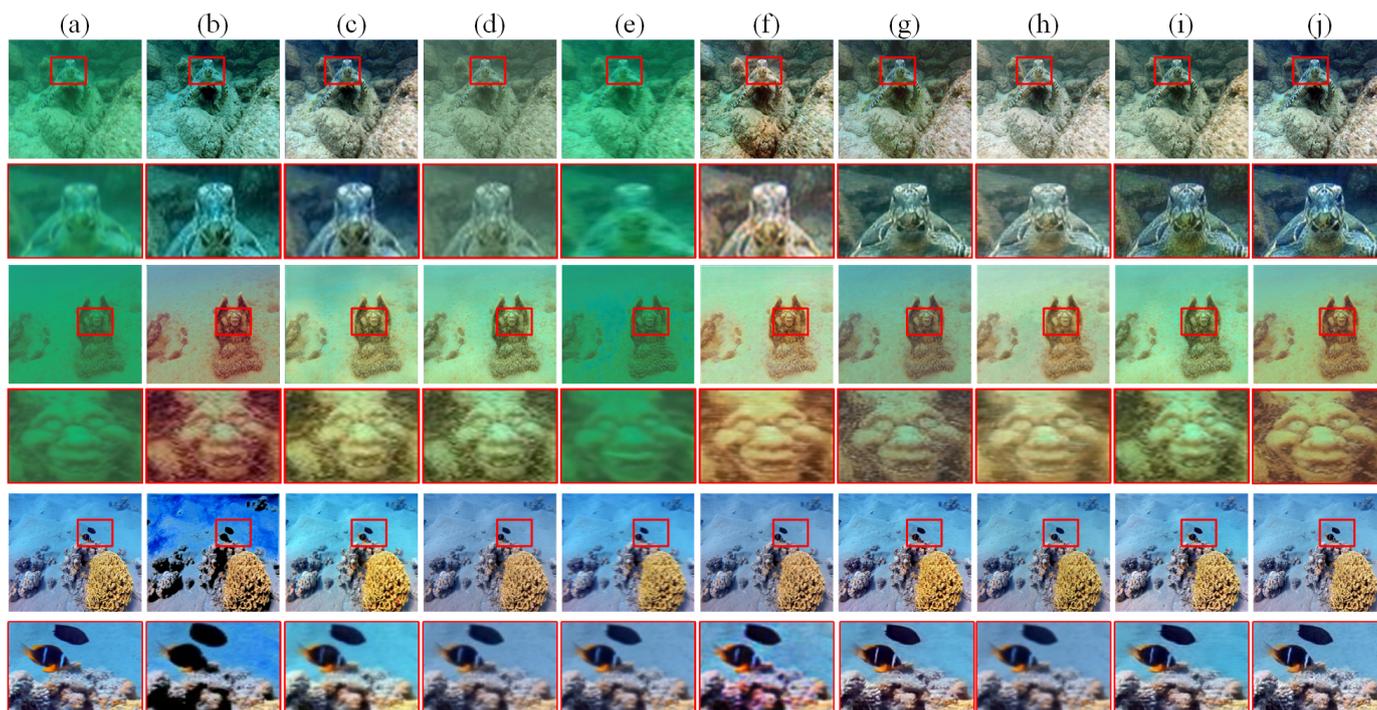


Figure 9. Comparison of TFCNet and the contrastive methods using the UIEBD-Blur dataset: (a) raw; (b) Deepwave [36]; (c) PUIENet [17]; (d) Shallow [37]; (e) FUnIE-GAN [22]; (f) DGUNet [38]; (g) MPRNet [32]; (h) UIRNet [7]; (i) TFCNet; (j) GT.

4.3.2. Quantitative Evaluation

We used standard metrics, such as PSNR, SSIM, and RMSE, to validate the superiority of our approach for full-reference evaluation [39]. The PSNR quantifies the pixel-level fidelity by measuring the logarithmic ratio between the maximum signal power and noise distortion, making it sensitive to the absolute error magnitude. The SSIM evaluates the perceptual quality through luminance, contrast, and structure comparisons, emphasizing local pattern preservation and visual perception. The RMSE provides a direct, interpretable measure of the average pixel-wise deviation, which is particularly useful for physical accuracy validation in scientific applications. Additionally, we used the UCIQE [40] and UIQM for no-reference evaluation, which are commonly used for assessing underwater image quality. The UIQM combines colorfulness, sharpness, and contrast measures to predict human visual preferences, while the UCIQE focuses on color distribution properties.

In the experimental evaluation using the UIEBD-Snow dataset in Table 1, the highest score is highlighted in red, and the second-highest score is highlighted in blue. The results processed using the TFCNet network had the best values in the PSNR, SSIM, and RMSE metrics, where it outperformed other methods. TFCNet did not attain the highest score or the second-highest score in the UIQM and UICQE metrics; nonetheless, its

performance remained close to the highest score. The four methods, namely, Deepwave [36], PUIENet [17], Shallow [37], and FUnIE-GAN [22], demonstrated distinct advantages when evaluated using the UIQM and UICQE metrics. However, when reference images are available, the UIQM and UICQE prioritize perceptual enhancements over physical fidelity, often assigning higher scores to artificially processed images despite significant deviations from the ground truth, where handcrafted features fail to capture the structural distortions measurable by the SSIM or RMSE. In addition, combined with a qualitative analysis, these methods are limited to underwater optical image enhancement and are ineffective in marine snow noise elimination tasks, making them unsuitable for underwater optical image multi-task restoration. In contrast, the TFCNet network excels in these scenarios, producing cleaner, more natural results with fine-grained textures, as shown in Figure 8.

Table 1. Quantitative comparison of TFCNet and the contrastive methods on the UIEBD-Snow dataset.

Method	PSNR↑	SSIM↑	RMSE↓	UIQM↑	UICQE↑
Deepwave [36]	15.374	0.556	23.276	4.620	0.507
PUIENet [17]	16.926	0.723	18.030	4.594	0.606
Shallow [37]	17.006	0.707	18.904	4.354	0.633
FUnIE-GAN [22]	15.619	0.486	24.762	5.106	0.596
DGUNet [38]	20.027	0.775	15.048	3.400	0.598
MPRNet [32]	20.711	0.795	14.068	3.363	0.597
UIRNet [7]	21.200	0.807	13.142	3.610	0.596
Ours	21.527	0.811	12.321	4.399	0.599

In evaluating the experimental results of the UIEBD-Blur dataset, as in Table 2, the highest score is highlighted in red, and the second-highest score is highlighted in blue. TFCNet demonstrated exceptional performance, surpassing other methods in the PSNR, SSIM, and RMSE. It achieved the second-highest score in the UIQM metric. TFCNet did not achieve the highest or second-highest score in the UICQE metric compared with Deepwave [36] and PUIENet [17]. Nonetheless, its results were close to the highest score. Combined with the performance Figure 9 illustrates, TFCNet performed the best, where it effectively restored underwater optical images with motion blur.

Table 2. Quantitative comparison of TFCNet and the contrastive methods using the UIEBD-Blur dataset.

Method	PSNR↑	SSIM↑	RMSE↓	UIQM↑	UICQE↑
Deepwave [36]	18.172	0.574	20.236	3.336	0.596
PUIENet [17]	17.668	0.585	21.130	3.218	0.594
Shallow [37]	17.693	0.572	20.831	3.108	0.581
FUnIE-GAN [22]	17.772	0.564	21.230	2.667	0.550
DGUNet [38]	20.069	0.698	15.297	2.891	0.562
MPRNet [32]	20.175	0.738	15.184	2.883	0.582
UIRNet [7]	21.509	0.775	13.628	3.054	0.584
Ours	21.810	0.787	12.755	3.239	0.589

Qualitative and quantitative analyses were conducted on the experimental results using the UIEBD-Snow and UIEBD-Blur datasets. Based on the comprehensive analysis of Figures 8 and 9 and Tables 1 and 2, our method considerably color-corrected and eliminated impurity and motion blur for underwater images, demonstrating practical relevance.

4.4. Ablation Study

Ablation experiments were conducted by designing four comparative models with approximately the same number of parameters to validate the effectiveness of the Swin Transformer-based encoder module and the activation-free CNN-based decoder module. These models include a Vision Transformer-based encoder combined with a CNN-based encoder, a pure CNN-based encoder–decoder structure, an original Swin Transformer-based encoder–decoder structure, and the proposed TFCNet network proposed in this paper.

A visual study was performed on the UIEBD-Snow dataset for a qualitative analysis. Figure 10 reveals that the model combining a Vision Transformer-based encoder with a CNN-based encoder did not effectively learn the mapping relationships, where it exhibited shallow learning capabilities during training and testing and performed poorly on challenging samples. The pure CNN-based encoder–decoder structure struggled to model the mapping between color and target images, which left considerable room for improvement in detail restoration. Combining the original Swin Transformer-based encoder with the CNN-based decoder improved the enhancement and restoration of underwater optical images to some extent; however, its overall performance remained measurably inferior to TFCNet.

As in Table 3, TFCNet demonstrated substantial performance improvements over other combinations, with increases of approximately 0.5 and 0.01 in the PSNR and SSIM, respectively. The results obtained using TFCNet’s hybrid architecture, combined with the visual performance in Figure 10, were more aligned with human visual perception and restored the complex underwater optical images better.

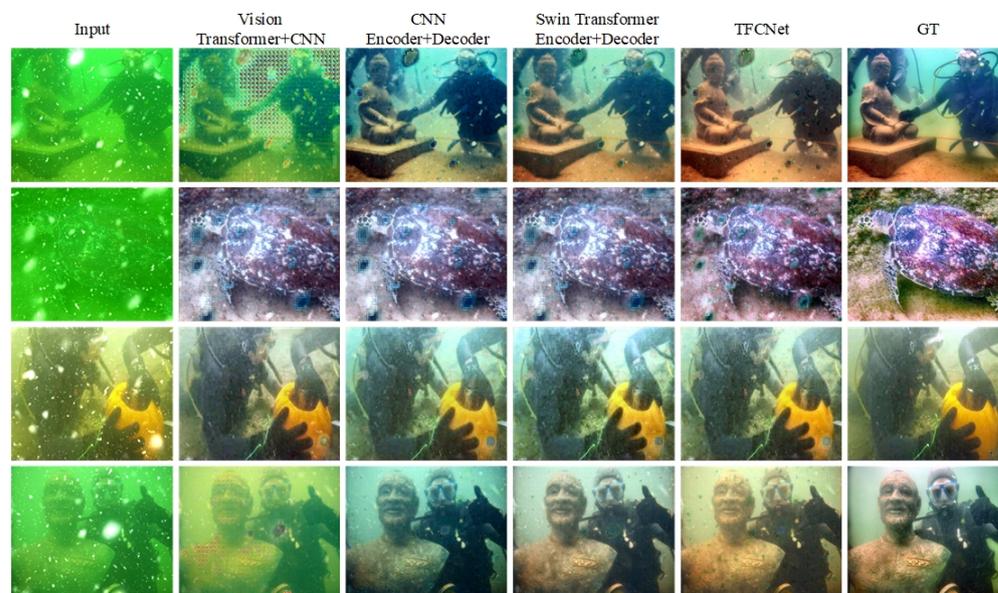


Figure 10. Ablation study of TFCNet hybrid-encoding mode.

To validate the design efficacy of the proposed CNN decoder in TFCNet, here we performed ablation experiments by substituting the decoder module with two alternatives, MPRNet [32] and UIRNet [7]. As demonstrated in Figure 11 and Table 4, the proposed TFCNet exhibited superior performance in both the qualitative and quantitative assessments.

Table 3. Quantitative results of the ablation study in TFCNet hybrid-encoding mode.

Parameter	Vision Transformer + CNN	CNN Encoder + Decoder	Swin Transformer Encoder + Decoder	TFCNet
PSNR	19.383	21.000	21.035	21.527
SSIM	0.742	0.792	0.801	0.810

Table 4. Quantitative results of the ablation study of CNN performance in hybrid-encoding mode.

Parameter	Swin Transformer + CNN(MPRNet)	Swin Transformer + CNN(UIRNet)	TFCNet
PSNR	21.037	21.091	21.525
SSIM	0.796	0.798	0.808

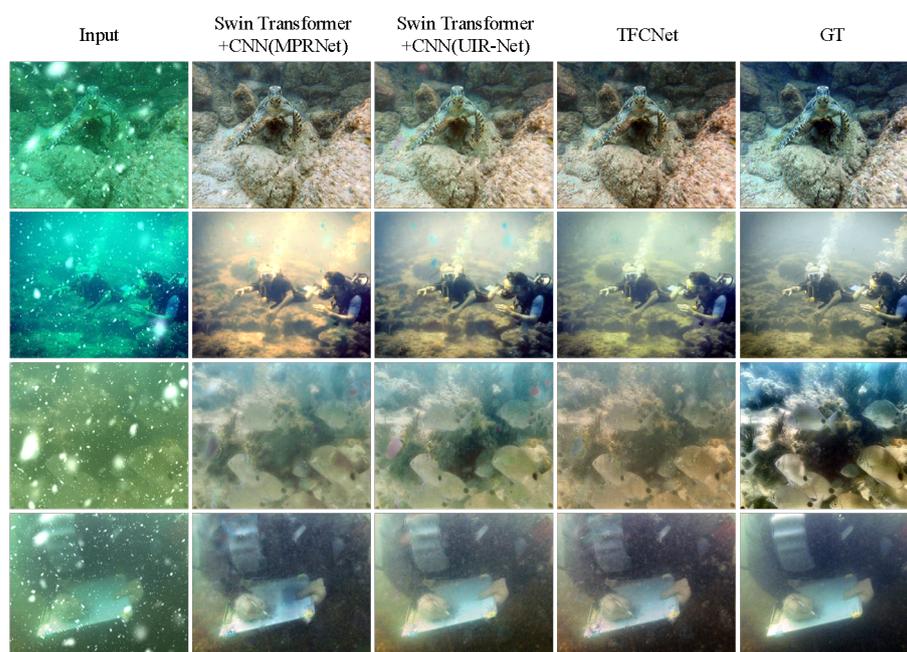


Figure 11. Ablation study of CNN performance in hybrid-encoding mode.

4.5. Application Testing

This section involves application tests using the training results from the TFCNet network on the UIEBD-Snow and UIEBD-Blur datasets to validate the effectiveness of TFCNet.

Initially, we focused on assessing a subset of images from the MSRB dataset [33], which is used for marine snow noise elimination in underwater optical imagery, along with its extended version, the MSIRB dataset [7]. Figure 12 illustrates the evaluation results, which involve real underwater optical images affected by marine snow noise.

Figure 12 reveals that the TFCNet method accomplished color correction in underwater optical images while also eliminating marine snow noise to a degree in the MSRB and MSIRB datasets. However, a closer examination revealed that the evaluation results of the TFCNet method on the MSRB dataset were less satisfactory than those obtained using the MSIRB dataset. This discrepancy can be attributed to the limitations in the datasets. When the morphology of marine snow noise does not align with the marine snow models in the dataset, the noise elimination is constrained, which is an issue that necessitates further in-depth research in future studies.

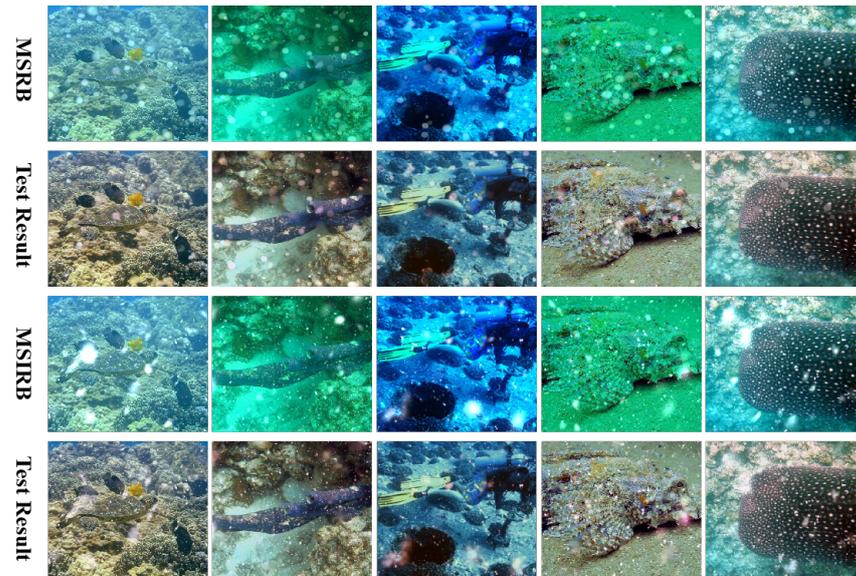


Figure 12. Assessing the performance of TFCNet on the MSRB and MSIRB datasets.

In addition, this section selected some complex underwater optical images that exhibited genuine marine snow noise and blurring phenomena for testing, with the results depicted in Figure 13. As illustrated, TFCNet performed color correction on underwater optical images (images (a) and (b)) while simultaneously mitigating the marine noise to some extent. However, the noise elimination effect was suboptimal for the densely concentrated dynamic turbidity presented in image (b). Furthermore, for images (c) and (d), TFCNet achieved color correction in the underwater optical imagery while partially addressing the blurring effects. Nevertheless, it remained limited in restoring severe motion blur, as exemplified in image (d).

In summary, TFCNet demonstrated commendable performance on the UIEBD-Snow and UIEBD-Blur datasets. Nonetheless, the processing outcomes for real-world images revealed certain inadequacies. This observation underscores the inherent limitations of the UIEBD-Snow and UIEBD-Blur datasets discussed in this section, indicating a need for further refinement and enhancement in subsequent research.

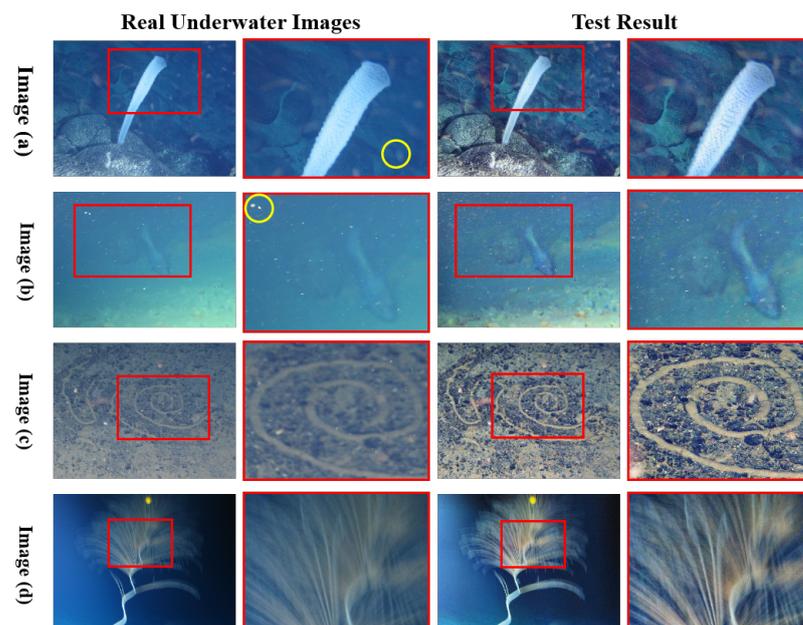


Figure 13. Evaluation of TFCNet’s performance on real complex underwater optical images.

5. Conclusions

This paper proposes TFCNet to address the limitations of existing image enhancement and restoration methods in handling the multi-task restoration of complex underwater optical images. The advantages of Transformers and the importance of hybrid architectures are discussed. The lack of relevant datasets was tackled by constructing the UIEBD-Blur dataset for blur restoration, extending the publicly available UIEB dataset designed for underwater optical image enhancement and restoration tasks. TFCNet leverages the strengths of Transformer and CNN architectures by integrating the global-information-focused Transformer structure with the CNN, which extracts low-level image features. This integration achieved the multi-task enhancement and restoration of complex underwater optical images. The experimental results demonstrated the superior performance of TFCNet on the UIEBD-Snow and UIEBD-Blur datasets, where it performed effective color correction while eliminating marine snow noise and addressing blur issues. Compared with the baseline methods, the proposed approach demonstrated consistent improvements, where it achieved minimum gains of 0.3 dB in the PSNR and 0.01 in the SSIM and a 0.8 reduction in the RMSE. This demonstrates its effectiveness in addressing the multi-task restoration challenges of complex underwater optical imaging to a considerable extent. Future work will focus on addressing the limitations identified in Section 4.5, particularly through methods such as domain adaptation, data augmentation, or transfer learning to improve the restoration effect of complex underwater images.

Author Contributions: Conceptualization, software, and methodology, S.Z. and X.M.; data curation, S.Z., H.Q. and X.M.; writing—original draft preparation, S.Z. and X.M.; writing—review and editing, X.Y. and S.G.; visualization, S.Z., H.Q. and X.M.; supervision, project administration, and funding acquisition, X.Y. All authors have read and agreed to the published version of this manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 42276187).

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: Xinkui Mei was employed by the Harbin Electric Science and Technology Company Limited. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Wang, S.; Li, W.; Xing, L. A Review on Marine Economics and Management: How to Exploit the Ocean Well. *Water* **2022**, *14*, 2626. [[CrossRef](#)]
2. Zhang, B.; Ji, D.; Liu, S.; Zhu, X.; Xu, W. Autonomous underwater vehicle navigation: A review. *Ocean Eng.* **2023**, *273*, 113861. [[CrossRef](#)]
3. Zhou, J.; Li, B.; Zhang, D.; Yuan, J.; Zhang, W.; Cai, Z.; Shi, J. UGIF-Net: An Efficient Fully Guided Information Flow Network for Underwater Image Enhancement. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–17. [[CrossRef](#)]
4. Xia, Q. An Overview of Underwater Vision Enhancement: From Traditional Methods to Recent Deep Learning. *J. Mar. Sci. Eng.* **2022**, *10*, 241.
5. Wei, X.; Ye, X.; Mei, X.; Wang, J.; Ma, H. Enforcing high frequency enhancement in deep networks for simultaneous depth estimation and dehazing. *Appl. Soft Comput.* **2024**, *163*, 111873. [[CrossRef](#)]
6. Ayilu, R.K.; Fabinyi, M.; Barclay, K. Small-scale fisheries in the blue economy: Review of scholarly papers and multilateral documents. *Ocean Coast. Manag.* **2022**, *216*, 105982. [[CrossRef](#)]
7. Mei, X.; Ye, X.; Zhang, X.; Liu, Y.; Wang, J.; Hou, J.; Wang, X. Uir-net: A simple and effective baseline for underwater image restoration and enhancement. *Remote Sens.* **2022**, *15*, 39. [[CrossRef](#)]
8. Zhou, J.; Yang, T.; Zhang, W. Underwater vision enhancement technologies: A comprehensive review, challenges, and recent trends. *Appl. Intell.* **2023**, *53*, 3594–3621. [[CrossRef](#)]

9. Peng, L.; Zhu, C.; Bian, L. U-shape transformer for underwater image enhancement. *IEEE Trans. Image Process.* **2023**, *32*, 3066–3079. [[CrossRef](#)] [[PubMed](#)]
10. Zhang, Y.; Chandler, D.M.; Leszczuk, M. Retinex-based underwater image enhancement via adaptive color correction and hierarchical U-shape transformer. *Opt. Express* **2024**, *32*, 24018–24040. [[CrossRef](#)]
11. Iqbal, K.; Odetayo, M.; James, A.; Salam, R.A.; Talib, A.Z.H. Enhancing the low quality images using unsupervised colour correction method. In Proceedings of the 2010 IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10–13 October 2010; pp. 1703–1709.
12. Hitam, M.S.; Awalludin, E.A.; Yussof, W.N.J.H.W.; Bachok, Z. Mixture contrast limited adaptive histogram equalization for underwater image enhancement. In Proceedings of the 2013 International Conference on Computer Applications Technology (ICCAT), Sousse, Tunisia, 20–22 January 2013; pp. 1–5.
13. Fu, X.; Zhuang, P.; Huang, Y.; Liao, Y.; Zhang, X.P.; Ding, X. A retinex-based enhancing approach for single underwater image. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 4572–4576.
14. Drews, P.L.; Nascimento, E.R.; Botelho, S.S.; Campos, M.F.M. Underwater depth estimation and image restoration based on single images. *IEEE Comput. Graph. Appl.* **2016**, *36*, 24–35. [[CrossRef](#)] [[PubMed](#)]
15. Peng, Y.T.; Cao, K.; Cosman, P.C. Generalization of the dark channel prior for single image restoration. *IEEE Trans. Image Process.* **2018**, *27*, 2856–2868. [[CrossRef](#)] [[PubMed](#)]
16. Mei, X.; Ye, X.; Wang, J.; Wang, X.; Huang, H.; Liu, Y.; Jia, Y.; Zhao, S. UIEOGP: An underwater image enhancement method based on optical geometric properties. *Opt. Express* **2023**, *31*, 36638–36655. [[CrossRef](#)]
17. Fu, Z.; Wang, W.; Huang, Y.; Ding, X.; Ma, K.K. Uncertainty inspired underwater image enhancement. In *Proceedings of the European Conference on Computer Vision*; Springer: Cham, Switzerland, 2022; pp. 465–482.
18. Li, J.; Skinner, K.A.; Eustice, R.M.; Johnson-Roberson, M. WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robot. Autom. Lett.* **2017**, *3*, 387–394. [[CrossRef](#)]
19. Wang, J.; Ye, X.; Liu, Y.; Mei, X.; Hou, J. Underwater self-supervised monocular depth estimation and its application in image enhancement. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105846. [[CrossRef](#)]
20. Fabbri, C.; Islam, M.J.; Sattar, J. Enhancing underwater imagery using generative adversarial networks. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 21–25 May 2018; pp. 7159–7165.
21. Han, J.; Shoeiby, M.; Malthus, T.; Botha, E.; Anstee, J.; Anwar, S.; Wei, R.; Petersson, L.; Armin, M.A. Single underwater image restoration by contrastive learning. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 2385–2388.
22. Islam, M.J.; Xia, Y.; Sattar, J. Fast underwater image enhancement for improved visual perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [[CrossRef](#)]
23. Zhang, S.; Zhao, S.; An, D.; Liu, J.; Wang, H.; Feng, Y.; Li, D.; Zhao, R. Visual SLAM for underwater vehicles: A survey. *Comput. Sci. Rev.* **2022**, *46*, 100510. [[CrossRef](#)]
24. Jiang, Q.; Chen, Y.; Wang, G.; Ji, T. A novel deep neural network for noise removal from underwater image. *Signal Process. Image Commun.* **2020**, *87*, 115921. [[CrossRef](#)]
25. Sun, K.; Meng, F.; Tian, Y. Underwater image enhancement based on noise residual and color correction aggregation network. *Digit. Signal Process.* **2022**, *129*, 103684. [[CrossRef](#)]
26. Sun, K.; Meng, F.; Tian, Y. Progressive multi-branch embedding fusion network for underwater image enhancement. *J. Vis. Commun. Image Represent.* **2022**, *87*, 103587. [[CrossRef](#)]
27. Farhadifard, F.; Radolko, M.; von Lukas, U.F. Single Image Marine Snow Removal based on a Supervised Median Filtering Scheme. In Proceedings of the VISIGRAPP (4: VISAPP), 2017: 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Porto, Portugal, 27 February–1 March 2017; pp. 280–287.
28. Fu, X.; Huang, J.; Zeng, D.; Huang, Y.; Ding, X.; Paisley, J. Removing rain from single images via a deep detail network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3855–3863.
29. Ren, D.; Zuo, W.; Hu, Q.; Zhu, P.; Meng, D. Progressive image deraining networks: A better and simpler baseline. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3937–3946.
30. Tu, Z.; Talebi, H.; Zhang, H.; Yang, F.; Milanfar, P.; Bovik, A.; Li, Y. Maxim: Multi-axis mlp for image processing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5769–5780.
31. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5728–5739.
32. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Multi-stage progressive image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 14821–14831.

33. Kaneko, R.; Sato, Y.; Ueda, T.; Higashi, H.; Tanaka, Y. Marine snow removal benchmarking dataset. In Proceedings of the 2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Taipei, Taiwan, 31 October–3 November 2023; pp. 771–778.
34. Chen, L.; Chu, X.; Zhang, X.; Sun, J. Simple baselines for image restoration. In *Proceedings of the European Conference on Computer Vision*; Springer: Cham, Switzerland, 2022; pp. 17–33.
35. 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](#)] [[PubMed](#)]
36. Sharma, P.; Bisht, I.; Sur, A. Wavelength-based attributed deep neural network for underwater image restoration. *ACM Trans. Multimed. Comput. Commun. Appl.* **2023**, *19*, 1–23. [[CrossRef](#)]
37. Naik, A.; Swarnakar, A.; Mittal, K. Shallow-UWnet: Compressed Model for Underwater Image Enhancement. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 15853–15854.
38. Mou, C.; Wang, Q.; Zhang, J. Deep generalized unfolding networks for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 17399–17410.
39. 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](#)]
40. Yang, M.; Sowmya, A. An underwater color image quality evaluation metric. *IEEE Trans. Image Process.* **2015**, *24*, 6062–6071. [[CrossRef](#)]

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.