Journal of *Marine Science and Engineering*

MDPI

Article

Multi-Level Wavelet-Based Network Embedded with Edge Enhancement Information for Underwater Image Enhancement

Kaichuan Sun [1,2,†], Fei Meng [3,†] and Yubo Tian [3,*]

[1] School of Ocean, Jiangsu University of Science and Technology, Zhenjiang 212100, China; kcsun991@gmail.com
[2] School of Computer and Information Engineering, Chuzhou University, Chuzhou 239000, China
[3] School of Information and Communication Engineering, Guangzhou Maritime University, Guangzhou 510725, China; mengfei@gzmtu.edu.cn
* Correspondence: tianyubo@just.edu.cn
† These authors contributed equally to this work.

Abstract: As an image processing method, underwater image enhancement (UIE) plays an important role in the field of underwater resource detection and engineering research. Currently, the convolutional neural network (CNN)- and Transformer-based methods are the mainstream methods for UIE. However, CNNs usually use pooling to expand the receptive field, which may lead to information loss that is not conducive to feature extraction and analysis. At the same time, edge blurring can easily occur in enhanced images obtained by the existing methods. To address this issue, this paper proposes a framework that combines CNN and Transformer, employs the wavelet transform and inverse wavelet transform for encoding and decoding, and progressively embeds the edge information on the raw image in the encoding process. Specifically, first, features of the raw image and its edge detection image are extracted step by step using the convolution module and the residual dense attention module, respectively, to obtain mixed feature maps of different resolutions. Next, the residual structure Swin Transformer group is used to extract global features. Then, the resulting feature map and the encoder's hybrid feature map are used for high-resolution feature map reconstruction by the decoder. The experimental results show that the proposed method can achieve an excellent effect in edge information protection and visual reconstruction of images. In addition, the effectiveness of each component of the proposed model is verified by ablation experiments.

Keywords: underwater image enhancement; wavelet transform; edge detection; Transformer

Citation: Sun, K.; Meng, F.; Tian, Y. Multi-Level Wavelet-Based Network Embedded with Edge Enhancement Information for Underwater Image Enhancement. *J. Mar. Sci. Eng.* **2022**, *10*, 884. https://doi.org/10.3390/jmse10070884

Academic Editors: Anna Nora Tassetti, Adriano Mancini and Pierluigi Penna

Received: 30 May 2022
Accepted: 22 June 2022
Published: 27 June 2022

1. Introduction

Underwater images play an essential role in the field of underwater resource detection and underwater engineering research [1,2]. However, images acquired underwater are often degraded by light absorption and scattering, such as blur, color cast, and marine snow. This has a severe impact on both underwater detection and engineering research. Traditional methods for underwater image enhancement (UIE) mostly use manual feature modeling methods to enhance visuals, such as prior knowledge-based methods [3,4], wavelet transform-based methods [5,6], and retinex-based method [7]. However, these methods are effective only in specific scenarios and have limited robustness in complex scenarios. With the rapid development of deep learning, convolutional neural network (CNN)- and Transformer-based methods have been widely used in the field of image processing [8–14].

At present, the CNN- and Transformer-based methods are the mainstream methods in the field of computer vision. Although the CNN-based methods have been shown to be able to capture local features between contexts, they cannot effectively model long-range dependencies due to the local property of convolutional kernels. In contrast, Transformer-based methods can capture global interactions between contexts and have shown promising

performance in long-range dependency modeling, but they perform poorly in capturing local contexts. Nevertheless, both local and global features are crucial for computer vision tasks, especially for image enhancement. Different from the existing CNN and Transformer combination methods, we add the Swin Transformer [15] group with residual structure in the transformation part of encoding and decoding, which can better integrate local and global information and improve the visual effect of enhanced images.

Due to its sparse representation ability for images and good reconstruction and time-frequency localization properties, wavelet transform has been widely used in image restoration and enhancement tasks [5,6]. In deep-learning-based methods, CNNs usually use pooling to expand the receptive field, but such an approach may lead to the loss of important information, such as the color or edge of an image, which is not conducive to feature extraction and analysis. To solve this problem, Liu et al. [16] proposed embedding wavelet transform into the CNN architecture to reduce the resolution of feature maps while increasing the receptive field. After that, many studies on computer vision have embedded wavelet transform into their network to recover detailed information from the raw image, and good results have been achieved [17–19], which has been an inspiration for the method proposed in this study.

Although the existing deep-learning-based methods can achieve excellent performance on UIE, they often produce blurred edges in enhanced images. Recently, edge information has been used in deep-learning-based models and applied to a wide range of computer vision tasks [20–22]. However, although edge information has the potential to improve the performance of UIE, it has not been commonly adopted by the CNN- and Transformer-based methods. To reinforce the ability of the UIE network to recover edge information, first, the edge information of the raw image should be obtained by the Sobel edge detector; then, both the raw image and the edge detection image should be used as the input of a deep-learning-based network to strengthen the learning effect of edge information in the encoding and decoding processes.

In this work, to obtain high-quality underwater visual images, a UIE method that uses CNN and Transformer as the main framework and combines the wavelet transform and edge detection, called WE-Net, is proposed. Specifically, first, the Sobel edge detection algorithm is applied to the raw image to obtain the corresponding edge image, and then these two images are used as input and sent to the hybrid encoder. In the hybrid encoder, a feature extraction module consisting of a convolutional block and a residual dense attention module (RDAM) is designed, and a discrete wavelet transform is employed to reduce the resolution of the feature map progressively. The RDAM is composed of a dense residual module and attention blocks. The effectiveness of dense residual blocks has been verified in [23,24], and it has been shown that they have a positive effect on color correction. In the attention blocks, we use pixel attention, channel attention, and spatial attention combined in a parallel manner, which is encouraged to focus on the effective information of the image. Next, the encoded feature maps are fed to the residual-structured Swin Transformer group to learn the global features of the context. Finally, the obtained feature maps and the mixed feature map obtained by the encoding process are sent to the decoder for resolution reconstruction. To be in accordance with the encoder, the decoder adopted in this study consists of deconvolutional and residual dense attention modules.

The main contributions of our work are as follows:

- An improved UIE model that combines CNN and Transformer is proposed, and discrete wavelet transform and edge detection are added to the network to improve its feature representation and edge enhancement performances;
- A dense residual attention module, which consists of a dense residual block and three attention modules, is designed and embedded into the encoder-decoder network for feature encoding and decoding;
- The effectiveness of the proposed WE-Net is verified by comparison with the existing methods on the full- and non-reference datasets for UIE. In addition, the robustness

of the proposed model is demonstrated by ablation studies, and quantitative and qualitative tests.

2. Related Work

As a branch of image processing, UIE is mainly intended to enhance the visual effect of degraded images. With the outstanding performance of deep learning, the focus of researchers has shifted from traditional hand-crafted feature-based modeling methods to data-driven deep-learning-based methods. In this section, some of the common methods related to the proposed method are introduced.

CNN-Transformer-Based Methods: CNN (encoder-decoder architecture) and Transformer, as essential deep learning models, have achieved encouraging results in the computer vision field. Recently, many studies have effectively integrated these two models, using their advantages to compensate for their shortcomings [25–28]. Li et al. [25] combined a Transformer with a 3D CNN to effectively model local and global features for medical image segmentation, while introducing a deformable bottleneck module to capture more shape-aware feature representations. Song et al. [26] proposed a hybrid network similar to the U-Net and combined the Transformer and CNN to extract global and local information for medical image registration. Gao et al. [27] proposed an efficient CNN-Transformer cooperation network for face super-resolution tasks, using the multi-scale connected encoder-decoder architecture as the backbone. Chen et al. [28] adopted a hybrid CNN-Transformer structure to exploit high-resolution spatial information from CNN features and Transformer-encoded global context. The aforementioned works use the advantages of the CNN and Transformer to capture the local and global information of the context well and obtain satisfactory results, which has also provided ideas for this work.

Wavelet-Based Methods: The wavelet transform decomposes the input signal into different frequency components and represents a powerful tool for image processing and time-frequency representation. Traditional wavelet transform-based image denoising and image restoration have been widely studied since before the advent of deep learning [5,6,29,30]. For instance, Zhou et al. [31] performed wavelet decomposition on color-corrected and contrast-stretched underwater images, and fused the decomposed components in equal proportions to improve the color recovery effect. However, traditional methods use hand-crafted features and cannot remove uneven noise information well. Wavelet transform has been embedded into deep learning models to perform various vision tasks, such as image inpainting [32], medical image super-resolution [33], image deraining [19], image denoising [16,18], image ISP [24], and underwater image enhancement [17]. Inspired by the previous work, this study reduces the resolution of feature maps and increases the receptive field using the wavelet transform in encoding, and reconstructs high-resolution feature maps using inverse wavelet transform in decoding.

Edge-Enhancement-Based Methods: The essence of UIE is to improve the clarity of an image, and edge information is one of the critical indicators of image enhancement. Many recent studies have incorporated edge prior knowledge into deep-learning-based models. For instance, Chen et al. [20] designed a novel deblurring model by explicitly modeling edge information as prior knowledge. Kim et al. [22] proposed combining a dense edge detection network and feature merge network to enhance edge information for image super-resolution. Liang et al. [21] proposed a densely connected CNN based on edge enhancement for low-dose CT denoising. Different from the above-mentioned embedding methods of edge information, this study uses the prior edge knowledge and a raw image as input to perform mixed encoding of the network and sends the fused information to the decoder for decoding through skip connections. This ensures that edge information can be paid attention to during both input feature extraction and reconstruction.

3. Proposed Method

In this section, the concept of discrete wavelet transform and Sobel edge detection are briefly introduced, and the main motivation for this work is explained. The proposed

WE-Net based on RDAM and its network architecture for UIE is presented. Finally, a loss function is introduced.

3.1. Proposed Architecture Background

Discrete Wavelet Transform: Wavelet transform has been an important method in image processing. It can decompose the image into independent sub-bands containing low- and high-frequency information; these sub-bands can provide important information for subsequent feature extraction and analysis. In this work, the Haar wavelet transform is used, where four filter kernels are used for image decomposition: $f_{LL} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $f_{LH} = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}$, $f_{HL} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}$, $f_{HH} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$, where indexes L and H represent low and high frequencies, respectively. The low-pass filter f_{LL} captures smooth surfaces and textures, while the other three high-pass filters extract vertical, horizontal, and diagonal edge information. Given an input image I, wavelet transform can be performed through convolution and down-sampling operations, which can be expressed by:

$$DWT_i = (I \otimes f_i) \downarrow 2, \quad i \in \{LL, LH, HL, HH\} \tag{1}$$

where $\otimes$ represents the convolution operation, and $\downarrow 2$ represents the standard down-sampling operator with a factor of two. According to Equation (1), the DWT can be deemed a convolution process on an input image I, which uses four 2×2 convolution kernels with fixed weights and a stride of two. Therefore, through the DWT process, four decomposed sub-bands $\{DWT_{LL}, DWT_{LH}, DWT_{HL}, DWT_{HH}\}$, are obtained, and each sub-band is half the size of I.

According to the orthogonal property of four filters, we can reconstruct the four sub-bands to the target image through IDWT without information loss. It is precisely because of the information-lossless properties of DWT and IDWT that they are widely used in CNNs for image processing. Inspired by [16,24], the DWT and IDWT can be employed to replace down-sampling and up-sampling operations in an encoder-decoder network. As shown in Figure 1, this study uses the DWT in the encoding module to reduce the feature map resolution while increasing the receptive field and the IDWT in the decoding module to reconstruct feature maps.

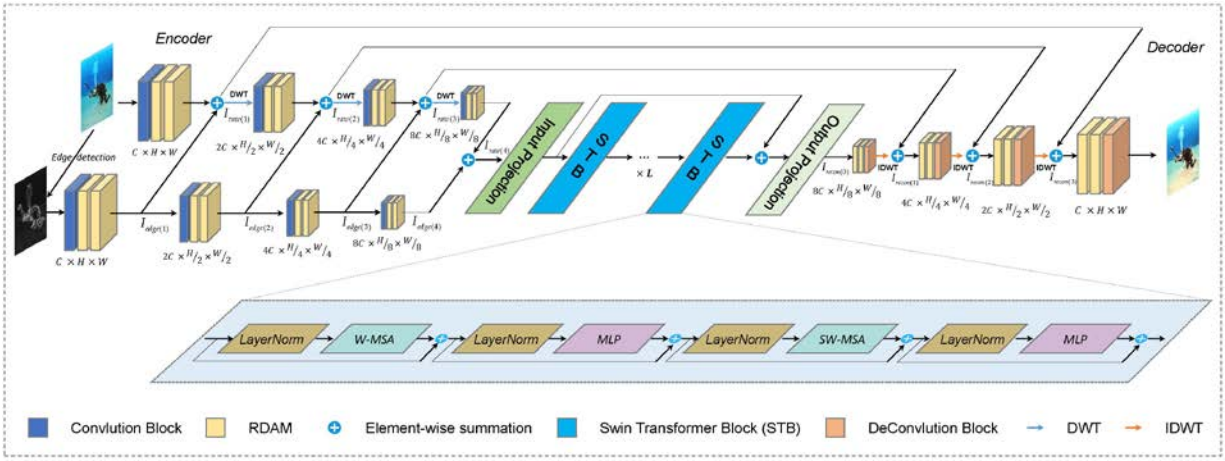

Figure 1. The overall architecture of the proposed WE-Net.

Edge Detection: The edge refers to the collection of pixels whose grayscale changes sharply in the image, and this is the most basic image feature. Edge detection has been a fundamental problem in the field of image processing and computer vision, and its purpose is to identify points in an image with significant changes in brightness. In deep-learning-based methods, the nonlinear mapping between the raw image and the corresponding

degraded image is usually learned in an end-to-end manner without considering that the image edge information can be easily lost during the learning process, which can make the enhanced image edge view unclear. Therefore, edge detection is performed on the raw image during the encoding process of the network, and the obtained image is used as an input of the encoder together with the raw image. In this study, the Sobel edge detection algorithm is used to obtain the edge images.

The Sobel operator represents a combination of Gaussian smoothing and differential operations. It has a strong anti-noise ability, and the resulting edges are smooth and continuous, so it has been widely used in edge-detection tasks. Given an input image I, the Sobel operator contains two 3×3 convolution filters, that is, $S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$, $S_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$. Then, the two convolution filters are convolved with an input image I to obtain the image grayscale values G_x and G_y for horizontal and vertical edge detection, respectively, which is given by:

$$G_i = S_i \otimes I, \; i \in \{x, y\} \tag{2}$$

Next, according to the obtained G_x and G_y, the magnitude of the gradient value G is calculated by:

$$G = \sqrt{G_x^2 + G_y^2} \tag{3}$$

In this work, the Sobel edge detection is implemented by calling the Sobel function from the Kornia library [34].

3.2. Overall Architecture

The structure of the proposed WE-Net is presented in Figure 1. Given an input underwater image $I_{raw} \in R^{C \times H \times W}$ and its edge image $I_{edge} \in R^{C \times H \times W}$ with a spatial resolution of $H \times W$ and C channels, first, feature maps are obtained from the input image and its edge images separately by a convolution block and the proposed RDAM, and the mixed feature maps are fed to the next level via DWT. After that, the Swin Transformer is employed to refine and enhance the mixed-encoded features further. Finally, the IDWT, RDAM, and deconvolutional block are repeatedly applied to gradually produce an enhanced result.

Encoder Stage: As mentioned above, the encoding stage is designed for feature extraction. First, the edge image I_{edge} is obtained by performing the Sobel edge detection method on the raw image I_{raw}. Then, a convolution block and two specially designed RDAMs are applied to each encoding stage to extract the features of the raw image branch and edge image branch, respectively. The convolution block consists of a 3×3 convolutional layer and has a *PReLU* activation function. Next, the feature maps of the raw image branch and the edge image branch are fused, and the feature map resolution is reduced by the DWT. Therefore, after each encoding stage, the size of the output feature maps is halved, while the number of output channels is doubled. Thus, the i-th stage of the encoder produces the feature maps $I_{E(i)} \in R^{iC \times \frac{H}{i} \times \frac{W}{i}}$. The mathematical process of the encoder stage can be expressed as follows:

$$\begin{aligned} I'_{raw(i)} &= RDAM(RDAM(ConvBlo(I_{raw(i-1)}))) \oplus RDAM(RDAM(ConvBlo(I_{edge(i-1)}))) \\ I_{raw(i)} &= DWT(I'_{raw(i)}) \; i \in \{1, 2, 3, 4\} \end{aligned} \tag{4}$$

where $ConvBlo(\cdot)$ represents the $PReLU(Conv(\cdot))$ operation sequence, and $I_{raw(0)}$ and $I_{edge(0)}$ represent the raw image and its edge image, respectively.

Bottleneck Stage: There exists a bottleneck stage between the encoding and decoding stages. To achieve better usage of these features in the decoding stage, the Swin Transformer block (STB) is used to refine and enhance the encoded features further. The Transformer is used mainly because it can compensate for the inability of the encoding stage to efficiently model long-term dependencies and learn global interactions. In the proposed design, the L Swin Transformers are used for concatenation, and skip connections are added at the beginning and end. This can not only effectively transform the encoded features but also retain the detailed features. The output of the s-th ($s \in [1, 2, \ldots, L]$) Transformer layer can be calculated by:

$$\begin{aligned} X'_s &= W - MSA(LN(X_{s-1})) + X_{s-1} \\ X_s &= MLP(LN(X'_s)) + X'_s \\ X'_{s+1} &= SW - MSA(LN(X_s) + X_s \\ X_{s+1} &= MLP(LN(X'_{s+1})) + X'_{s+1} \end{aligned} \tag{5}$$

where X'_s and X_s represent the output features of the window-based multi-head self-attention (*W-MSA*) and *MLP* module of block s, respectively, while $LN(\cdot)$ denotes layer normalization. The specific implementation details of the *W-MSA* and *SW-MSA* can be found in [15].

Decoder stage: To generate full-resolution enhanced results as the raw image space ($3 \times H \times W$), a decoder is introduced to reconstruct feature maps, which consist of the RDAM, deconvolution block, and IDWT. Specifically, the decoder uses the bottleneck layer output and hybrid feature of the encoder as inputs and progressively fuses them through the RDAM and deconvolution block to reconstruct high-quality representations. The deconvolution block consists of a 3×3 deconvolutional layer with a stride of one and the *PReLU* activation function. In accordance with the encoding stage, the IDWT is used for feature map up-sampling in the decoding stage. Therefore, each decoder halves the number of output feature channels while doubling the size of output feature maps, as shown in Figure 1. The mathematical expression of the decoder stage is given by:

$$I_{recon(i)} = IDWT(DeConvBlo(RDAM(RDAM(I_{recon(i-1)})))) \oplus I'_{raw(4-i)} i \in \{1,2,3\} \tag{6}$$

where $DeConvBlo(\cdot)$ represents the $PReLU(DeConv(\cdot))$ operation sequence, and $I_{recon(0)}$ represents the output feature map of the bottleneck layer.

3.3. Residual Dense Attention Module (RDAM)

As one of the most important modules in WE-Net, the RDAM is designed for feature extraction and reconstruction using convolution and deconvolution blocks. The RDAM consists of a residual dense block and a triple attention module (TAM). Residual dense blocks have been widely used in computer vision tasks and have been proven to be effective in feature extraction; learning residual information helps to improve color mapping performance [23,24,35]. In the proposed model, the residual dense block consists of five 3×3 convolutional layers, four *PReLU* activation functions, and four *BatchNorm* layers, as shown in Figure 2. In the convolutional layer, the first four layers aim to increase the number of feature maps, while the last layer concatenates all feature maps generated from the first four convolutional layers, *PReLU* activation function, and *BatchNorm* layers. At the end of the residual dense block, the TAM is introduced to encourage the network to learn the key spatial-, pixel-, and channel-wise information. As shown in Figure 2, the TAM includes channel attention, spatial attention [36], and pixel attention [37], which has been shown to be an effective combination in a parallel manner in our previous work [38]. Given an input feature map $F_{in} \in R^{C \times H \times W}$, the TAM is obtained as follows:

$$\begin{aligned} F_{RDB(1)} &= CRB(F_{in}) + F_{in} \\ F_{RDB(2)} &= CRB(F_{RDB(1)}) + F_{RDB(1)} + F_{in} \\ F_{RDB(i)} &= CRB(F_{RDB(i-1)}) + F_{RDB(i-1)} + \cdots + F_{RDB1} + F_{in} \\ F_{RDB(5)} &= Conv(CRB(F_{RDB(4)}) + F_{RDB(4)} + F_{RDB(3)} + F_{RDB(2)} + F_{RDB(1)} + F_{in}) \end{aligned} \tag{7}$$

$$
\begin{aligned}
F_{CA} &= \sigma(Conv(\mathrm{ReLU}(Conv(GAP(F_{RDB(5)}))))) \otimes F_{RDB(5)} \\
F_{PA} &= \sigma(Conv(\mathrm{ReLU}(Conv(F_{RDB(5)})))) \otimes F_{RDB(5)} \\
F_{SA} &= \sigma(Conv([GAP\left(F_{RDB(5)}\right); GMP(F_{RDB(5)})])) \otimes F_{RDB(5)} \\
F_{out} &= Conv(Conv(Conv(F_{RDB(5)} \oplus F_{CA}) \oplus F_{PA}) \oplus F_{SA}) \oplus F_{in}
\end{aligned}
\tag{8}
$$

where $CRB(\cdot)$ represents the operation sequence $PReLU(BatchNorm(Conv(\cdot)))$; $F_{RDB(i)}$ denotes the i-th layer of the residual dense block; $F_{RDB(5)}$ represents the output of the residual dense block; σ is the sigmoid activation function; GAP and GMP stand for the global average pooling and the global max pooling, respectively; and F_{CA}, F_{PA}, and F_{SA} represent the feature maps obtained based on the output of channel, pixel, and spatial attention sub-modules, respectively.

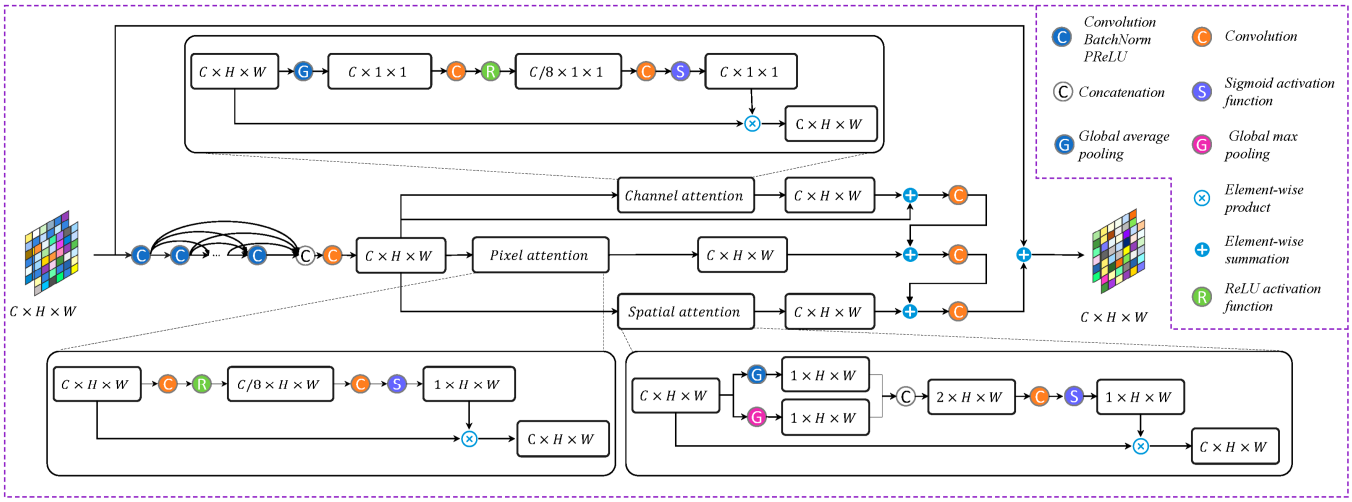

Figure 2. Illustration of the RDAM.

3.4. Loss Function

In this work, the WE-Net parameters are optimized by minimizing the pixel loss L_{l1} as follows:

$$
L_{l1} = \sum_{x=1}^{H} \sum_{y=1}^{W} \left| I_{recon} - I_{ref} \right| \tag{9}
$$

where I_{recon} and I_{ref} represent the reconstructed images and the corresponding reference images, respectively. For the UIE task, this study uses only the naïve L_{l1} pixel loss to demonstrate the effectiveness of the proposed network.

4. Experiments

In this section, first, the implementation details are introduced; then, the proposed method is compared with the state-of-the-art (SOTA) methods both qualitatively and quantitatively on the full- and non-reference datasets, respectively. Ablation experiments are conducted to validate the effectiveness of each component of the proposed WE-Net.

4.1. Implementation Details

Datasets: To train WE-Net, 10,090 pairs of underwater images from the EUVP datasets [39], 1120 pairs of underwater images from the UFO-120 datasets [40], and 790 pairs of underwater images from the UIEB datasets [41] were randomly selected. Therefore, a total of 12,000 pairs of underwater images were used for model training. For testing, the remaining 100 pairs of underwater images from the UIEB dataset, 500 pairs of underwater images from the UFO-120 dataset, and 1345 pairs of underwater images of the EUVP datasets were used. To verify the robustness of the proposed network, comprehensive experiments were conducted on the non-reference datasets (i.e., the Test-C76 and RUIE datasets [42]).

The Test-C76 dataset contained 60 underwater images without reference images provided in the UIEB dataset and 16 representative examples presented on the project page of the SQUID [43]. For the RUIE dataset, its subset UTTS, which contained a total of 300 images, was used as a test set named RUIE-UTTS. For training and testing datasets, the resolution of the images was adjusted to 256 × 256.

Experimental Settings: To implement the proposed network, we use Pytorch as the deep learning framework on an Intel i9-10900X CPU with 32Gb RAM, and a Nvidia GeForce GTX 3090 GPU with 24 Gb of VRAM. To optimize the proposed model, the Adam optimizer was adopted with a momentum of 0.9. The learning rate used a cosine annealing strategy [44], with an initial value of 1e-4. The number of epochs was set to 200, and the batch size was set to eight. The number of STBs was empirically set to eight. In all comparative experiments, the same training and testing datasets were used.

Comparison Methods: WE-Net was compared with eight methods, including two traditional methods (UDCP [3] and ULAP [45]), a residual-network-based method (UResnet [8]), a shallow-network-based method (shallow-UWnet [9]), a color-balance-based method (UIEC2Net [11]), a physical model and CNN-fusion-based method (Chen et al. [46]), a multi-stage method (Deep-WaveNet [10]), and a DWT-based method (Ma et al. [17]).

Evaluation Metrics: For full-reference datasets, full-reference evaluations were conducted using the PSNR [47], SSIM [48], PCQI [49], and MSE [50] metrics. The PSNR and MSE were used to assess content similarities between the enhanced and reference images; a higher PSNR value (a lower MSE) indicated that the result was closer to the reference in terms of image content. The SSIM and PCQI were used to assess contrast and structure similarity; a higher SSIM value (a higher PCQI value) indicated that the result was more similar to the reference in terms of image structure and texture. For non-reference datasets that did not have reference images, the non-reference evaluation metrics UIQM and UCIQE were used to measure the performance of the methods. UIQM [51] and UCIQE [52] were used to assess the non-uniform color cast and contrast of enhanced images; a higher UIQM value (a higher UCIQE value) indicated better results.

4.2. Comparisons with SOTA Methods on Full-Reference Datasets

Visual Comparisons: Due to the absorption of light as it propagated through water (red corresponds to the longest wavelength, which is absorbed first, followed by orange and yellow), images captured underwater were predominantly bluish, greenish, and yellowish, as shown in Figure 3a, Figure 4a, and Figure 5a, respectively. However, images captured underwater were also affected by other factors, such as light source and sea area. Based on this aspect, images were divided into low-light and shallow-water images, as shown in Figures 6a and 7a. To illustrate the effect of the proposed network on different types of underwater images better, inspired by [11], the test images were divided into five types: bluish underwater images, greenish underwater images, yellowish underwater images, low-illuminated underwater images, and shallow-water images, as shown in Figures 3–7, respectively. The ULAP mostly relied on underwater imaging models and prior knowledge, which made it less robust to complex scenes and even aggravated the effect of the color cast. As shown in Figures 3–5, the Shallow-UWnet and UResnet could not effectively remove color casts from bluish, greenish, and yellowish underwater images. The models of Chen et al., Ma et al., Deep WaveNet, UIEC2Net, and the proposed network performed comparatively better for color correction. Among them, the proposed method performed best regarding both color correction and detail preservation. As shown in Figure 6, neither the traditional method nor the deep learning method performed very well when the raw image was collected under low-illuminated conditions. However, compared to the other methods, the results of the proposed method were the closest to the reference image. As shown in Figure 8, the method of Ma et al. and the proposed method performed better in image detail information preservation when the raw image was a shallow-water image. Moreover, in Figures 3–7, it can be seen that the proposed method achieved the highest PSNR score among all methods. This verified the effectiveness of the proposed method.

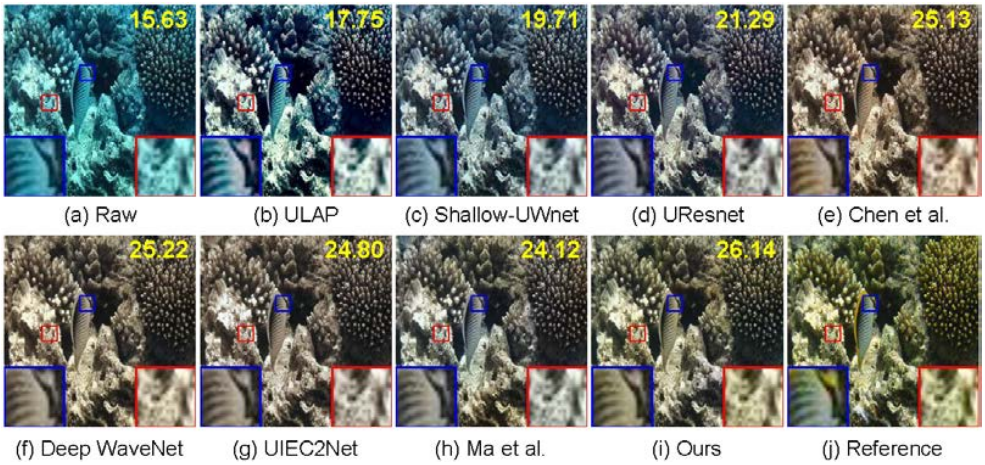

Figure 3. The visual comparison results of different methods on the bluish underwater images. The number presented on the top-right corner of each image refers to its PSNR. More results are shown in Figure A1 [8–11,17,45,46].

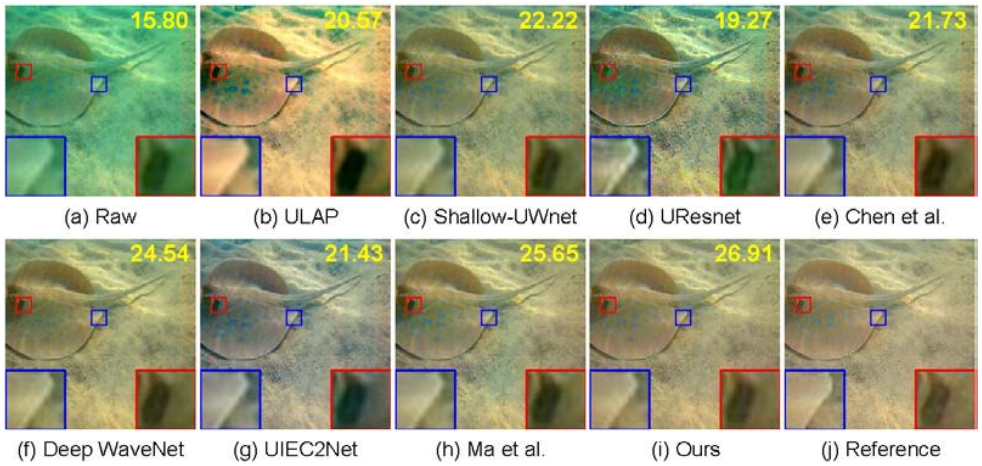

Figure 4. The visual comparison results of different methods on the greenish underwater images. The number presented on the top-right corner of each image refers to its PSNR. More results are shown in Figure A2 [8–11,17,45,46].

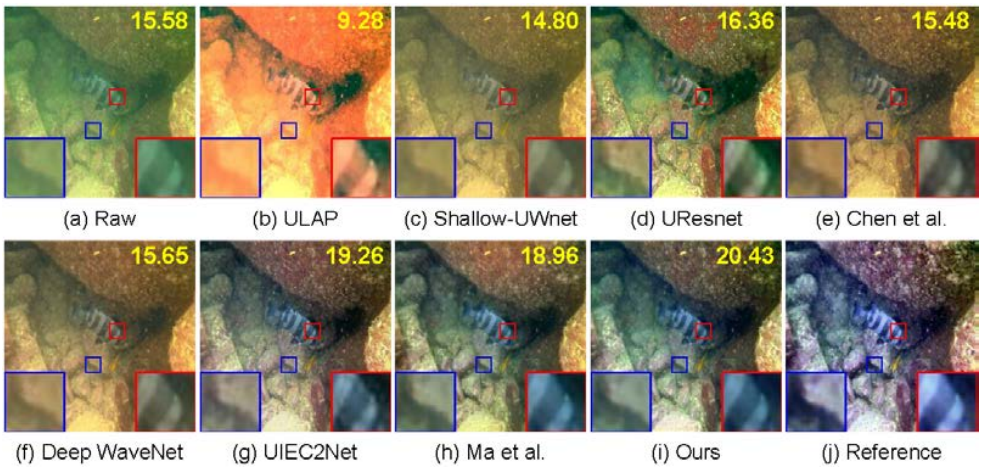

Figure 5. The visual comparison results of different methods on the yellowish underwater images. The number presented on the top-right corner of each image refers to its PSNR. More results are shown in Figure A3 [8–11,17,45,46].

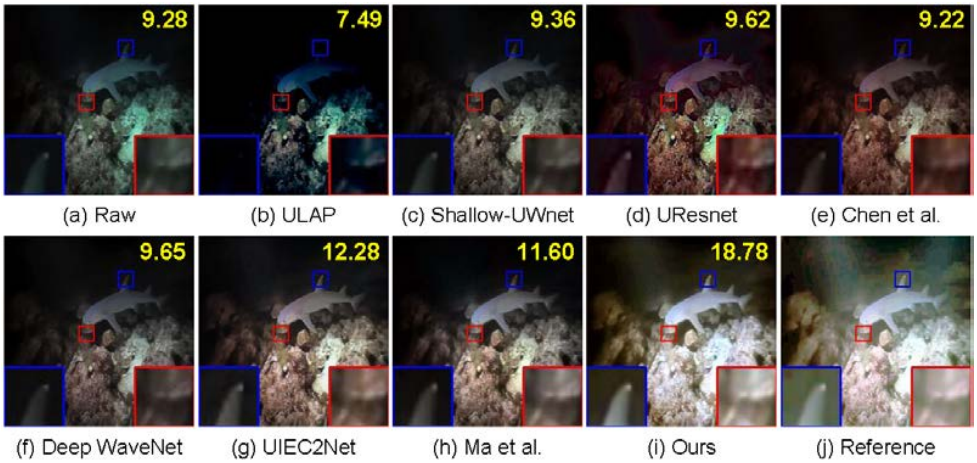

Figure 6. The visual comparison results of different methods on the low-illuminated underwater images. The number presented on the top-right corner of each image refers to its PSNR. More results are shown in Figure A4 [8–11,17,45,46].

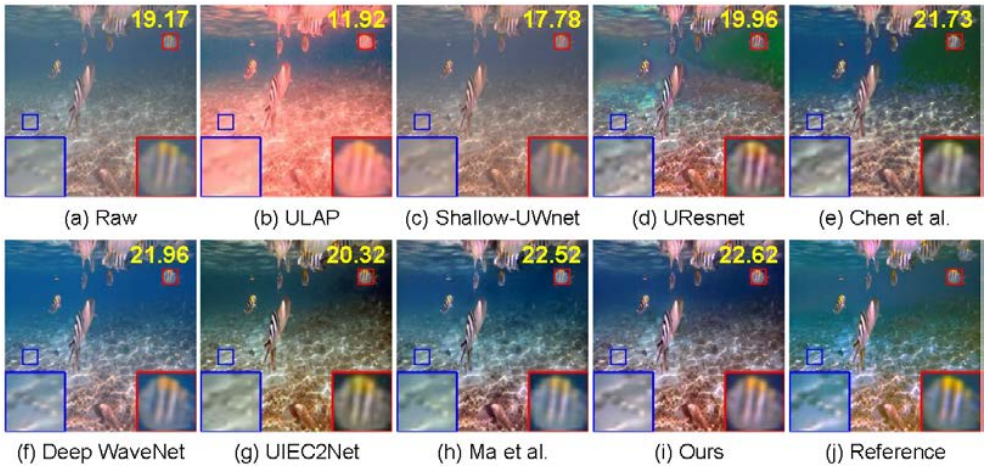

Figure 7. The visual comparison results of different methods on the shallow-water images. The number presented on the top-right corner of each image refers to its PSNR. More results are shown in Figure A5 [8–11,17,45,46].

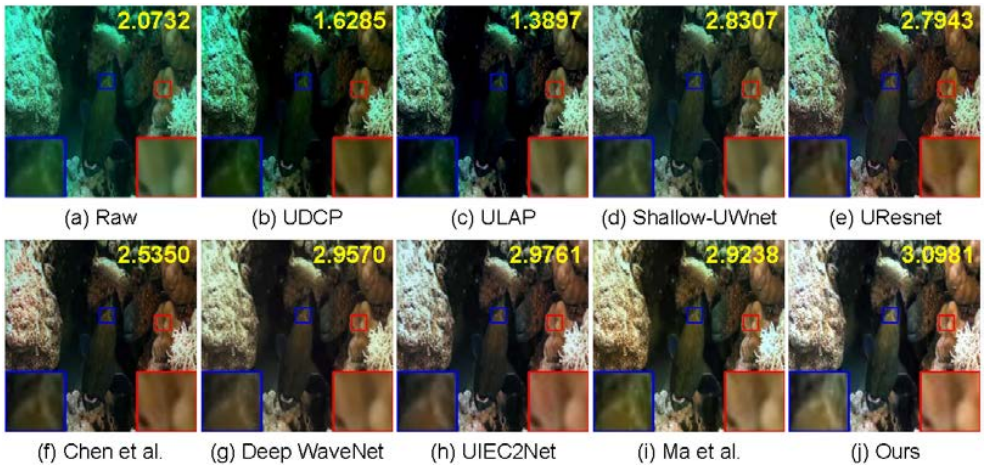

Figure 8. The visual comparison results of different methods on the underwater images from the Test76 dataset. The number presented on the top-right corner of each image refers to its UIQM. More results are shown in Figure A6 [3,8–11,17,45,46].

Quantitative Comparisons: A quantitative comparison of the methods was performed on the UIEB, UFO-120, and EUVP datasets in terms of average values of the PSNR, SSIM, MSE, and PCQI metrics. The quantitative results are presented in Tables 1–3. As presented in Table 1, the proposed WE-Net outperformed all methods in terms of the PSNR, SSIM, and MSE metrics on the UIEB dataset. Compared with the second-best-performing method, the proposed method achieved improvements of 5.95%, 4.33%, and 21.06% in terms of the PSNR, SSIM, and MSE metrics, respectively. As presented in Tables 2 and 3, the proposed WE-Net outperformed all competing methods on all metrics on the UFO-120 and EUVP datasets. Compared with the second-best-performing method, the proposed method achieved improvements of 3.42%, 1.55%, 15.83%, and 1.27%, and 3.42%, 1.12%, 11.82%, and 0.8% in terms of PSNR, SSIM, MSE, and PCQI metrics on the UFO-120 and EUVP, respectively. According to the results on different datasets, the proposed network had significant advantages compared to the competing methods.

Table 1. The evaluation of different methods on the UIEB dataset in terms of average PSNR (dB), SSIM, MSE, and PCQI. Bold and underlined values denote the best and second-best results, respectively.

Method	PSNR↑	SSIM↑	MSE↓	PCQI↑
UDCP [3]	11.68	0.5362	5.1172	0.8521
ULAP [45]	15.59	0.7345	2.8694	0.9177
Shallow-UWnet [9]	17.36	0.7686	1.7166	**1.0816**
UResnet [8]	17.91	0.7498	1.4731	0.7698
Chen et al. [46]	17.81	0.7552	1.5475	0.8876
Deep WaveNet [10]	18.71	0.8127	1.3427	<u>1.0178</u>
UIEC^{2}Net [11]	<u>21.31</u>	0.8310	<u>0.7739</u>	0.8429
Ma et al. [17]	19.87	<u>0.8536</u>	1.0449	0.9618
Ours	**22.58**	**0.8906**	**0.6109**	0.9206

Table 2. The evaluation of different methods on the UFO-120 dataset in terms of average PSNR (dB), SSIM, MSE, and PCQI. Bold and underlined values denote the best and second-best results, respectively.

Method	PSNR↑	SSIM↑	MSE↓	PCQI↑
UDCP [3]	14.48	0.5252	2.8719	0.6968
ULAP [45]	19.47	0.6952	0.8744	0.6660
Shallow-UWnet [9]	23.56	0.7629	0.3142	0.7619
UResnet [8]	23.30	0.7686	0.3346	0.6782
Chen et al. [46]	23.21	0.7589	0.3666	0.7404
Deep WaveNet [10]	24.42	0.7791	0.2625	0.7663
UIEC^{2}Net [11]	24.15	0.8033	0.2806	0.7213
Ma et al. [17]	<u>26.30</u>	<u>0.8055</u>	<u>0.1768</u>	<u>0.7609</u>
Ours	**27.20**	**0.8180**	**0.1488**	**0.7706**

Table 3. The evaluation of different methods on the EUVP dataset in terms of average PSNR (dB), SSIM, MSE, and PCQI. Bold and underlined values denote the best and second-best results, respectively.

Method	PSNR↑	SSIM↑	MSE↓	PCQI↑
UDCP [3]	13.81	0.6174	3.2111	0.7340
ULAP [45]	18.10	0.7384	1.2602	0.7063
Shallow-UWnet [9]	21.17	0.8406	0.6862	0.8524
UResnet [8]	21.32	0.8256	0.6431	0.7678
Chen et al. [46]	22.92	0.8566	0.4675	0.8520
Deep WaveNet [10]	23.13	0.8601	0.4676	0.8696
UIEC^{2}Net [11]	23.35	<u>0.8740</u>	0.4286	0.8454
Ma et al. [17]	<u>23.96</u>	0.8728	<u>0.3900</u>	<u>0.8721</u>
Ours	**24.78**	**0.8838**	**0.3439**	**0.8791**

4.3. Comparisons with SOTA Methods on Non-Reference Datasets

To demonstrate the robustness of the proposed network, comparative experiments on non-reference datasets Test76 and RUIE-UTTS were conducted. The qualitative results

are shown in Figures 8 and 9, where it can be observed that the enhanced results obtained by the proposed network preserved both essential colors and detailed image information. Moreover, the proposed method achieved the highest UIQM (Figure 8) and the highest UCIQE (Figure 9) among all competing methods. The quantitative comparison results of different methods on the Test76 and RUIE-UTTS datasets are given in Table 4, where the average values of the non-reference evaluation metrics UIQM and UCIQE of different methods are presented. As shown in Table 4, the proposed model achieved the best quantitative performance on the non-reference datasets among all competing methods, except for the UIQM metric on the RUIE-UTTS, which was lower than those of the UResnet, UIEC^{2}Net, and method of Ma et al. According to the quantitative and qualitative results, the proposed WE-Net had high effectiveness.

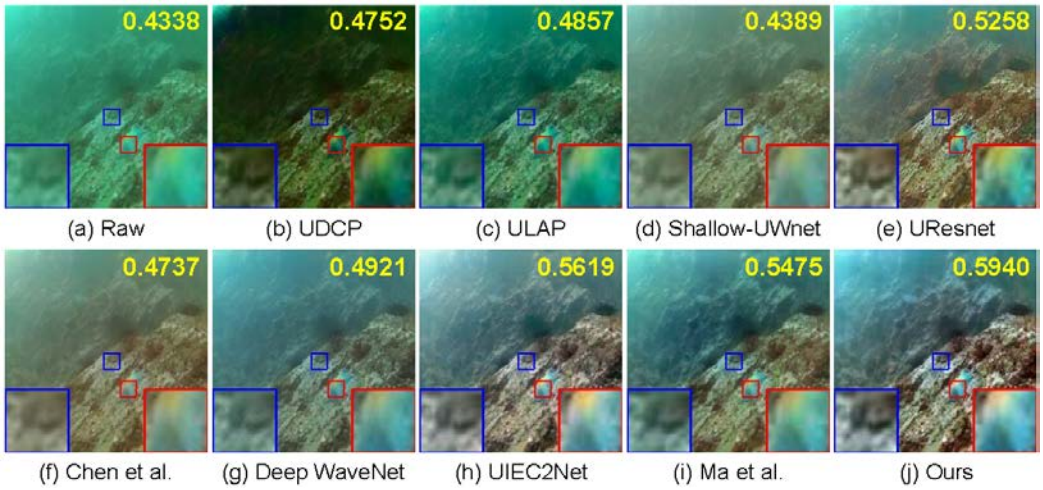

Figure 9. The visual comparison results of different methods on the underwater images from the RUIE-UTTS dataset. The numbers presented on the top-right corner of each image refer to its UCIQE. More results are shown in Figure A7 [3,8–11,17,45,46].

Table 4. The evaluation of different methods on the Test76 and RUIE-UTTS datasets in terms of average UIQM and UCIQE. Bold and underlined values denote the best and second-best results, respectively.

Method	Test76		RUIE-UTTS	
	UIQM↑	UCIQE↑	UIQM↑	UCIQE↑
UDCP [3]	1.3489	0.5386	2.2369	0.5201
ULAP [45]	1.5708	0.5222	2.6003	<u>0.5275</u>
Shallow-UWnet [9]	2.1192	0.4659	2.8849	0.4577
UResnet [8]	2.3534	0.5218	**3.0769**	0.5076
Chen et al. [46]	2.2492	0.4993	2.8633	0.4850
Deep WaveNet [10]	2.3492	0.4977	2.9712	0.4788
UIEC^{2}Net [11]	<u>2.5421</u>	<u>0.5473</u>	<u>3.0514</u>	0.5181
Ma et al. [17]	2.4884	0.5361	3.0436	0.5200
Ours	**2.5596**	**0.5589**	3.0229	**0.5425**

4.4. Ablation Studies

Extensive ablation experiments were performed to analyze the effects of the main components of the proposed WE-Net, including the residual-structured Swin Transformer group (RSTG), the edge enhancement branch (EEB), the DWT, and the RDAM. More specifically, w/o RSTG denotes the proposed WE-Net without the residual-structured Swin Transformer group; w/o EEB denotes the proposed WE-Net without the edge enhancement branch; and w/o DWT denotes the proposed WE-Net without both the DWT and the IDWT (the convolution and deconvolution stride values were adjusted to replace down- and

up-sampling operations, respectively); w/o RDAM denotes the proposed WE-Net without the RDAM; and lastly, w/one RDAM represents the proposed WE-Net with one RDAM.

As presented in Table 5, the full proposed model achieved the best quantitative performance on the EUVP and UFO-120 datasets when compared with the ablated models, except that the MSE metric on the EUVP dataset and the PCQI on the UFO-120 dataset were lower than those of w/o DWT and w/o RSTG, respectively. As shown in Figure 10, the full proposed model achieved the best visual results in terms of essential color recovery and detailed information preservation among all methods. As shown in Figure 10, for w/o RSTG, w/o DWT, w/o EEB, and w/o RDAM, although the noise in the enhanced images was significantly reduced, the color bias was severe. Generally, the improvements in the color cast of w/o RDAM and the proposed method were more obvious and closer to the reference image than those of the other methods. The qualitative and quantitative results indicated the high effectiveness of each component in the proposed model.

Table 5. The quantitative results of the ablation study in terms of average PSNR (dB), SSIM, MSE, and PCQI values. Bold values show the best performer.

Method	EUVP				UFO-120			
	PSNR↑	SSIM↑	MSE↓	PCQI↑	PSNR↑	SSIM↑	MSE↓	PCQI↑
w/o RSTG	24.44	0.8812	0.3668	0.8742	27.07	0.8179	0.1512	**0.7718**
w/o EEB	24.55	0.8812	0.3600	0.8778	26.90	0.8178	0.1565	0.7632
w/o RDAM	24.52	0.8792	0.3605	0.8698	26.93	0.8108	0.1581	0.7651
w/one RDAM	24.75	0.8830	0.3445	0.8765	27.14	0.8148	0.1504	0.7682
w/o DWT	24.76	0.8828	**0.3421**	0.8782	27.04	0.8143	0.1527	0.7710
Full model	**24.78**	**0.8838**	0.3439	**0.8791**	**27.20**	**0.8180**	**0.1488**	0.7706

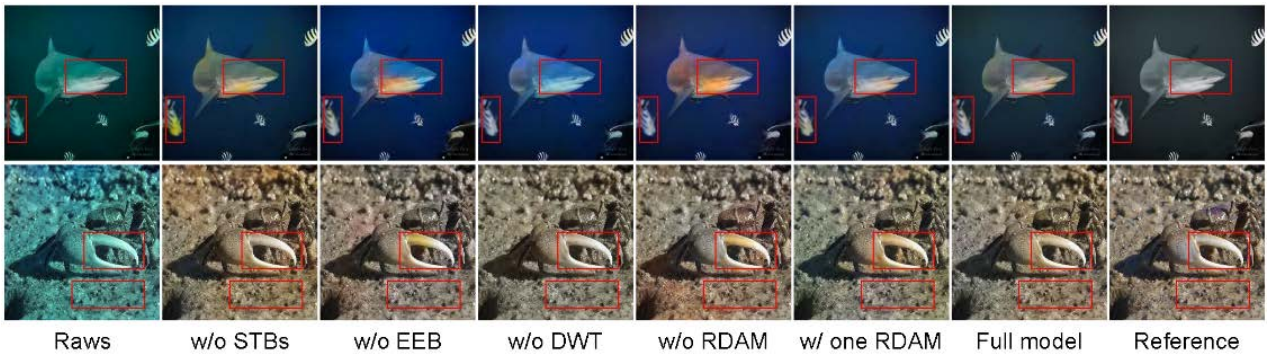

Figure 10. The visual comparison results of ablation study for the contributions of each component of WE-Net.

4.5. Limitations

From the above qualitative and quantitative experiments, it is observed that our proposed algorithm is significantly better than other methods in terms of visual effects, especially in dealing with bluish, greenish, yellowish, low-illuminated, and shallow-water images. This is mainly due to four parts: first, we introduced DWT and IDWT to replace traditional down-sampling and up-sampling operations, which can alleviate the information loss caused by the change of feature map; second, we introduced EEB to strengthen the protection of edge information; third, we introduced an RDAM, which can improve the ability of network color correction and detail information learning; finally, we added an RSTG module to the encoder-decoder to effectively learn the global features of the context.

However, it is the addition of these modules that also brings some limitations, including complexity and execution time. As shown in Figure 11, we compare the test time of our method with six SOTA methods based on the full- and non-reference datasets via a PC with a single Nvidia GeForce GTX 3090 GPU. From that, we observe that the test time of

our proposed method on five test datasets is higher than other methods. As we all know, the Transformer model has good global feature learning ability, but its model complexity is significantly higher. We added EEB and RSTG to the proposed model, which also makes our model more complex than other models. In the future, we will improve the enhanced visual effect by designing a low-complexity multi-path network structure to grade details, color, and noise through a separate pipeline.

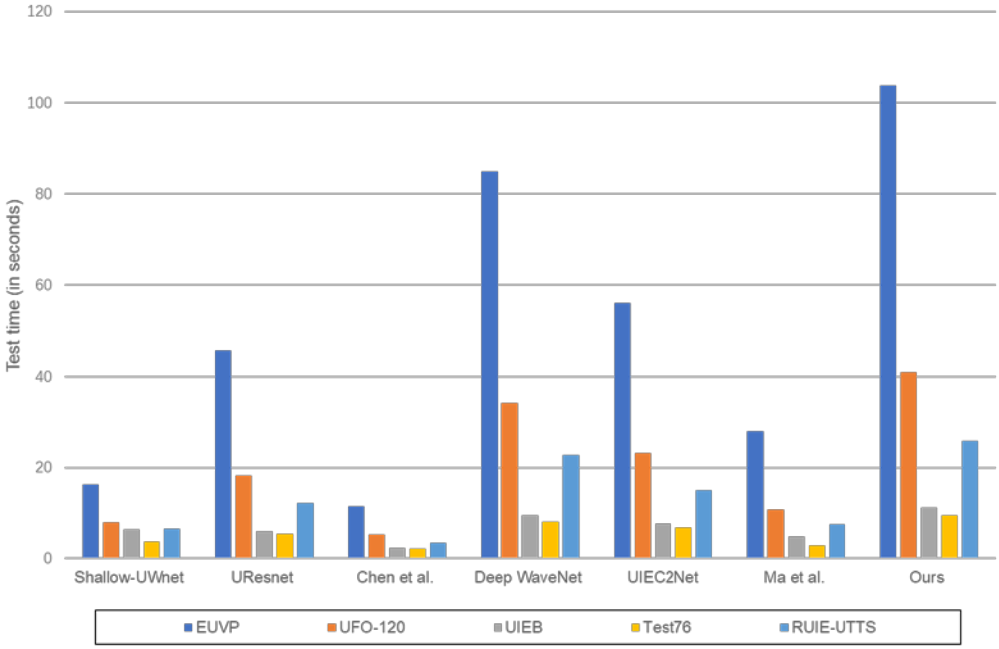

Figure 11. Comparison of our proposed network with other methods in terms of test time on five test datasets [8–11,17,46].

5. Conclusions

In this paper, WE-Net, which combines discrete wavelet transform and edge enhancement information, is proposed for underwater image enhancement. In the encoding stage, first, the edge image is obtained by performing the edge detection on a raw image and then used as input of a hybrid encoder, which can strengthen the fusion of edge information in the original feature maps. Then, the encoded feature map is fed to the residual-structured Swin Transformer group to obtain global features. Finally, feature decoding is performed to reconstruct the image. In the encoding and decoding processes, an improved residual dense attention module is used to extract and reconstruct features. Experiments on the full- and non-reference datasets show that the proposed method has excellent performance. The effectiveness of each component of the proposed model is verified by ablation experiments.

Author Contributions: All authors have made great contributions to this paper. Individual contributions are as follows: conceptualization, methodology, software, validation, writing—original draft preparation, K.S. and F.M.; formal analysis, investigation, data curation, writing—review and editing, Y.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Scientific Research Capacity Improvement Project of Key Developing Disciplines in Guangdong Province of China, grant number: 2021ZDJS057; National Natural Science Foundation of China, grant number: 61771225; and Natural Science Research Project of Anhui Education Department, grant number: KJ2020B07.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The open real-world underwater image datasets used in this paper were collected from the internet. EUVP dataset: http://irvlab.cs.umn.edu/resources/euvp-dataset (accessed on 25 March 2021); UFO-120 dataset: http://irvlab.cs.umn.edu/resources/ufo-120-datase (accessed on 25 March 2021); UIEB dataset: https://li-chongyi.github.io/proj_benchmark.html (accessed on 18 June 2021)); SQUID: http://csms.haifa.ac.il/profiles/tTreibitz/datasets/ambient_forwardlooking/index.html (accessed on 22 August 2021); RUIE dataset: https://github.com/dlut-dimt/RealworldUnderwater-Image-Enhancement-RUIE-Benchmark (accessed on 22 August 2021).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

As shown in Figures A1–A7, we provide more visual results of our method and others as a supplement to the visualizations in the main paper.

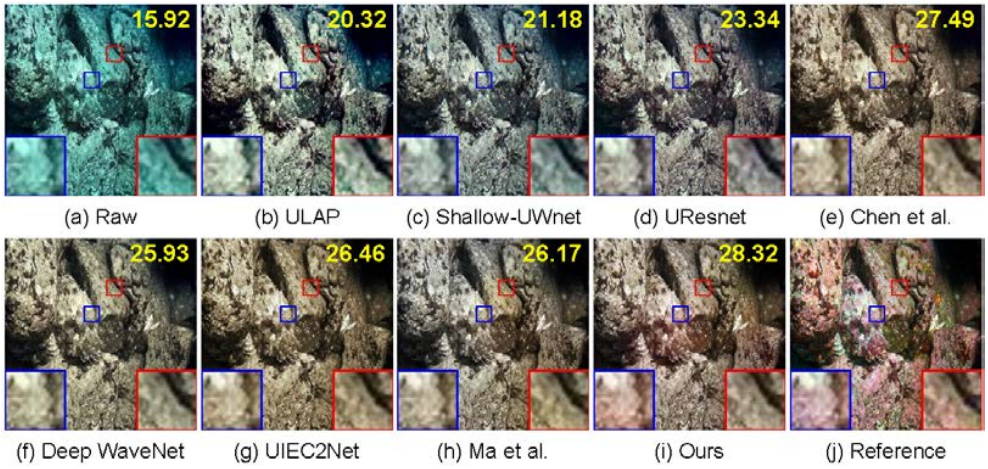

Figure A1. More visual comparison of different methods on bluish underwater images. The number presented on the top-right corner of each image refers to its PSNR [8–11,17,45,46].

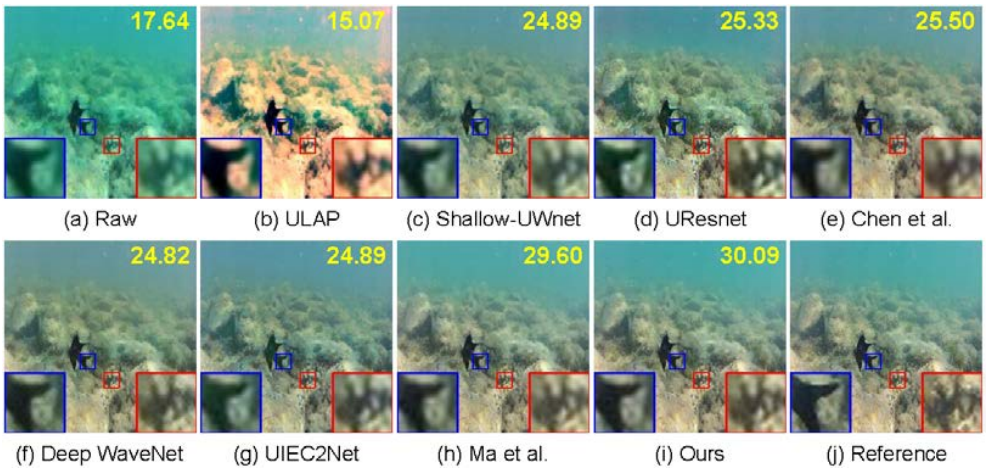

Figure A2. More visual comparison of different methods on greenish underwater images. The number presented on the top-right corner of each image refers to its PSNR [8–11,17,45,46].

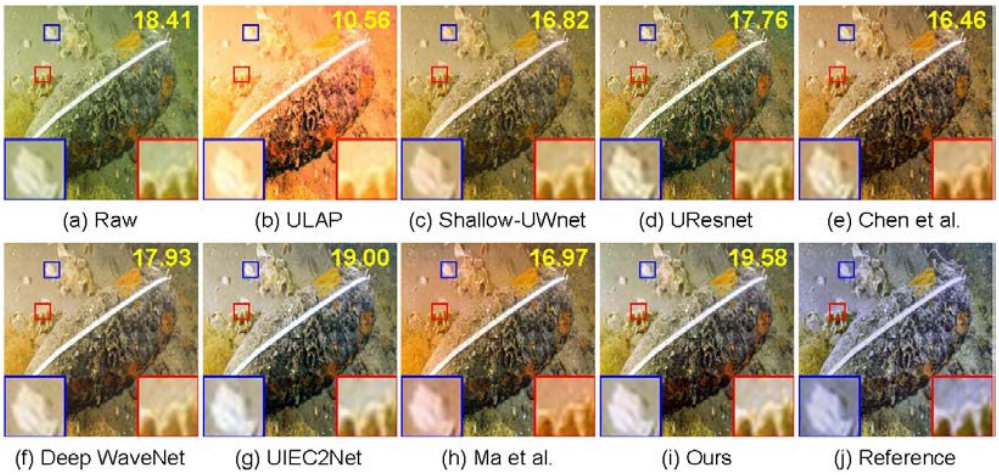

Figure A3. More visual comparison of different methods on yellowish underwater images. The number presented on the top-right corner of each image refers to its PSNR [8–11,17,45,46].

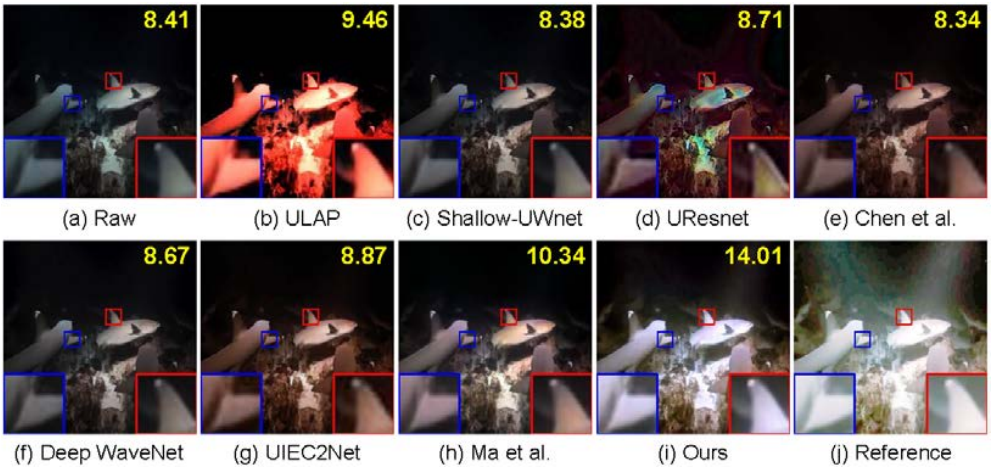

Figure A4. More visual comparison of different methods on low-illuminated underwater images. The number presented on the top-right corner of each image refers to its PSNR [8–11,17,45,46].

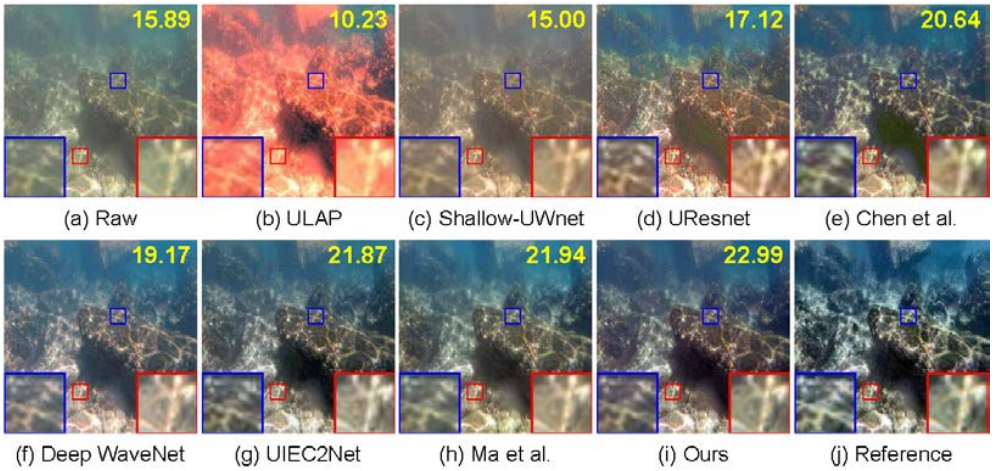

Figure A5. More visual comparison of different methods on the shallow water images. The number presented on the top-right corner of each image refers to its PSNR [8–11,17,45,46].

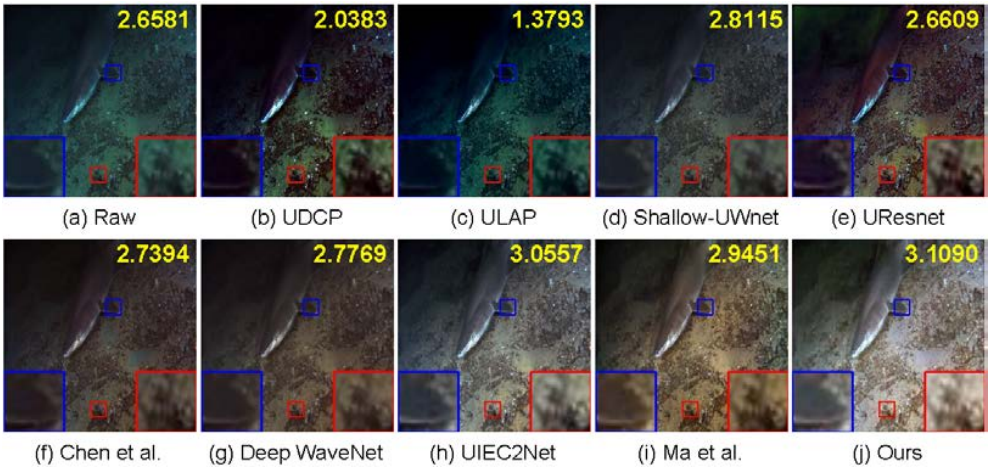

Figure A6. More visual comparison results of different methods on the underwater images from the Test76 dataset. The number presented on the top-right corner of each image refers to its UIQM [3,8–11,17,45,46].

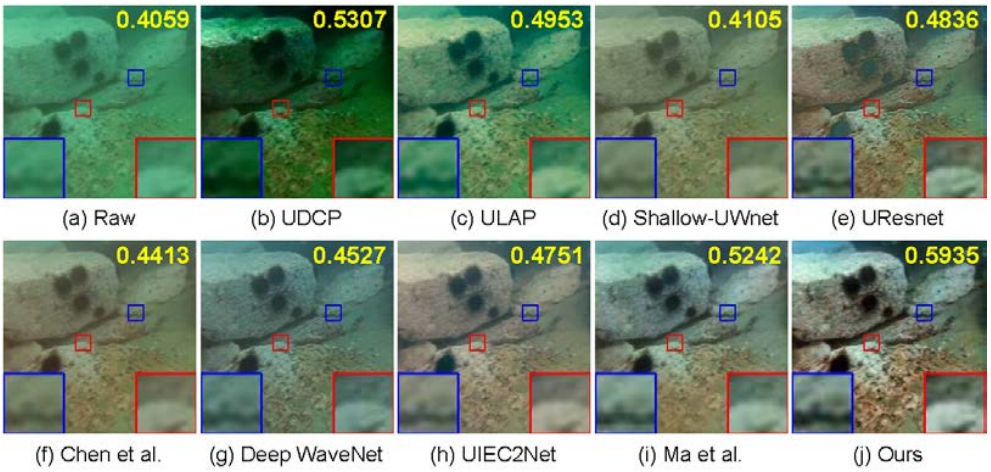

Figure A7. More visual comparison results of different methods on the underwater images from the RUIE-UTTS dataset. The number presented on the top-right corner of each image refers to its UCIQE [3,8–11,17,45,46].

References

1. Bonin-Font, F.; Oliver, G.; Wirth, S.; Massot, M.; Negre, P.L.; Beltran, J. Visual Sensing for Autonomous Underwater Exploration and Intervention Tasks. *Ocean Eng.* **2015**, *93*, 25–44. [CrossRef]
2. Li, A.; Yu, L.; Tian, S. Underwater Biological Detection Based on YOLOv4 Combined with Channel Attention. *J. Mar. Sci. Eng.* **2022**, *10*, 469. [CrossRef]
3. Drews, P.; Nascimento, E.; Moraes, F.; Botelho, S.; Campos, M. Transmission Estimation in Underwater Single Images. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Sydney, Australia, 1–8 December 2013; pp. 825–830.
4. Zhang, W.; Liu, W.; Li, L. Underwater Single-Image Restoration with Transmission Estimation Using Color Constancy. *J. Mar. Sci. Eng.* **2022**, *10*, 430. [CrossRef]
5. Figueiredo, M.A.T.; Nowak, R.D. An EM Algorithm for Wavelet-based Image Restoration. *IEEE Trans. Image Process.* **2003**, *12*, 906–916. [CrossRef] [PubMed]
6. Figueiredo, M.A.T.; Bioucas-Dias, J.M.; Nowak, R.D. Majorization–minimization Algorithms for Wavelet-based Image Restoration. *IEEE Trans. Image Process.* **2007**, *16*, 2980–2991. [CrossRef]
7. Zhang, S.; Wang, T.; Dong, J.; Yu, H. Underwater Image Enhancement Via Extended Multi-scale Retinex. *Neurocomputing* **2017**, *245*, 1–9. [CrossRef]
8. Liu, P.; Wang, G.; Qi, H.; Zhang, C.; Zheng, H.; Yu, Z. Underwater Image Enhancement with a Deep Residual Framework. *IEEE Access* **2019**, *7*, 94614–94629. [CrossRef]
9. Naik, A.; Swarnakar, A.; Mittal, K. Shallow-UWnet: Compressed Model for Underwater Image Enhancement. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 2–9 February 2021; pp. 15853–15854.

10. Sharma, P.K.; Bisht, I.; Sur, A. Wavelength-based Attributed Deep Neural Network for Underwater Image Restoration. *arXiv* **2021**, arXiv:2106.07910.
11. Wang, Y.; Guo, J.; Gao, H.; Yue, H. UIEC^{2}Net: CNN-based Underwater Image Enhancement Using Two Color Space. *Signal Process. Image Commun.* **2021**, *96*, 116250. [CrossRef]
12. Peng, L.; Zhu, C.; Bian, L. U-shape Transformer for Underwater Image Enhancement. *arXiv* **2021**, arXiv:2111.11843.
13. Hu, K.; Weng, C.; Zhang, Y.; Jin, J.; Xia, Q. An Overview of Underwater Vision Enhancement: From Traditional Methods to Recent Deep Learning. *J. Mar. Sci. Eng.* **2022**, *10*, 241. [CrossRef]
14. Liu, X.; Pedersen, M.; Wang, R. Survey of Natural Image Enhancement Techniques: Classification, Evaluation, Challenges, and Perspectives. *Digit. Signal Process.* **2022**, *127*, 103547. [CrossRef]
15. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.X.; Zhang, Z.; Lin, S.; Guo, B.N. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021; pp. 10012–10022.
16. Liu, P.; Zhang, H.; Lian, W.; Zuo, W. Multi-level Wavelet Convolutional Neural Networks. *IEEE Access* **2019**, *7*, 74973–74985. [CrossRef]
17. Ma, Z.; Oh, C. A Wavelet-based Dual-stream Network for Underwater Image Enhancement. *arXiv* **2022**, arXiv:2202.08758.
18. Aytekin, C.; Alenius, S.; Paliy, D.; Gren, J. A Sub-band Approach to Deep Denoising Wavelet Networks and a Frequency-adaptive Loss for Perceptual Quality. *arXiv* **2021**, arXiv:2102.07973.
19. Yang, H.H.; Yang, C.H.H.; Wang, Y.C.F. Wavelet Channel Attention Module with a Fusion Network for Single Image Deraining. In Proceedings of the IEEE International Conference on Image Processing, Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 883–887.
20. Chen, Y.; Huang, J.; Wang, J.; Xie, X. Edge Prior Augmented Networks for Motion Deblurring on Naturally Blurry Images. *arXiv* **2021**, arXiv:2109.08915.
21. Liang, T.; Jin, Y.; Li, Y.; Wang, T. EDCNN: Edge Enhancement-based Densely Connected Network with Compound Loss for Low-dose CT Denoising. In Proceedings of the IEEE International Conference on Signal Processing, Beijing, China, 6–9 December 2020; pp. 193–198.
22. Kim, K.; Chun, S.Y. SREdgeNet: Edge Enhanced Single Image Super Resolution Using Dense Edge Detection Network and Feature Merge Network. *arXiv* **2018**, arXiv:1812.07174.
23. Liu, X.; Ma, Y.; Shi, Z.; Chen, J. GridDehazeNet: Attention-based Multi-scale Network for Image Dehazing. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 7314–7323.
24. Dai, L.; Liu, X.; Li, C.; Chen, J. AWNet: Attentive Wavelet Network for Image ISP. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 185–201.
25. Li, J.; Wang, W.; Chen, C.; Zhang, T.X.; Zha, S.; Wang, J.; Yu, H. TransBTSV2: Wider Instead of Deeper Transformer for Medical Image Segmentation. *arXiv* **2022**, arXiv:2201.12785.
26. Song, L.; Liu, G.; Ma, M. TD-Net: Unsupervised Medical Image Registration Network Based on Transformer and CNN. *Appl. Intell.* **2022**, *52*, 1–9.
27. Gao, G.; Xu, Z.; Li, J.; Yang, J.; Zeng, T.; Qi, G.J. CTCNet: A CNN-Transformer Cooperation Network for Face Image Super-Resolution. *arXiv* **2022**, arXiv:2204.08696.
28. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.D.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y.Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306.
29. Ruikar, S.D.; Doye, D.D. Wavelet Based Image Denoising Technique. *Int. J. Adv. Comput. Sci. Appl.* **2011**, *2*, 49–53.
30. Gnanadurai, D.; Sadasivam, V. An Efficient Adaptive Thresholding Technique for Wavelet Based Image Denoising. *Int. J. Electron. Commun. Eng.* **2008**, *2*, 1703–1708.
31. Zhou, J.; Wei, X.; Shi, J.; Chu, W.; Lin, Y. Underwater Image Enhancement Via Two-level Wavelet Decomposition Maximum Brightness Color Restoration and Edge Refinement Histogram Stretching. *Opt. Express* **2022**, *30*, 17290–17306. [CrossRef]
32. Yu, Y.; Zhan, F.; Lu, S.; Pan, J.X.; Ma, F.Y.; Xie, X.S.; Miao, C.Y. WaveFill: A Wavelet-based Generation Network for Image Inpainting. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 14114–14123.
33. Dharejo, F.A.; Zawish, M.; Zhou, F.D.Y.; Dev, K.; Khowaja, S.A.; Qureshi, N.M.F. Multimodal-Boost: Multimodal Medical Image Super-Resolution using Multi-Attention Network with Wavelet Transform. *arXiv* **2021**, arXiv:2110.11684.
34. Riba, E.; Mishkin, D.; Shi, J.; Ponsa, D.; Moreno-Noguer, F.; Bradski, G. A Survey on Kornia: An Open Source Differentiable Computer Vision Library for PyTorch. *arXiv* **2020**, arXiv:2009.10521.
35. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual Dense Network for Image Super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2472–2481.
36. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
37. Qin, X.; Wang, Z.L.; Bai, Y.C.; Xie, X.D.; Jia, H.Z. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 11908–11915.
38. Sun, K.C.; Meng, F.; Tian, Y.B. Progressive Multi-branch Embedding Fusion Network for Underwater Image Enhancement. *J. Vis. Common. Image R.*. (Minor Revise).

39. Islam, M.J.; Xia, Y.; Sattar, J. Fast Underwater Image Enhancement for Improved Visual Perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [CrossRef]
40. Islam, M.J.; Luo, P.; Sattar, J. Simultaneous Enhancement and Super-resolution of Underwater Imagery for Improved Visual Perception. *arXiv* **2020**, arXiv:2002.01155.
41. 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]
42. Liu, R.; Fan, X.; Zhu, M.; Hou, M.; Luo, Z. Real-World Underwater Enhancement: Challenges, Benchmarks, and Solutions Under Natural Light. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 4861–4875. [CrossRef]
43. Berman, D.; Levy, D.; Avidan, S.; Treibitz, T. Underwater Single Image Color Restoration Using Haze-lines and a New Quantitative Dataset. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 2822–2837. [CrossRef]
44. Loshchilov, I.; Hutter, F. SGDR: Stochastic Gradient Descent with Warm Restarts. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
45. Song, W.; Wang, Y.; Huang, D.; Tjondronegoro, D. A Rapid Scene Depth Estimation Model Based on Underwater Light Attenuation Prior for Underwater Image Restoration. In Proceedings of the Pacific Rim Conference on Multimedia, Hefei, China, 21–22 September 2018; pp. 678–688.
46. Chen, X.; Zhang, P.; Quan, L.; Yi, C.; Lu, C. Underwater Image Enhancement Based on Deep Learning and Image Formation Model. *arXiv* **2021**, arXiv:2101.00991.
47. Hore, A.; Ziou, D. Image Quality Metrics: PSNR vs. SSIM. In Proceedings of the International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.
48. 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] [PubMed]
49. Wang, S.; Ma, K.; Yeganeh, H.; Wang, Z.; Lin, W. A Patch-structure Representation Method for Quality Assessment of Contrast Changed Images. *IEEE Signal Process. Lett.* **2015**, *22*, 2387–2390. [CrossRef]
50. Hunt, B.R. The Application of Constrained Least Squares Estimation to Image Restoration by Digital Computer. *IEEE Trans. Comput.* **1973**, *100*, 805–812. [CrossRef]
51. Panetta, K.; Gao, C.; Agaian, S. Human-visual-system-inspired Underwater Image Quality Measures. *IEEE J. Ocean. Eng.* **2015**, *41*, 541–551. [CrossRef]
52. Yang, M.; Sowmya, A. An Underwater Color Image Quality Evaluation Metric. *IEEE Trans. Image Process.* **2015**, *24*, 6062–6071. [CrossRef]