

UHD Underwater Image Enhancement via Frequency-Spatial Domain Aware Network

Yiwen Wei, Zhuoran Zheng, and Xiuyi Jia[✉]

School of Computer Science and Engineering, Nanjing University of Science and
Technology, Nanjing, China
jiaxy@njjust.edu.cn

Abstract. Currently, carrying ultra high definition (UHD) imaging equipment to record rich environmental conditions in deep water has become a hot issue in underwater exploration. However, due to the poor light transmission in deep water spaces and the large number of impurity particles, UHD underwater imaging is often plagued by low contrast and blur. To overcome these challenges, we propose an efficient two-path model (UHD-SFNet) that recovers the color and the texture of an underwater blurred image in the frequency and the spatial domains. Specifically, the method consists of two branches: in the first branch, we use a bilateral enhancement pipeline that extracts the frequency domain information of a degraded image to reconstruct clear textures. In the pipeline, we embed 1D convolutional layers in the MLP-based framework to capture the local characteristics of the token sequence. In the second branch, we develop U-RSGNet to capture the color features of the image after Gaussian blurring to generate a feature map rich in color information. Finally, the extracted texture features are fused with the color features to produce a clear underwater image. In addition, to construct paired high-quality underwater image enhancement dataset, we propose UHD-CycleGAN with the help of domain adaptation to produce more realistic UHD synthetic images. Experimental results show that our algorithm outperforms existing methods significantly in underwater image enhancement on a single GPU with 24G RAM. Codes are available at <https://github.com/wyw0112/UHD-SFNet>.

1 Introduction

Over the past few years, underwater image enhancement has received increasing attention as a fundamental task to improve advanced marine applications and services. Especially with the popularity of ultra high definition (UHD) imaging devices, there is an increasing demand for clear UHD underwater images for marine applications and services. Unfortunately, light in deep water is affected by wavelength-dependent absorption and scattering [1–4], resulting in problems such as low contrast and color cast in underwater images. In addition, underwater microorganisms and impurities can further enhance light scattering,

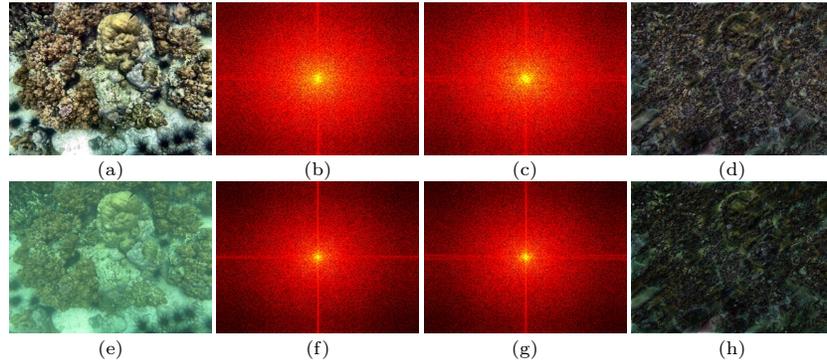


Fig. 1. The figure shows the difference in the normalized Fourier coefficients and the high frequency information between a pair of clear and blurred images. Where (a) is the clear image, (b) and (c) are the real and imaginary parts of (a) in the frequency domain after conversion to a grayscale map, respectively, (d) is the high-frequency part of (a), (e) is the blurred image, and (f)-(h) as above.

which can seriously interfere with the subsequent analysis of the computer vision. Therefore, it is difficult but very important to design efficient underwater image enhancement methods.

Traditional methods [3, 5–7] implement the task of underwater image enhancement with the help of statistical properties of the image and physical assumptions. However, these methods solve the color cast of images mainly by using manual priors, which make it difficult to correct low contrast and missing texture information in dynamic scenes. Recently, deep learning methods have been widely used to enhance underwater images, including standard convolution methods [8, 9], methods with physical priors [10], and GAN-based methods [11–13]. Although these methods accomplish state-of-the-art results, their modeling pattern is to stack a large number of convolution kernels on the spatial domain to reconstruct the missing details, the masking of impurities, and the bluish (greenish) tones, ignoring the role of information in the frequency domain.

In order to use the spatial properties in the frequency domain to better learn feature patterns that are difficult to learn in the spatial domain, another research line [14–16] try to enhance images in the frequency domain, such as underwater image enhancement [16], and image deblurring [14]. These methods usually use Fourier transform or wavelet transform to obtain the frequency domain coefficients of the degraded image, and later use methods such as thresholding, filtering or deep learning to bridge the gap between the degraded image and the clear image. However, most of these methods ignore the role played by the real and imaginary parts of the frequency domain coefficients (complex tensor). The common practice of these methods is to assign a weight to a set of complex numbers, or discretize an image to obtain high-frequency and low-frequency features and process them separately. However, using fixed weights will result in

the model being not able to reconcile the different semantic information of the images well, so we use a dynamic aggregation token method to process the real and imaginary parts of the frequency domain features. The viewpoint can be supported by Wave-MLP [17], which obtains a sizable performance benefit by splitting the real and imaginary parts of the estimated Fourier coefficients and refining them separately. Although PRWNet [16] deals with high-frequency and low-frequency features after wavelet transform respectively, the recovery of the frequency-domain characteristics of the image requires long-distance dependent features of the image. So it needs to stack a lot of convolutions to recover vibrant color and textures, which results in its inability to process UHD images on a single 24G RAM GPU. Since the framework of MLP is easy to establish long-distance dependence on the image, in the paper, we use the MLP-based model to extract features on the frequency domain to recover the blurred texture information. Furthermore, as shown in Figure 1, we demonstrate the difference in the normalized Fourier coefficients and the high frequency information between a pair of clear and blurred images.

To solve the above problems, we propose UHD-CycleGAN to generate more realistic underwater blurred images on a single GPU with 24G RAM, and develop UHD-SFNet to generate high quality underwater enhanced images. To better recover the image texture information, our frequency domain perception sub-network based on MLP extracts the frequency domain information of an image by embedding the spatial induction bias in the bilateral technique. It is worth noting that by using a parallel twin channel, the real and imaginary parts of the estimated Fourier coefficients are refined separately to produce a high-quality bilateral grid. For the spatial domain branch, we use U-RSGNet to reconstruct the color information of the blurred underwater image by ingesting a gated residual to enhance the channel domain properties.

The contributions of this paper are summarized as follows:

- We propose a new network framework that can enhance UHD underwater images by extracting information in the spatial and the frequency domains. The framework complements the frequency and the spatial domain information of the image to better recover the color and texture information of the image, and the framework can process UHD underwater images on a single 24G RAM GPU.
- In the frequency domain branch, we embed the 1D convolution in the MLP-based method to capture the local features inside the image patches as a complement to the global frequency domain information that can recover the sharp edges of the image. In the spatial domain branch, we introduce the gated residual to enhance the properties of the channel domain to restore the natural color.
- We propose the novel UHD-CycleGAN to generate the U-Water100 dataset using 100 high-quality underwater clear images of different sizes. A large number of experimental results demonstrate the SOTA results achieved by our underwater image enhancement method.

2 Related Work

2.1 Underwater Image Enhancement

The early research methods were mainly physical model-free and physical model-based methods. The physical model-free methods aim to modify the image pixel values to improve the contrast of underwater blurred images. Iqbal et al. [18] stretched the dynamic pixel range of RGB color space and HSV color space to improve the contrast and saturation of underwater images. Ghani and Isa [19,20] made improvement for the problem that Iqbal et al. [18] method causes over-enhancement or under-enhancement. Ancuti et al. [21] recovered clear images by mixing contrast-enhanced images and color-corrected images in a multiscale fusion strategy. Fu et al. [22] proposed a new retinex-based enhancement method to enhance a single underwater image.

The physical model-based methods treat underwater image enhancement as an uncertain inverse problem where handcrafted priors are used to estimate the potential parameters of the image formation model. Chiang et al. [5] and Drews-Jr et al. [6] implemented underwater image enhancement by modifying the dark channel prior (DCP) proposed by He et al. [23]. Li et al. [7] proposed a hybrid method including color correction and underwater image deblurring to improve the visual quality of degraded underwater images. Akkaynak et al. [3] proposed a modified underwater image formation equation which is a physically more accurate model.

With the continuous development of deep learning and the accumulation of large datasets, data-driven methods have become increasingly popular in recent years. These methods mainly use convolutional operations to extract image features instead of using various manually extracted prior features. Due to the lack of underwater image enhancement datasets, early work used generative adversarial networks (GAN) to generate datasets or perform unpaired learning. Li et al. [24] first applied GAN to the generation of underwater blurred images. Jamadandi et al. [12] used wavelet transform to reconstruct the signal better. Uplavikar et al. [13] enabled the model to better discriminate between different types of underwater blur by introducing a classifier. Li et al. [9] constructed an underwater image enhancement benchmark (UIEB) and proposed a convolutional neural network trained on this benchmark. Li et al. [10] proposed an underwater image enhancement network with multicolor spatial embedding guided by medium transport, combining the advantages of physical models to deal with off-color and low-contrast problems. Huo et al. [16] enabled the network to progressively refine the underwater images in the spatial and the frequency domains by using a wavelet boosting learning strategy.

2.2 UHD Image Processing

Gabiger-Rose et al. [25] used a simultaneous field-programmable gate array implementation of bilateral filters to make more efficient use of dedicated resources. Jie et al. [26] proposed the Laplace Pyramid Translation Network (LPTN) to

avoid direct processing of high resolution images by feature extraction and fusion of multiple low resolution images. Lin et al. [27] used a base network to compute a low-resolution result and a second network to refine selected patches at high resolution to perform image keying in real time. Wang et al. [28] used a light weight two-head weight predictor with two outputs to perform fast image enhancement.

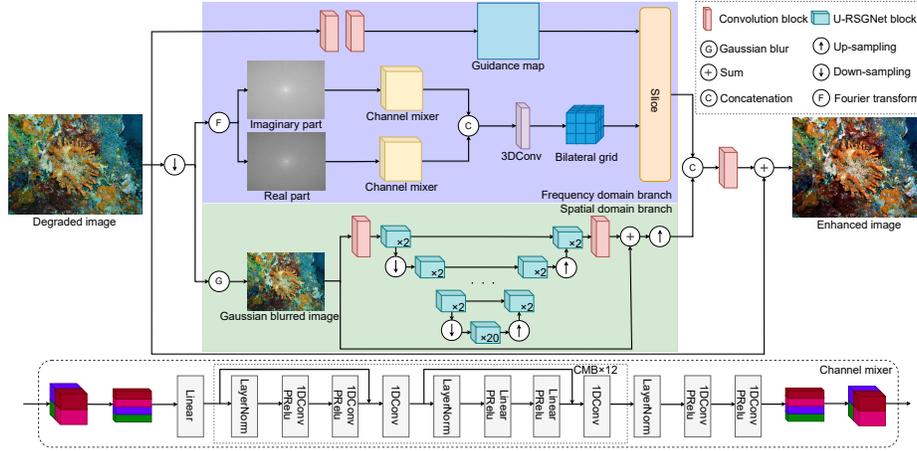


Fig. 2. In the purple region, the frequency domain coefficients are divided into the real and imaginary parts. The real and imaginary domain parts are processed separately in a pair of channel mixers to extract the frequency domain information of the image. Afterwards, the generated pairs of frequency domain feature maps are concatenated together and filtered using 3D convolution kernel to produce a bilateral grid. Next, the bilateral grid is queried using the guidance map to reconstruct the texture of the clear image. In the light green region, the spatial domain branch uses an encoder-decoder structure with gated residuals to process the blurred image to extract the color information of the image. Finally, the extracted frequency domain information and spatial domain information are fused to produce a clear image.

3 Proposed Method

As shown in Figure 2, in order to efficiently reconstruct the texture and color of a degraded image, we develop a two-path network dealing with the spatial and the frequency features. Specifically, we first use bilinear interpolation to downsample an underwater blurred image of arbitrary size to improve the speed of the model in extracting features in the frequency and the spatial domains. Next, in the spatial domain branch, we use Gaussian blur for the low-resolution input image (focusing on reconstructing the color information [29]), and then extract the spatial domain information of the image with the help of U-RSGNet.

In the frequency domain branch, we use the fast Fourier transform to obtain the Fourier coefficients of an image. In particular, we extract features from the real and imaginary parts of the obtained Fourier coefficients using a pair of channel mixers to construct a high-quality bilateral grid. Then we slice [30] the obtained bilateral grid with the guidance map to obtain a frequency domain feature map rich in texture information. Finally, the output feature maps in the spatial and the frequency domains are fused with the input image to obtain a clear and colorful image.

3.1 Frequency Domain Information Extraction

Inspired by Wave-MLP, the frequency domain features can easily be modelled by MLPs to obtain accurate results. However, the MLP-based methods do not take into account the inductive bias of the local space of patches within the image, which is undoubtedly a loss for extracting local information from the image. For this reason, we use a pair of channel mixers to extract the frequency domain features of the degraded image. First, the input image I is converted into Fourier coefficients F by fast Fourier transform. Next, we divide the Fourier coefficients F into a real part $real \in \mathbb{R}^{(C \times H \times W)}$ and an imaginary part $imag \in \mathbb{R}^{(C \times H \times W)}$, where H , W and C are the length, width and channel of the image respectively. In addition, we divide $real$ and $imag$ into patches of length and width P and stretch them into 1D sequences of token embeddings $T_{real} \in \mathbb{R}^{((H/P) \times (W/P)) \times (C \times P \times P)}$ and $T_{imag} \in \mathbb{R}^{((H/P) \times (W/P)) \times (C \times P \times P)}$. Then, they are processed using channel mixers to obtain a real part feature F_r and an imaginary part feature F_i , respectively. The Channel Mixer Block (CMB) consists of several LayerNorm layers, linear layers, 1D convolutional layers, and PReLU activation function layers. The details are as follows:

$$T = CN_3(I + P(CN_2(P(CN_1(N(I)))))) , \quad (1)$$

$$T' = CN_4(T + P(L_2(P(L_1(N(T)))))) , \quad (2)$$

where I is the input feature map, N is the layer normalization, CN is the 1D convolutional layer, L is the fully connected layer, and P is the PReLU activation function. The overall channel mixer first processes the input image by using a linear layer, and then feeds the feature maps into twelve CMBs to extract the image frequency domain information. The output feature map of the last CMB is processed by using a layer normalization, two 1D convolutions and PReLU activation functions to refine the high frequency features to obtain the final output.

We concatenate the real part features F_r and the imaginary part features F_i . Then we compress the frequency domain features into a four-dimensional affine bilateral grid after filtering by a 3D convolution kernel. The coordinates of the grid are in three dimensions. The reshaped frequency domain features can be viewed as a $16 \times 16 \times 16$ bilateral grid B , where each grid cell contains 3 digits. We process the original resolution image using convolution blocks to generate a

guidance map G with bootstrap functionality. We use the slicing operation [30] to generate a high-quality feature map F_f by querying the bilateral grid through the guidance map.

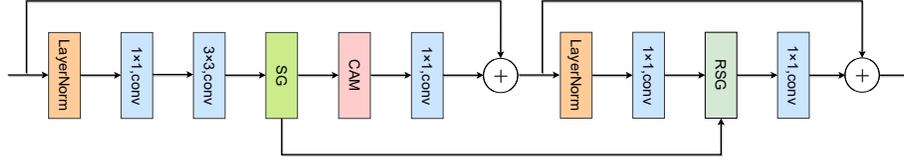


Fig. 3. The proposed structure of the U-RSGNet block in which the gated residual is ingested to enhance the channel domain characteristics.

3.2 Spatial Domain Information Extraction

To further utilize the rich color information in the spatial domain to recover colorful images, we introduce the spatial domain branch. This branch utilizes U-RSGNet to progressively enhance the channel information (color information) of the input features. The branch is a U-shaped structure consisting of a contracting path and an expansive path. The branch has five layers, in the first four layers each path is composed of two U-RSGNet blocks, and in the last layer twenty U-RSGNet blocks are used to mix the channel information of the lowest resolution feature map. In order to shrink or expand the feature map, a 2D convolution or sub-pixel convolution is inserted after each processing stage of the path. The structure of the U-RSGNet block is shown in Figure 3. Because gated linear unit [31] can effectively improve the ability of the model to handle the long-distance dependence, we incorporate gated linear units in the network. The structure of the gated linear unit is as follows:

$$\text{Gate}(X) = f(X) \otimes \sigma(g(X)) , \quad (3)$$

where X is the feature map, f and g denote linear transformations, \otimes denotes element multiplication, and σ denotes the nonlinear activation function. Although the gating method with GELU activation function can enhance the modeling capability of the model, this will undoubtedly lead to higher model complexity. This runs counter to our philosophy of designing a lightweight model. To alleviate this problem, we use SimpleGate (SG) [32] as the basic feature enhancement unit to replace the traditional gating strategy. The details are as follows:

$$T_1, T_2 = \text{Chunk}(X) , \quad (4)$$

$$O = T_1 \otimes T_2 , \quad (5)$$

where the *Chunk* operation is to cut $X \in \mathbb{R}^{(C \times H \times W)}$ directly in the channel dimension into feature sub-images $T_1 \in \mathbb{R}^{((C/2) \times H \times W)}$ and $T_2 \in \mathbb{R}^{((C/2) \times H \times W)}$.

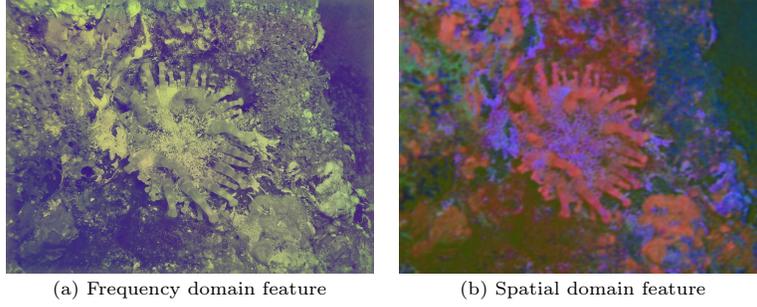


Fig. 4. (a) is the normalized output feature map of the frequency domain branch, and (b) is the normalized output feature map of the spatial domain branch.

T_1 and T_2 have the same size in each dimension. To further improve the U-RSGNet’s ability to model the channel information to reconstruct the color of a blurred image, we propose the Residual Simple Gate (RSG) as follows:

$$RSG(X, Y) = SG(Y + \beta * SG(X)) , \quad (6)$$

where $X \in \mathbb{R}^{(C \times H \times W)}$ and $Y \in \mathbb{R}^{((C/2) \times H \times W)}$ are the input feature map of the simple gate and the output feature map of the previous layer of the residual simple gate, respectively, $*$ is the channelwise product operation, and $\beta \in \mathbb{R}^{((C/2) \times 1 \times 1)}$ is the learnable channel attention vector. To further enhance the global attention capability of the model to obtain richer color information, our channel attention module (CAM) performs a global average pooling operation on the feature map to turn it into a channel attention vector $A \in \mathbb{R}^{(C \times 1 \times 1)}$, and then transposes A to $A^T \in \mathbb{R}^{(1 \times 1 \times C)}$, where C is the number of channels. Then, A^T is processed through a linear layer to enhance the global attention capability of the model. Finally, the transposed A^T is multiplied in the channel dimension with the input feature map to obtain the feature map rich in global information. The details are as follows:

$$CAM(X) = X * L(pooling(X)) , \quad (7)$$

where X is the input feature map, L is the fully connected layer, and $pooling$ is the global average pooling.

The whole process can be described as an input image I is Gaussian blurred to obtain a color image I_g , and then the color image is input to U-RSGNet to obtain a color-enhanced feature map F_c .

3.3 Spatial-Frequency Feature Fusion

To efficiently obtain a clear and colorful enhanced image, we use a standard feature fusion strategy at the end of the algorithm. Specifically, we concatenate the frequency domain feature map F_f and the spatial domain feature map F_c in the channel dimension and process them with a standard convolution block

(containing 3×3 and 1×1 convolutions with PReLU activation functions behind) to obtain a 3-channel feature map F_o . Finally, the final feature map F_o is summed with the original input image I with weights to obtain an enhanced image with sharp edges and rich color.

As seen in Figure 4(a), the output feature map of the frequency domain branch has sharper edges. In contrast, the recovery of an image’s color requires the network to have the ability to accurately extract channel features, which is capable of capturing the image’s long-distance dependent features. Therefore, we develop U-RSGNet to enhance the color characteristics of degraded images. In detail, we add gated residuals to enhance the characteristics of the channel domain to improve the color reconstruction capability of the model. As seen in Figure 4(b), the spatial domain branch focuses more on the color features of the image.

4 Experimental Results

In this section, we evaluate the proposed method by conducting experiments on synthetic datasets and real-world images. All results are compared with six state-of-the-art underwater image enhancement methods and one generic image enhancement method. These include two traditional methods (Ancuti et al. [21], Berman et al. [33]), a GAN-based method (FUnIE-GAN [11]), and four CNN-based methods (WaterNet [9], Ucolor [10], PRWNet [16], and NAFNet [32]). In addition, we perform ablation studies to show the effectiveness of each module of our network.

4.1 Training Data

To train and evaluate the proposed network as well as the comparison methods, we propose UHD-CycleGAN for migrating the style of underwater clear images to the style of blurred images on a single GPU. To generate high-quality UHD underwater images without noise, our generator first downsamples the images and feeds them into a convolutional network with an encoding-decoding structure (this structure can effectively mitigate the noise interference [34]). To avoid checkerboard artifacts caused by transposed convolution, we use bilinear interpolation for upsampling and downsampling, and add a 1×1 convolution with PReLU activation function to fill the gaps after the upsampling and downsampling operations. Specifically, we use all the blurred images of UIEBD as one style (including challenging images without corresponding reference) and clear images from U-Water100 and clear images from UIEBD as another style. We then migrate their styles to each other. Then a discriminator is used to determine whether the conversion is realistic or not. Notably, we also convert the converted images to the original style by a reverse generator and calculate their L1 loss from the original images to further promote more realistic conversion. We migrate the styles of 100 selected underwater clear images to the styles of underwater real blurred images in the UIEB dataset [9] to generate the

U-Water100 dataset. Our training dataset consists of a total of 890 underwater images. Among them, 800 are from UIEBD [9], and 90 are from U-Water100 that we produce. Accordingly, we use 90 images from UIEBD (T90) and 10 images from U-Water100 as the test set.

4.2 Implementation Details

The models are implemented in PyTorch and the networks are trained using the Adam optimizer. In this case, we use images of size 512×512 as input to train networks, with the batch size of 6. The initial learning rates for UHD-CycleGAN and UHD-SFNet are set to 0.00002 and 0.001, respectively. We train the UHD-SFNet for 400 epochs and UHD-CycleGAN for 10 epochs respectively. For WaterNet [9] and Ucolor [10], we fine-tune the networks using the same training data as ours based on the official models provided by the authors. For PRWNet [16] and NAFNet [32], we train their networks using the same dataset and the same experimental setup as ours. Notably, to allow the network to train and test UHD images on a single 24G RAM GPU, we downsample the inputs of Ucolor [10], PRWNet [16], and NAFNet [32] to 512×512 resolution and upsample them to the original resolution at the end.

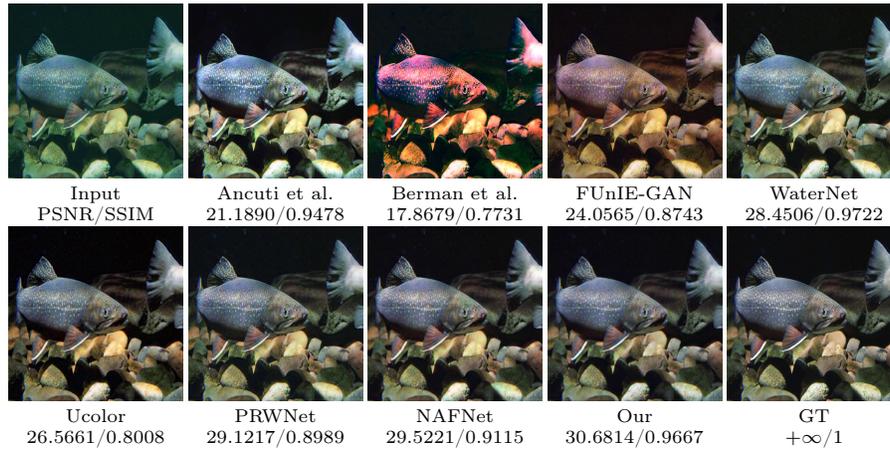


Fig. 5. Results of underwater image enhancement on the U-Water100 test dataset. Compared to other state-of-the-art methods, our method obtains better visual quality and recovers more image details.

4.3 Evaluation and Results

Quantitative Evaluation. Our proposed method is evaluated on two datasets, namely UIEB and U-Water100 datasets. All CNN-based methods are fine-tuned

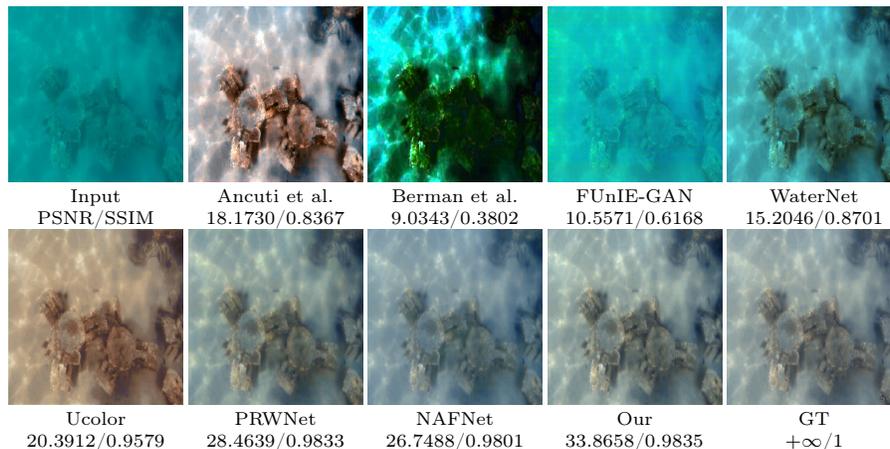


Fig. 6. Results of underwater image enhancement on the T90 dataset. Compared to other state-of-the-art methods, our method obtains better visual quality and recovers more image details.

or trained on UIEBD and U-Water100. We use PSNR and SSIM as evaluation metrics for full-reference images. For the reference-free real-world images, we use UIQM [35] as the evaluation metric. For all the three metrics, the higher score means better image quality. As can be seen in Table 1, we achieve the best results for both SSIM and PSNR metrics, but we only achieve the second best results for UIQM. However, due to the limited applicability of UIQM, it can only be used as a reference.

Qualitative Evaluation. Figures 5 and 6 show the results of the proposed method and the comparison methods on one UHD image from the U-Water100 and one image from the UIEB dataset, respectively. Figure 7 shows the results of the proposed method and the comparison methods on two challenging and unreferenced images from the UIEB dataset. It can be seen that the conventional methods [21, 33] tend to over-enhance the results and lead to color distortion. The GAN-based method [11] has weak color recovery capability and is prone to generate pseudo-streaks. The recent deep learning models [10, 16, 32] still have some ambiguity and color distortion in the results due to the lack of modeling capability. And the deep learning models [16, 32] stack a lot of convolutions in order to have higher performance resulting in the inability to process UHD images on a single 24G RAM GPU directly. Our algorithm is able to directly process UHD images and better recover the color and edges of the image. The enhanced underwater images produced by our algorithm in Figure 5-6 are close to the ground truth clear images. The images generated by our algorithm in Figure 7 have more realistic color and sharp edges.

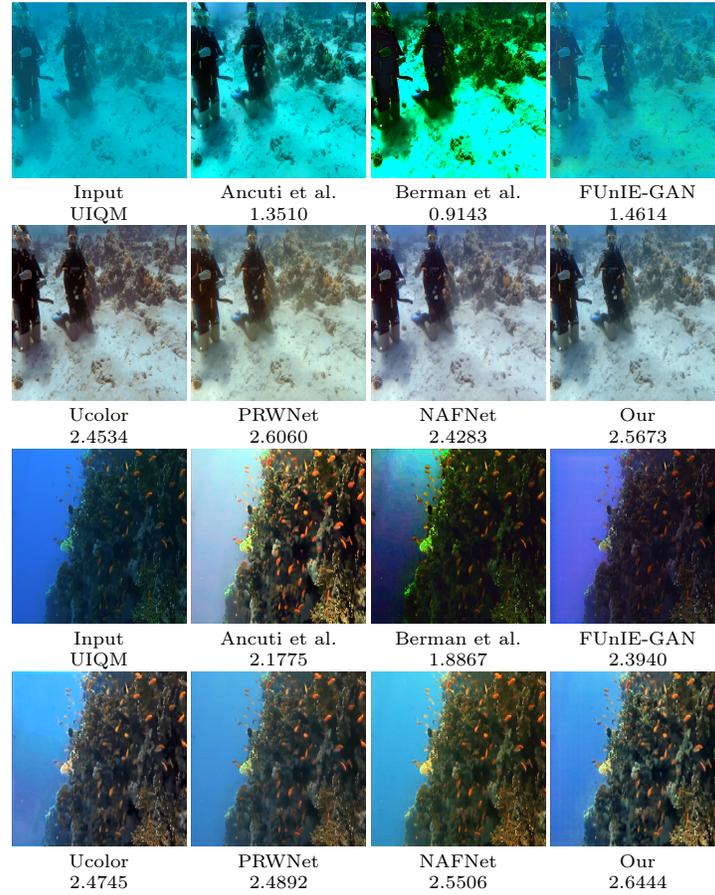


Fig. 7. Results of underwater image enhancement on the T60 dataset. Compared to other state-of-the-art methods, our method obtains better visual quality and recovers more image details.

4.4 Ablation Study

To demonstrate the effectiveness of each module introduced in the proposed network, we perform an ablation study including the following six experiments.

The Effectiveness of the Network of the Frequency Domain Branch.

We keep the structure of the network of the spatial domain branch unchanged and replace the network of the frequency domain branch with the network of the spatial domain branch. The images are processed with the new two-branch network to obtain the final outputs.

The Effectiveness of the Network of the Spatial Domain Branch.

We keep the structure of the network of the frequency domain branch unchanged

Table 1. Quantitative evaluation of the T60 dataset on UIQM metrics and quantitative evaluation of the T90 and U-Water100 datasets on PSNR and SSIM metrics.

	U-Water100		T90		T60
	PSNR	SSIM	PSNR	SSIM	UIQM
Input	21.8028	0.8955	17.5543	0.8191	1.6961
Ancuti et al. [21]	20.6852	0.8779	23.1564	0.9180	2.0978
Berman et al. [33]	16.1752	0.7197	15.5902	0.7205	1.5523
FUnIE-GAN [11]	20.4901	0.7891	17.9076	0.8057	2.3789
WaterNet [9]	22.3721	0.8859	19.5718	0.9057	2.1338
Ucolor [10]	23.7193	0.8578	20.3287	0.8538	2.2145
PRWNet [16]	23.5255	0.8390	23.8358	0.9293	2.4209
NAFNet [32]	24.2579	0.8615	24.3782	0.9288	2.2959
UHD-SFNet	25.0462	0.9158	25.2020	0.9426	2.4037

Table 2. The ablation studies of the network of the frequency domain branch, the network of the spatial domain branch, 1D convolution, residual simple gate, spatial domain information and frequency domain information are denoted as A, B, C, D, E and F, respectively.

	U-Water100		T90	
	PSNR	SSIM	PSNR	SSIM
A	20.4543	0.8741	23.6186	0.8890
B	22.9176	0.8749	19.8691	0.8997
C	24.7120	0.9085	22.0638	0.8482
D	24.4870	0.8957	24.2489	0.9210
E	16.2886	0.8599	18.9768	0.8771
F	24.5104	0.9089	24.1519	0.9229
Our	25.0462	0.9158	25.2020	0.9426

and replace the network of the spatial domain branch with the network of the frequency domain branch. The images are processed with the new two-branch network to obtain the final outputs.

The Effectiveness of the 1D Convolution. To ensure that the parameters and computational complexity of the model are similar to our method, we replace the 1D convolution in the channel mixer with the linear layer and compare it with our proposed method.

The Effectiveness of the Residual Simple Gate. We replace the residual simple gate with the simple gate in the network and compare it with our proposed method.

The Effectiveness of the Spatial Domain Information. The inputs to both branches of the network are images in the frequency domain. In this case, the input of the spatial domain branch is replaced by the real and imaginary parts of the frequency domain of the image with the same resolution as the inputs of the spatial domain branch.

The Effectiveness of the Frequency Domain Information. The inputs to both branches of the network are images in the spatial domain. In this case, the inputs of both the real and imaginary parts of the frequency domain branch are replaced by the images in the spatial domain with the same resolution as the inputs of the frequency domain branch.

Table 2 shows the results of our method compared with these six baselines on the U-Water100 and UIEB datasets. As can be seen in Table 2, the network of the spatial domain branch can achieve better metrics to some extent by relying on the color features of the images, but its edge recovery capability is still insufficient. Using only the network of the frequency domain branch does not produce satisfactory results, but its better ability to extract high-frequency information can be complemented by the ability to recover color from the network of the spatial-domain branch to produce enhanced images with rich color and sharp edges. Because 1D convolution can enhance the ability of the network to extract high-frequency features, it is used to compensate for the loss of local features caused by processing channel features with only linear layers. RSG can further improve the color recovery capability of the network by enhancing the channel domain features. Using the frequency domain information and the spatial domain information to complement each other can further refine the image edges and color to obtain a clearer and more colorful image.

5 Conclusion

In this paper, we propose a new framework for UHD underwater image enhancement in the spatial and the frequency domains. Our algorithm learns the features of the real and imaginary parts of the Fourier coefficients of the image to reconstruct the details of the image with the help of the channel mixer. In addition, we use the U-RSGNet with the RSG to recover the color information of the images. Quantitative and qualitative results show that our proposed network compares well with state-of-the-art underwater enhancement methods in terms of accuracy and produces visually pleasing results on real-world underwater images.

Acknowledgements. This work was supported by National Key Research and Development Program of China (2019YFB1706900), National Natural Science Foundation of China (62176123), Fundamental Research Funds for the Central Universities (30920021131) and Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX22_0461).

References

1. McGlamery, B.: A computer model for underwater camera systems. In: *Ocean Optics VI*. vol. 208, pp. 221–231 (1980)
2. Jaffe, J.S.: Computer modeling and the design of optimal underwater imaging systems. *JOE* **15**(2), 101–111 (1990)
3. Akkaynak, D., Treibitz, T.: A revised underwater image formation model. In: *CVPR*. pp. 6723–6732 (2018)
4. Akkaynak, D., Treibitz, T., Shlesinger, T., Loya, Y., Tamir, R., Iluz, D.: What is the space of attenuation coefficients in underwater computer vision? In: *CVPR*. pp. 4931–4940 (2017)
5. Chiang, J.Y., Chen, Y.C.: Underwater image enhancement by wavelength compensation and dehazing. *TIP* **21**(4), 1756–1769 (2011)
6. Drews, P.L., Nascimento, E.R., Botelho, S.S., Campos, M.F.M.: Underwater depth estimation and image restoration based on single images. *CG&A* **36**(2), 24–35 (2016)
7. Li, C., Guo, J., Guo, C., Cong, R., Gong, J.: A hybrid method for underwater image correction. *PRL* **94**, 62–67 (2017)
8. Li, C., Anwar, S., Porikli, F.: Underwater scene prior inspired deep underwater image and video enhancement. *PR* **98**, 107038 (2020)
9. Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., Tao, D.: An underwater image enhancement benchmark dataset and beyond. *TIP* **29**, 4376–4389 (2019)
10. Li, C., Anwar, S., Hou, J., Cong, R., Guo, C., Ren, W.: Underwater image enhancement via medium transmission-guided multi-color space embedding. *TIP* **30**, 4985–5000 (2021)
11. Islam, M.J., Xia, Y., Sattar, J.: Fast underwater image enhancement for improved visual perception. *RA-L* **5**(2), 3227–3234 (2020)
12. Jamadandi, A., Mudenagudi, U.: Exemplar-based underwater image enhancement augmented by wavelet corrected transforms. In: *CVPRW*. pp. 11–17 (2019)
13. Uplavikar, P.M., Wu, Z., Wang, Z.: All-in-one underwater image enhancement using domain-adversarial learning. In: *CVPRW*. pp. 1–8 (2019)
14. Mao, X., Liu, Y., Shen, W., Li, Q., Wang, Y.: Deep residual fourier transformation for single image deblurring. *arXiv preprint arXiv:2111.11745* (2021)
15. Qin, F., Li, C., Cao, L., Zhu, L., Zou, X., Li, X., Zhang, T., Xue, Y.: Blind image restoration with defocus blur by estimating point spread function in frequency domain. In: *ICAIP*. pp. 62–67 (2021)
16. Huo, F., Li, B., Zhu, X.: Efficient wavelet boost learning-based multi-stage progressive refinement network for underwater image enhancement. In: *ICCV*. pp. 1944–1952 (2021)
17. Tang, Y., Han, K., Guo, J., Xu, C., Li, Y., Xu, C., Wang, Y.: An image patch is a wave: Phase-aware vision mlp. In: *CVPR*. pp. 10935–10944 (2022)
18. Iqbal, K., Odetayo, M., James, A., Salam, R.A., Talib, A.Z.H.: Enhancing the low quality images using unsupervised colour correction method. In: *SMC*. pp. 1703–1709 (2010)
19. Ghani, A.S.A., Isa, N.A.M.: Underwater image quality enhancement through integrated color model with rayleigh distribution. *Applied Soft Computing* **27**, 219–230 (2015)
20. Ghani, A.S.A., Isa, N.A.M.: Enhancement of low quality underwater image through integrated global and local contrast correction. *Applied Soft Computing* **37**, 332–344 (2015)

21. Ancuti, C., Ancuti, C.O., Haber, T., Bekaert, P.: Enhancing underwater images and videos by fusion. In: CVPR. pp. 81–88 (2012)
22. Fu, X., Zhuang, P., Huang, Y., Liao, Y., Zhang, X.P., Ding, X.: A retinex-based enhancing approach for single underwater image. In: ICIP. pp. 4572–4576 (2014)
23. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. TPAMI **33**(12), 2341–2353 (2010)
24. 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. RA-L **3**(1), 387–394 (2017)
25. Gabiger-Rose, A., Kube, M., Weigel, R., Rose, R.: An fpga-based fully synchronized design of a bilateral filter for real-time image denoising. TIE **61**(8), 4093–4104 (2013)
26. Liang, J., Zeng, H., Zhang, L.: High-resolution photorealistic image translation in real-time: A laplacian pyramid translation network. In: CVPR. pp. 9392–9400 (2021)
27. Lin, S., Ryabtsev, A., Sengupta, S., Curless, B.L., Seitz, S.M., Kemelmacher-Shlizerman, I.: Real-time high-resolution background matting. In: CVPR. pp. 8762–8771 (2021)
28. Wang, T., Li, Y., Peng, J., Ma, Y., Wang, X., Song, F., Yan, Y.: Real-time image enhancer via learnable spatial-aware 3d lookup tables. In: ICCV. pp. 2471–2480 (2021)
29. Zheng, Z., Ren, W., Cao, X., Wang, T., Jia, X.: Ultra-high-definition image hdr reconstruction via collaborative bilateral learning. In: ICCV. pp. 4449–4458 (2021)
30. Gharbi, M., Chen, J., Barron, J.T., Hasinoff, S.W., Durand, F.: Deep bilateral learning for real-time image enhancement. TOG **36**(4), 1–12 (2017)
31. Dauphin, Y.N., Fan, A., Auli, M., Grangier, D.: Language modeling with gated convolutional networks. In: ICML. pp. 933–941 (2017)
32. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. arXiv preprint arXiv:2204.04676 (2022)
33. Berman, D., Treibitz, T., Avidan, S.: Diving into haze-lines: Color restoration of underwater images. In: BMVC (2017)
34. Li, S., Liu, X., Jiang, R., Zhou, F., Chen, Y.: Dilated residual encode–decode networks for image denoising. JEI **27**(6), 063005 (2018)
35. Panetta, K., Gao, C., Agaian, S.: Human-visual-system-inspired underwater image quality measures. JOE **41**(3), 541–551 (2015)