Signal, Image and Video Processing (2024) 18:3411–3420
https://doi.org/10.1007/s11760-024-03004-8

ORIGINAL PAPER

Underwater image enhancement using scale-patch synergy transformer

Lu Fan[1] · Bo Wang[1]

Received: 17 September 2023 / Revised: 25 December 2023 / Accepted: 3 January 2024 / Published online: 13 February 2024

Abstract
In the past few years, underwater image enhancement has attracted an increasing amount of research work because it plays an important role in computer vision related underwater tasks, such as aquatic robotics and marine engineering. However, wavelength-dependent light absorption and scattering introduces unpleasant color distortion and reduces the visibility of images in underwater scenes. In this paper, we propose a two-branch multi-scale (MSN) and multi-patch (MPN) synergy network, called Multi-SPNet, which aims to improve the contrast, brightness, and eliminate color distortion of non-uniform degraded underwater images. Specifically, the features extracted from multi-scale and multi-patch branches are interweaved for progressive image enhancement, where the upper and the lower branches utilize efficient Transformer blocks for learning multi-scale representation from low-to-high resolution and aggregating features via multiple image patches from fine to coarse level, respectively. The complementary branches can construct a synergistic merge to employ their mutual benefits for local and non-local pixel interactions. Extensive experiments on synthetic and real-world underwater image datasets clearly prove the effectiveness and superiority of the proposed Multi-SPNet against the state-of-the-art models both qualitatively and quantitatively.

Keywords Underwater image enhancement · Multi-scale and multi-patch · Synergy network · Transformer blocks

1 Introduction

Recently, underwater image enhancement has received much attention in computer vision, because the clear underwater images and videos are very important for the perception and understanding of underwater scenes [1, 2]. However, it is difficult to obtain good visual quality of underwater images due to wavelength-dependent light absorption as well as forward and backward scattering. Usually, these degraded images have lower visibility, reduced contrast, and color distortion, which seriously limited their practical applications for downstream vison tasks, such as target tracking, classification, detection, etc. [3, 4].

In the past few years, various methods for underwater image enhancement can be divided into three categories: physical model-based methods, non-physical model-based methods, and deep learning-based methods [5–7]. Physical model-based methods aim to perform defogging and visibility recovery based on prior assumptions, mathematically modeling the degradation process of underwater images, and estimating the parameters of optical imaging models so as to obtain corresponding clear images. For instance, Drews et al. [17]. proposed UDCP, which is a physical model-based restoration method restoring media transfer graphs and scene depth. However, when there are white objects or artificial light sources in the underwater environment, the results recovered by the algorithm are not optimal. While non-physical model-based methods improve the contrast and brightness of images by adjusting the image pixel values, and do not rely on underwater optical imaging models. For example, Peng et al. [19] proposed an underwater image recovery algorithm (IBLA) based on light absorption, which is used to obtain transmission maps by estimating atmospheric light values and depth maps of fuzzy images. This algorithm greatly improves image quality using the estimated depth, but does not solve the problem of color deviation. Although these methods can improve the image quality of underwater scenes to some extent, neither of them can meet

✉ Bo Wang
tjuwb@nxu.edu.cn

[1] School of Electronic and Electrical Engineering, Ningxia University, Yinchuan 750021, People's Republic of China

the requirements of real-world underwater image enhancement. In contrast, existing deep learning-based methods have achieved impressive performance in recent years [8, 9]. However, the mainstream of these methods generally adopted convolutional neural networks (CNNs), which have inherent shortcomings including limited receptive fields and static weights at inference stage.

To address these problems above, we propose a scale-patch synergy networks for underwater image enhancement, employing a two-branch multi-scale and multi-patch network with efficient Restormer to deal with color distortion and visibility degradation of underwater images. The multi-scale branch tends to eliminate color distortion by aggregating global features, while the multi-patch branch tends to learn spatial variant degradations and thus recover local details [10]. In a nutshell, our main contributions can be summarized as follows:

1. We propose a two-branch multi-scale and multi-patch transformer, where the multi-scale transformer fully removes color distortion from underwater images by aggregating global information, while the multi-patch transformer restores local details by learning spatial variance degradation.
2. We provide a synthetic underwater image dataset consisting of 400 synthetic underwater images classified into eight categories based on different degrees of degradation. Mimicking the real-world underwater images is beneficial for future research in underwater image enhancement.
3. Our proposed network can generalize well on both synthetic and real-world underwater images. Moreover, an efficient interweaved network structure can achieve decent results with only 5.19 M parameters.

2 Related work

2.1 Traditional methods

In order to enhance the visibility of underwater images, some traditional underwater image enhancement methods were proposed. Li et al. [11] proposed a method based on minimum information loss and histogram distribution prior to improve the contrast and brightness of underwater images, but it is prone to color bias and excessive local contrast. Song et al. [12] proposed a fast and effective underwater scene depth estimation model based on underwater light attenuation prior (ULAP) to recover the true scene radiance underwater, but its drawback is that it does not solve the color bias problem. Drews et al. [13] proposed an underwater dark channel prior method (UDCP) based on physical model to recover the medium transmission map and scene depth. However, when there are white objects or artificial light sources in the underwater environment, the results are not optimal.

2.2 Deep learning-based methods

In recent years, deep learning has made significant progress on low-vision tasks. For example, a multi-scale dense GAN network for underwater image enhancement was proposed by Li et al. [14]. The UWCNN proposed by Li et al. [15] synthesizes underwater images based on a modified underwater image formation model and the corresponding underwater scene parameters, but the method is less robust and cannot adapt to diverse underwater environments, and its performance results are often unsatisfactory especially for real underwater scenes. Meanwhile, Li et al. [16] proposed the first comprehensive perceptual study and analysis of underwater image enhancement using a large-scale real-world image dataset UIEB. Li et al. [17] proposed a Ucolor network based on multi-color space, and the transmission image is embedded as the weight of the image recovery area to prevent over-enhancement of the underwater image. The network has achieved great improvement in terms of visual quality and quantitative metric. However, when it is faced with very limited illumination in an underwater environment, this method cannot provide satisfied results. Shen et al. [18] presented UDAformer, a dual attention transformer-based method for underwater image enhancement. Although the UDAformer is prone to better reconstruct resulting images, it brings over-exposure and exhibits poor generalization.

2.3 Transformer-based methods

As an emerging scheme, Transformer-based methods showed remarkable performance in computer vision. Liu et al. [19] proposed the Shift Window Transformer (Swin Transformer), where shift windows bring higher efficiency by specifying self-attentive computations in non-overlapping local windows while allowing cross-window connections. Wang et al. [20] proposed Uformer, which introduces a new locally enhanced window Transformer block for denoising and deblurring, which largely reduces the computational complexity of feature maps. Zamir et al. [21] proposed Restormer, building two key modules: a multi-headed attention mechanism and a gated feedforward neural network. Restormer was designed as an efficient Transformer model and thus suitable for high-resolution images. Peng et al. [22] proposed a U-type Transformer based on an integrated channel-based multi-scale feature fusion converter module and a global feature modeling Transformer for underwater image enhancement. However, these methods cannot take full advantage of modeling image and patch relations in local and global scopes, thus ignore local and non-local pixel interactions.

Therefore, in order to solve the shortcomings, we propose a two-branch synergy network based on efficient Restormer for underwater images enhancement.

3 Proposed method

The overall framework of our proposed method and the detailed architecture of Restormer are shown in Figs. 1 and 2, respectively. The multi-scale network eliminates color distortion in different underwater scenes by aggregating global features, while the multi-patch network learns spatial degradation so as to recover local detail features.

3.1 Multi-scale branch

The upper multi-scale branch has three-stage architecture. Each level is composed of multiple Transformer blocks connected to learn at different scales. Specifically, the input image I is down sampled by 2 and 4 to form an image pyramid and the images are labeled as $I_{0.5}^{H}$ and $I_{0.25}^{H}$, respectively.

$$F_1 = \mathrm{Enc} - \mathrm{Dec}(I_{0.25}^{H}) \tag{1}$$

$$\mathrm{Skip1} = \mathrm{up4}(F_1) \tag{2}$$

where Enc – Dec indicates the whole process from encoder to decoder, up4(·) indicates an upsampling operation multiplier of 4. This is represented by the first green line in Fig. 1.

In the second stage, the feature map F_1 is enlarged by a factor of 2, with Skip*2* as part of the second level of input. Then, the feature maps F_2 are obtained by encoder–decoder with adding input $I_{0.5}^{H}$.

$$P_1 = \mathrm{up2}(F_1) \tag{3}$$

$$F_2 = \mathrm{Enc} - \mathrm{Dec}(P_1 + \mathrm{Skip2} + I_{0.5}^{H}) \tag{4}$$

where up2(·) indicates an up-sampling operation multiplier of 2.

In the third stage, Skip*3* is obtained by magnifying the feature map F_2 by a factor of 2, as input to the third stage part of the second branch, which is represented by the second green line in Fig. 1. magnifying the feature map F_2 of the second stage by a factor of 2, adding it to the original image I and feeding it together to the codec to obtain the feature map F_3:

$$\mathrm{Skip3} = \mathrm{up2}(F_2) \tag{5}$$

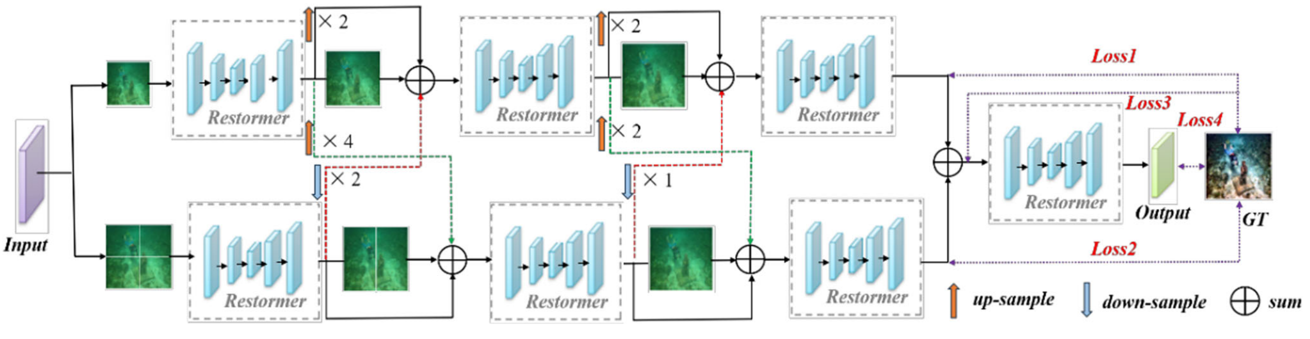

Fig. 1 Overview architecture of Multi-SPNet (Color figure online)

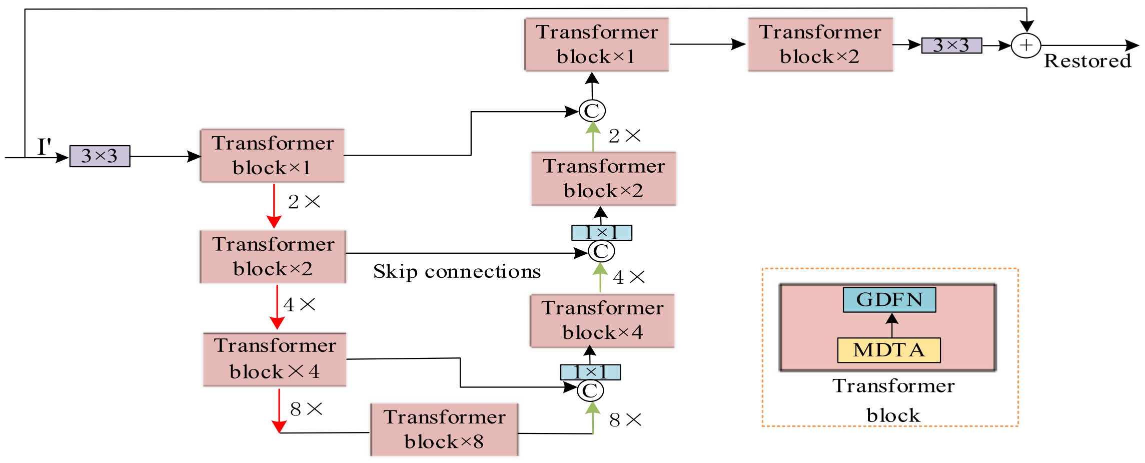

Fig. 2 The overall architecture of Restormer

$$P_2 = \mathrm{up2}(F_2) \tag{6}$$

$$F_3 = \mathrm{Enc} - \mathrm{Dec}(P_2 + \mathrm{Skip4} + I) \tag{7}$$

3.2 Multi-patch branch

The lower multi-patch branch is also a three-stage structure, with each stage working on different patches. Firstly, the input images are pre-processed: at the first level only one patch per image is considered. Secondly, the images are divided horizontally into two patches, denoted as $I^H_{2,1}$ and $I^H_{2,2}$. Finally, at the third level, it is further divided vertically into four patches based on the previous level, denoted as $I^H_{3,1}$, $I^H_{3,2}$, $I^H_{3,3}$, $I^H_{3,4}$ respectively. Thus, the network follows a bottom-up learning process of information flow. The patch from the third level is fed into the codec network to obtain the corresponding feature map F_4.

$$F_{4,j} = \mathrm{Enc} - \mathrm{Dec}(I^H_{3,j}), \; j = 1, 2, 3, 4 \tag{8}$$

We concatenate the horizontal feature mappings to obtain a new feature representation.

$$P_{4,j} = [F_{4,2j-1}, F_{4,2j}], \; j = 1, 2 \tag{9}$$

The pair is down-sampled as partial input to the second stage of the first branch, which is represented by the first red line in Fig. 1:

$$\mathrm{Skip2} = \mathrm{down}(P_4) \tag{10}$$

In the second stage, the output $P_{4,j}$ of the first layer, Skip1 of the first branch network and the input image $I^H_{2,i}, \; i = 1, 2$ are fed separately to the codec, and the output features are summed to obtain the feature map F_5.

$$F^*_{4,1} = \mathrm{Enc} - \mathrm{Dec}(P_{4,1} + I^H_{2,1}) \tag{11}$$

$$F^*_{4,2} = \mathrm{Enc} - \mathrm{Dec}(P_{4,2} + I^H_{2,2}) \tag{12}$$

$$F^*_{4,3} = \mathrm{Enc} - \mathrm{Dec}(\mathrm{Skip1}) \tag{13}$$

$$F_5 = [F^*_{4,1}, F^*_{4,2}] + F^*_{4,3} \tag{14}$$

At the first stage, a skip connection Skip4 is represented by the second red line in Fig. 1. which is used as part of the input to the third stage of the first branch, while the previous stage feature output, F_5 the skip connection Skip3 and the image I are summed and sent to the codec to obtain the feature map F_6.

$$\mathrm{Skip4} = \mathrm{up}(F_5) \tag{15}$$

$$F_6 = \mathrm{Enc} - \mathrm{Dec}(F_5 + \mathrm{Skip3} + I) \tag{16}$$

3.3 Fusion block

The final step of fusion for the above two branches is to cascade the feature maps F_3 and F_6 from the two-branch output to obtain feature map F_7, which is fed into the codec to obtain the final output image $\hat{I}$.

$$F_7 = F_3 + F_6 \tag{17}$$

$$\hat{I} = \mathrm{Enc} - \mathrm{Dec}(F_7) \tag{18}$$

3.4 Restormer structure

The model starts with a 3 × 3 convolution-based overlapping image patch embedding to obtain low level features, and these shallow features are converted to deep features by a 4-stage symmetric encoder-decoder. Starting with a high-resolution input, the encoder is applied to reduce the spatial size while expand the channel capacity, taking low-resolution features as input and progressively recovering the high resolution. Each level of the codec includes multiple Transformer blocks, which consists of two modules, i.e., multi-head separable convolutional transposition self-attention mechanism (MDTA) and the gated feedforward neural network (GDFN), and the increasing number of heads for multi-headed attention is [1, 2, 4, 8]. Finally, the refined image is concatenated with the original image as the output of the whole module.

3.5 Loss function

To obtain better visual quality and perceptual scores, we used a linear combination of three loss functions, including MSE loss, TV loss, and Charbonnier loss.

MSE loss: Mean Square Error (MSE) loss is commonly used for regression loss function. The smaller the MSE, the better the quality of the model, given an original image I of size $m * n$ and a noisy image K after adding noise to it. The formula is shown below:

$$\mathrm{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \tag{19}$$

TV loss: Differences in adjacent pixel values in an image can be resolved to some extent by reducing the TV loss, thus maintaining the smoothness of the output image. The TV loss can be expressed as:

$$L_{\mathrm{tv}} = ||\nabla_x \hat{I}||_2 + ||\nabla_y \hat{I}||_2 \tag{20}$$

Charbonnier loss: The advantage of Charbonnier loss over $L1$ loss is that the curve is smoother and the gradient near the zero value is not too small, which avoids the gradient explosion. The loss can be expressed as:

$$\ell(I', \hat{I}) = \sqrt{||I' - \hat{I}||^2 + \varepsilon^2} \tag{21}$$

where $\hat{I}$ denotes the GT image, I' is the enhanced image, as $\varepsilon = 10^{-3}$ is usually the case.

We calculate the total loss of the enhanced image $\hat{I}$ with each of the four branches, and the expressions are:

$$\mathrm{Loss}[i] = c_1 \mathrm{MSE}[i] + c_2 \mathrm{TV}[i] + c_3 \mathrm{Cha}[i] \quad i = 1, 2, 3, 4 \tag{22}$$

where $c_1 = 0.006$, $c_2 = 2e - 8$, $c_3 = 1$. When i is 1, it represents the weighted loss with F_3 and GT, and when i is 2, it represents the weighted loss with F_6 and GT. When i is 3, it represents the weighted loss with F_7 and GT, and when i is 4, it represents the weighted loss with $\hat{I}$ and GT.

In this network, the total loss function is a linear combination of the loss functions of four branches, and the loss function of each branch can be expressed as a linear weighting of the above three loss functions. The formula is expressed as follows:

$$\mathrm{Loss} = \mathrm{loss1} + \mathrm{loss2} + \mathrm{loss3} + \mathrm{loss4} \tag{23}$$

4 Experiments

All experiments are conducted using Pytorch and NVIDIA GeForce RTX 2080 GPU. To increase the training data, we cropped the images into blocks with a resolution of 256 × 256, and set the batch size to 2, the learning rate to 0.0003. The optimizer was performed using Adam, and the total training epochs is 100.

4.1 Datasets

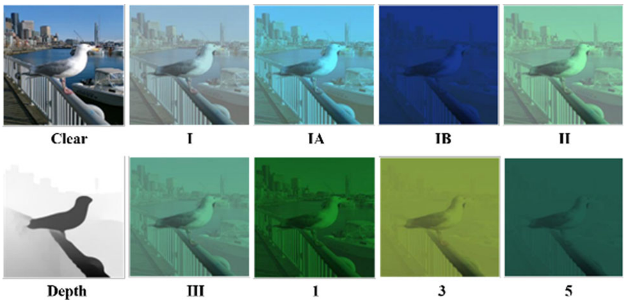

Fig. 3 Eight types of synthesized underwater images (Color figure online)

Inspired by Li et al. [15], we synthesize 400 pairs with eight types of underwater images according to the atmospheric light scattering model, simulating the real-world underwater images. As shown in Fig. 3, I is light blue, IA is medium blue, IB is dark blue, II is light green, III is medium green, 1 is dark green, 3 is light yellow, and 5 is dark black. It is necessary to estimate the global atmospheric light value, and then modify the depth to 50 m to adapt to different types of underwater images by changing the values of R, G, and B channels. As shown in Table 1.

$$T_\lambda(x) = 10^{-A \cdot d(x)} = \frac{E_\lambda(x, d(x))}{E_\lambda(x, 0)} = N_\lambda(\mathrm{d}(x)) \tag{24}$$

Where $T_\lambda(x)$ is a function of the wavelength and distance $\mathrm{d}(x)$ of the λ light from the scene point x to the camera. $N_\lambda(\mathrm{d}(x))$ denotes the ratio of the energy after the $\mathrm{d}(x)$ distance traveled by the light to the energy when it initially enters the dust particles.

For training, we selected 1200 pairs of images, including 774 pairs of randomly selected images from the UIEB datasets and 426 pairs of synthetic images. As for synthetic images, 400 pairs were synthesized by our method and the remaining 26 pairs were randomly selected in [15]. Then, we used three publicly available datasets for testing, including UIEB-890, Challenging-60, Test-R90, and Color Checker datasets.

4.2 Comparison with state-of-the-art methods

We compare our proposed model with eight state-of-the-art methods, both subjectively and objectively, including three traditional methods (UDCP [13], Two-Step [6], L2UWE [23]), and five deep learning-based methods (Water-Net [16], UWCNN [15], FUnIE-GAN [24], USLN [25], MTUR-Net [26]).

4.2.1 Visual comparisons

To verify the superiority of the proposed method, subjective visual effects were used to evaluate our method. As shown in Fig. 4a, the input images were selected from the Test-R90, UIEB-890, and Challenging-60 datasets, all the input images have very low visibility, contrast and severe color distortion.

Table 1 Eight kinds of synthesized images corresponding to different N_λ

Types	I	IA	IB	II	III	1	3	5
Red	0.890	0.780	0.710	0.800	0.820	0.700	0.710	0.850
Green	0.921	0.875	0.880	0.925	0.850	0.780	0.820	0.930
Blue	0.922	0.835	0.750	0.940	0.710	0.880	0.800	0.880

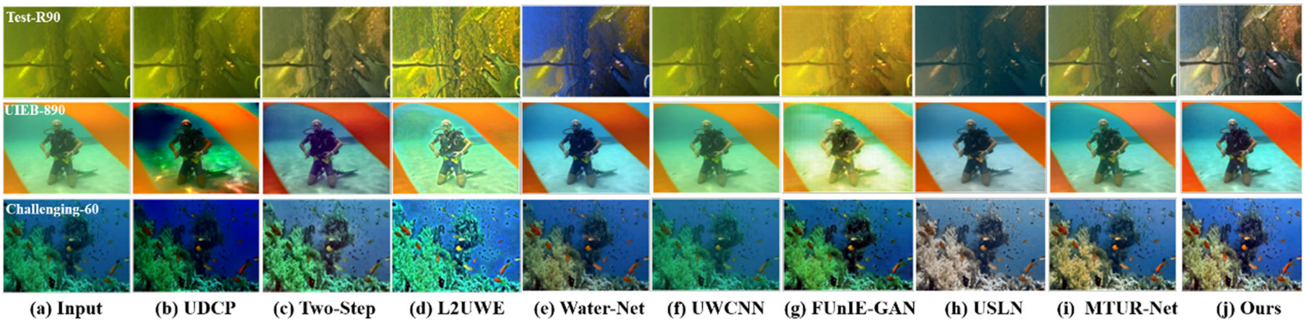

Fig. 4 Subjective comparisons on Test-R90, UIEB-890, and Challenging-60 datasets

The effectiveness of the proposed Multi-SPNet is verified by comparing it with the other eight state-of-the-art methods, which are shown from Fig. 4b–j.

As for traditional methods, UDCP [13] and two-step [6] cannot remove the color distortion of the underwater images, and the results even brings additional over-enhanced artifacts, as shown in Fig. 4b and c, while L2UWE [23] is mainly appropriate for low light scenes of underwater images, it is not effective in recovering those degraded images with severe color distortion and very hazy scenes, as shown in Fig. 4d.

As to deep learning-based methods, the results of Water-Net [16] show improvement in terms of color balance, but the enhanced images have low contrast issues. The other competing methods introduce additional artifacts, unexpected colors, and low brightness (e.g., UWCNN [15] and FUnIE-GAN [24], USLN [25], and MTUR-Net [26]), while our method can effectively remit color casts and remove the haze on the underwater images as shown in Fig. 4j.

To further demonstrate the ability of color casts removal of our method, we used Color Checker dataset, containing seven underwater color images taken by different cameras, and compared with the other five deep learning-based methods for fairness. As can be seen from Fig. 5, all visual comparisons show that our method provides a visually pleasing effect with natural appearance, genuine color and more details than other five competing methods.

4.2.2 Quantitative comparisons

As for quantitative comparisons, we employ full-reference and non-reference evaluations, as well as inference speed and parameters of the models to compare and discuss the performance of different methods.

PSNR and SSIM metrics are usually adopted as full-reference evaluation, a higher score of PSNR and SSIM means a better result. In addition, we also use UCIQE and entropy metrics as non-reference evaluations. It is worth noting that the higher the UCIQE score, the better the balance between the standard deviation of chromaticity, contrast and saturation averages. While the higher the entropy value, the more information the image contains.

In Table 2, we provide the results of eight different methods on Test-R90 dataset, and it can be found that our method achieves the highest PSNR and SSIM scores. Meanwhile, we also test the proposed model on UIEB-890 dataset, as shown in Table 3, the best scores with PSNR and SSIM are achieved. This indicates that our model has a significant improvement on performance compared to the state-of-the-art methods. In the next step, our proposed model obtains the best scores in terms of UCIQE and Entropy on Challenging-60 dataset, which is shown in Table 4.

Moreover, we compare our method with the deep learning-based methods in terms of the model parameters as well as the inference time as shown in Table 5. Although UWCNN has an advantage in terms of parameters, both performance and generalization ability of UWCNN is poor. In contrast, the proposed Multi-SPNet ranks the second in terms of model parameters and inference time is less than 1 s, which demonstrates the balance performance of Multi-SPNet.

4.2.3 Ablation study

To verify the effectiveness of each branch, an ablation study is performed on Test-R90 datasets by removing multi-scale branch. The results are shown in Table 6 and Fig. 6, where we can observe that the resulting images without (w/o) MPN

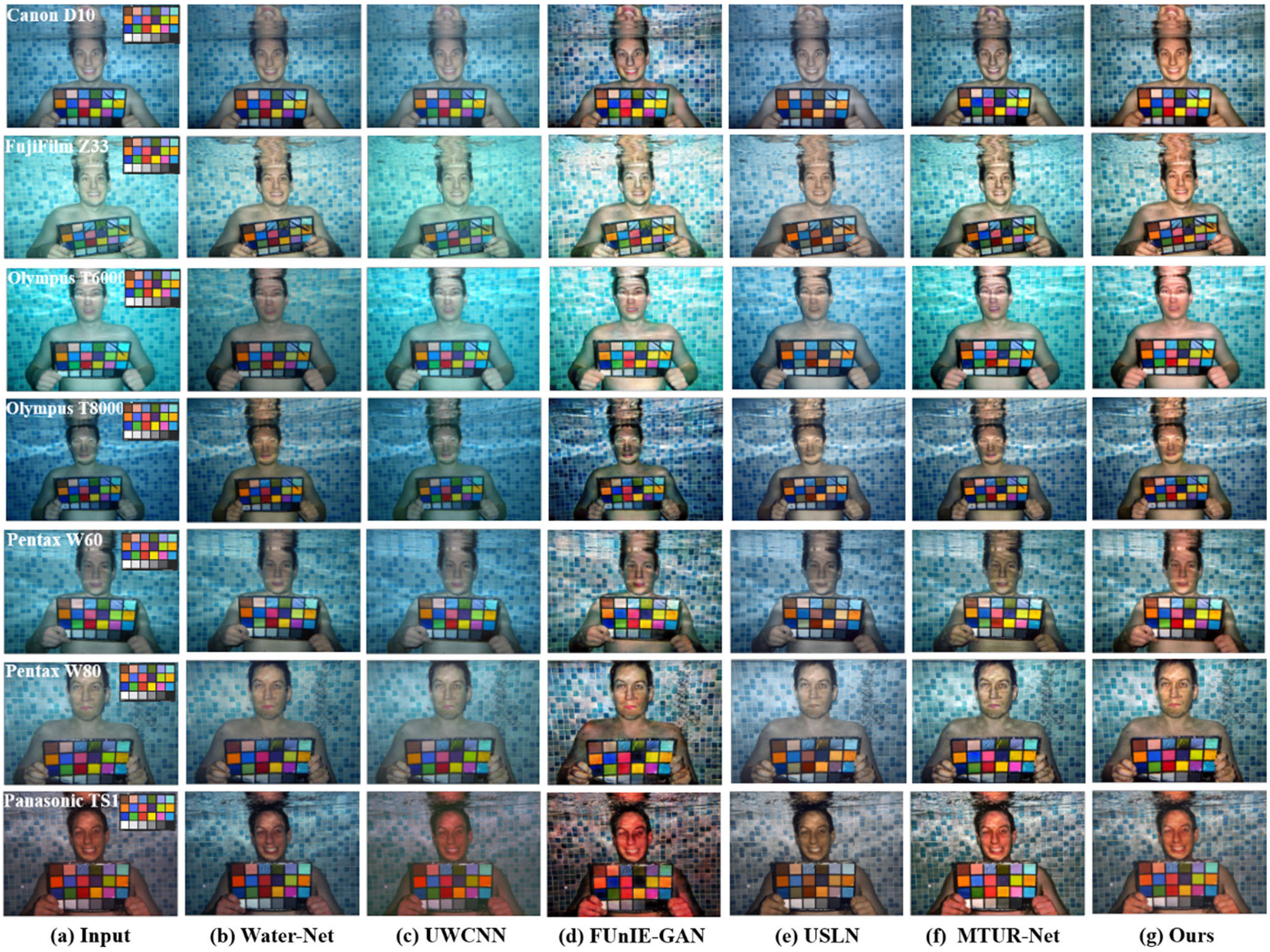

Fig. 5 Subjective comparisons on Color Checker dataset (Color figure online)

Table 2 Quantitative comparisons using full-reference and non-reference metrics on Test-R90 dataset

Methods	Test-R90 dataset			
	PSNR↑	SSIM↑	UCIQE↑	Entropy
UDCP	10.96	0.5148	0.5803	6.3558
Tow-step	19.98	0.8569	0.5706	7.2625
L2UWE	13.83	0.7241	0.5484	7.2628
Water-Net	19.61	0.8290	0.6032	7.3116
UWCNN	16.73	0.7620	0.5143	6.8051
FUnIE-GAN	16.66	0.7617	0.5841	7.2589
USLN	21.00	0.8666	0.5800	7.3224
MUTR-Net	22.36	0.8781	0.6134	7.4864
Our	22.86	0.8947	0.6247	7.5602

Table 3 Quantitative comparisons using full-reference and non-reference metrics on UIEB-890 dataset

Methods	UIEB-890 dataset			
	PSNR↑	SSIM↑	UCIQF↑	Entropy↑
UDC'P	11.25	0.5958	0.5889	6.0180
Two-Step	19.89	0.7123	0.5758	7.2464
L2UWE	13.83	0.8066	0.5501	7.1162
Water-Net	20.24	0.9027	0.5879	7.2613
UWCNN	17.49	0.8492	0.5195	6.8690
FUnlE-GAN	18.01	0.8698	0.5904	7.3132
USLN	21.80	0.9268	0.5820	7.3568
MUTR-Net	23.88	0.9465	0.6080	7.4413
Our	23.96	0.9471	0.6264	7.5695

Table 4 Quantitative comparisons using non-reference metrics on Challenging-60 dataset

Methods	Challenging-60 dataset	
	UCIQE↑	Entropy↑
UDCP	0.5241	5.6842
Two-Step	0.5449	6.9874
L2UWE	0.5326	7.0602
Watcr-Nct	0.5675	7.0416
UWCNN	0.4850	6.4408
FUnlE-GAN	0.5500	6.7900
USLN	0.5636	7.3363
MTUR-Nct	0.5832	7.2752
Our	0.5987	7.3678

Table 5 Comparisons of inference time and parameters

Models	Inference time (s)	Parameters (M)
Water-Net	1.17	24.8
UWCNN	0.04	0.2
FUnIE-GAN	0.09	7.0
Ours	0.95	5.19

Table 6 Ablation study on different branches and loss functions on the Test-R90 dataset

Methods	MSN	MPN	MSE loss	TV loss	Charbonnier loss	PSNR↑	SSIM↑	UCIQE↑	Entropy
1	w/o	✓	✓	✓	✓	22.25	0.8839	0.6182	7.5100
2	✓	w/o	✓	✓	✓	22.35	0.8933	0.6200	7.5427
3	✓	✓	w/o	✓	✓	22.45	0.8915	0.6108	7.5198
4	✓	✓	✓	w/o	✓	22.71	0.8923	0.6131	7.4879
5	✓	✓	✓	✓	w/o	22.78	0.8928	0.6148	7.5155
Multi-SPNet	✓	✓	✓	✓	✓	22.86	0.8947	0.6247	7.5602

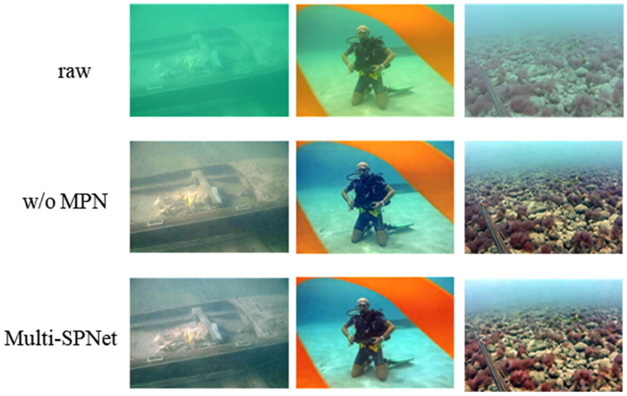

Fig. 6 Ablation study of MPN

have low contrast and color bias. While the results with full model show better color balance and high contrast. Thus, it can be proved that MPN effectively removes color bias and significantly improve contrast.

Table 6 analyzes the effectiveness of two branches and three loss functions quantitatively. It can be seen that different branches including MSN and MPN of network and total loss functions in the proposed method have achieved the highest score, indicating that the proposed Multi-SPNet has the best performance.

The highest score indicating that the proposed Multi-SPNet has the best performance.

5 Conclusion

In this paper, an effective and efficient scale-patch synergy Transformer is proposed for underwater image enhancement. The multi-scale network pays attention to global information and effectively eliminates severe color casts, while the multi-patch network aims to improve the contrast and recover local details. Thus, the proposed Multi-SPNet combines the advantages of multi-scale and multi-patch to obtain better performance on four publicly used datasets both qualitatively and quantitatively. More importantly, our method has fewer parameters than other competitors. In the future, we will try to investigate more effective networks for the challenging underwater image enhancement.

Author contributions L.F. contributed to the study conception, design and wrote original draft. B.W. completed review and editing of abstract, introduction, related works, methods, experiments, conclusion and references. All authors reviewed the manuscript.

Funding This work was supported by the Natural Science Foundation of Ningxia Province (No. 2023AAC03023).

Availability of data and materials The raw data can be shared if the researchers need to do research on relevant topic.

Declarations

Competing interests The authors declare no competing interests.

Ethical approval Not applicable.

References

1. Jaffe, J.S.: Underwater optical imaging: the past, the present, and the prospects. IEEE J. Ocean. Eng. **40**(3), 683–700 (2014)
2. Sheinin, M., Schechner, Y.Y.: The next best underwater view. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 3764–3773 (2016)
3. Lin, W.H., Zhong, J.X., Liu, S., Li, T., Li, G.: Roimix: proposal-fusion among multiple images for underwater object detection. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2588–2592 (2020)
4. Jesus, A., Zito, C., Tortorici, C., Roura, E., De Masi, G.: Underwater object classification and detection: first results and open challenges. In: OCEANS, pp. 1–6 (2022)
5. Akkaynak, D., Treibitz, T.: A revised underwater image formation model. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 6723–6732 (2018)
6. Fu, X., Fan, Z., Ling, M., Huang, Y., Ding, X.: Two-step approach for single underwater image enhancement. In: IEEE international symposium on intelligent signal processing and communication systems (ISPACS), pp. 789–794 (2017)
7. Fu, X., Cao, X.: Underwater image enhancement with global-local networks and compressed-histogram equalization. Signal Process. Image Commun. **86**, 115892 (2020)
8. Jiang, Q., Zhang, Y., Bao, F., Zhao, X., Zhang, C., Liu, P.: Two-step domain adaptation for underwater image enhancement. Pattern Recogn. **122**, 108324 (2022)
9. Yin, S., Hu, S., Wang, Y., Wang, W., Li, C., Yang, Y.H.: Degradation-aware and color-corrected network for underwater image enhancement. Knowl.-Based Syst. **258**, 109997 (2022)
10. Liang, P., Ding, W., Fan, L., Wang, H., Li, Z., Yang, F., Wang, B., Li, C.: Multi-scale and multi-patch transformer for sandstorm image enhancement. J. Vis. Commun. Image Represent. **89**, 103662 (2022)
11. Li, C.Y., Guo, J.C., Cong, R.M., Pang, Y.W., Wang, B.: Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. IEEE Trans. Image Process. **25**(12), 5664–5677 (2016)
12. 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: Pacific Rim Conference on Multimedia (PCM), pp. 678–688 (2018)
13. Drews, P.L., Nascimento, E.R., Botelho, S.S., Campos, M.F.M.: Underwater depth estimation and image restoration based on single images. IEEE Comput. Gr. Appl. **36**(2), 24–35 (2016)
14. Li, J., Skinner, K.A., Eustice, R.M., Johnson-Roberson, M.: 2017 WaterGAN: unsupervised generative network to enable real-time color correction of monocular underwater images. IEEE Robot. Autom. Lett. **3**(1), 387–394 (2017)
15. Li, C., Anwar, S., Porikli, F.: Underwater scene prior inspired deep underwater image and video enhancement. Pattern Recogn. **98**, 107038 (2020)
16. 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. **29**, 4376–4389 (2019)
17. Li, C., Anwar, S., Hou, J., Cong, R., Guo, C., Ren, W.: Underwater image enhancement via medium transmission-guided multi-color

space embedding. IEEE Trans. Image Process. **30**, 4985–5000 (2021)

18. Shen, Z., Xu, H., Luo, T., Song, Y., He, Z.: UDAformer: underwater image enhancement based on dual attention transformer. Comput. Gr. **111**, 77–88 (2023)
19. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE international conference on computer vision (ICCV), pp. 10012–10022 (2021)
20. Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H.: Uformer: a general u-shaped transformer for image restoration. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 17683–17693 (2022)
21. Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M. H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 5728–5739 (2022)
22. Peng, L., Zhu, C., Bian, L.: U-shape transformer for underwater image enhancement. IEEE Trans. Image Process. **32**, 3066–3079 (2023)
23. Marques, T. P., Albu, A. B.: L2uwe: a framework for the efficient enhancement of low-light underwater images using local contrast and multi-scale fusion. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 538–539 (2020)
24. Islam, M.J., Xia, Y., Sattar, J.: Fast underwater image enhancement for improved visual perception. IEEE Robot. Autom. Lett. **5**(2), 3227–3234 (2020)
25. Xiao, Z., Han, Y., Rahardja, S., Ma, Y.: USLN: A statistically guided lightweight network for underwater image enhancement via dual-statistic white balance and multi-color space stretch. arXiv preprint at arXiv:2209.02221. (2022)
26. Yan, K., Liang, L., Zheng, Z., Wang, G., Yang, Y.: Medium transmission map matters for learning to restore real-world underwater images. Appl. Sci. **12**(11), 5420 (2022)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.