

## Research Article

## MCRNet: Underwater image enhancement using multi-color space residual network

Ningwei Qin, Junjun Wu<sup>\*</sup>, Xilin Liu, Zeqin Lin, Zhifeng Wang

School of Mechatronic Engineering and Automation, Foshan University, Foshan 528225, China

## ARTICLE INFO

## Article history:

Received 6 May 2024

Revised 24 May 2024

Accepted 4 June 2024

Available online 17 June 2024

## Keywords:

Underwater image enhancement

Deep learning

Color correction

Underwater robots

## ABSTRACT

The selective attenuation and scattering of light in underwater environments cause color distortion and contrast reduction in underwater images, which can impede the ever-growing demand for underwater robot operations. To address these issues, we propose a Multi-Color space Residual Network (MCRNet) for underwater image enhancement. Our method takes advantage of the unique features of color representation in the RGB, HSV, and Lab color spaces. By utilizing the distinct feature representations of images in different color spaces, we can highlight and fuse the most informative features of the three color spaces. Our approach employs a self-attention mechanism in the multi-color space feature fusion module. Extensive experiments demonstrate that our method achieves satisfactory results in color correction and contrast improvement of underwater images, particularly in severely degraded scenes. Consequently, our method outperforms state-of-the-art methods in both subjective visual comparison and objective evaluation metrics.

© 2024 The Author(s). Published by Elsevier B.V. on behalf of Shandong University. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In recent years, the depletion of land resources has led to a greater emphasis on the development of marine resources. Underwater robots are increasingly required in various fields such as aquaculture, submarine cable and pipeline laying, offshore facility inspection, and underwater scientific research. However, the underwater environment is more complex than the terrestrial environment, presenting challenges for underwater operations. For instance, the water's different attenuation rates for various wavelengths of light result in underwater images often suffering from greenish or bluish color distortion. Additionally, suspended particles in seawater cause forward and backward scattering, resulting in low contrast and hazing issues in underwater images. These factors hinder underwater robots from achieving high-quality underwater images, which limits their ability to perform high-level underwater vision tasks, such as object detection [1,2], semantic segmentation [3], and image classification [4], many approaches aim to enhance the performance of simultaneous localization and mapping (SLAM) for robots [5,6]. As a low-level underwater vision task, underwater image enhancement can overcome the problem of limited and unclear vision for underwater robots, and it plays an important role in improving the performance of high-level underwater vision tasks [7].

Early underwater image enhancement techniques relied on conventional methods categorized into two types: image formation model-based (IFM-based) methods and IFM-free methods. Nevertheless, the intricate underwater environment limits the efficacy of a single imaging physical model or prior knowledge-based methods for modifying the distribution of pixel values in certain degraded scenarios. Consequently, such approaches may cause under-/over-enhancement or introduce additional artifacts.

Deep learning has become increasingly important in low-level vision tasks, including image super resolution [8,9], image denoising [10], image deraining [11] and other fields. While deep learning is also making strides in underwater image enhancement tasks [12], current methods have limited ability to eliminate color distortion and improve contrast in certain scenes, and their generalization ability requires improvement. This is because most methods solely rely on the RGB color space and overlook the rich feature representation available in the HSV and Lab color spaces. To address these challenges, we propose a novel solution called the Multi-Color Space Residual Network (MCRNet) for underwater image enhancement. The key contributions of this paper are as follows:

- We introduce the MCRNet for enhancing underwater images. Unlike most methods that work only with RGB color space, MCRNet leverages multiple color spaces to fuse feature representations, resulting in better color correction and contrast improvement of underwater images.
- To overcome the limitations of traditional fusion methods like concatenation and summation, we propose the Multi-Color Space Feature Fusion (MCFF) module. MCFF employs a

<sup>\*</sup> Corresponding author.

E-mail address: [jjunwu@fosu.edu.cn](mailto:jjunwu@fosu.edu.cn) (J. Wu).

self-attention mechanism to selectively highlight the most informative feature representations in each color space and fuse them.

- Our approach is evaluated through both qualitative visual analysis and quantitative metrics on multiple datasets. Our results demonstrate that MCRNet outperforms state-of-the-art underwater image enhancement methods and exhibits excellent color correction ability even in extremely degraded scenes.

## 2. Related work

Underwater image enhancement can be divided into IFM-based restoration methods, IFM-free enhancement methods and deep learning-based methods.

**IFM-based Methods:** This kind of method regards underwater image restoration as a reversible problem, and solves the problem by establishing a physical model based on prior knowledge. Given that the formation model of underwater images is comparable to that of atmospheric images, many researchers are inspired by the dark channel prior (DCP) [13]. For example, Drews et al. [14] proposed the underwater dark channel prior (UDCP) according to the characteristics that the red channel attenuates fastest in the underwater environment. Galdran et al. [15] proposed the red dark channel prior (RDCP) based on the blue channel, green channel and inverse red channel. Peng et al. [16] proposed a depth estimation method for underwater images based on image blurriness and light absorption (IBLA). Akkaynak et al. [17] proposed a modified model named Sea-thru to obtain high-quality underwater images, but it requires much additional information. Song et al. [18] obtained the scene depth estimation model based on the underwater light attenuation prior (ULAP), and further enhanced the underwater image according to the depth map. Recently, Zhou et al. [19] used backscatter pixel prior to remove backscatter and performed color compensation to remove red artifacts in underwater images. Since the complexity of the underwater environment, there is no single physical model that can cover all degradation scenarios, which may result in noticeable noise or artifacts in some cases.

**IFM-free Methods:** Such methods produce visually appealing underwater images by selectively modifying pixel values. One line of work is based on the fusion strategy. Ancuti et al. [20] performed the underwater image enhancement by fusing underwater images after color correction and contrast enhancement. Subsequently, Ancuti et al. [21] further used a multi-scale fusion strategy of underwater images based on color compensation and an improved white balance algorithm. Another line of work is based on Retinex theory. Fu et al. [22] used Retinex algorithm to decompose the luminance layer to solve the problem of underexposure and fuzzy in underwater images. Zhuang et al. [23] proposed the Bayesian retinex algorithm with multiorder gradient priors to enhance the underwater images. The other direction is based on the minimum loss principle. Li et al. [24] used the minimum information loss principle to dehaze the underwater image and improved the contrast by adjusting the histogram distribution. Zhang et al. [25] used minimum color loss and locally adaptive contrast enhancement for underwater image color correction. Although enhancement methods are generally faster than model-based approaches, they do not take into account the imaging mechanism of underwater images, and may result in under-/over-enhancement.

**Deep learning-based Methods:** In recent years, researchers have increasingly applied deep learning to the underwater image enhancement task [26–28], thanks in part to the growing computing power of GPUs and the popularity of deep learning techniques. Fabbri et al. [29] proposed UGAN, which included a generator based on the U-Net architecture and a discriminator

modeled as the patchGAN. Li et al. [30] proposed a dataset called UIEB and used it to train the Water-Net, which is based on a gated fusion network. Later, Li et al. [31] proposed UWCNN which has 10 different models that are trained on different degraded scenes. Liu et al. [7] proposed an object-guided twin adversarial contrastive learning network to achieve task-orientated enhancement. Recently, Shen et al. [32] proposed a dual attention transformer-based method to enhance underwater images. Deep learning-based approaches rely on real, accurate, and extensive image datasets, which are typically easy to obtain for tasks like super-resolution and denoising. However, the unique characteristics of the underwater environment make it challenging to obtain reference images of underwater scenes. Currently, most underwater image datasets with reference images are generated by existing underwater image enhancement methods and then manually selecting reference images [30,33]. This approach often leads to datasets with uneven reference image quality.

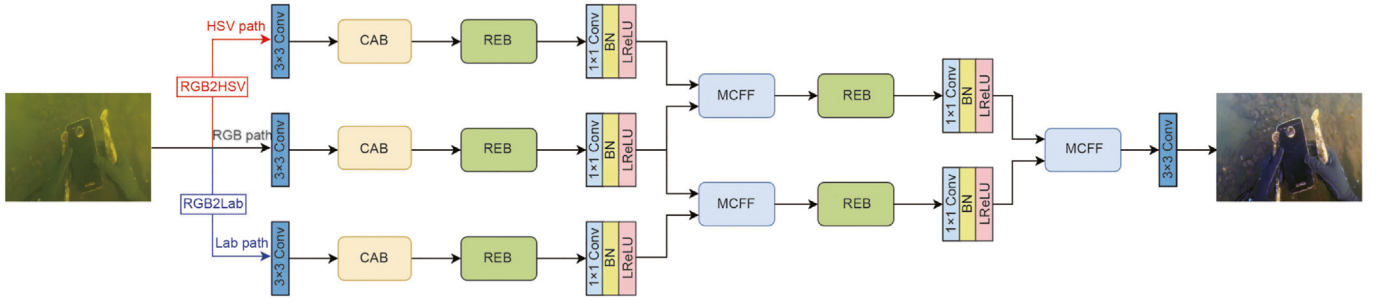
**Application of multi-color space in underwater image enhancement:** Numerous methods for enhancing underwater images utilize the multi-color space mechanism. Iqbal et al. [34] performed contrast correction and color balance on RGB and HSI color spaces and proposed a unsupervised color correction method. Deperlioglu et al. [35] proposed HSV color space with the extension of V component and histogram equalization method to enhance underwater images. Hu et al. [36] decoupled HS and V channels and passed them to the color cast removal branch and contrast enhancement branch respectively, then merged the two branches and converted them back to RGB color space. Lyu et al. [37] proposed a network that first enhances the input through CNN-based residual groups, and then converts the result to the YUV color space for post-processing. Li et al. [38] through multi-color space embedding to extract multiple features and then guide the network through medium transmission to enhance attention and proposed Ucolor. However, Ucolor is not an end-to-end method. It needs to generate the transmission map as input using the general dark channel prior [39], which can be time-consuming.

## 3. Proposed method

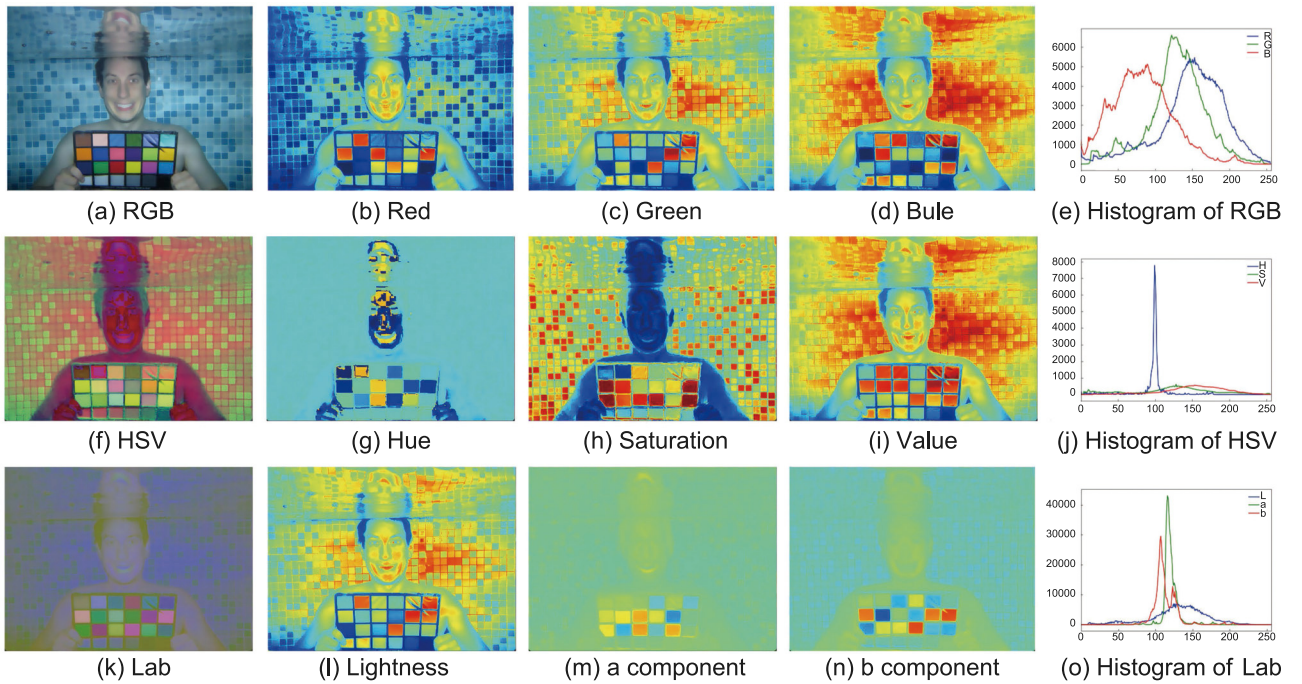
Fig. 1 illustrates MCRNet, an end-to-end network that consists of three color space branches that does not require any down-sampling operation. After color space conversion, the original underwater image is divided into the HSV, RGB, and Lab paths. Next, we apply the channel attention block [40] and the residual enhancement block, followed by a  $1 \times 1$  convolution layer, batch normalization, and Leaky ReLU activation on each path to extract the most representative features. After that, we aggregate the features from the HSV and Lab paths with the RGB path using the multi-color space feature fusion module. We then pass the two aggregated features through the REB, a  $1 \times 1$  convolution layer, batch normalization, and Leaky ReLU activation before fusing them with MCFF. Finally, we extract deep features using a  $3 \times 3$  convolution layer to obtain the enhanced underwater image.

### 3.1. Multi-color space residual network

Underwater images often suffer from color distortion and low contrast due to the selective attenuation of light in aquatic environments. To address these issues, we drew inspiration from traditional image enhancement techniques that rely on multi-color spaces [35]. We extracted feature representations from HSV, RGB, and Lab color spaces to better correct the color deviation in underwater images. As depicted in Fig. 2, the same underwater image can have different visual representations in different color spaces and their components. RGB color space is convenient



**Fig. 1.** Overview of the proposed MCRNet. MCRNet comprises three different paths: the HSV path, the RGB path, and the Lab path. Within each path, we utilize a channel attention block (CAB) and a residual enhancement block (REB) for feature extraction. These features from different color spaces are then fused using the multi-color space feature fusion module (MCFE).



**Fig. 2.** Visual examples of underwater images in different color spaces and their components. For better visualization, (b)–(d), (g)–(i), and (l)–(n) are represented by heatmaps. The image is from [21].

for hardware implementation and commonly used in computer vision. However, its three components are all used to represent the color, therefore, if the value of one component is changed, the color of the pixel will be changed accordingly. In contrast, HSV color space is closer to the human visual system, and its three components can intuitively express the hue, saturation and brightness of an image. Lab color space is designed based on human color perception and offers uniform perception and a wide color gamut. By combining feature representations from these three color spaces, MCRNet can learn the degradation mechanism of underwater images effectively and achieve satisfactory color correction results. In the following sections, we will describe each module in detail.

### 3.1.1. Multi-color space feature fusion

The most common ways of multi-color space fusion are typically achieved through simple concatenation or summation, but these fusion approaches cannot provide expressive characteristics to the network. In the case of color correction of underwater images, each color space’s features make an independent contribution. To fuse these features from different color spaces, inspired

by [41], we introduce nonlinear factors into the feature fusion from different color spaces using the self-attention mechanism, which is called multi-color space feature fusion (MCFE).

Fig. 3 shows the details of the MCFE module, which adaptively selects the features of the most informative color space by the Fuse and the Select operations. **Fuse:** MCFE first combines the features from two different color space branches using the element-wise summation operation expressed as:  $C = C_1 + C_2$ . Next, global average pooling (GAP) is used to embed global information and generate channel-wise statistics  $s$ . We then use a  $1 \times 1$  convolution layer to reduce channel, followed by a  $3 \times 3$  depthwise convolution to emphasize local context information, and activated with the GELU non-linearity to produce a compact feature  $z$ . Finally, two global feature descriptors  $a, b$  are obtained by passing the feature vectors  $z$  to two parallel  $1 \times 1$  convolution layers corresponding to two color space branches for channel-upscaling. **Select:** Feature descriptors  $a$  and  $b$  are passed through the softmax function to generate attention activations  $s_1$  and  $s_2$  that we use to adaptively recalibrate the feature map and from two color space branches. The final output feature map is defined as:  $V = C_1 * s_1 + C_2 * s_2$ .

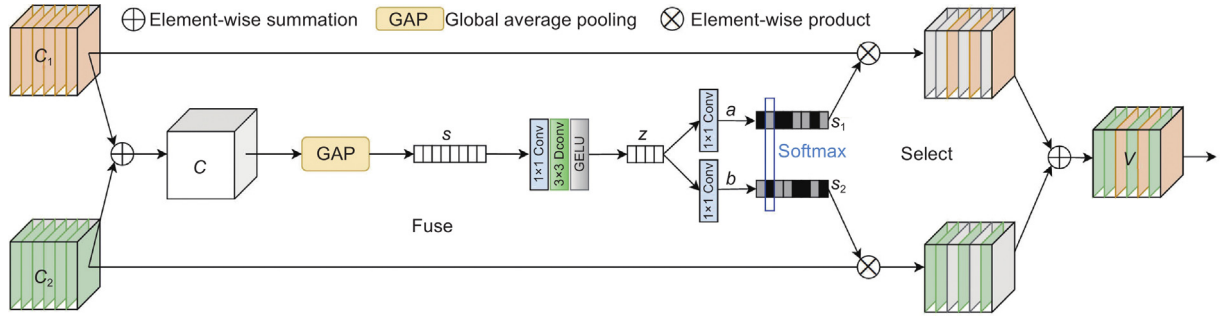


Fig. 3. Multi-color space feature fusion module.

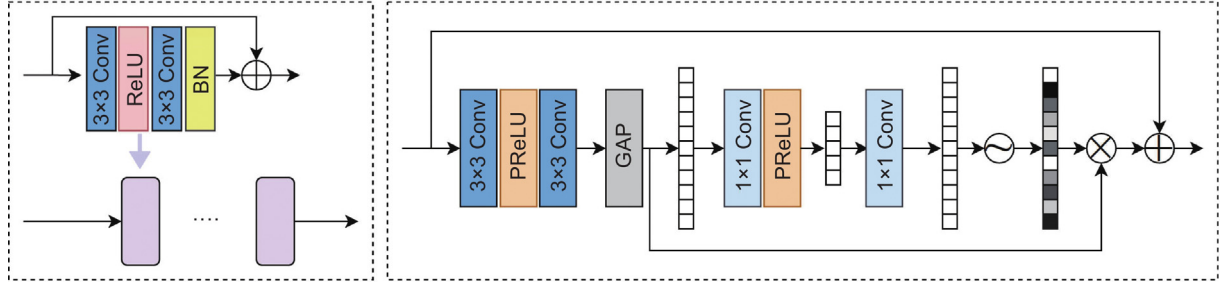


Fig. 4. (a) Residual enhancement block. (b) Channel attention block.

### 3.1.2. Residual enhancement block

As shown in Fig. 4(a), the residual enhancement block is composed of several residual subblocks. The residual network can solve the problems of performance degradation and gradient vanishing of deep neural networks, and further improves the feature extraction ability of the network [42,43]. Each residual subblock in the residual enhancement block consists of a convolution layer and a ReLU activation, followed by a convolution layer and batch normalization. Finally, a skip connection is applied. For our experiment, we set the kernel size of all convolution layers to  $3 \times 3$  with a stride of 1. Additionally, we included 4 residual subblocks in each REB. As the number of layers in the network increases, the receptive fields expand, allowing for better capture of global features in underwater images, such as overall color distortion. Since our network does not use any downsampling operation, it retains the fine details of the image.

### 3.2. Loss function

To train MCRNet, we use the following loss function to train our model. We use  $\hat{I}$  to represent the enhanced image, and  $I^*$  to represent the corresponding ground truth image.

**Charbonnier Loss:** Charbonnier Loss [44] is similar to L1 loss, which is used to calculate the distance between the enhanced image and the corresponding ground truth image.

$$\mathcal{L}_{char} = \sqrt{\|\hat{I} - I^*\|^2 + \varepsilon^2} \quad (1)$$

**SSIM Loss:** Structural SIMilarity Index (SSIM) compares luminance, contrast, and structure between two images. The formula of SSIM is as follows:

$$SSIM(\hat{I}, I^*) = \frac{(2\mu_{\hat{I}}\mu_{I^*} + c_1)(2\sigma_{\hat{I}I^*} + c_2)}{(\mu_{\hat{I}}^2 + \mu_{I^*}^2 + c_1)(\sigma_{\hat{I}}^2 + \sigma_{I^*}^2 + c_2)} \quad (2)$$

Where  $\mu_{\hat{I}}$  and  $\mu_{I^*}$  denote the mean of the enhanced image and the corresponding ground truth image respectively,  $\sigma_{\hat{I}}^2$  and  $\sigma_{I^*}^2$  represent the variance of the two images, and  $\sigma_{\hat{I}I^*}$  is the covariance.  $c_1$  and  $c_2$  are two constants to ensure stability in case of the

denominator is 0. We set  $c_1 = (255 \times 0.01)^2$ ,  $c_2 = (255 \times 0.03)^2$ . Then, the SSIM loss can be expressed as

$$\mathcal{L}_{SSIM} = 1 - SSIM(\hat{I}, I^*) \quad (3)$$

**Edge Loss:** Similar to the image super-resolution task, the high-frequency edge details of the underwater image will be lost due to various degradation problems in the underwater environment. Therefore, we use edge loss [45] to optimize the edge information of underwater images:

$$\mathcal{L}_{edge} = \sqrt{\|\Delta(\hat{I}) - \Delta(I^*)\|^2 + \varepsilon^2} \quad (4)$$

where the  $\varepsilon$  in Charbonnier Loss and Edge Loss is empirically set to  $10^{-3}$ .

**Perceptual loss:** Perception loss uses the features extracted from the VGG network [46] pre-trained on the ImageNet dataset [47] to supervise the features generated by the model to make it have high-level perception information. We use the conv5\_4 layer in the VGG19 network to calculate the perception loss:

$$\mathcal{L}_{per} = \|\Phi(\hat{I}) - \Phi(I^*)\|_2 \quad (5)$$

Where  $\Phi$  denotes the high-level feature extracted from the conv5\_4 layer in the pre-trained VGG19 network.

**Total Loss:** We get the total loss function by the linear combination of the three loss components:

$$\mathcal{L}_{total} = \mathcal{L}_{char} + \mathcal{L}_{SSIM} + \lambda_e \mathcal{L}_{edge} + \lambda_{perc} \mathcal{L}_{perc} \quad (6)$$

Where we set  $\lambda_e = \lambda_{perc} = 0.05$  by experience.

### 3.3. Progressive learning strategy

The existing underwater image enhancement methods typically randomly crop or resize input images to a fixed patch size for training. While small patches focus on capturing local details like edges and textures and can speed up training, they may sacrifice the global features of the image. On the other hand, larger patches better learn global features like color deviation

and haze effects but require more training time and memory resources. To address these tradeoffs, we propose a progressive learning strategy that better handle the texture details and color correction of underwater images, and balance the performance and training time of the model.

Specifically, during the early epochs, our model is trained on small patches using a large batch size. Later, the patch size gradually increases, and the batch size decreases. This approach effectively speeds up training and balances the local and global features of underwater images. Compared to continuous training on large patches, our strategy reduces training time with almost no performance drop. Additionally, our mixed-size training strategy leads to better generalization when testing on images of different sizes.

## 4. Experiments

In this section, we first introduce the implementation details and experimental settings, and then we conduct comprehensive experiments on our method with representative traditional and deep learning-based underwater image enhancement methods. We also provide two application tests of underwater feature point detection and underwater edge detection. Finally, the ablation study was performed to verify the effectiveness of each module in our approach.

### 4.1. Implementation details

For training, we use 800 pairs of real-world underwater images from the UIEB dataset [30], and since real underwater reference images could not be obtained, we augmented our dataset with 1250 pairs of synthetic images from the NYU-v2 RGB-D dataset [46], which simulate ten different underwater degraded scenes. That is total of 2050 pairs of images were used to train our model.

Our implementation of MCRNet was done using the PyTorch framework on the NVIDIA GeForce RTX 3090Ti platform. We employed the Adam optimizer for 150K iterations at a fixed learning rate of  $1e-4$ . For progressive learning, we use image patch sizes of 256, 320, 384, 448 with batch sizes of 16, 8, 4, 2, respectively, and we use vertical and horizontal flips for data augmentation.

### 4.2. Experiment settings

**Compared Methods.** We compared MCRNet with nine methods, including three IFM-based methods (UDCP [14], IBLA [16], ULAP [18]), one IFM-free methods (Fusion18 [21]), and six deep learning-based methods (Water-Net [30], UWCNN [31], FUnIE-GAN [48], Ucolor [38], PUIE\_MP [49] and NU<sup>2</sup>Net [50]). For all deep learning-based methods, we adopt the pre-trained models provided by the corresponding authors to generate the best objective evaluation results. Among them, UWCNN has 10 different pre-trained models according to different underwater scenes, we only chose type-1 for comparison, so it will show poor generalization in some scenes, which also limits the method's applicability in actual underwater operations. In contrast, our method has only one pre-trained model but performs well in color correction of underwater images in various scenes.

**Benchmarks.** For testing, we use the remaining 90 pairs of real underwater images with reference images in the UIEB dataset [30], denote as Test-U90. We also conduct experiments on two non-reference test datasets: Test-C60 [30] and Test-SQUID [51]. Test-C60 is a challenging dataset that includes 60 underwater images with no ground truth images from the UIEB dataset, which cannot be fully enhanced by existing algorithms.

The Test-SQUID dataset selects 16 representative images from [51], featuring four different diving sites.

**Evaluation Metrics.** To assess the quality of non-reference underwater image datasets, we use three specific evaluation metrics: UIQM [52], UCIQE [53], and Twice Mixing [54].

- UIQM comprises three measures: colorfulness (UICM), sharpness (UISM), and contrast (UIConM). A higher UIQM indicates that an image is closer to the human visual system.
- UCIQE calculates the standard deviation of chroma and contrast in the CIE Lab color space, as well as the mean value of saturation in the HSV color space. These three components are then combined linearly to obtain the final result. A higher UCIQE score indicates better enhancement effects of underwater images.

However, both UIQM and UCIQE are not always accurate [19, 30]. UIQM tends to give high scores to images with red artifacts due to over-enhancement, while UCIQE prefers vivid images with high contrast.

- Twice Mixing (TM) is a self-supervised underwater image evaluation method based on deep learning. It generates two random mixing ratios to create input data and guide network learning. A higher TM score indicates greater alignment with subjective judgments.

For the full reference test dataset Test-U90, we additionally use Peak Signal to Noise Ratio (PSNR) and Structural SIMilarity index (SSIM) to evaluate the quality of images generated by each method.

- PSNR score is directly proportional to the correlation of pixel content between the ground truth images and the generated results.
- SSIM measures the similarity of the generated results with the reference image in terms of structure, contrast, and luminance. It is more consistent with the human visual system than PSNR [55].

### 4.3. Experiments on underwater datasets

For the real underwater datasets, we first tested on the Test-U90 with reference images. We selected 8 representative images depicting 4 different underwater degradation scenes, which present in Fig. 5(a). From top to bottom, these scenes include shallow water, low illumination, blue distortion, and green distortion. In the shallow water images, both Water-Net and Ucolor failed to completely remove the hazy effect, while Fusion18 and MCRNet effectively improve contrast and correct color in both images. For the low-illuminated images, FUnIE-GAN introduced red artifacts, but Fusion18, Ucolor and MCRNet all improved contrast, with MCRNet achieving the most significant improvement in brightness. It is worth noting that the reference images for this scene were overexposed, resulting in the loss of texture details on the shark. On the other hand, Fusion18, Ucolor, and MCRNet all successfully removed the blue distortion, among which Fusion18 and MCRNet achieving a more thorough removal. MCRNet restored color better than Fusion18 and achieved better color correction on coral in the sixth image, which had a greenish tone in the reference image. In the green distorted images, Fusion18 shows purple artifacts on the background. PUIE\_MP exhibits excellent color correction effects, but the overall results tend to be pale, lacking in contrast, and unable to achieve a satisfactory dehazing result. MCRNet and NU<sup>2</sup>Net have clearer visual effects than the reference image. The overall results of

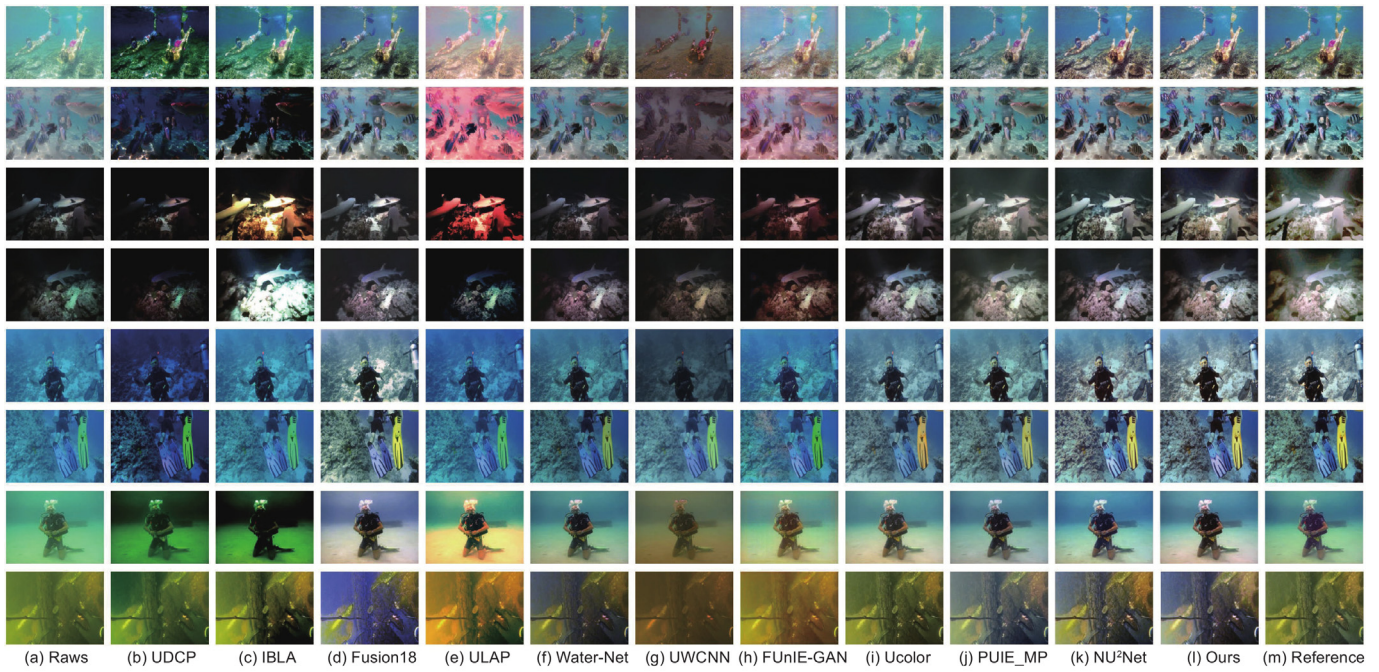


Fig. 5. Visual comparison of different methods on real underwater dataset Test-U90. It is suggested to zoom in for a better view.

**Table 1**  
Full-reference metrics in terms of SSIM and PSNR and non-reference metrics in terms of UCIQE, UIQM and Twice Mixing (TM) for each method on Test-U90.

Methods	Params.↓	SSIM	PSNR	UCIQE	UIQM	TM
Raws	-	0.750	16.111	0.543	2.336	0.743
UDCP [14]	-	0.472	10.493	0.574	1.632	0.613
IBLA [16]	-	0.652	15.100	0.609	1.696	0.955
Fusion18 [21]	-	0.878	20.663	0.593	2.808	1.127
ULAP [18]	-	0.720	15.049	<b>0.613</b>	1.919	0.969
Water-Net [30]	25M	0.858	19.595	0.587	2.829	0.995
UWCNN [31]	39.97M	0.665	13.433	0.529	2.489	0.453
FUNIE-GAN [48]	7.02M	0.747	16.769	0.575	<b>3.035</b>	0.992
Ucolor [38]	157.4M	0.868	20.768	0.583	2.864	1.003
PUIE_MP [49]	1.4M	0.879	21.281	0.575	2.844	1.044
NU <sup>2</sup> Net [50]	3.15M	<b>0.892</b>	<b>22.793</b>	0.602	<b>2.894</b>	<b>1.163</b>
<b>Ours</b>	1.49M	<b>0.918</b>	<b>22.755</b>	<b>0.620</b>	2.880	<b>1.305</b>
<b>Reference</b>	-	-	-	<b>0.623</b>	<b>2.811</b>	<b>1.296</b>

Water-Net are relatively dark, and the results of Ucolor are hazy. NU<sup>2</sup>Net introduces additional artifacts in the second and last images. MCRNet exhibits the best generalization ability in all results, demonstrating excellent color restoration and contrast enhancement capabilities.

Table 1 reports the average quantitative metrics of different methods in the Test-U90 dataset. We highlight the best performance in red and the second best in blue. For full-reference metrics SSIM and PSNR, PUIE can significantly enhance image contrast and correct most color casts, but its results generally lack saturation, leading to its third-place ranking in PSNR and SSIM scores. NU<sup>2</sup>Net provides stable enhancement results, but it may introduce additional artifacts, leading to the second-best SSIM and best PSNR performance. MCRNet ranks first in SSIM and second in PSNR for its excellent color correction performance and generalization. In terms of non-reference metrics, the proposed method ranks first in UCIQE and Twice Mixing. FUNIE-GAN has achieved first place in UIQM, even surpassing the reference images. However, as mentioned in Section 4.2, UIQM prefers those images with over-enhancement effects in some cases.

Next, we will test on more challenging underwater datasets. We select 4 challenging images from Test-C60 shown in Fig. 6(a).

None of the comparison methods were able to achieve satisfactory results on all of the images. Many methods even introduce artifacts, resulting in a worse visual perception than the original inputs, such as UDCP, IBLA, ULAP, UWCNN and FUNIE-GAN. In Fig. 6(d), Fusion18 achieves good enhancement results on the first two images, but the last severely green distorted images introduce purple artifacts. Water-Net can remove most of the severe greenish color deviation, as shown in the last image in Fig. 6(f). However, its results lack contrast, the overall brightness of the image is dark, and the enhancement effect on low-illuminated underwater images is not obvious. Ucolor improve the visual effect in all images except the first shallow water image where red artifacts were introduced. NU<sup>2</sup>Net performs well in removing background color casts, it mistakenly restores colors on second and third images. MCRNet performs color correction and improved contrast in every challenging image, showing great generalization.

In Fig. 7, we present the results of each method in the Test-SQUID dataset. It is evident that the majority of methods struggle to eliminate the blue distortion. While Fusion18 successfully removed the blue distortion, they produce purple artifacts in the background, Water-Net introduce green artifacts. NU<sup>2</sup>Net excessively enhanced the colors of the coral in the first image. MCRNet produces satisfactory results by effectively removing the blue distortion while preserving more image details and generating cleaner and brighter results.

Table 2 reports the average quantitative metrics of each method in Test-C60 and Test-SQUID datasets. Notably, UDCP ranks first UCIQE ranking on the Test-SQUID, but as shown in Fig. 7(b), its results exhibit significantly lower visual quality than the input images. For the Twice Fixing, our method achieves the highest score on both datasets. These findings highlight the challenges of Underwater Image Enhancement Image Quality Assessment, which still requires further improvement.

Then we verify the accuracy of each method in color restoration. As can be seen from Fig. 8, although FUNIE-GAN removed the blue artifact, they also introduced color deviation. Fusion18,

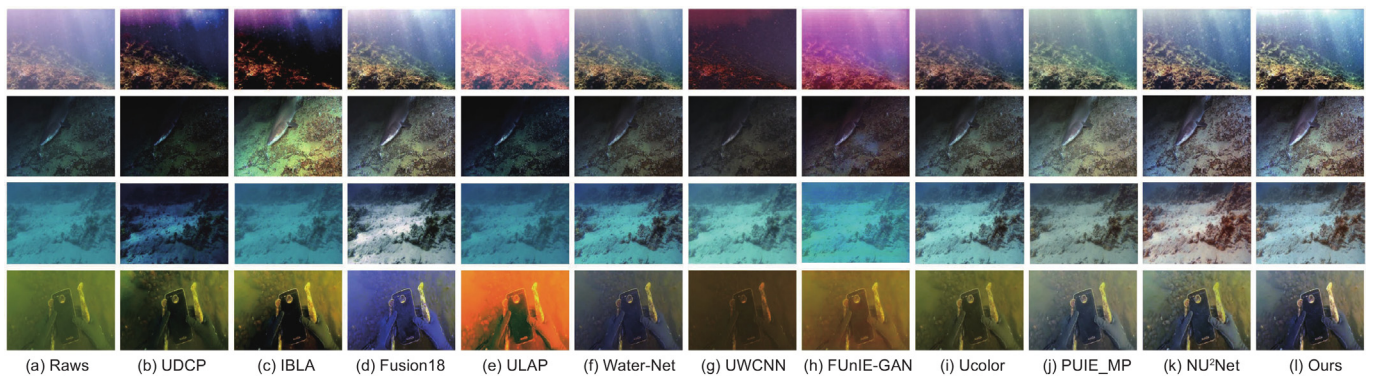


Fig. 6. Visual comparison of different methods on real underwater dataset Test-C60. It is suggested to zoom in for a better view.

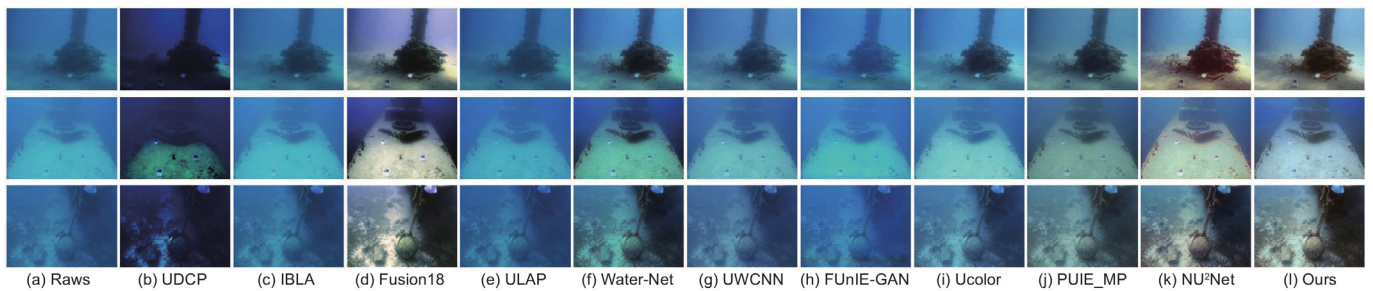


Fig. 7. Visual comparison of different methods on real underwater dataset Test-SQUID. It is suggested to zoom in for a better view.

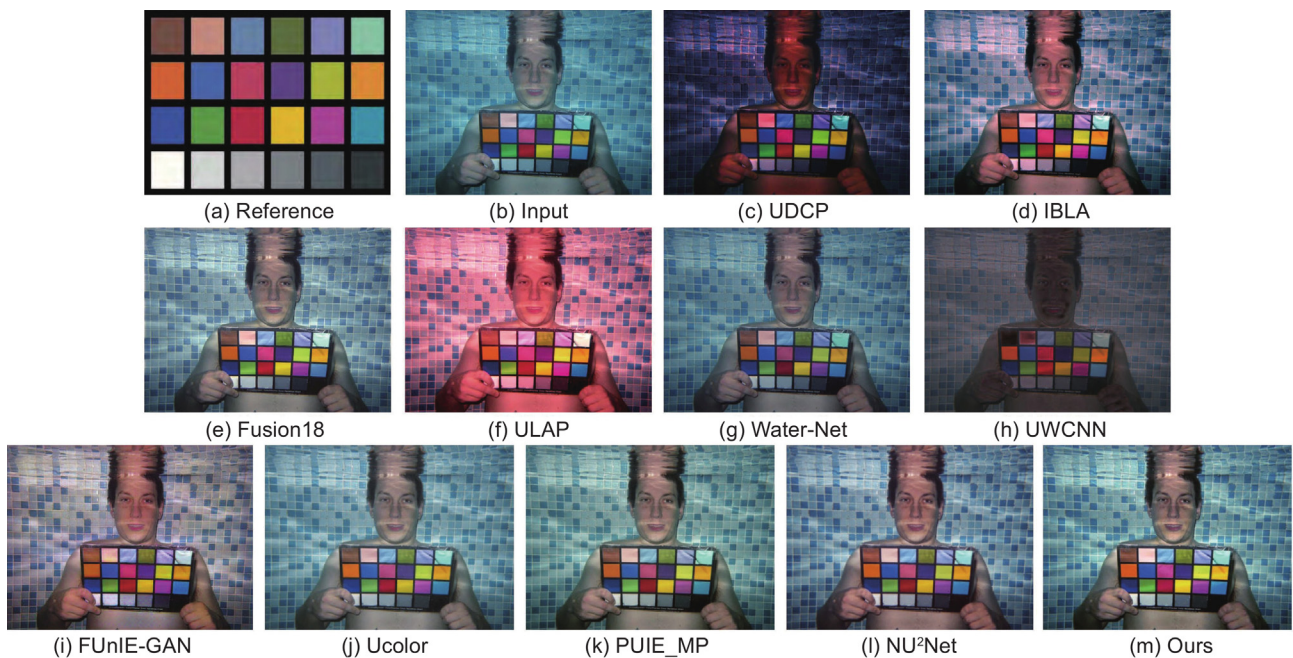


Fig. 8. Visual comparison of different methods on a Color Checker image taken by the Pentax W60 camera from [21].

Ucolor, NU<sup>2</sup>Net and MCRNet restored the color without introducing color deviation. Among them, MCRNet stands out for achieving the most accurate color restoration based on the reference color checker in Fig. 8(a).

Finally, we made a comparison of the enlarged local details for each deep learning-based method, as shown in Fig. 9. UWCNN removed some of the blue artifacts, but in the second image, like FUnIE-GAN, it also produced additional color distortion. Ucolor's results go a step further than Water-Net's, but neither can completely eliminate color deviation. In contrast, MCRNet removes

most of the blue color distortion, and retains the most details of the color board. In the second image, compared with Ucolor, the results of PUIE\_MP and MCRNet are more thorough in removing yellow distortion. However, PUIE\_MP has low contrast issue and lacks saturation in its result. NU<sup>2</sup>Net exhibits a phenomenon of over-enhancement, where its zoomed-in areas show excessively high saturation. The result of MCRNet has richer details, as can be seen from the enlarged areas. The above experimental examples demonstrate the superiority of the multi-color space

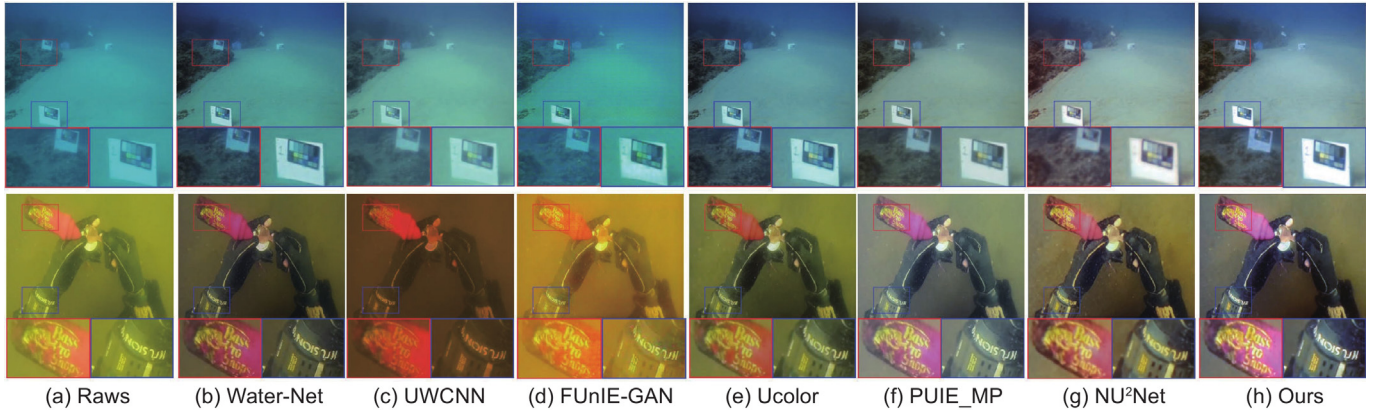


Fig. 9. Visual comparison of the enlarged local details.

Table 2

Quantitative comparison of different methods on Test-C60 and Test-SQUID.

Methods	Test-C60			Test-SQUID		
	UCIQE	UIQM	TM	UCIQE	UIQM	TM
Raw	0.515	1.856	0.330	0.500	0.655	0.355
UDCP [14]	0.529	1.084	0.214	<b>0.576</b>	0.782	0.197
IBLA [16]	<b>0.585</b>	1.579	0.534	0.528	0.733	0.422
Fusion18 [21]	0.574	2.373	0.594	<b>0.571</b>	2.116	0.644
ULAP [18]	0.566	1.372	0.468	0.526	0.734	0.436
Water-Net [30]	0.560	2.385	0.547	0.556	2.131	0.625
UWCNN [31]	0.505	2.151	0.317	0.485	2.035	0.421
FUnIE-GAN [48]	0.543	<b>2.863</b>	0.523	0.513	1.941	0.560
Ucolor [38]	0.551	2.414	0.531	0.541	2.049	<b>0.671</b>
PUIE_MP [49]	0.553	2.357	0.585	0.543	2.112	0.533
NU <sup>2</sup> Net [50]	0.571	<b>2.508</b>	<b>0.660</b>	0.562	<b>2.202</b>	0.635
<b>Ours</b>	<b>0.594</b>	2.483	<b>0.736</b>	<b>0.571</b>	<b>2.206</b>	<b>0.683</b>

fusion mechanism in color correction and contrast improvement of underwater images.

#### 4.4. Application tests

To verify that our method can better promote the visual perception of underwater robots, we conducted two application tests of underwater feature point detection and edge detection. We selected the test images from the Test-U90 and PKU datasets [56].

**Underwater feature point detection:** In order to realize autonomous positioning and navigation in the underwater environment, underwater robots need to use abundant semantic information, and the number of underwater feature points can reflect that to some extent. In Fig. 10, we conduct SIFT [57] feature points detection for the enhanced results of all deep learning-based methods. MCRNet detected the most feature points with its excellent color correction ability.

**Underwater edge detection:** The quality of the underwater image determines its sharpness of the edge. Detecting sufficient number of accurate edges is beneficial to the robot to perceive the underwater environment. We use Canny Edge Detector [58] to detect the edge contour of the input image and enhanced results of the deep learning-based methods. As shown in Fig. 11, compared with other methods, MCRNet's results have the most complete edge contour. Experimental results show that the proposed method can recover the edges of underwater images well.

Table 3

Comparison of average quantitative metrics of the ablation study on Test-U90.

Methods	SSIM	PSNR	UCIQE	UIQM	TM
w/ 3-RGB	0.876	20.030	0.576	2.856	1.250
w/o Lab	0.902	21.724	0.607	<b>2.906</b>	<b>1.319</b>
w/o HSV	0.886	20.924	0.615	2.852	1.211
w/o MCFF (concat)	<b>0.912</b>	<b>22.433</b>	<b>0.618</b>	2.862	<b>1.325</b>
w/o MCFF (sum)	0.896	21.575	0.620	2.849	1.271
<b>Full Model</b>	<b>0.918</b>	<b>22.755</b>	<b>0.620</b>	<b>2.880</b>	1.305

#### 4.5. Ablation study

We conducted a comprehensive ablation study to demonstrate the contribution and effectiveness of each module. The experimental settings are as follows:

- **w/o HSV:** MCRNet without the HSV color space path.
- **w/o Lab:** MCRNet without the Lab color space path.
- **w/ 3-RGB:** All input paths are RGB images.
- **w/o MCFF (concat):** Instead of using the multi-color space feature fusion module, MCRNet uses the simple concatenation operation.
- **w/o MCFF (sum):** Instead of using the multi-color space feature fusion module, MCRNet uses the simple summation operation.

In Fig. 12, we present the enhanced results of each ablation method for two severely distorted images. The 3RGB and w/o Lab lose correct color information, while w/o HSV and w/o MCFF (concat) partially remove the color deviation, resulting in yellowish or reddish images. On the other hand, w/o MCFF (sum) introduces green color deviation in edge details of the image, such as the edge of the phone in the first image and the joint of the gloves in the second image. The Full Model utilizes features from RGB, HSV, and Lab color spaces, and leverages the multi-color space feature fusion module to integrate these features. As a result, it completely eliminates color distortion, improves image contrast, and makes the underwater image more transparent.

Table 3 reports the quantitative results of our ablation study using the Test-U90 dataset, and the Full Model achieves the best performance across most metrics. The effectiveness of our multi-color space feature fusion mechanism is demonstrated by both visual comparisons and quantitative analysis.

## 5. Conclusion

This paper proposes an underwater image enhancement model based on the multi-color space feature fusion mechanism. The

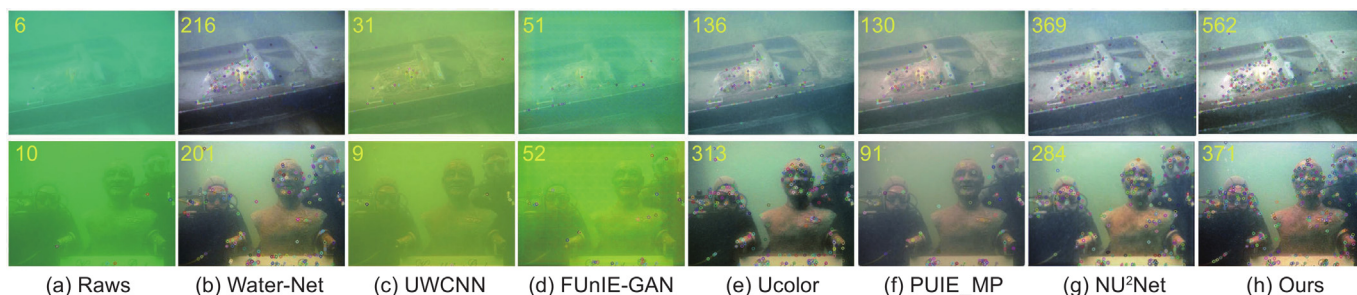


Fig. 10. Underwater feature point detection. In the top left corner is the number of feature points for each method.

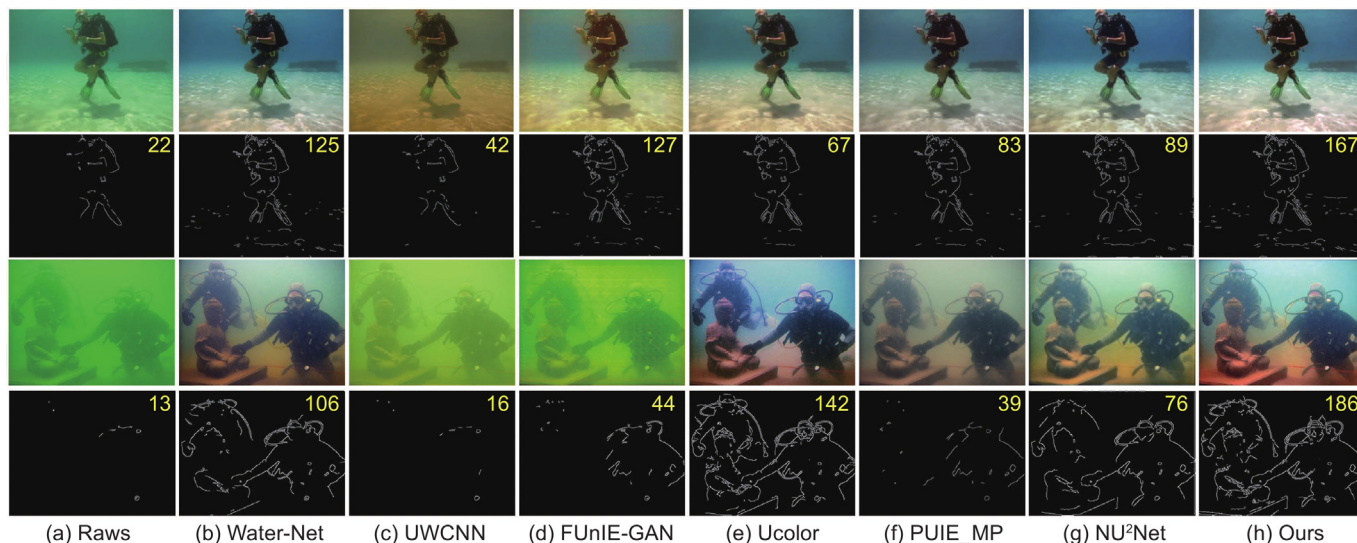


Fig. 11. Underwater edge detection. In the top right corner is the number of edges for each method.

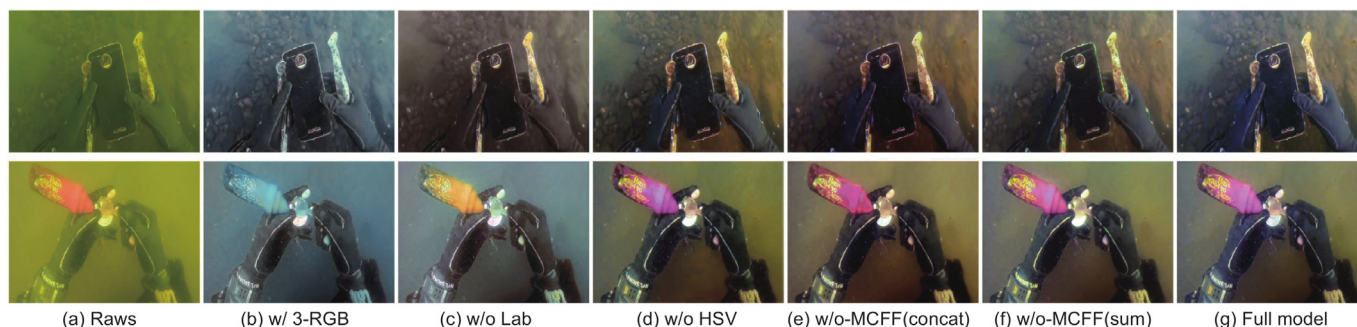


Fig. 12. Visual comparison of the ablation study. It is suggested to zoom in for a better view.

model leverages features from multiple color spaces and adaptively selects the most informative feature representation through a self-attention-based multi-color space feature fusion module. Extensive experiments show that our method can restore color distortion and improve the contrast of underwater images while preserving image details. Moreover, our application test demonstrates that our method has a positive impact on other visual tasks. Additionally, ablation studies demonstrate the contribution of each module in our method. However, we acknowledge that the instability of existing underwater image evaluation metrics and the challenges in acquiring reference images for underwater datasets are two urgent problems that need to be addressed in future work.

### CRediT authorship contribution statement

**Ningwei Qin:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis. **Junjun Wu:** Supervision, Project administration. **Xilin Liu:** Writing – review & editing, Validation, Methodology. **Zeqin Lin:** Supervision, Conceptualization. **Zhifeng Wang:** Supervision.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported in part by the National Key R&D Program of China (2022YFB4702300), in part by the National Natural Science Foundation of China(62273097), in part by the Guangdong Basic and Applied Basic Research Foundation, China (2022A1515140044, 2019A1515110304, 2020A1515110255, and 2021B1515120017), in part by the Research Foundation of Universities of Guangdong Province, China (2019KZDZX1026, 2020KCXTD015, and 2021KCXTD083), in part by the Foshan Key Area Technology Research Foundation, China (2120001011009), and in part by the Guangdong Philosophy and Social Science Program, China (GD23XTS03).

## References

- [1] W.-H. Lin, J.-X. Zhong, S. Liu, T. Li, G. Li, RoIMix: proposal-fusion among multiple images for underwater object detection, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2020, pp. 2588–2592.
- [2] X. Liang, P. Song, Excavating roi attention for underwater object detection, in: 2022 IEEE International Conference on Image Processing, ICIP, IEEE, 2022, pp. 2651–2655.
- [3] N. Nezla, T.M. Haridas, M. Supriya, Semantic segmentation of underwater images using unet architecture based deep convolutional encoder decoder model, in: 2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS, vol. 1, IEEE, 2021, pp. 28–33.
- [4] S. Mittal, S. Srivastava, J.P. Jayanth, A survey of deep learning techniques for underwater image classification, IEEE Trans. Neural Netw. Learn. Syst. (2022).
- [5] X. Liu, S. Wen, Z. Pan, C. Xu, J. Hu, H. Meng, Vision-IMU multi-sensor fusion semantic topological map based on ratslam, Measurement 220 (2023) 113335.
- [6] S. Wen, Z. Zhang, C. Ma, Y. Wang, H. Wang, An extended Kalman filter-simultaneous localization and mapping method with Harris-scale-invariant feature transform feature recognition and laser mapping for humanoid robot navigation in unknown environment, Int. J. Adv. Robot. Syst. 14 (6) (2017) 1729881417744747.
- [7] R. Liu, Z. Jiang, S. Yang, X. Fan, Twin adversarial contrastive learning for underwater image enhancement and beyond, IEEE Trans. Image Process. 31 (2022) 4922–4936.
- [8] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, R. Timofte, Swinir: Image restoration using swin transformer, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1833–1844.
- [9] D. Song, Y. Wang, H. Chen, C. Xu, C. Xu, D. Tao, Adders: Towards energy efficient image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 15648–15657.
- [10] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.-H. Yang, L. Shao, Multi-stage progressive image restoration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14821–14831.
- [11] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.-H. Yang, Restormer: Efficient transformer for high-resolution image restoration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5728–5739.
- [12] S. Raveendran, M.D. Patil, G.K. Birajdar, Underwater image enhancement: A comprehensive review, recent trends, challenges and applications, Artif. Intell. Rev. 54 (7) (2021) 5413–5467.
- [13] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2010) 2341–2353.
- [14] P. Drews, E. Nascimento, F. Moraes, S. Botelho, M. Campos, Transmission estimation in underwater single images, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2013, pp. 825–830.
- [15] A. Galdran, D. Pardo, A. Picón, A. Alvarez-Gila, Automatic red-channel underwater image restoration, J. Vis. Commun. Image Represent. 26 (2015) 132–145.
- [16] Y.-T. Peng, P.C. Cosman, Underwater image restoration based on image blurriness and light absorption, IEEE Trans. Image Process. 26 (4) (2017) 1579–1594.
- [17] D. Akkaynak, T. Treibitz, Sea-thru: A method for removing water from underwater images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 1682–1691.
- [18] W. Song, Y. Wang, D. Huang, D. Tjondronegoro, A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration, in: Pacific Rim Conference on Multimedia, Springer, 2018, pp. 678–688.
- [19] J. Zhou, T. Yang, W. Chu, W. Zhang, Underwater image restoration via backscatter pixel prior and color compensation, Eng. Appl. Artif. Intell. 111 (2022) 104785.
- [20] C. Ancuti, C.O. Ancuti, T. Haber, P. Bekaert, Enhancing underwater images and videos by fusion, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 81–88.
- [21] C.O. Ancuti, C. Ancuti, C. De Vleeschouwer, P. Bekaert, Color balance and fusion for underwater image enhancement, IEEE Trans. Image Process. 27 (1) (2017) 379–393.
- [22] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, X. Ding, A retinex-based enhancing approach for single underwater image, in: 2014 IEEE International Conference on Image Processing, ICIP, IEEE, 2014, pp. 4572–4576.
- [23] P. Zhuang, C. Li, J. Wu, Bayesian retinex underwater image enhancement, Eng. Appl. Artif. Intell. 101 (2021) 104171.
- [24] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, B. Wang, Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior, IEEE Trans. Image Process. 25 (12) (2016) 5664–5677.
- [25] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, C. Li, Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement, IEEE Trans. Image Process. 31 (2022) 3997–4010.
- [26] J. Wu, X. Liu, Q. Lu, Z. Lin, N. Qin, Q. Shi, FW-GAN: Underwater image enhancement using generative adversarial network with multi-scale fusion, Signal Process., Image Commun. 109 (2022) 116855.
- [27] Q. Jiang, Y. Zhang, F. Bao, X. Zhao, C. Zhang, P. Liu, Two-step domain adaptation for underwater image enhancement, Pattern Recognit. 122 (2022) 108324.
- [28] H. Li, P. Zhuang, Dewaternet: A fusion adversarial real underwater image enhancement network, Signal Process., Image Commun. 95 (2021) 116248.
- [29] C. Fabbri, M.J. Islam, J. Sattar, Enhancing underwater imagery using generative adversarial networks, in: 2018 IEEE International Conference on Robotics and Automation, ICRA, IEEE, 2018, pp. 7159–7165.
- [30] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, D. Tao, An underwater image enhancement benchmark dataset and beyond, IEEE Trans. Image Process. 29 (2019) 4376–4389.
- [31] C. Li, S. Anwar, F. Porikli, Underwater scene prior inspired deep underwater image and video enhancement, Pattern Recognit. 98 (2020) 107038.
- [32] Z. Shen, H. Xu, T. Luo, Y. Song, Z. He, UDAformer: Underwater image enhancement based on dual attention transformer, Comput. Graph. (2023).
- [33] L. Peng, C. Zhu, L. Bian, U-shape transformer for underwater image enhancement, 2021, arXiv preprint arXiv:2111.11843.
- [34] K. Iqbal, M. Odetayo, A. James, R.A. Salam, A.Z.H. Talib, Enhancing the low quality images using unsupervised colour correction method, in: 2010 IEEE International Conference on Systems, Man and Cybernetics, IEEE, 2010, pp. 1703–1709.
- [35] O. Deperlioglu, U. Kose, G.E. Guraksin, Underwater image enhancement with HSV and histogram equalization, Image 1 (4) (2018) 461–465.
- [36] J. Hu, Q. Jiang, R. Cong, W. Gao, F. Shao, Two-branch deep neural network for underwater image enhancement in HSV color space, IEEE Signal Process. Lett. 28 (2021) 2152–2156.
- [37] Z. Lyu, A. Peng, Q. Wang, D. Ding, An efficient learning-based method for underwater image enhancement, Displays 74 (2022) 102174.
- [38] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, W. Ren, Underwater image enhancement via medium transmission-guided multi-color space embedding, IEEE Trans. Image Process. 30 (2021) 4985–5000.
- [39] Y.-T. Peng, K. Cao, P.C. Cosman, Generalization of the dark channel prior for single image restoration, IEEE Trans. Image Process. 27 (6) (2018) 2856–2868.
- [40] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 286–301.
- [41] X. Li, W. Wang, X. Hu, J. Yang, Selective kernel networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 510–519.
- [42] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
- [43] B. Lim, S. Son, H. Kim, S. Nah, K. Mu Lee, Enhanced deep residual networks for single image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 136–144.
- [44] P. Charbonnier, L. Blanc-Feraud, G. Aubert, M. Barlaud, Two deterministic half-quadratic regularization algorithms for computed imaging, in: Proceedings of 1st International Conference on Image Processing, vol. 2, IEEE, 1994, pp. 168–172.
- [45] G. Seif, D. Androutsos, Edge-based loss function for single image super-resolution, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2018, pp. 1468–1472.

- [46] N. Silberman, D. Hoiem, P. Kohli, R. Fergus, Indoor segmentation and support inference from RGBD images, in: *European Conference on Computer Vision*, Springer, 2012, pp. 746–760.
- [47] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.
- [48] M.J. Islam, Y. Xia, J. Sattar, Fast underwater image enhancement for improved visual perception, *IEEE Robot. Autom. Lett.* 5 (2) (2020) 3227–3234.
- [49] Z. Fu, W. Wang, Y. Huang, X. Ding, K.-K. Ma, Uncertainty inspired underwater image enhancement, in: *European Conference on Computer Vision*, Springer, 2022, pp. 465–482.
- [50] C. Guo, R. Wu, X. Jin, L. Han, W. Zhang, Z. Chai, C. Li, Underwater ranker: Learn which is better and how to be better, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, (no. 1) 2023, pp. 702–709.
- [51] D. Berman, D. Levy, S. Avidan, T. Treibitz, Underwater single image color restoration using haze-lines and a new quantitative dataset, *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (8) (2020) 2822–2837.
- [52] K. Panetta, C. Gao, S. Aghaian, Human-visual-system-inspired underwater image quality measures, *IEEE J. Ocean. Eng.* 41 (3) (2015) 541–551.
- [53] M. Yang, A. Sowmya, An underwater color image quality evaluation metric, *IEEE Trans. Image Process.* 24 (12) (2015) 6062–6071.
- [54] Z. Fu, X. Fu, Y. Huang, X. Ding, Twice mixing: A rank learning based quality assessment approach for underwater image enhancement, *Signal Process., Image Commun.* 102 (2022) 116622.
- [55] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (4) (2004) 600–612.
- [56] Z. Chen, T. Jiang, Y. Tian, Quality assessment for comparing image enhancement algorithms, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3003–3010.
- [57] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vis.* 60 (2) (2004) 91–110.
- [58] J. Canny, A computational approach to edge detection, *IEEE Trans. Pattern Anal. Mach. Intell.* (6) (1986) 679–698.