



Learning multiscale pipeline gated fusion for underwater image enhancement

Xu Liu¹ · Sen Lin² · Zhiyong Tao³

Received: 15 January 2022 / Revised: 10 May 2022 / Accepted: 3 February 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Evidence suggests that vision is among the most critical factors in marine information exploration. Instead, underwater images are generally poor quality due to color casts, lack of texture details, and blurred edges. Therefore, we propose the Multiscale Gated Fusion conditional GAN (MGF-cGAN) for underwater image enhancement. The generator of MGF-cGAN consists of Multiscale Feature Extract Module (Ms-FEM) and Gated Fusion Module (GFM). In Ms-FEM, we use three different parallel subnets to extract feature information, which can extract richer features than a single branch. The GFM can adaptively fuse the three outputs from Ms-FEM. GFM generates better chromaticity and contrast than other fusion ways. Additionally, we add the Multiscale Structural Similarity Index Measure (MS-SSIM) loss to train the network, which is highly similar to human perception. Extensive experiments across three benchmark underwater image datasets corroborate that MGF-cGAN can generate images with better visual perception than classical and State-Of-The-Art (SOTA) methods. It achieves 27.1078dB PSNR and 11.9437 RMSE on EUVP dataset. More significantly, enhanced results of MGF-cGAN also provide excellent performance in underwear saliency detection, SURF key matching test, and so on. Based on this study, MGF-cGAN is found to be suitable for data preprocessing in an underwater multimedia system.

Keywords Underwater image enhancement · Multiscale feature extraction · Gated fusion · MS-SSIM loss · Conditional GAN

✉ Sen Lin
lin.sen6@126.com

Xu Liu
dalong.xu.liu@ieee.org

Zhiyong Tao
xyzmail@126.com

¹ School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, 230601, China

² School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang, 110159, China

³ School of Electronic and Information Engineering, Liaoning Technical University, Huludao, 125105, China

1 Introduction

With the advent of marine research, underwater information processing has increasingly become an indispensable technique [1, 25, 55]. Unfortunately, the quality and visual experience of the original marine image is poor owing to light absorbance, and refraction by seawater [13, 47], resulting in substantial difficulties while presenting aquatic information. Whether in ocean exploration or resource development, research on automatic underwater image enhancement is paramount. This research is mainly aimed at the data preprocessing of the underwater multimedia system. The images processed by our method are clearer and can reflect richer environmental information (Fig. 1).

Deep learning-driven means are being used to realize image processing [8, 9, 39, 40, 45], for example, underwater image enhancement, image dehazing, image super-resolution, etc. Wang et al. [46] presented a CNN-based underwater image enhancement network (UIE-Net), which can solve both color cast and hazy problems. Since real-world training sets are lacking, the network must be trained on synthetic images. To alleviate the aforementioned congestion, Li et al. [28] established the real-world underwater image enhancement dataset (UIEB), which provides a benchmark for training and evaluating. They also proposed a CNN-based multi-input fusion framework (Water-Net), enhancing the visual quality of images effectively. In practice, the network requires three preprocessed images, which greatly slows down the network. Li et al. [30] proposed a GAN-based fusion method (DewaterNet) for enhancing underwater images. Combined with basic blocks and multi-term loss function, it can also produce visually appealing results. However, this method still needs to preprocess the input image, and the effect of fusion is not obvious. Liu et al. [31] designed

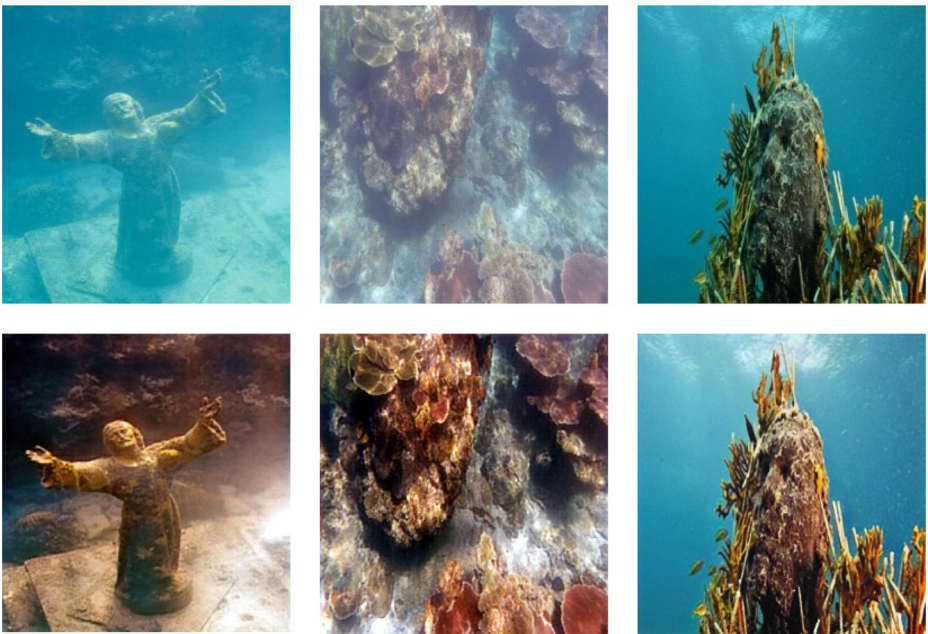


Fig. 1 In the top row are raw images taken underwater, while in the bottom row are the images enhanced by our method

a CNN-based deep residual network (UResnet), which works well with color balancing and detail improvement. Instead, the network uses the same convolution layers which may lead to problems such as insufficient feature learning. Zong et al. [56] presented a GAN-based cycle network (Local-CycleGAN) that enhances images acquired in complicated environments. The network can simultaneously supervise local and global information through multiple losses.

Presently, deep learning-based underwater image enhancement has been the subject of many research studies. Even though these methods contribute to the overall quality of the image, there are several issues to consider:

- 1) Some algorithms produce weak enhancement effects for images with problems such as color cast, lack of details.
- 2) Due to the use of the same convolution layers, the extracted features are not rich, resulting in poor learning ability and insufficient performance (e.g. UResnet).
- 3) There are few studies on fusion module, and some algorithms require image preprocessing before fusion (e.g. Water-Net, DewaterNet).
- 4) The loss function of most image enhancement networks has a fixed shape, and new loss functions are rarely employed.

To address these issues, we propose the multiscale gated fusion conditional GAN (MGF-cGAN) for underwater image enhancement. By comparison with classical and State-Of-The-Art (SOTA) methods, the proposed method shows more promising results and is effective in multiple underwater applications. The following are our major contributions:

- 1) We present a network called MGF-cGAN, which can correct colors, enhance details. In this framework, the generator enhances the underwater image, and the discriminator determines whether the image is generated or ground truth.
- 2) In the generator, a multiscale feature extract module (Ms-FEM) is designed to extract the features. Ms-FEM consists of three different parallel subnets, which can extract richer features than a single branch.
- 3) We also propose the gated fusion module (GFM) to synthesize clear images by adaptively fusing the features from parallel subnets. GFM generates better chromaticity and contrast than other fusion ways.
- 4) Considering the impact of resolution on image clarity, we use the loss functions with MS-SSIM, which is also more suitable for human perception.

In this paper, Section 2 outlines the background related to our method. Section 3 describes our algorithm in detail. Section 4 indicates the qualitative and quantitative analyses. Section 5 depicts a comprehensive ablation study. Section 6 discusses the advantages and limitations of our method. Section 7 offers a conclusion of our work.

2 Background

2.1 Deep learning-driven underwater image enhancement algorithms

Traditional methods use specific filters [37] and morphological techniques [34] to balance local colors, enhance contrast and brightness. Unlike traditional methods, deep learning-driven underwater image enhancement algorithms learn the paired training data through the network, achieving better performance. Deep learning-related methods can be classified into two types: *CNN-based model* [10, 27–29, 31, 44, 46]; *GAN-based model* [5, 14, 19,

30, 56]. CNN-based models prioritize the original underwater images, whereas GAN-based models emphasize image quality.

Methods with CNN-based model Li et al. [27] proposed a scene prior inspired deep network (UWCNN). Underwater image reconstruction is done directly by the network, rather than estimating model parameters, which may be beneficial from a synthetic underwater image training set. The model is lightweight, but it does not do well for enhancing image details. To address color casts and low contrast problems, Li et al. [29] presented Ucolor, which fully exploits the characteristics of multiple color spaces. Ucolor can effectively improve the visual quality of underwater images. Unfortunately, Ucolor fails to work on some underwater images with high turbidity.

Methods with GAN-based model Despite the development of various algorithms, real-time and adaptable approaches remain inadequate for real-world tasks. Chen et al. [5] created an underwater image restoration scheme (GAN-RS), which can automatically enhance underwater vision in real time. This method has been proven to be effective in improving the target grasping ability of underwater robots. Nevertheless, the images generated by this method tend to be gray and white, and the colors are not vivid. To correct color distortion, underexposure, and fuzziness, Guo et al. [14] developed a multiscale dense network (UWGAN), which can provide greater detail information and take advantage of previously features. But this algorithm cannot generate images that are aesthetically pleasing.

Light propagates through water similarly to how it does in the atmosphere, but the underwater image has its unique optical model. So we will introduce the underwater optical imaging model. Additionally, pixel-to-pixel deep network (P2P) and conditional GAN (cGAN) are also two basic models we will describe.

2.2 Basic models according to our work

Underwater optical imaging model It can be represented with three separate components according to the McGlamery model [32] in Fig. 2. Direct transmission is calculated by using the infrared back of the aiming reticule. The forward scattering component describes a portion of the reflected light dispersed at a slight angle. The background scattering component represents the background light dispersed by the remaining small particles.

When the $d(x)$ is short enough, forward scattering can be ignored by imaging systems [24]. As a consequence, the simplified model is specified:

$$I_{\beta}(z) = J_{\beta}(z) \cdot t_{\beta}(z) + B_{\beta} \cdot (1 - t_{\beta}(z)) \quad (1)$$

where $I_{\beta}(z)$ is the intensity observed at the pixel z and it is formed by mixing the background light B_{β} in relation to the transmission map $t_{\beta}(z)$.

Pixel-to-pixel deep network (P2P) Researchers commonly use convolution layers to filter out the noise and keep the crucial components of the input. In underwater image processing, continuous convolution layers can extract most features. However, these layers cannot recover the details of low definition images. Therefore, Isola et al. [23] introduced the deconvolution layers to refine the details and proposed the pixel-to-pixel deep network (P2P). Sun et al. [41] applied the P2P to underwater image enhancement. This model enhances the image pixel-by-pixel rather than considering the physical parameters of the

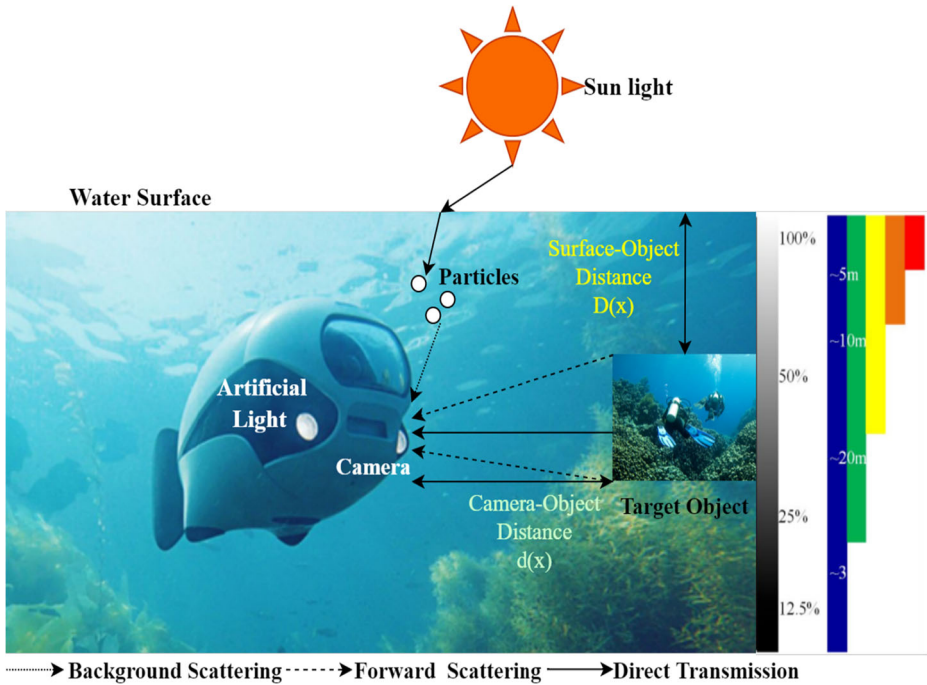


Fig. 2 Underwater optical imaging model

image. Inspired by P2P, we design three different parallel subnets, which can extract and refine the details of the input.

Conditional GAN (cGAN) On the other hand, GAN is too free, uncontrolled for larger images with more pixels. Comparatively, due to the conditional information, cGAN [33] offers better stability and representation flexibility than the original GAN during the process of image enhancement. Li et al. [26] presented an encoder-decoder generator network (Dehaze-cGAN) for image dehazing, which extends the basic cGAN formulation to help generate realistic clear images. To tackle the difficulty of removing raindrops, Zhang et al. [53] employed cGAN. When compared to the GAN-based method, cGAN delivers superior outcomes. In [43], Wang et al. established a cGAN-based high-resolution image synthesis network, which generates high-quality images by using a modified loss function. Employing a similar methodology, MGF-cGAN builds upon cGAN to produce high-resolution results and improve robustness.

3 Methods and materials

The MGF-cGAN includes two parts, as shown in Fig. 3. The first part is the generator responsible for generating clear images. Part two involves a discriminator capable of distinguishing the images produced by the generator from the corresponding target images. By using discriminator-supervised generator learning, the network is able to enhance images well.

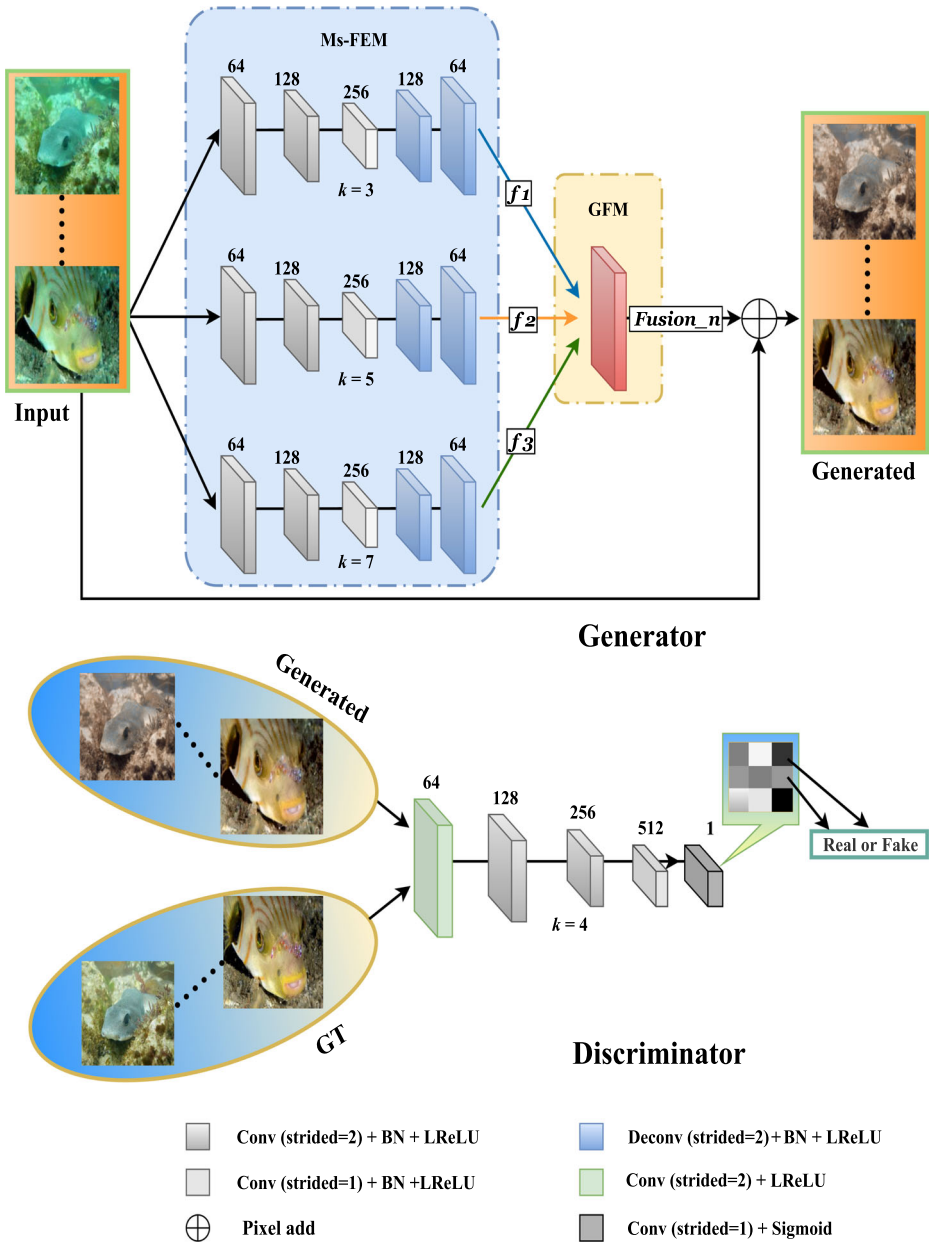


Fig. 3 Overview of the proposed method. k is the kernel size of convolution layers (conv) or deconvolution layers (deconv), above each block is the number of channels, GT means target image (ground truth), BN represents the batch normalization, LReLU represents the leaky ReLU function

3.1 Generator

In Fig. 4, the generator is composed of multiscale feature extract module (Ms-FEM) and gated fusion module (GFM). Ms-FEM and GFM complete feature extraction and fusion tasks respectively.

Specifically, the Ms-FEM comprises three different parallel subnets ($k=3, k=5, \text{ and } k=7$). Each subnet consists of a symmetric structure shaped like a P2P network. The subnet is composed of encoders and decoders, which are responsible for extracting and refining features. Additionally, the batch normalization (BN) [21] and the leaky ReLU (LReLU) function [15] in each block reduce the number of parameters and improve network convergence respectively. Moreover, we add a convolution layer (stride=1) in the middle of each subnet to help increase the nonlinearity and make the subnet deeper while maintaining the size of feature maps. GFM adaptively fuses the features extracted by Ms-FEM, and its relevant information will be introduced in detail below.

3.2 Discriminator

The discriminator is similar to the PatchGAN [49], as shown in Fig. 4. In the discriminator, each pixel block is determined by whether each $M \times M$ pixel block is real or created by the generator. The last layer of the discriminator uses the sigmoid function to convert the output range to $[0, 1]$, and then gets the score of authenticity. Each pixel block represents the difference between the local receptive fields of the input. As the model of GAN progresses and matures, PatchGAN is extended to serve as the basis for the cycle adversarial network (CycleGAN) [51]. The discriminator parameters are less than the full image discriminator.

3.3 Gated Fusion Module (GFM)

Nowadays, the gate module [38] has a significant effect on the fusion of essential features. Inspired by the gate module, we propose the gated fusion module (GFM)

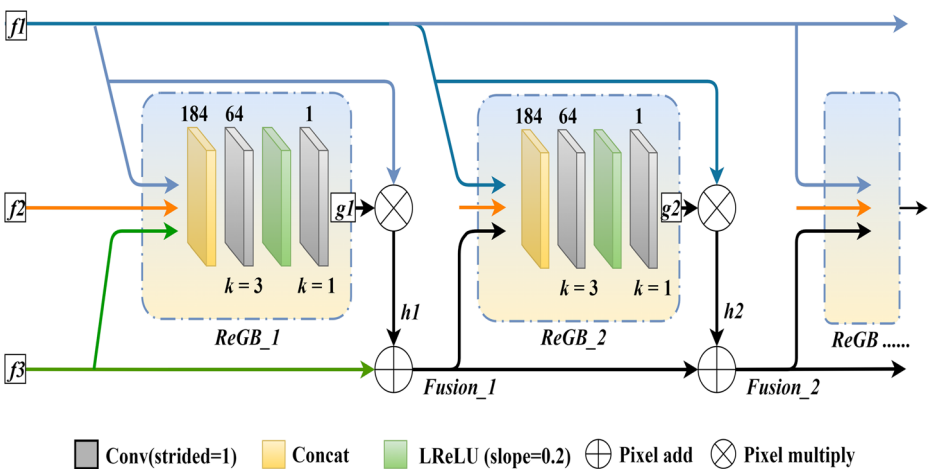


Fig. 4 Structure of gated fusion module (GFM). k is the kernel size of convolution layers (conv) or deconvolution layers (deconv), above each block is the number of channels, concat represents concatenate layer

that adopts a recursive strategy to fuse the features extracted from parallel subnets gradually.

As shown in Fig. 4, the GFM is composed of multiple sets of recursive gate block (ReGB). In the first ReGB, ϕ_{f1} , ϕ_{f2} and ϕ_{f3} represents the features extracted from the three parallel subnets. Parallel subnet ($k=3$) has a small receptive field for extracting detailed features, while parallel subnet ($k=7$) possesses a larger receptive field for extracting structural features. Between parallel subnet ($k=3$) and parallel subnet ($k=7$), parallel subnet ($k=5$) can extract most of the features.

Hence, we first obtain ϕ_{g1} through the ReGB:

$$\phi_{g1} = ReGB_{-1}(\phi_{f1}, \phi_{f2}, \phi_{f3}) \tag{2}$$

second, we multiply ϕ_{g1} with the ϕ_{f1} acquired from the parallel subnet ($k=3$) to enhance detailed features:

$$\phi_{h1} = \phi_{g1} \otimes \phi_{f1} \tag{3}$$

where \otimes means pixel by pixel multiplication.

Require: $\phi_{f1}, \phi_{f2}, Fusion_{n-1}$

Ensure: $Fusion_n$

$$m = 1, k = 1$$

for m in $1 \dots n$ **do**

(I) Specific operations in formula (5) and layers in ReGB.

$$Concat(\phi_{f1}, \phi_{f2}, Fusion_{n-1})$$

for k in $1, 2$ **do**

if $k = 1$ **then**

$$LReLU(Conv_{3 \times 3}(stride=1))$$

else

$$Conv_{1 \times 1}(stride=1)$$

end if

end for

obtain ϕ_{gn}

(II) Processed as formula (6), (7).

end for

return $Fusion_n$

Algorithm 1 Our proposed GFM scheme.

Finally, the results represented by ϕ_{h1} and ϕ_{f3} (from parallel subnet ($k=7$)) are then added together to enhance structural features:

$$Fusion_{-1} = \phi_{h1} + \phi_{f3} \tag{4}$$

where $+$ means pixel by pixel addition.

As evidenced above, we proposed a recursive strategy to fuse the observed feature by using the dependency of three parallel subnets. Through stacking n ReGBs in a clump,

each ReGB has the same goal of integrating the features ϕ_{f1} and ϕ_{f3} into the mainstream $Fusion_{n-1}$ in an adaptive way.

Every fusion operation uses the features obtained from the last recursive round. Thus, the feature $Fusion_{n-1}$ learned from different rounds has been fused continuously:

$$\phi_{gn} = ReGB_{-n}(\phi_{f1}, \phi_{f2}, Fusion_{n-1}) \quad (5)$$

$$\phi_{hn} = \phi_{gn} \otimes \phi_{f1} \quad (6)$$

$$Fusion_n = \phi_{hn} + Fusion_{n-1} \quad (7)$$

DGFN [38] uses a module to fuse the three enhanced images to generate a clear image. However, due to DGFN requiring pre-processing of the input, adaption to other imaging fields is entirely unfeasible. In contrast, the proposed GFM is a generalized image fusion module that does not rely on prior knowledge and pre-processes. GFM is designed to fuse the features collected by Ms-FEM adaptively.

3.4 The choice of ReGB number

In determining the best number of ReGB in GFM, we use 245 underwater images from UIEB [28] (188 images), EUVP [22] (48 images) and U45 [2] (9 images) for the testing of this section. The implementation details will be introduced in Section 4.

A comparison of test images and their corresponding gradient maps between pixel by pixel addition (ADD), concatenate fusion (CONCAT), and GFM with different ReGB numbers are shown in Figs. 5 and 6. Before the fusion, the output shape of the last layer is set as $h \times w \times 3$.

In Fig. 5, we can obviously see that most results generated by ReGB will be better than those from ADD and CONCAT. Significantly, the Image3, Image4, and Image5 generated from CONCAT are shrouded in a green mask, and there is no noticeable change in the ADD results. The surface in Image1, Image3, and Image4 suffers from severe exposure when $ReGB = 1$. In contrast, the images generated from $ReGB = 2$ are bright, with strong contrast. From the $ReGB = 3$, we find that its visual effect is worse than $ReGB = 2$. Furthermore, few results of the $ReGB = 4$ have seriously color cast and uneven noise.

In Fig. 6, images with CONCAT and ReGB are shown to have good texture features relative to the configuration of using ADD. With the increase of the number of ReGBs, the description of gradient maps corresponding to Fig. 6 become complex, and the texture of the results are clear. Moreover, it can be proved that the $ReGB = 3$ has a advantage of details compared with other situations.

To further highlight the effectiveness of GFM, we not only use information entropy (IE) [42] to describe the richness of features but also use color colorfulness index (CCI) [16] to assess the hue of images.

In Fig. 7, the values of IE show an upward trend before the $ReGB = 2$. Additionally, a remarkable advantage is evident by $ReGB = 1$ or $= 2$ when measured with the CCI. And we can clearly see that when the $ReGB = 2$, the values of IE and CCI are the highest points in their respective curves.

The above comparison results indicate that $ReGB = 2$ can produce the highest visual effect and image quality at the same time. Therefore, based on the comparative results above, the $ReGB = 2$ is selected as the primary network architecture of the GFM.

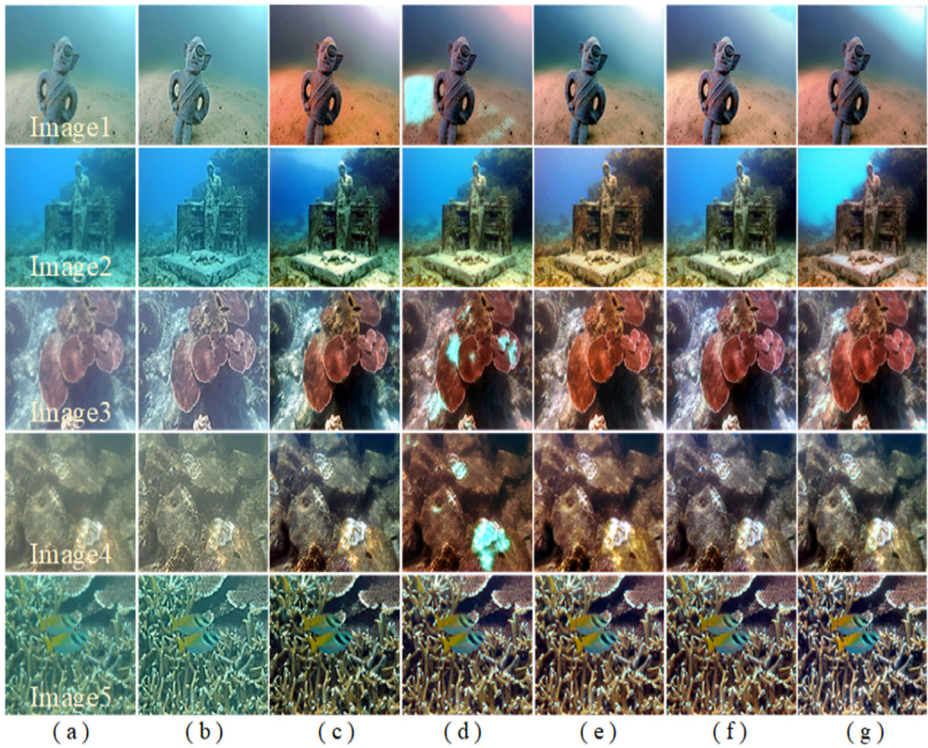


Fig. 5 Comparison of processed images between ADD, CONCAT and GFM with different numbers of ReGB. (a) Input, (b) ADD, (c) CONCAT, (d) ReGB = 1, (e) ReGB = 2, (f) ReGB = 3, (g) ReGB = 4

3.5 Loss function

(1) Wasserstein GAN with gradient penalty (WGAN-GP) loss

Considering the challenging problem during training the original GAN, namely, a bursting or diminishing gradient, we use the WGAN-GP loss [12] and change the condition variable:

$$\begin{aligned} \mathcal{L}_{\text{WGAN-GP}} = E_{u,v}[D(u, v)] - E_u[D(u, G(u))] \\ + \xi E_{\hat{u}} \left[(\|\hat{u} D(\hat{u})\|_2 - 1)^2 \right] \end{aligned} \tag{8}$$

where u and v are the input and ground truth, respectively, \hat{u} are the samples at the spatial distance between $G(u)$ and v , ξ represents the ratio.

(2) L_1 loss

L_1 loss [17] is selected to reduce the amount artifacts:

$$L_1 = \frac{1}{N} \sum_{e \in S} |x(e) - y(e)| \tag{9}$$

where e is the pixel, S is the color block represented by the pixel, $x(e)$ and $y(e)$ represent the pixel value of the generated and ground truth, respectively.

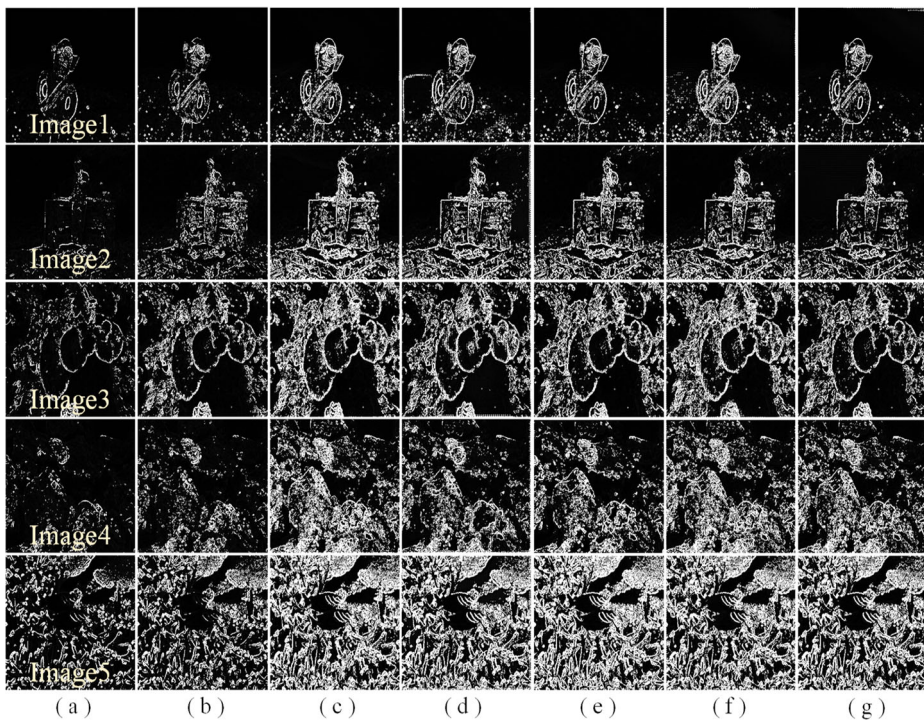
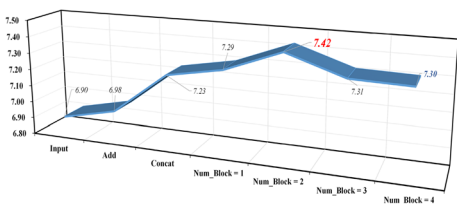


Fig. 6 Comparison of gradient maps between ADD, CONCAT and GFM with different numbers of ReGB. (a) Input, (b) ADD, (c) CONCAT, (d) ReGB = 1, (e) ReGB = 2, (f) ReGB = 3, (g) ReGB = 4

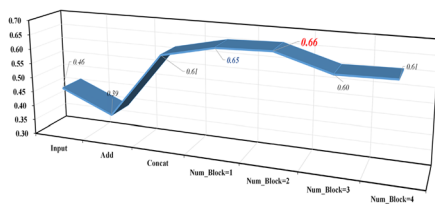
(3) Multiscale structural similarity index measure (MS-SSIM) loss

MS-SSIM loss [54] is an SSIM loss based on a multilayer. The MS-SSIM loss is calculated as follows for each level of the binary pyramid:

$$V_{MS-SSIM}(p) = I_M^\alpha(p) \cdot \prod_{j=1}^W cs_j^{\beta_j}(p) \tag{10}$$



(a) Comparison of Information Entropy (IE)



(b) Comparison of Color Colorfulness Index (CCI)

Fig. 7 IE and CCI comparison between ADD, CONCAT, and GFM with different numbers of ReGB. Num_Block is the number of ReGB. The red and blue figures indicate the highest and second-highest values, respectively

for convenience, we set $\alpha = \beta_j = 1$, $j = \{1, \dots, W\}$. The loss of color blocks P and the loss of its central pixel \tilde{p} can be roughly calculated as:

$$L_{\text{MS-SSIM}}(P) = 1 - V_{\text{MS-SSIM}}(\tilde{p}) \quad (11)$$

MS-SSIM loss may result in brightness changes and color deviation. To compensate for the aberration caused by MS-SSIM loss, we use L_1 loss to maintain brightness and color simultaneously.

(4) Total loss

Ultimately, the aggregate of three components means the total loss:

$$L = \lambda \mathcal{L}_{\text{WGAN-GP}} + \alpha L_1 + (100 - \alpha) L_{\text{MS-SSIM}} \quad (12)$$

where λ , α are constant, we have set to 2, 18, respectively.

4 Experiments

For this study, we choose 6000 paired samples from the EUVP [22] dataset (5500 paired images) and UIEB [28] dataset (500 paired images) to train MGF-cGAN, ensuring that the ratio of the sample taken out to the total sample is similar in the two datasets. In the EUVP dataset, there are numerous images of poor and good perceptual quality. A total of 11950 pairs of underwater images were collected during ocean explorations under varying circumstances. The UIEB dataset is a large-scale real-world underwater image enhancement benchmark consisting of 950 paired images. The underwater images were likely taken in natural light, artificial light, or a mixture of the two.

For testing, we selected the remaining 515 and 450 paired samples of the EUVP and UIEB datasets, respectively. Additionally, to further evaluate the ability of our method, we add U45 [2] dataset for testing the challenging samples. The U45 dataset contains underwater images of three classic scenes (greenish, bluish and hazy), which is very difficult to enhance.

The training process involves resizing samples to 256×256 and normalizing the pixel value to $[-1, 1]$. We set $\text{ReGB} = 2$, as proved by Section 3.4. On an NVIDIA Titan Xp (12G), the proposed MGF-cGAN is built by using Tensorflow and trained for 100 epochs. Training is sped up using the Adam optimizer with learning rate set at 0.0001, where β_1 and β_2 have default values of 0.5 and 0.99.

4.1 Subjective evaluation

Our purpose in this section is to conduct visual comparisons including color card and real-world underwater images. To examine the color recovery ability of MGF-cGAN, we conducted a comparison of underwater color card restoration. We evaluate proposed method against the following classical and state-of-the-art (SOTA) approaches: relative global histogram stretching (RGHS) [18], de-scattering and enhancing using dehazenet and HWD (DehazeNet & HWD) [35], CNN-based multi-input fusion framework (Water-Net) [28], CNN-based deep residual network (UResnet) [31], GAN-based fast underwater image enhancement (FUnIE-GAN) [22], conditional GAN-based underwater image restoration (UGAN) [50], GAN-based underwater image restoration scheme (GAN-RS) [6] in Fig. 8.

It can be clearly seen that the image processed by DehazeNet & HWD is generally gray and white, and the color card is dark. Results of FUnIE-GAN, UGAN, GAN-RS have low

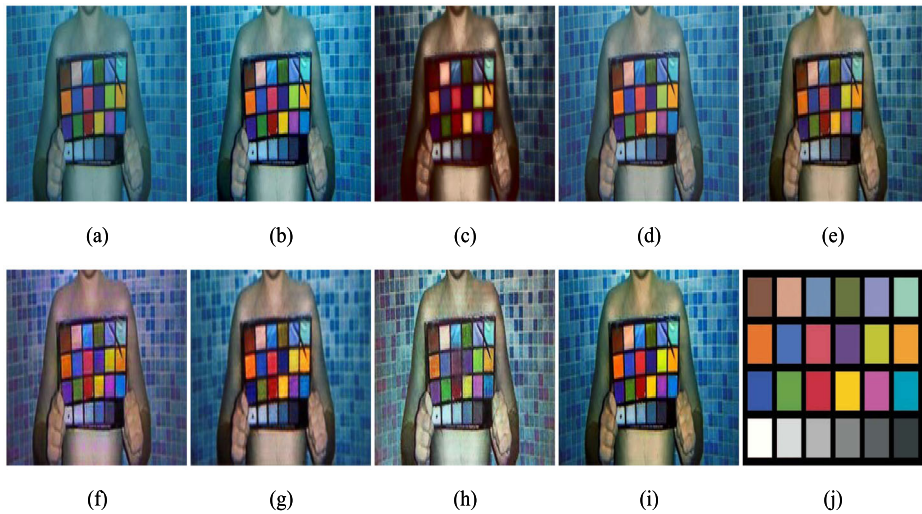


Fig. 8 Results of color card restoration. (a) Input, (b) RGHS, (c) DehazeNet&HWD, (d) Water-Net, (e) UResnet, (f) FUNIE-GAN, (g) UGAN, (h) GAN-RS, (i) MGF-cGAN, (j) Standard color card

contrast, and some blocks are indistinguishable. RGHS, Water-Net, UResnet and MGF-cGAN can correct the color effectively, while MGF-cGAN has more vivid colors and higher brightness.

In order to show enhancement of image color and details using MGF-cGAN, we compare it with the classic and SOTA algorithms mentioned above in Fig. 9. The RGHS can improve the contrast of images, but fails to address the color cast. And the details are blurry in Image4 and Image6. The color of DehazeNet&HWD results are relatively single, and the overall brightness of them are low. Furthermore, images enhanced by DehazeNet&HWD lack some texture features. The Water-Net cannot balance the color well, and outputs are too dark with poor visual effect. The UResnet, FUNIE-GAN and UGAN are not capable of removing haze from Image2, Image3, Image4, Image7 and there are red artifacts in some images processed by FUNIE-GAN. The UGAN has obvious yellowish artifacts in Image5. In contrast, both GAN-RS and MGF-cGAN can effectively enhance image details, while the images processed by MGF-cGAN have more balanced chromaticity and better contrast.

4.2 Objective evaluation

The subjective evaluation finds that the MGF-cGAN corrects the color and enhances the detail of images. To reaffirm the effects, we use structural similarity (SSIM) [52], peak signal-to-noise ratio (PSNR) [20], root mean square error (RMSE) [7] in UIEB [28] and EUVP [22] dataset, UCIQE [48], UIQM [36] in U45 [22] dataset to provide a objective evaluation of the image quality. The SSIM, PSNR and RMSE can determine the difference between the generated image and the ground truth. In addition, UCIQE and UIQM can comprehensively evaluate the quality of underwater images. Tables 1 and 2 list the average SSIM and PSNR values on UIEB and EUVP dataset. Table 3 shows the average UCIQE and UIQM values on U45 dataset.

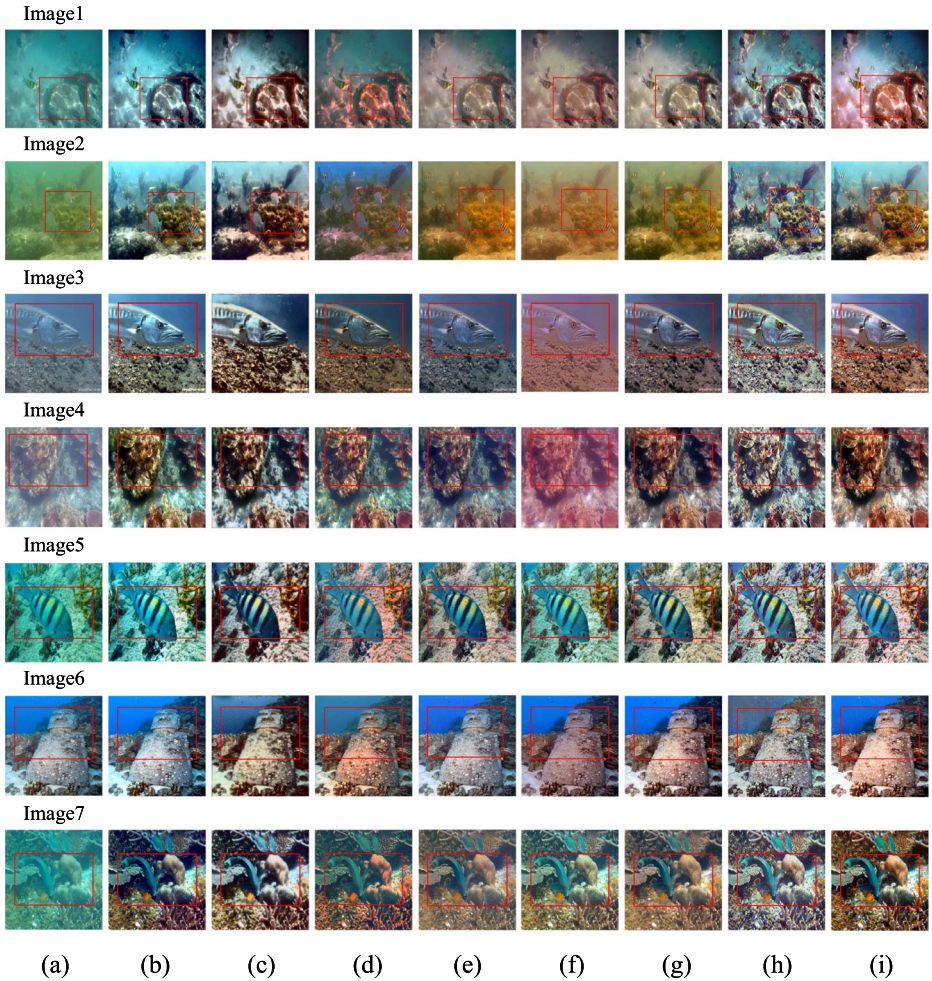


Fig. 9 Visual comparison results with red boxes which mark the important region. (a) Input, (b) RGHS, (c) DehazeNet&HWD, (d) Water-Net, (e) UResnet, (f) FUInE-GAN, (g) UGAN, (h) GAN-RS, (i) MGF-cGAN

Among the compared methods, the proposed MGF-cGAN outperforms all of them in most cases, as can be seen. Despite some cases where our model is not the best or second-best, its performance is still very close to the top results in the majority of quality metrics. Therefore, it can be concluded that our method outperforms the other models for its superior performance in most quality metrics.

4.3 Specific applications in related fields

To investigate the applicability of MGF-cGAN in related domains, saliency detection [11], SURF key matching test [3] and key point detection [4] are used to perform for underwater visual tasks, as illustrated in Figs. 10, 11 and 12.

Table 1 Statistical results applying SSIM, PSNR and RMSE in UIEB dataset

Methods	SSIM $\uparrow \pm$ Variance \downarrow	PSNR $\uparrow \pm$ Variance \downarrow	RMSE $\downarrow \pm$ Variance \downarrow
RGHS	0.7753 \pm 0.0926	23.3994 \pm 4.8660	19.9955 \pm 10.9782
DehazeNet & HWD	0.7248 \pm 0.0805	19.4252 \pm 2.8684	28.7892 \pm 9.9312
Water-Net	0.8058 \pm 0.0848	22.4686 \pm 4.1820	21.4522 \pm 10.3325
UResnet	0.8545 \pm 0.0623	23.4128 \pm 4.1753	19.2873 \pm 9.5281
FUnIE-GAN	0.7309 \pm 0.1002	21.1926 \pm 3.8942	24.5889 \pm 11.8037
UGAN	0.8120 \pm 0.0531	25.2241 \pm 3.3161	15.0648 \pm 6.2261
GAN-RS	0.6749 \pm 0.0863	18.5851 \pm 2.3539	31.1522 \pm 8.7727
MGF-cGAN	0.8654 \pm 0.0493	24.0672 \pm 2.3020	18.6340 \pm 5.9968

The red and blue figures indicate the best and second-best values, respectively. Variance is the number following the mean

Table 2 Statistical results applying SSIM, PSNR and RMSE in EUVP dataset

Methods	SSIM $\uparrow \pm$ Variance \downarrow	PSNR $\uparrow \pm$ Variance \downarrow	RMSE $\downarrow \pm$ Variance \downarrow
RGHS	0.7385 \pm 0.0680	21.6699 \pm 3.7493	23.0962 \pm 10.3917
DehazeNet & HWD	0.6419 \pm 0.0952	17.8813 \pm 3.0135	34.5297 \pm 12.0481
Water-Net	0.7893 \pm 0.07451	24.3033 \pm 3.8019	17.1410 \pm 8.0754
UResnet	0.7288 \pm 0.0954	19.6279 \pm 3.8939	29.4150 \pm 13.4542
FUnIE-GAN	0.7927 \pm 0.0610	26.2057 \pm 3.0589	13.3546 \pm 5.6165
UGAN	0.7365 \pm 0.0772	21.1100 \pm 3.0747	23.9354 \pm 9.0420
GAN-RS	0.6009 \pm 0.0934	16.6111 \pm 2.4113	39.1675 \pm 11.2318
MGF-cGAN	0.7877 \pm 0.0845	27.1078 \pm 3.0021	11.9437 \pm 4.3490

The red and blue figures indicate the best and second-best values, respectively. Variance is the number following the mean

Table 3 Statistical results applying the UCIQE, UIQM in U45 dataset

Methods	UCIQE $\uparrow \pm$ Variance \downarrow	UIQM $\uparrow \pm$ Variance \downarrow
RGHS	0.6369 \pm 0.0376	2.5688 \pm 0.7354
DehazeNet & HWD	0.6150 \pm 0.0226	3.2113 \pm 0.2506
Water-Net	0.6132 \pm 0.0315	3.3925 \pm 0.2412
UResnet	0.6024 \pm 0.0257	3.1916 \pm 0.2539
FUnIE-GAN	0.5563 \pm 0.0568	3.0543 \pm 0.4917
UGAN	0.6167 \pm 0.0346	3.4534 \pm 0.1555
GAN-RS	0.5765 \pm 0.0213	3.4274 \pm 0.2864
MGF-cGAN	0.6343 \pm 0.0322	3.5053 \pm 0.2370

The red and blue figures indicate the best and second-best values, respectively. Variance is the number following the mean

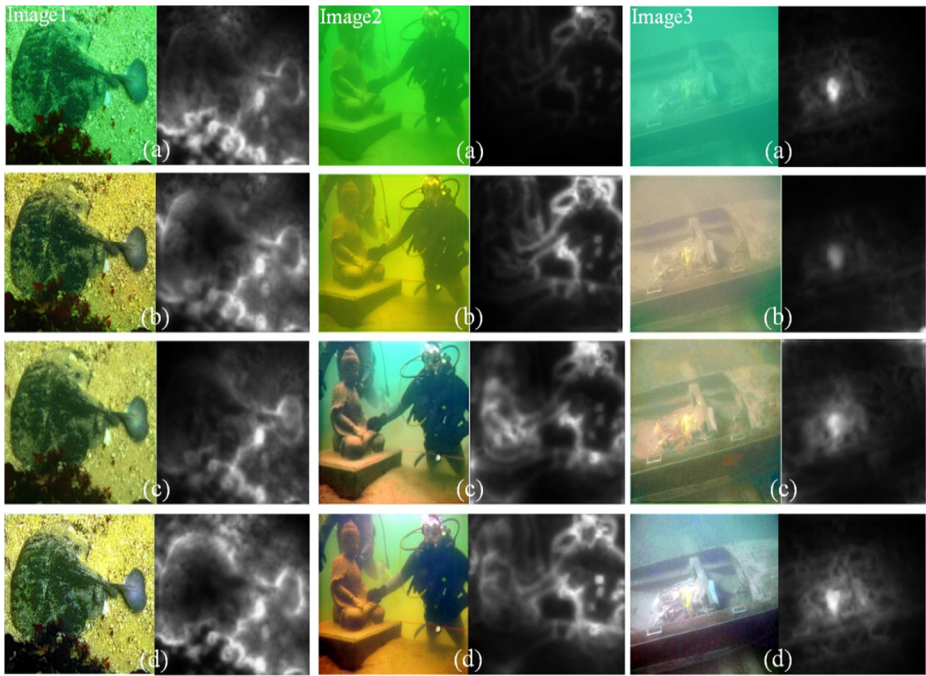


Fig. 10 Comparison of saliency detection, (a) Input, (b) FUNIE-GAN, (c) UResnet, (d) MGF-cGAN

In Fig. 10, the FUNIE-GAN and MGF-cGAN are shown to be able to enhance the edge of the fish in Image1. The UResnet and MGF-cGAN can define the contour features of the portrait and diver in Image2. Compared with other algorithms, the details of the wooden boat are clearer when we use MGF-cGAN. In Fig. 11, the proposed method can match more keypoints. And the diver image processed by our algorithm can detect more points to describe the human body, in Fig. 12. Therefore, it can be proved that MGF-cGAN has good applicability.

5 Ablation study

As our proposed GFM is based on several essential components from different P-sns in Ms-FEM, we use ablation studies to analyze the main components. More specifically, we denote the abbreviations as below, Table 4 and Fig. 13:

- 1) M1: Our proposed network with ϕ_{f1} .
- 2) M2: Our proposed network with ϕ_{f2} .
- 3) M3: Our proposed network with ϕ_{f3} .
- 4) M4: Our proposed network with ϕ_{f1} and ϕ_{f2} .
- 5) M5: Our proposed network with ϕ_{f2} and ϕ_{f3} .
- 6) M6: Our proposed network with ϕ_{f1} and ϕ_{f3} .
- 7) **Proposed**: Our proposed network with ϕ_{f1} , ϕ_{f2} and ϕ_{f3} .

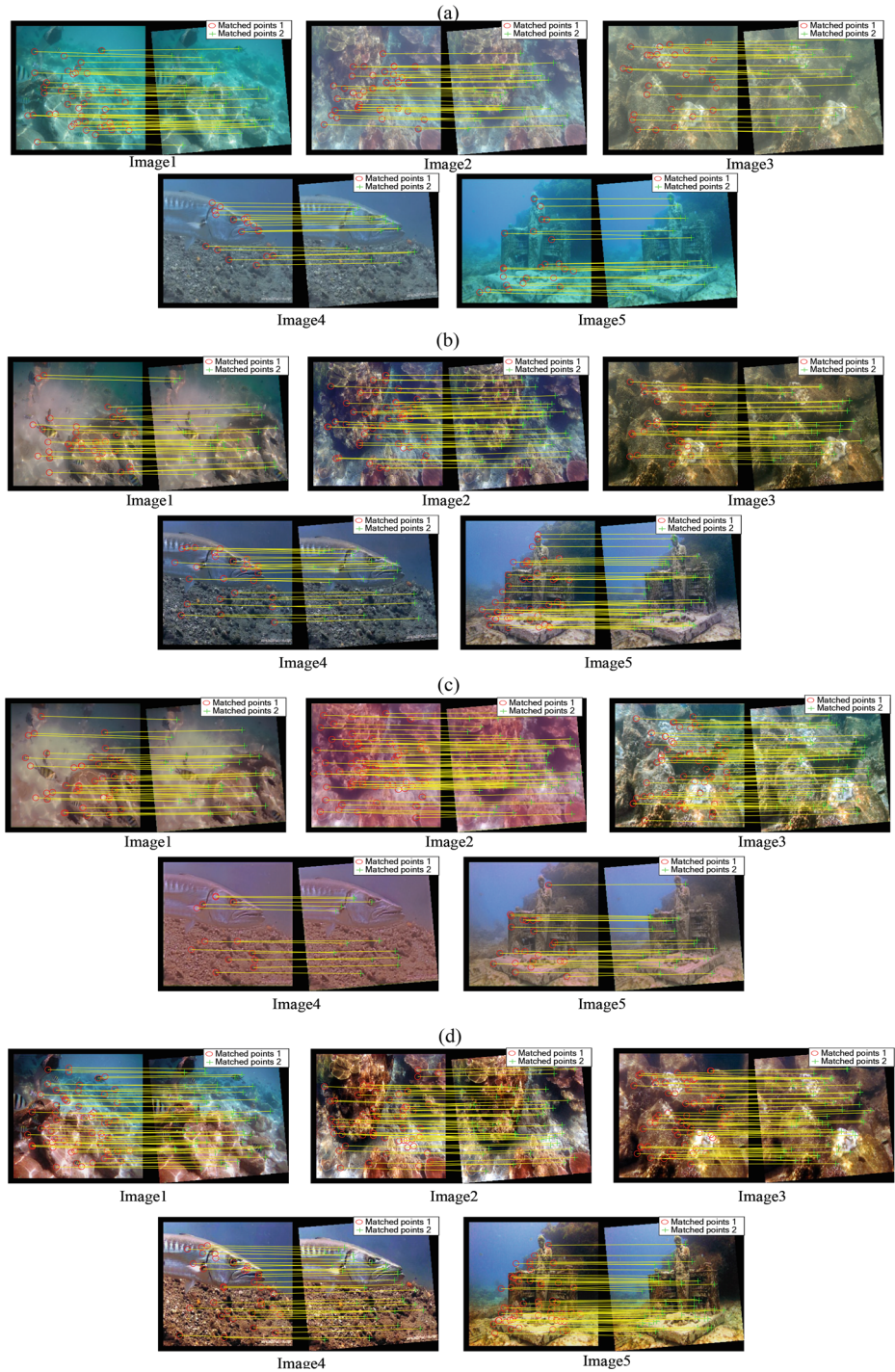


Fig. 11 Comparison of SURF key matching test, (a) Input, (b) FUNIE-GAN, (c) UResnet, (d) MGF-cGAN

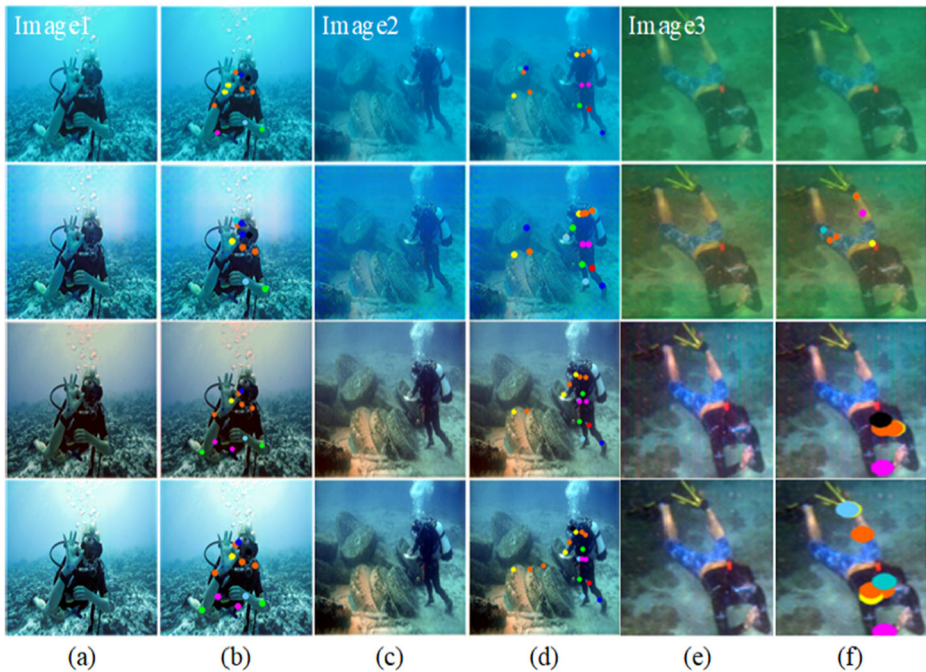


Fig. 12 Comparison of key point detection, (a) Original underwater images, (b) Results. From the first row to the last row are Input, FUNIE-GAN, UResnet, MGF-cGAN, respectively

The ablation study experiments are carried out on the EUVP-testset with the same settings as in Section 4. We set $ReGB = 2$, as proved by Section 3.4. Quantitative evaluations are reported in Fig. 14.

Each component of our proposed method is valuable, as can be shown in Fig. 14. Especially, M4 and M5, i.e., our proposed network without ϕ_{f3} and our proposed network without ϕ_{f1} , have much lower performance, which proves the importance of Multiscale information from different P-sns playing an essential role in our method, although the variance in M2 is good. Furthermore, we present a graphic illustration as a comparison of the effectiveness of every component in Fig. 15. We can obviously see that the outputs of our proposed network have superior visual perception than other results.

Table 4 The various components of our proposed network

Components	M1	M2	M3	M4	M5	M6	Proposed
ϕ_{f1}	✓			✓		✓	✓
ϕ_{f2}		✓		✓	✓		✓
ϕ_{f3}			✓		✓	✓	✓

The ✓ sign denotes the use of the corresponding component

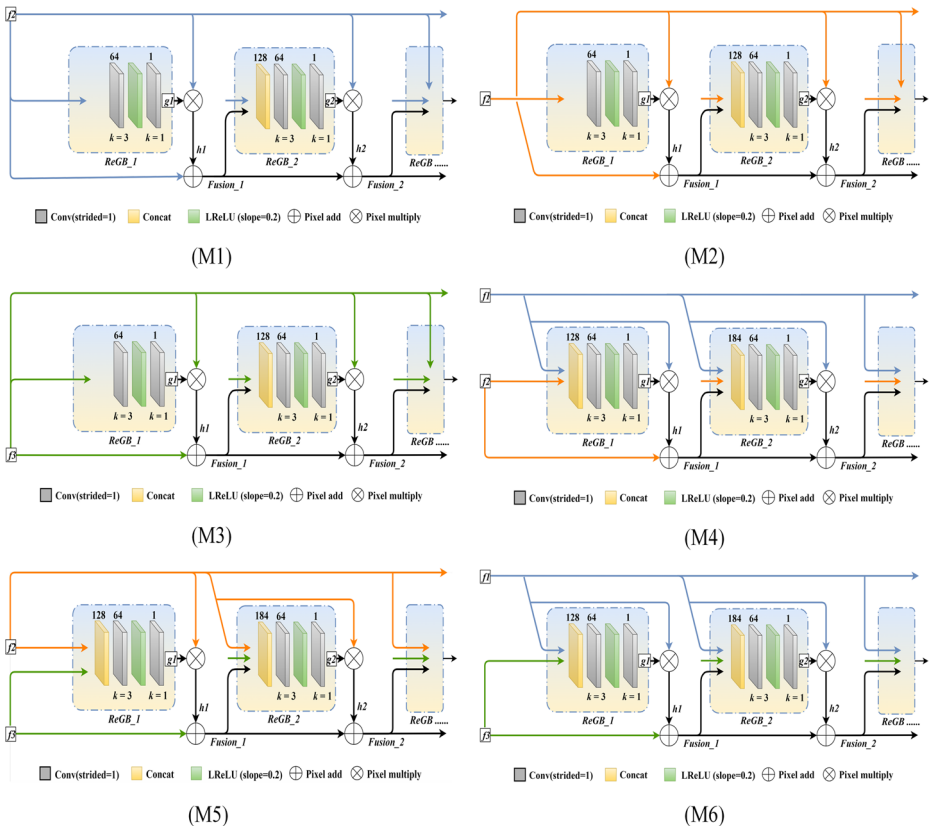


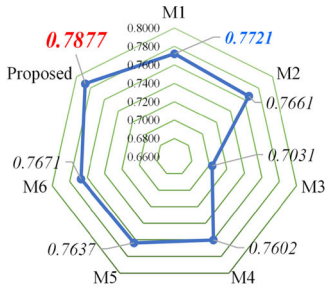
Fig. 13 GFM architecture of different components in Table 4

6 Discussion

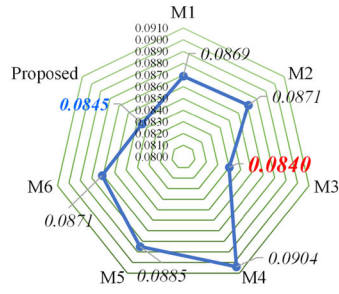
For the experiment, we first conduct a subjective evaluation. By comparing the proposed method with classical and SOTA approaches, we find that it can correct color in underwater images and enhance the details. We then perform an objective evaluation, and conclude that the MGF-cGAN is better than other models in terms of various image quality metrics. Furthermore, enhanced results of MGF-cGAN also demonstrate outstanding performance in underwater saliency detection, the SURF key matching test, and other visual perception tasks. Finally, we compare the effect of various architecture in the ablation study. However, since the generator of MGF-cGAN is complex and the discriminator is simple, the two parts cannot be balanced, which easily causes the discriminator to be unstable. To strengthen the discriminator, we are considering the multi-branch discriminator in future research.

7 Conclusion

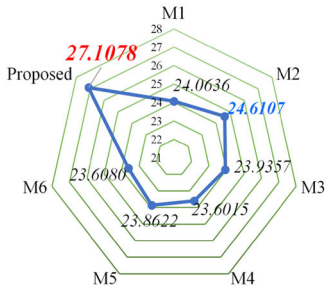
In this paper, we propose the MGF-cGAN to overcome the challenges of color cast, texture details, and blurred edges. The model uses conditional generative adversarial networks. We



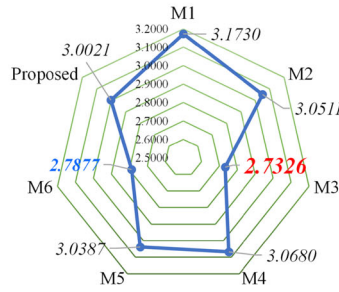
(SSIM)



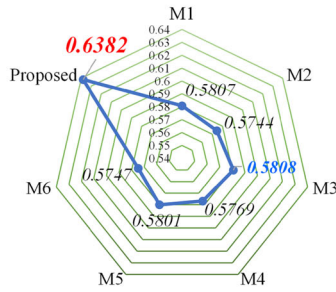
(SSIM-Variance)



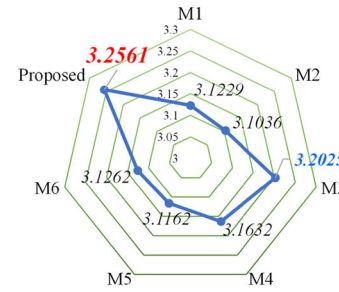
(PSNR)



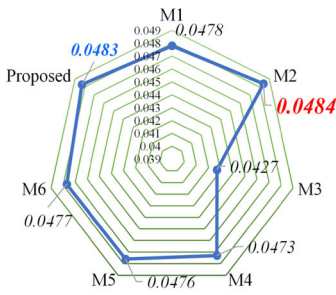
(PSNR-Variance)



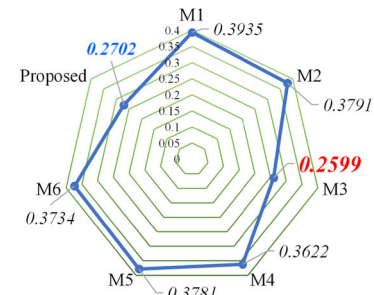
(UCIQE)



(UCIQE-Variance)



(UIQM)



(UIQM-Variance)

Fig. 14 Evaluation of ablation study, the red and blue figures indicate the best and second-best values, respectively

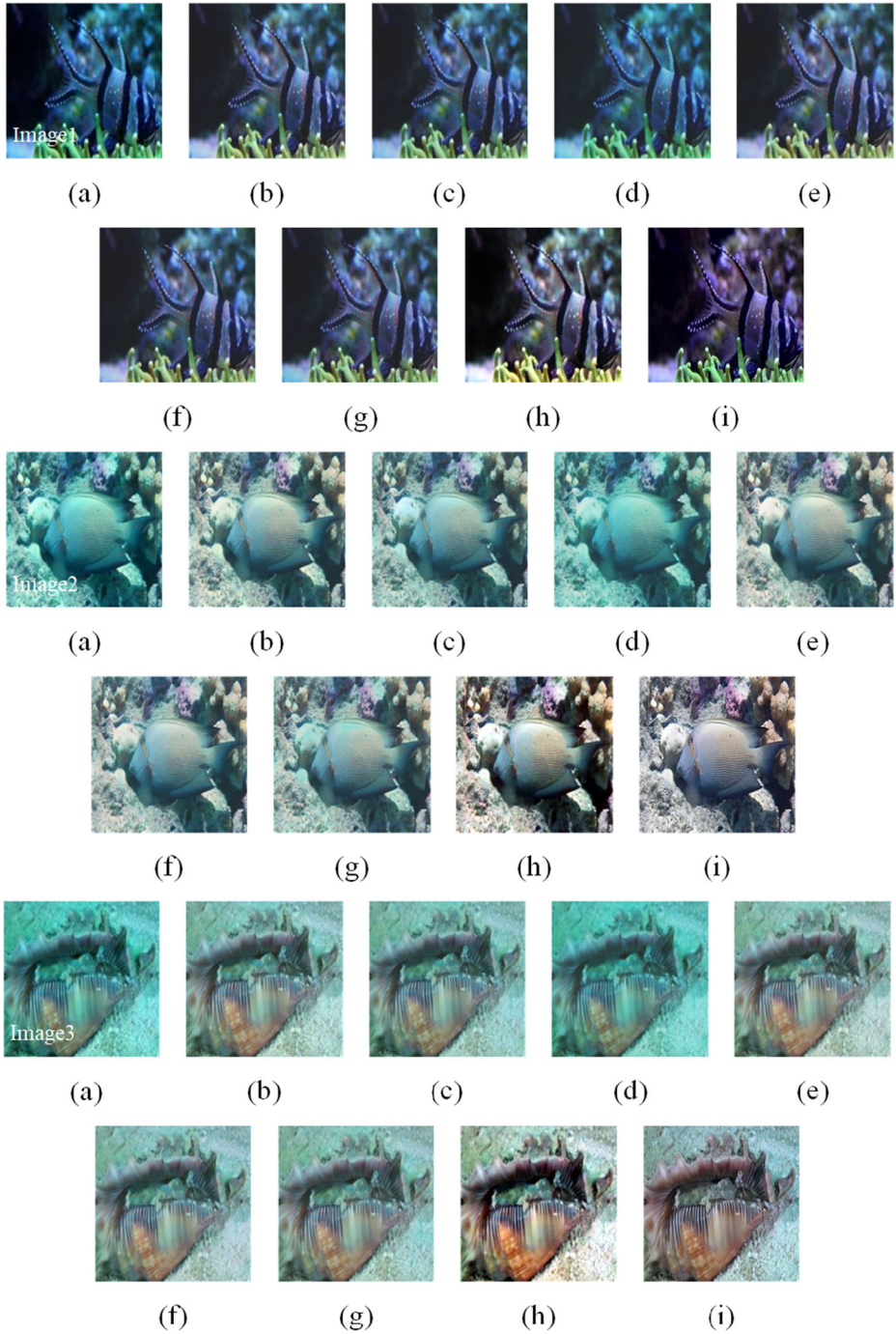


Fig. 15 Visual examples of ablation study comparisons of our proposed method, (a) Input, (b) M1, (c) M2, (d) M3, (e) M4, (f) M5, (g) M6, (h) Proposed, (i) Ground truth

design a Ms-FEM to extract the features of the input via three P-sns, which extract richer features than a single branch. Moreover, we propose a GFM for an adaptive combination of the three branches. Compared to other fusion operations, GFM results in better chromaticity, sharpness, and contrast. Extensive experiments performed on three benchmark underwater image datasets indicated that the proposed MGF-cGAN is superior to classical and SOTA methods for enhancing underwater images, and it can be applied to some underwater visual tasks.

Acknowledgements This work is supported by the National Key Research and Development Program of China under Grant (No. 2018YFB1403303), the Key Research and Development Program of Liaoning Province under Grant (No. 2019JH2/10100014).

Author Contributions Xu Liu implemented the algorithm, performed the experiments, and wrote the manuscript. Sen Lin revised the paper and provided funding support. Zhiyong Tao provided funding support.

Data Availability The authors confirm that the data generated or analyzed and supporting the findings of this study are available within the article.

Declarations

Conflict of Interests The authors declare no conflict of interest.

References

1. Ahn J, Yasukawa S, Sonoda T et al (2018) An optical image transmission system for deep sea creature sampling missions using autonomous underwater vehicle. *IEEE J Ocean Eng* 45:350–361
2. Anwar S, Li C (2020) Diving deeper into underwater image enhancement: a survey. *Signal Process Image Commun* 89:115978
3. Bay H, Ess A, Tuytelaars T, Van Gool L (2008) Speeded-up robust features (SURF). *Comput Vis image Underst* 110:346–359
4. Cao Z, Hidalgo G, Simon T et al (2019) Openpose: realtime multi-person 2D pose estimation using part affinity fields. *IEEE Trans Pattern Anal Mach Intell* 43:172–186
5. Chen X, Yu J, Kong S et al (2019) Towards real-time advancement of underwater visual quality with GAN. *IEEE Trans Ind Electron* 66:9350–9359
6. Chen X, Yu J, Kong S et al (2019) Towards real-time advancement of underwater visual quality with GAN. *IEEE Trans Ind Electron* 66:9350–9359
7. Fan G-F, Yu M, Dong S-Q et al (2021) Forecasting short-term electricity load using hybrid support vector regression with grey catastrophe and random forest modeling. *Util Policy* 73:101294
8. Fazlali H, Shirani S, McDonald M et al (2020) Aerial image dehazing using a deep convolutional autoencoder. *Multimed Tools Appl* 79:29493–29511
9. Feng X, Li J, Hua Z (2020) Low-light image enhancement algorithm based on an atmospheric physical model. *Multimed Tools Appl* 79:32973–32997
10. Fu X, Cao X (2020) Underwater image enhancement with global–local networks and compressed-histogram equalization. *Signal Process Image Commun* 86:115892
11. Goferman S, Zelnik-Manor L, Tal A (2011) Context-aware saliency detection. *IEEE Trans Pattern Anal Mach Intell* 34:1915–1926
12. Gulrajani I, Ahmed F, Arjovsky M et al (2017) Improved training of Wasserstein GANs. In: *Proceedings of the international conference on neural information processing systems*, pp 5769–5779
13. Guo JC, Li CY, Guo CL, Chen SJ (2017) Research progress of underwater image enhancement and restoration methods. *J Image Graph* 22:273–287
14. Guo Y, Li H, Zhuang P (2019) Underwater image enhancement using a multiscale dense generative adversarial network. *IEEE J Ocean Eng* 45:862–870
15. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: *Proceedings of the IEEE international conference on computer vision*, pp 1026–1034

16. Huang K-Q, Wang Q, Wu Z-Y (2006) Natural color image enhancement and evaluation algorithm based on human visual system. *Comput Vis Image Underst* 103:52–63
17. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 4700–4708
18. Huang D, Wang Y, Song W et al (2018) Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition. In: *Proceedings of the international conference on multimedia modeling*, pp 453–465
19. Huang Y, Liu M, Yuan F (2021) Color correction and restoration based on multi-scale recursive network for underwater optical image. *Signal Process Image Commun* 93:116174
20. Huynh-Thu Q, Ghanbari M (2008) Scope of validity of PSNR in image/video quality assessment. *Electron Lett* 44:800–801
21. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the international conference on machine learning*, pp 448–456
22. Islam MJ, Xia Y, Sattar J (2020) Fast underwater image enhancement for improved visual perception. *IEEE Robot Autom Lett* 5:3227–3234
23. Isola P, Zhu J-Y, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1125–1134
24. Jaffe JS (1990) Computer modeling and the design of optimal underwater imaging systems. *IEEE J Ocean Eng* 15:101–111
25. Li S, Liu f, Wei J (2022), Underwater image restoration based on exponentiated mean local variance and extrinsic prior. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-021-11269-1>
26. Li R, Pan J, Li Z, Tang J (2018) Single image dehazing via conditional generative adversarial network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 8202–8211
27. Li C, Anwar S, Porikli F (2020) Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognit* 98:107038
28. Li C, Guo C, Ren W et al (2020) An underwater image enhancement benchmark dataset and beyond. *IEEE Trans Image Process* 29:4376–4389
29. Li C, Anwar S, Hou J et al (2021) Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Trans Image Process* 30:4985–5000
30. Li H, Zhuang P (2021) Dewaternet: a fusion adversarial real underwater image enhancement network. *Signal Process Image Commun* 95:116248
31. Liu P, Wang G, Qi H et al (2019) Underwater image enhancement with a deep residual framework. *IEEE Access* 7:94614–94629
32. McGlamery BL (1980) A computer model for underwater camera systems. In: *Ocean optics VI. International society for optics and photonics*, pp 221–231
33. Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv: [1411.1784](https://arxiv.org/abs/1411.1784)
34. Naga Srinivasu P, Balas VE (2021) Performance measurement of various hybridized kernels for noise normalization and enhancement in high-resolution MR images. *Bio-inspired Neurocomputing*. Springer, Berlin, pp 1–24
35. Pan PW, Yuan F, Cheng E (2018) Underwater image de-scattering and enhancing using dehazenet and HWD. *J Mar Sci Technol* 26:531–540
36. Panetta K, Gao C, Agaian S (2015) Human-visual-system-inspired underwater image quality measures. *IEEE J Ocean Eng* 41:541–551
37. Rahman Z, Jobson DJ, Woodell GA (2004) Retinex processing for automatic image enhancement. *J Electron Imaging* 13:100–110
38. Ren W, Ma L, Zhang J et al (2018) Gated fusion network for single image dehazing. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 3253–3261
39. Sharma T, Agrawal I, Verma NK (2020) CSIDNet: compact single image dehazing network for outdoor scene enhancement. *Multimed Tools Appl* 79:30769–30784
40. Shen P, Zhang L, Wang M, Yin G (2021) Deeper super-resolution generative adversarial network with gradient penalty for sonar image enhancement. *Multimed Tools Appl* 80:28087–28107
41. Sun X, Liu L, Li Q et al (2019) Deep pixel-to-pixel network for underwater image enhancement and restoration. *IET Image Process* 13:469–474
42. Tsai D-Y, Lee Y, Matsuyama E (2008) Information entropy measure for evaluation of image quality. *J Digit Imaging* 21:338–347
43. Wang T-C, Liu M-Y, Zhu J-Y et al (2018) High-resolution image synthesis and semantic manipulation with conditional gans. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 8798–880

44. Wang Y, Guo J, Gao H, Yue H (2021) UIEC²-Net: CNN-based underwater image enhancement using two color space. *Signal Process Image Commun* 96:116250
45. Wang Y, Yin S, Basu A (2021) A multi-scale attentive recurrent network for image dehazing. *Multimed Tools Appl* 80:32539–32565
46. Wang Y, Zhang J, Cao Y, Wang Z (2017) A deep CNN method for underwater image enhancement. In: *IEEE International conference on image processing*, pp 1382–1386
47. Wu Q, Guo Y, Hou J et al (2021) Underwater optical image processing based on double threshold judgements and optimized red dark channel prior method. *Multimed Tools Appl* 80:29985–30002
48. Yang M, Sowmya A (2015) An underwater color image quality evaluation metric. *IEEE Trans Image Process* 24:6062–6071
49. Yang M, Hu K, Du Y et al (2020) Underwater image enhancement based on conditional generative adversarial network. *Signal Process Image Commun* 81:115723
50. Yu X, Qu Y, Hong M (2018) Underwater-GAN: underwater image restoration via conditional generative adversarial network. In: *Proceedings of the international conference on pattern recognition*, pp 66–75
51. Yuan Q, Li J, Zhang L et al (2020) Blind motion deblurring with cycle generative adversarial networks. *Vis Comput* 36:1591–1601
52. Z Wang A, C Bovik H, Sheikh R, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans image Process* 13:600–612
53. Zhang H, Sindagi V, Patel VM (2019) Image de-raining using a conditional generative adversarial network. *IEEE Trans Circ Syst Video Technol* 30:3943–3956
54. Zhao H, Gallo O, Frosio I, Kautz J (2016) Loss functions for image restoration with neural networks. *IEEE Trans Comput imaging* 3:47–57
55. Zhou J, Yao J, Zhang w, Zhang D (2021) Multi-scale retinex-based adaptive gray-scale transformation method for underwater image enhancement. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-021-11327-8>
56. Zong X, Chen Z, Wang D (2021) Local-cycleGAN: a general end-to-end network for visual enhancement in complex deep-water environment. *Appl Intell* 51:1947–1958

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.