



OPEN ACCESS

EDITED BY

Xuemin Cheng,
Tsinghua University, China

REVIEWED BY

Lina Zhou,
Hong Kong Polytechnic University,
Hong Kong SAR, China
Ning Wang,
Dalian Maritime University, China

*CORRESPONDENCE

Guoqiang Zhong
✉ gqzhong@ouc.edu.cn
Dong Wang
✉ wangdong@ouc.edu.cn

RECEIVED 09 February 2023

ACCEPTED 29 May 2023

PUBLISHED 21 June 2023

CITATION

Wang H, Zhong G, Sun J, Chen Y, Zhao Y,
Li S and Wang D (2023) Simultaneous
restoration and super-resolution GAN for
underwater image enhancement.
Front. Mar. Sci. 10:1162295.
doi: 10.3389/fmars.2023.1162295

COPYRIGHT

© 2023 Wang, Zhong, Sun, Chen, Zhao, Li
and Wang. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Simultaneous restoration and super-resolution GAN for underwater image enhancement

Huiqiang Wang, Guoqiang Zhong*, Jinxuan Sun, Yang Chen,
Yuxiao Zhao, Shu Li and Dong Wang*

College of Computer Science and Technology, Ocean University of China, Qingdao, China

Underwater images are generally of low quality, limiting the performance of subsequent perceptual tasks, such as underwater object detection and recognition. However, only a few methods can improve the quality of underwater images by simultaneously restoring and super-resolving underwater images. In this paper, we propose an end-to-end trainable model based on generative adversarial networks (GANs) called Simultaneous Restoration and Super-Resolution GAN (SRSRGAN) to obtain clear super-resolution underwater images automatically. In particular, our model leverages a cascaded architecture with two stages of carefully designed generative adversarial networks to restore and super-resolve corrupted underwater images in a coarse-to-fine manner. The major advantages of SRSRGAN are twofold. First, it is a unified solution that can simultaneously restore and super-resolve images. Second, SRSRGAN is not limited by the prior experience of the types and levels of underwater degraded images but can perform the inference using only observed corrupted images. These two advantages enable SRSRGAN to enjoy better flexibility and higher practicability in realistic underwater scenarios. Extensive experimental results demonstrate the superiority of SRSRGAN in underwater image restoration, super-resolution, and simultaneous restoration and super-resolution.

KEYWORDS

image enhancement, generative adversarial network, simultaneous restoration and super-resolution, deep learning, underwater images

1 Introduction

With 70% of the earth's surface covered by water, there is great potential for exploiting underwater resources. The underwater environment offers numerous valuable resources, such as marine biology, mineral resources, and tidal energy. However, there is a wide gap between the plentiful marine resources and their exploitation. To this end, various kinds of methods have been proposed to obtain information about the underwater environment to promote the use of marine resources. Among others, a crucial way to obtain information from the underwater environment is image understanding, while the images captured in

realistic underwater scenarios usually have severe defects, such as blurriness, noise, and color distortion (Soni and Kumare, 2020).

Specifically, underwater image defects are caused by various factors. Light rays exponentially decay as the underwater depth increases, which makes underwater images of low contrast and darkness (Ancuti et al., 2018). Furthermore, lights of different colors have different absorption rates underwater, depending on the wavelengths, which results in color distortion of underwater images (Chiang and Chen, 2012). In addition, bubbles and suspended particles in water may cause noise in underwater images (Lu et al., 2017b). Hence, the poor visibility of underwater images seriously affects the exploration of the underwater environment. On the other hand, high-resolution images are essential in many realistic underwater applications, such as marine animal recognition (Chen et al., 2021; Wang et al., 2023b), seabed detection, and deep ocean resources exploration (Lu et al., 2017a). Therefore, the critical tasks for underwater image enhancement are eliminating defects and obtaining super-resolution (SR) images.

To the best of our knowledge, only a few approaches can simultaneously restore and super-resolve underwater images. In particular, Cheng et al. (2018) propose a method that restores underwater images by the white balance (Liu et al., 1995) with the contrast limited adaptive histogram equalization (CLAHE) (Reza, 2004) and super-resolves the restored image by a super-resolution generative adversarial network (GAN) (Ledig et al., 2017). However, due to the fact that this method only utilizes traditional color correction as a preprocessing step for the input image of the super-resolution model during the restoration stage, it limits its ability to remove other types of degradation features. Recently, Islam et al.

(2020a) also introduce an approach to learning enhanced super-resolution underwater images. However, their proposed method is limited in its ability to model complex degradation features due to its lack of consideration for capturing multi-scale features in the network architecture design. Due to these limitations, these methods are difficult to generate high-fidelity and high-quality super-resolution images in underwater image enhancement in real-world scenarios.

To address the above issues, we propose simultaneous restoration and super-resolution GAN (SRSRGAN) to obtain underwater images of high visual quality, which is an end-to-end trainable model based on GAN. With a two-stage design, SRSRGAN captures underwater degradation information and fine-grained high-frequency information in the restoration stage and the super-resolution stage, respectively. In the restoration stage, benefiting from the superior structure of the proposed multi-level degradation restoration generator (MLDRG), our model leverages degradation information among different scales, positions, and channels to transform degraded images to clean images. In the super-resolution stage, the high frequency learning module (HFLM) excavates fine-grained high-frequency information to super-resolve clean images. In addition, we adopt a relativistic discriminator to further enhance the quality of our generated underwater images. Thanks to the corporation of the restoration stage and the super-resolution stage, SRSRGAN enjoys two highly expected merits, i.e., i) it provides a unified solution for simultaneous underwater image restoration and super-resolution reconstruction; ii) it is free from the prior of the underwater corruption types and ratios. Extensive experimental results show that SRSRGAN is superior to the state-of-the-art (SOTA) methods

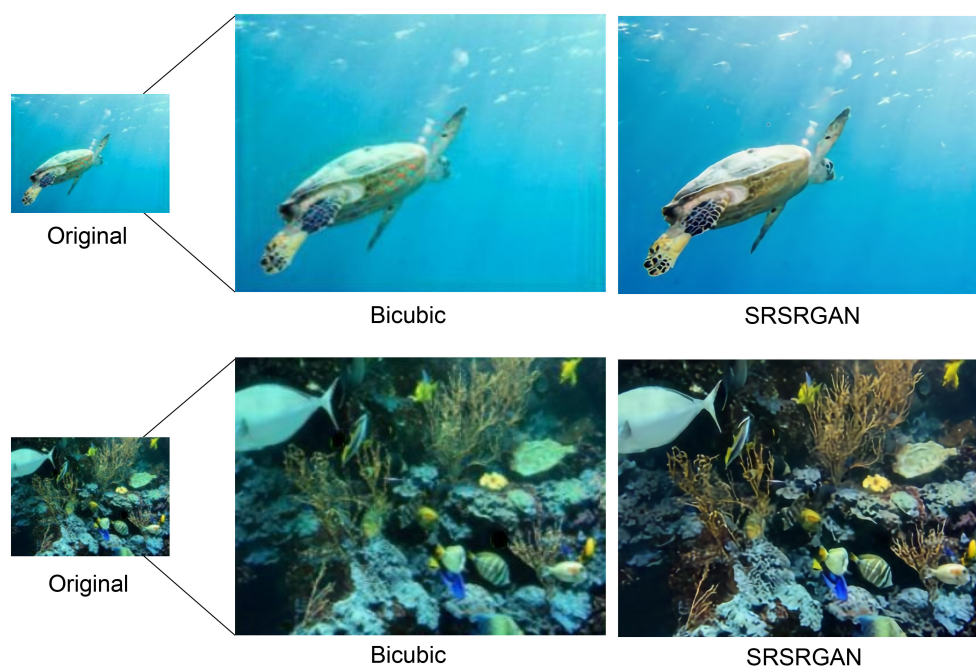


FIGURE 1

The proposed SRSRGAN model provides realistic underwater image enhancement results through an effective inference. The first and second rows, from left to right, are the original underwater image, the bicubic interpolation enhanced image, and the image enhanced by our method.

in underwater image restoration/enhancement, single image super-resolution (SISR), and simultaneous restoration and super-resolution. The qualitative enhancement effect of SRSRGAN is shown in [Figure 1](#).

In summary, our main contributions are as follows:

- We propose a new underwater image enhancement model for simultaneous restoration and super-resolution of underwater images, called SRSRGAN, which does not require any prior image degradation information to perform inference in advance. Therefore, it can flexibly deal with complex underwater scenarios.
- We design a two-stage framework to process underwater images. In the restoration stage, the proposed MLDRG leverages degradation information from different scales, positions, and channels to transform degraded underwater images to clean images. Moreover, the super-resolution stage is presented to enhance the representational ability of high-frequency features for underwater image super-resolution. In addition, this staged design improves the flexibility of the network model.
- Qualitative and quantitative comparisons among SRSRGAN, underwater image restoration/enhancement methods, SISR methods, and existing simultaneous restoration and super-resolution methods show the superiority of SRSRGAN.

2 Related work

In the existing work, few methods are available for simultaneous restoration and super-resolution of underwater images, except the ones mentioned in Section 1 ([Ledig et al., 2017](#); [Cheng et al., 2018](#)). Therefore, in this section, we mainly review the research progress of underwater image restoration/enhancement and SISR.

2.1 Underwater image restoration/enhancement

Traditional underwater image restoration/enhancement algorithms aim to recover a clean image from the degraded observation, including automatic white balance ([Liu et al., 1995](#)), histogram equalization ([Hummel, 1977](#)), and CLAHE ([Reza, 2004](#)). Although these methods improve the quality of underwater images to a certain extent, there are still various problems, such as color deviation, artificial artifacts, and noise amplification. Inspired by the morphology and function of the teleost fish retina, [Gao et al. \(2019\)](#) propose an underwater image enhancement model to solve the problems of blurring and nonuniform color biasing in underwater images. Moreover, several methods are proposed inspired by the dark channel prior ([He et al., 2011](#)). Particularly, [Drews et al. \(2013\)](#) consider underwater images' blue and green channels as underwater visual information sources, and apply a dark channel method to process underwater visual information. [Galdran et al. \(2015\)](#)

propose a dark channel variant called the red channel method to restore the lost contrast and colors associated to short wavelengths in underwater images. Recently, [Li et al. \(2022\)](#) propose a framework called ACCE-D that uses multiple filters and adaptive color and contrast enhancement strategies to enhance underwater images. In addition, [Alenezi et al. \(2022\)](#) propose a method to enhance underwater images by estimating global background light and transmission maps. However, these methods have a common limitation in that the prior assumptions may be invalid with the changes in environmental status.

As convolutional neural networks develop rapidly, some deep networks are used to establish mapping relationships from an underwater image to the clear one ([Hou et al., 2018](#); [Lu et al., 2018](#)). In particular, [Li et al. \(2020\)](#) give an overview of the previous work for underwater image restoration and establish a CNN model named Water-Net to get restored underwater images. Additionally, the emergence of GAN ([Goodfellow et al., 2014](#)) provides more chances for underwater image restoration. For example, [Li et al. \(2018\)](#) propose WaterGAN to generate underwater images from in-air images, which uses two fully convolutional networks to estimate the depth of the generated underwater images and correct their color, respectively. Different from it, UGAN uses two GAN-based models for underwater image generation and color correction, respectively ([Fabbri et al., 2018](#)). In recent work, Underwater GAN ([Yu et al., 2019](#)) uses Wasserstein GAN-GP ([Gulrajani et al., 2017](#)) as the network's backbone for underwater image restoration. Additionally, [Guo et al. \(2020\)](#) propose a multi-scale dense GAN (UWGAN) for underwater image enhancement. These methods improve the quality of underwater images to a certain extent. However, they only focus on restoring the color contrast and color distortion of underwater images and do not further improve the image quality by improving the image resolution. [Liu et al. \(2022\)](#) propose a twin adversarial contrastive learning method to enhance the visual quality of underwater images. Many previous underwater image enhancement methods have only focused on restoring the color contrast and color distortion of underwater images. However, their method has limited ability to remove noise in underwater images. Therefore, [Wang et al. \(2023a\)](#) propose an end-to-end underwater attention generative adversarial network to alleviate the influence of underwater noise problem. These methods improve the quality of underwater images to some extent.

2.2 Single image super-resolution

In some early survey papers on image super-resolution ([Nasrollahi and Moeslund, 2014](#); [Köhler et al., 2017](#); [Yang et al., 2019](#)), there are two principal categories of image super-resolution: multiple image super-resolution (MISR) ([Tsai, 1984](#); [Capel and Zisserman, 2001](#); [Caner et al., 2003](#); [Farsiu et al., 2004](#); [Harmeling et al., 2010](#)) and SISR ([Storkey, 2002](#); [Lian, 2006](#); [Yang et al., 2008](#); [Yang et al., 2010](#); [Dong et al., 2016](#)). Here, we mainly introduce SISR, as the number of underwater images is still very small in general.

Interpolation-based SR methods are typically used to increase the resolution of an image, such as bicubic interpolation ([Keys,](#)

1982) and Lanczos filtering (Duchon, 1979). The reconstructed edges are generally blurred in the super-resolution images obtained by these methods. These methods obtain the SR image, while the reconstructed edges are generally blurry. Subsequent methods focus on matching the edges of the low-resolution (LR) and high-resolution (HR) images (Li and Orchard, 2000; Muresan, 2005). However, the HR images they generate still suffer from blurring and artifacts.

Sparse representation SR methods are based on the sparse signal representation and compressed sensing theory. Yang et al. (Yang et al., 2008; Yang et al., 2010) train two dictionaries for LR and HR patches jointly. They consider the sparse representations of LR and HR images and utilize the sparse representations of the LR images to obtain the HR images. Moreover, the natural image prior framework is added to guide the sparse representation SR method (Kim and Kwon, 2010). However, such an SR method based on the sparse representation needs a long time to train the sparse coding dictionary. More recently, Timofte et al. (Timofte et al., 2013; Timofte et al., 2014) improve the efficiency of sparse representation SR methods using anchored neighborhood regression on the LR patch in the dictionary. Nevertheless, the texture details are generally absent from the generated SR images.

With the rapid development of deep learning, many SR methods based on deep learning have emerged and achieved excellent performance in recent years. Dong et al. (Dong et al., 2014; Dong et al., 2016) propose a fully convolutional network to establish a mapping between the LR and HR images, which has great superiority over the previous approaches. Later, Shi et al. (2016) propose a sub-pixel convolutional neural network, which expands the channels of output features by the convolutional layers and then rearranges the tensor to obtain the HR images. With the depth of neural networks increasing, Kim et al. (2016) use a deep neural network similar to VGG-net to generate SR images. In addition, some researchers propose to utilize the residual networks to achieve an excellent SR effect (Lim et al., 2017; Zhang et al., 2018b; Zhang et al., 2018c; Chen et al., 2019). Furthermore, with the flourishing of GAN-based models, recent

work has shown great success in SISR. Ledig et al. (2017) propose a super-resolution generative adversarial network (SRGAN) to recover SR images from LR images. Wang et al. (2018) enhance SRGAN by modifying the generator with residual-in-residual dense blocks, which can generate realistic images with natural textures. Unfortunately, these methods are only suitable for images taken in the air but cannot perform well on underwater images.

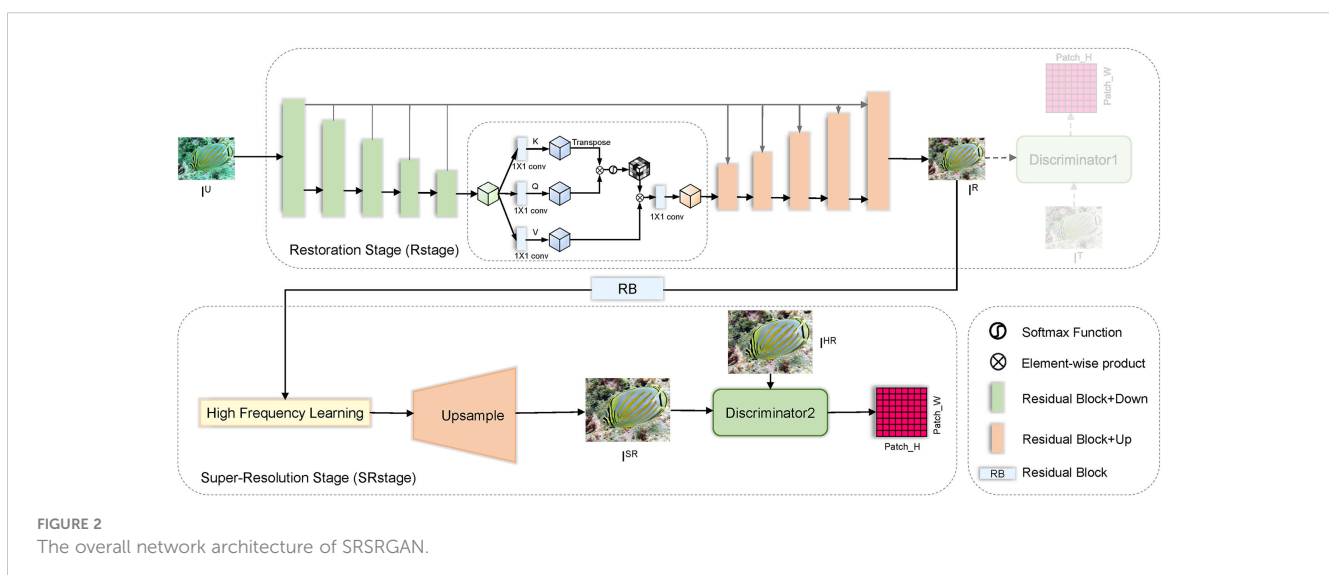
At present, a few researchers are involved in the field of underwater image super-resolution. Particularly, Lu et al. (2017a) propose a two-step method for underwater image super-resolution. Firstly, they obtain a scattered HR image and a descattered HR image by self-similarity SR methods; secondly, they fuse the two HR images to obtain the final image. More recently, Islam et al. (2020b) propose a fully convolutional neural network using residual learning for underwater SISR, called super-resolution using deep residual multipliers (SRDRM). In addition, they also formulate an adversarial training pipeline (SRDRM-GAN). However, the generated images by these methods have limited image quality and visual perception. In this paper, we propose an end-to-end trainable GAN-based model called SRSRGAN, which can simultaneously restore and super-resolve underwater images.

3 The proposed model

SRSRGAN aims to build an effective simultaneous restoration and super-resolution model for underwater image enhancement, which obtains an input underwater image I^U and outputs a clear super-resolution image I^{SR} .

In this section, we elaborate on the proposed end-to-end trainable model SRSRGAN, which consists of a restoration stage (Rstage, $f_R(\cdot)$) and a super-resolution stage (SRstage, $f_{SR}(\cdot)$), as shown in Figure 2.

In brief, SRSRGAN first feeds it into $f_R(\cdot)$ to learn the clear image I^R for a given degraded image I^U . Then, I^R is further passed through $f_{SR}(\cdot)$ to obtain the super-resolution image I^{SR} . In the following part, we first illustrate the restoration stage in Section 3.1.



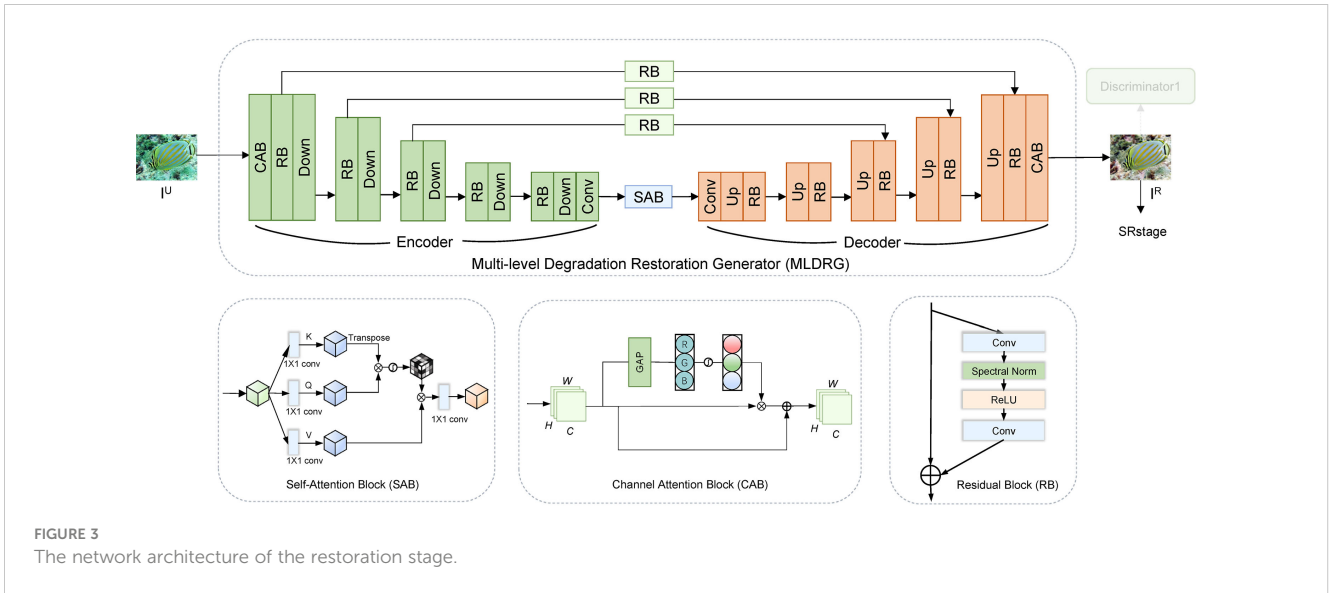


FIGURE 3 The network architecture of the restoration stage.

Then, we describe the super-resolution stage in Section 3.2. Finally, we introduce our end-to-end trainable framework of SRSRGAN in Section 3.3.

3.1 Restoration stage

The light passing through the water attenuates as the depth increases, and the background light will also affect the underwater images. In order to restore a clear image I^R from an underwater image I^U containing noise and distortion, we propose a GAN-based model in the restoration stage. Mathematically,

$$\min_{G_R} \max_{D_R} V(G_R, D_R) = E_{I^T \sim p_{data}(I^T)} [\log D_R(I^T)] + E_{I^U \sim p_G(I^U)} [\log (1 - D_R(G_R(I^U)))] \tag{1}$$

where I^U and I^T denote the image to be restored and the ground-truth image, respectively.

3.1.1 Network architecture

The overall network architecture of our proposed restoration model is shown in Figure 3, which is designed to restore the clear

image I^R from the noisy and distorted underwater image I^U . To make SRSRGAN better deal with underwater degradation, we carefully design a multi-level degradation restoration generator (MLDRG) consisting of an encoder and a decoder.

3.1.1.1 Multi-level degradation restoration generator

Our MLDRG focuses on dealing with noise and color degradation because color degradation and noise degradation are frequent in underwater images.

Specifically, color distortion is an extremely important issue when processing underwater images. Due to the unique properties of the underwater environment, color distortion in underwater images is even more severe, which has an adverse impact on the quality and usability of the images. To address this issue, we utilize the channel attention blocks (CAB) (Hu et al., 2018) to enhance the network’s focus on color information, thereby improving the network’s color restoration capability. To better handle the color restoration of shallow feature colors, we place the CAB in the first and last part of the decoder. The role of the CAB is to enhance the network’s focus on shallow features, thereby improving the network’s color restoration capability. In the first part of the decoder, the CAB can assist the network in better capturing the color information of shallow features, providing a better foundation

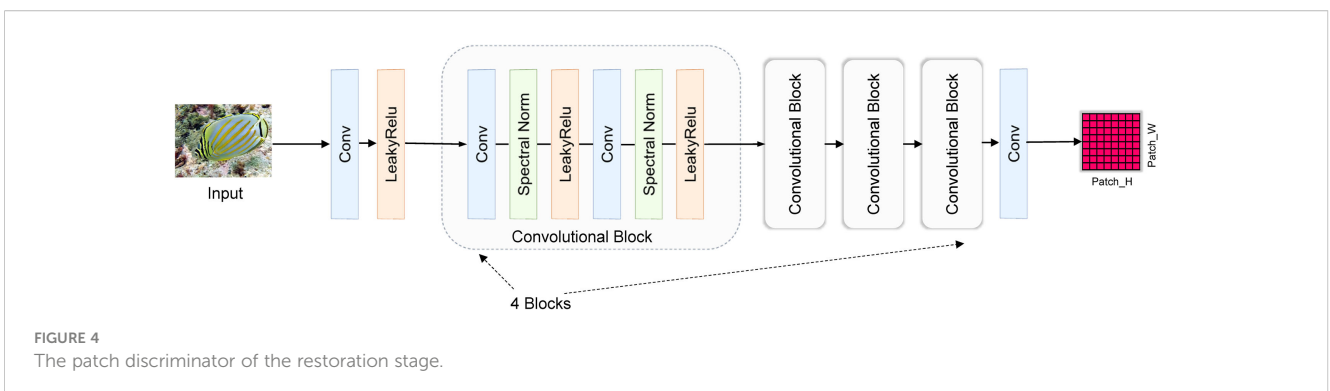


FIGURE 4 The patch discriminator of the restoration stage.

for subsequent color restoration. In the last part of the decoder, the CAB can further enhance the network's focus on shallow features, thereby improving the network's color restoration capability.

On the other hand, due to the random, irregular, and uneven distribution of noise in images, the presence of noise can increase the difficulty of model restoration and negatively impact image quality. Therefore, we implement a self-attention block (SAB) (Vaswani et al., 2017) in front of the decoder to remove the noise with different characteristics in underwater images. Specifically, SAB can determine the weight of each pixel by calculating its similarity with other pixels, allowing the network to focus more on the contextual information related to the current pixel. This can help the network better understand the structural information in the image and remove the random and irregular noise. Hence, the generator has the capacity to better handle the noise of underwater images than that without the attention block. For implementation details, the global feature maps are coded into queries, keys, and values in d_q , d_k , and d_v dimensions, respectively. The attention function is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (2)$$

where Q is a matrix for queries, K and V are matrices for keys and values, respectively.

In addition, we implement the spectral normalization (SN) (Miyato et al., 2018) after each convolutional layer in MLDRG. SN layers control the Lipschitz constant of MLDRG by constraining the spectral norm ($\sigma(W) = 1$) of each layer. In particular, the Lipschitz constant describes the intensity of the output as it changes with the input. For the Lipschitz continuous function ϕ , if it satisfies

$$\frac{\|\phi(x') - \phi(x)\|_2}{\|x' - x\|_2} \leq k, \quad (3)$$

then k ($k \geq 0$) is called the Lipschitz constant. In other words, MLDRG is insensitive to the perturbation of the inputs. Hence, it can better handle noisy underwater images than that without the SN layer. This discovery has the same viewpoint as (Lin et al., 2019) that the Lipschitz continuity is effective for image-denoising tasks.

Ideally, the generator should be able to retain more multi-scale information and spatial context information while providing flexibility for the super-resolution stage. To this end, we employ the residual block with the SN layer to extract features at 5 scales. In order to switch scales in our restoration framework, we use a 3×3 convolutional layer with stride 2 for downsampling and a bilinear interpolation algorithm for upsampling after a 3×3 convolutional layer with stride 1. Considering the intrinsic information loss in downsampling and upsampling, we add a residual block at each scale to fuse useful information from the encoder to the decoder.

3.1.1.2 Restoration discriminator

Figure 4 shows the discriminator of the restoration stage. The discriminator D_R distinguishes the ground-truth image I^T from the generated image I^R at the level of image patches. Specifically, given an input image I_d to be discriminated, it first extracts the shallow

features F_0 of the discriminated image by a 3×3 convolutional layer. Mathematically,

$$F_0 = \text{Conv}(I_d). \quad (4)$$

Since the extraction of deep features is essential to discriminate the images, we design 4 convolutional blocks, each containing a 3×3 convolutional layer with stride 1 and a 3×3 convolutional layer with stride 2. In addition, considering the Lipschitz continuity of the discriminator, we add an SN layer and a Leaky ReLU function after each convolutional layer to stabilize its network training. Therefore, we feed the previously extracted shallow feature F_0 into 4 designed convolutional blocks to further excavate the deep features of the input image. Mathematically,

$$F_i = H_{\text{Block}_i}(F_{i-1}), \quad i = 1, 2, \dots, N, \quad (5)$$

where H_{Block_i} denotes the i -th convolutional block in the path discriminator, and N denotes the number of convolutional blocks. Hence, F_N is the final output of the patch discriminator, and $F_1 \sim F_{N-1}$ are intermediate feature maps extracted from our convolutional blocks.

Finally, we utilize a 3×3 convolutional layer to predict a 16×16 probability matrix P_{output} for image patch discrimination. Through the probability matrix, we can increase the discriminator's sensitivity to image patch detail discrimination, thus forcing MLDRG to generate more realistic details. Mathematically,

$$P_{\text{output}} = \text{Conv}(F_N). \quad (6)$$

Moreover, the discriminator loss in the restoration stage can be defined as:

$$L_D^R = -E_{I^T \sim p_{\text{data}}(I^T)}[\log D_R(I^T)] + E_{I^U \sim p_G(I^U)}[\log (D_R(G_R(I^U)))]. \quad (7)$$

3.1.2 Loss function

To remove the corruption from the observed underwater images, we formulate some objective functions. First, the adversarial loss for MLDRG can be formulated as:

$$L_{Adv}^R = -E_{I^U \sim p_G(I^U)}[\log (D_R(G_R(I^U)))]. \quad (8)$$

In the restoration stage, we define the mean absolute error (MAE) loss to measure the pixel gap between the generated images and the target images. Mathematically,

$$L_{MAE}^R = \frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H |I_{ij}^T - G_R(I^U)_{ij}|. \quad (9)$$

Moreover, in order to enhance the human visual quality of reconstructed images, we formulate a perceptual loss to measure the distance between the restored images and the ground-truth images on the perceptual feature space. It can be formulated as:

$$L_{Perceptual}^R = \frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H |\phi_{x,y}(I^T)_{ij} - \phi_{x,y}(G_R(I^U))_{ij}|, \quad (10)$$

where $\phi_{x,y}$ symbol denotes a pre-trained VGG-net feature extractor, which can obtain the feature map of the y -th convolutional layer before the x -th max-pooling layer (Simonyan and Zisserman, 2015), while W and H symbols denote the width and height of the obtained corresponding feature map, respectively.

3.2 Super-resolution stage

SISR is aimed at generating the high-resolution image I^{SR} from the low-resolution image I^{LR} . Generally, the degradation process from I^{HR} to I^{LR} is unknown and can be affected by various factors, such as defocusing and noise. Following the common practice (Zhang et al., 2017; Zhang et al., 2018a; Liang et al., 2022), we obtain I^{LR} by a downsampling operation with the scaling factor r . For an image with C channels, the I^{LR} and I^{SR} are described as a $C \times W \times H$ tensor and a $C \times rW \times rH$ tensor, respectively.

To add the texture details to the restored image I^R fed from the restoration stage, we propose another GAN-based model, which is aimed at generating the corresponding super-resolution image I^{SR} from the restored image I^R . The objective function of the super-resolution stage is formulated as:

$$\min_{G_{SR}} \max_{D_{SR}} V(G_{SR}, D_{SR}) = E_{I^{HR} \sim p_{data}(I^{HR})} [\log D_{SR}(I^{HR})] + E_{I^{LR} \sim p_G(I^{LR})} [\log (1 - D_{SR}(G_{SR}(I^{LR})))], \quad (11)$$

where I^{LR} and I^{HR} symbols denote the corresponding high-resolution and low-resolution images, respectively.

3.2.1 Network architecture

The generator of the super-resolution stage is illustrated in Figure 5, which consists of a high frequency learning module (HFLM) and an upsampling module (UM). Due to the effective collaboration of HFLM and UM, our super-resolution stage can restore many high-frequency details from low-resolution underwater images.

3.2.1.1 High frequency learning module

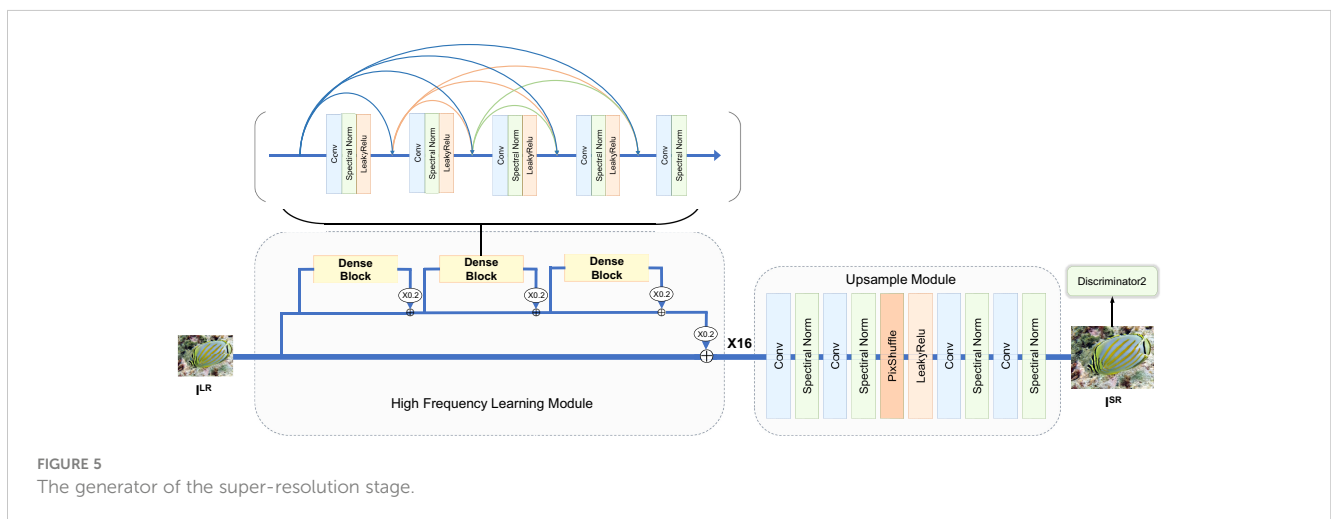
After obtaining the restored images after the restoration stage, we further seek to excavate the high-frequency information of underwater images. High-frequency information can be described as Figure 6, where I^{LR} is obtained by bicubic interpolation of the restored image, and I^{HR} is the high-resolution ground truth image. Our task in the super-resolution stage is to learn these high-frequency information. The task of our high-frequency feature learning module is to learn these high-frequency information. To this end, HFLM first directly transmits the low frequency information of the low resolution image to the upsampling module through a connection, and then learns the high frequency information of the image through 16 residual-in-residual blocks (Wang et al., 2018). In each dense block, it captures and transmits high frequency information by establishing dense residual connections in the network. Specifically, each layer is connected to all previous layers, making it easier for high frequency features to propagate throughout the network and be better captured and represented. Mathematically,

$$F_i^H = \begin{cases} \rho \cdot H_{D_i}(I_{LR}) + I_{LR}, & i = 1; \\ \rho \cdot H_{D_i}(F_{i-1}) + F_{i-1}^H, & i = 2, \dots, N, \end{cases} \quad (12)$$

where I_{LR} and ρ denote the output of the restoration stage and the residual scaling parameter, respectively, H_{D_i} represents the i -th dense block, N represents the number of the dense blocks, and F_i^H represents the i -th intermediate high-frequency feature. Specifically, we set ρ and N to 0.2 and 16, respectively. Furthermore, we add an SN layer after each convolutional layer to constrain the Lipschitz continuity of HFLM, and the leak rate of the Leaky Relu activation function is set to 0.2.

3.2.1.2 Upsampling module

After obtaining the high-frequency information provided F_N^H by the HFLM, we adopt a pixshuffle layer for upsampling, passing the F_N^H through convolutional layers and inter-channel recombination to obtain a high-resolution feature map I^{SR} . Similar to HFLM, we



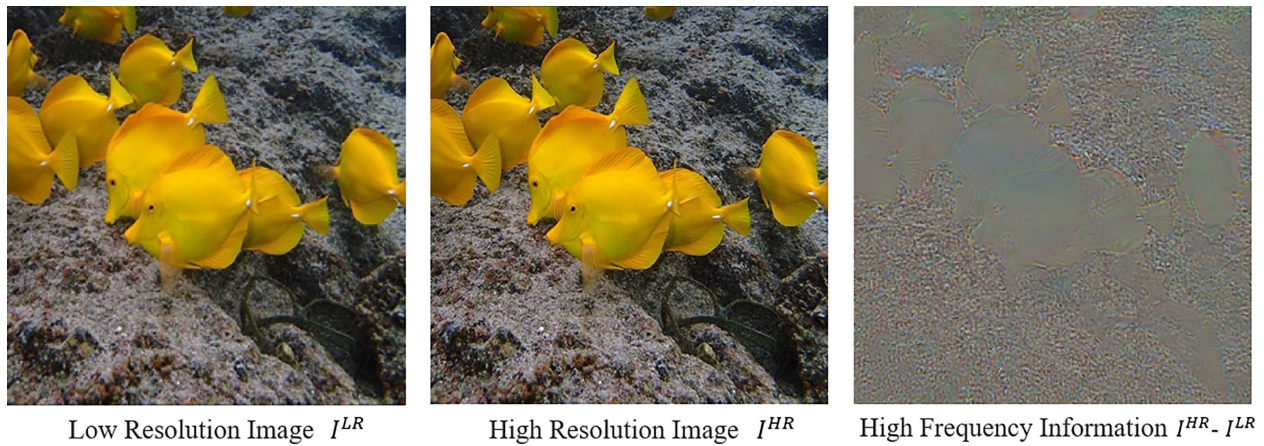


FIGURE 6
The representation of high frequency information.

add an SN layer after each convolutional layer to constrain the Lipschitz continuity of UM and stabilize the network training. Mathematically,

$$I^{SR} = UM_1(F_N^H + I^{LR}). \quad (13)$$

3.2.1.3 Super-resolution discriminator

The architecture of the super-resolution stage's discriminator is similar to that of the restoration stage. However, for image super-resolution, we expect the output by its discriminator to be the probability that the real image I^{HR} is relatively more realistic than the fake image I^{SR} . To this end, we use a relativistic discriminator (Jolicœur-Martineau, 2018), which is defined as:

$$D_{SR}(I^{HR}, I^{SR}) = \sigma(P_{SR}(I^{HR}) - E[P_{SR}(I^{SR})]), \quad (14)$$

$$D_{SR}(I^{SR}, I^{HR}) = \sigma(P_{SR}(I^{SR}) - E[P_{SR}(I^{HR})]), \quad (15)$$

where I^{SR} denotes the output of the generator in the super-resolution stage, $P_{SR}(I)$ denotes the probability output of the patch discriminator, $E[\cdot]$ represents the operation of taking the average probability output obtained from mini-batch images, and σ denotes the sigmoid activation function. Then, we formulate the discriminator loss as follows:

$$L_D^{SR} = -E_{I^{HR} \sim p_{data}(I^{HR})}[\log D_{SR}(I^{HR}, G_{SR}(I^{LR}))] + \\ - E_{I^{LR} \sim p_G(I^{LR})}[\log (1 - D_{SR}(G_{SR}(I^{LR}), I^{HR}))]. \quad (16)$$

3.2.2 Loss function

Similar to the restoration stage, The MAE loss and perception loss are both used to optimize its generator in the super-resolution stage for a better reconstruction effect. Mathematically,

$$L_{MAE}^{SR} = \frac{1}{r^2 WH} \sum_{i=1}^{rW} \sum_{j=1}^{rH} |I_{ij}^{HR} - G_{SR}(I^{LR})_{ij}|, \quad (17)$$

$$L_{Perceptual}^{SR} = \frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H |\phi_{x,y}(I^{HR})_{ij} - \phi_{x,y}(G_{SR}(I^{LR}))_{ij}|. \quad (18)$$

In addition, we formulate the adversarial loss for the generator as follows:

$$L_{Adv}^{SR} = -E_{I^{HR} \sim p_{data}(I^{HR})}[\log (1 - D_{SR}(I^{HR}, G_{SR}(I^{LR})))] + \\ - E_{I^{LR} \sim p_G(I^{LR})}[\log D_{SR}(G_{SR}(I^{LR}), I^{HR})]. \quad (19)$$

It can be clearly seen that the adversarial loss in our super-resolution stage includes both I^{HR} and $I^{SR} = G_{SR}(I^{LR})$. Hence, the gradient of the generator in the super-resolution stage benefits from both the generated images and the ground-truth images. In contrast, the gradient of the generator in the previous stage only benefits from the generated images.

3.3 SRSRGAN

SRSRGAN combines the restoration stage and the super-resolution stage into an end-to-end trainable model. Concretely, the generator of SRSRGAN combines generators of the restoration stage and the super-resolution stage. For training, we adopt a two-stage training strategy. In the first stage, we use a restoration discriminator to supervise the restoration stage generator's training, which serves as a pre-training for the restoration stage generating adversarial network. In the second stage, we directly use the super-resolution stage discriminator to supervise the entire SRSRGAN model's training. Finally, we can train SRSRGAN as an end-to-end GAN-based model.

During inference, by feeding degraded underwater images into the SRSRGAN model, we can obtain clean high-resolution underwater images end-to-end.

In addition, by doing so, SRSRGAN has the following advantages in addition to the benefits brought by its well-designed model structure:

- Removing noise from images will generally introduce artifacts to the images, while the devised super resolution stage in SRSRGAN can generate texture details to avoid the artifacts.
- The generator of SRSRGAN benefits from both the image restoration and image super-resolution tasks, leading to better performance than using the restoration stage and super-resolution stage sequentially.
- During inference, degraded images only require one forward pass through the network to complete both image restoration and super-resolution reconstruction.

4 Experiments

We applied our SRSRGAN to underwater image restoration/enhancement, SISR, and simultaneous restoration and super-resolution. We also made a comparison with the state-of-the-art (SOTA) methods for underwater images.

We took the underwater image I^U as input in the restoration stage and the ground-truth image I^T for model training. I^R was the restored image generated in the restoration stage. In the super-resolution stage, we downsampled I^R (I^{HR}) with a scaling factor $r = 2$ to get I^{LR} . I^{SR} was the super-resolved image generated in the super-resolution stage.

We conducted experiments in PyTorch on NVIDIA GeForce RTX 3090 GPUs. To optimize SRSRGAN, we employed the Adam (Kingma and Ba, 2014) optimizer to perform global iterative learning with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and its learning rate was set to 2×10^{-4} . Considering the model depth, we adopted the warming-up strategy (He et al., 2016) to improve the learning rate gradually.

4.1 Data and metrics

4.1.1 Dataset

We used 790 images from the UIEBD (Li et al., 2020) dataset and 1500 images from the UFO-120 (Islam et al., 2020a) dataset to train SRSRGAN and the compared methods, except Gao et al. (2019) and SESR (Islam et al., 2020a). For Gao et al. (2019)'s method, we downloaded the results from the author's GitHub webpage; for SESR, we downloaded the released well-trained model from the author's GitHub webpage. In addition, we employed various datasets to test them, including the other 100 images in the UIEBD dataset with the corresponding reference images, 120 images in the UFO-120 dataset with the corresponding reference images, the same underwater scene shot by seven different professional cameras (Ancuti et al., 2018), 248 images in the USR-248 (Islam et al., 2020b) dataset, 25 images previously used for the evaluation in related papers (Emberton et al., 2015; Galdran et al., 2015; Ancuti et al., 2018; Guo et al., 2020), and 19 real underwater images we collected from the Internet.

4.1.2 Full-reference metrics

We performed a full-reference evaluation of underwater images with widely used metrics, i.e., peak signal to noise ratio (PSNR) and structural similarity index (SSIM) (Wang et al., 2004). We treated the clear HR image as the ground-truth image. The higher the PSNR value is, the closer the enhanced image is to the ground-truth image in terms of image content. Similarly, the higher the SSIM value is, the closer the enhanced image is to the ground-truth image in terms of image texture and structure.

4.1.3 Non-reference metrics

We adopted two commonly used non-reference metrics for underwater image quality evaluation, i.e., UCIQE (Yang and Sowmya, 2015) and UIQM (Panetta et al., 2016). A higher UCIQE score indicates that the enhanced image has less color cast, less blur, and better contrast. Meanwhile, a higher UIQM score indicates that the enhanced image is more in line with human perception.

4.2 Evaluation on underwater image restoration

We first qualitatively compared SRSRGAN with several SOTA underwater image restoration/enhancement methods. As Figure 7 illustrates, FUnIE-GAN (Islam et al., 2020c), CycleGAN (Zhu et al., 2017), and Gao et al. (2019)'s method have limited positive effects on the greenish water image, while FUnIE-GAN (Islam et al., 2020c) has a less positive effect on the bluish water image. Pix2Pix (Isola et al., 2017) has an obvious reddish color shift. UGAN (Fabbri et al., 2018) and Gao et al. (2019)'s method aggravate the noise effect that introduces light spots in the first image. In addition, TAEL (Liu et al., 2022)'s ability to correct the green and blue tones of underwater images is limited. In contrast, SRSRGAN can rectify the greenish and bluish hue of the images, and eliminate the blurring and noise on the images.

In addition, Figure 8 shows that images contain the standard *Macbeth Color Checker* taken by seven different professional cameras, i.e., Panasonic TS1, Pentax W80, Olympus Tough 8000, Pentax W60, Olympus Tough 6000, FujiFilm Z33, and Canon D10. The images processed by Gao et al. (2019)'s method still suffer from obvious color distortion. FUnIE-GAN (Islam et al., 2020c) deals well with the bluish color deviation but produces a reddish color shift when handling the dark image. On the contrary, SRSRGAN obtains the best color correction for different cameras.

The performance of SRSRGAN and its comparison methods is quantitatively evaluated in terms of the full-reference and non-reference metrics, as shown in Table 1. For the full-reference evaluation, the results are obtained by comparing the results of each method with the corresponding ground truth (reference) images. It can be seen that SRSRGAN achieves the highest PSNR and SSIM value, which means that the images generated by SRSRGAN have the closest content and structure to the ground-truth images. Moreover, SRSRGAN obtains the highest UCIQE and



FIGURE 7 Qualitative comparison between SRSRGAN and the SOTA restoration/enhancement methods. From left to right are the original underwater images, the results of UGAN (Fabbri et al., 2018), Pix2Pix (Isola et al., 2017), CycleGAN (Zhu et al., 2017), Gao et al. (2019)'s method, TACL (Liu et al., 2022), FUnIE-GAN (Islam et al., 2020c), and SRSRGAN. Our sample image is sourced from the public dataset UIEBD (Li et al., 2020).

UIQM score, which indicates that the images generated by SRSRGAN have the best color and human visual perception.

4.3 Evaluation on underwater image super-resolution

Following the same procedure, we evaluated the qualitative and quantitative SR performance of SRSRGAN, respectively. In Particular, we took the existing underwater SISR methods for comparison, i.e., SRDRM (Islam et al., 2020b) and SRDRM-GAN (Islam et al., 2020b). In addition, we compared SRSRGAN with some SOTA SISR methods (for images taken in the air), including

VDSR (Kim et al., 2016), EDSR (Lim et al., 2017), DBPN (Haris et al., 2018), SRCNN (Dong et al., 2016), SRGAN (Ledig et al., 2017), and ESRGAN (Wang et al., 2018). From Figure 9, we can find that ESRGAN achieves the best results among the SISR methods for images taken in the air, while SRDRM and SRDRM-GAN can better handle underwater images than ESRGAN. In contrast, SRSRGAN generates clear super-resolution images with the correct color and sharp texture details.

Table 2 illustrates the quantitative evaluation on SRSRGAN and the compared methods. It is obvious that SRSRGAN obtains the highest score for both PSNR and SSIM, which indicates that the images generated by SRSRGAN have the highest pixel similarity and structure consistent with the ground-truth images.

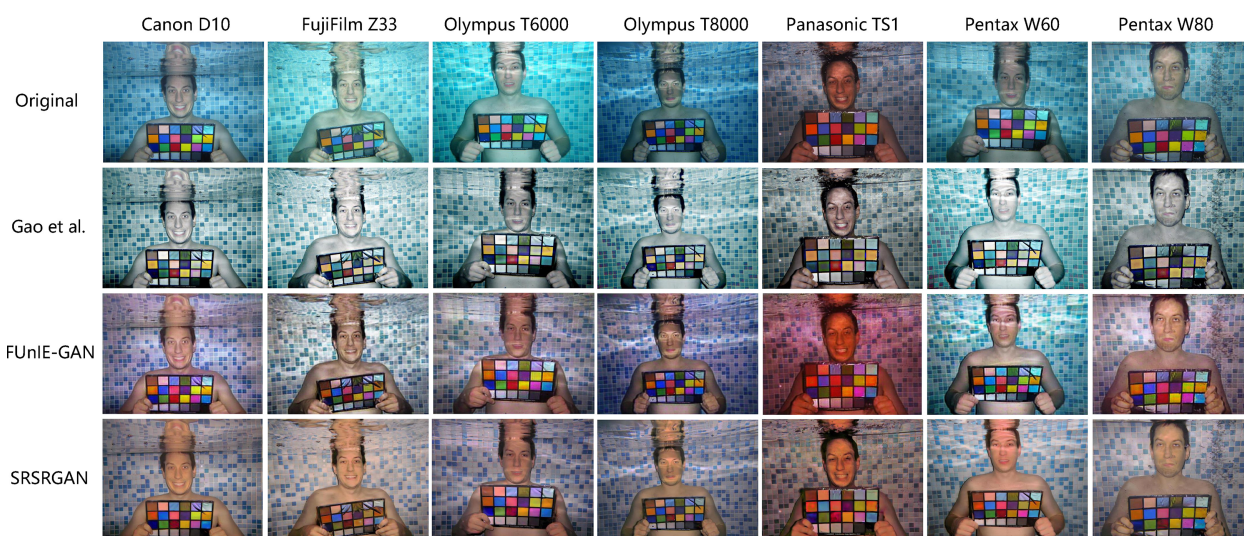


FIGURE 8 The results of SRSRGAN and the compared methods on a set of underwater images taken by different professional cameras, which contain the standard Macbeth Color Checker (Ancuti et al., 2018). The names of the cameras used to take the photos are listed at the top of each column. From top to bottom are original underwater images, the results of the Gao et al. (2019)'s method, FUnIE-GAN (Islam et al., 2020c), and SRSRGAN, respectively. Our sample image is sourced from publicly available image data in the paper (Ancuti et al., 2018). Our sample image is sourced from the public dataset UIEBD (Li et al., 2020).

TABLE 1 Quantitative evaluation on the restored images generated by SRSRGAN and the compared methods on the UIEBD dataset.

Method	PSNR/SSIM	UCIQE	UIQM
UGAN (Fabbri et al., 2018)	19.79/0.7108	0.6299	3.3218
Pix2Pix (Isola et al., 2017)	20.02/0.7230	0.5941	3.1349
CycleGAN (Zhu et al., 2017)	18.71/0.7547	0.5941	3.0144
Water-Net (Li et al., 2020)	19.13/0.7471	0.5721	3.0593
FUnIE-GAN (Islam et al., 2020c)	20.44/0.7257	0.5541	3.1255
Shallow-UWnet (Naik et al., 2021)	20.11/0.728	0.5123	3.0730
TACL (Liu et al., 2022)	20.41/0.733	0.5447	3.168
SRSRGAN	20.92/0.7731	0.6453	3.3467

The best results are shown in boldface.

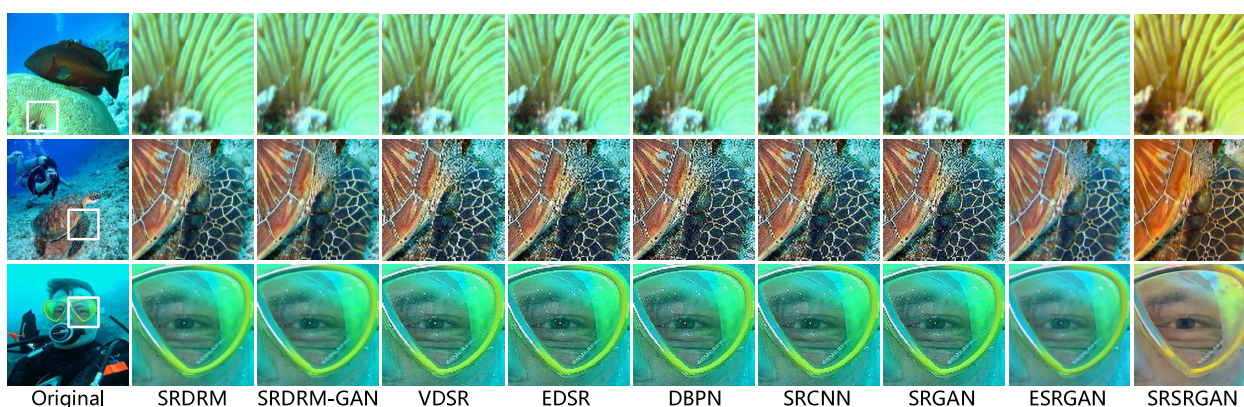


FIGURE 9

Visual comparison with several SOTA SISR methods, including SRDRM (Islam et al., 2020b), SRDRM-GAN (Islam et al., 2020b), VDSR (Kim et al., 2016), EDSR (Lim et al., 2017), DBPN (Haris et al., 2018), SRCNN (Dong et al., 2016), SRGAN (Ledig et al., 2017), and ESRGAN (Wang et al., 2018). Our sample image is sourced from the public dataset UIEBD (Li et al., 2020).

4.4 Evaluation on simultaneous restoration and super-resolution

In this experiment, we compared SRSRGAN with existing methods for simultaneous restoration and super-resolution of underwater images, i.e., Cheng et al. (2018)'s method and SESR (Islam et al., 2020a). The results of the qualitative comparison are shown in Figure 10. It can be seen that Cheng et al. (2018)'s method increases the contrast and brightness of underwater images while the images still have a bluish shift in some patches. It can also be seen from Figure 10 that SESR (Islam et al., 2020a) tends to produce artifacts on the enhanced images. SRSRGAN is more effective in restoring the colors

and increasing the resolution of underwater images. This is because the end-to-end trainable model forces the generator of SRSRGAN to complete the ultimate task that restores and super-resolves underwater images simultaneously. In other words, the generator of SRSRGAN benefits from both the restoration stage and the super-resolution stage, so that it can better adapt to the simultaneous restoration and super-resolution task for underwater images.

The quantitative evaluation of SRSRGAN and its comparison methods are shown in Tables 3 and 4. It is obvious that SRSRGAN is effective for color correction, deblurring, and contrast restoration, with the highest scores. In addition, SRSRGAN delivers sharpness and fine-grained texture details with the highest PSNR and SSIM.

TABLE 2 Quantitative evaluation on the underwater image super-resolution on the UIEBD dataset.

Method	PSNR	SSIM	Method	PSNR	SSIM
SRDRM (Islam et al., 2020b)	18.85	0.7102	SRDRM-GAN (Islam et al., 2020b)	18.93	0.7210
VDSR (Kim et al., 2016)	18.36	0.6440	EDSR (Lim et al., 2017)	18.83	0.7166
SRCNN (Dong et al., 2016)	18.78	0.7197	SRGAN (Ledig et al., 2017)	18.54	0.7159
ESRGAN (Wang et al., 2018)	18.22	0.6462	SRSRGAN	20.92	0.7731

The best results are shown in boldface.

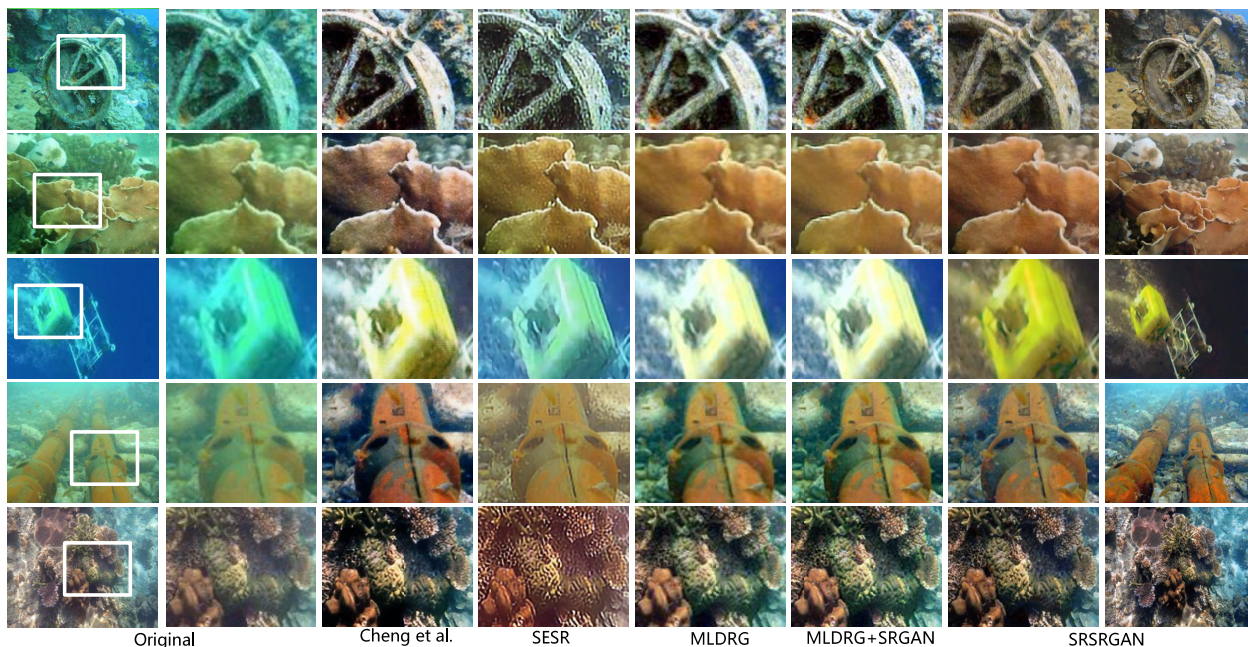


FIGURE 10 Visual comparison between SRSRGAN and several SOTA methods on simultaneous restoration and super-resolution of underwater images, i.e., Cheng et al. (2018) and SESR (Islam et al., 2020a).

4.5 Ablation study

For the proposed SRSRGAN, we added a self-attention block (SAB) to the generator of the restoration stage to learn the important parts of the global features. Meanwhile, we used the spectral normalization (SN) to constrain the Lipschitz continuity of the generators and the discriminators. Furthermore, in the super-resolution stage, we employed the high frequency learning module (HFLM) to excavate fine-grained features from the input LR images. To test the effects of these components in SRSRGAN, we conducted an ablation study to verify their effectiveness.

Table 5 illustrates the performance of SRSRGAN and its variants with different components in terms of PSNR and SSIM. We can see that SAB makes a slight improvement to the performance of SRSRGAN than that without SAB. This is due to the fact that the attention block learns the important parts of the underwater images. Hence, the generator can make more efforts to restore the important parts. In addition, we can see that removing SN from SRSRGAN greatly degrades the performance of SRSRGAN. The reason behind this is that SN effectively guarantees the Lipschitz continuity of SRSRGAN. To be specific,

SN stabilizes the training of SRSRGAN by constraining the Lipschitz continuity of its generator and discriminator.

Last but not least, HFLM plays an important role in SRSRGAN, and the variant without HFLM gets the lowest score for both PSNR and SSIM. This can be attributed to the fact that HFLM effectively extracts features of the input images. As a result, with the valid extraction features, the generator of the super-resolution stage can accurately increase the resolution of the images.

5 Conclusion

In this paper, we propose an end-to-end trainable model called SRSRGAN, which is free from the prior of corruption types and levels of underwater images. Meanwhile, SRSRGAN is a unified solution for simultaneous restoration and super-resolution of underwater images. Specifically, it captures underwater degradation information and fine-grained high-frequency information in two stages. Moreover, benefiting from the superior structure of the proposed MLDRG, our model leverages degradation information among different scales, positions, and

TABLE 3 Quantitative evaluation of SRSRGAN and the compared methods on the UFO-120 and USR-248 datasets.

Method	UFO-120		USR-248	
	PSNR	SSIM	PSNR	SSIM
Cheng et al. (2018)	26.03	0.77	27.23	0.81
SESR (Islam et al., 2020a)	27.17	0.77	26.16	0.77
SRSRGAN	28.08	0.78	29.13	0.85

The best results are shown in boldface.

TABLE 4 Quantitative evaluation of SRSRGAN and the compared methods on the UIEBD dataset.

Method	PSNR/SSIM	UCIQE	UIQM
Cheng et al. (2018)	19.92/0.7381	0.5792	2.7404
SESR (Islam et al., 2020a)	18.19/0.6917	0.5385	2.7064
MLDRG	18.55/0.6698	0.5389	2.5356
MLDRG+SRGAN	18.36/0.6592	0.5185	2.6286
SRSRGAN	20.92/0.7731	0.6453	3.3467

The best results are shown in boldface.

TABLE 5 Comparisons of the performance of SRSRGAN and its variants with different components on the UIEBD dataset.

Method	PSNR	SSIM
SRSRGAN without SAB and SN	19.29dB	0.7204
SRSRGAN without SAB	19.96dB	0.7528
SRSRGAN without SN	19.75dB	0.7335
SRSRGAN without HFLM	18.79dB	0.7153
SRSRGAN	20.92dB	0.7731

The best results are shown in boldface.

channels in the restoration stage to transform degraded images to clean images. Besides, the HFLM excavates fine-grained high-frequency information to super-resolve clean images. Extensive experimental results demonstrate the superiority of SRSRGAN in underwater image restoration, super-resolution, and simultaneous restoration and super-resolution.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

HW: Conceptualization, Methodology, Model design, Investigation, Writing - Original Draft. GZ and DW: Methodology, Writing - Review and Editing, Funding acquisition, Supervision. JS and YC: Conceptualization, Methodology, Investigation, Writing - Original Draft. YZ: Formal analysis, Writing - Review and Editing. SL: Visualization, Validation. All authors contributed to the article and approved the submitted version.

Funding

This work was partially supported by the National Key Research and Development Program of China under Grant No. 2018AAA0100400, HY Project under Grant No. LZ2022033004,

the Natural Science Foundation of Shandong Province under Grants No. ZR2020MF131 and No. ZR2021ZD19, Project of the Marine Science and Technology cooperative Innovation Center under Grant No. 22-05-CXZX-04-03-17, the Science and Technology Program of Qingdao under Grant No. 21-1-4-ny-19-nsh, and Project of Associative Training of Ocean University of China under Grant No. 202265007.

Acknowledgments

We want to thank “Qingdao AI Computing Center” and “Eco-Innovation Center” for providing inclusive computing power and technical support of MindSpore during the completion of this paper.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Alenezi, F., Armghan, A., and Santosh, K. (2022). Underwater image dehazing using global color features. *Eng. Appl. Artif. Intell.* 116, 105489. doi: 10.1016/j.engappai.2022.105489
- Ancuti, C. O., Ancuti, C., De Vleeschouwer, C., and Bekaert, P. (2018). Color balance and fusion for underwater image enhancement. *IEEE Trans. Image Process.* 27, 379–393. doi: 10.1109/TIP.2017.2759252
- Caner, G., Tekalp, A. M., and Heinzelman, W. B. (2003). “Super resolution recovery for multi-camera surveillance imaging,” in *International Conference on Multimedia and Expo* (Baltimore, MD, USA: IEEE), Vol. 1. 109–112. doi: 10.1109/ICME.2003.1220866
- Capel, D., and Zisserman, A. (2001). “Super-resolution from multiple views using learnt image models,” in *IEEE Conference on Computer Vision and Pattern Recognition* (Kauai, HI, USA: IEEE), Vol. 2. II–II. doi: 10.1109/CVPR.2001.991022
- Chen, Y., Sun, J., Jiao, W., and Zhong, G. (2019). Recovering super-resolution generative adversarial network for underwater images. *Neural Information Processing 4*, 75–83. doi: 10.1007/978-3-030-36808-1
- Chen, T., Wang, N., Wang, R., Zhao, H., and Zhang, G. (2021). One-stage cnn detector-based benthonic organisms detection with limited training dataset. *Neural Networks* 144, 247–259. doi: 10.1016/j.neunet.2021.08.014
- Cheng, N., Zhao, T., Chen, Z., and Fu, X. (2018). “Enhancement of underwater images by super-resolution generative adversarial networks,” in *International Conference on Internet Multimedia Computing and Service* (Nanjing, Jiangsu, China: IEEE). doi: 10.1145/3240876.3240881
- Chiang, J. Y., and Chen, Y.-C. (2012). Underwater image enhancement by wavelength compensation and dehazing. *IEEE Trans. Image Process.* 21, 1756–1769. doi: 10.1109/TIP.2011.2179666
- Dong, C., Loy, C. C., He, K., and Tang, X. (2014). “Learning a deep convolutional network for image super-resolution,” in *European Conference on Computer Vision* (Zurich, Switzerland: Springer). 184–199.
- Dong, C., Loy, C. C., He, K., and Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 295–307. doi: 10.1109/TPAMI.2015.2439281
- Drewns, J., do Nascimento, E., Moraes, F., Botelho, S., and Campos, M. (2013). “Transmission estimation in underwater single images,” in *IEEE International Conference on Computer Vision Workshops* (Sydney, Australia: IEEE). 825–830. doi: 10.1109/ICCVW.2013.113DBLP:conf/iccvw/DrewnsNMBBC13
- Duchon, C. (1979). Lanczos filtering in one and two dimensions. *J. Appl. meteorol.* 18, 1016–1022. doi: 10.1175/1520-0450(1979)018<1016:LFIOAT>2.0.CO;2duchon1979lanczos
- Emberton, S., Chittka, L., and Cavallaro, A. (2015). *British Machine Vision Conference (BMVC)* (Swansea, UK: BMVA Press), Vol. 125. 1–125.
- Fabbri, C., Islam, M. J., and Sattar, J. (2018). “Enhancing underwater imagery using generative adversarial networks,” in *IEEE International Conference on Robotics and Automation* (Brisbane, Australia: IEEE). 7159–7165. doi: 10.1109/ICRA.2018.8460552DBLP:conf/icra/FabbriIS18
- Farsiou, S., Robinson, M. D., Elad, M., and Milanfar, P. (2004). Fast and robust multiframe super resolution. *IEEE Trans. Image Process.* 13, 1327–1344. doi: 10.1109/TIP.2004.834669DBLP:journals/tip/FarsiouREM04
- Galdran, A., Pardo, D., Picón, A., and Alvarez-Gila, A. (2015). Automatic red-channel underwater image restoration. *J. Visual Communication Image Represent.* 26, 132–145. doi: 10.1016/j.jvcir.2014.11.006DBLP:journals/jvcir/GaldranPPA15
- Gao, S.-B., Zhang, M., Zhao, Q., Zhang, X.-S., and Li, Y.-J. (2019). Underwater image enhancement using adaptive retinal mechanisms. *IEEE Trans. Image Process.* 28, 5580–5595. doi: 10.1109/TIP.2019.2919947DBLP:journals/tip/GaoZZLL19
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial networks. *Adv. Neural Inf. Process. Syst.* 3, 139–144. doi: 10.1145/3422622DBLP:journals/corr/GoodfellowPMXWOCB14
- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. (2017). Improved training of Wasserstein GANs. *Neural Inf. Process. Syst. (NIPS)*, 5769–5779. doi: 10.48550/arXiv.1704.00028
- Guo, Y., Li, H., and Zhuang, P. (2020). Underwater image enhancement using a multiscale dense generative adversarial network. *IEEE J. Oceanic Eng.* 45, 862–870. doi: 10.1109/JOE.2019.29114478730425
- Haris, M., Shakhnarovich, G., and Ukita, N. (2018). “Deep back-projection networks for super-resolution,” in *IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, Utah: IEEE). 1664–1673. doi: 10.1109/CVPR.2018.00179haris2018deep
- Harmeling, S., Sra, S., Hirsch, M., and Schölkopf, B. (2010). “Multiframe blind deconvolution, super-resolution, and saturation correction via incremental em,” in *IEEE International Conference on Image Processing (ICIP)* (Hong Kong, China: IEEE). 3313–3316. doi: 10.1109/ICIP.2010.5651650DBLP:conf/icip/HarmelingSHS10
- He, K., Sun, J., and Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 2341–2353. doi: 10.1109/TPAMI.2010.168DBLP:journals/pami/HeOT11
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, Nevada: IEEE). 770–778. doi: 10.1109/CVPR.2016.90DBLP:conf/cvpr/kaimingHe2016
- Hou, M., Liu, R., Fan, X., and Luo, Z. (2018). “Joint residual learning for underwater image enhancement,” in *IEEE International Conference on Image Processing* (Athens, Greece: IEEE). 4043–4047. doi: 10.1109/ICIP.2018.8451209DBLP:conf/icip/HouLFL18
- Hu, J., Shen, L., and Sun, G. (2018). “Squeeze-and-excitation networks,” in *IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT, USA: IEEE). 7132–7141. doi: 10.1109/CVPR.2018.00745DBLP:conf/ieee/transenet
- Hummel, R. (1977). Image enhancement by histogram transformation. *Comput. Graphics Image Process.* 6, 184–195. doi: 10.1016/S0146-664X(77)80011-7hummel1975image
- Islam, M., Luo, P., and Sattar, J. (2020a). Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. *Robot.: Sci. Sys.* doi: 10.15607/RSS.2020.XVI.018DBLP:journals/corr/abs-2002-01155
- Islam, M. J., Sakib Enan, S., Luo, P., and Sattar, J. (2020b). “Underwater image super-resolution using deep residual multipliers,” in *IEEE International Conference on Robotics and Automation* (Cambridge, Massachusetts, USA: MIT Press). 900–906. doi: 10.1109/ICRA40945.2020.9197213DBLP:conf/icra/IslamELS20
- Islam, M. J., Xia, Y., and Sattar, J. (2020c). Fast underwater image enhancement for improved visual perception. *IEEE Robot. Auto. Lett.* 5, 3227–3234. doi: 10.1109/LRA.2020.2974710DBLP:journals/ral/IslamXS20
- Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). “Image-to-image translation with conditional adversarial networks,” in *IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI, USA: IEEE). 5967–5976. doi: 10.1109/CVPR.2017.632DBLP:conf/cvpr/IsolaZZE17
- Jolicœur-Martineau, A. (2018). The relativistic discriminator: a key element missing from standard gan. *arXiv*. doi: 10.48550/arXiv.1704.00028
- Keys, R. (1982). Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoustics Speech Signal Process.* 29, 1153–1160. doi: 10.1109/TASSP.1981.1163711keys1981cubic
- Kim, K. I., and Kwon, Y. (2010). Single-image super-resolution using sparse regression and natural image prior. *IEEE Trans. Pattern Anal. Mach. Intell.* (Las Vegas, NV, USA: IEEE) 32, 1127–1133. doi: 10.1109/TPAMI.2010.25DBLP:journals/pami/KimK10
- Kim, J., Lee, J., and Lee, K. M. (2016). “Accurate image super-resolution using very deep convolutional networks,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1646–1654. doi: 10.1109/CVPR.2016.182DBLP:conf/cvpr/KimLL16a
- Kingma, D., and Ba, J. (2014). “Adam: A method for stochastic optimization,” in *International Conference on Learning Representations* (Banff, Canada: OpenReview.net).
- Köhler, T., Bätz, M., Naderi, F., Kaup, A., Maier, A. K., and Riess, C. (2017). Benchmarking super-resolution algorithms on real data. *CoRR* abs/1904.08444. 10.48550/arXiv.1709.04881
- Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., et al. (2017). “Photo-realistic single image super-resolution using a generative adversarial network,” in *IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI, USA: IEEE). 105–114. doi: 10.1109/CVPR.2017.19DBLP:conf/cvpr/LedigTHCCAATTWS17
- Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., et al. (2020). An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.* 29, 4376–4389. doi: 10.1109/TIP.2019.2955241DBLP:journals/tip/LiGRCHKT20
- Li, X., Hou, G., Li, K., and Pan, Z. (2022). Enhancing underwater image via adaptive color and contrast enhancement, and denoising. *Eng. Appl. Artif. Intell.* 111, 104759. doi: 10.1016/j.engappai.2022.104759DBLP:journals/EAAI/xinjieLi22
- Li, X., and Orchard, M. (2000). “New edge directed interpolation,” in *IEEE International Conference on Image Processing* (Paris, France: IEEE), Vol. 2. 311–314. doi: 10.1109/ICIP.2000.899369DBLP:conf/icip/LiO00
- Li, J., Skinner, K. A., Eustice, R. M., and Johnson-Roberson, M. (2018). Watergan: unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robot. Auto. Lett.* 3, 387–394. doi: 10.1109/LRA.2017.2730363DBLP:journals/ral/LiSEJ18
- Lian, H. (2006). “Variational local structure estimation for image super-resolution,” in *International Conference on Image Processing* (Atlanta, Georgia, USA: IEEE). 1721–1724. doi: 10.1109/ICIP.2006.312713DBLP:conf/icip/Lian06
- Liang, J., Zeng, H., and Zhang, L. (2022). “Details or artifacts: a locally discriminative learning approach to realistic image super-resolution,” in *IEEE Conference on Computer Vision and Pattern Recognition* (New Orleans, LA, US: IEEE). 5647–5656. doi: 10.1109/CVPR52688.2022.00557DBLP:conf/cvpr/jielLiangDA2022
- Lim, B., Son, S., Kim, H., Nah, S., and Lee, K. M. (2017). “Enhanced deep residual networks for single image super-resolution,” in *IEEE Conference on Computer Vision and Pattern Recognition Workshops* (Honolulu, HI, USA: IEEE). 1132–1140. doi: 10.1109/CVPRW.2017.151DBLP:conf/cvpr/LimSKNL17
- Lin, J., Gan, C., and Han, S. (2019). Defensive quantization: when efficiency meets robustness. doi: 10.48550/arXiv.1904.08444DBLP:journals/corr/abs-1904-08444

- Liu, Y.-C., Chan, W.-H., and Chen, Y.-Q. (1995). Automatic white balance for digital still camera. *IEEE Trans. Consumer Electron.* 41, 460–466. doi: 10.1109/30.468045liu1995automatic
- Liu, R., Jiang, Z., Yang, S., and Fan, X. (2022). Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Trans. Image Process.* 31, 4922–4936. doi: 10.1109/tip.2022.3190209liu2022twin
- Lu, H., Li, Y., Nakashima, S., Kim, H., and Serikawa, S. (2017a). Underwater image super-resolution by descattering and fusion. *IEEE Access* 5, 670–679. doi: 10.1109/ACCESS.2017.2648845DBLP:journals/access/LuLNKS17
- Lu, H., Li, Y., Uemura, T., Kim, H., and Serikawa, S. (2018). Low illumination underwater light field images reconstruction using deep convolutional neural networks. *Future Gen. Comput. Syst.* 82, 142–148. doi: 10.1016/j.future.2018.01.001DBLP:journals/fgcs/LuLUKS18
- Lu, H., Li, Y., Zhang, Y., Chen, M., Serikawa, S., and Kim, H. (2017b). Underwater optical image processing: a comprehensive review. *mob. netw. Appl.* 22, 1204–1211. doi: 10.1007/s11036-017-0863-4DBLP:journals/monet/LuLZCSK17
- Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. *Int. Conf. Learn. Represent. (ICLR)*. doi: 10.48550/arXiv.1802.05957
- Muresan, D. D. (2005). Fast edge directed polynomial interpolation 2, II–990. doi: 10.1109/ICIP.2005.1530224DBLP:conf/icip/Muresan05
- Naik, A., Swarnakar, A., and Mittal, K. (2021). “Shallow-uwnet: compressed model for underwater image enhancement,” in *Proceedings of the AAAI Conference on Artificial Intelligence* (Vancouver, Canada: AAAI), Vol. 35. 15853–15854. doi: 10.1609/aaai.v35i18.17923naik2021shallow
- Nasrollahi, K., and Moeslund, T. (2014). Super-resolution: a comprehensive survey. *Mach. Vision Appl.* 25, 1423–1468. doi: 10.1007/s00138-014-0623-4DBLP:journals/mva/NasrollahiM14
- Panetta, K., Gao, C., and Agaian, S. (2016). Human-visual-system-inspired underwater image quality measures. *IEEE J. Oceanic Eng.* 41, 541–551. doi: 10.1109/JOE.2015.2469915panetta2015human
- Reza, A. M. (2004). Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *J. VLSI Signal Process. Syst. signal image video Technol.* 38, 35–44. doi: 10.1023/B:VLSI.0000028532.53893.82DBLP:journals/vlisp/Reza04
- Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A. P., Bishop, R., et al. (2016). “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in *IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV, USA: IEEE). 1874–1883. doi: 10.1109/CVPR.2016.207DBLP:conf/cvpr/ShiCHTABRW16
- Simonyan, K., and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. doi: 10.48550/arXiv.1409.1556
- Soni, O. K., and Kumare, J. S. (2020). “A survey on underwater images enhancement techniques,” in *IEEE 9th International Conference on Communication Systems and Network Technologies* (Gwalior, India: IEEE). 333–338. doi: 10.1109/CSNT48778.2020.9115732sahu2014survey
- Storkey, A. J. (2002). Dynamic structure super-resolution. *Neural Inf. Process. Syst. (NIPS)*.
- Timofte, R., De, V., and Gool, L. V. (2013). “Anchored neighborhood regression for fast example-based super-resolution,” in *IEEE International Conference on Computer Vision* (Sydney, Australia: IEEE) 15, 1920–1927. doi: 10.1109/ICCV.2013.241DBLP:conf/iccv/TimofteDG13
- Timofte, R., Smet, V. D., and Gool, L. J. V. (2014). “A+: adjusted anchored neighborhood regression for fast super-resolution,” in *Asian Conference on Computer Vision* (Singapore: Springer), Vol. 9006. 111–126.
- Tsai, R. (1984). Multiframe image restoration and registration. *Adv. Comput. Visual Image Process.* 1, 317–339.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). “Attention is all you need,” in *Conference on Neural Information Processing Systems* (Long Beach, USA: Curran Associates Inc). 6000–6010.
- Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612. doi: 10.1109/TIP.2003.819861DBLP:journals/tip/WangBSS04
- Wang, N., Chen, T., Kong, X., Chen, Y., Wang, R., Gong, Y., et al. (2023a). Underwater attentional generative adversarial networks for image enhancement. *IEEE Trans. Human-Machine Syst.*, 1–11. doi: 10.1109/THMS.2023.3261341WangUnderwater
- Wang, N., Chen, T., Liu, S., Wang, R., Karimi, H. R., and Lin, Y. (2023b). Deep learning-based visual detection of marine organisms: a survey. *Neurocomputing* 532, 1–32. doi: 10.1016/j.neucom.2023.02.018WANG20231
- Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., et al. (2018). “ESRGAN: enhanced super-resolution generative adversarial networks,” in *European Conference on Computer Vision* (Piscataway, NJ, USA: IEEE). 63–79.
- Yang, M., and Sowmya, A. (2015). An underwater color image quality evaluation metric. *IEEE Trans. Image Process.* 24, 6062–6071. doi: 10.1109/TIP.2015.2491020DBLP:journals/tip/YangS15
- Yang, J., Wright, J., Huang, T., and Ma, Y. (2008). “Image super-resolution as sparse representation of raw image patches,” in *IEEE Conference on Computer Vision and Pattern Recognition* (Anchorage, Alaska, USA: IEEE). 1–8. doi: 10.1109/CVPR.2008.4587647DBLP:conf/cvpr/YangWHM08
- Yang, J., Wright, J., Huang, T. S., and Ma, Y. (2010). Image super-resolution via sparse representation. *IEEE Trans. Image Process.* 19, 2861–2873. doi: 10.1109/TIP.2010.2050625DBLP:journals/tip/YangWHM10
- Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.-H., and Liao, Q. (2019). Deep learning for single image super-resolution: a brief review. *IEEE Trans. Multimedia* 21, 3106–3121. doi: 10.1109/TMM.2019.2919431DBLP:journals/corr/abs-1808-03344
- Yu, X., Qu, Y., and Hong, M. (2019). Underwater-GAN: underwater image restoration via conditional generative adversarial network. *International Conference on Pattern Recognition*, 66–75. doi: 10.1007/978-3-030-05792-3_7
- Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y. (2018b). “Image super-resolution using very deep residual channel attention networks,” in *European Conference on Computer Vision* (Munich, Germany: Springer). 294–310. doi: 10.1007/978-3-030-01234-2_1
- Zhang, Y., Tian, Y., Kong, Y., Zhong, B., and Fu, Y. (2018c). “Residual dense network for image super-resolution,” in *IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT, USA: IEEE). 2472–2481. doi: 10.1109/CVPR.2018.00262DBLP:conf/cvpr/ZhangTKZ018
- Zhang, K., Zuo, W., Gu, S., and Zhang, L. (2017). “Learning deep cnn denoiser prior for image restoration,” in *IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI, USA: IEEE). 2808–2817. doi: 10.1109/CVPR.2017.300DBLP:conf/cvpr/kaiZhangLDPI2017
- Zhang, K., Zuo, W., and Zhang, L. (2018a). “Learning a single convolutional super-resolution network for multiple degradations,” in *IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT, USA: IEEE). doi: 10.1109/CVPR.2018.00344DBLP:conf/cvpr/kaiZhangDS2018
- Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *IEEE International Conference on Computer Vision* (Venice, Italy: IEEE). 2242–2251. doi: 10.1109/ICCV.2017.244DBLP:conf/iccv/ZhuPIE17