Research Article

Super-Resolution Reconstruction of Underwater Image Based on Image Sequence Generative Adversarial Network

Li Li,[1,2] Zijia Fan [ID],[3] Mingyang Zhao [ID],[3] Xinlei Wang [ID],[4] Zhongyang Wang [ID],[3] Zhiqiong Wang [ID],[3] and Longxiang Guo [ID][1,2]

[1]*Acoustics Science and Technology Laboratory, Harbin Engineering University, Harbin 150001, China*
[2]*College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China*
[3]*College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110169, China*
[4]*School of Computer Science and Engineering, Northeastern University, Shenyang 110169, China*

Correspondence should be addressed to Longxiang Guo; heu503@hrbeu.edu.cn

Since the underwater image is not clear and difficult to recognize, it is necessary to obtain a clear image with the super-resolution (SR) method to further study underwater images. The obtained images with conventional underwater image super-resolution methods lack detailed information, which results in errors in subsequent recognition and other processes. Therefore, we propose an image sequence generative adversarial network (ISGAN) method for super-resolution based on underwater image sequences collected by multifocus from the same angle, which can obtain more details and improve the resolution of the image. At the same time, a dual generator method is used in order to optimize the network architecture and improve the stability of the generator. The preprocessed images are, respectively, passed through the dual generator, one of which is used as the main generator to generate the SR image of sequence images, and the other is used as the auxiliary generator to prevent the training from crashing or generating redundant details. Experimental results show that the proposed method can be improved on both peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) compared to the traditional GAN method in underwater image SR.

1. Introduction

Due to the complexity of the underwater imaging environment, the underwater image distortion is severe, and it is difficult to obtain clear and high-quality images [1]. In order to solve this problem, high-resolution images can be obtained by hardware and software. But as we all know, the hardware is relatively expensive and difficult to implement, so the super-resolution (SR) technology of underwater images is a necessary job.

Conventional super-resolution methods include interpolation, sparse representation, deep learning, etc. Kumudham and Rajendran [2] proposed a sparse representation algorithm because of the sparsity of the high-dimensional sonar image data. The image is divided into many blocks including dictionaries of low-resolution image blocks and high-resolution image blocks are created, and each block is represented using a sparse coefficient and a dictionary to obtain a high-resolution image. However, interpolation and sparse representation can lead to blurred edges in the process of obtaining super-resolution and the image information is reduced. In order to solve the problem, Lu et al. [3] used the SR algorithm based on self-similarity to obtain scattered high-resolution (HR) images and applied convex fusion rules to recover the final HR images. The experimental results show superiority, and the edges of images are significantly enhanced. However, the high-resolution image produced by the interpolation method often has some errors, which cause problems such as blockiness or detail degradation, and the improvement of image edges is not obvious. Besides, sparse representation causes blurring due to overfitting or underfitting. In recent years, deep learning methods can solve these problems well and have also been used in super-resolution for better results. Ding

et al. [4] first used an adaptive color correction algorithm to compensate for color cast and produced a natural color corrected image. Secondly, the super-resolution convolutional neural network is applied to the image to eliminate blurring of images. The experiment shows that the proposed network can learn the image deblurring from a large amount of images and the corresponding sharp image and effectively improve the quality of the underwater image. Islam et al. [5] provided a deep residual network-based generation model for single-image super-resolution (SISR) of underwater images and a countertraining pipeline for learning SISR from the paired data. At the same time, an objective function is also developed in order to supervise the training, which evaluates the perceptive quality of the image according to the overall content, color, and local style information of the image. Liu et al. [6] proposed an underwater image enhancement method through a deep residual network. Firstly, the synthetic underwater image is generated as the training data of the convolutional neural network model with a cycle-consistent generative adversarial network (CycleGAN). Secondly, the underwater residual neural network (RESNET) model for the underwater image enhancement is proposed by applying the very deep super-resolution (VDSR) reconstruction model to the application of the underwater image resolution. In addition, the loss function is also improved to form a multiple loss function with mean square error (MSE) loss and edge difference loss. However, the deep learning method lacks high frequency and details, which results in an incomplete representation of the image.

In order to obtain the image with more details, the generative adversarial network is proposed for super-resolution of images. Cheng et al. [7] proposed a new underwater image enhancement framework. The image preprocessing and deblurring are first performed with an improved super-resolution GAN. Then, the improved super-resolution GAN is used to deblur and enhance the preprocessed image. On the basis of the GAN, the loss function is corrected to sharpen the preprocessed image. Experimental results show that the enhanced GAN method effectively improves the quality of underwater images. Sung et al. [8] proposed a method for improving the resolution of underwater sonar images based on GAN. First, a network with 16 residual blocks and 8 convolutional layers is built, and then the network is trained with the sonar images intercepted in several ways. The results show that the method can improve the resolution of the sonar image and obtain a higher peak signal-to-noise ratio (PSNR) compared with the interpolation method. Furthermore, in video SR, Lucas et al. [9] designed a 15-residual neural network SRResNet for video SR, which is pretrained on MSE loss and fine-tuned in the feature-space loss. Wang et al. [10] designed a GAN using the space adaptive loss function to improve the network based on spatial activities. Chu et al. [11] proposed a GAN that obtains time coherence without loss of spatial information, and a new loss function is proposed based on this. Liu and Li [12] proposed an improved image super-resolution method based on Wasserstein distance and gradient penalty to generate a GAN to improve the gradient disappearance problem. Shamsolmoali et al. [13] proposed a GAN

that can be learned step by step, which can generate complete information and improve network stability and image quality. Xie et al. [14] proposed a method for generating SR images using time-coherent three-dimensional volume data and a novel temporal discriminator for identification. Bulat and Tzimiropoulos [15] proposed a new residual-based architecture that integrates facial information spectrum and structural information to improve SR image.

Since GAN is very excellent for super-resolution of images, it is used for super-resolution of underwater images. In the super-resolution of underwater images, a single-image super-resolution is usually processed. But in the actual situation, underwater images can generate multiple images in the same scene. Furthermore, in recent years, image fusion has been widely used because it can collect a lot of information which is applied to images [16]. Therefore, considering multiple images in the process of obtaining super-resolution images by the means of ISGAN will greatly improve the resolution of the image and make the image more detailed. In addition, due to a lot of interference and low resolution of underwater images, a single generator cannot capture full details when generating SR images, which has certain instability. In order to eliminate the deviation of the single generator to generate images and enhance the robustness, the dual generator model of image sequence is put forward, which can combine the characteristics of two generators to generate SR images with different effects and increase the quality and diversity of SR images.

In this paper, we have the contributions as follows:

(i) We design the image sequence GAN for super-resolution of underwater images and obtain high-quality images by fusing the image sequence and generating and discriminating SR images.

(ii) The dual generator including main generator and auxiliary generator is used to improve the stability of generator and optimize the structure of network.

(iii) The proposed method is evaluated experimentally, and the experimental results show that this method can obtain images with more details and higher resolution.

2. Methods

In order to solve the super-resolution problem of the image sequence, the SRGAN network [17] structure is improved so that the image can adapt to the underwater image and the information of multiple images can be obtained. Therefore, the resolution of the underwater image can be improved on the basis of adding more image details.

2.1. ISGAN Method. We divide the ISGAN method into two steps: preprocessing and ISGAN structure. In the process of preprocessing, the color of images is corrected and the contrast of images is improved for the convenience of the following training. At the same time, the images are pretrained to ensure the stability of the network and improve

the training speed. In the ISGAN structure, image fusion is carried out and the method of a dual generator is used to ensure the accuracy and clarity of the generated images.

2.1.1. Preprocessing. Because of the serious distortion and low contrast of underwater images, the white balance and contrast limited adaptive histogram equalization (CLAHE) are used to preprocess the images. The white balance is used to correct the color of the seafloor in order to create a normal underwater scene, and the CLAHE is used to improve the visibility of underwater organisms to get the enhanced image. Therefore, the results of the image preprocessing are as shown in Figure 1.

Then, the pretraining is conducted in order to increase the speed of training in discriminator and maintain the stability of the generator. Before training, part of the HR image training set is put into the discriminator D for pretraining; the prior training ensures early identification capability of the discriminator and maintains the training intensity and efficiency of the generator [18]. Furthermore, the pretraining prevents the collapse of the training mode that leads to the continuously failed generation of SR images and ensures the stability and the training speed of the discriminator and generator, which is convenient for the following training strategy adjustment.

2.1.2. ISGAN Architecture. After images are preprocessed, they are sent to the ISGAN for training and the SR image is generated by the generator. In the generator, the image sequence is fused firstly. Because there is a certain offset between the image sequences, the image needs to be registered by using the geometric registration (SURF), linear photometric model, and affine motion.

Then, the image fusion is performed to collect the information of all images. In order to fully represent the detailed features of all image sequences, the fusion process of image sequences is added to the network structure, and the resolution of the image is improved by blending the sharpest part of each image sequence. Firstly, the image is decomposed into four sub-bands by stationary wavelet transform (SWT), which are low-low (LL) sub-band, low-high (LH) sub-band, high-low (HL) sub-band, and high-high (HH) sub-band, where the LL sub-band is an approximation coefficient containing the original detail of the image, and the remaining LH, HL, and HH sub-bands represent the detail parameters of the original image. The process of dividing the sub-bands by SWT is shown in Figure 2.

After all the images are divided into different sub-bands, the principal component analysis (PCA) is performed on the sub-bands. This method can find the best feature for the data, which is the clearest part of the image, and represent this part as the first feature. The principal component, represented by the data with the largest variance in the calculation, is a good representation of the data. In signal processing, it is generally believed that the signal has a large variance, while the noise has a small variance. The variance ratio between the signal and the noise is defined as the signal-to-noise ratio. Therefore, the variance is usually used

to judge whether the signal is useful information. Similarly, such an idea is also adopted in image processing. It is generally considered that the useful part in the image has a large variance so that it is taken as the principal component, while noise is generally considered as redundant information. Subsequently, the first principal component is sorted and selected and the first principal component of each sub-band is fused. In this process, the fusion rule is to multiply all the pixels of the sub-band of each image by the largest eigenvector of the sub-band. Finally, the processing for each sub-band is repeated, and new fused sub-bands LL, HL, LH, and HH are, respectively, established, as shown in the following equation:

$$
\begin{aligned}
\mathrm{LL}(i, j) &= p_1 \mathrm{LL}_1(i, j) + p_2 \mathrm{LL}_2(i, j) + \cdots + p_n \mathrm{LL}_n(i, j), \\
\mathrm{LH}(i, j) &= q_1 \mathrm{LH}_1(i, j) + q_2 \mathrm{LH}_2(i, j) + \cdots + q_n \mathrm{LH}_n(i, j), \\
\mathrm{HL}(i, j) &= u_1 \mathrm{HL}_1(i, j) + u_2 \mathrm{HL}_2(i, j) + \cdots + u_n \mathrm{HL}_n(i, j), \\
\mathrm{HH}(i, j) &= v_1 \mathrm{HH}_1(i, j) + v_2 \mathrm{HH}_2(i, j) + \cdots + v_n \mathrm{HH}_n(i, j),
\end{aligned}
\tag{1}
$$

where the size of each image is $M \times N$, n represents the number of image sequences, and i and j represent pixel locations, where $i = 1, 2, \ldots, M$, $j = 1, 2, \ldots, N$. Besides, p, q, u, and v represent the largest eigenvectors of the four sub-bands in the source image, and the four sub-bands LL, LH, HL, and HH, respectively, represent four sub-bands after fusion according to the fusion rule. Finally, the four fusion sub-bands are reconstructed by inverse stationary wavelet transform (ISWT) to obtain the refused image I^{RE}. The refused image is iterated through the generator and discriminator, and the image is learned to obtain a super-resolution image. In the process of learning, the main generation of the confrontation network model proposed by Ledig et al. [17] is used for learning. The generated confrontation network model can be expressed as follows:

$$
\begin{aligned}
\min_{\theta_G} \max_{\theta_D} &\ E_{I^{\mathrm{HR}} \sim p_{\mathrm{train}}(I^{\mathrm{HR}})} \left[\log D_{\theta_D}\left(I^{\mathrm{HR}}\right) \right] + E_{I^{\mathrm{RE}} \sim p_G(I^{\mathrm{RE}})} \\
&\ \left[\log\left(1 - D_{\theta_D}\left(I^{\mathrm{RE}}\right)\right) \right].
\end{aligned}
\tag{2}
$$

The equation is expressed as it allows the training generation model G to fool the discriminator D that distinguishes the super-resolution image from the real image by training and obtain the super-resolution image by continuously learning the fused image and finally determine the super-resolution image. In this way, our generator can learn to create and gradually optimize SR images so that the discriminator cannot distinguish between real and fake images, which makes the generated images more and more similar to real images.

At the same time, when the main generator generates SR images, an auxiliary generator is also used to generate a group of SR images. In the auxiliary generator, in order to reduce artifacts, improve generalization ability, and reduce computational complexity, the BN layer is removed to improve training stability and performance. Then, these two sets of SR images and HR images are mixed as an input of the
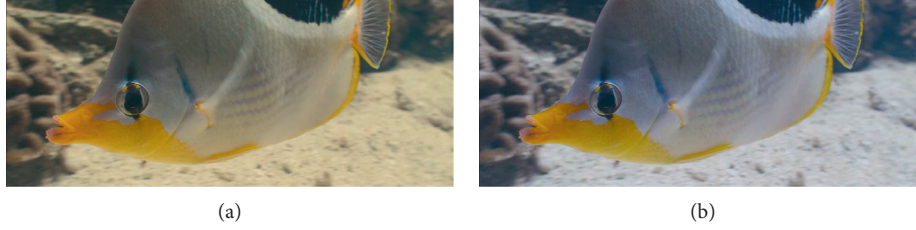
FIGURE 1: The results of the image preprocessing. (a) Original underwater image. (b) Preprocessed underwater image.
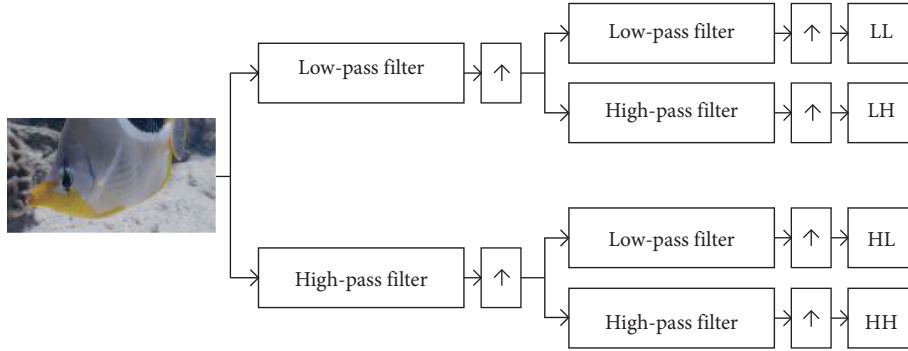

FIGURE 2: The process of sub-band decomposition.

discriminator so that it can enhance the robustness of results and make the resulting images more reliable.

In the ISGAN model, the generator uses two convolutional layers as the activation function, where the convolutional layer has a small 3×3 convolution kernel and 64 feature maps, followed by a batch normalization (BN) layer and parametric rectified linear unit (PReLU) layer and two trained subpixel convolution layers to improve the resolution of the input image. The discriminator uses the Leaky ReLU activation layer to avoid maximum pooling in the entire network. It contains 8 convolutional layers, adding 3×3 filter kernels, increasing from 64 to 512 to obtain the probability of sample classification. Through such a network model, the resolution of the image can be significantly improved, and a better super-resolution reconstruction result can be obtained. The network architecture is shown in Figure 3.

2.2. Loss Function. In the ISGAN model, the perceptual loss is capable of enriching the details in the image. Since the perceptual loss function is critical to the performance of the generator, it is expressed as a weighted sum of the content loss and the adversarial loss according to the proposed ISGAN model. Among them, the content loss includes the mean square error loss (MSE) and the VGG loss, and the adversarial loss is used to confuse whether the SR image generated by the generator is a real image. The loss function is shown in the following equation:

$$l^{\mathrm{SR}} = \underbrace{\underbrace{l^{\mathrm{SR}}_{\mathrm{Con}}}_{\text{content loss}} + \underbrace{10^{-3} l^{\mathrm{SR}}_{\mathrm{Gen}}}_{\text{adversarial loss}}}_{\text{perceptual loss}}. \tag{3}$$

2.2.1. Content Loss. The content loss includes MSE loss and VGG loss. The MSE loss is the most widely used optimization target in image super-resolution and represents the expected value of the square of the difference between the estimated value and the true value. MSE can evaluate the degree of change of the data, which is a convenient method to measure the "average error." The smaller the value of MSE, the better the accuracy of the prediction model to describe the experimental data. In the proposed ISGAN model, the MSE loss is defined as follows:

$$l^{\mathrm{SR}}_{\mathrm{MSE}} = \frac{1}{r^2 W H} \sum_{x=1}^{rW} \sum_{y=1}^{rH} \left(I^{\mathrm{HR}}_{x,y} - G_{\theta_G} \left(I^{\mathrm{RE}} \right)_{x,y} \right)^2. \tag{4}$$

However, while achieving a particularly high PSNR, MSE loss usually results in the lack of high-frequency content in the generated SR image so that the image will produce a smooth texture. Therefore, the VGG loss is added, which is defined by training the ReLU activation layer of the VGG network. It can be defined as the Euclidean distance between the feature representation of the reconstructed image $G_{\theta_G}(I^{\mathrm{RE}})$ and the real reference image (I^{HR}). The feature map of a layer is extracted on the already trained VGG network, and this feature map of the generated image is compared with the real image, as shown in the following equation:

$$l^{\mathrm{SR}}_{\mathrm{VGG}} = \frac{1}{W H} \sum_{x=1}^{W} \sum_{y=1}^{H} \left(\phi \left(I^{\mathrm{HR}} \right)_{x,y} - \phi \left(G_{\theta_G} \left(I^{\mathrm{RE}} \right) \right)_{x,y} \right)^2, \tag{5}$$

where W and H represent the dimensions of the corresponding feature mapping in the VGG network.

To sum it up, the content loss of the ISGAN model can be defined by MSE loss and VGG loss, as shown in the following equation:
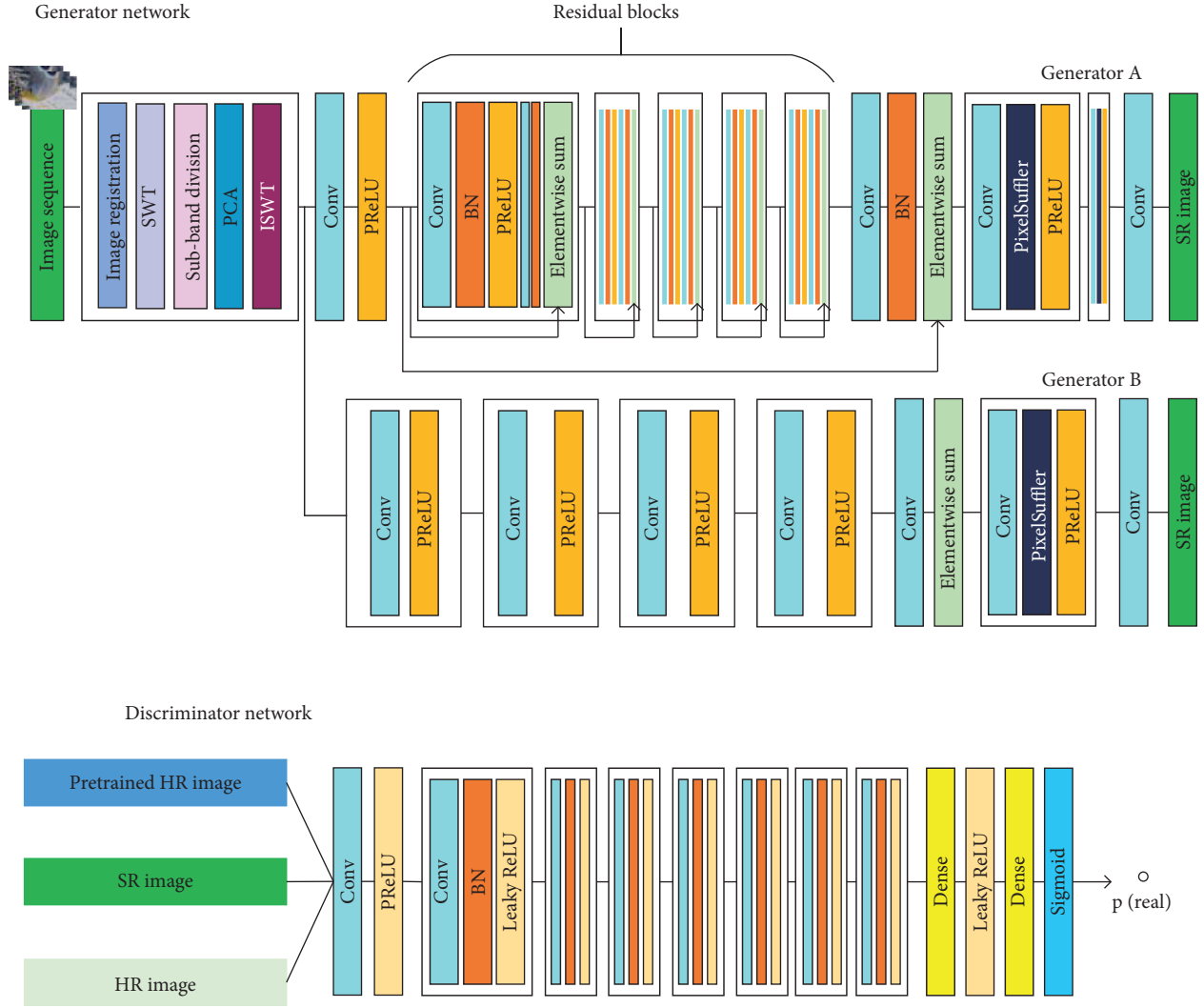
FIGURE 3: The network architecture of ISGAN.

$$l_{\text{Con}}^{\text{SR}} = l_{\text{MSE}}^{\text{SR}} + l_{\text{VGG}}^{\text{SR}}, \tag{6}$$

where $l_{\text{MSE}}^{\text{SR}}$ and $l_{\text{VGG}}^{\text{SR}}$ denote the MSE loss and VGG loss in the above definitions, respectively. Therefore, the content loss defined in this way makes the reconstructed image as similar as possible to the high-resolution image and has similar characteristics to the low-resolution original image.

2.2.2. Adversarial Loss. In addition to the above content loss, the adversarial loss is also important to the perceptual loss. Its purpose is to fool the discriminator to determine the generated super-resolution image so that it can generate a data distribution that the discriminator cannot distinguish and thus cannot judge whether the image is a real image. In the proposed ISGAN model, the adversarial loss can be defined as follows:

$$l_{\text{Gen}}^{\text{SR}} = \sum_{n=1}^{N} -\log D_{\theta_D}\big(G_{\theta_G}\big(I^{\text{RE}}\big)\big), \tag{7}$$

where $D_{\theta_D}(G_{\theta_G}(I^{\text{RE}}))$ represents the probability that the reconstructed image $G_{\theta_G}(I^{\text{RE}})$ is judged to be a high-resolution image, and in order to obtain a better gradient characteristic, we reduce $\log[1 - D_{\theta_D}(G_{\theta_G}(I^{\text{RE}}))]$ to $-\log D_{\theta_D}(G_{\theta_G}(I^{\text{RE}}))$.

With this loss function, the discriminator's ultimate goal is to output 1 for all real pictures, and for all fake images, the output is 0. On the contrary, the goal of the generator is to fool the discriminator, which is to output 1 for the generated image. In this way, the process of alternating iterative training can be achieved, and the images that can fool the discriminator are obtained, which is the resulting super-resolution image.

3. Results and Discussion

3.1. Training and Parameters. Our training dataset is collected from NTIRE database, which is different from the testing data. In the experiments, we obtain the low-resolution (LR) image from the high-resolution (HR) images by downsampling with a factor of 16. The size of HR image is
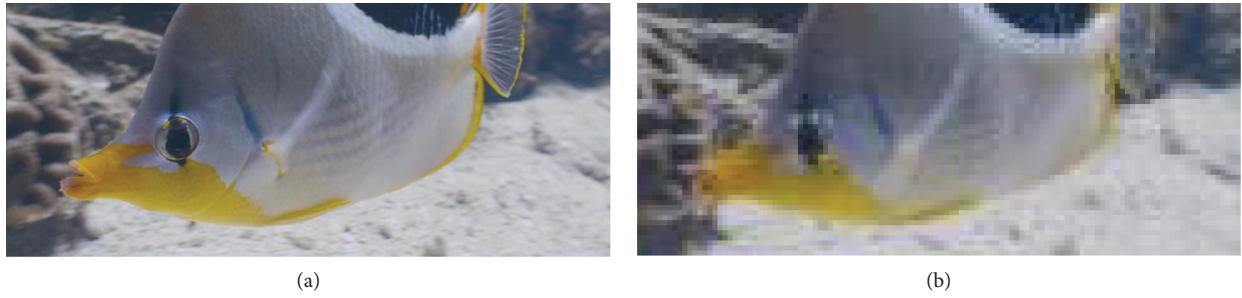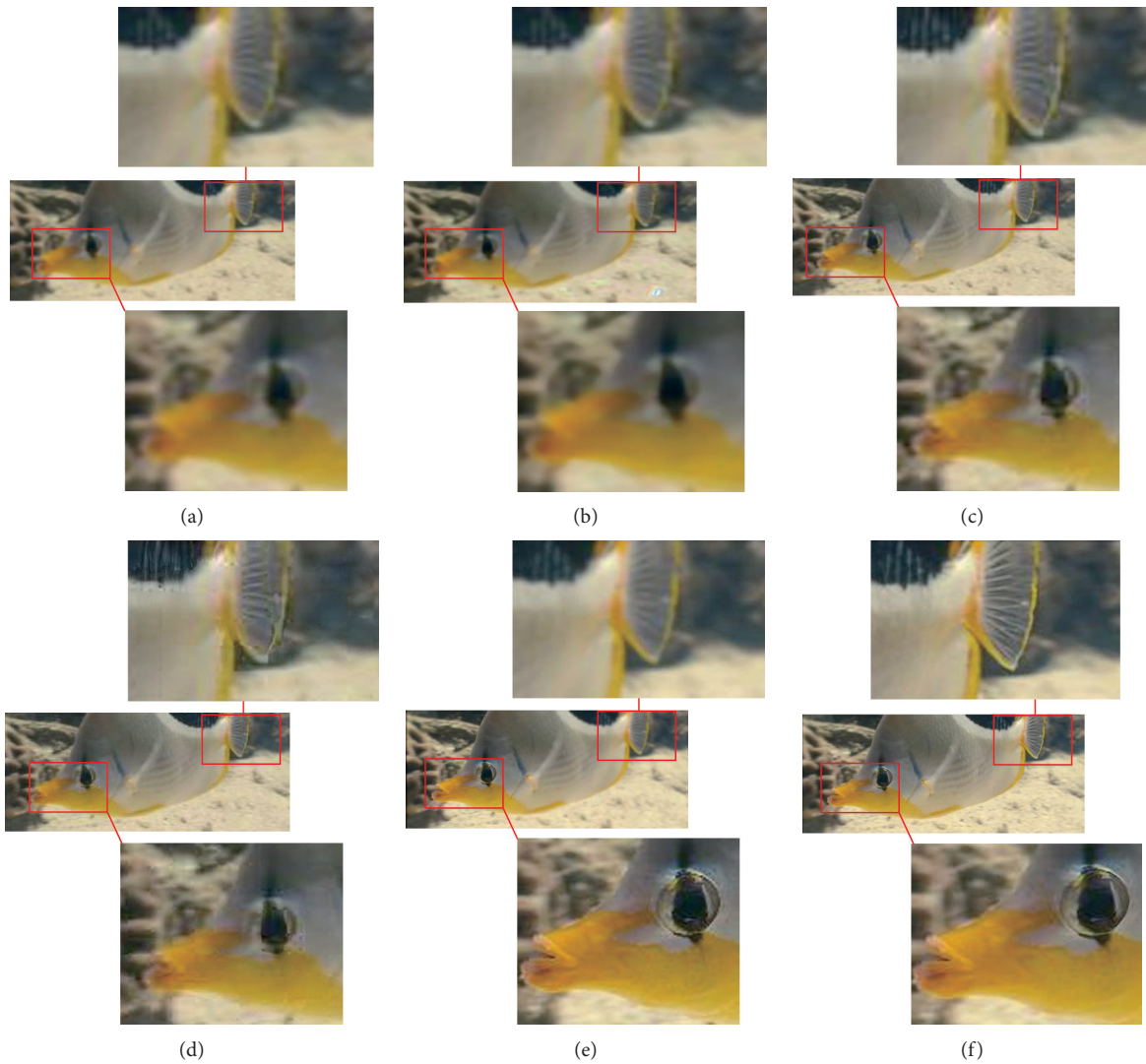
(a) (b)

FIGURE 4: (a) HR image and (b) LR image.

(a) (b) (c)

(d) (e) (f)

FIGURE 5: SR results on image Fish 1. (a) Bicubic. (b) USIGAN. (c) EGAN. (d) GGAN. (e) VDSR. (f) ISGAN.

2040×1404, as shown in Figure 4. For each minibatch, 16 random HR subimages are cropped, which is not only to increase the amount of data but also to weaken data noise and increase model stability. For optimization, we use Adam [19] with $\beta_1 = 0.9$. In addition, the networks are trained with a learning rate of 10^{-4} and 10^3 update iterations.

3.2. Evaluation Results. To verify the performance of the proposed method, several experiments are performed for comparison. We test three kinds of fish images named Fish 1, Fish 2, and Fish 3 separately and compare the performance of the bicubic interpolation, underwater sonar image GAN (USIGAN) method [8], enhancement GAN (EGAN)
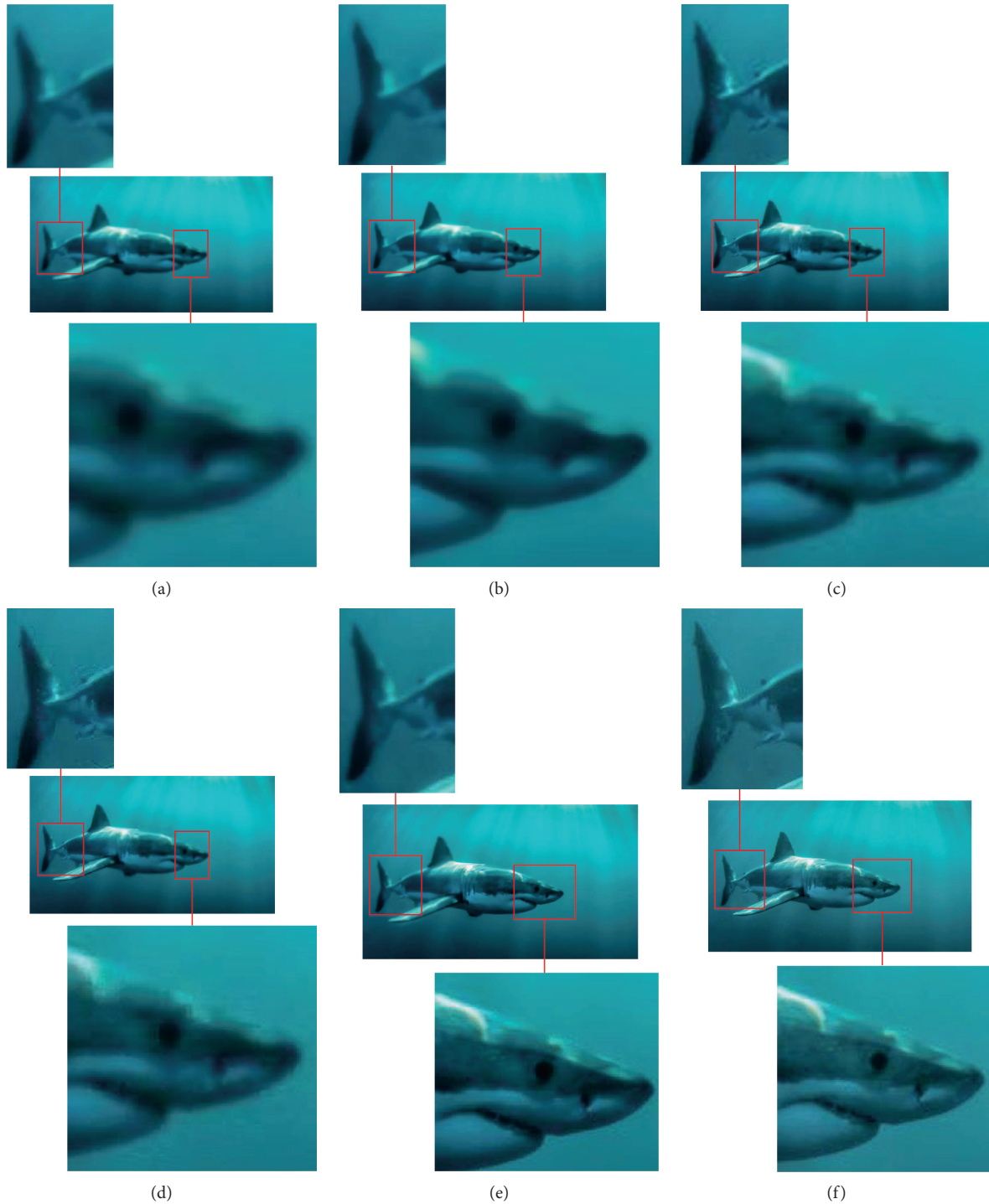
Figure 6: SR results on image Fish 2.(a) Bicubic. (b) USIGAN. (c) EGAN. (d) GGAN. (e) VDSR. (f) ISGAN.

method [7], gradual GAN (GGAN) method [13], and very deep super-resolution (VDSR) method [6]. Here, the bicubic interpolation method is the most traditional and classic underwater image super-resolution method. The USIGAN method is the traditional GAN method for underwater sonar images and the EGAN and GGAN methods are the improved GAN methods. Besides, the VDSR method is one of the deep learning methods for super-resolution of

underwater images. Figures 5–7 show the SR results obtained by different methods.

It can be seen from the figure that the proposed method has the best effect, which can clearly show the details of each part of the image and also have a higher resolution. The bicubic interpolation method can improve the resolution of the image but cannot restore the full details of the image. Both USIGAN method and EGAN method can obtain better

FIGURE 7: SR results on image Fish 3. (a) Bicubic. (b) USIGAN. (c) EGAN. (d) GGAN. (e) VDSR. (f) ISGAN.

results than the bicubic method, but some details are still unclear. In addition, GGAN and VDSR can get high-resolution images with sufficient clarity but the unfocused areas cannot be clearly restored. Our proposed ISGAN method can not only get the clearest images but also reflect the information of the whole image completely.

To further verify the effectiveness of the proposed method, we consider two evaluation index indicators, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), which are calculated as objective measurements, as shown in Tables 1 and 2. PSNR is one of the most common and widely used objective criteria for evaluating images, which is defined based on MSE, as shown in the following equation:

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2, \qquad (8)$$

TABLE 1: PSNR of SR images with different methods.

	Bicubic	USIGAN	EGAN	GGAN	VDSR	ISGAN
Fish 1	24.5151	24.8645	24.9737	25.379	25.663	26.347
Fish 2	30.9937	34.257	34.5471	35.672	37.936	39.7221
Fish 3	23.1508	23.415	23.6668	24.383	24.509	25.202

TABLE 2: SSIM of SR images with different methods.

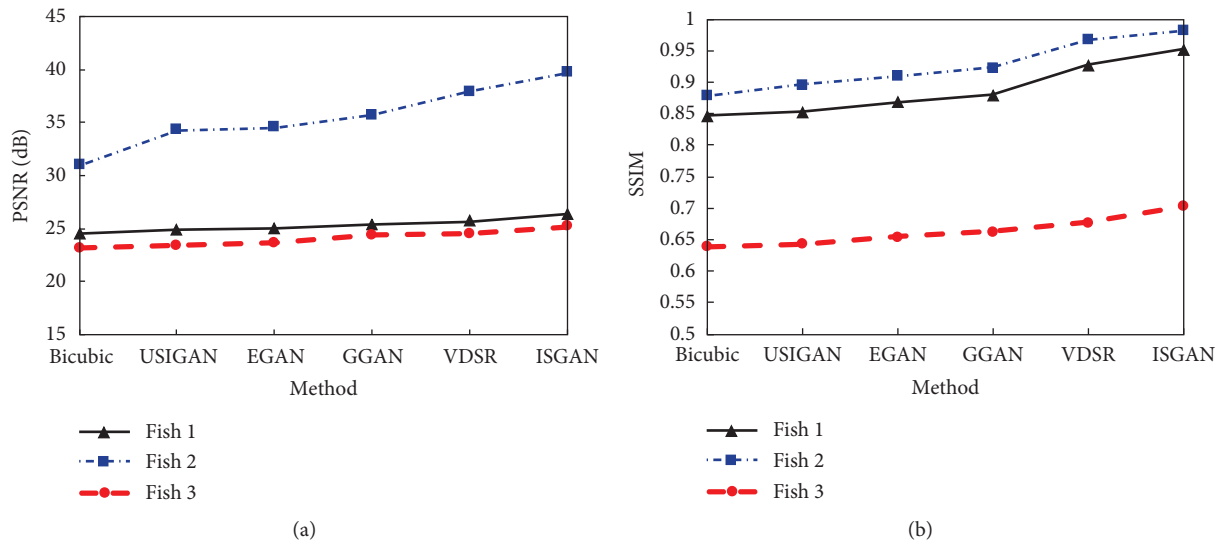|	Bicubic	USIGAN	EGAN	GGAN	VDSR	ISGAN
Fish 1	0.847	0.853	0.8689	0.8803	0.9276	0.9528
Fish 2	0.879	0.897	0.91	0.923	0.9675	0.9827
Fish 3	0.639	0.643	0.655	0.6631	0.6772	0.703

FIGURE 8: Evaluation index. (a) PSNR. (b) SSIM.

where K represents the LR image and I represents the HR image of size $M \times N$. Then, PSNR is defined as

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{255}{\text{MSE}}\right). \quad (9)$$

SSIM is a similarity determined by three measures of LR image and HR image, and the three measures are brightness, contrast, and structure, respectively, expressed as

$$l(I, K) = \frac{2\mu_I \mu_K + c_1}{\mu_I^2 + \mu_K^2 + c_1},$$

$$c(I, K) = \frac{2\sigma_I \sigma_K + c_2}{\sigma_I^2 + \sigma_K^2 + c_2}, \quad (10)$$

$$s(I, K) = \frac{\sigma_{IK} + c_3}{\sigma_I \sigma_K + c_3},$$

which generally take $c_3 = c_2/2$. In the equation, μ_I and μ_K are the mean values of I and K, σ_I^2 and σ_K^2 are the variances of I and K separately, σ_{xy} is the covariance of I and K, and

Besides, in order to avoid calculation errors caused by the denominator being 0 in the formula, c_1 and c_2 are set as nonzero constants to ensure the stability of the result. Therefore, SSIM can be expressed as

$$\text{SSIM}(x, y) = \left[l(x, y)^{\alpha} \cdot c(x, y)^{\beta} \cdot s(x, y)^{\gamma} \right]. \quad (11)$$

The comparison results are shown in Figure 8 according to PSNR and SSIM.

The results show the superiority of the proposed method in the testing data. It can be seen from the figure that the proposed method performs best in both the PSNR and SSIM evaluation indexes. Besides, the bicubic method gets the lowest value of PSNR and SSIM, which means this method cannot restore enough information. USIGAN and EGAN methods have higher PSNR and SSIM values than the bicubic method, but they still cannot reflect the complete details. In addition, GGAN and VDSR methods have higher PSNR and SSIM values close to those of the ISGAN method. Although the GGAN and VDSR methods can obtain a clear image, the missing part of the detail cannot be supplemented by these two methods. Therefore, the proposed ISGAN

method can accomplish two tasks at the same time and obtain an image with the best effect.

4. Conclusion

The super-resolution reconstruction is performed by using the underwater image sequence through the improvement of the existing GAN model, where the fusion step of the image sequence is added in the generator, and the loss function is changed accordingly. Therefore, it can be more suitable for the super-resolution reconstruction method for underwater images, which combines image sequence information to acquire features in more images, resulting in clearer and more detailed super-resolution underwater images. Experimental results show that the proposed ISGAN method can improve image resolution and display complete image information.

Data Availability

The training data used to support the findings of this study were collected from the NTIRE public database.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

References

[1] S. Anwar, C. Li, and F. Porikli, "Deep underwater image enhancement," 2018, https://arxiv.org/abs/1807.03528.

[2] R. Kumudham and V. Rajendran, "Super resolution enhancement of underwater sonar images," *SN Applied Sciences*, vol. 1, no. 8, 2019.

[3] H. Lu, Y. Li, S. Nakashima, H. Kim, and S. Serikawa, "Underwater image super-resolution by descattering and fusion," *IEEE Access*, vol. 5, pp. 670–679, 2017.

[4] X. Ding, Y. Wang, L. Zheng, J. Zhang, and X. Fu, "Towards underwater image enhancement using super-resolution convolutional neural networks," in *Proceedings of the International Conference on Internet Multimedia Computing and Service*, Qingdao, China, August 2017.

[5] M. J. Islam, S. S. Enan, P. Luo, and J. Sattar, "Underwater image super-resolution using deep residual multipliers," *Clinical Orthopaedics and Related Research*, 2019, https://arxiv.org/abs/1909.09437.

[6] P. Liu, G. Wang, H. Qi, C. Zhang, H. Zheng, and Z. Yu, "Underwater image enhancement with a deep residual framework," *IEEE Access*, vol. 7, pp. 94614–94629, 2019.

[7] N. Cheng, T. Zhao, Z. Chen, and X. Fu, "Enhancement of underwater images by super-resolution generative adversarial networks," in *Proceedings of the 10th International Conference on Internet Multimedia Computing and Service*, vol. 22, no. 1–22, Nanjing, China, August 2018.

[8] M. Sung, J. Kim, and S. C. Yu, "Image-based super resolution of underwater sonar images using generative adversarial network," in *Proceedings of the IEEE Region 10 Conference TENCON 2018*, pp. 457–461, Jeju, Republic of Korea, October 2018.

[9] A. Lucas, S. Lopez-Tapia, R. Molina, and A. K. Katsaggelos, "Generative adversarial networks and perceptual losses for video super-resolution," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3312–3327, 2019.

[10] X. Wang, A. Lucas, S. Lopez-Tapia, X. Wu, R. Molina, and A. K. Katsaggelos, "Spatially adaptive losses for video super-resolution with GANs," in *Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1697–1701, Brighton, UK, May 2019.

[11] M. Chu, Y. Xie, J. Mayer, L. Leal-Taixé, and N. Thuerey, "Temporally coherent GANs for video super-resolution (TecoGAN)," 2018, https://arxiv.org/abs/1811.09393.

[12] S. Liu and X. Li, "A novel image super-resolution reconstruction algorithm based on improved GANs and gradient penalty," *International Journal of Intelligent Computing and Cybernetics*, vol. 12, no. 3, pp. 400–413, 2019.

[13] P. Shamsolmoali, M. Zareapoor, R. Wang, D. K. Jain, and J. Yang, "G-GANISR: gradual generative adversarial network for image super resolution," *Neurocomputing*, vol. 366, pp. 140–153, 2019.

[14] Y. Xie, E. Franz, M. Chu, and N. Thuerey, "tempoGAN: a temporally coherent, volumetric GAN for super-resolution fluid flow," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, p. 95, 2018.

[15] A. Bulat and G. Tzimiropoulos, "Super-FAN: integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 109–117, Salt Lake City, AL, USA, June 2018.

[16] F. Zhang and K. Zhang, "Superpixel guided structure sparsity for multispectral and hyperspectral image fusion over couple dictionary," *Multimedia Tools and Applications*, vol. 79, no. 7-8, pp. 4949–4964, 2020.

[17] C. Ledig, L. Theis, and F. Huszar, "Photo-realistic single image super-resolution using a generative adversarial network," *CVPR*, vol. 105–114, 2017.

[18] A. Bulat, J. Yang, and G. Tzimiropoulos, "To learn image super-resolution, use a gan to learn how to do image degradation first," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 185–200, Munich, Germany, September 2018.

[19] D. Kingma, J. Ba, and Adam, "A method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, May 2015.