Recovering Super-Resolution Generative Adversarial Network for Underwater Images

Yang Chen, Jinxuan Sun, Wencong Jiao, and Guoqiang Zhong[✉]

Department of Computer Science and Technology, Ocean University of China,
238 Songling Road, Qingdao 266100, China
gqzhong@ouc.edu.cn

Abstract. In this paper, we propose an end-to-end Recovering Super-Resolution Generative Adversarial Network (RSRGAN) to automatically learn super-resolution underwater images. RSRGAN mainly includes two parts. The first part is a Recovering GAN, aiming at color correction and removing noise in the images. The generator of Recovering GAN is based on an encoder-decoder network with self-attention on the global feature. The second part is a Super-Resolution GAN, which adopts the residual-in-residual dense block in its generator, to add details onto the results fed from the Recovering GAN. Both qualitative and quantitative experimental results show the advantage of RSRGAN over the state-of-the-art approaches for underwater image super-resolution.

Keywords: Underwater images · Super-resolution · Generative adversarial network

1   Introduction

Underwater images generally suffer from severe degradation, such as lack of contrast, color casting and noise. The poor visibility of underwater images limits the performance of subsequent vision tasks. Hence, high-resolution (HR) images are desirable for many underwater applications.

In this paper, we propose an end-to-end Recovering Super-Resolution Generative Adversarial Network (RSRGAN) to generate the super-resolution (SR) underwater images. We solve the problem in two stages. In the first stage, we use the first part of RSRGAN, called Recovering GAN, to correct the color and remove the noise of the underwater images. In the second stage, we use the second part of RSRGAN, called Super-Resolution GAN, to enrich the fine texture details of the images restored by the Recovering GAN. RSRGAN combines the generators of both GAN models and fine-tunes the entire model with the Super-Resolution GAN's discriminator. Experimental results show that RSRGAN outperforms the state-of-the-art methods for underwater image super-resolution.

2   Related Work

As far as we know, there are very few super-resolution methods for underwater images. In this section, we mainly introduce the recent approaches for underwater image restoration and single image super-resolution (SISR).

2.1   Underwater Images Restoration

Typical underwater restoration algorithms, such as histogram equalization and automatic white balance [9], improve the visual quality to some extent, but they suffer from noise amplification and color deviations problems. The emergence of generative adversarial network (GAN) [4] provides a new chance for underwater image restoration problem. Fabbri et al. designed a GAN model with a fully convolutional encoder-decoder generator to restore underwater images [3]. However, this method cannot perform well on the heavy noise images.

2.2   Single Image Super-Resolution (SISR)

Interpolation and sparse representation learning are widely used SR methods [17]. However, the generated SR images generally lack detailed textures. With the development of deep learning, Dong et al. [2] used a three-layer fully convolutional network to get HR images. Kim et al. used a 20-layer VGG [13] to obtain SR images [6]. In [8], the authors proposed the enhanced deep residual networks for SISR. As GAN-based models boom, Ledig et al. [7] and Wang et al. [15] employed GAN with the perceptual loss for the applications of image SR. But unfortunately, these methods are only performed on the natural images. Particularly, Lu et al. [10] applied denoising and descattering methods to the SR underwater images. However, denoising and descattering brought additional blur to the SR images.
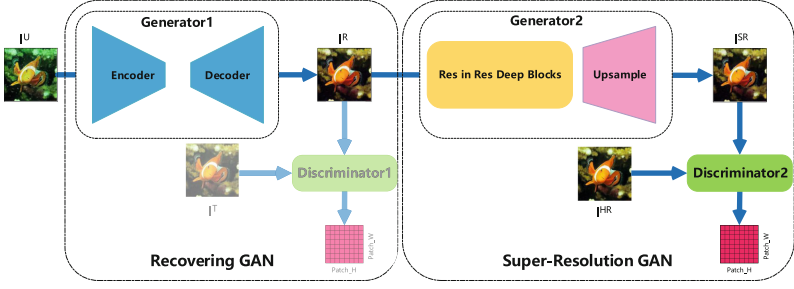
3   The Proposed Model

The goal of this work is to establish an underwater super-resolution system. The proposed RSRGAN includes two parts, Recovering GAN and Super-Resolution GAN. Figure 1 shows the architecture of RSRGAN.

3.1   Recovering GAN

To recover the clear image I^R from the noisy and degraded image I^U, we propose a GAN-based model, Recovering GAN, which can be formalized as:
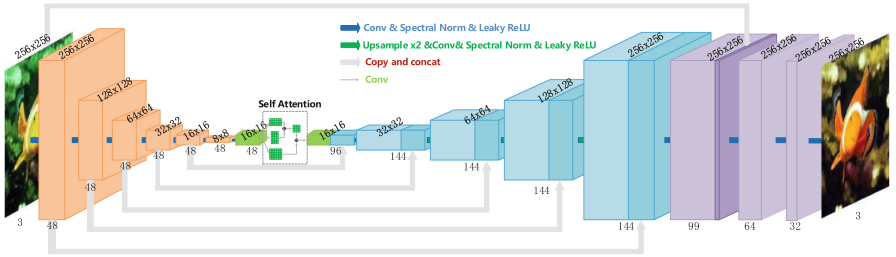
$$\min_{G_R} \max_{D_R} V(G_R, D_R) = E_{I^T \sim p_{train}(I^T)}[\log D_R(I^T)]+$$
$$E_{I^U \sim p_G(I^U)}[\log(1 - D_R(G_R(I^U)))], \qquad (1)$$

where I^T is the clear image (ground-truth) corresponding to I^U.

Fig. 1. The architecture of RSRGAN. The pretrained generators of both GAN models are combined as the generator of RSRGAN. The discriminator of the Super-Resolution GAN is used as the discriminator of RSRGAN. The light-colored parts are the discriminator of Recovering GAN only used for its pre-training.

Network Architecture. The generator of Recovering GAN is a fully convolutional encoder-decoder, as shown in Fig. 2. Every step in the encoder consists of a 3×3 convolutions with stride 1 and a 3×3 convolutions with stride 2 (in orange color). Every step in the decoder consists of an upsampling of the feature map followed by two 3×3 convolutions, a concatenation with the corresponding cropped feature map from the encoder, and a 3×3 convolution (in blue color). The other three 3×3 convolutions process the feature maps to an image (in purple color). Each convolutional layer is followed by the spectral normalization (SN) [11] and Leaky ReLU activation ($\alpha = 0.2$). Furthermore, we implement a self-attention block [14] on the global feature map before the decoder.

Fig. 2. The generator of Recovering GAN. Orange and blue boxes correspond to the feature maps in the encoder and decoder, respectively. The number of channels is denoted below the box. The size of the feature maps is denoted on the top of the box. Green boxes represent the global feature. Dark blue boxes represent copied feature maps, while purple boxes display the process from the feature maps to an image. (Color figure online)

The discriminator network D_R contains 4 convolutional blocks followed by a 3×3 convolution to obtain a 16×16 probability matrix for image patches classification. Here, each convolutional block consists of a 3×3 convolution with

stride 1 and a 3×3 convolution with stride 2. The numbers of feature maps of the blocks increase by a factor of 2 from 64 to 512. Each convolutional layer is followed by the SN and Leaky ReLU activation ($\alpha = 0.2$). The discriminator loss is defined as:

$$L_D^R = -E_{I^T \sim p_{train}(I^T)}[\log D_R(I^T)] + E_{I^U \sim p_G(I^U)}[\log(D_R(G_R(I^U)))]. \qquad (2)$$

Loss Function. We define the adversarial loss of the generator as:

$$L_{Adv}^R = -E_{I^U \sim p_G(I^U)}[\log(D_R(G_R(I^U)))]. \qquad (3)$$

In Recovering GAN, we adopt the Mean Absolute Error (MAE) loss to measure the similarity between pixels:

$$L_{MAE}^R = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} |I_{x,y}^R - G_R(I^U)_{x,y}|. \qquad (4)$$

Furthermore, we define the perceptual loss as:

$$L_{VGG}^R = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} |\phi_{i,j}(I^R)_{x,y} - \phi_{i,j}(G_R(I^U))_{x,y}|, \qquad (5)$$

where $\phi_{i,j}$ is the feature map obtained by the j-th convolution (after activation) before the i-th max pooling layer within the pre-trained VGG-19 network [13], while $W_{i,j}$ and $H_{i,j}$ are the dimensions of the respective feature maps.

3.2   Super-Resolution GAN

The Super-Resolution GAN is trained to generate corresponding I^{SR} given I^{LR} (I^R). The objective function of Super-Resolution GAN can be formalized as:
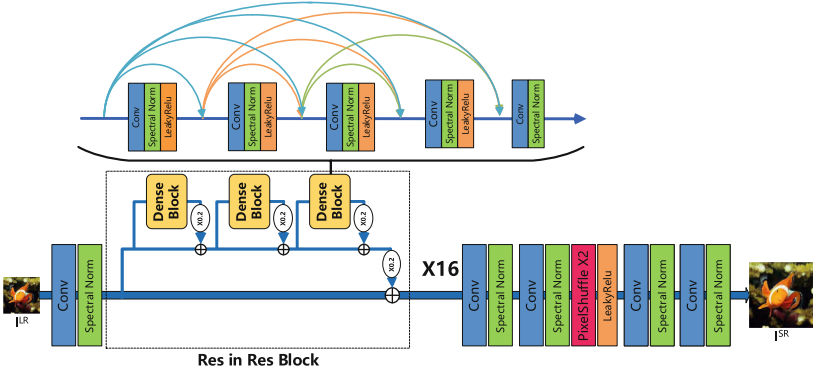
$$\min_{G_{SR}} \max_{D_{SR}} V(G_{SR}, D_{SR}) = E_{I^{HR} \sim p_{train}(I^{HR})}[\log D_{SR}(I^{HR})] +$$
$$E_{I^{LR} \sim p_G(I^{LR})}[\log(1 - D_{SR}(G_{SR}(I^{LR})))]. \qquad (6)$$

Network Architecture. The generator of Super-Resolution GAN includes 16 residual-in-residual blocks [15], as shown in Fig. 3. Specifically, each of the blocks consists of 3 dense blocks and each dense block has 5 convolutional layers. The convolutions are 3×3 with stride 1 and the residual scaling parameter is 0.2. The pixel-shuffle layer increases the resolution of the input image. In addition, each convolutional layer in the generator is followed by the SN.

The discriminator's structure of Super-Resolution GAN is similar to that in Recovering GAN. However, we want to predict the probability that a real image I^{HR} is relatively more realistic than a fake one I^{SR}. The relativistic discriminator is formalized as:

$$D_{SR}(I^{HR}, I^{SR}) = \sigma(P_{SR}(I^{HR}) - E[P_{SR}(I^{SR})]), \qquad (7)$$
$$D_{SR}(I^{SR}, I^{HR}) = \sigma(P_{SR}(I^{SR}) - E[P_{SR}(I^{HR})]), \qquad (8)$$

Fig. 3. The generator of Super-Resolution GAN.

where $P_{SR}(I)$ is the patch discriminator output, σ is the sigmoid function, $E[\cdot]$ takes an average for the images in the batch, and $I^{SR} = G_{SR}(I^{LR})$. The discriminator loss is then defined as:

$$L_D^{SR} = E_{I^{HR} \sim p_{train}(I^{HR})}[\log D_{SR}(I^{HR}, G_{SR}(I^{LR}))] +$$
$$E_{I^{LR} \sim p_G(I^{LR})}[\log(1 - D_{SR}(G_{SR}(I^{LR}), I^{HR}))]. \tag{9}$$

Loss Function. Similar to the Recovering GAN, Super-Resolution GAN uses the MAE loss and perceptual loss to optimize the generator:

$$L_{MAE}^{SR} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} |I_{x,y}^{HR} - G_{SR}(I^{LR})_{x,y}|, \tag{10}$$

$$L_{VGG}^{SR} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} |\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{SR}(I^{LR}))_{x,y}|. \tag{11}$$

Meanwhile, the adversarial loss for the generator is:

$$L_{Adv}^{SR} = E_{I^{HR} \sim p_{train}(I^{HR})}[\log(1 - D_{SR}(I^{HR}, G_{SR}(I^{LR})))] +$$
$$E_{I^{LR} \sim p_G(I^{LR})}[\log D_{SR}(G_{SR}(I^{LR}), I^{HR})]. \tag{12}$$

We can see that, the adversarial loss for the generator benefits from the gradients from both I^{HR} and $I^{SR} = G_{SR}(I^{LR})$, while previous GAN generator is only benefited from the generated data.

3.3   Recovering Super-Resolution GAN (RSRGAN)

We combine the generators of pre-trained Recovering GAN and Super-Resoluiton GAN as the generator of RSRGAN. Furthermore, we use the discriminator of Super-Resolution GAN as the discriminator of RSRGAN. Finally, we fine-tune RSRGAN as an end-to-end network.

The loss function of RSRGAN can be defined as the weighted sum of the losses aforementioned:

$$L^{SR} = \lambda_{R1}L_{MAE}^R + \lambda_{R2}L_{VGG}^R + \lambda_{R3}L_{Adv}^R + \lambda_{R4}L_D^R +$$
$$\lambda_{SR1}L_{MAE}^{SR} + \lambda_{SR2}L_{VGG}^{SR} + \lambda_{SR3}L_{Adv}^{SR} + \lambda_{SR4}L_D^{SR}, \tag{13}$$

where $\lambda_{R1} \sim \lambda_{R4}$ and $\lambda_{SR1} \sim \lambda_{SR4}$ denote the weights for Recovering GAN's loss and Super-Resolution GAN's loss, respectively.

In general, removing noise from images may introduce artifacts. The Super-Resolution GAN in RSRGAN can generate textural details to avoid the artifacts. Moreover, the end-to-end training of RSRGAN can lead to better performance than employing Recovering GAN and Super-Resolution GAN separately.
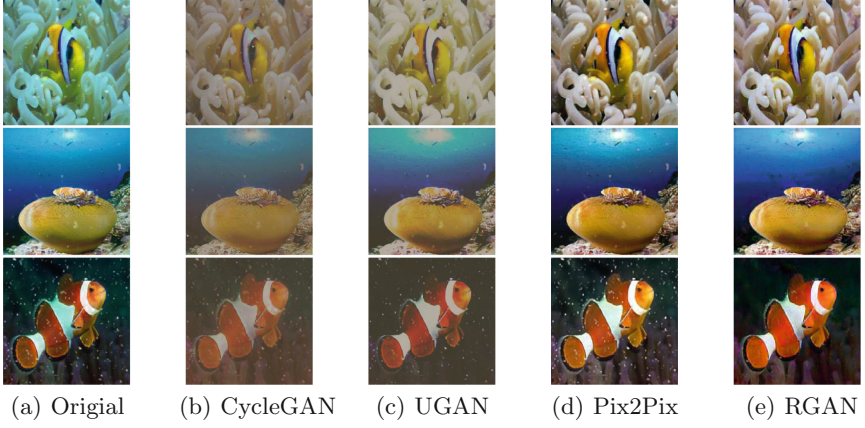
4   Experiments

In this section, we compared RSRGAN with several state-of-the-art methods for underwater image super-resolution. For the parameters, we empirically set $\lambda_{R2} = \lambda_{SR2} = 2 \times 10^{-2}$, $\lambda_{R3} = \lambda_{SR3} = 1 \times 10^{-2}$. In addition, we set $\lambda_{R1} = \lambda_{R4} = \lambda_{SR1} = \lambda_{SR4} = 1$ as normal GAN-based model. The learning rate was set to 2×10^{-5}, and the Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ was employed for the network training.

4.1   Dataset

For this research, there are no available datasets yet, which contain pairs of clear ground-truth and corresponding low quality underwater images. Following [3], we used images from ImageNet to train a CycleGAN that learned the mapping from natural to underwater images. After that, we used the CycleGAN to generate underwater images with those containing marine creatures. Finally, we added marine snow noise [1] to the underwater images. Concretely, we used 5000 image pairs for training and 1100 image pairs for test.

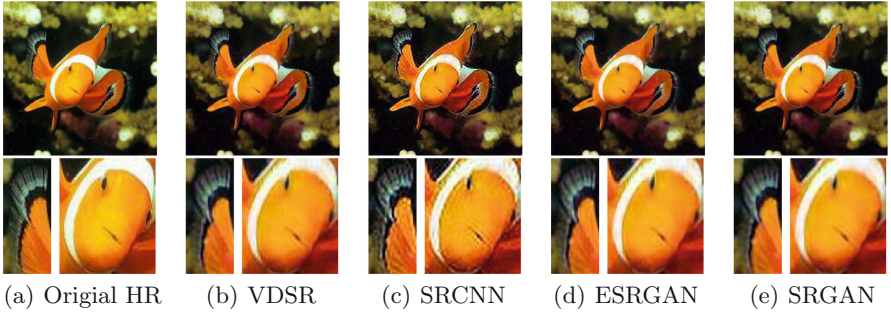### 4.2   Evaluation of the Underwater Image Restoration

Figure 4 shows some samples of the original images and the restored images obtained by Recovering GAN and some state-of-the-art image restoration methods, feeding them with the noisy underwater images. It is obvious that the restored images by CycleGAN [18] and UGAN [3] lack brightness. Pix2Pix [5] recovered the color of the images well, but the noise still remained. In contrast, Recovering GAN achieved the best performance. It not only recovered the color of the images, but also removed the noise in the images. Hence, Recovering GAN is beneficial for underwater image super-resolution.

(a) Origial   (b) CycleGAN   (c) UGAN   (d) Pix2Pix   (e) RGAN

Fig. 4. Comparison between Recovering GAN (RGAN for short here) and the state-of-the-art methods for underwater image restoration.

4.3 Evaluation of the Underwater Image Super-Resolution

Figure 5 shows the super-resolution images generated by Super-Resolution GAN, VDSR [6], SRCNN [2] and ESRGAN [2], feeding them with the restored images by Recovering GAN. We can see that Super-Resolution GAN outperforms the compared SR methods in both sharpness and details. For instance, Super-Resolution GAN can produce sharper and more natural fins and contour line of the fish's face than the compared methods.

(a) Origial HR   (b) VDSR   (c) SRCNN   (d) ESRGAN   (e) SRGAN

Fig. 5. Comparison between Super-Resolution GAN (SRGAN for short here) and the state-of-the-art methods for image SR.

4.4 Performance of RSRGAN

To quantitatively compare RSRGAN and the state-of-the-art methods, the performance of some image restoration methods followed by several SR methods

were evaluated in terms of peak signal to noise ratio (PSNR) and structural
similarity index (SSIM) [16]. The results are shown in Table 1. It is easy to see
that, the results from Recovering GAN in the last column are obviously higher
than those in the other columns. Furthermore, Super-Resolution GAN is more
powerful than the other compared SR methods. More importantly, RSRGAN
delivers the best performance in terms of both PSNR and SSIM. Particularly,
it performs better than employing Recovering GAN and Super-Resolution GAN
separately, which confirms the effectiveness of the end-to-end architecture.

Table 1. Comparison of underwater image super-resolution results in terms of PSNR
(dB)/SSIM. The first row shows the names of the image restoration methods, while
the first column shows the names of the image super-resolution methods.

SR	Restoration			
	CycleGAN [18]	UGAN [3]	Pix2Pix [5]	Recovering GAN
bicubic	16.70dB/0.5575	21.76dB/0.7195	18.87dB/0.6210	22.78dB/0.7322
VDSR [6]	16.85dB/0.5957	22.65dB/0.7811	19.21dB/0.6725	23.10dB/0.7689
EDSR [8]	16.84dB/0.5776	19.97dB/0.6404	18.67dB/0.5992	21.12dB/0.6378
ESPCN [12]	16.76dB/0.5735	21.47dB/0.6494	18.56dB/0.5978	21.12dB/0.6378
SRCNN [2]	16.72dB/0.5723	20.13dB/0.6462	18.55dB/0.5960	21.34dB/0.6440
SRGAN [7]	16.71dB/0.5747	19.68dB/0.6393	18.42dB/0.5945	20.71dB/0.6312
ESRGAN [15]	**16.87dB**/0.5906	22.54dB/0.7738	**19.23dB**/0.6633	**23.66dB/0.7806**
our SRGAN	-	-	-	**24.04dB/0.7832**
RSRGAN	**24.16dB/0.7886**			

5   Conclution

We propose an end-to-end RSRGAN model for underwater image super-
resolution. RSRGAN includes two parts: Recovering GAN and Super-Resolution
GAN. Recovering GAN corrects the color distortion and removes the noise in the
images, while Super-Resolution GAN enriches the texture details to the results
fed from Recovering GAN. RSRGAN combines the generators of Recovering
GAN and Super-Resolution GAN, while fine-tunes the entire model with the
discriminator of Super-Resolution GAN. The qualitative and quantitative com-
parison results demonstrate the superiority of RSRGAN over the state-of-the-art
methods for underwater image super-resolution.

References

1. Cyganek, B., Gongola, K.: Real-time marine snow noise removal from underwater video sequences. J. Electron. Imaging **27**(04), 043002 (2018)
2. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2016)
3. Fabbri, C., Islam, M.J., Sattar, J.: Enhancing underwater imagery using generative adversarial networks. In: ICRA, pp. 7159–7165 (2018)
4. Goodfellow, I.J., et al.: Generative adversarial networks. CoRR abs/1406.2661 (2014)
5. Isola, P., Zhu, J., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: CVPR, pp. 5967–5976 (2017)
6. Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. In: CVPR, pp. 1646–1654 (2016)
7. Ledig, C., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: CVPR, pp. 105–114 (2017)
8. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: CVPR, pp. 1132–1140 (2017)
9. Liu, Y.C., Chan, W.H., Chen, Y.Q.: Automatic white balance for digital still camera. IEEE Trans. Consum. Electron. **41**(3), 460–466 (1995)
10. Lu, H., Li, Y., Nakashima, S., Kim, H., Serikawa, S.: Underwater image super-resolution by descattering and fusion. IEEE Access **5**, 670–679 (2017)
11. Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y.: Spectral normalization for generative adversarial networks. In: ICLR (2018)
12. Shi, W., et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: CVPR, pp. 1874–1883 (2016)
13. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ICLR (2015)
14. Vaswani, A., et al.: Attention is all you need. In: NIPS, pp. 6000–6010 (2017)
15. Wang, X., et al.: ESRGAN: enhanced super-resolution generative adversarial networks. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018. LNCS, vol. 11133, pp. 63–79. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11021-5_5
16. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Processing **13**(4), 600–612 (2004)
17. Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.: Deep learning for single image super-resolution: a brief review. CoRR abs/1808.03344 (2018)
18. Li, M., Huang, H., Ma, L., Liu, W., Zhang, T., Jiang, Y.: Unsupervised image-to-image translation with stacked cycle-consistent adversarial networks. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11213, pp. 186–201. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01240-3_12