

## Image Super-Resolution Based on Generative Adversarial Networks: A Brief Review

Kui Fu<sup>1</sup>, Jiansheng Peng<sup>1, 2, \*</sup>, Hanxiao Zhang<sup>2</sup>, Xiaoliang Wang<sup>3</sup> and Frank Jiang<sup>4</sup>

**Abstract:** Single image super resolution (SISR) is an important research content in the field of computer vision and image processing. With the rapid development of deep neural networks, different image super-resolution models have emerged. Compared to some traditional SISR methods, deep learning-based methods can complete the super-resolution tasks through a single image. In addition, compared with the SISR methods using traditional convolutional neural networks, SISR based on generative adversarial networks (GAN) has achieved the most advanced visual performance. In this review, we first explore the challenges faced by SISR and introduce some common datasets and evaluation metrics. Then, we review the improved network structures and loss functions of GAN-based perceptual SISR. Subsequently, the advantages and disadvantages of different networks are analyzed by multiple comparative experiments. Finally, we summarize the paper and look forward to the future development trends of GAN-based perceptual SISR.

**Keywords:** Single image super-resolution, generative adversarial networks, deep learning, computer vision.

### 1 Introduction

Image super-resolution reconstruction technology is a research hotspot in the field of computer vision, and plays an important role in remote sensing images [Haut, Fernandez-Beltran, Paoletti et al. (2018)], medical images [Nie, Trullo, Lian et al. (2017); Guo, Cui, Yang et al. (2019); Mahapatra, Bozorgtabar and Garnavi (2019)], video surveillance and so on. Image super-resolution refers to the recovery of a corresponding high-resolution image from a low-resolution single image or a sequence of low-resolution images. Traditional single image super resolution requires multiple low-resolution images to

---

<sup>1</sup> School of Physics and Mechanical and Electronic Engineering, Hechi University, Yizhou, 546300, China.

<sup>2</sup> School of Electrical and Information Engineering, Guangxi University of Science and Technology, Liuzhou, 545006, China.

<sup>3</sup> School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan, 411100, China.

<sup>4</sup> School of Information Technology, Deakin University, Geelong, VIC 3220, Australia.

\* Corresponding Author: Jiansheng Peng. Email: sheng120410@163.com.

Received: 24 January 2020; Accepted: 05 April 2020.

restore high-frequency details of high-resolution images. However, it is sometimes difficult to obtain multiple images in some real-world scenarios.

The super resolution of a single image is an inverse problem. Since a single low-resolution image loses a lot of high-frequency information, there are many possibilities for reconstruction results. In recent years, the deep learning method has achieved remarkable results in image processing [He, Zhang, Ren et al. (2016); Huang, Liu, Maaten et al. (2017)] due to its powerful learning ability. Therefore, most of the researches attempt to use the deep learning method to perform a single image super-resolution task and has achieved good results. Moreover, a network of adversarial learning has been proposed in recent years: GAN. The GAN-based method can achieve lifelike results on image synthesis [Karras, Laine and Aila (2019)]. Most single image super resolution methods based on GAN models get lower scores in peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) evaluation experiments than traditional methods, but have higher mean-opinion-score (MOS).

In this paper, our main goal is to provide a survey of perceptual SISR based on GAN models in recent years. We focus on the challenges, structural approaches, and results of the SISR based on the GAN models. These represent the problems to be solved and possible development methods for using GAN method on image super-resolution. The following chapters are organized as follows: In Section 2, we try to explore the challenges of single image super resolution. In Section 3, the datasets and evaluation metrics for a single image super-resolution experiment are presented. In Section 4, we review the recent study of single image perceptual super-resolution using GAN methods. In Section 5, the results of a single image super-resolution reconstruction based on GAN are summarized and analyzed. In Section 6, we summarize the full text and look to the future.

## 2 Standard GAN model and SSIR challenges

The single image super-resolution based on GAN trains neural networks with the idea of zero-sum game. The generator tries to learn the distribution of the real data, and the discriminator tries to judge the authenticity of the data from the generator. In the end, both are continually optimized to achieve Nash equilibrium.

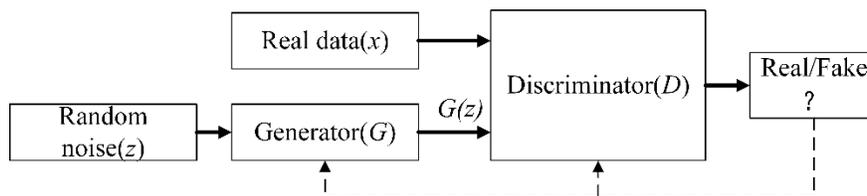


Figure 1: SGAN model

### 2.1 Standard GAN model

The standard GAN (SGAN) model [Goodfellow, Pouget-Abadie, Mirza et al. (2014)] includes two neural networks, a Generator ( $G$ ) and a Discriminator ( $D$ ). The SGAN model is shown in Fig. 1. The objective function can be described as follows:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (1)$$

where  $x$  represents a real dataset,  $p_{data}(x)$  represents a real sample distribution,  $z$  represents a random noise,  $p_z(z)$  represents a prior distribution (generally assumed to be a Gaussian distribution), and  $E(\square)$  represents an operation to calculate expected value.

### 2.2 GAN model challenge

First, the GAN-based model uses an adversarial idea to train neural networks. This method can complete the single-image super-resolution task by continuously optimizing the generator and discriminator. Second, due to the adversarial idea, this approach avoids the need to design complex loss functions to optimize the model. However, due to this adversarial idea, it has caused problems such as unstable training, gradient dispersion, and model collapse. For this open-ended problem, the optimization direction of most GAN models is in terms of architecture and loss function [Pan, Yu, Yi et al. (2019)]. Therefore, a naive idea is to apply different GAN models to SISR. However, the convergence time, memory requirements and image quality of different GAN models have advantages and disadvantages.

### 2.3 Evaluation metric challenge

In general, in the image super-resolution experiments, most image super-resolution models use the mean square error (MSE) to optimize the distance between reconstructed high-resolution images and ground truth (GT) images. This optimization method itself is beneficial to traditional super-resolution (SR) evaluation metrics: peak signal-to-noise ratio (PSNR), structural similarity (SSIM). However, using MSE to optimize models will make the reconstructed high-resolution images tend to smooth. Moreover, on the mean opinion scores (MOS), the images reconstructed by the traditional SISR method are lower than the SISR based on the GAN model. However, MOS cannot quantitatively analyze the perceptual quality of an image. Therefore, it is a better choice to apply the evaluation metric of perception combined with distortion to evaluate the perception quality of the image [Blau, Mechrez, Timofte et al. (2018)].

### 2.4 Real-world dataset challenge

Most of the existing research low-resolution images are sampled from high-resolution images using bicubic interpolation. In an unknown down-sampling experiment, the single image super-resolution evaluation metric score will decrease [Timofte, Agustsson, Van Gool et al. (2017)]. In addition, in the real world, low resolution images tend to have more complex motion. Usually, building a large-scale training set of real-world super-resolution images may solve this problem to a certain extent. Furthermore, there are also methods that use GAN architecture to learn image degradation and then adapt to real datasets [Bulat and Yang (2018)].

### 2.5 Up-sampling multiple challenge

In super-resolution experiments, the  $\times 4$  up-sampling factor is currently the mainstream study. A single image super-resolution for a higher up-sampling factor is a very

challenging problem, because an image with a higher up-sampling factor needs more high-frequency information. Therefore, it is crucial how to recover the detail information of high-resolution images in a single image higher resolution reconstruction experiment. For the problem of higher resolution image reconstruction, it can be considered to integrate the reconstruction information of different scales by using a multi-scale reconstruction method [Wang, Perazzi, McWilliams et al. (2018)].

### 3 Datasets and evaluation metrics

In order to train and evaluate a single image super-resolution model, some available datasets are needed for experiments. So far, the commonly used experimental datasets for single image super-resolution based on GAN model include: DIV2K [Agustsson and Timofte (2017)], Set5 [Bevilacqua, Roumy, Guillemot et al. (2012)], Set14 [Zeyde, Elad and Protter (2010)], BSD100 [Martin, Fowlkes, Tal et al. (2001)], Urban100 [Huang, Singh and Ahuja (2015)], PIRM [Blau, Mechrez, Timofte et al. (2018)]. Tab. 1 provides a brief summary of these datasets.

**Table 1: SSIR public datasets**

| <b>Dataset</b> | <b>Classes</b> | <b>Contents</b>                        | <b>Download link</b>  |
|----------------|----------------|--|---|
| DIV2K          | 1000           | people, cities, fauna, sceneries, etc. | <a href="https://data.vision.ee.ethz.ch/cvl/DIV2K/">https://data.vision.ee.ethz.ch/cvl/DIV2K/</a> |
| Set5           | 5              | baby, bird, butterfly, head, etc.      |   |
| Set14          | 14             | animals, humans, insects, etc.         | <a href="https://github.com/jbhua ng0604/SelfExSR">https://github.com/jbhua ng0604/SelfExSR</a>   |
| BSD100         | 100            | animals, buildings, people, etc.       |   |
| Urban100       | 100            | cities, urban, architectures, etc.     |   |
| PIRM           | 200            | people, objects, environments, etc.    | <a href="https://pirm.github.io/">https://pirm.github.io/</a>                                     |

#### 3.1 Public datasets

The following sections introduce the common training and evaluation datasets for single-image super-resolution based on the GAN model.

##### 3.1.1 Training dataset

**DIV2K dataset.** The dataset consists of high-quality images collected on the Internet, and it contains 1,000 RGB images with a resolution of 2 K. The DIV2 K dataset contains a variety of categories of content, including people, natural environments, flora and fauna, and more. The dataset is divided into a training set, a validation set, and a test set, and the numbers are 800, 100, and 100, respectively. The training set, validation set, and test set all have three down-sampling factors, which are  $\times 2$ ,  $\times 3$ , and  $\times 4$ , respectively.

##### 3.1.2 Evaluation datasets

**Set5 dataset.** The Set5 dataset contains 5 images, and the size of the images is generally small. This dataset contains high-low resolution images, and its down-sampling factors

include  $\times 2$ ,  $\times 3$ ,  $\times 4$ .

**Set14 dataset.** The Set14 dataset contains 14 images. The high-low resolution images in this dataset contain three down-sampling factors, which are  $\times 2$ ,  $\times 3$ , and  $\times 4$ , respectively.

**BSD100 dataset.** The BSD100 dataset contains 100 kinds of images from real-life scenes. There are three different high-low resolution down-sampling factors in this dataset, which are  $\times 2$ ,  $\times 3$  and  $\times 4$ , respectively.

**Urban100 dataset.** The Urban100 dataset contains 100 high-resolution images from urban scenes. The high-low resolution images in this dataset contain two down-sampling factors, which are  $\times 2$  and  $\times 4$ , respectively.

**PIRM dataset.** The PIRM dataset contains a validation set and a test set, each containing 100 pairs of high-low resolution images. Furthermore, the down-sampling factor for high-low resolution images is only  $\times 4$ . This dataset is currently used primarily to measure the perception-distortion of images.

### 3.2 Evaluation metrics

Existing image super-resolution evaluation metrics can be divided into two categories: based on reconstruction accuracy (PSNR, SSIM) and based on visual perception (PI, Root Mean Square Error).

**Peak signal to noise ratio (PSNR).** PSNR represents the ratio of the maximum power of the signal to the power of the signal noise. Given a real image  $x$  and a reconstructed image  $y$  whose pixels are both  $N$ , then MSE and PSNR can be defined as:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (x - y)^2 \quad (2)$$

$$\text{PSNR} = 10 \log_{10} \left( \frac{L^2}{\text{MSE}} \right) \quad (3)$$

Here,  $L$  is generally 255, which represents the largest pixel value of an 8-bit image.

**Structural similarity (SSIM).** SSIM is used to measure the similarity of two images, and its value is 0 to 1. A larger SSIM value indicates a smaller degree of image distortion, which means that the image quality is better. Given a real image  $x$  and a reconstructed image  $y$ , SSIM can be defined as:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\mu_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (4)$$

Here,  $\mu_x$  and  $\mu_y$ ,  $\sigma_x^2$  and  $\sigma_y^2$  represent the corresponding mean and variance of the image  $x$  and the image  $y$ , respectively. The  $\mu_{xy}$  is expressed as the covariance of  $x$  and  $y$ . For the constants  $c_1$  and  $c_2$ , they are constrained by  $c_1 = (k_1 L)^2$  and  $c_2 = (k_2 L)^2$ , respectively. Where  $k_1$  and  $k_2$  take a smaller constant (e.g., 0.01) and  $L$  represents the largest possible pixel value.

**Perceptual index (PI) and root mean square error (RMSE).** The perceptual-distortion plane consists of the PI of the vertical axis and the RMSE of the horizontal axis. By setting the RMSE value, it can be divided into three regions (regions 1/2/3 were defined by  $\text{RMSE} \leq 11.5/12.5/16$  respectively) as shown in Fig. 2. In Fig. 2, the model cannot be implemented in the area below the curve, which shows that perception and distortion can only achieve better performance by trade-off. PI combines the non-reference image

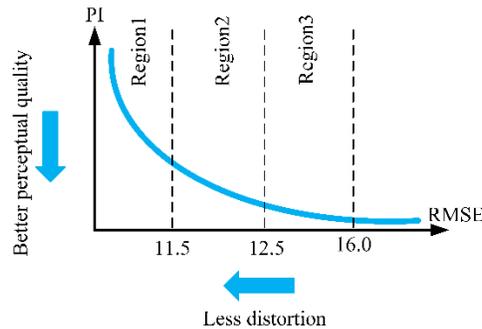
quality measurements of Ma et al. [Ma, Yang, Yang et al. (2017)] and NIQE [Mittal, Soundararajan and Bovik (2012)], which can be expressed as:

$$PI = \frac{1}{2}((10 - Ma) + NIQE) \quad (5)$$

The RMSE is calculated as the square root of the Mean Squared Error (MSE) of all pixels in all images and can be expressed as:

$$RMSE = \left( \frac{1}{M} \sum_{i=1}^M \frac{1}{N_i} \|x_i^{HR} - x_i^{EST}\|^2 \right)^{1/2} \quad (6)$$

where  $x_i^{HR}$  and  $x_i^{EST}$  represent the ground truth and estimated image, respectively.  $N_i$  represents the number of pixels in  $x_i^{HR}$  and  $M$  represents the number of images in the test set.



**Figure 2:** Perceptual-distortion plane

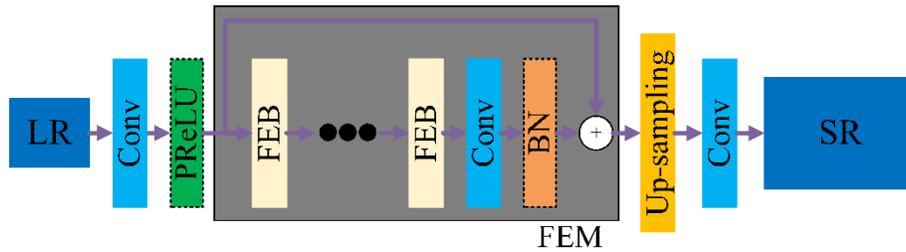
#### 4 SISR reconstruction based on GAN model

The traditional SISR method training neural networks use the  $L_1/L_2$  cost function, which makes the reconstruction results have higher PSNR and SSIM. However, the reconstructed high-resolution images lack rich details. At present, generative adversarial network (GAN) and its variants have achieved remarkable results in the field of images. The GAN-based method uses a discriminator to judge the authenticity of reconstructed high-resolution images, which makes the generated high-resolution images closer to the real images as a whole. Besides, the reconstructed images have more details and are more consistent with human visual perception. The GAN-based image super-resolution network structure is improved on the SGAN. The basic generator network and discriminator network are shown in Figs. 3 and 4, respectively. In Fig. 3, the gray box represents a feature extraction module (FEM), which includes multiple feature extraction blocks and a convolution layer (Conv) and an optional batch normalization layer (BN). In most GAN-based image super-resolution methods, their main improvement directions are discriminator networks, generator networks, and loss functions. Moreover, there are other improvement directions. For example, Shama et al. [Shama, Mechrez, Shoshan et al. (2019)] adds an adversarial feedback loop (AFL) to the SGAN. In other words, they add a feedback module based on the discriminator and generator, which enables the model to use the discriminator information during the testing stage. Next, we will introduce improvements in discriminator networks, generator networks, and loss functions. Besides, in order to maintain a good balance between RMSE and PI, Luo et al. [Luo,

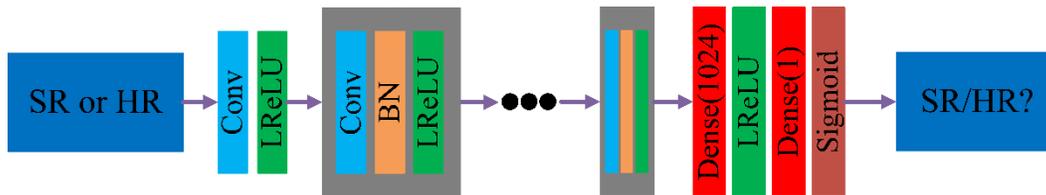
Chen, Xie et al. (2018)] used two complementary GAN networks, named Bi-GANs-ST. One memory residual GAN (MR-GAN) is used to reduce the RMSE score, and another weight-aware GAN (WP-GAN) is used to reduce the PI score. Zhong et al. [Zhong and Zhou (2020)] used a latent spatial regularization (LSR) generator network in the proposed LSRGAN. They added a companion encoder to apply regularization conditions to the GAN to generate a more ideal image manifold.

**4.1 Discriminator network**

Currently, part of the work is improved on the discriminator network [Lee, Lee, Lee et al. (2019); Park, Son, Cho et al. (2018); Wang, Perazzi, McWilliams et al. (2018)]. Lee et al. [Lee, Lee, Lee et al. (2019)] studied the discriminator of SRGAN [Ledig, Theis, Huszár et al. (2017)] and found that strided convolution layers and maxpooling layers of the VGG network would cause detailed information losses and visual artifacts. Therefore, they proposed a resolution-preserving SRGAN (RPSRGAN). First, in order to keep the details in the discriminator network, they set the stride size of the convolutional layers to 1. Then, the maxpooling layers are removed after training the VGG network. To reduce the small amount of high-frequency noise generated in the SISR based on the GAN model, Park et al. [Park, Son, Cho et al. (2018)] used an additional discriminator in the feature domain. First, a traditional image discriminator is used to process pixel-level images. Besides, a feature discriminator attempts to distinguish the authenticity of the reconstructed image based on the extracted feature map. Similarly, Zhu et al. [Zhu, Chen, Peng et al. (2020)] also used multiple feature discriminators in the proposed GAN-IMC. They used three discriminators to distinguish the three aspects of the input image: image, morphology, and color. From the experimental results of these improved discriminators, the perceptual quality of reconstructed images can be improved to a certain extent.



**Figure 3:** Generator network



**Figure 4:** Discriminator network

## 4.2 Generator network

At present, most researches on improving the GAN generator network mainly enhance the FEM of the single generator network in Fig. 3. Unlike improving a single generator network, a multi-scale generator network extends a single generator network to multiple scales for reconstruction. Next, we will introduce the improvement methods of these two parts.

### 4.2.1 FEM improvements

**Residual block (RB) with/without batch normalization.** Yu et al. [Yu and Porikli (2016)] earlier used a generative adversarial network for super-resolution reconstruction of images. They used multiple convolutional layers to build a generator network, and output face images with a  $\times 8$  up-sampling factor. It is well known that deeper networks can improve the performance of the model, but as the number of network layers increases, it will lead to gradient dispersion or gradient explosion. This problem can be solved using normalized initialization and intermediate normalization layers. However, when the network reaches a certain depth, it will cause model degradation. With the introduction of the deep residual network, this problem is solved. Therefore, early improved generator networks mostly use the residual network. Ledig et al. [Ledig, Theis, Huszár et al. (2017)] proposed SRGAN earlier, and the generator network is shown in Fig. 3. They used 16 residual blocks with BN layers in the feature extraction block (FEB), as shown in Fig. 5(a). Similarly, Lee et al. [Lee, Lee, Lee et al. (2019)] also use the same generator network.

Recently, some experimental results show that removing BN layers can improve the performance of generator networks and reduce the computational complexity of generator networks. Therefore, subsequent work removed all BN layers from the FEM. The EUSR-PCL proposed by Cheon et al. [Cheon, Kim, Choi et al. (2018)] used FEB as shown in Fig. 5(b), and removed PReLU in the dashed box in Fig. 3. Furthermore, Choi et al. [Choi, Kim, Cheon et al. (2019)] used a similar architecture in the proposed 4PP-EUSR.

**Residual block with scaling (RB-Scaling).** Model performance can be improved by increasing the number of feature maps. However, too many feature maps will cause the network training process to be unstable. Therefore, some models add a residual scaling factor (e.g., 0.1) behind the last convolution layer of the residual block to stabilize the network training process. PESR Vu et al. [Vu, Luu, Yoo et al. (2018)] and EPSR Vasu et al. [Vasu, Thekke and Rajagopalan (2018)] both used scaled residual blocks, as shown in Fig. 5 (c).

**Residual dense block (RDB) with scaling.** Recent research results show that densely connected neural network layers can alleviate the gradient dispersion problem. In addition, the densely connected networks can better reuse features. However, densely connected networks cannot be designed as deeply as residual networks. Therefore, part of the work combined the residual network and the densely connected network into a residual dense block (RDB). Wang et al. [Wang, Yu, Wu et al. (2018)] used residual dense blocks with scaling factors to improve FEB in the proposed ESRGAN, as shown in Fig. 5(d). Chen et al. [Chen, Liu, Liu et al. (2019); Shama, Mechrez, Shoshan et al. (2019)] used a similar approach. Among them, Chen et al. [Chen, Liu, Liu et al. (2019)] made a change on the densely connection network of Fig. 5(d). They move the residual scaling to skip-connection.

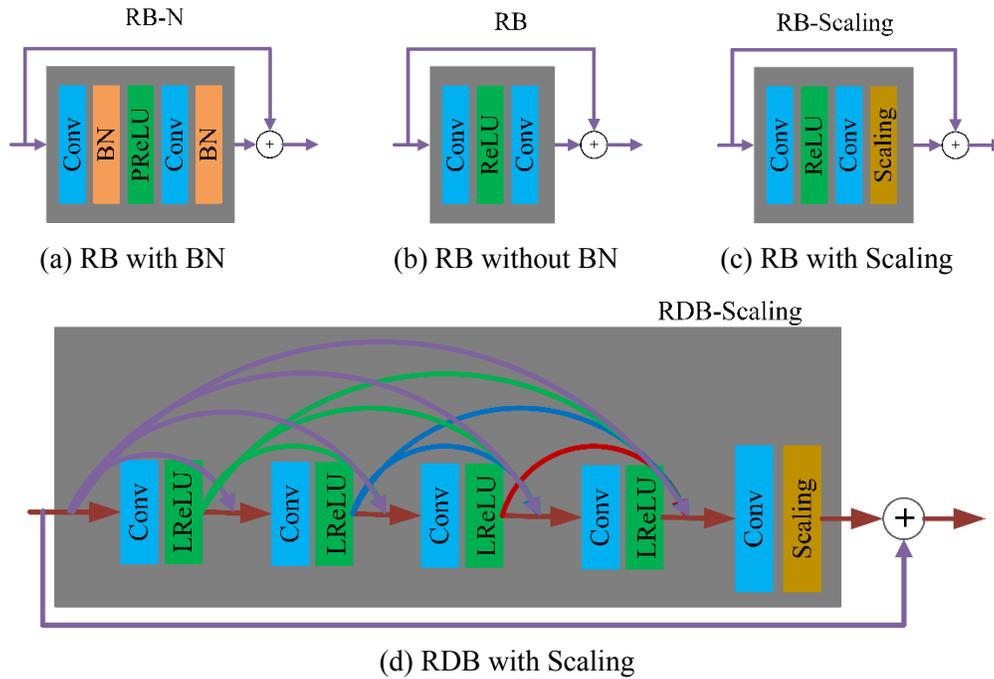


Figure 5: Different feature extraction blocks

4.2.2 Multi-scale generator network

The multi-scale generator network is shown in Fig. 6. The network can complete image super-resolution tasks in three different scales. In Fig. 6, FEB and FEM can refer to Figs. 5(b) and 3, respectively.

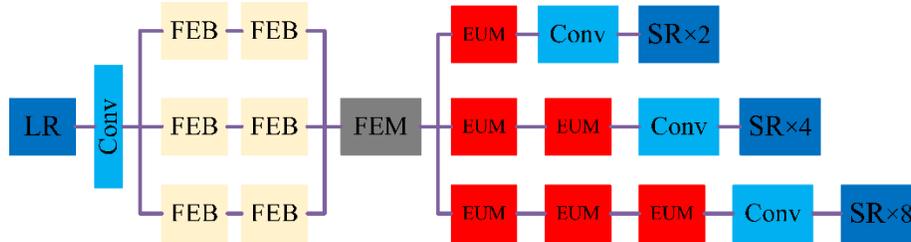
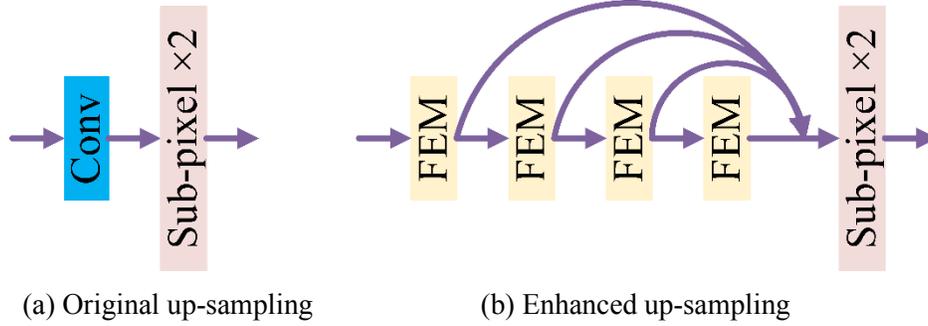


Figure 6: Multi-scale generator network

**Improved up-sampling.** The sub-pixel convolution layer was first proposed by Shi et al. [Shi, Caballero, Huszár et al. (2016)]. They fed a single LR of size  $H \times W \times C$  into convolutional layers to generate  $r^2C$  feature maps, all of which are  $H \times W$  in size. Where,  $H$ ,  $W$ , and  $C$  are the height, width, and number of channels of the LR, respectively. Moreover,  $r$  represents an up-sampling factor. The sub-pixel convolution operator rearranges generated feature maps into a single Super-resolution image of size  $rH \times rW \times C$ , as shown in Fig. 7(a). Unlike the method of generating  $r^2C$  feature maps directly through convolutional layers, Cheon et al. [Cheon, Kim, Choi et al. (2018); Choi, Kim, Cheon et

al. (2019)] both use enhanced up-sampling modules [Kim and Lee (2018)]. Specifically, they replaced Conv in the original up-sampling with 4 FEMs, as shown in Fig. 7(b). It should be noted that each FEM generates the same number of feature maps as the input.



**Figure 7:** Different up-sampling modules

### 4.3 Loss function

GAN-based perception SISR can generate natural images mainly due to the adversarial network. In order to further improve the perceptual quality of images, most studies apply different loss functions while improving the GAN network. Ledig et al. [Ledig, Theis, Huszár et al. (2017)] used a perceptual loss function in SRGAN, which includes a content loss and an adversarial loss. The content loss includes a pixel-based MSE loss ( $L_2$  loss) and a feature space-based VGG loss. Generally, the network uses MSE to ensure that the reconstructed image is similar to the ground truth image. These losses are defined as:

$$L_{MSE} = \|I_{SR} - I_{HR}\|_2^2 \quad (7)$$

Here,  $I_{SR}$  represents a reconstructed super-resolution image, and  $I_{HR}$  represents a ground truth image.

Besides, the network also uses VGG loss to improve image perceptual quality. The loss function can be expressed as:

$$L_{VGG} = \|\phi(I_{SR}) - \phi(I_{HR})\|_2^2 \quad (8)$$

where,  $\phi$  indicates the VGG19 network feature layers.

In addition to the content loss introduced above, there are the adversarial loss  $L_{G\_adv}$  of the GAN generator and the adversarial loss  $L_{D\_adv}$  of the discriminator. Their loss function is defined as:

$$L_{G\_adv} = -E_{I_{LR}} \log D(G(I_{LR})) \quad (9)$$

$$L_{D\_adv} = -E_{I_{HR}} \log D(I_{HR}) - E_{I_{LR}} \log D(1 - G(I_{LR})) \quad (10)$$

Here,  $G$  and  $D$  represent a generator and a discriminator, respectively.

Since SRGAN was proposed, most studies have attempted to improve its content loss and adversarial loss. According to the expressions in the existing literature, we reclassify the content loss into content loss (e.g.,  $L_2$  loss) and perception loss (e.g.,  $L_{VGG}$  loss). Next, we will introduce some improved loss functions.

### 4.3.1 Improved content loss

**$L_1$  loss.** In addition to using  $L_2$  loss, some work also uses  $L_1$  loss to evaluate the 1-norm distance between the reconstructed super-resolution image and the ground truth image [Cheon, Kim, Choi et al. (2018); Wang, Yu, Wu et al. (2018); Chen, Liu, Liu et al. (2019)]. The  $L_1$  loss can be calculated as:

$$L_1 = E_{I_{LR}} \|G(I_{LR}) - I_{HR}\|_1 \quad (11)$$

where,  $I_{LR}$  means a low-resolution image, and  $I_{HR}$  means a ground truth image.

**Total variance loss.** To alleviate the problem of high frequency noise amplification in GAN, Vu et al. [Vu, Luu, Yoo et al. (2018)] used a total variance loss function. The loss function is expressed as:

$$L_{TV} = -\sum_{i,j,k} (|I_{i,j+1,k}^{SR} - I_{i,j,k}^{SR}| + |I_{i,j,k+1}^{SR} - I_{i,j,k}^{SR}|) \quad (12)$$

**Rank-content loss.** In traditional image accuracy SR experiments, using MSE as a loss function is beneficial to PSNR. However, in perceptual super-resolution experiments, there is no direct loss function to optimize PI. Therefore, most perceptual SR methods cannot show stable performance. For example, different SR methods have advantages and disadvantages on different test sets. In order to alleviate this problem, Zhang et al. [Zhang, Liu, Dong et al. (2019)] proposed RankSRGAN. They train a Ranker that can learn perceptual metric behavior, and then use a Rank-content loss to optimize perceptual quality. The loss function is shown as:

$$L_{Rank} = \text{sigmoid}(R(G(I_{LR}))) \quad (13)$$

Here,  $R(G(I_{LR}))$  represents the ranking score of  $I_{LR}$ .

### 4.3.2 Improved perceptual loss

**VGG19 loss combined with Resnet50 loss.** Most GAN-based image super-resolution uses a pre-trained VGG19 network to calculate the perceptual loss. They calculate the MSE of  $I_{SR}$  and  $I_{HR}$  in the feature space, as shown in Eq. (8). The difference is that Chen et al. [Chen, Liu, Liu et al. (2019)] calculated the 1-norm distance of  $I_{SR}$  and  $I_{HR}$  in the feature space. Moreover, they combined VGG19 and Resnet50 to jointly develop a new perceptual loss. The perceptual loss can be defined as:

$$L_{VGG-Res} = E \|VGG(I_{SR}) - VGG(I_{HR})\|_1 + E \|Resnet(I_{SR}) - Resnet(I_{HR})\|_1 \quad (14)$$

**Differential content loss.** Cheon et al. [Cheon, Kim, Choi et al. (2018)] used a differential content loss to evaluate the distance between the reconstructed super-resolution image and the ground truth image at a deeper level. The loss function can be calculated as:

$$L_{DC} = \frac{1}{WH} (\sum_w |d_x I_w^{HR} - d_x I_w^{SR}| + \sum_h |d_y I_h^{HR} - d_y I_h^{SR}|) \quad (15)$$

where,  $d_x$  and  $d_y$  represent horizontal and vertical differential operators, respectively.

**Discrete cosine transform loss.** For traditional content loss, improving the distortion-based performance of the image will reduce the perceptual quality. However, perception loss can better alleviate this situation. Therefore, the application of new perceptual loss is a key to improve distortion and perception. In this problem, Cheon et al. [Cheon, Kim, Choi et al.

(2018)] used a discrete cosine transform loss function to compare the differences between two images in the frequency domain. The DCT loss function can be expressed as:

$$L_{DCT} = \frac{1}{WH} \left( \sum_w \sum_h \|DCT(I^{HR})_{w,h} - DCT(I^{SR})_{w,h}\|^2 \right) \quad (16)$$

Here,  $DCT(I)$  represents the DCT coefficient of the image  $I$ .

#### 4.3.3 Improved adversarial loss

**Relativistic GAN loss.** Because SRGAN uses an adversarial loss function of binary cross entropy, this will cause model training to be unstable. Therefore, some studies have tried to stabilize the training process of GAN models with other adversarial losses [Vu, Luu, Yoo et al. (2018); Wang, Yu, Wu et al. (2018)]. In SRGAN, the discriminator can be expressed as  $D(x)=\text{sigmoid}(C(x))$ , where  $C(x)$  represents the output of a no-transformed discriminator. Vu et al. [Vu, Luu, Yoo et al. (2018)] and Wang et al. [Wang, Yu, Wu et al. (2018)] both introduced Relativistic GAN [Jolicoeur-Martineau (2018)] for improved networks. Among them, Eqs. (17) and (18) are relativistic GAN (RGAN) loss, and Eqs. (19) and (20) are relativistic average GAN loss. These loss functions are defined as:

$$L_{G,R} = -E_{(I_{LR}, I_{HR})} [\log(\text{sigmoid}(C(G(I_{LR})) - C(I_{HR}))) \quad (17)$$

$$L_{D,R} = -E_{(I_{LR}, I_{HR})} [\log(\text{sigmoid}(C(I_{HR}) - C(G(I_{LR})))) \quad (18)$$

$$L_{G,Ra} = -E_{I_{HR}} [\log(1 - D_{Ra}(I_{HR}, G(I_{LR}))) - E_{I_{LR}} [\log(D_{Ra}(G(I_{LR}), I_{HR})) \quad (19)$$

$$L_{D,Ra} = -E_{I_{HR}} [\log(D_{Ra}(I_{HR}, G(I_{LR}))) - E_{I_{LR}} [\log(1 - D_{Ra}(G(I_{LR}), I_{HR})) \quad (20)$$

where,  $D_{Ra}(I_{HR}, G(I_{LR})) = \text{sigmoid}(C(I_{HR}) - E_{I_{LR}} [C(G(I_{LR}))])$ .

**Focal RGAN loss.** In a single-image super-resolution experiment, images with rich textures and patches are more difficult to reconstruct. To solve this problem, Vu et al. [Vu, Luu, Yoo et al. (2018)] used a focal loss to emphasize difficult training samples. They emphasize difficult samples and reduce the weight of simple samples, which increases the texture reconstruction in the images. The loss function is expressed as:

$$L_{F,RGAN} = -\sum (1-p)^\gamma \log(p) \quad (21)$$

Here,  $\gamma$  is a focusing parameter,  $p = \sigma(C(G(I_{LR})) - C(I_{HR}))$ ,  $C$  is obtained before the last sigmoid function  $\sigma$  of the discriminator.

**Other GAN losses.** In addition to the above improved adversarial losses, there are some studies that also use GAN losses in other variants. Purohit et al. [Purohit, Mandal, Rajagopalan et al. (2018)] used conditional GAN loss [Mirza and Osindero (2014)] ( $L_{CGAN}$ ) in the proposed MRDN-GAN, Lee et al. [Lee, Chuang and Wang (2019)] used Wasserstein GAN loss [Arjovsky, Chintala and Bottou (2017)] ( $L_{WGAN}$ ) in the proposed RCAN-GAN.

## 5 Results of SISR reconstruction in recent year

In this section, we analyze the advantages and disadvantages of different GAN-based SISR methods through multiple comparative experiments. It should be noted that the experimental up-sampling factor for all the analyses below is  $\times 4$ .

### 5.1 Comparison of training time and model parameters

First, we compare the training time and parameters of different models, as shown in Tab. 2. The lower the PI calculated on the PIRM-self validation set, the better the perceptual quality. “Parameters” of Tab. 2 indicates the size of the generator network. “\*” of Tab. 2 indicates that the experimental result is from RPSRGAN. As can be seen from Tab. 2, EPSR and ESRGAN can achieve better perceptual quality. However, their model parameters are 28 times and 33 times larger than that of SRGAN, respectively. Moreover, EPSR takes about 3 times more training time than SRGAN. Furthermore, it can be seen from Tab. 2 that the currently widely used training dataset is the DIV2K dataset. Since the dataset has fewer classes, the Flickr2K dataset [Lim, Son, Kim et al. (2017)] is also used in the ESRGAN experiment. They found that using additional training data with rich textures was more conducive to restoring the texture of the image.

**Table 2:** Comparison of models parameters and training time

| Method   | Training dataset | GPU           | Parameters(M) | Time(h) | PI    |
|----------|------------------|---------------|---------------|---------|-------|
| SRGAN*   | DIV2K            | TitanX Pascal | 1.54          | 12      | 2.238 |
| EPSR     | DIV2K            | TitanX Pascal | 43.07         | 45      | 2.069 |
| ESRGAN   | DIV2K            | Titan Xp      | 51.02         | -       | 2.040 |
| EUSP-PCL | DIV2K            | GTX 1080      | -             | 48      | 2.818 |
| MRDN-GAN | DIV2K            | Titan X       | >5.0          | 66      | 2.190 |
| PESR     | DIV2K            | Titan Xp      | -             | 20      | 2.130 |
| RPSRGAN  | DIV2K+Flickr2K   | TitanX Pascal | 1.54          | 26      | 2.196 |

### 5.2 Comparison of PI scores for PSNR models

At present, most GAN-based perceptual SISR studies have used PSNR-oriented network structures as generators of GAN models. Tab. 3 briefly summarizes the PSNR and PI values obtained by some PSNR models on different datasets. Where, the results of the three models of SRCNN [Dong, Loy, He et al. (2014)], VDSR [Kim, Lee, Lee et al. (2016)] and EDSR [Lim, Son, Kim et al. (2017)] are from Vu et al. [Vu, Luu, Yoo et al. (2018)]. The DBPN [Haris, Shakhnarovich and Ukita (2018)] and RCAN [Zhang, Li, Li et al. (2018)] results are from Cheon et al. [Cheon, Kim, Choi et al. (2018)] and Lee et al. [Lee, Chuang and Wang (2019)], respectively. It can be seen from Tab. 3 that the PSNR-oriented model can achieve a lower PI value to a certain extent while achieving a higher PSNR value. This shows that the PSNR and PI are not completely opposite, and better PSNR models can be used in the future to perceptual SISR experiments. Besides, it is necessary to see that the PI scores obtained by the PSNR-oriented models is far less than the PI values calculated by the high resolution datasets. This also shows that most of the current PSNR-oriented models cannot adapt well to perceptual SISR experiments. However, on the other hand, using a better PSNR model for GAN-based perceptual SISR architecture may further improve the images perceptual quality.

**Table 3:** PI comparison for PSNR-oriented models

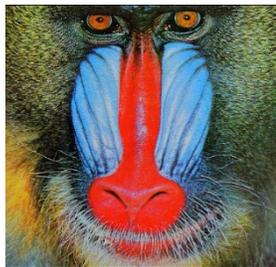
| Method | Set5       | Set14      | BSD100     | Urban100   |
|--------|------------|------------|------------|------------|
| SRCNN  | 30.48/6.79 | 27.50/6.03 | 26.90/6.04 | 24.52/5.94 |
| VDSR   | 31.35/6.45 | 28.03/5.77 | 27.29/5.70 | 25.18/5.54 |
| DBPN   | 32.47/6.10 | 28.82/5.57 | 27.72/5.54 | 26.38/5.19 |
| EDSR   | 32.46/6.00 | 28.80/5.52 | 27.71/5.40 | 26.64/5.14 |
| RCAN   | 32.63/-    | 28.87/-    | 27.77/5.14 | 26.82/4.98 |
| HR     | -/3.62     | -/3.48     | -/2.30     | -/3.70     |

### 5.3 Comparison of perceptual SISR models

Tab. 4 compares the PSNR and PI of different models on Set5, Set14, BSD100 (B100), and Urban100 (U100). It can be seen from Tab. 4 that models with lower PI scores tend to have lower PSNR scores. Combining Tabs. 3 and 4 shows that the PI values obtained by most models on different datasets are already lower than the PI scores calculated by the high-resolution datasets themselves. This shows the point that GAN-based perceptual SISR may generate excessively realistic images. Moreover, it can be seen from Fig. 8 that the image reconstructed by the GAN method is sharper and has a lower PI than the image reconstructed by the Bicubic method. However, the PSNR obtained by the GAN method is lower in terms of image accuracy (PSNR). This shows that it is significant to apply PI to evaluate the perceptual quality of the generated images.

Furthermore, SRResNet in Tab. 4 is a generator network in SRGAN, which is a PSNR-oriented model. It can also be seen from Tab. 4 that the super-resolution images generated by the GAN architecture is inferior to the PSNR model generated images in the PSNR values. However, GAN-based perceptual SISR can balance PSNR and PI through perceptual trade-off schemes. This method is more advantageous for tasks that require image quality and reconstruction accuracy in the future.

From the above analysis, for the perceptual SR methods, pursuing only the perceptual quality or distortion of the image is more one-sided. When a good reconstructed image requires good perceptual quality, it also needs good reconstruction accuracy. Therefore, a solution that can balance the accuracy of image reconstruction and the perceptual quality of the image may have more potential.



Baboon (PSNR/PI)



Bicubic (20.25/6.70)



GAN (20.14/1.95)

**Figure 8:** Visualized super-resolution results for “Baboon”

**Table 4:** Comparison of different datasets

| Method     | Datasets             | PSNR                    | PI                           |
|------------|----------------------|-------------------------|------------------------------|
| SRResNet   | Set5/Set14/B100      | 32.05/28.49/27.59       | -/-/-                        |
| EUSR-PCL   | Set5/Set14/B100      | 31.60/28.20/27.10       | 4.57/3.55/2.82               |
| GAN-IMC    | Set5/Set14/B100      | 27.89/25.67/24.20       | 3.23/2.86/2.49               |
| 4PP-EUSR   | Set5/Set14/B100      | 31.37/27.97/26.90       | 4.24/3.35/2.96               |
| EPSR       | Set5/Set14/B100/U100 | 29.6/25.5/25.0/22.8     | 3.26/2.70/ <b>2.20</b> /3.33 |
| PESR       | Set5/Set14/B100/U100 | -/-/-                   | 3.42/ <b>2.66</b> /2.25/3.41 |
| RCAN-GAN   | Set5/Set14/B100/U100 | -/-/-                   | 3.76/3.00/2.46/3.46          |
| Bi-GANs-ST | Set5/Set14/B100/U100 | 29.54/26.01/24.54/23.90 | 3.53/2.87/2.38/ <b>3.27</b>  |
| MRDN-GAN   | Set5/Set14/B100/U100 | 30.08/26.67/25.74/24.54 | 3.43/2.82/2.37/3.55          |
| LSRGAN     | Set5/Set14/B100/U100 | 30.19/26.52/25.53/24.76 | <b>3.13</b> /2.82/2.29/3.51  |

#### 5.4 Comparison of different network structures and loss functions

In this experiment, we compare the network structure and loss function of different methods. Tab. 5 shows the RMSE and PI on PIRM test set. It can be seen from Tab. 5 that the generator loss of most models currently uses a combination of pixel loss, perceptual loss, and adversarial loss, and they are also trying other losses. At present, the method adopted by ESRGAN achieved better perceptual quality in the third region. ESRGAN combined the residual dense network with the relativistic GAN loss and used some training trick. The difference is that EPSR uses a slightly worse adversarial loss function and generator network. In the end, EPSR achieved performance close to ESRGAN by setting non-negative scale factors for the generator loss. With this method, EPSR can achieve better perceptual quality in multiple regions of the perceptual-distortion plane. In a word, different architectures combined with different loss functions may achieve the most advanced results in a certain region of perceptual-distortion. However, considering the trade-off between perceptual-distortion is more likely to prevail over the entire perceptual-distortion plane.

**Table 5:** Comparison of different network structures and loss functions

| Method     | Network Description                                  | Generator Loss  | PI (R1; R2; R3)            |
|------------|--|---|----------------------------|
| EPSR       | FEM (RB-Scaling)+discriminator (Fine-tuning Fig. 4)  | $\alpha_1 L_{G\_adv} + \alpha_2 L_{VGG} + \alpha_3 L_2$                         | (2.71; 2.28; 2.01)         |
| ESRGAN     | FEM (RDB-Scaling)+discriminator (Fig. 4)             | $\alpha_1 L_{G\_Ra} + \alpha_2 L_1 + L_{VGG}$<br>(before the activation layers) | (2.93; 2.42; <b>1.98</b> ) |
| EUSR-PCL   | FEM(RB)+discriminator (Fine-tuning Fig. 4)           | $L_{G\_adv} + L_1 + L_{DC} + L_{DCT}$   | (2.75; 2.28; 2.14)         |
| PESR       | FEM (RB-Scaling)+discriminator (Typical Conv)        | $\alpha_1 L_{VGG} + \alpha_2 L_{Focal} + \alpha_3 L_{TV}$                       | (3.32; 2.60; 2.01)         |
| Bi-GANs-ST | FEM (Fine-tuning RB)+discriminator (Fig. 4)          | $\alpha_1 L_{G\_adv} + \alpha_2 L_{VGG} + L_2$                                  | (4.44; -; 2.07)            |
| MRDN-GAN   | FEM (Fine-tuning RDB-Scaling)+discriminator (Fig. 4) | $L_{G\_CGAN} + L_{VGG} + L_1$   | (3.83; 2.64; 2.13)         |

### 5.5 Comparison of real dataset results

Currently, most GAN-based perceptual SISRs are studied on bicubic degradation datasets. They use known degradation datasets in both training and testing. In order to briefly analyze the performance of the perceptual SISR based on the GAN model on the real dataset, we list some results in Tab. 6. The results of RCAN-Bic, RCAN-Real and DA-SRNet in Tab. 6 are from Gong et al. [Gong, Sun, Shi et al. (2020)]. The results of RCAN-Bic1 are from Lee et al. [Lee, Chuang and Wang (2019)]. Among them, DA-SRNet is a perceptual SISR method based on GAN network, and the other is a method oriented to PSNR model. In addition, RCAN-Bic and RCAN-Bic1 represent models trained on a bicubic degradation (known degradation) dataset, respectively. RCAN-Real and DA-SRNet represent models trained on a real dataset (unknown degradation), respectively. It can be seen from Tab. 6 that the model trained on the known degradation dataset has a lower performance on the real dataset. However, RCAN-Real trained on a real dataset can achieve better performance than RCAN-Bic trained on a known degradation dataset. Furthermore, it should be noted that the GAN-based perceptual SISR can achieve a lower PI score.

**Table 6:** Real dataset reconstruction results

| Method    | Test dataset    | Test degradation | PSNR        | PI        |
|-----------|-----------------|------------------|-------------|-----------|
| RCAN-Bic1 | BSD100/Urban100 | Bicubic          | 27.77/26.82 | 5.14/4.98 |
| RCAN-Bic1 | RealSR          | Unknown          | 27.65       | 7.84      |
| RCAN-Bic1 | RealSR          | Unknown          | 29.49       | 7.18      |
| DA-SRNet  | RealSR          | Unknown          | 27.72       | 5.24      |

## 6 Conclusion and future direction

In this paper, a single image super-resolution reconstruction based on generative adversarial network is reviewed in recent years. We introduce the challenges that faced by SSIR and also introduced common evaluation datasets. In addition, we review the direction of perceptual SSIR optimization based on the GAN model and summarize the results of partial reconstruction in recent years. Although the perceptual SSIR based on the GAN model has achieved considerable success, the super-resolution of the single image using the generative adversarial network has just started, so there is still a certain gap between the reconstructed image and the real result. Secondly, due to the problems of GAN itself, there is not much breakthrough in the latest research. However, we believe that these problems can be further solved with the optimization of neural networks. Based on the review of this paper, we propose the following possible future directions for research:

- (1) SSIR reconstruction of video. Most studies currently use DIV2K [Agustsson and Timofte (2017)] image datasets for training. At present, there are too few studies on image super-resolution using image sequences (video files). In the future, it may be considered to use video files to assist in super-resolution reconstruction of a single image.
- (2) Improving efficiency of models. Most studies pursue the expression ability of the model, but ignore the time-consuming of the model. The current state-of-the-art model (EPSR) has a very high runtime. Therefore, how to reduce the calculation amount of the model and improve the speed of the algorithm is a direction worth considering.
- (3) Establishment of new assessment measures. Most of the current research is focused on improving the PSNR and SSIM of reconstructed images, but these indexes do not correspond to the visual perception of the human eye. Although perception and distortion are used to assess the perceptual quality of an image, a low perception score may be an overly realistic result. Therefore, a more just image index is needed to measure the quality of the reconstructed image in the future.
- (4) SSIR for unknown degradation. Most current methods degrade images using known algorithms to obtain low-resolution images (e.g., bicubic interpolation). However, these methods cannot adapt well to unknown degraded images in the real world. Therefore, reconstruction of low resolution images of unknown degraded can be considered in the future [Bulat, Yang, Tzimiropoulos et al. (2018); Gong, Sun, Shi et al. (2020)].
- (5) Design of GAN models. On the one hand, the performance improvement of perceptual super-resolution comes from loss functions. On the other hand, the performance improvement of neural networks comes from design of GAN models. In perceptual SR, it is a natural choice to use the latest image feature extraction networks to improve the GAN model. Therefore, in the future, other advanced neural network structures or new network structures can be considered to extract image features.

**Funding Statement:** The authors are highly thankful to the Development Research Center of Guangxi Relatively Sparse-populated Minorities (ID: GXRKJSZ201901), to the Natural Science Foundation of Guangxi Province (No. 2018GXNSFAA281164). This research was financially supported by the project of outstanding thousand young teachers' training in higher education institutions of Guangxi, Guangxi Colleges and Universities Key

Laboratory Breeding Base of System Control and Information Processing.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- Agustsson, E.; Timofte, R.** (2017): Ntire 2017 challenge on single image super-resolution: dataset and study. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 126-135.
- Arjovsky, M.; Chintala, S.; Bottou, L.** (2017): Wasserstein gan. *arXiv preprint arXiv:1701.07875*.
- Bevilacqua, M.; Roumy, A.; Guillemot, C.; Alberi Morel, M. L.** (2012): Low-complexity single-image super-resolution based on nonnegative neighbor embedding. *British Machine Vision Conference*.
- Blau, Y.; Mechrez, R.; Timofte, R.; Michaeli, T.; Zelnikmanor, L. et al.** (2018): The 2018 PIRM challenge on perceptual image super-resolution. *European Conference on Computer Vision*, pp. 334-355.
- Bulat, A.; Yang, J.; Tzimiropoulos, G.** (2018): To learn image super-resolution, use a gan to learn how to do image degradation first. *European Conference on Computer Vision*, pp. 185-200.
- Chen, B. X.; Liu, T. J.; Liu, K. H.; Liu, H. H.; Pei, S. C.** (2019): Image super-resolution using complex dense block on generative adversarial networks. *IEEE International Conference on Image Processing*, pp. 2866-2870.
- Cheon, M.; Kim, J. H.; Choi, J. H.; Lee, J. S.** (2018): Generative adversarial network-based image super-resolution using perceptual content losses. *European Conference on Computer Vision*, pp. 51-62.
- Choi, J. H.; Kim, J. H.; Cheon, M.; Lee, J. S.** (2019): Deep learning-based image super-resolution considering quantitative and perceptual quality. *Neurocomputing*, vol. 398, pp. 347-359.
- Dong, C.; Loy, C. C.; He, K.; Tang, X.** (2014): Learning a deep convolutional network for image super-resolution. *European Conference on Computer Vision*. Springer, Cham, pp. 184-199.
- Gong, D.; Sun, W.; Shi, Q.; Hengel, A. V. D.; Zhang, Y.** (2020): Learning to Zoom-in via Learning to Zoom-out: real-world super-resolution by generating and adapting degradation. *arXiv preprint arXiv:2001.02381*.
- Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D. et al.** (2014): Generative adversarial nets. *Advances in Neural Information Processing Systems*, pp. 2672-2680.
- Guo, Y.; Cui, Z.; Yang, Z.; Wu, X.; Madani, S.** (2019): Non-Local DWI image super-resolution with joint information based on GPU implementation. *Computers, Materials & Continua*, vol. 61, no. 3, pp. 1217-1231.

- Haris, M.; Shakhnarovich, G.; Ukita, N.** (2018): Deep back-projection networks for super-resolution. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1664-1673.
- Haut, J. M.; Fernandez-Beltran, R.; Paoletti, M. E.; Plaza, J.; Plaza, A. et al.** (2018): A new deep generative network for unsupervised remote sensing single-image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 11, pp. 6792-6810.
- He, K.; Zhang, X.; Ren, S.; Sun, J.** (2016): Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778.
- Huang, G.; Liu, Z.; Maaten, L. V. D.; Weinberger, K. Q.** (2017): Densely connected convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700-4708.
- Huang, J. B.; Singh, A.; Ahuja, N.** (2015): Single image super-resolution from transformed self-exemplars. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5197-5206.
- Jolicœur-Martineau, A.** (2018): The relativistic discriminator: a key element missing from standard GAN. *arXiv preprint arXiv:1807.00734*.
- Karras, T.; Laine, S.; Aila, T.** (2019): A style-based generator architecture for generative adversarial networks. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4401-4410.
- Kim, J. H.; Lee, J. S.** (2018): Deep residual network with enhanced upscaling module for super-resolution. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 800-808.
- Kim, J.; Lee, J. K.; Lee, K. M.** (2016): Accurate image super-resolution using very deep convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1646-1654.
- Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A. et al.** (2017): Photo-realistic single image super-resolution using a generative adversarial network. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4681-4690.
- Lee, D.; Lee, S.; Lee, H.; Lee, K.; Lee, H. J.** (2019): Resolution-preserving generative adversarial networks for image enhancement. *IEEE Access*, vol. 7, pp. 110344-110357.
- Lee, W. Y.; Chuang, P. Y.; Wang, Y. C. F.** (2019): Perceptual quality preserving image super-resolution via channel attention. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1737-1741.
- Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K. M.** (2017): Enhanced deep residual networks for single image super-resolution. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 136-144.
- Luo, X.; Chen, R.; Xie, Y.; Qu, Y.; Li, C.** (2018): Bi-GANs-ST for perceptual image super-resolution. *European Conference on Computer Vision*, pp. 20-34.

- Ma, C.; Yang, C. Y.; Yang, X.; Yang, M. H.** (2017): Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, vol. 158, pp. 1-16.
- Mahapatra, D.; Bozorgtabar, B.; Garnavi, R.** (2019): Image super-resolution using progressive generative adversarial networks for medical image analysis. *Computerized Medical Imaging and Graphics*, vol. 71, pp. 30-39.
- Martin, D.; Fowlkes, C.; Tal, D.; Malik, J.** (2001): A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *IEEE International Conference on Computer Vision*, pp. 416-423.
- Mirza, M.; Osindero, S.** (2014): Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.
- Mittal, A.; Soundararajan, R.; Bovik, A. C.** (2012): Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209-212.
- Nie, D.; Trullo, R.; Lian, J.; Petitjean, C.; Ruan, S. et al.** (2017): Medical image synthesis with context-aware generative adversarial networks. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 417-425.
- Pan, Z.; Yu, W.; Yi, X.; Khan, A.; Yuan, F. et al.** (2019): Recent progress on generative adversarial networks (GANs): a survey. *IEEE Access*, vol. 7, pp. 36322-36333.
- Park, S. J.; Son, H.; Cho, S.; Hong, K. S.; Lee, S.** (2018): Srfeat: single image super-resolution with feature discrimination. *European Conference on Computer Vision*, pp. 439-455.
- Purohit, K.; Mandal, S.; Rajagopalan, A. N.** (2018): Scale-recurrent multi-residual dense network for image super-resolution. *European Conference on Computer Vision*.
- Shama, F.; Mechrez, R.; Shoshan, A.; Zelnik-Manor, L.** (2019): Adversarial feedback loop. *IEEE International Conference on Computer Vision*, pp. 3205-3214.
- Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A. P. et al.** (2016): Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1874-1883.
- Timofte, R.; Agustsson, E.; Van Gool, L.; Yang, M. H.; Zhang, L.** (2017): Ntire 2017 challenge on single image super-resolution: Methods and results. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 114-125.
- Vasu, S.; Thekke, M. N.; Rajagopalan, A. N.** (2018): Analyzing perception-distortion tradeoff using enhanced perceptual super-resolution network. *European Conference on Computer Vision*, pp. 114-131.
- Vu, T.; Luu, T. M.; Yoo, C. D.** (2018): Perception-enhanced image super-resolution via relativistic generative adversarial networks. *European Conference on Computer Vision*, pp. 98-113.
- Wang, X.; Yu, K.; Wu, S.; Gu, J.; Dong, C. et al.** (2018): Esrgan: enhanced super-resolution generative adversarial networks. *European Conference on Computer Vision*, pp. 63-79.

- Wang, Y.; Perazzi, F.; McWilliams, B.; Sorkine-Hornung, A.; Sorkine-Hornung, O. et al.** (2018): A fully progressive approach to single-image super-resolution. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 864-873.
- Yu, X.; Porikli, F.** (2016): Ultra-resolving face images by discriminative generative networks. *European Conference on Computer Vision*, pp. 318-333.
- Zeyde, R.; Elad, M.; Protter, M.** (2010): On single image scale-up using sparse-representations. *International Conference on Curves and Surfaces*, pp. 711-730.
- Zhang, W.; Liu, Y.; Dong, C.; Qiao, Y.** (2019): Ranksgan: Generative adversarial networks with ranker for image super-resolution. *IEEE International Conference on Computer Vision*, pp. 3096-3105.
- Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B. et al.** (2018): Image super-resolution using very deep residual channel attention networks. *European Conference on Computer Vision*, pp. 286-301.
- Zhong, S.; Zhou, S.** (2020): Optimizing generative adversarial networks for image super resolution via latent space regularization. *arXiv preprint arXiv:2001.08126*.
- Zhu, X.; Cheng, Y.; Peng, J.; Wang, R.; Le, M. et al.** (2020): Super-resolved image perceptual quality improvement via multifeature discriminators. *Journal of Electronic Imaging*, vol. 29, no. 1, pp. 013017.