IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 20, 2020, accepted October 6, 2020, date of publication October 12, 2020, date of current version October 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3030044

Least Squares Relativistic Generative Adversarial Network for Perceptual Super-Resolution Imaging

SANYOU ZHANG[1,2], DEQIANG CHENG[1], DAIHONG JIANG[3], AND QIQI KOU[1]

[1]School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221000, China
[2]Department of Science and Technology, Suzhou Wujiang District Public Security Bureau Population Management Brigade, Suzhou 215200, China
[3]Information and Electrical Engineering College, Xuzhou University of Technology, Xuzhou 221000, China

Corresponding author: Deqiang Cheng (chengdq@cumt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51774281, in part by the National Key Research and Development Program of China under Grant 2018YFC0808302, in part by the Major Project of Natural Science Research of the Jiangsu Higher Education Institutions of China under Grant 18KJA520012, and in part by the Xuzhou Science and Technology Plan Project under Grant KC19197.

ABSTRACT Currently, deep-learning-based methods have been the most popular super-resolution techniques owing to the improvement of super-resolution performance. However, they are still lack perceptual fine details and thus result in unsatisfying visual quality. This article proposes a novel method for high-quality perceptual super-resolution imaging, named SRLRGAN-SN. It aims to recovery visually plausible images with perceptual texture details by using the least squares relativistic generative adversarial network (GAN). The method applies the spectral normalization on the network with the target of enhancing the performance of GAN for super-resolution task. The least squares relativistic discriminator is designed to drive reconstruction images approximating high-quality perceptual manifold. Besides, a novel perceptual loss assembly is proposed to preserve structural texture details as much as possible. Results of experiment show that our method can not only recovery more visually realistic details, but also outperforms other popular methods regarding to quantitative metrics and perceptual evaluations.

INDEX TERMS Generative adversarial network, super-resolution imaging, relativistic discriminator, perceptual quality, spectral normalization.

LIST OF ACRONYMS

GAN	Generative Adversarial Network
CNN	Convolutional Neural Network
PSNR	Peak Signal to Noise Ratio
SSIM	Structural Similarity Index Measure
ReLU	Rectified Linear Unit
PReLU	Parametric Rectified Linear Unit
FID	Frechet Inception Distance
MSE	Mean Square Error
PI	Perceptual Index
SN	Spectral Normalization
JS	Jensen-Shannon
LR	Low-Resolution
HR	High-Resolution
SR	Super-Resolution
SRCNN	SR CNN
VDSR	Very Deep SR network
EDSR	Enhanced Deep SR network
DRRN	Deep Recursive Residual Network
CARN	Cascading Residual Network
SRGAN	SR GAN
ESRGAN	Enhanced SRGAN
PESRGAN	Perception-Enhanced SRGAN
NatSRGAN	Natural SRGAN
WGAN	Wasserstein GAN
WGAN-GP	Wasserstein GAN-Gradient Penalty
LSGAN	Least Squares GAN
RaGAN	Relativistic Average GAN

The associate editor coordinating the review of this manuscript and approving it for publication was Andrea F. Abate.

I. INTRODUCTION

As a typical ill-posed issue during restoration [1], [2], super-resolution for imaging reconstruction aims to improve

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

spatial resolution by digital signal processing without changing the existing hardware. Due to the strong fitting ability of deep learning, super-resolution methods for imaging task realize a great leap of improvement. The applications are extensive from surveillance imaging enhancement, remote sensing system, object target recognition and other computer vision scenarios [3], [4]. Recently, super-resolution techniques for imaging with convolutional neural network (CNN) have shown better performance than traditional methods [5], [6]. Most of CNN-based super-resolution methods use pixel loss for training to seek improvements of typically quantitative metrics in terms of peak signal-to-noise ratio (PSNR). While pixel loss can be easily optimized, it usually fails to provide pleasant realistic details in accordance with perceptual vision, which trends toward distortion especially for large scale factors. Due to the dramatic development of generative adversarial network [7], [8] in generating photo-realistic images, it provides a new approach for perceptual super-resolution imaging. For the benefit of perceptual loss and adversarial loss employed in loss minimization, GAN-based super-resolution methods achieve greatly improvements of abundant visual perception compared with CNN-based methods. However, owing to the drawbacks of training GAN problems such as vanishing gradient, difficult optimization, mode collapse, among others, GAN-based methods suffers from limitations as following issues [9], [10]. GANs are quite difficult to train and tend quickly to collapse because the overfitting of one of the networks it comprises. The discriminator should be trained suitably, neither too well nor too bad. If the discriminator trained too well, the gradient will disappear easily. If the discriminator trained too bad, it is difficult to distinguish between the real true sample and the generated fake distribution. The most ideal state is the discriminator trained to be just right, but it is difficult to grasp the state during the training process. Firstly, original GAN for super-resolution imaging excruciatingly difficult to train for its fickle and unstable inherent property. Secondly, When training the generator, the real high-resolution samples are not involved, so the discriminator must remember all attributes about real samples, resulting in the performance bottleneck of guiding the generator to further produce more realistic images. Thirdly, for the lack of texture guided optimization in the loss function, it can't fully maintain structural information of geometric textures. Besides, common metrics including PSNR and SSIM exist some disadvantages that they are not suitable for measuring perceptual similarity that human visions can received.

In this article, to cope with these issues indicated above, the least squares relativistic generative adversarial network is proposed for single image super-resolution. It aims to generate visually plausible images with perceptual texture details. Major contributions of this study are outlined below:

1) The spectral normalization is applied on the network with the target of enhancing the performance of GAN for super-resolution task.
2) The least squares relativistic discriminator is designed to drive reconstruction images approximating high-quality perceptual manifold.
3) A novel perceptual loss assembly is proposed to enhance realistic texture details as much as possible with the weighting sum of the content loss, feature loss, texture loss, and least squares relativistic adversarial loss.

II. RELATED WORK

A. SUPER-RESOLUTION FOR IMAGING

The goal of the super-resolution task is to recover details and enhance resolution of images. There are various techniques for super-resolution tasks [11]. Early classical solutions are interpolation-based methods like bilinear and bicubic methods and so on [12], which are easy to realize but trend to recover blurry images. More sophisticated approaches are learning-based methods such as neighborhood embedding algorithms, example-based algorithms and so on [13], which are basically follow the framework of sparse coding but the produced reconstruction results are not ideal. Currently, deep-learning-based methods have been the most popular techniques owing to greatly improve the performance of super-resolution tasks. There are two main approaches to deep-learning-based methods. The first approach is CNN-based methods, which employ various convolutional neural networks along with skip connections [14]. One of the first literature put forward by Dong *et al.* [15] is the prominent model named as SRCNN. Kim *et al.* [16] put forward the VDSR model to gain superior super-resolution performance by training a deeper super-resolution network. Lim *et al.* [17] propose the EDSR method to obtain higher accuracy by using an enhanced deep residual network. The research of reference [18] provides a novel approach for super-resolution imaging to significantly improve the training process of network, which uses visual attention component within a deep residual network. Tai *et al.* [19] propose the DRRN method, which the depth of convolutional network up to 52 layers with the global and local residual learning. Ahn *et al.* [20] propose the CARN method, which connects all the layers densely to achieve lightweight super-resolution with cascading residual network. CNN-based methods always try to minimize the pixel loss to achieve high quantitative value in terms of PSNR and SSIM [21]. However, they are still lack fine details and thus result in unsatisfying perceptual quality. To overcome these issues, the second approach is GAN-based methods, aiming to provide novel solutions for generating plausible images with photo-realistic perceptual quality. GAN is first introduced in the super-resolution field by Ledig *et al.* [22], also known as SRGAN, which outperforms previous super-resolution techniques in perceptual vision. Recent progress in GAN-based methods sparks various new solutions. Sajjadi *et al.* [23] propose the EnhanceNet method using feed-forward fully network, with the benefits of combining feature-space loss and texture-matching loss for better image quality at high magnification scales. Wang *et al.* [24]

further extend the SRGAN architecture and propose the ESRGAN method, which employs residual-scaling dense network and relativistic discriminator to improve the realism of images. Vu *et al.* [25] employ the relativistic discriminator and put forward the PESRGAN model coupling with focal loss and total variance loss. Soh *et al.* [26] put forward the NatSRGAN model through the use of the natural manifold discriminator. In addition to the typical reconstruction loss, naturalness loss and relativistic adversarial loss are added in the NatSRGAN to further encourage images towards more natural manifold. Haut *et al.* [27] propose a new generative network for unsupervised super-resolution task by using convolutional generator model, and apply the proposed technique to super-resolve remote sensing imagery.

B. SPECTRAL NORMALIZATION FOR GAN

The instability is the well-known issue of GAN's training. Arjovsky *et al.* [28] analyze the reason that the objective function of GAN is based on the Jensen-Shannon (JS) divergence. To deal with the issue, Arjovsky *et al.* put forward the solution of Wasserstein GAN (WGAN) [28] and its improved approach WGAN-GP [29], which replace JS divergence with Wasserstein distance to stabilize the training of the discriminator. However, there are also some problems in the training of WGAN and WGAN-GP, such as slow convergence, high computational complexity and weight clipping resulting gradient explosion. Recently, Miyato *et al.* [30] propose SNGAN by restricting the spectral normalization of each layer to ensure the discriminator consistent with the Lipschitz continuity, where the spectral normalization is the largest singular value of the weight matrix. The integration of the spectral normalization technique enables the the discriminator network to be trained stablely. It has been showed by various studies that spectral normalization for GAN is more effective and efficient than previous approaches to enhance the optimization process. Moreover, Odena *et al.* [31] prove that the well-conditioned generator can help to enhance GAN's performance. In view of this conclusion, Zhang *et al.* [32] propose SAGAN by employing spectral normalization in both the generator and the discriminator network for stabilizing training, which not only owns low computational cost of optimization, but also increases the stability of training process. Hence, following from these studies, spectral normalization is applied on both networks of SRLRGAN-SN for more stable and efficient training.

C. PERCEPTUAL QUALITY ASSESSMENT

Traditionally, quantitative metrics such as PSNR and SSIM are extensively used to evaluate image quality. However, none of them can consistently correlate very well with human perception [33]. Thus, these metrics alone are not suitable for differentiating and evaluating image perceptual quality. To precisely assess perceptual quality of super-resolution methods, a perception-driven metric referred as perceptual index (PI) [33] is proposed for perceptual super-resolution tasks. As shown in [33], the PI indicator is more fit to assess GAN-based approaches, and the less the value acquired, the higher the quality exhibited. PI is calculated as follows:

$$PI(\mathbf{I}) = \frac{1}{2}((10 - \mathrm{Ma}(\mathbf{I})) + \mathrm{NIQE}(\mathbf{I})) \tag{1}$$

where $\mathbf{I}$ is the image, Ma(.) and NIQE(.) represent the non-reference quality assessments proposed in [34] and [35], respectively. There is a high correlation between PI and human perception. In order to ensure that the resolved image is similar to the real sample in pix-wise content, this study explores the metric PI couple with PSNR to measure the reconstructed performance.

III. SRLRGAN-SN METHOD

The overall model architecture of SRLRGAN-SN method extends from SRGAN. Compared with SRGAN, the spectral normalization is applied on both the generative and discriminative network with the target of enhancing the performance of GAN for super-resolution task. In addition, there is a least squares relativistic discriminator used to differ between the generated super-resolution images and original high-resolution samples. It identifies the relative boundary difference instead of the dichotomy difference between generated images and real samples. Furthermore, a novel perceptual loss assembly is employed to further drive images towards more realistic manifold. SRLRGAN-SN is involved two adversarial networks comprised of generator and discriminator depicted in Figure 1. Network topology for generator and discriminator are described in Table 1 and Table 2, respectively. Layers with the same type are signed in the same color, where k, n, s represent the size of convolution kernel, the number of feature maps and the size of convolution stride, respectively. The generator network is mainly composed of 5 residual blocks for producing high-quality images. Each residual block consists of two convolution layers, two spectral normalization layers and one activation layer with PReLU function in contact with skip connection in ResNet [36]. In front of residual blocks, a convolution layer and an activation layer are employed for feature extraction. Two sub-pixel convolution layers in the tail are used for up-sampling. The network of discriminator is in charge of distinguishing produced images from true samples. The discriminator is increased from 64 to 512 feature maps with the network mainly consisting of 8 convolutional layers. Each convolutional layer is followed by one spectral normalization layer and activated with Leaky ReLU function [22]. The network of discriminator ends with two dense layers to return the probability that the true high-resolution sample is more plausible than the fake produced image.

It is well known that designing an effective loss function is very key for outputting pleasant results. The multiple perceptual loss combination for optimizing is integrated to the SRLRGAN-SN framework. Apart from the least squares relativistic adversarial loss in providing better overall visual quality for super-resolution, the error of the generator is also calculated through the content loss, feature loss as well

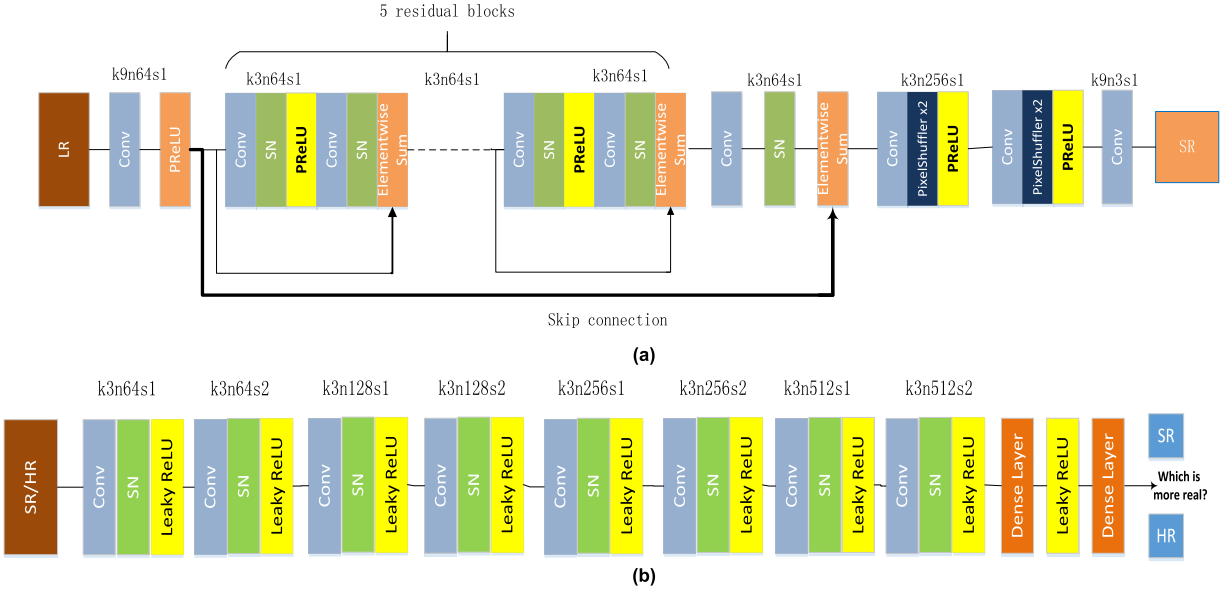

FIGURE 1. SRLRGAN-SN overall model architecture. (a) Network of generator; (b) Network of discriminator.

TABLE 1. Network topology for generator.

CONV_ID	Kernel size	Number of kernels	Stride
Conv before Residual Block			
conv_1	9x9	64	1
Residual Block			
conv_2	3x3	64	1
conv_3	3x3	64	1
...	...	...	...
conv_11	3x3	64	1
Conv after Residual Block			
conv_12	3x3	64	1
Up-sampling			
conv_13	3x3	256	1
conv_14	3x3	256	1
Output			
conv_15	9x9	3	1

TABLE 2. Network topology for discriminator.

CONV_ID	Kernel size	Number of kernels	Stride
conv_1	3x3	64	1
conv_2	3x3	64	2
conv_3	3x3	128	1
conv_4	3x3	128	2
conv_5	3x3	256	1
conv_6	3x3	256	2
conv_7	3x3	512	1
conv_8	3x3	512	2

as texture loss to recover finer texture details. In this case, each loss function in the combination provides an unique perspective on GAN-based perceptual super-resolution tasks.

A. CONTENT LOSS

Conventional content loss generally employs mean-square-error(MSE) loss to measure the pixel-based content similarity. It is prone to causing overly smooth results. In this study, we introduce the Charbonnier loss [37] as content loss to maintain edge details. As shown in [37], it provides pixel-space regularization for loss optimization and contributes to the qualitative improvement. The content loss is calculated as follows:

$$L_{con} = \frac{1}{T}\sum_{t=1}^{T}\sqrt{\left(G_{\theta_G}\left(I_t^{LR}\right) - I_t^{HR}\right)^2 + \varepsilon^2} \tag{2}$$

L_{con} represents the content loss. Where $G_{\theta_G}\left(I_t^{LR}\right)$ is the produced image and $I_t^{HR}, t = 1, \ldots, T$ is the true original sample corresponding with low-resolution sample $I_t^{LR}, t = 1, \ldots, T$. θ represents network parameters of the generator. ε is a minute constant term near to 0, which denotes the influence of the Charbonnier penalty.

B. FEATURE LOSS

Conventional feature loss for measuring the semantic perception is following in [38]. It employs feature maps extracted after activation in perceptual loss. However, Wang *et al.* [24] have demonstrated that employing feature maps before

activation can producing more accurate texture details. Thus, following [24], we use the pre-trained vgg-19 model to extract the feature representation of images generated by SRLRGAN-SN and the true original samples. The feature loss is defined as follows:

$$L_{fea} = \left\| \phi\left(I^{HR}\right) - \phi\left(G_{\theta_G}\left(I^{LR}\right)\right) \right\|_2^2 \tag{3}$$

L_{fea} represents the feature loss. Where $G_{\theta_G}\left(I^{LR}\right)$ and I^{HR} are the produced image and the true original sample, respectively. ϕ denotes the feature mapping from the vgg-19 model.

C. TEXTURE LOSS

Texture loss is the metric measuring the structural style similarity. It is presented by Gatys *et al.* [39] to enable texture details to look convincingly in their style and context. By adding this loss, it is motivated to to further push images towards realistic textures and visually much closer style as much as possible. The texture loss is defined as the correlation between feature maps described as follows:

$$L_{tex} = \left\| Gram\left(\phi\left(G_{\theta_G}\left(I^{LR}\right)\right)\right) - Gram\left(\phi\left(I^{HR}\right)\right) \right\|_2^2 \tag{4}$$

L_{tex} represents the texture loss. Where $G_{\theta_G}\left(I^{LR}\right)$ and I^{HR} are the produced image and the true original sample, respectively. ϕ denotes the feature mapping from the vgg-19 model. $Gram(F) = FF^T$ represents the Gram Matrix of feature layer F multiplied on transposed self.

D. LEAST SQUARES RELATIVISTIC ADVERSARIAL LOSS

In the original GAN for super-resolution tasks, the discriminator maximizes its capability to only distinguish real or fake images. Recently, the relativistic discriminator is proposed by Jolicoeur-Martineau [40], which can not only generate more higher-quality images than original GAN but also stand out the impressive stabilization for super-resolution tasks in [24], [25]. Unlike original GAN, relativistic GAN estimates the probability that the real sample is more realistic than the generated image, which not only real samples but also generated images are involved in the adversarial learning process. The relativistic average GAN (RaGAN) presented in [40] for adversarial training is calculated as follows:

$$\begin{aligned} L_D^{RaGAN} &= -\mathbb{E}_{x_r \sim P}\left[\log\left(\tilde{D}\left(x_r\right)\right)\right] \\ &\quad - \mathbb{E}_{x_f \sim Q}\left[\log\left(1 - \tilde{D}\left(x_f\right)\right)\right] \end{aligned} \tag{5}$$

$$\begin{aligned} L_G^{RaGAN} &= -\mathbb{E}_{x_f \sim Q}\left[\log\left(\tilde{D}\left(x_f\right)\right)\right] \\ &\quad - \mathbb{E}_{x_r \sim P}\left[\log\left(1 - \tilde{D}\left(x_r\right)\right)\right] \end{aligned} \tag{6}$$

$$\tilde{D}\left(x_r\right) = \sigma\left(C\left(x_r\right) - \mathbb{E}_{x_f \sim Q} C\left(x_f\right)\right) \tag{7}$$

$$\tilde{D}\left(x_f\right) = \sigma\left(C\left(x_f\right) - \mathbb{E}_{x_r \sim P} C\left(x_r\right)\right) \tag{8}$$

where $x_r \sim P$ and $x_f \sim Q$ represent the distribution of real samples and generated images in [40], respectively. $C(.)$ signifies the output of non-transformed discriminator. σ refers to the sigmoid function.

To further improve the ability of the relativistic discriminator for super-resolution imaging, the least squares relativistic discriminator is employed to drive produced images highly approximating the quality of real high-resolution samples. The Least Squares GAN (LSGAN) proposed by Mao *et al.* [41] is an extension to the original GAN, which penalizes generated samples according their distance from the decision boundary to encourage higher-quality generation. Inspired by the idea of LSGAN devising the boundary to best separate real samples and generated images, the least squares relativistic adversarial loss derived from RaGAN is designed to drive images approximating high-quality perceptual manifold, which is defined as follows:

$$\begin{aligned} L_D^{ADV} &= \mathbb{E}\left[\left(C\left(I^{HR}\right) - \mathbb{E}\left[C\left(G_{\theta_G}\left(I^{LR}\right)\right)\right] - 1\right)^2\right] \\ &\quad + \mathbb{E}\left[\left(C\left(G_{\theta_G}\left(I^{LR}\right)\right) - \mathbb{E}\left[C\left(I^{HR}\right)\right] + 1\right)^2\right] \end{aligned} \tag{9}$$

$$\begin{aligned} L_G^{ADV} &= \mathbb{E}\left[\left(C\left(G_{\theta_G}\left(I^{LR}\right)\right) - \mathbb{E}\left[C\left(I^{HR}\right)\right] - 1\right)^2\right] \\ &\quad + \mathbb{E}\left[\left(C\left(I^{HR}\right) - \mathbb{E}\left[C\left(G_{\theta_G}\left(I^{LR}\right)\right)\right] + 1\right)^2\right] \end{aligned} \tag{10}$$

where $G_{\theta_G}\left(I^{LR}\right)$ and I^{HR} are the produced image and the true original sample, respectively. $C(.)$ signifies the output of non-transformed discriminator.

E. TOTAL LOSS

Hence, the cost of total loss is calculated by weighted summing up four losses indicated above, which is calculated as follows:

$$L_G = \alpha L_{con} + \beta L_{fea} + \gamma L_{tex} + \delta L_G^{ADV} \tag{11}$$

$$L_D = \delta L_D^{ADV} \tag{12}$$

where α, β, γ and δ are the weights given to the content loss, feature loss, texture loss and least squares relativistic adversarial loss, respectively, which enable to satisfy aspects contributed by the combination of multiple loss functions simultaneously.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. TRAINING DETAILS

The environment is configured as Table 3. The DIV2K dataset is employed to train our SRLRGAN-SN model, which contains 800 images for training and 100 images for validation. Data augmentation for raising the diversity of training images with random 90°, 180°, 270° rotations and horizontal flips. The Adam [42] optimizer is employed with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We initialize the learning rate to 10^{-4} and halve over every 50000 iterations. The proposed model is optimized with the total loss in equations (11) and (12), where $\alpha = 10^{-2}, \beta = 1, \gamma = 1$ and $\delta = 10^{-3}$ are empirically set to be optimal respectively. The feature mapping from the

TABLE 3. Configuration of experiment environment.

Configuration	Version
Memory	64G
Graphics card	P4_16G
Operating system	Ubuntu 16.04
TensorFlow	tensorflow_gpu 1.8
Cuda&cuDNN	CUDA 9.0 & cuDNN 7.1
Python	Python 3.6

vgg-19 model before activation is gained from the *4-th* layer in front of the *5-th* pooling layer.

B. COMPARISON OF GAN PERFORMANCE

To access the GAN performance of our SRLRGAN-SN model, the Frechet Inception Distance(FID) proposed by Martin *et al.* [43] is used. The FID is a popular metric for evaluating the performance of GAN, and lower score is better correlating well with higher generative quality. In addition, smaller score of the FID indicates more stable training of the model [43]. The FID score is calculated based on a pre-trained Inception v3 model $\psi(.)$ [44], which is defined as follows:

$$\begin{aligned}&\mathrm{FID}(\psi(I^{HR}), \psi(G_{\theta_G}\left(I^{LR}\right)))\\&= \left\|\mu_{I^{HR}} - \mu_{G_{\theta_G}(I^{LR})}\right\|_2^2\\&\quad + \mathrm{Tr}\left(\Sigma_{I^{HR}} + \Sigma_{G_{\theta_G}(I^{LR})} - 2\left(\Sigma_{I^{HR}}\Sigma_{G_{\theta_G}(I^{LR})}\right)^{\frac{1}{2}}\right)\end{aligned} \tag{13}$$

where $G_{\theta_G}\left(I^{LR}\right)$ and I^{HR} are the produced image and the true original sample, respectively. μ refers to the mean feature obtain from image. Σ refers to the covariance matrix for the feature vector of image. *Tr* is the trace linear algebra operation to sums up all the diagonal elements. SRLRGAN-SN is compared with GAN-based super-resolution methods including SRGAN [22], ESRGAN [24], PESRGAN [25], NatSRGAN [26], and SRLRGAN-SN without spectral normalization (SRLRGAN). We use available source codes and open results of indicated methods above respectively to compare the performance in experiments. From Table 4, SRGAN indicates the lowest GAN performance in terms of FID, which the value is 6.83. SRLRGAN-SN shows the best GAN performance in terms of FID, which the value is 6.22. The value of our method is 0.61, 0.50, 0.41, 0.36 less than that of SRGAN, ESRGAN, PESRGAN and NatSRGAN separately. It shows that the least squares relativistic discriminator is effective at improving the model's performance. Besides, the result of SRLRGAN-SN method applied the spectral normalization is more better than that without spectral normalization. It demonstrates the effectness of least squares relativistic generative adversarial network with spectral normalization to enhance the performance of GAN for super-resolution task from the quantitative evaluation perspective.

TABLE 4. Comparison of GAN performance.

Method	SRGAN	ESRGAN	PESRGAN	NatSRGAN	SRLRGAN	SRLRGAN-SN
FID	6.83	6.72	6.63	6.58	6.50	6.22

C. COMPARISON OF OTHER POPULAR METHODS

To validate the performance of our method for super-resolution imaging, our SRLRGAN-SN method is compared with other state-of-the-art(SOAT) methods regarding to quantitative metric and perceptual quality. All results are achieved on Set5, Set14, BSD100 and Urban100 dataset [22], respectively. The SOAT methods include VDSR [16], EDSR [17], SRGAN [22], EnhanceNet [23], ESRGAN [24], PESRGAN [25], NatSRGAN [26]. VDSR and EDSR are typical PSNR-driven CNN-based super-resolution methods, which are aiming for high PSNR value rather than realistic visual quality. SRGAN, EnhanceNet, ESRGAN, PESRGAN and NatSRGAN are popular perception-driven GAN-based super-resolution methods in recent years. We use available source codes and open results of indicated methods above respectively to compare the performance in experiments.

For perceptual super-resolution imaging, good perceptual quality is crucial from a perspective of human vision. In the aspect of quantitative metric, we use PI in equations (1) couple with PSNR to evaluate the objective performance [45], although the PSNR is not as effective as the PI metric in terms of perceptual quality. Thus, PSNR is used to offer a relatively minor point of referential estimation for the loss of quality between the pixel values, on account of not indicating perceptual quality well. PI offers a primary point of reference for evaluating the perceptual reconstruction quality from the perspective of satisfying the requirements of the perceptual assessment. Quantitative results compared to SOAT methods regarding to PSNR and PI are shown in Table 8. Bold values signified the best performance are highlighted. As shown in Table 5, at a magnification factor of 4x, EDSR obtains the highest average value of PSNR, and our method rises to the top in terms of average PI. As expected, the PSNR values obtained by PSNR-driven methods are higher than these of GAN-based methods, but the PI indicators correlating with perceptual quality universally far lag behind these of GAN-based methods. The SRLRGAN-SN method wins the first place in SOAT methods regarding to average PI, where the mean value is 3.14, 2.79, 0.24, 0.36, 0.37, 0.21 and 0.58 lower than that of VDSR [16], EDSR [17], SRGAN [22], EnhanceNet [23], ESRGAN [24], PESRGAN [25] and NatSRGAN [26] separately. Compared with PSNR-driven methods, Our proposed method also achieves comparable performance in terms of PSNR. The graphs about epochs versus considered metrics in terms of average PSNR and PI are shown in Figure 2. To further compare the SRLRGAN-SN with more methods considering different magnification factor, comparison results for 2x upscaling are depicted in Table 6. Clearly from Table 5 and Table 6, it signifies that SRLRGAN-SN can effectively obtain competitive

TABLE 5. Comparison of SOAT methods in terms of PSNR/PI (4x upscaling).

Dataset	VDSR	EDSR	SRGAN	EnhanceNet	ESRGAN	PESRGAN	NatSRGAN	SRLRGAN-SN
Set5	31.35/6.45	**32.46**/6.00	29.41/3.18	28.56/**2.93**	30.32/3.32	28.79/3.42	30.98/3.59	30.57/3.02
Set14	28.02/5.77	**28.71**/5.52	26.02/2.80	25.67/3.02	26.41/2.93	25.92/2.66	27.42/3.11	27.18/**2.54**
BSD100	27.29/5.70	**27.72**/5.40	25.18/2.59	24.93/2.91	24.48/2.34	25.24/2.25	26.44/2.77	26.23/**2.11**
Urban 100	25.18/5.54	**26.64**/5.14	25.11/3.30	23.54/3.47	24.36/3.77	24.06/3.41	25.46/3.75	25.32/**3.23**
Average	27.96/5.87	**28.88**/5.52	26.43/2.97	25.68/3.08	26.39/3.09	26.00/2.94	27.58/3.31	27.33/**2.73**

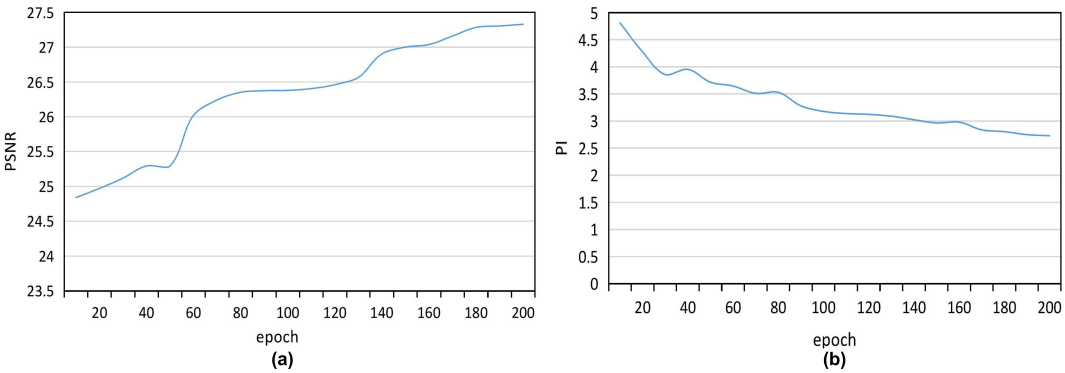

FIGURE 2. The graphs about epochs versus considered metrics for 4x upscaling. (a)average PSNR; (b)average PI.

TABLE 6. Comparison of SOAT methods in terms of PSNR/PI (2x upscaling).

Dataset	VDSR	EDSR	SRGAN	EnhanceNet	ESRGAN	PESRGAN	NatSRGAN	SRLRGAN-SN
Set5	37.53/6.11	**38.11**/5.83	36.65/3.05	36.21/2.90	35.11/3.15	33.81/3.28	37.94/3.46	37.29/2.94
Set14	33.03/5.57	**33.92**/5.29	31.14/2.62	31.07/2.87	31.82/2.85	30.65/2.57	31.26/3.07	31.02/**2.36**
BSD100	31.90/5.53	**32.32**/5.36	30.90/2.36	29.05/2.74	28.82/2.21	28.54/2.13	32.15/2.55	32.06/**2.06**
Urban 100	30.76/5.19	**32.93**/5.03	30.81/3.18	28.68/3.25	29.76/3.54	29.12/3.21	30.69/3.43	30.54/**3.12**
Average	33.31/5.60	**34.32**/5.38	32.38/2.80	31.25/2.94	31.38/2.94	30.53/2.80	33.01/3.13	32.73 /**2.62**

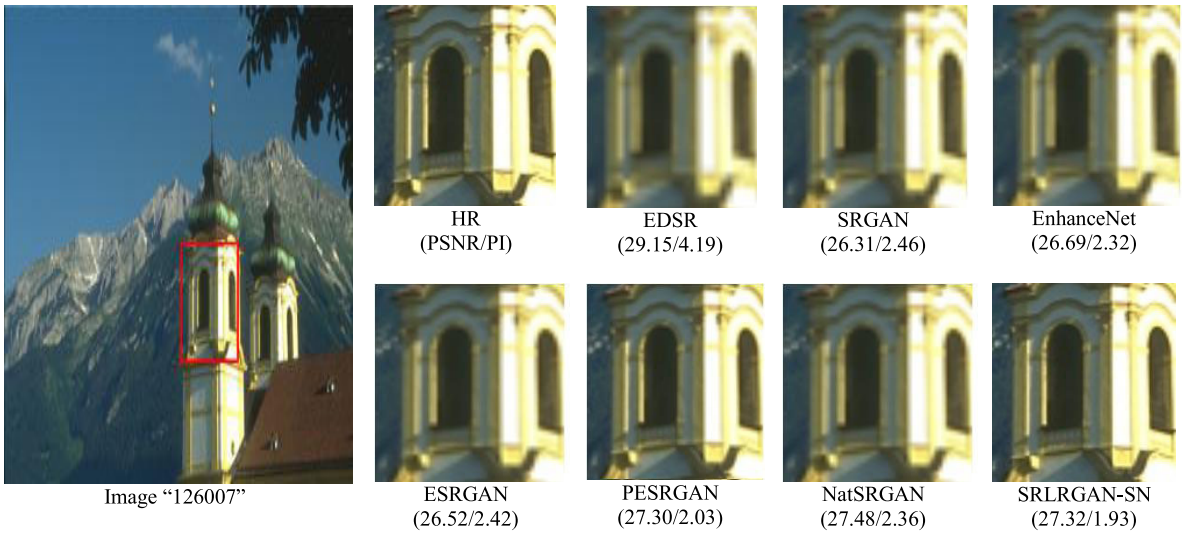

FIGURE 3. Comparison of perceptual quality with SOAT super-resolution methods for image "126007" from BSD100 dataset (4x upscaling).

advantages in terms of quantitative metric compared with other SOAT methods.

In the aspect of perceptual quality, the comparison of visual perception is used to rank the perceptual performance. The

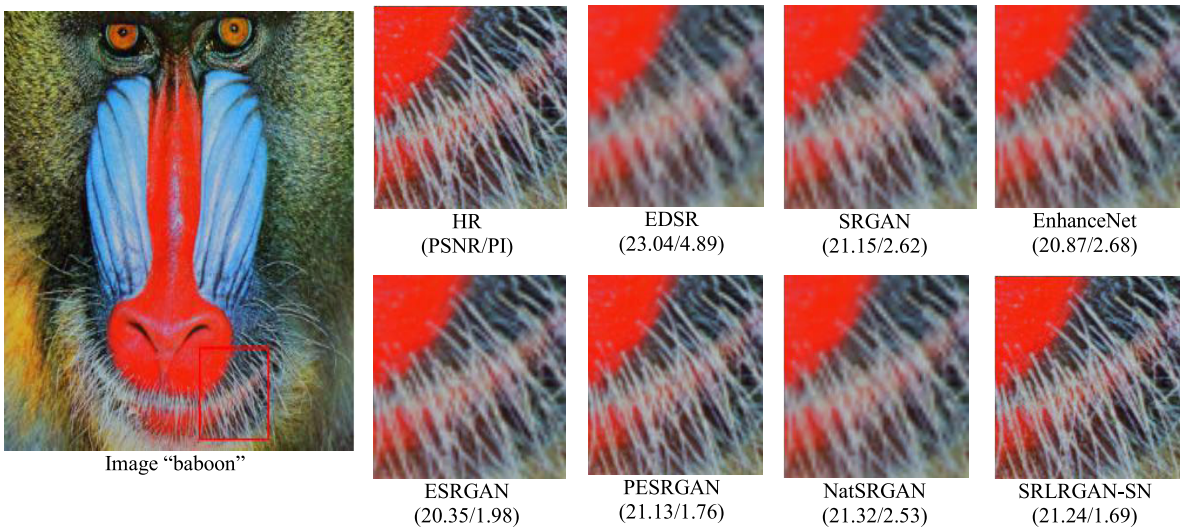

FIGURE 4. Comparison of perceptual quality with SOAT super-resolution methods for image "baboon" from Set14 dataset (4x upscaling).

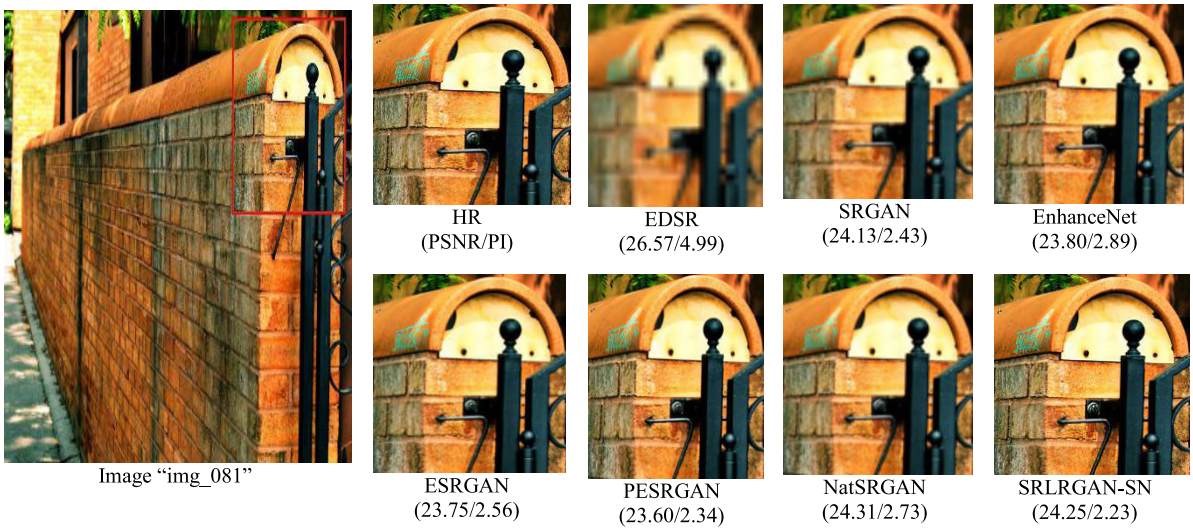

FIGURE 5. Comparison of perceptual quality with SOAT super-resolution methods for Image "img_081" from Urban dataset (4x upscaling).

comparison results of perceptual quality for 4x upscaling are illustrated in the Figure 3 to Figure 5. Although EDSR trained by pixel loss achieves the highest value of PSNR, it obviously compares unfavourably with other GAN-based methods on the overall reconstructed perceptual vision. It supports the point that PSNR-driven methods are always prone to reconstruct blurry and smooth images. Compared with other SOAT methods in visual verisimilitude, we can also visually observe that our SRLRGAN-SN method not only reconstructs textures sharper and more natural, but also looks less distinguishable from the true high-resolution image. For image ''126007'' from BSD100 dataset, it can be seen that most of SOAT methods suffer from blurry artifacts and cannot do well to sharpen up the feature-rich structures and regions between the building. By comparison, our SRLRGAN-SN method seems impressive sharpening around detailed structures of the building and provide natural and realistic textures. For image ''baboon'' from Set14 dataset, it can be seen that most of SOAT methods fail to preserve fine details of whiskers and background textures. By comparison, the visual effect of our SRLRGAN-SN method synthesising plausible textures is very noticeable. For image ''img_081'' from Urban dataset, it also can be seen that SRLRGAN-SN method generates visually plausible image with more perceptual texture details. Impressive results validate that our proposed method indeed achieve the superior outperformance correlating well with visual perception.

D. ABLATION ANALYSIS

To prove the effectiveness of each module of SRLRGAN-SN, the ablation study is performed to compare their impact on contributions for super-resolution imaging. Results of ablation regarding to PI on BSD100 set are depicted in Table 7. Bold values signified the best performance are highlighted.

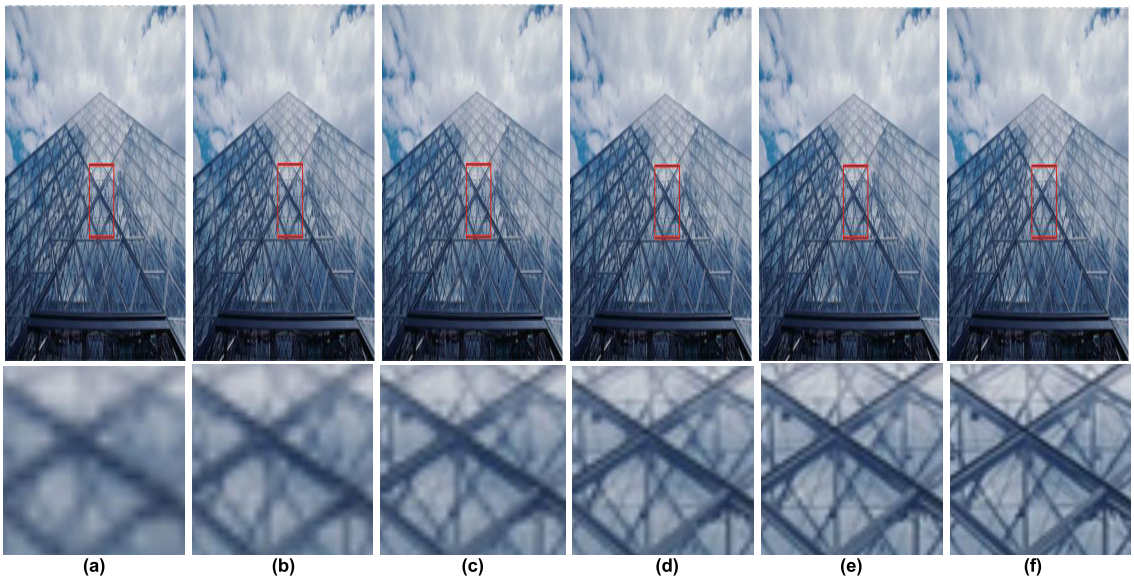

FIGURE 6. Comparison of perceptual texture details of different modular settings for image "223061" from BSD100 dataset. The subgroup (a) to (e) are corresponding to the settings (a) to (e) indicated in Table 7, respectively.

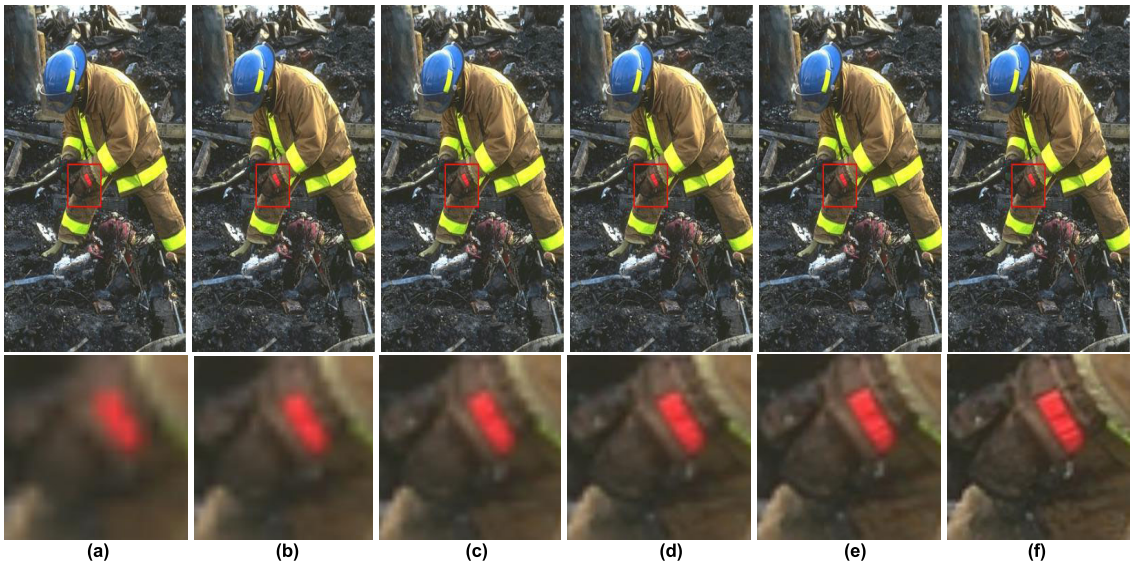

FIGURE 7. Comparison of perceptual texture details of different modular settings for image "285079" from BSD100 dataset. The subgroup (a) to (e) are corresponding to the settings (a) to (e) indicated in Table 7, respectively.

The content loss L_{con}, feature loss L_{fea}, texture loss L_{tex}, proposed adversarial loss L^{ADV}, and spectral normalization operation SN are conducted ablation analysis through different modular settings in Table 7. To validate the performance of our proposed adversarial loss derived from RaGAN, the adversarial loss L^{RaGAN} based on RaGAN is added by contrast. The model only trained by L_{con} in setting (a) gains the worst performance. After the L_{fea} is added in setting (b), the loss combination provide significant improvements to 3.29 in terms of perceptual index. When the L_{tex} is added in setting (c), the perceptual quality is further enhanced to 3.20. Higher perceptual quality can be gained in setting (e) by adding proposed L^{ADV} in contrast with adding L^{RaGAN} in setting (d). Besides, the performance of setting (f) applied the spectral normalization on the basis of setting (e) achieves the best result.

Figure 6 and Figure 7 show the comparison of perceptual texture details of different modular settings on BSD100 dataset. The subgroup (a) to (e) are corresponding to settings (a) to (e) indicated in Table 7, respectively. In each column, the original high-resolution sample is displayed on the top, and the corresponding super-resolution image is displayed on the below. As anticipated, the model only trained by setting (a) shows the fuzziest vision effect with poor perceptual quality from a perspective of appealing to a vision viewer. After the feature loss is added in setting (b), the vision effect has some improvement, but the texture details are not obvious. When the texture loss is added in setting (c), the

TABLE 7. Results of ablation analysis in terms of PI on BSD100 dataset.

Setting	L_{con}	L_{fea}	L_{tex}	L^{RaGAN}	L^{ADV}	SN	PI
(a)	√	×	×	×	×	×	5.37
(b)	√	√	×	×	×	×	3.29
(c)	√	√	√	×	×	×	3.20
(d)	√	√	√	√	×	×	2.45
(e)	√	√	√	×	√	×	2.23
(f)default	√	√	√	×	√	√	**2.11**

TABLE 8. Comparison of model complexity.

Model	VDSR	EDSR	SRGAN	EnhanceNet	ESRGAN	PESRGAN	NatSR	Ours
FLOP (G)	174.25	823.32	113.20	121.01	1034.10	3056.32	328.64	420.46
Param (M)	0.67	43.09	1.51	0.85	16.70	43.09	4.90	6.50

texture details have some improvement, but the perceptual quality is not pleasant. When the RaGAN adversarial loss is added in setting (d), the perceptual quality is improved obviously from the vision perspective. Furthermore, When the RaGAN adversarial loss is replaced by our proposed least squares relativistic adversarial loss in setting (e), it is more outstanding in perceptual quality than the setting (d). The setting (f) employed in this study with spectral normalization achieve optimal performance in terms of PI and yields the most visually realistic details. To sum up, it shows that all these modules indicated in Table 7 have contributed to increase the capability for perceptual super-resolution imaging. The least squares relativistic generative adversarial network with spectral normalization to be effective at improving the model's performance at generating high-quality perceptual images from the vision perspective.

E. COMPLEXITY ANALYSIS

To prove the computational ability of SRLRGAN-SN, a study about the number of FLOP and the parameters is conducted. It is compared with the VDSR [16], EDSR [17], SRGAN [22], EnhanceNet [23], ESRGAN [24], PESRGAN [25], NatSRGAN [26]. Results of complexity analysis are depicted in Table 8. Obviously from Table 8 that SRLRGAN-SN can achieve competitive performance, the number of FLOP and the parameters are far less than models that own excellent performance in terms of PSNR or PI, such as EDSR, ESRGAN and PESRGAN. While the number of FLOP and the parameters of SRLRGAN-SN are not the lowest, it gets a balance between perceptual quality and model complexity successfully. It verifies that SRLRGAN-SN is efficient for super-resolution imaging with lightweight of complexity.

V. CONCLUSION

A novel method named SRLRGAN-SN for super-resolution imaging is proposed combined with spectral normalization and least squares relativistic discriminator. On the one hand, the spectral normalization is applied on the network to enhance the performance of GAN for super-resolution task. On the other hand, the least squares relativistic discriminator is used to drive generated images towards more perceptual manifold. We also adopt a new combination of multiple loss functions to produce visually plausible images with realistic texture details as much as possible. Results of experiment signify that SRLRGAN-SN can recover more fine texture details and acquire better performance compared with other SOAT for super-resolution imaging. In the future work, we will explore more optimization functions and adopt more network structures such as dense residual aggregation network to further increase the capability of GAN for super-resolution imaging.

REFERENCES

[1] D. Liu, Z. Wang, B. Wen, J. Yang, W. Han, and T. S. Huang, "Robust single image super-resolution via deep networks with sparse prior," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3194–3207, Jul. 2016.

[2] J.-T. Hsu, C.-H. Kuo, and D.-W. Chen, "Image super-resolution using capsule neural networks," *IEEE Access*, vol. 8, pp. 9751–9759, 2020.

[3] V. K. Ha, J. Ren, X. Xu, S. Zhao, G. Xie, and V. M. Vargas, "Deep learning based single image super-resolution: A survey," *Int. J. Automat. Comput.*, vol. 16, no. 4, pp. 413–426, 2019.

[4] J. Dai-Hong, D. Lei, L. Dan, and Z. San-You, "Moving-object tracking algorithm based on PCA-SIFT and optimization for underground coal mines," *IEEE Access*, vol. 7, pp. 35556–35563, 2019.

[5] L. Zhang, P. Wang, C. Shen, L. Liu, W. Wei, Y. Zhang, and A. Van Den Hengel, "Adaptive importance learning for improving lightweight image super-resolution network," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 479–499, Feb. 2020.

[6] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3867–3876.

[7] M. Zamorski, A. Zdobylak, M. Zięba, and J. Świątek, "Generative Adversarial Networks: Recent developments," in *Proc. Int. Conf. Artif. Intell. Soft Comput.* Cham, Switzerland: Springer, 2019, pp. 248–258.

[8] H. Alqahtani, M. Kavakli-Thorne, and G. Kumar, "Applications of generative adversarial networks (GANs): An updated review," *Arch. Comput. Methods Eng.*, vol. 3, pp. 1–28, Dec. 2019.

[9] A. Fatir Ansari, J. Scarlett, and H. Soh, "A characteristic function approach to deep implicit generative modeling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7478–7487.

[10] C. Wang, C. Xu, X. Yao, and D. Tao, "Evolutionary generative adversarial networks," *IEEE Trans. Evol. Comput.*, vol. 23, no. 6, pp. 921–934, Jan. 2019.

[11] Q. Kou, D. Cheng, H. Zhuang, and R. Gao, "Cross-complementary local binary pattern for robust texture classification," *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 129–133, Jan. 2019.

[12] F. Zhou, W. Yang, and Q. Liao, "Interpolation-based image super-resolution using multisurface fitting," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3312–3318, Jul. 2012.

[13] C. Kim, K. Choi, H.-Y. Lee, K. Hwang, and J. B. Ra, "Robust learning-based super-resolution," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 2017–2020.

[14] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4799–4807.

[15] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 184–199.
[16] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.
[17] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.
[18] J. M. Haut, R. Fernandez-Beltran, M. E. Paoletti, J. Plaza, and A. Plaza, "Remote sensing image superresolution using deep residual channel attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9277–9289, Nov. 2019.
[19] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3147–3155.
[20] N. Ahn, B. Kang, and K. A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 252–268.
[21] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 334–355.
[22] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.
[23] M. S. M. Sajjadi, B. Scholkopf, and M. Hirsch, "EnhanceNet: Single image super-resolution through automated texture synthesis," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4491–4500.
[24] N. C. Rakotonirina and A. Rasoanaivo, "ESRGAN+ : Further improving enhanced super-resolution generative adversarial network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1–17.
[25] T. Vu, T. M. Luu, and C. D. Yoo, "Perception-enhanced image super-resolution via relativistic generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 98–113.
[26] J. W. Soh, G. Y. Park, J. Jo, and N. I. Cho, "Natural and realistic single image super-resolution with explicit natural manifold discrimination," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8122–8131.
[27] J. M. Haut, R. Fernandez-Beltran, M. E. Paoletti, J. Plaza, A. Plaza, and F. Pla, "A new deep generative network for unsupervised remote sensing single-image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6792–6810, Nov. 2018.
[28] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.
[29] I. Gulrajani, F. Ahmed, and M. Arjovsky, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.
[30] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," 2018, *arXiv:1802.05957*. [Online]. Available: http://arxiv.org/abs/1802.05957
[31] A. Odena, J. Buckman, C. Olsson, T. B. Brown, C. Olah, C. Raffel, and I. Goodfellow, "Is generator conditioning causally related to GAN performance?" 2018, *arXiv:1802.08768*. [Online]. Available: http://arxiv.org/abs/1802.08768
[32] H. Zhang, I. Goodfellow, D. Metaxas, And A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7354–7363.
[33] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6228–6237.
[34] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, "Learning a no-reference quality metric for single-image super-resolution," *Comput. Vis. Image Understand.*, vol. 158, pp. 1–16, May 2017.
[35] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'Completely Blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[37] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 624–632.
[38] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 694–711.
[39] L. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.* 2015, pp. 262–270.
[40] A. Jolicoeur-Martineau, "The relativistic discriminator: A key element missing from standard GAN," 2018, *arXiv:1807.00734*. [Online]. Available: http://arxiv.org/abs/1807.00734
[41] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2794–2802.
[42] I. K. M. Jais, A. R. Ismail, and S. Q. Nisa, "Adam optimization algorithm for wide and deep neural network," *Knowl. Eng. Data Sci.*, vol. 2, no. 1, pp. 41–46, 2019.
[43] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6626–6637.
[44] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
[45] Q. Kou, D. Cheng, L. Chen, and K. Zhao, "A multiresolution gray-scale and rotation invariant descriptor for texture classification," *IEEE Access*, vol. 6, pp. 30691–30701, 2018.

SANYOU ZHANG was born in 1989. He received the master's degree in computer application technology from the China University of Mining and Technology, where he is currently pursuing the Ph.D. degree. He graduated from Soochow University, in 2013. He has worked with the Suzhou Wujiang District Public Security Bureau. His main research interest includes computer vision.

DEQIANG CHENG was born in Henan, China, in 1979. He is currently a Professor and the Ph.D. Supervisor with the School of Information and Control Engineering, China University of Mining and Technology. His research interests include machine learning, video coding, image processing, and pattern recognition.

DAIHONG JIANG was born in Hunan, China, in 1969. She received the Ph.D. degree in communication and information systems. She is currently a Professor of communication and information systems. She graduated from the China University of Mining and Technology, in 2015. She has worked with the Xuzhou University of Technology. Her main research interests include intelligent computation and database technology.

QIQI KOU received the Ph.D. degree from the China University of Mining and Technology, in 2019. He is currently a Lecturer with the School of Computer Science and Technology, China University of Mining and Technology. His research interests include image processing and pattern recognition.