SRDN: A Unified Super-Resolution and Motion Deblurring Network for Space Image Restoration

Xi Yang, *Member, IEEE*, Xiaoqi Wang, Nannan Wang, *Member, IEEE*, and Xinbo Gao, *Senior Member, IEEE*

***Abstract*— Space target super-resolution (SR) is a domain-specific single image SR problem aiming to help distinguish the satellite and spacecrafts from numerous space debris. Compared to the other object SR problem, images for space target are always in low quality with varies of degradation condition, as a result of long distance and motion blur, which significantly reduces the manual classification reliability, especially for these small targets, e.g., satellite payloads. To address this challenge, we present an end-to-end SR and deblurring network (SRDN). Concretely, focusing on the low-resolution (LR) space target images with blind motion blur, we integrate the SR and deblur function together, improving the image quality by a unified generative adversarial network (GAN)-based framework. We implement a deblur module by using contrastive learning to extract degradation feature and add symmetrical downsampling and upsampling modules to the SR network in order to restore texture information, while shortcut connections are redesigned to maintain the global similarity. Extensive experiments on the public satellite dataset, BUAA-SID-share1.5, demonstrate that our network outperforms the state-of-the-art SR and deblur methods.**

***Index Terms*— Artificial intelligence, artificial neural networks, high-resolution (HR) imaging, image denoising.**

I. Introduction

As THE aerospace industry develops, the increasing number of on-orbit spacecrafts and debris leads the near-Earth space to a minefield of potential accidents. In this situation, tracing and distinguishing the debris and spacecrafts from space target images start to play an important role in guarding space security. However, for either traditional manual interpretation method or the automatic deep-learning-based method, the limited image quality will lead to inevitable recognition errors. One instinct solution is to restore space target images with better quality, which is exactly the core concept of super-resolution (SR).

SR aims to reconstruct the high-resolution (HR) image from the corresponding low-resolution (LR) image. Considered as an important research area of computer vision, SR is applicable for surveillance devices, space tracking systems, medical images, and so on. Traditional SR methods for space target recognition include the sparse coding method [1], [2], the anchored neighborhood regression method [3], the self-exemplars method [4], the Bayes method [5], and the deep convolutional neural network-based method [6], [7]. However, these traditional learning-based SR methods tend to generate indistinguishable images once the upscale factor is greater than or equal to 4, whereas, in practical satellite monitor tasks, the resolution of original images is ultralow, which requires the SR methods to upscale the images greatly.

Since the deep-learning-based SR method, SRCNN [8], is proposed, SR methods have been used extensively in the image processing area. After that, numerous PSNR-oriented SR networks and their corresponding training strategies are proposed [9]–[11], achieving a remarkable improvement on either output image quality or image processing speed. Unfortunately, the PSNR-oriented methods will reconstruct an oversmooth HR image, whereas the original one lacks high-frequency details. Similarly, since the objective index, PSNR, is theoretically different from our subjective feelings, the output HR image cannot fully satisfy the practical demand.

Since perceptual loss [12] is proposed, the perceptual-driven method becomes increasingly popular. As the perceptual loss is calculated at feature extraction layers, these methods can collect more semantic information of the given images, reconstruct vivid texture details, and, finally, generate more natural HR images. Based on SRGAN [12], ESRGAN [13] improves its perceptual loss and structure of generator and discriminator, dramatically increases the visual perception, and sustains high PSNR result via L1 loss.

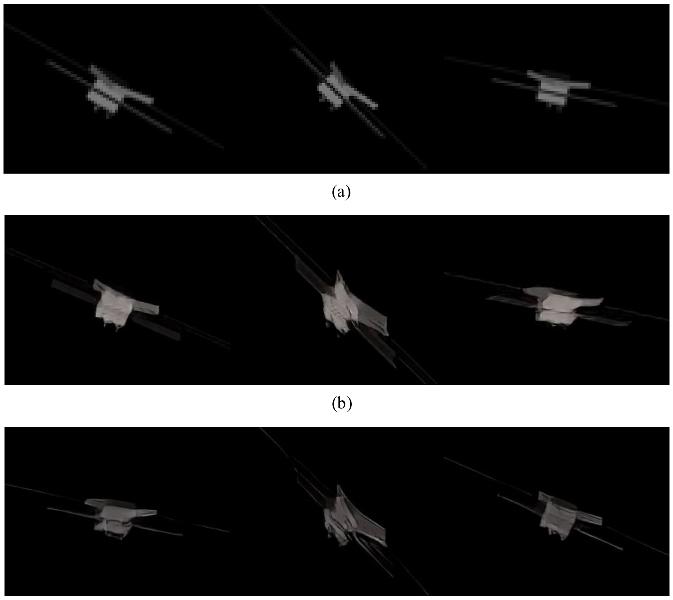
Generally, the image quality of space target is relatively poorer due to many factors, e.g., the defocus, motion blur, and the vast distance between the camera and the target, which also means object detection and component recognition on the images of space target becomes difficult. Thus, the most instinct solution is to use the SR method to clarify the LR images. However, as shown in Fig. 1, with motion blur, the space target images can hardly be recognized after SR

Manuscript received September 29, 2021; revised November 1, 2021; accepted November 24, 2021. Date of publication November 29, 2021; date of current version February 21, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61976166, Grant 62036007, Grant 62176195, Grant 61922066, and Grant 61876142; in part by the Key Research and Development Program of Shaanxi under Grant 2021GY-030; in part by the Innovation Capacity Support Plan of Shaanxi Province under Grant 2020KJXX-027; in part by the Fundamental Research Funds for the Central Universities under Grant JB210115; and in part by Open Research Projects of Zhejiang Laboratory under Grant 2021KG0AB01. *(Corresponding author: Xinbo Gao.)*

Xi Yang, Xiaoqi Wang, and Nannan Wang are with the State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xi’an 710071, China (e-mail: yangx@xidian.edu.cn; xqwang14@stu.xidian.edu.cn; nnwang@xidian.edu.cn).

Xinbo Gao is with the State Key Laboratory of Integrated Services Networks, School of Electronic Engineering, Xidian University, Xi’an 710071, China, and also with the Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: xbgao@mail.xidian.edu.cn).

Digital Object Identifier 10.1109/TGRS.2021.3131264



(a)

(b)

(c)

Fig. 1. Simple example of the output images reconstructed by (a) original (without motion blur), (b) SR(ESRGAN), and (c) SRDN.

preprocessing. Thus, the preprocessing procedures, such as motion deblurring, are always necessary. As far as we know, there is currently no deep-learning-based method to solve motion blur and SR problems simultaneously on space targets, which brings the LR image clearness problem of fast-moving space targets more complex.

This article is exactly established on this real demand. In this article, we propose an end-to-end SR and deblurring network (SRDN). Our contribution can be summarized as follows.

We redesigned the SR network to better extract the global feature; the concrete solutions include the following.

1) Add the symmetrical shrink and expand layers at both ends of residual in residual dense blocks (RRDB [13]) to reduce feature sizes, permitting a larger batchsize and ensuring the temporal efficiency.
2) Based on our first step, compared to the other SR methods [14], [15], we further enlarge the range of shortcut connections in the network. Low-frequency information in the LR and HR images is highly similar, which means that the network only needs to learn the high-frequency residual between HR and LR images; thus, enlarging the range of shortcut connections helps reconstruct the HR image from the global scale.
3) We change the number of basic blocks to better fit the SR and deblurring task on space target images.

Meanwhile, we design an individual degradation feature learning module, using contrastive learning to solve motion blur. Concretely, by maximizing the mutual information between patches cropped from the same image, we can extract the degradation feature from the blurred image. Focusing on the space target images, we propose a special data augmentation method to minimize the mutual information in the first place.

Extensive experiments show that our network can reconstruct HR images with higher PSNR and SSIM results than the state-of-the-art SR methods from LR images of space targets with motion blur.

In this way, we can upscale and deblur an image in a single network and by using a unify loss function to reduce the mutual interference between two tasks. Also, our method shows a better result on either objective indexes (PSNR and SSIM) or the subjective index [perceptual index (PI)], i.e., **0.63/0.0152/0.49** (PSNR/SSIM/PI) ahead of state of the art.

II. Related Works

A. Space Target Recognition

Space target recognition aims to identify and distinguish spacecrafts from various meteoroids and debris. As the neural network technique develops rapidly, an increasing number of neural network-based space target recognition methods are applied. Ma *et al.* [16] proposed a space target recognition algorithm based on the 2-D wavelet transform. In this network, the space target image was decomposed based on a 2-D wavelet transform; then, the singular value feature was extracted from approximate and detailed parts. Finally, the radial base function (RBF) neural network was used to classify space targets. Meanwhile, Zhang and Zhou [17] used the backpropagation fuzzy neural network (BPFNN) classifier to recognize space targets.

Since Hinton *et al.* [18] proposed the Alexnet that first introduces the DCNN structure to images classification, more recognition works start to leverage the DCNN backbone [19], [20]. The first DCNN-based space target recognition is proposed by Zeng and Xia [21], achieving impressive accuracy on the public datasets. However, the DCNN-based methods always require a large dataset to train the network, and thus, the D2N4 [22] proposed a few-shot learning method to minimize the accuracy loss caused by the lack of dataset. To make the space target recognition more automatic, composing of locating and recognition function, T-SCNN [23] is able to classify all the targets from a larger space image. Unfortunately, most of the aforementioned methods require clear and HR images in a public dataset, and the images of LR will largely limit their performance, leading to error identification. Therefore, SR methods and motion deblur are necessary for space target recognition.

B. Super-Resolution

The deep-learning-based algorithms start from SRCNN [8]. The SRCNN includes three processing procedures: feature extracting layer, nonlinear mapping layer, and deconvolutional layer, achieving an end-to-end SR network that can generate HR images directly from LR images. After that, Shi *et al.* [24] proposed ESPCN, which first introduces the subpixel convolutional layer to achieve the mapping procedure from LR image to HR image. Kim *et al.* [25] reckon that the thought of residual network is suitable for solving SR problem, so the proposed VDSR network only learns the residual between the original HR images and the upscaled LR images. Lim *et al.* [26] proposed EDSR that removes the batch-normalization (BN) layer from the SRResNet, saving the graphic memory and permitting to construct a larger and more complex network to get better performance. These thoughts deeply influence the afterward deep-learning-based SR algorithms.

As proposed by Goodfellow *et al.* [57], the generative adversarial network (GAN) is gradually applied to the area of image recovery. Ledig *et al.* [12] proposed SRGAN is based on GAN, using perceptual loss and adversarial loss to recover the feeling of reality of images. Wang *et al.* [13] proposed ESRGAN, which introduces the RRDB module to improve the structure of the model. Furthermore, they use the adversarial loss in relative average GAN (RaGAN) to supervise the discriminator and promote the perceptual loss by combining the image restoration loss and GAN loss of the generator [27].

In recent years, focusing on the different practical scenarios, many improvements of SR methods are directly based on the real demand of different applications. CSNLN [28] not only considers both attention on global feature but also features in different scales, contributing an efficient and effective network; deep generative prior [29] explores prior information in images, creating a comprehensive network, which could handle multiple image restoration tasks.

Focusing on the oversmooth problem of PSNR-oriented SR, EEGAN [30] proposes an edge-enhance method by using the Laplace operator. To deal with the different degradation situations in real-world scenarios, BSRGAN [31] collects all the degradation models from different images and substitutes the degradation model on images randomly, thus augmenting degradation data. Similarly, DRN [32] proposes a dual regression modal, achieving a closed loop in the SR problem, which means that it not only learns the common projection from the LR image to the HR image but also learns a degradation regression from HR to LR images, which helps to solve the real-world SR problem.

Meanwhile, the traditional SR methods are also under great development, e.g., GFN [33] introduces the recursive gate block and HDRN [34] introduces a hierarchical dense recursive network for image restoration. RFANet [35] introduces a residual feature aggregation framework and enhanced spatial attention block to the SR network.

C. Motion Deblur

The working principle of the camera is to record the scene information by a light-sensitive surface during the exposure time. In the case of motion, the object will move during the exposure; thus, the blurs will appear on the edge of moving objects. As for the space target images captured by satellites, the relative speed would be extremely high; thus, the motion blur is a crucial problem for space target recognition.

The family of deblurring problems is divided into two types: blind and nonblind deblurrings. Early work mostly focused on nonblind deblurring. Under an assumption that the blur kernels $k(M)$ are known, the motion blur could be added to the original images by using linear convolution and vice versa. However, this inverse filtering method is sensitive to noise. In the case when blur kernel $K(M)$ is very small, the direct deconvolution of the blurred image and the kernel actually amplified the noise. Therefore, the nonblind deblurring mostly relies on those less sensitive algorithms, e.g., classical Lucy–Richardson algorithm [36] and the Wiener [37] or Tikhonov filter [38] to perform the deconvolution operation and obtain I_S estimate. In practical scenarios, commonly, the blur kernel is unknown, and blind deblurring algorithms estimate both the latent image I_S and blur kernels $k(M)$.

As the research on computer vision develops, the convolutional neural network (CNN) is considered the most promising method. Those families of methods address the blur caused by camera shake by considering blur to be uniform across the image. First, the camera motion is estimated in terms of the induced blur kernel, and then, the effect is reversed by performing a deconvolution operation. Starting with the success of Zhou *et al.* [39], several methods have developed over the last several decades. Many of the current deblur methods are based on the estimation of the blur kernels [40]–[42]. However, the running time, as well as the stopping criterion, is a significant problem for those kinds of algorithms.

Similarly, GAN shows extraordinary performance in deblur area. Kupyn *et al.* [43] proposed DeblurGAN and its improved version, DeblurGAN-v2 [44], to apply the GAN to motion deblur problem, achieving a deep-learning-based end-to-end deblur network, and it is five times (5×) faster than the fastest competitor.

Our main purpose is to improve the quality of space target image with motion blur. Thus, we create a unified space target image restoration network, jointly solve SR and deblurring problems.

III. Proposed Algorithm

In this session, we first introduce our network structure, then discuss our improvements, and introduce the loss function of our network.

A. Network Structure

As shown in Fig. 2, the framework of the network can be roughly divided into two parts: the degradation learning module and the SR module; the input images are first put into the degradation module to extract degradation features in order to reduce the influence of motion blur, and then, the processed feature is sent into SR module to create HR images.

Fig. 3 shows the structure of our generator. It is an end-to-end network module containing three parts: the preprocessing unit, the feature extracting unit, and the image upscaling unit. By inserting the image into the generator, we can obtain the HR image at the output layer.

As the RRDB in ESRGAN [13] shows extraordinary performance in the image restoration area, we decide to adopt it as the basic block of our network. It combines multiple dense connected networks to achieve strong learning ability and capability to learn the mapping function between the LR and HR images. Meanwhile, though the BN layer can normalize the data according to the average value and its variance to increase the network performance, for SR tasks

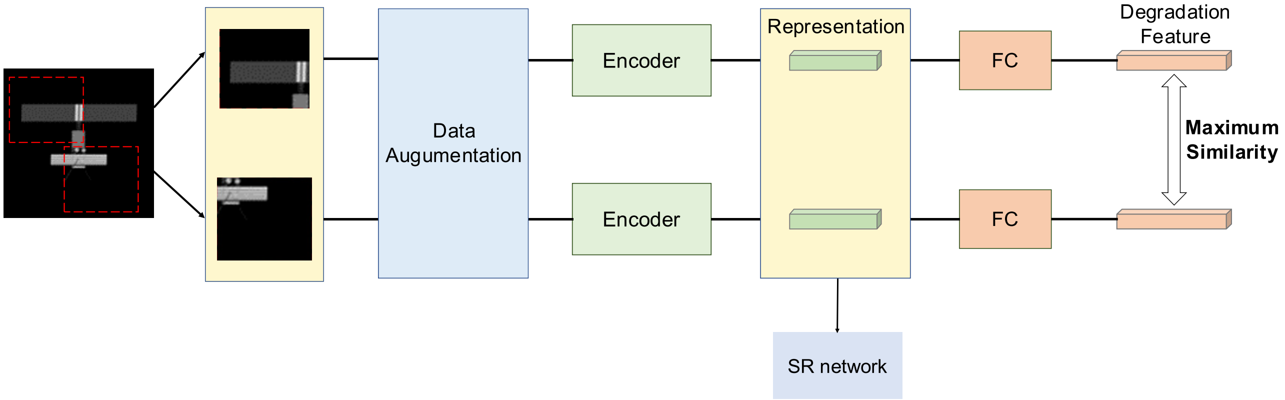


Fig. 2. Framework of our degradation extraction module, where encoder and FC both consist of fully connected layers. As the mutual information between two image patches is exactly the degradation feature, maximizing their similarity can decouple the degradation feature from the image.

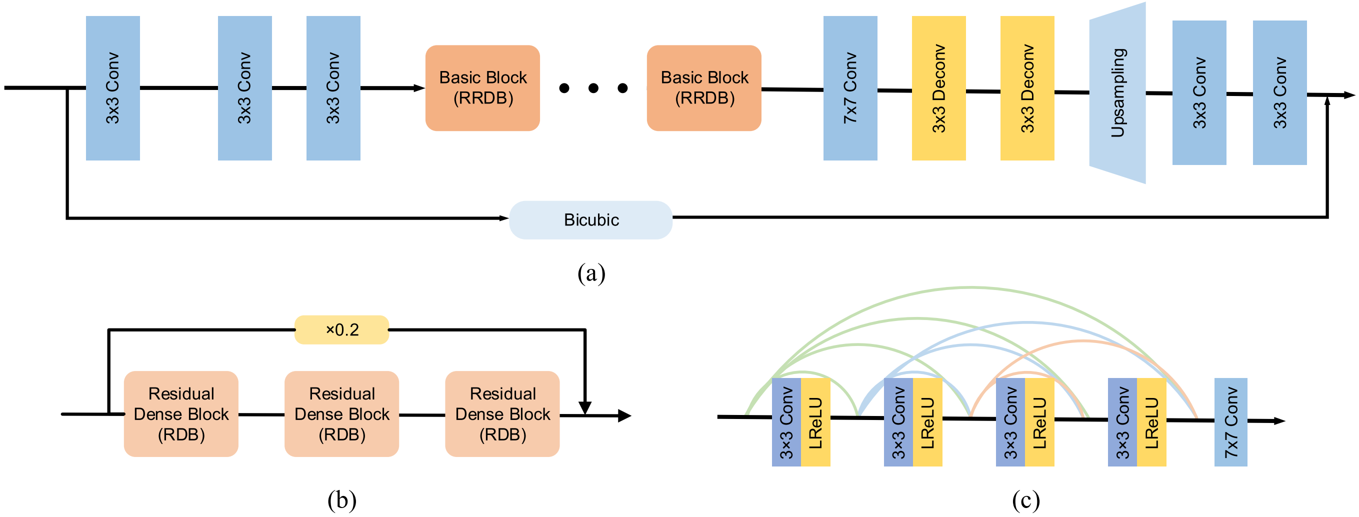


Fig. 3. Generator architecture of our network. (a) Generator of our SRDN. (b) Structure of the RRDB module. (c) Structure of the RDB module.

having highly different training sets and testing sets, the BN layer will somehow limit the performance of our network and even add artifacts onto the output image. Taking this into consideration, we remove the BN layer in our SRDN.

Based on RaGAN, our discriminator will predict the possibility of how the original HR image contains more information than our output HR image. Compared with the standard GAN, our network is more suitable to deal with the gradient between generated and real-world images, helping the generator learn a more detailed edge and texture. To stabilize this deep network, we use the scaled residual in shortcut connection, i.e., the residual will multiply a constant before it is added onto the main path.

B. Network Improvements

At the very beginning of the experiment, we use ESRGAN to process the LR images. After 30 000 iterations' training, ESRGAN could show the best PSNR and SSIM result on BUAA-SID 1.5 datasets [45], [46], whereas, in the experiments on the real-world images of space target, ESRGAN will approach its best performance when training after 270 000 iterations, of which the time cost is unacceptable. Thus, from a temporal perspective, we add two symmetrical convolutional and deconvolutional layers at both ends of the RRDB modules. In this way, the input feature size shrinks from $1 \times 128 \times 128$ to $1 \times 32 \times 32$, therefore reducing the time complexity. Meanwhile, the smaller feature image brings less redundancy than the larger one.

Second, we enlarge the range of shortcut connections. Couples of works [15], [47] illustrate that the shortcut connection that directly connects the input and output layers could be more effective than the smaller one on local features, i.e., as the input and output images in SR tasks are highly related, so using the shortcut connection to connect the input layer to the output layer could help the network focus on the difference, especially the high-frequency information. In this way, we not only reduce the learning cost but also improve the detail and texture restoration, which is essential for space targets. Also, to address the SR and motion deblurring problem simultaneously, we add an additional module for deblurring.

The common formulation of the nonuniform blur model is given in the following:

$$I_B = k(M) * I_S + N \tag{1}$$

where I_B is a blurry image and $k(M)$ is an unknown blur determined by motion field M. I_S is the sharp latent image, N is an additive noise, and $*$ denotes the convolution operation. In our hypothesis, we ignore the additional noise but only consider the motion blur. Therefore, there is actually a linear correlation between a clean image and the corresponding blurred one. Also, following the previous deblur work [48], [49], we assume that the degradation model on one image is the same and varies on different images. Concretely, we use six linear convolutional layers to extract the degradation feature because of the linear correlation between the original image and the motion blur; in addition, we remove the activation functions between these convolutional layers.

C. Loss Function

The loss function could be divided into two parts: contrastive loss for deblur task and the consistency loss for the SR task. In the training phase of one batch, we first use the contrastive loss to maximize the mutual information between patches from the same image and minimize it between patches from different images in this batch. After the degradation offset, we use the consistency loss to maintain the consistency between grand truth and the LR images.

1) Contrastive Loss: In a practical scenario, due to numerous interfering factors, the motion blur for each image is always different; thus, our degradation feature learning module assumes that the degradation model for one image is the same and varies when images are different. Under this assumption, if we crop two disconnected patches from one image, their only mutual information is the degradation model. The contrastive loss can be formulated as

$$L_x = -\log \frac{\exp\left(x \cdot x^{+}/\tau\right)}{\sum_{n=1}^{N} \exp\left(x \cdot x_n^{-}/\tau\right)}. \tag{2}$$

The x^+ denotes the positive pairs, while x_n^- denotes other negative pairs from the same batch, and τ is a temperature hyperparameter. Following the assumption above, we regard the patches from the same image as the positive pairs and those from different images as the negative pairs. Thus, by maximizing the mutual information from two disconnected patches from the same image, we can train a degradation learning module that can extract degradation features from blurred images. Then, before the SR process, we subtract the image by its degradation model; thus, theoretically, we obtain the image without the degradation, i.e., the motion blur. After that, we use the consistency loss to finish the SR task.

As for the consistency loss, to increase the perceptual and objective quality simultaneously, we directly combined the losses; thus, the overall loss function is

$$L_G = L_{\text{percep}} + \lambda L_G^{Ra} + \eta L_1. \tag{3}$$

It includes three parts: perceptual loss L_{percep}, adversarial loss L_G^{Ra}, and L1 loss L_1. λ and η are constants to balance the different loss functions. L_1 loss is the most common loss function in the SR area, which calculates the pixelwise differences between images, while perceptual loss compares high-level information to draw the distance between images and, thus, to get a better visual result. The adversarial loss establishes a constraint to keep the generated images and the real images in the same distribution.

2) Perceptual Loss: We calculate the perceptual loss by using the feature image before the activation layer. In the very deep network, the activated features are always few and scattered, so the activation layer will actually compromise the effectiveness of the perceptual loss. Furthermore, the activated feature may also cause the lightness difference between the reconstructed images and the origin images. The expression of perceptual loss is given as follows:

$$l_{\text{percept}} = \underbrace{l_{\text{X}}^{\text{SR}}}_{\text{content loss.}} + \underbrace{10^{-3} l_{\text{Gen}}^{\text{SR}}}_{\text{adversarial loss}}. \tag{4}$$

The perceptual loss consists of two parts: the content loss for image restoration and a part of generative loss. In our network, we choose L2 loss as the content loss in the perceptual loss, as the L1 loss already exists in our overall loss. By combining these two losses, we want our network to be able to restore more high-frequency details of the images. The adversarial loss in perceptual loss can be expressed as

$$l_{\text{Gen}}^{\text{SR}} = \sum_{n=1}^{N} -\log D_{\theta_D}\left(G_{\theta_G}\left(I^{\text{LR}}\right)\right) \tag{5}$$

where $D_{\theta_D}(G_{\theta_G}(I^{\text{LR}}))$ is the probability that the upscaled LR image is recognized as a natural image by the discriminator.

3) Adversarial Loss: Different from the perceptual loss, the adversarial loss for the generator is the Euclidean distance between the input LR images and the target HR image, i.e.,

$$L_G^{Ra} = -\mathbb{E}_{X_r}\left[\log(1 - D_{Ra}(X_r, X_i))\right] - \mathbb{E}_{X_f}\left[\log(D_{Ra}(X_i, X_r))\right] \tag{6}$$

where X_r represents the target HR image, G represents the generator, X_i represents the input LR image, D_{Ra} represents the discriminator of our network, and that for the discriminator is the symmetrical version

$$L_D^{Ra} = -\mathbb{E}_{X_r}\left[\log(1 - D_{Ra}(X_r, X_i))\right] - \mathbb{E}_{X_f}\left[\log(1 - D_{Ra}(X_i, X_r))\right]. \tag{7}$$

4) L1 Loss: Besides perceptual loss, we also need a PSNR-oriented loss to maintain the performance of our network on objective indexes. There are, in fact, two pixelwise loss functions; L1 and L2 losses can be used here to compare the images. As the L2 loss is a part of perceptual loss, we no longer need an extra L2 loss here. Thus, we use the L1 loss to calculate the pixelwise loss, of which the equation is

$$L = \frac{1}{N} \sum_{i=1}^{W} \sum_{j=1}^{H} \left\| I_{i,j}^{\text{HR}} - I_{i,j}^{\text{SR}} \right\|_1 \tag{8}$$

where $I_{i,j}^{\text{HR}}$ and $I_{i,j}^{\text{SR}}$ represent the (i, j) pixel on the target HR image and the generated SR image. The L1 loss will compare every single pixel of images and take the average.

IV. Experiments

A. Datasets and Implementation Details

We use the 2-D subset of BUAA-SID 1.5 [45], [46] as the training and testing sets, which includes 2042 gray-scale images of a single satellite captured from different viewpoints on a view sphere with the pitch angle of $y = [-0.5\pi, 0.5\pi]$ and the yaw angle of $\theta = [0, 2\pi]$. Similar to the other SR algorithm, the upscale factor is set to 4×. We randomly choose 1842 images from 2042 images as the training set, 100 images as the validation set, and 100 images as the testing set.

As for 4× scale, we first downscale the image to 1/4 of its original resolution by bicubic interpolation, use the random generation method of motion trail provided by DeblurGAN [43] to add random motion blur onto the LR images, and, finally, get 2042 pairs of corresponding LR and HR images for the experiments. Then, we convert these images to the RGB format and apply data augmentation, i.e., random rotating, flipping, and downscaling, on the datasets. The batch size is set to 32. Though contrastive learning is demonstrated to be effective in extracting degradation feature, as for space target images, most part of the image is in black; thus, random cropped patches in the same image could contain more mutual information than the degradation feature. Therefore, we implement an image augmentation method in the contrastive learning part; by setting a threshold, these pixels under the threshold will be changed to random color to minimize the mutual information in the first place.

We train our network by using (3) as the overall loss function to train the generator and set $\lambda = 0.005$ and $\eta = 0.01$. The training iteration is 200k and half the learning rate, which is initially 10^{-4}, at [10k, 20k, 50k, 100k]. We prefer a relatively lower learning rate because a high learning rate will lead to severe shaking in the training process, making the output image distorted.

We use the Adam optimizer to train the network, where β_1 is set to 0.9 and β_2 is set to 0.999. The activation function of our network is ReLU, and therefore, we use the Kaiming initialization [50] to initialize all the trainable parameters in our network. As for other SR methods, we train them by their default settings.

B. Image Quality Assessment (IQA)

In the following experiments, we use the most common evaluation metric, i.e., PSNR, SSIM, and PI. PSNR is based on the error between the corresponding pixels. SSIM compares the brightness, contrast, and structural differences between two images. Both PSNR and SSIM are positive evaluation metrics, which means that the larger the value is, the less the images distort. Also, to comprehensively compare the state-of-the-art SR methods and ours, we introduce PI [51] to evaluate the performance of these methods. Combined with Ma *et al.* [52] and NIQE [53], the formula of PI is

$$\mathrm{PI} = \frac{1}{2}((10 - \mathrm{Ma}) + \mathrm{NIQE}). \tag{9}$$

Ma and NIQE are both the no-reference IQA index focusing on visual comfort. With the two indexes combined, PI shows competitive accuracy on subjective evaluation. Thus, PI is to evaluate the perceptual quality of images. A low PI value indicates better image quality.

TABLE I

PSNR, SSIM Evaluation and Time Consumption of Different Network Structures: A: ESRGAN; B: Additional Convolutional and Deconvolutional Layers; C: Enlarged Shortcut Connection; and D: Changing in the Number of RRDB Modules

Methods	PSNR(dB)	SSIM	Time(s)	Flops(G)
A	31.06	0.9010	0.079	294.02
A+B	31.10	0.9029	**0.052**	230.59
A+E	31.66	0.9131	0.087	294.63
A+B+C	31.32	0.9093	**0.052**	230.59
A+B+C+4D	31.69	0.9162	0.055	267.34
A+B+E	31.54	0.9133	0.065	232.63
A+B+C+E	31.57	0.9159	0.065	232.63
A+B+C+2D+E	31.79	0.9188	0.068	251.01
A+B+C+4D+E	**31.83**	**0.9192**	0.070	269.38
A+B+C+6D+E	31.64	0.9160	0.073	287.76

RMSE is also a commonly used objective evaluate metric, of which the equation is given as follows:

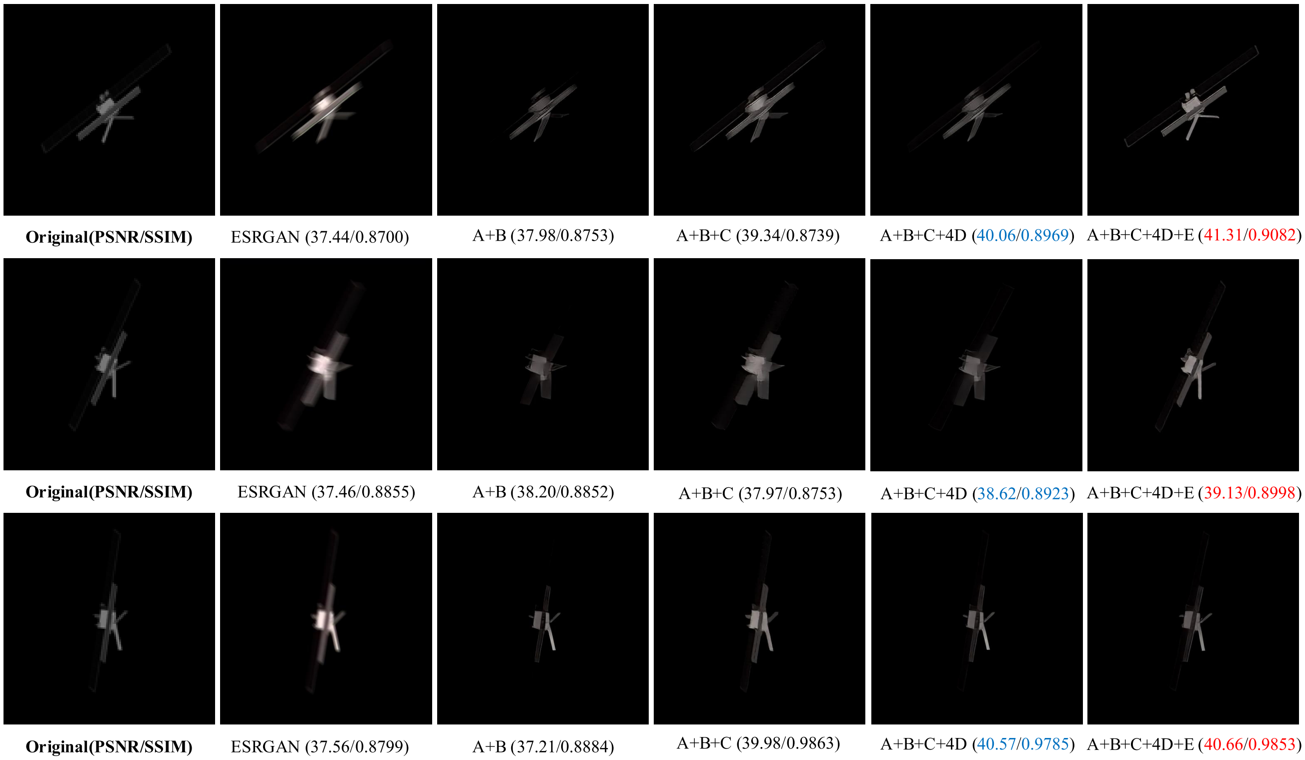
$$\sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}. \tag{10}$$

To better balance the visual quality and objective results, in the following experiments, we will compare the PI and RMSE simultaneously.

C. Ablation Study

In order to demonstrate the effectiveness of each part of our improvements, we define the original ESRGAN as the initial network and add each of our improvements step by step. We divide our network as following parts: A: the original ESRGAN; B: the symmetrical convolutional and deconvolutional layers on both ends of all the RRDB modules; C: enlarged shortcut connection; and D: changing the number of RRDB modules by 27. In the following experiment, we evaluate the effectiveness of each change by objective indexes, i.e., PSNR and SSIM. For example, A + B means ESRGAN with additional layers and all of the other implementations, i.e., loss functions, parameters, and so on, but without enlarged shortcut connection and change of the number of RRDB modules.

Table I shows the average PSNR, SSIM, and time consumption of different network structures for images in the validation set. It is obvious that the enlarged shortcut connection and additional RRDB module play important roles in improving the network performance as the PSNR result increases dramatically. As illustrated in Fig. 4, A + E (the degradation feature learning module) outperforms the ESRGAN on PSNR/SSIM by 0.60/0.121, which could demonstrate that the original ESRGAN cannot effectively reduce the influence of motion



Original(PSNR/SSIM) ESRGAN (37.44/0.8700) A+B (37.98/0.8753) A+B+C (39.34/0.8739) A+B+C+4D (40.06/0.8969) A+B+C+4D+E (41.31/0.9082)

Original(PSNR/SSIM) ESRGAN (37.46/0.8855) A+B (38.20/0.8852) A+B+C (37.97/0.8753) A+B+C+4D (38.62/0.8923) A+B+C+4D+E (39.13/0.8998)

Original(PSNR/SSIM) ESRGAN (37.56/0.8799) A+B (37.21/0.8884) A+B+C (39.98/0.9863) A+B+C+4D (40.57/0.9785) A+B+C+4D+E (40.66/0.9853)

Fig. 4. Ablation study on our network with different design choices and their corresponding visual and PSNR/SSIM evaluation results.

blur, but, once we enlarge the range of shortcut connection, the sense of vision gets much better.

As the task changes from a single SR task to both SR and deblur, the requirement of the learning ability of the network also increases. In practice, we tried to increase the number of the RRDB module to increase the network learning ability, To verify which is the most suitable number of RRDB modules, we designed a deeper experiment to evaluate the network performance, in which the module number varies from 23 to 29, and we find that, when the number of RRDB modules is 27 ($D = 4$), the network performs the best. As illustrated in Table I, solely changing the number of RRDB modules from 23 to 27 can increase the PSNR result by **0.37**.

Also, it is noticeable that the original ESRGAN costs more time than ours in the training phase. For the entire training set of the BUAA-SID 1.5 dataset, the processing time is about 153 s. For 5k epochs, the entire training procedure needs approximately 11 h. Thus, we need to shrink the feature map in the RRDB module. Similar to deblurGAN [43], we accept the additional convolutional and deconvolutional layers to accelerate the training speed. As the feature maps are shrunk, we have larger flexibility and adjustable space for the batch size to pursue better training results. In this way, as illustrated in Table I, shrinking the feature map incredibly could increase the temporal efficiency on the premise of not losing the reconstruction accuracy.

To better exhibit the performance difference between our network and the ESRGAN, we sampled a few results in the training progress as the training dynamics. As illustrated in Fig. 5, both ESRGAN and our network converge and perform well at the first validation (5k iteration), but ours performs better as the iteration continues. However, to pursue better results on the real dataset, we still need to train the model over 10k iteration. As the iteration increases, ESRGAN performs unstable on the PSNR result. During this period, from the validation set, we discover that the color of the output image dramatically changes, while ours performs stable. As both ESRGAN and SRDN drop BN layers away, the resistance on varying images also decreases. In our experiment, space target images with severe motion blur are used, which could lead to the stability of such network decrease. However, with the help of the degradation feature extraction module, our SRDN can avoid instability for the most part.

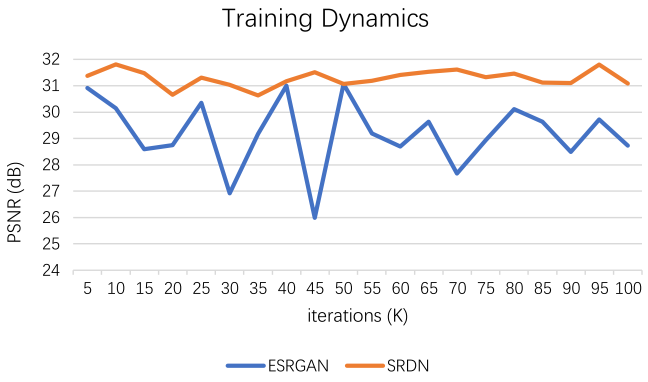


Fig. 5. Training dynamics between ESRGAN and SRDN.

D. Sample Results

To demonstrate the effectiveness of our method, we compare it with state-of-the-art methods, including

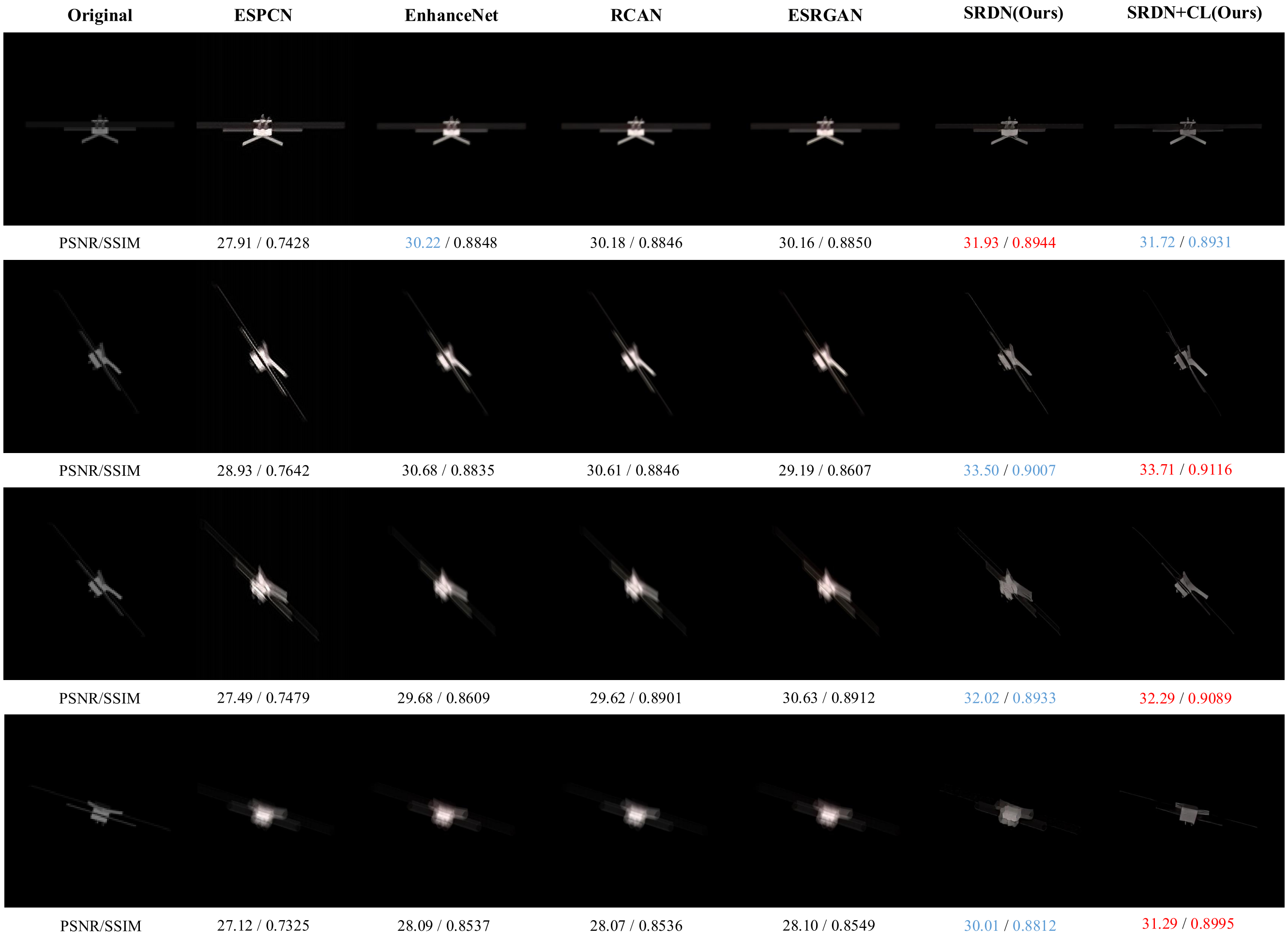


Fig. 6. Visual results with PSNR/SSIM evaluation of state-of-the-art SR methods and ours for satellite images with varying degree of motion blur.

SRCNN [8], ESPCN [24], EnhanceNet [54], RCAN [15], and ESRGAN [13].

Fig. 6 shows some sample images reconstructed by the classical SR methods and our network. It is noticeable that, in some cases, images reconstructed by classical SR methods achieve good results on PSNR evaluation but still leave severe motion blur. That is mainly because the traditional SR methods cannot recover the texture of the space target. In order to pursue better PSNR results, the PSNR-oriented SR methods tend to output oversmoothed images. Interpreting with the existing motion blur, finally, leads to the images becoming much smoother than its original HR image and lack of high-frequency details, which is far from our satisfaction. With a deeper network structure and better learning ability, our SRDN can reconstruct the image with much slighter motion blur, achieving better performance not only on PSNR results but also on visual feelings.

Also, in order to give a better view of the effectiveness of those SR methods on motion deblur, Fig. 6 shows some of the output satellite images with varying degrees of motion blur, i.e., the first row has no motion blur, and motion blur gets stronger row by row. Thus, we can better judge all kinds of images by the subjective view.

From the images, we can see that the two most important differences between the classical SR methods and ours are the hue and the degree of motion blur. In most cases for images generated by classical SR methods, the color is obviously much brighter than the original one. As the SR methods cannot recognize the motion blur, they consider it as a part of the original image, and thus, the color of output images is significantly different from the original one.

Even taking the color issue out of consideration, the current SR methods cannot recover the images well once the original images are with serious motion blur. In contrast, compared to SRDN, SRDN + CL does not perform better when the motion blur is in a low range, e.g., the first column, but outperforms all other methods when the blur is severe, which demonstrates that our degradation feature learning module works properly in extracting the degradation features. Therefore, ours can significantly reduce the influence of motion blur, achieving ideal results not only on PSNR evaluation but also the sense of vision.

E. Statistic Results

In Table II, we compare state-of-the-art SR methods with our network, SRDN, and ours outperforms all of them on

TABLE II

PSNR AND SSIM EVALUATIONS OF THE EFFECTIVENESS OF STATE-OF-THE-ART METHODS AND OURS IN DEALING WITH SATELLITE IMAGES WITH RANDOM MOTION BLUR

Scale	Methods	PSNR(dB)	SSIM
2×	Bicubic	33.61	0.9057
	ESPCN	31.99	0.8124
	EnhanceNet	34.79	0.9167
	RCAN	36.14	0.9328
	ESRGAN	36.29	0.9317
	USRNet	36.06	**0.9361**
	DRN	36.31	0.9350
	HAN	37.10	0.9369
	BSRGAN	37.34	**0.9372**
	SRDN(ours)	37.22	0.9336
	SRDN+CL(ours)	**37.39**	0.9362
4×	Bicubic	27.11	0.8110
	ESPCN	24.81	0.7468
	EnhanceNet	27.64	0.8836
	RCAN	30.76	0.9049
	ESRGAN	31.06	0.9010
	USRNet	29.32	0.9056
	DRN	30.88	0.9183
	HAN	31.41	0.9098
	BSRGAN	31.79	**0.9224**
	SRDN(ours)	31.69	0.9162
	SRDN+CL(ours)	**31.83**	0.9192

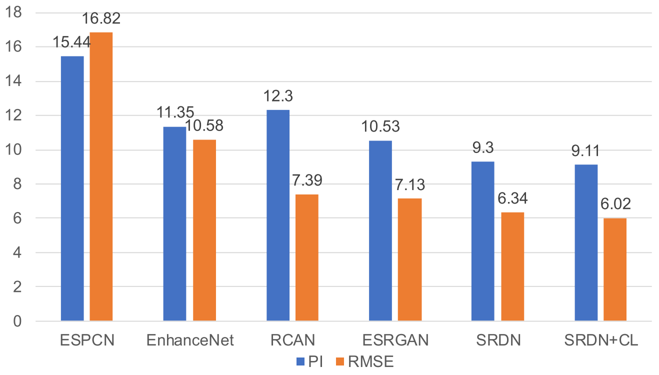


Fig. 7. PI and RMSE comparison between current SR methods (the lower the better).

average PSNR and SSIM results. Specifically, compared with the state-of-the-art SR methods, i.e., ESPCN [24], EnhanceNet [54], RCAN [15], ESRGAN [13], USRNet [55], HAN [56], DRN [32], and BSRGAN [31], on the BUAA-SID 1.5 dataset, our model achieves 37.39/0.9362 (PSNR/SSIM), when the scale is 2×, and 31.83/0.9192, when the scale is 4×. Compared with the solely SR methods, i.e., RCAN, ESRGAN, HAN, and so on, in 2× scale comparison, a significant improvement is shown on PSNR result, which leads by 0.29 than the second best, while, in 4× scale comparison, our network by leads 0.42 on PSNR, respectively. Compared with the joint methods, i.e., DRN and BSRGAN, we achieve 0.04/0.05 lead on PSNR result in 2×/4× comparison.

In addition, the current PSNR-oriented SR methods may provide better PSNR results but ignore the subjective feeling and, thus, output oversmooth HR images. Thus, besides the PSNR and SSIM, as Blau *et al.* [51] indicate, we evaluate the SR methods by PI and RMSE (see Fig. 7). Our method achieves the best both on RMSE and PI results. Compared to ESRGAN, we get a 1.42 lead on PI evaluation, which illustrates that our model could be able to restore space target images with a better visual impression.

Overall, images with poorer RMSE results can hardly achieve ideal results on PI. However, the oversmooth images, e.g., the images reconstructed by EnhenceNet, will not get a good result on PI. By contrast, with the help of perceptual loss, both ESRGAN and our network achieve good results on PI assessment.

TABLE III

CLASSIFICATION ACCURACY OF SPACE TARGET RECOGNITION ON THE BUAA-SID 1.5 DATASET FOR DIFFERENT IMAGE PREPROCESSING METHODS. THE BEST IS IN **BOLD**, AND THE SECOND IS UNDERLINED

Metrics(%)	Rank-1	Rank-5	mAP
Bicubic	94.04	96.98	86.07
ESRGAN	94.22	96.85	86.79
BSRGAN	94.62	97.57	87.64
SRDN	94.27	97.33	87.10
SRDN+CL	**94.71**	**98.14**	**88.23**

Table III shows the evaluation of effectiveness of our method on space target recognition. In the experiment, we use ten classes of the satellite in BUAA-SID 1.5-1d and evaluate the top-*k* accuracy and mean average precision by ResNet-50. From the table, we can see that, compared to the interpolation method, i.e., bicubic, with the help of SR methods, the classification accuracy raises in varying degrees. Comparing our method to the SR method, ESRGAN, our SRDN outperforms 0.05%/0.31% on rank-1 and mAP results, respectively. Compared to the joint method, BSRGAN, our SRDN + CL outperforms 0.09%/0.59%, respectively, which could demonstrate that both SR and the deblurring method are effective in improving the classification accuracy on blurred LR space target images.

V. CONCLUSION

Traditionally, to restore a better image from an LR image with motion blur, the only approach is to deblur the image and then upscale it. In order to reduce the complexity and mutual interference, we propose a new SR network called SRDN, especially for the space target with motion blur, which combines the deblur task and the SR task into a single network.

We first downscale the feature image inside the network by using multiple convolutional layers before the feature extraction unit of the network; thus, we can increase the batch size to accelerate the training process and receiving better output images. The range of the shortcut connection is also enlarged to help the network learn the residual part between the input and output, and finally, we increase the number of RRDB modules in ESRGAN, which significantly raises the computational capability and learning ability of the network.

Extensive experiments prove that, compared to other SR algorithms, SRDN can restore the space target images with motion blur effectively and achieve better results on general evaluation metrics. For every degree of blur, the perceptual quality of the images restored by our network is the best.

References

[1] L. Wang, Z. Huang, Y. Gong, and C. Pan, "Ensemble based deep networks for image super-resolution," *Pattern Recognit.*, vol. 68, pp. 191–198, Aug. 2017.

[2] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, "Convolutional sparse coding for image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1823–1831.

[3] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920–1927.

[4] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5197–5206.

[5] J. Salvador and E. Perez-Pellitero, "Naive Bayes super-resolution forest," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 325–333.

[6] W. Xie, X. Zhang, Y. Li, J. Lei, J. Li, and Q. Du, "Weakly supervised low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 51, no. 8, pp. 3889–3900, Aug. 2021.

[7] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1529–1543, May 2020.

[8] C. Dong, C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199.

[9] S. Lei, Z. Shi, and Z. Zou, "Coupled adversarial training for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3633–3643, May 2020.

[10] S. Mei, R. Jiang, X. Li, and Q. Du, "Spatial and spectral joint super-resolution using convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4590–4603, Jul. 2020.

[11] C. Dong, C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 391–407.

[12] C. Ledig *et al.*, "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.

[13] X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 63–79.

[14] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.

[15] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. SPIEEuropean Conf. Comput. Vis.*, 2018, pp. 286–301.

[16] J. Ma, H. Zhang, B. Li, and Y. Wang, "Space target recognition algorithm based on two-dimensional wavelet transform," *J.-Nat. Univ. Defense Technol.*, vol. 28, no. 1, pp. 57–61, 2006.

[17] J. Zhang and X. Zhou, "Research on feature recognition algorithm for space target," *Proc. SPIE*, vol. 6786, Nov. 2007, Art. no. 678616.

[18] K. Alex, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[19] X. Yang, T. Wu, N. Wang, Y. Huang, B. Song, and X. Gao, "HCNN-PSI: A hybrid CNN with partial semantic information for space target recognition," *Pattern Recognit.*, vol. 108, Dec. 2020, Art. no. 107531.

[20] W. Xie, J. Lei, S. Fang, Y. Li, X. Jia, and M. Li, "Dual feature extraction network for hyperspectral image analysis," *Pattern Recognit.*, vol. 118, Apr. 2021, Art. no. 107992.

[21] H. Zeng and Y. Xia, "Space target recognition based on deep learning," in *Proc. 20th Int. Conf. Inf. Fusion (Fusion)*, Jul. 2017, pp. 1–5.

[22] X. Yang, X. Nan, and B. Song, "D2N4: A discriminative deep nearest neighbor neural network for few-shot space target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3667–3676, May 2020.

[23] T. Wu *et al.*, "T-SCNN: A two-stage convolutional neural network for space target recognition," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul./Aug. 2019, pp. 1334–1338.

[24] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.

[25] 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.

[26] 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.

[27] L. Wang, S. Guo, W. Huang, and Y. Qiao, "Places205-VGGNet models for scene recognition," 2015, *arXiv:1508.01667*.

[28] Y. Mei, Y. Fan, Y. Zhou, L. Huang, T. S. Huang, and H. Shi, "Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5690–5699.

[29] X. Pan, X. Zhan, B. Dai, D. Lin, C. C. Loy, and P. Luo, "Exploiting deep generative prior for versatile image restoration and manipulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Sep. 24, 2021, doi: 10.1109/TPAMI.2021.3115428.

[30] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced GAN for remote sensing image superresolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5799–5812, Aug. 2019.

[31] K. Zhang, J. Liang, L. Van Gool, and R. Timofte, "Designing a practical degradation model for deep blind image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2021, pp. 4791–4800.

[32] Y. Guo *et al.*, "Closed-loop matters: Dual regression networks for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5407–5416.

[33] X. Zhang, H. Dong, Z. Hu, W. Lai, F. Wang, and M. Yang, "Gated fusion network for degraded image super resolution," *Int. J. Comput. Vis.*, vol. 128, pp. 1699–1721, Jan. 2020.

[34] K. Jiang, Z. Wang, P. Yi, and J. Jiang, "Hierarchical dense recursive network for image super-resolution," *Pattern Recognit.*, vol. 107, Nov. 2020, Art. no. 107475.

[35] J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, "Residual feature aggregation network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2359–2368.

[36] L. Lucy, "An iterative technique for the rectification of observed distributions," *Astronomical J.*, vol. 79, pp. 745–754, 1974.

[37] S. Suresh, S. Lal, C. Chen, and T. Celik, "Multispectral satellite image denoising via adaptive cuckoo search-based Wiener filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4334–4345, Aug. 2018.

[38] P. Meincke, "Linear GPR inversion for lossy soil and a planar air-soil interface," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 12, pp. 2713–2721, Dec. 2001.

[39] Y.-T. Zhou, R. Chellappa, A. Vaid, and B. K. Jenkins, "Image restoration using a neural network," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 36, no. 7, pp. 1141–1151, Jul. 1988.

[40] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 787–794, Jul. 2006.

[41] L. Xu, S. Zheng, and J. Jia, "Unnatural L0 sparse representation for natural image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1107–1114.

[42] S. Xiang, G. Meng, Y. Wang, C. Pan, and C. Zhang, "Image deblurring with matrix regression and gradient evolution," *Pattern Recognit.*, vol. 45, no. 6, pp. 2164–2179, Jun. 2012.

[43] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8183–8192.

[44] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8878–8887.

[45] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, and C. Pan, "Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3322–3337, Jun. 2017.

[46] H. Zhang and J. Zhiguo, "Multi-view space object recognition and pose estimation based on Kernel regression," *Chin. J. Aeronaut.*, vol. 27, no. 5, pp. 1233–1241, 2014.

[47] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 723–731.

[48] S. Bell-Kligler, A. Shocher, and M. Irani, "Blind super-resolution kernel estimation using an internal-GAN," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 284–293.

[49] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.

[50] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[51] 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. Workshops*, 2018, pp. 334–355.

[52] 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.

[53] A. Mittal, R. Soundararajan, and A. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[54] 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.

[55] K. Zhang, L. Van Gool, and R. Timofte, "Deep unfolding network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3217–3226.

[56] B. Niu *et al.*, "Single image super-resolution via a holistic attention network," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 191–207.

[57] I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, vol. 27. 2014.

Xi Yang (Member, IEEE) received the B.Eng. degree in electronic information engineering and the Ph.D. degree in pattern recognition and intelligence system from Xidian University, Xi'an, China, in 2010 and 2015, respectively.

From 2013 to 2014, she was a Visiting Ph.D. Student with the Department of Computer Science, The University of Texas at San Antonio, San Antonio, TX, USA. In 2015, she joined the State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, where she is currently an Associate Professor of communications and information systems. Her research interests include image/video processing, computer vision, and multimedia information retrieval.

Xiaoqi Wang received the B.E. degree in communications engineering from Xidian University, Xi'an, China, in 2019, where he is currently pursuing the M.S. degree in communication and information system.

His research interests include deep learning, image processing, and person reidentification.

Nannan Wang (Member, IEEE) received the B.Sc. degree in information and computation science from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2009, and the Ph.D. degree in information and telecommunications engineering from Xidian University, Xi'an, in 2015.

From September 2011 to September 2013, he was a Visiting Ph.D. Student with the University of Technology Sydney, Ultimo, NSW, Australia. He is currently a Professor with the State Key Laboratory of Integrated Services Networks, Xidian University. He has published over 100 articles in refereed journals and proceedings, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (T-PAMI), *International Journal of Computer Vision* (IJCV), and Conference on Neural Information Processing Systems (NeurIPS). His research interests include computer vision, pattern recognition, and machine learning.

Xinbo Gao (Senior Member, IEEE) received the B.Eng., M.Sc., and Ph.D. degrees in electronic engineering, signal, and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively.

From 1997 to 1998, he was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Post-Doctoral Research Fellow with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong. Since 2001, he has been with the School of Electronic Engineering, Xidian University. He is currently a Cheung Kong Professor of the Ministry of Education of China and a Professor of pattern recognition and intelligent system with Xidian University and a Professor of computer science and technology with the Chongqing University of Posts and Telecommunications, Chongqing, China. He has published six books and around 300 technical articles in refereed journals and proceedings. His research interests include image processing, computer vision, multimedia analysis, machine learning, and pattern recognition.

Prof. Gao is also a fellow of the Institute of Engineering and Technology and the Chinese Institute of Electronics. He has served as the general chair/co-chair, the program committee chair/co-chair, or a PC member of around 30 major international conferences. He is also on the editorial boards of several journals, including *Signal Processing* (Elsevier) and *Neurocomputing* (Elsevier).