IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 7, 2021, accepted July 25, 2021, date of publication August 2, 2021, date of current version August 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3101858

Mars Image Super-Resolution Based on Generative Adversarial Network

CONG WANG[1,2], YIN ZHANG[1], YONGQIANG ZHANG[1], RUI TIAN[1], AND MINGLI DING[1]
[1]School of Instrument Science and Engineering, Harbin Institute of Technology (HIT), Harbin 150001, China
[2]Shanghai Institute of Satellite Engineering, Shanghai 200240, China

Corresponding authors: Mingli Ding (mingli.ding.hit@gmail.com) and Yongqiang Zhang (yongqiang.zhang.hit@gmail.com)

This work was supported in part by China Postdoctoral Science Foundation under Grant 259822, in part by the National Postdoctoral Program for Innovative Talents under Grant BX20200108, in part by the National Science Foundation of China under Grant 61976070, and in part by the Science Foundation of Heilongjiang Province under Grant LH2021F024.

ABSTRACT High-resolution (HR) Mars images have great significance for studying the land-form features of Mars and analyzing the climate on Mars. Nowadays, the mainstream image super-resolution methods are based on deep learning or CNNs, which are better than traditional methods. However, these deep learning based methods obtain low-resolution(LR) images usually by using an ideal down-sampling method (*e.g.* bicubic interpolation). There are two limitations in the existing SR methods: 1) The paired LR-HR data by using such methods can achieve a satisfactory results when tested on an ideal datasets. But, these methods always fail in real Mars image super-resolution, since real Mars images rarely obey an ideal down-sampling rule. 2) The LR images obtained by ideal down-sampling methods have no noise while real Mars images usually have noise, which leads to the super-resolved images are not realistic in texture details. To solve the above-mentioned problems, in this article, we propose a novel two-step framework for Mars image super-resolution. Specifically, to address limitation 1), we focus on designing a new degradation framework by estimating blur-kernels. To address limitation 2), a Generative Adversarial Network (GAN) is trained to generate noise distribution. Extensive experiments on the Mars32k dataset demonstrate the effectiveness of the proposed method, and we achieve better qualitative and quantitative results compared to other SOTA methods.

INDEX TERMS Generative adversarial network, kernel estimation, mars image super-resolution, noise model.

I. INTRODUCTION

Among the eight major planets, Mars is the most similar to the Earth, and it is also considered the most likely to give birth to life. The natural environment of Mars is also similar to the earth, therefore, Mars becomes the main target for deep space exploration in this century.

In the process of exploring Mars, obtaining high-resolution images of Mars is of great significance for subsequent scientific research. For example, high-resolution Mars images are used to identify rock types, compositions, rock texture, structure and characteristics, *etc.* . Moreover, high-resolution Mars images can also be used for the missions of navigating and exploring the Mars rover, *i.e.* , through the obtained high-resolution Mars images, the path and distance of the rover can be further planned. Furthermore, during the process of transmission from Mars to earth, high-resolution images may face problems of slow transmission speed and transmission failure. Here, lower-resolution images can be used during transmission to reduce communication delays, and super-resolution algorithms can be used on the earth to finish the specific computation tasks.

Moreover, Mars often has dust storms and the atmosphere on the Mars is thin, the atmosphere and the environment in outer space is unpredictable, and the temperature changes drastically, which makes the image obtained from the camera on Mars rover board unstable. At the same time, the Mars rover imaging system will be affected by cosmic radiation, which also makes the images of Mars taken by the probe are blurry. Generally, improving the hardware equipment can increase the resolution of Mars images, which is also the simplest and most intuitive method. However, this type of method has a high cost and a long development period, and increasing the complexity of imaging equipment may bring

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Liu.

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

negative effects such as noise and slowing down the transmission rate. Therefore, from a long-term perspective, it is of high practical value and necessity to improve the resolution of Mars images through algorithms.

Super-resolution methods aim to reconstruct a High-resolution (HR) image from a Low-resolution (LR) image by recovering high-frequency details. Among them, a boost in performance was achieved in past few years by introducing deep learning based methods [1]–[9]. These methods assume that LR images are obtained by the following degradation method:

$$I_{LR} = (I_{HR} * k) \downarrow_s + n \tag{1}$$

where k and n indicate super-resolution (SR) kernel and noise respectively, and the goal is to recover HR images I_{HR} from the given LR images I_{LR}.

However, these deep learning based super-resolution methods usually cannot achieve satisfied results when directly applied to real Mars images. Specifically, they may suffer from two limitations. First, most of them use an ideal down-sampling methods (*e.g.* bicubic) to prepare the corresponding LR images when training the SR networks. However, the HR and LR images of real Mars do not obey the those ideal interpolation relation. The reason is that the ideal down-sampled images (LR) do not belong to the same domain as the original Mars images, and the high-frequency details of images will be lost. Figure 1 shows the results that directly use the deep learning based methods on 'non-ideal' LR images. From Figure 1, we can see that EDSR [5] and ESRGAN [10] produce unsatisfied super-resolution results for the real Mars images.

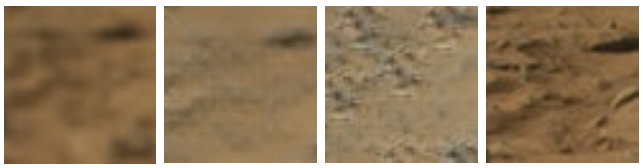

FIGURE 1. Visualization comparison results of EDSR [5], ESRGAN [10], and ours.

The second limitation is that the LR images obtained by ideal interpolation methods have no noise, while real Mars images usually have noise. Moreover, the noise distribution of the real Mars images are very complex so that some methods [11]–[13] add a known noise (*e.g.* Gaussian noise) to address this issue. Therefore, one solution for this problem is to learn a noise distribution over Mars images and inject it to LR images.

In this paper, we propose a new degradation framework for Mars image Super-Resolution, which contains a kernel estimation and a noise model. To address the first limitation, we use the kernel estimation method to generate realistic LR Mars images. To address the second limitation, we use the noise extraction algorithm to collect noise from the original images and add it to the down-sampled images (LR). In this way, the burden of feature extraction is reduced as the model has rich prior information from real Mars images. To verify the effectiveness of our proposed method, we conducted extensive experiments on the Mars32k dataset. The experimental results show that each component of our method is helpful to improve the quality of the super-resolution images, and our proposed method achieve a better result compared current SOTA methods.

In summary, our contributions are three fold:

- We design a kernel estimation method to obtain realistic LR images which can remove artifacts in SR images.
- A noise extraction algorithm and a GAN-based method are utilized for noise modeling and sampling, which make the degraded images have a similar noise distribution with the real image.
- We demonstrate the effectiveness of our proposed method in Mars image super-resolution and show that a better result is achieved over previous state-of-the-art approaches on Mars32k dataset (PSNR increase 2.074dB by kernel estimation and 0.198dB by noise injection).

The rest of the paper is organized as follows. We review the related work in Section 2. In Section 3, the kernel estimation, noise extraction algorithm, the detailed architecture of our noise model, and the SR model are described. In Section 4, we present some experiments and ablation studies on Mars32k dataset. Finally, the conclusions are provided in Section 5.

II. RELATED WORKS

The method of obtaining super-resolution Mars images mainly relies on optical cameras. In November 1996, the United States launched the Mars Global Surveyor probe [14], [15], which carries a Mars orbiting camera and a narrow angle camera that obtained gray-scale high resolution images (typically 1.5 to 12 m per pixel), and includes red and blue wide angle cameras for collecting context (240 m per pixel) and daily global images (7.5 km per pixel). Europe's first Mars exploration project (*i.e.* Mars Express) [16] was launched in June 2003, its high-resolution stereo camera can achieve full-surface shooting of Mars with a resolution of 10m, and it can achieve 2 m ultra-high resolution [17] on the local surface of Mars.

Usually using the ability of hardware system to improve resolution of images is limited by factors such as volume and quality, and the sub-pixel imaging technology has become an effective method of using algorithms to improve image resolution under the limitations of the current hardware systems. Since most of images obtained by the probe are under-sampled, the sub-pixel technology is a method of increasing the sampling rate, and then using a series of sub-pixel processing (such as interpolation and recursive iterative operations) to obtain clear Mars images. In following, we main introduce some related works about the image super-resolution.

A. IMAGE SUPER-RESOLUTION

Interpolation is the earliest image super-resolution algorithm. [18] first proposes the image super-resolution algorithm, which was implemented in the frequency domain using Fourier transform. In the frequency domain, the under-sampling samples of the image appear as spectrum aliasing, and finally the super-resolution data is computed by the under-sampling data. Based on [18], Kim [19] proposes blur noise to improve the super-resolution algorithm. The reconstruction-based methods realizes the registration of LR images and HR images through prior knowledge. The traditional learning-based methods [20], [21] mainly use HR images to obtain LR images through some degradation processing, or through algorithm to learn the mapping from LR images to HR images, and then use this mapping for image super-resolution reconstruction.

In recent years, with the development of deep learning theory, neural networks have demonstrated their excellent performance. SRCNN [1] uses a convolutional neural network for image super-resolution, and only uses three convolutional layers to achieve a satisfactory super-resolution results. The three convolutional layers correspond to feature extraction, nonlinear mapping and finally image reconstruction, respectively. A jump connection in the residual network [22] promotes the performance of convolutional neural networks. Many scholars use this framework in image super-resolution and achieve good results. VDSR [4] is proposed to use ResNet to accelerate the network convergence, and by expanding the receptive field and increasing the depth of the network to improve the super-resolution results.

However, these SR methods assume that the LR images are down-sampled from the HR images by a fixed bicubic interpolation. In this paper, we focus on using GAN to estimate the process of down-sampling.

B. NOISE MODELING

Noise modeling approaches are usually used to address image blind de-noising problems [23]–[28]. These methods include the conjunction of noise modeling and an adaptive de-noising algorithm generally. For example, Multiscale [23] is an adoptive of Non-local Bayes approach [29] which assumes the noise model of each patch and its nearby patches to be zero-mean correlated Gaussian distribution. NMBD [27] is proposed to model image noise with mixture of Gaussian (MoG) and developed a Low-rank MoG filter to recover the clean images. However, these methods only utilize the internal information of a single input image and explicitly define the noise model, which may limit the capability of noise modeling. Unlike the previous methods, in order to utilize more external information, we use a noise extraction algorithm to get noise blocks from Mars images and use an GAN to generate similar noise distribution.

C. GENERATIVE ADVERSARIAL NETWORK

Recently, GAN [30] have attracted widespread attention [31]–[34]. The essence of GAN is to generate similar distributions through adversarial learning strategy. On the one hand, this feature makes it is possible to train a GAN to obtain a mapping from a pair of LR and HR images for image super-resolution. SRGAN [2] is the first work that uses of GAN for super resolution. The authors find that the high PSNR results will cause discomfort in human visual perception, and propose to use the perceptual loss of the VGG network to replace the original MSE loss, thereby improving the visual experience of the SR images. Wang *et al.* improves the network structure on the basis of SRGAN and proposes ESRGAN [10]. They remove batch normalization layer (BN), and the network structure is changed to more complex RRDB (Residual in Residual Dense Block) and relative discriminators is used to further improve the quality of super-resolution images. Reference [35] gives a survey for image synthesis with adversarial networks, readers can refer this survey for more details about GANs.

On the other hand, a lot of works [36]–[40] have proved that GANs could learn complex distribution. Therefore, we think GAN has great capability to learn the latent noise model. However, training a GAN is tricky and GAN not easy to converge. In this paper, we choose DCGAN [36] to generate noise distribution and use WGAN [41], [42] to overcome the balance between the generative network and the discriminative network.

III. PROPOSED METHOD

In this section, we describe the details of the proposed degradation framework for Mars image Super-Resolution. First, we analyze the difference between Mars images and Earth images. Then, we introduce the pipeline and the overview of the proposed degradation framework, as shown in Figure 2 and Algorithm 1 respectively. Finally, a noise extraction algorithm and noise model are designed to obtain a more realistic LR image.

Algorithm 1 Degradation Framework

Input: Real Mars images X, HR Mars images Y
  Init Kernel Bank $\mathbf{K} = null$ and Noise Bank $\mathbf{N} = null$
  for all $i_{src} \in X$ **do**
    Estimate k from i_{src}
    Add k to $\mathbf{K}$.
    Extract n from i_{src}
    if n meets Equal 4 and 3 **then**
      Add n to $\mathbf{N}$
    end if
  end for
  for all $I_{HR} \in Y$ **do**
    Randomly select $k_i \in \mathbf{K}$
    $I_k = (I_{HR} * k_i) \downarrow_s$ where s is a downsampling scale factor
    Randomly select $n_i \in \mathbf{N}$
    $I_{LR} = I_k + n_i$
  end for
Output: Realistic paired images $\{I_{LR}, I_{HR}\}$

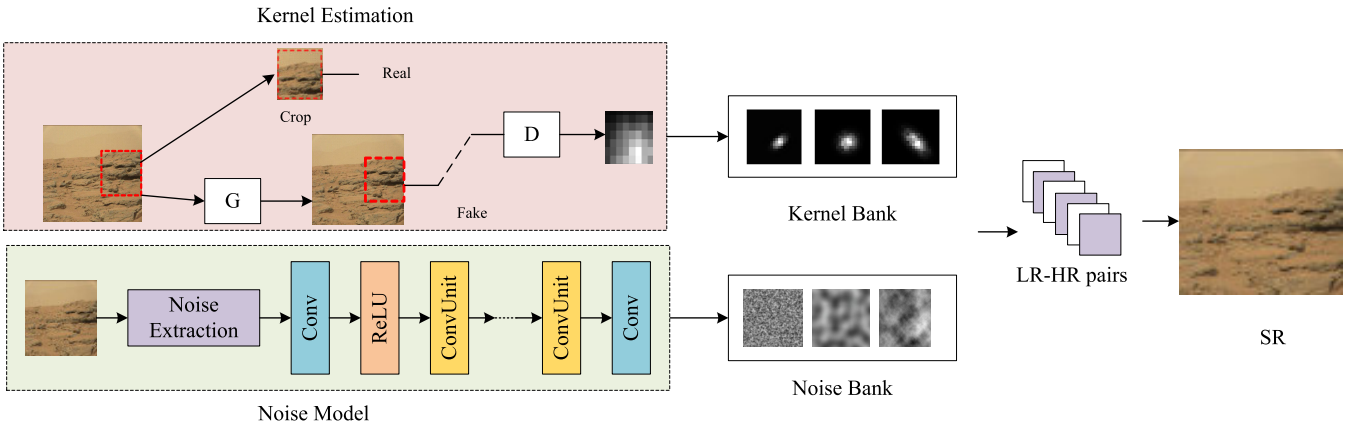

FIGURE 2. The pipeline of the proposed Mars Image Super-Resolution method. Kernel estimation obtains blur-kernel from real Mars images and puts them into a kernel bank', noise model extracts noise from images and puts them into a 'noise bank'. Finally, randomly selecting a blur-kernel and noise from 'kernel bank' and 'noise bank' to produced realistic LR-HR image pairs for training an SR network.

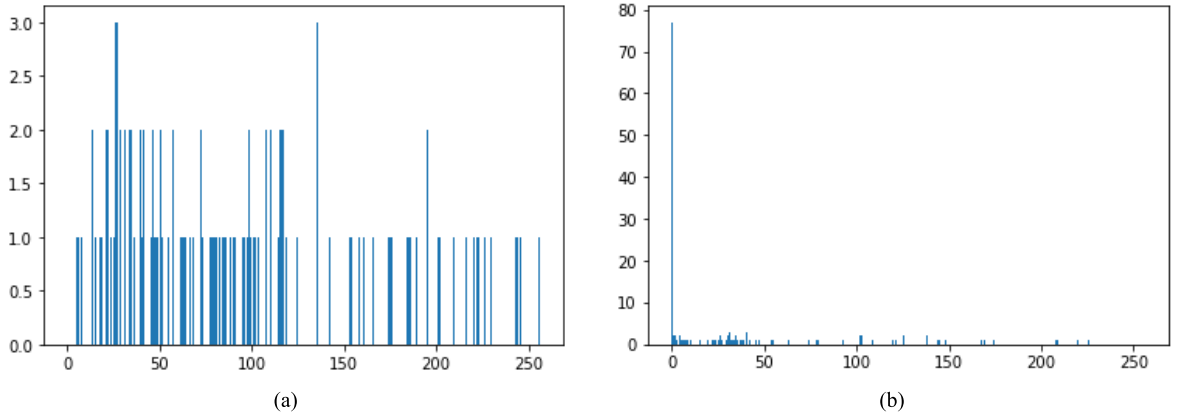

FIGURE 3. Comparison of gray-scale between Earth images and Mars images. This is a gray-scale histogram which (a) is the Earth image and (b) is the Mars image. The horizontal direction represents pixel values from 0 to 256 and the vertical direction represents the number of pixels.

A. DIFFERENCE BETWEEN MARS IMAGES AND EARTH IMAGES

In order to study whether it is possible to directly apply the existing earth image super-resolution methods to the Mars images, we first analyze the gray-scale of the earth image and the gray-scale of the Mars image, as shown in Figure 3.

DIV2K and Mars32k datasets are use to analyze the difference between Earth and Mars images. From Figure 3, we can find that the gray-scale distribution of the Earth image is uniform, while the gray-scale range of the Mars image is narrow, mostly between 0-40. This means Mars images lack surface information, thus to extract the feature of Mars images is very difficult.

To further verify the difference between the images of Mars and Earth, we perform feature extraction experiments on these two type of images. We select SURF [43], which has a better execution efficiency than the SIFT algorithm [44], to calculate the number of feature points in the Earth and Mars images, as shown in Figure 4. From Figure 4, We find that the number of feature points in the Earth image is 752, and the number of feature points in the Mars image is 256. The feature points of Mars image are obviously less than that of Earth image.

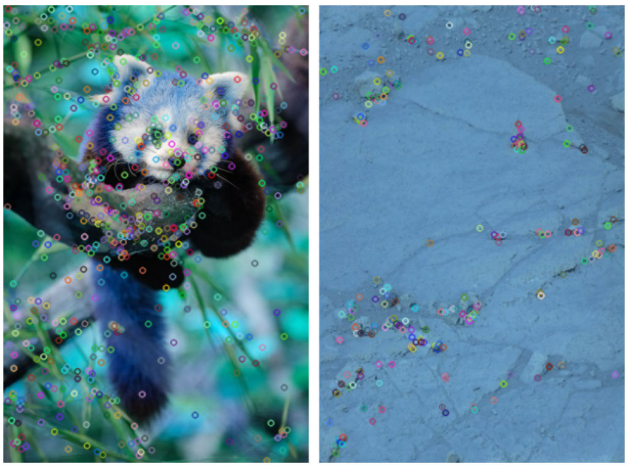

FIGURE 4. Feature distribution of Earth images and Mars images.

We draw conclusions through experiments that the earth image has many feature points and rich texture features, and the foreground and background of images are obviously different. The number of feature points in the Mars images is relatively small, mainly concentrated in areas with complex land-form. Moreover, most of Mars is still

plain and with sparse texture features, which brings great difficulties and challenges to the task of Mars images super-resolution. In the following parts, we mainly describe the details of our proposed Mars images super-resolution method.

B. OVERVIEW OF DEGRADATION FRAMEWORK

For image super-resolution, an important task is to obtain HR-LR image pairs. Generally, one of the easiest methods to obtain LR images is image sub-sampling. However, in the process of down-sampling, the high-frequency details are lost, at the same time, the noise distribution of images is also changed. In this paper, we use GAN to learn the distribution of real Mars images and to generate a noise distribution similar to the real Mars image.

As shown in Figure 2, there are three components in the proposed architecture. The first component is the kernel estimation, which uses s generator G to generate a down-sampled image that its distribution is as close as possible to that of the LR image. The output of the kernel estimation is underlying SR kernel and then put them into a 'kernel bank'. The second component is the noise model, in which the noise is collected through the noise extraction algorithm, and then put into a 'noise bank'. The last component is to obtain final LR image, we randomly select kernel from 'kernel bank' and randomly select noise from the 'noise bank' to generate the final LR Mars images by Equal 1. To describe our method concisely, we show this process as an algorithm as in Algorithm 1. In the following parts, we will detailed introduce the kernel estimation and noise model, respectively.

C. KERNEL ESTIMATION

We use KernelGAN [45] to explicitly estimate kernels from real Mars images, and the appropriate parameters are set based on real Mars images. The Generator G constitutes the down-sampling model, which is a 6 layers convolutional network with 64 channels and without any activation layer. Therefore the parameters of all layers can be combined into a fixed kernel. The filters are $7 \times 7, 5 \times 5, 3 \times 3, 1 \times 1, 1 \times 1, 1 \times 1$, respectively.

The Discriminator D is a fully convolutional network, and the architecture is a 7×7 convolution filter followed by six 1×1 convolutions, including Batch normalization(BN) [46], ReLU [47] and a Sigmoid [48] activation function. The goal of the discriminator D is to learn the distribution of patches of the input image I_{LR} and to discriminate the real patches belonging to this distribution or the fake patches generated by G.

The kernel is implicitly captured by the trained weights of G. Convolving all the filters of G sequentially with stride 1 can obtain the kernel k. When the kernel is obtained, it is a small array that can be supplied to SR algorithms. In order to prevent the optimization process from producing too scattered and smooth kernels, the estimated kernel needs to meet the following constraints:

$$\arg\min_{k} ||(I_{src} * k) \downarrow_s - I_{src} \downarrow_s ||_1 + |1 - \sum k_{i,j}| + \sum k_{i,j} \cdot m_{i,j} + |1 - D((I_{src} * k) \downarrow_s)| \quad (2)$$

where $(I_{src} * k) \downarrow_s$ is the down-sampled LR image with kernel k, and $I_{src} \downarrow_s$ is the down-sampled image with ideal kernel. Our purpose is to minimize this error to encourage the down-sampled image to preserve the low-frequency information of real images. What's more, the second term of Eq. 2 is to constrain the sum of k is equal to 1. The third term of Eq. 2 is to limit the boundary of k. Finally, the discriminator $D(\cdot)$ is to ensure the consistency of original Mars images.

D. NOISE MODEL

1) NOISE EXTRACTION ALGORITHM

In order to model the noise distribution, noise patches need to be extracted from real Mars images. We choose the weak background in the given real Mars images to reduce the impact of the Mars image background. By this way, GAN can focus on learning the distribution of noise, which makes the distribution more accurate.

Inspired by GCBD [37], we adopt a noise extract algorithm as follows. Let p_i denotes an image block with a size of $m \times m$ extracted from a given Mars image with step s_g, q_j^i represents an image block with a size of $n \times n$ extracted from the obtained p_i block with the step s_l. A filtering rule is used to extract noise with mean and variance in a certain range, which can be formulated as follows:

$$|Mean(q_j^i) - Mean(p_i)| \leq \mu \cdot Mean(p_i) \quad (3)$$

and

$$|Var(q_j^i) - Var(p_i)| \leq \gamma \cdot Var(p_i) \quad (4)$$

where $Mean(\cdot)$ and $Var(\cdot)$ denote the function to calculate mean and variance respectively. μ and γ are hyper-parameters, where $\mu, \gamma \in (0, 1)$. For each block $n \times n$ j, if the Eq 3 and Eq 4 are satisfied the condition, p_i will be regarded as a smooth patch and added to the noise bank $S = \{s_1, s_2, \ldots, s_t\}$. Then the noise patches $N = \{n_1, n_2, \ldots, n_t\}$ can be obtained by $n_i = s_i - Mean(s_i)$.

2) ENLARGE NOISE BANK WITH GAN

When input images are not enough, the extracted noise patches might lack of diversity. However, GAN can leverage the great capability of CNN to learn the noise model implicitly and to capture more features of noises without human knowledge of image priors. Therefore, in our method, the GAN is introduced to generate more noise data, and the result is that a larger 'noise bank' is obtained.

We adopt the similar network as DCGAN [36] to generate noise samples for enlarging N, and finally we get a large 'noise bank' $N' = \{n'_1, n'_2, \ldots, n'_t\}$. The network structure is illustrated in Figure 5, the Generator is used to generate noise to enlarge 'noise bank' which consists of a feature extraction

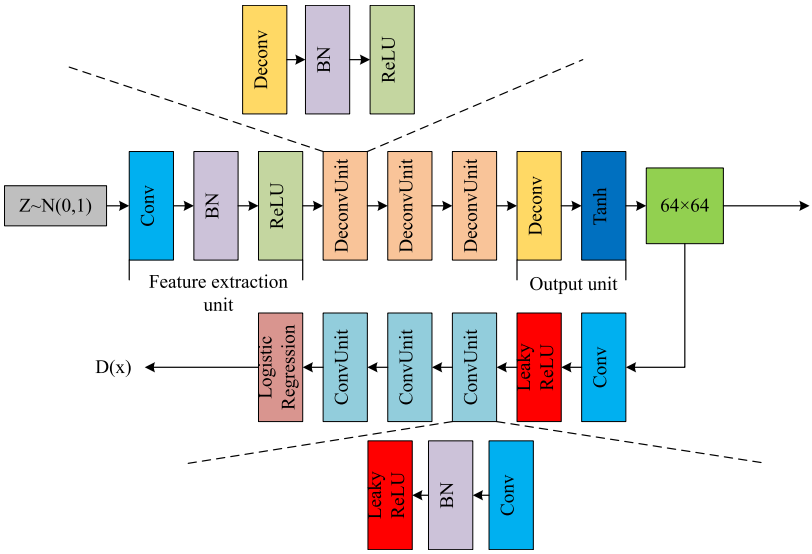

FIGURE 5. The network architecture of the noise model. Generator includes three de-convolution units which can enlarge the receptive field, so that each convolution output contains a larger range of information, while keeping the number of parameters unchanged. Discriminator includes logistic regression at the end of the network to realize two classification. The filter number of the generative network from the second to the last unit is 256, 128, 64, and is equal to the output channel number respectively. The filter number of the discriminator network from the first to the fourth unit is 64, 128, 256, and 512 respectively.

unit, three de-convolution units and an output unit. The feature extraction unit including a 5×5 convolution, a BN [46] layer and a ReLU [47] activate function. De-convolution units including a 5×5 fractionally-strided convolution, a BN layer and a ReLU activate function. Output unit including a 5×5 fractionally-strided convolution and a Tanh activate function. The Discriminator is try to distinguish real noise from those noise generated by Generator. The structure of the discriminator is composed of a 5×5 convolution layer, a convolution unit and an logistic regression as output. The convolution unit including a 5×5 convolution, a BN layer and a ReLU activate function.

E. SUPER-RESOLUTION MODEL

Following ESRGAN [10], we implement an SR network and train it on constructed paired HR-LR data. Specifically, the SR network contains several RRDB [10] blocks and the BN [46] layers are removed. In addition, some advanced techniques including residual scaling and smaller initialization are used to facilitate the training of a deeper model. Moreover, relativistic average GAN is used to improve the performance of the discriminator, and it is designed to 'discriminate whether an image is more real than another' instead of 'an image is real or fake'. This improvement helps the generator to recover more realistic texture details. Furthermore, the perceptual loss is proposed on the basis of SRGAN, and the VGG feature is used before the activation layer instead of after the activation layer, which can generate clearer edge and a good visual results. To sum up, the final loss function of the proposed SR network is the weight sum of three losses:

$$L_{total} = \lambda_1 \cdot L_1 + \lambda_{per} \cdot L_{per} + \lambda_{adv} \cdot L_{adv} \tag{5}$$

whereL_1, L_{per} and L_{adv} denote pixel loss, perceptual loss and adversarial loss respectively. λ_1, λ_{per} and λ_{adv} are set as 0.01, 1, 0.005 empirically.

IV. EXPERIMENTS

In this section, we conduct experiments to validate our proposed framework on Mars32k. First, we analyze kernel estimation accuracy and compare with two common blur kernel with our proposed method. Then, we feed the LR-HR pairs into SR network to analyze the SR performance. Finally, we conduct some ablation studies to verify the effectiveness of degradation framework and each component in our pipeline.

A. DATASETS AND EVALUATION METRICS

1) Mars32k

Mars32k dataset consists of about 32,000 color images collected by the Curiosity rover on Mars between August 2012 and November 2018. The images show various geographical and geological features of Mars such as mountains and valleys, craters, dunes and rocky terrain. All images have been scaled down using linear interpolation to 560×500 pixels (some images have been cropped). The dataset is intended for unsupervised learning and the images are only labeled with the date they are taken on. This dataset only contains photos taken with Curiosity's Mastcam camera and all gray scale or other images are removed.

2) EVALUATION METRICS

We choose PSNR and SSIM to evaluate the results of our method. PSNR and SSIM are commonly-used evaluation metrics for image super-resolution. PSNR is mainly an

evaluation metric to measure the difference between pixels of two images. SSIM is an evaluation index obtained by weighted multiplication of three relatively independent attributes of brightness, contrast and structure. Since SSIM is more in line with visual perception, SSIM is also widely used. The bigger the PSNR and SSIM are, the closer the generated images is to the ground truth.

B. IMPLEMENTATION DETAILS

1) KERNEL ESTIMATION

We initialize learning rate for Generator and Discriminator with $2e^{-4}$ and train the networks with 3000 iterations. In kernel post processing, we set small filtering values of the kernel to 40.

2) NOISE MODEL

In noise extraction algorithm, we set m, s_g to 64, 32 respectively, n, s_l is 16, 16 respectively. μ and *gamma* is 0.1 and 0.25. For enlarging noise bank with GAN, we follow the parameter settings in DCGAN [36].

3) SUPER-RESOLUTION MODEL

The entire super-resolution network is been trained for 3000 generations, using the Adam optimizer [49] ($\beta_1 = 0.5$, $\beta_2 = 0.999$), the learning rate is set to $2e^{-4}$, and it decreases by 0.1 times every 750 generations. Our system is implemented in PyTorch on an NVIDIA RTX3090 GPU.

C. BLUR KERNEL ESTIMATION

To verify the effectiveness of the kernel estimation model, we first show the kernel estimation results compared with ESRGAN [10] in Figure 6. It can be seen from Figure 6 that the blur kernel images obtained through learning strategy in our method is quite different from the Bicubic kernel, and it is similar to the blur kernel of real images.

Furthermore, we also show the quantitative results of kernel estimation, and compare with SRGAN [2] and ESRGAN [10]. Similar to our method, SRGAN [2] and ESRGAN [10] can receive the down-sampled LR image using the blur kernel as inputs, but three different blur kernels are used as down-sampling kernels: Gaussian kernel, Bicubic interpolation kernel and the kernel that obtained by learning strategy, respectively.

Table 1 shows that in the ESRGAN [10] network, using blur kernel estimation is 10dB higher than the Gauss kernel and 5dB higher than the Bicubic interpolation kernel. Moreover, the results show that using the blur kernel estimation is much better than using other kernels for image super-resolution.

D. RESULTS ON MARS IMAGES SUPER-RESOLUTION

In order to verify the effectiveness of the degenerate framework, the method proposed in this paper is compared with the some classic methods including EDSR [5], ESRGAN [10] and ZSSR [50]. We verify our proposed method on Mars32k

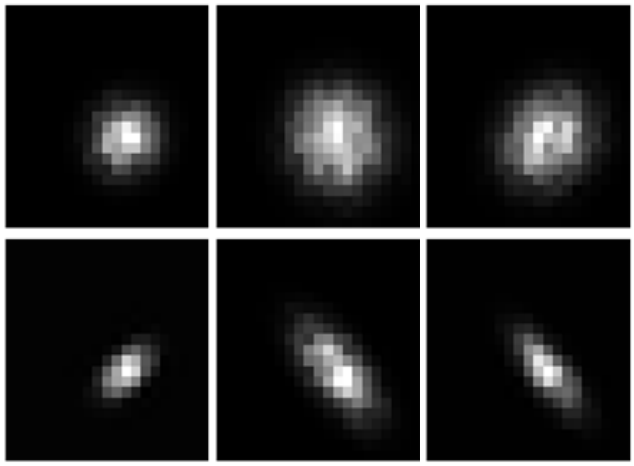

FIGURE 6. SR kernel estimation results.

TABLE 1. Results of Blur-Kernel Estimation. The first two lines are LR images obtained by two traditional methods, the third line is that LR images obtained by blur kernel estimation (Ours) and then the LR images are used as input for training.

Blur-Kernel	Scale	SRGAN [2]	ESRGAN [10]
Gauss	×4	15.12	15.46
Bicubic	×4	18.45	20.225
Ours	×4	**23.61**	**25.52**

verification sets, and the visualization results are shown and the evaluation indicators PSNR and SSIM are calculated in Figure 7 and Table 2, respectively.

Figure 7 shows the comparison of visualization results between our proposed method and other methods. In Figure 7, we find our method achieves the best performance. Compared to EDSR [5] and ESRGAN [10], our method reconstructs Mars image with less artifact and the color is close to ground truth. Compared our proposed method with ZSSR [50], although ZSSR [50] achieve a good result, our method has rich texture details and less noise.

For quantitative results, as show in Table 2, our proposed method also achieves the best performance in both $\times 2$ and $\times 4$ scales. Compared with GAN-based method like ESRGAN [10], by using kernel estimation and noise model, our method has a large margin improvement, *i.e.* 2.98dB, 5.22dB improvement in PSNR for scales $\times 2$, $\times 4$ respectively. Also, our method gets better performance than ZSSR [50] and EDSR [5], because we use perceptual loss so that the SR images are more in line with human perception.

E. ABLATION STUDY

1) USE BICUBIC INTERPOLATION ONLY

In this case, the LR images is down-sampled by bicubic interpolation, and then directly use these HR-LR image pairs for training the super-resolution network, *i.e.* without using kernel estimation and noise model, other parameters are set as default, and the results are shown in Figure 8. Such conditions

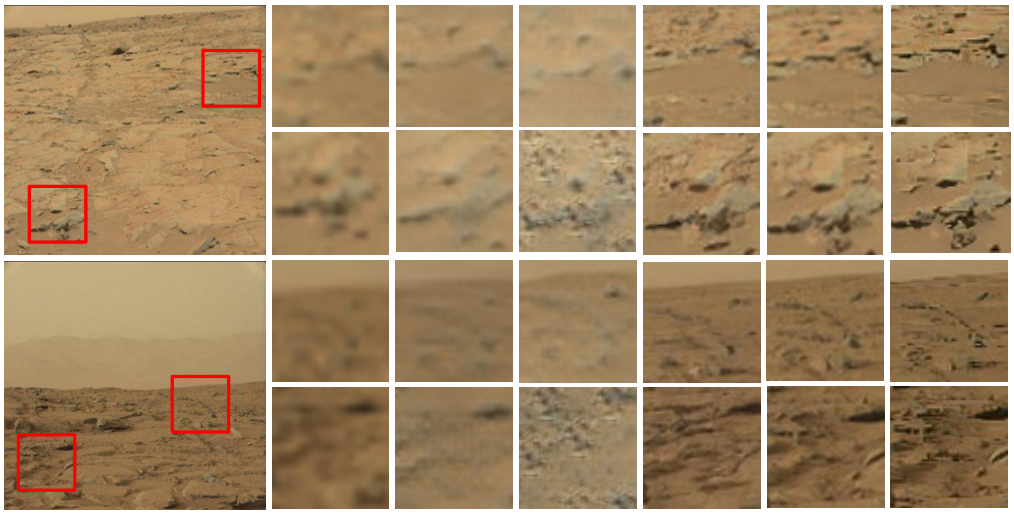

FIGURE 7. Qualitative super-resolution results on Mars32k dataset. We compare our proposed method with EDSR [5], ESRGAN [10], ZSSR [50].

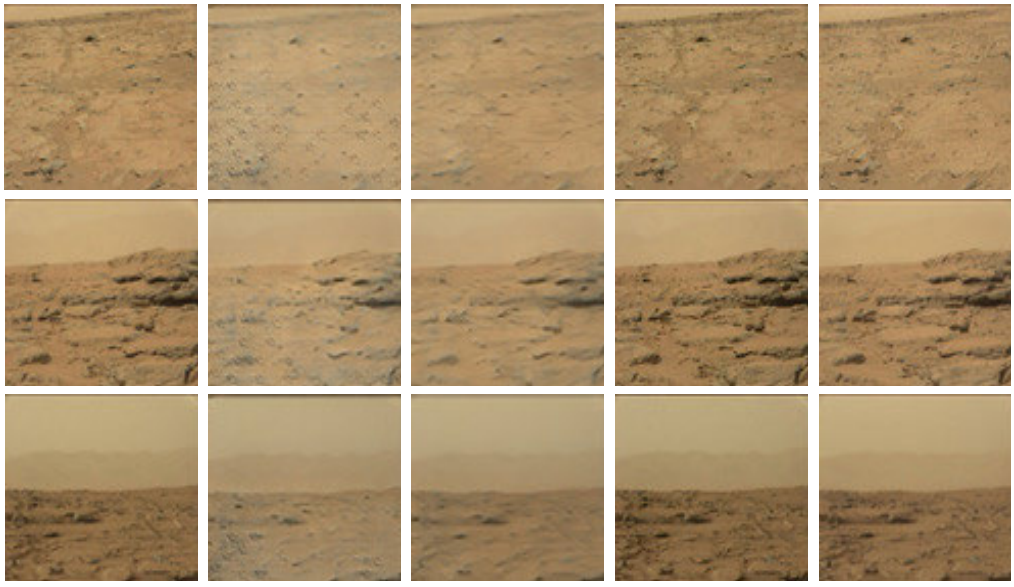

FIGURE 8. Comparison results on Mars32k dataset. We compare our method (Ours) with 'Bicubic', 'No-kernel', 'No-noise' and 'GT'.

TABLE 2. Quantitative results for super-resolution on Mars32k. Compared our proposed method with EDSR [5], ESRGAN [10], ZSSR [50] in PSNR/SSIM and we provide both ×2 and ×4 scales in SR performance.

Method/Scale	×2	×4
EDSR [5]	26.03/0.49	23.06/0.42
ESRGAN [10]	23.26/0.58	20.22/0.51
ZSSR [50]	25.52/0.62	23.58/0.58
Ours	**26.24/0.72**	**25.74/0.628**

are equal to using the ESRGAN [10] for image super-resolution. From the experiment results in the second column of Figure 8, it can be found that the evaluation indicators and visual effects are bad, the PSNR is only 20.225dB, and the SSIM is 0.512dB. At the same time, the sand grains in the Mars image have not been reconstructed, resulting in circular pits and the color of the super-resolution image becomes lighter, which is quite different from the real Mars images.

2) WITHOUT BLUR KERNEL ESTIMATION

In this case, noise is only added during the down-sampling process. Because blur kernel estimation is not used, this experiment can verify the effectiveness of the blur kernel estimation. From Figure 8, we can find that when using the blur kernel estimation, the PSNR 2.047dB and the SSIM 0.045dB are higher than that without blur kernel estimation (as shown in (b) and (d)), which proves that the blur kernel estimation is very important for Mars image super-resolution.

From the above experiment, it is concluded that in the down-sampling process, the LR images obtained by using the blur kernel generated from GAN is better than the direct bicubic interpolation. Furthermore, by adding noise, the color of the image is improved compared to only using bicubic interpolation. And by adding noise, the color of the image is improved compared with that of only using bicubic interpolation.

In order to further verify the influence of blur kernel estimation and noise model in the down-sampling process, we conduct ablation studies on the Mars32k dataset.

3) WITHOUT NOISE MODEL

In this case, only blur kernel estimation is used in the down-sampling process and without adding noise, so this experiment can verify the effect of the noise model. From the experiment results in the fourth column(c) of Figure 8, the PSNR value increases by 0.198dB, therefore, it is necessary to add noise during the down-sampling process.

The complete experimental results in the last column of Figure 8 show that the blur kernel obtained by GAN, and then feed the achieved HR-LR image pairs to ESRGAN [10] for super-resolution. The texture details and colors of the Mars images are very close to the real image, and the evaluation indicators of PSNR and SSIM have also reach the highest compared with other methods.

V. CONCLUSION

In this paper, we analyze the difference between Mars image and earth image firstly. Since the LR images obtained by the traditional bicubic interpolation method does not conform to the down-sampling process of the real images, we propose a new degradation framework for Mars image super-resolution which includes kernel estimation and noise model. By using this degradation framework, we acquire more realistic low-resolution Mars images. Then, combining the obtained low-resolution images with the corresponding high-resolution images to form LR-HR pairs, and finally feed them to super-resolution network. Experiment result shows that our proposed method has a better performance than other classic methods. We also conduct ablation experiments to prove that our proposed components are helpful for the Mars image super-resolution.

ACKNOWLEDGMENT

(Cong Wang, Yin Zang, and Yongqiang Zhang contributed equally to this work.)

REFERENCES

[1] 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.

[2] C. Ledig, L. Theis, F. Huszar, 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.

[3] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang, "FSRNet: End-to-end learning face super-resolution with facial priors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2492–2501.

[4] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1637–1645.

[5] 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.

[6] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1664–1673.

[7] Y. Zhang, M. Ding, Y. Bai, M. Xu, and B. Ghanem, "Beyond weakly supervised: Pseudo ground truths mining for missing bounding-boxes object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 4, pp. 983–997, Apr. 2020.

[8] Y. Zhang, Y. Bai, M. Ding, Y. Li, and B. Ghanem, "W2F: A weakly-supervised to fully-supervised framework for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 928–936.

[9] Y. Zhang, Y. Bai, M. Ding, Y. Li, and B. Ghanem, "Weakly-supervised object detection via mining pseudo ground truth bounding-boxes," *Pattern Recognit.*, vol. 84, pp. 68–81, Dec. 2018.

[10] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, Sep. 2018, pp. 63–79.

[11] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1256–1272, Jun. 2017.

[12] U. Schmidt, J. Jancsary, S. Nowozin, S. Roth, and C. Rother, "Cascades of regression tree fields for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 4, pp. 677–689, Apr. 2016.

[13] U. Schmidt and S. Roth, "Shrinkage fields for effective image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2774–2781.

[14] T. N. Harrison, G. R. Osinski, L. L. Tornabene, and E. Jones, "Global documentation of gullies with the Mars reconnaissance orbiter context camera and implications for their formation," *Icarus*, vol. 252, pp. 236–254, May 2015.

[15] J. E. Graf, R. W. Zurek, J. K. Erickson, B. Jai, M. D. Johnston, and R. de Paula, "Status of Mars reconnaissance orbiter mission," *Acta Astronautica*, vol. 61, nos. 1–6, pp. 44–51, Jun. 2007.

[16] R. Schmidt, J. Credland, A. Chicarro, and P. Moulinier, "ESA's Mars express mission—Europe on its way to Mars," *ESA Bull.-European Space Agency ESA BULL-EUR Space Agency*, vol. 98, pp. 56–66, Jun. 1999.

[17] G. Neukum and R. Jaumann, "HRSC co-investigator team HRSC: The high resolution stereo camera of Mars express," *Mars Exp., Sci. Payload. ESA SP*, vol. 1240, pp. 17–36, Jan. 2004.

[18] R. Tsai and T. Huang, "Multiframe image restoration and registration," *Adv. Comput. Vis. Image Process*, vol. 1, pp. 317–339, Jan. 1984.

[19] S. Kim, N. Bose, and H. Valenzuela, "Reconstruction of high resolution image from noise undersampled frames," in *Advances in Communications and Signal Processing*. Berlin, Germany: Springer, 1989, pp. 315–326.

[20] C.-Y. Yang and M.-H. Yang, "Fast direct super-resolution by simple functions," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 561–568.

[21] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2004, pp. 275–282.

[22] 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.

[23] M. Lebrun, M. Colom, and J.-M. Morel, "Multiscale image blind denoising," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3149–3161, Oct. 2015.

[24] C. Liu, R. Szeliski, S. B. Kang, C. L. Zitnick, and W. T. Freeman, "Automatic estimation and removal of noise from a single image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 299–314, Feb. 2008.

[25] Y. Zhang, M. Ding, Y. Bai, D. Liu, and B. Ghanem, "Learning a strong detector for action localization in videos," *Pattern Recognit. Lett.*, vol. 128, pp. 407–413, Dec. 2019.

[26] T. Rabie, "Robust estimation approach for blind denoising," *IEEE Trans. image Process.*, vol. 14, no. 11, pp. 1755–1765, Nov. 2005.

[27] F. Zhu, G. Chen, and P. A. Heng, "From noise modeling to blind image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 420–429.

[28] A. B. Baral and M. Torlak, "Impact of number of noise eigenvectors used on the resolution probability of MUSIC," *IEEE Access*, vol. 7, pp. 20023–20039, 2019.

[29] M. Lebrun, A. Buades, and J. M. Morel, "A nonlocal Bayesian image denoising algorithm," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1665–1688, Jan. 2013.
[30] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," Jun. 2014, *arXiv:1406.2661*. [Online]. Available: https://arxiv.org/abs/1406.2661
[31] Y. Zhang, M. Ding, Y. Bai, and B. Ghanem, "Detecting small faces in the wild based on generative adversarial network and contextual information," *Pattern Recognit.*, vol. 94, pp. 74–86, Oct. 2019.
[32] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, "SOD-MTGAN: Small object detection via multi-task generative adversarial network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 206–221.
[33] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, "Finding tiny faces in the wild with generative adversarial network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 21–30.
[34] Y. Wu, L. Lan, H. Long, G. Kong, X. Duan, and C. Xu, "Image super-resolution reconstruction based on a generative adversarial network," *IEEE Access*, vol. 8, pp. 215133–215144, 2020.
[35] P. Shamsolmoali, M. Zareapoor, E. Granger, H. Zhou, R. Wang, M. E. Celebi, and J. Yang, "Image synthesis with adversarial networks: A comprehensive survey and case studies," *Inf. Fusion*, vol. 72, pp. 126–146, Aug. 2021.
[36] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: http://arxiv.org/abs/1511.06434
[37] J. Chen, J. Chen, H. Chao, and M. Yang, "Image blind denoising with generative adversarial network based noise modeling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3155–3164.
[38] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2223–2232.
[39] M. Adil, S. Mamoon, A. Zakir, M. A. Manzoor, and Z. Lian, "Multi scale-adaptive super-resolution person re-identification using GAN," *IEEE Access*, vol. 8, pp. 177351–177362, 2020.
[40] O.-Y. Lee, Y.-H. Shin, and J.-O. Kim, "Multi-perspective discriminators-based generative adversarial network for image super resolution," *IEEE Access*, vol. 7, pp. 136496–136510, 2019.
[41] J. Adler and S. Lunz, "Banach Wasserstein GAN," 2018, *arXiv:1806.06621*. [Online]. Available: http://arxiv.org/abs/1806.06621
[42] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," 2017, *arXiv:1704.00028*. [Online]. Available: http://arxiv.org/abs/1704.00028
[43] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2006, pp. 404–417.
[44] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[45] S. Bell-Kligler, A. Shocher, and M. Irani, "Blind super-resolution kernel estimation using an internal-GAN," 2019, *arXiv:1909.06581*. [Online]. Available: http://arxiv.org/abs/1909.06581
[46] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[47] K. Jarrett, K. Kavukcuoglu, M. A. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?" in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2146–2153.
[48] X. Yin, J. Goudriaan, E. A. Lantinga, J. Vos, and H. J. Spiertz, "A flexible sigmoid function of determinate growth," *Ann. Botany*, vol. 91, no. 3, pp. 361–371, 2003.
[49] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, Dec. 2014, pp. 1–15.
[50] A. Shocher, N. Cohen, and M. Irani, "Zero-shot super-resolution using deep internal learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3118–3126.

CONG WANG is currently pursuing the Ph.D. degree in instrument science and technology with HIT. He is a Professor at Shanghai Institute of Satellite Engineering. His research interests include computer vision, machine learning, and deep learning, mainly weakly/fully-supervised object detection, image super-resolution, and image generation.

YIN ZHANG received the M.S. degree in instrument science and technology from Harbin Institute of Technology (HIT), Harbin, China, in 2021, where he is currently pursuing the Ph.D. degree. His research interests include computer vision, pattern recognition, machine learning, and deep learning, mainly image super-resolution, image denoising, and generative adversarial networks.

YONGQIANG ZHANG received the M.S. and Ph.D. degrees in instrument science and technology from Harbin Institute of Technology (HIT), Harbin, China, in 2015 and 2020, respectively. From 2017 to 2018, he worked as a Visiting Student at King Abdullah University of Science and Technology (KAUST). He is currently an Assistant Professor with the School of Instrumentation Science and Engineering, HIT. His research interests include computer vision, pattern recognition, machine learning, and deep learning, mainly face detection, weakly/fully-supervised object detection, activity detection, and image and video understanding in the real-world.

RUI TIAN received the bachelor's degree from Harbin University of Science and Technology, in 2021, where he is currently pursuing the master's degree. His research interests include computer vision, machine learning, deep learning, and spacecraft measurement and control technology.

MINGLI DING received the B.S., M.S., and Ph.D. degrees in instrument science and technology from Harbin Institute of Technology (HIT), Harbin, China, in 1996, 1997, and 2001, respectively. From 2009 to 2010, he worked as a visiting scholar in France. He is currently a Professor with the School of Electrical Engineering and Automation, HIT. He has published over 40 papers in peer-reviewed journals and conferences. His research interests include intelligence tests and information processing, automation test technology, computer vision, and machine learning.