IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 5, 2020, accepted March 14, 2020, date of publication March 18, 2020, date of current version April 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2981726

Adaptive Diagonal Total-Variation Generative Adversarial Network for Super-Resolution Imaging

ZHANG SAN-YOU[1,2], CHENG DE-QIANG[1], JIANG DAI-HONG[3], KOU QI-QI[1], AND MA LU[1]

[1]School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221000, China
[2]Department of Science and Technology, Suzhou Wujiang District Public Security Bureau, Suzhou 215200, China
[3]Key Laboratory of Intelligent Industrial Control Technology of Jiangsu Province, Information and Electrical Engineering College, Xuzhou University of Technology, Xuzhou 221000, China

Corresponding author: Cheng De-Qiang (chengdq@cumt.edu.cn)

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

ABSTRACT To address problems that the loss function does not correlate well with perceptual vision in super-resolution methods based on the convolutional neural network(CNN), a novel model called the ADTV-SRGAN is designed based on the adaptive diagonal total-variation generative adversarial network. Combined with global perception and the local structure adaptive method, spatial loss based on the diagonal variation model is proposed to make the loss function can be adjusted according to the spatial features. Pixel loss and characteristic loss are in combination with the spatial loss for the fusing optimization of the total loss function such that high-frequency details of the images are maintained to improve their quality. The results of experiment show that the proposed method can obtain competitive results in objective evaluations. In subjective assessment, images reconstructed by it are clear, delicate, and natural, and it preserved edge- and texture-related details.

INDEX TERMS Generative adversarial network, super-resolution imaging, image reconstruction, total variation, loss function.

I. INTRODUCTION

Methods to reconstruct super-resolution images can be classified into five categories. The first consists of interpolation methods [1] that estimate the pixel values of interpolation points by using information on the neighborhood of known pixel points. The cons of interpolation methods is that although they have very less complexity, however, poor selection of neighboring pixels for interpolation often leads to the creation of artifacts [2]. The second consists of two-step upscaling methods [3] that apply simple interpolation methods in first stage to upscale the image followed by edge refinement to reduce the artifacts created due to the first stage interpolation. The issue with such methods is that they often over-smooth the edges during refinement process. Especially, for interpolation with large scale, they convert the pointy edges into curves. The third consists of reconstruction methods [4] that establish an observation model and solve it using the inverse process to implement image reconstruction. Because the degradation of images is complex and diverse, it is difficult to define for humans to comprehensively define the observation model. With the increase in the magnification of images, the effects of restoring them are not ideal. The fourth class of methods to reconstruct super-resolution images is exemplified in the proposal in [5], which can extract the non-linear mapping between LR-images and HR-images. However, the effect of such methods remains poor in terms of the magnification factor and complex scenes, and is further limited by their poor capabilities of extraction and representation. The fifth class consists of deep learning methods [6] that can compensate for the lack of representation ability of shallow learning, has better generalization ability, and can better deal with complex image features than other methods. Deep learning methods based CNN are first introduced

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

in super-resolution field. The pioneer study is proposed by Dong *et al.* [7] which is named as SRCNN. It uses only three layers of the CNN to fit the non-linear mapping and feature transformation. To accelerate the model, Dong *et al.* [8] propose the FSRCNN model, which uses a small convolution layer to replace the large one in SRCNN. A lot of effective improvements over the CNN-based methods have been proposed to raise the performance of super-resolution ever since [9], [10]. However, those methods commonly encounter problems concerning image smoothing, step effects, insufficient expressive accuracy of textural features, and distortion in case of high magnification.

With the emergence of the generative adversarial network (GAN) [11], its powerful capability to generate realistic texture provides a new solution for super-resolution imaging. To overcome those problems in CNN-based methods and further improve the perceptual quality of the reconstructed image, a novel model called the ADTV-SRGAN is designed based on the adaptive diagonal total-variation generative adversarial network. It employs pixel loss, characteristic loss, spatial loss, and adversarial loss to recreate realistic details and avoid the phenomenon whereby the use of mean square error(MSE) based loss function leads to the excessive smoothing of image texture. The main contributions of this paper are as follows:

1. A multi-loss ensemble is used to produce visually satisfactory super-resolution results with the combination of pixel loss, characteristic loss, spatial loss, and adversarial loss, which can obtain a balance between objective evaluation and perceptual quality.

2. A new adaptive model based on diagonal total variation is proposed to keep high-frequency texture details and achieve better results.

3. A novel strategy is introduced to improve the performance of super-resolution based on GAN coupled with the total-variation based model.

II. RELATED WORK

At present many CNN-based methods have been proposed by researchers to solve super-resolution problems. Shi *et al.* [12] propose ESPCN-based super-resolution method that uses sub-pixel convolution to extract features directly. Aiming at problems as slow convergence and inability to perform multi-scale tasks, Kim *et al.* [13] propose the VDSR-based super-resolution method that uses a residual network ResNet to train a deeper super-resolution network model to achieve higher accuracy. In light of problems in sample learning in complex mapping, Lai *et al.* [14] propose LapSRN-based super-resolution method combining anti-convolution and residual learning. It uses the Laplacian pyramidal structure and hierarchical upper-sampling method to complete high multiple learning, using two-times sampling each time for super-resolution imaging through gradual sampling and step-by-step prediction of the residuals. It achieves a good perceptual reconstruction effect. Chen *et al.* [15] propose a GuideAE-based method for super-resolution imaging that combines a statistical model and the auto-encoder network to restore the image.

The above methods mostly use MSE-based loss for training to obtain a high value of quantitative metrics. While MSE-based loss function can be easy to optimize, it often fails to target a diversity of image features and struggles to accurately restore image detail [16]–[19]. As a result, the super-resolution image obtained is poor. In particular at high magnification, visual perception is prone to distortion. To overcome such issues, GAN for image super-resolution has attracted more and more attentions. Recently, GAN-based super-resolution methods such as SRGAN [20], EnhanceNet [21], PESRGAN and ISRGAN [23] have been proposed for generating better perceptual quality. Ledig *et al.* [20] introduce GAN into super-resolution for the first time, and proposed SRGAN method. SRGAN focuses on the use of adversarial and perceptual loss to enhance a realistic texture of super-resolution images that are consistent with our understanding of visual perception. Sajjadi *et al.* [21] propose EnhanceNet by using GAN with the concept of fully covolutional neural networks in the adversarial training. EnhanceNet introduces an additional texture loss in combination with adversarial and perceptual loss for better realistic textures. Vu *et al.* [22] propose PESRGAN by using a relativistic discriminator in the adversarial training. PESRGAN uses relativistic loss function in combination with content loss and total variance loss to improve the super-resolution quality. Chudasama and Ulpa [23] propose ISRGAN by using GAN with the concept of densely connected deep convolutional networks to recover the high-frequency texture details. Instead of relying only on MSE-based loss, ISRGAN is trained by the combination of of VGG based perceptual loss and adversarial loss. To further improve the quality of super-resolution images, inspired by adaptive total variation [24], [25] and diagonal total variation [26], [27] model to take full advantage of directional information of edges and textures, a novel strategy is introduced to improve the performance of super-resolution based on GAN coupled with the total-variation based model. Furthermore, a new adaptive model based on diagonal total variation is proposed to keep texture details. Combined with pixel loss characteristic loss and adversarial loss, the spatial loss based on the diagonal total-variation model is introduced to optimize the loss function adaptively adjusted according to the spatial features. The multi-loss ensemble can help the ADTV-SRGAN model preserve more high-frequency details and achieve better qualitative and quantitative super-resolution
performance.

III. ADTV-SRGAN METHOD

The key framework of ADTV-SRGAN consists of two models, the generative model G for generating high-resolution images to fool the discriminative model, and the discriminative model D that identifies whether the input images are produced by the generative model or obtained from

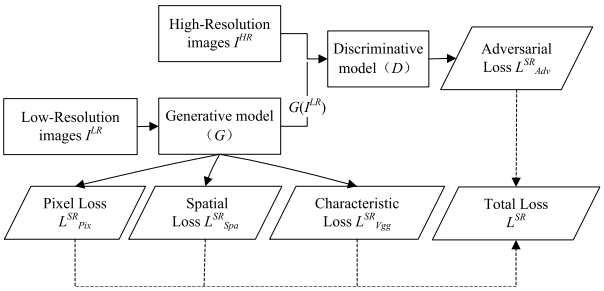


FIGURE 1. The schematic process of training the ADTV-SRGAN.

high-resolution samples. The adversarial learning is finished until the discriminative model cannot distinguish between generated image and the real sample. The architecture of ADTV-SRGAN is derived from SRGAN, but the loss function is different from it, and this is also the major improvement. The generative and discriminative network of ADTV-SRGAN are the same structure as that of the SRGAN except the loss function. As shown in Figure 1, ADTV-SRGAN is trained with a multi-loss ensemble with the combination of pixel loss L^{SR}_{Pix}, characteristic loss L^{SR}_{Vgg}, spatial loss L^{SR}_{Spa}, and adversarial loss L^{SR}_{Adv}, which are described in details separately in the following subsections.

The goal of optimizing the generative model G is to enable the generated image $G(I^{LR})$ to approach the HR-image I^{HR}, and that of the discriminative model D is to distinguish it from I^{HR}. The optimization is a process of a minimax game involving the two models, where the function $V(D, G)$ is as follows:

$$V(D,G) = \min_{\theta_G}\max_{\theta_D} E_{I^{HR}\sim p_{train}(I^{HR})}[\log D_{\theta_D}(I^{HR})] + E_{I^{LR}\sim p_G(I^{LR})}[\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})))] \quad (1)$$

where θ_G and θ_D are the parameters of the networks of the generative model and discriminative model, respectively. To train the SR-image I^{SR} using the LR-image I^{LR}, it is necessary to solve for the optimal generation model parameters $\hat{\theta}_G$ as follows:

$$\hat{\theta}_G = \arg\min_{\theta_G}\frac{1}{N}\sum_{n=1}^{N} L^{SR}\left(G_{\theta_G}\left(I_n^{LR}\right), I_n^{HR}\right) \quad (2)$$

where L^{SR} is total loss that determines whether the generative model can generate super-resolution images similar to the real samples. It contains pixel loss, characteristic loss, adversarial loss, and spatial loss. To retain textural details and ensure that the generative image is more real, this paper introduces spatial loss L^{SR}_{Spa} based on the adaptive diagonal total-variation model, so that it can be adjusted according to the spatial structure of the image. The new pixel loss L^{SR}_{Pix} and characteristic loss L^{SR}_{Vgg} are used to implement the fusion optimization of total loss L^{SR}:

$$L^{SR} = L^{SR}_{Pix} + L^{SR}_{Vgg} + L^{SR}_{Adv} + L^{SR}_{Spa} \quad (3)$$

A. PIXEL LOSS

Pixel loss is used to assess the consistency of content between the HR-image and the real sample. Traditional CNN-based methods of reconstruction mostly use pixel loss based on the MSE, which causes the reconstructed image to be too smooth and appear unrealistic. Lai *et al.* [14] proposed the Charbonnier loss which is a differentiable variant of L1 loss [28], [29]. To improve such details as the edges and contours of the reconstructed image, pixel loss uses Charbonnier loss to replace the MSE-based content loss function in SRGAN. The pixel loss L^{SR}_{Pix} is calculated as follows:

$$L^{SR}_{Pix} = \frac{1}{N}\sum_{n=1}^{N}\sqrt{\left(G_{\theta_G}\left(I_n^{LR}\right) - I_n^{HR}\right)^2 + \varepsilon^2} \quad (4)$$

where $G_{\theta G}(I_n^{LR})$ is the generated HR-image and I_n^{HR} the real sample. $\varepsilon(0 < \varepsilon < 1)$ is the constant term of the Charbonnier penalty. Following approach [14], ε is empirically set to be 0.001 in this paper.

B. CHARACTERISTIC LOSS

Characteristic loss is used to assess semantic similarity between the HR-image and the real sample. Johnson *et al.* [17] propose the perceptual loss to provide edges preserved possible solutions. Most traditional deep learning-based methods of reconstruction use layer 4 of the pre-trained VGG [17] network to extract feature maps. Features extracted by the VGG-16, which is an image classification network, can help retain the contours of edges of the image, but the effect of the reconstruction of local textural details is not realistic by adopting low-level feature maps. In accordance with recent approaches [18], [19], [30], [31], it can obtain better texture details using the high-level VGG perceptual features compared to low-level perceptual features. To extract more hierarchical semantic characteristics and recover closer textures to the original high-resolution image, following approach [31], this paper uses layer 13 of the pre-trained VGG-16 network, and calculates the Euclidean distance of high-level features as characteristic loss, which is more global and invariant. The characteristic loss L^{SR}_{Vgg} is calculated as follows:

$$L^{SR}_{Vgg} = \frac{1}{S_{i,j}H_{i,j}}\sum_{x=1}^{S_{i,j}}\sum_{y=1}^{H_{i,j}}\left(\phi_{i,j}\left(I^{HR}\right)_{x,y} - \phi_{i,j}\left(G_{\theta_G}\left(I^{LR}\right)\right)_{x,y}\right)^2 \quad (5)$$

where $\phi_{i,j}$ is the feature mapped to the j-th convolution layer in front of the i-th pooling layer. $S_{i,j}$ and $H_{i,j}$ respectively represent the length and width of the feature map.

C. ADVERSARIAL LOSS

Adversarial loss [20] represents the probability that the generated HR-image is a real sample given by the discriminative

model. Adversarial loss L_{Adv}^{SR} is the cross-entropy:

$$L_{Adv}^{SR} = \sum_{n=1}^{N} -\log D_{\theta_D}\left(G_{\theta_G}\left(I^{LR}\right)\right) \tag{6}$$

where $D_{\theta_D}(\cdot)$ is the probability that an image is a real sample and $G_{\theta_G}\left(I^{LR}\right)$ is the generated HR-image.

D. SPATIAL LOSS

Spatial loss is based on the total-variation model [32], [33] to reconstruct an image that maintains features of the edges and texture to the greatest extent. The total-variation model is a classic model in image restoration that exhibits good edge retention characteristics [24]–[27]. Details of the image can be preserved while smoothing. The total-variation model is as follows:

$$T(x) = \sum_{i=1}^{H}\sum_{j=1}^{W}\left|\nabla_{i,j}x\right| = \sum_{i=1}^{H}\sum_{j=1}^{W}\sqrt{\left(\nabla_{i,j}^{p}\right)^2 + \left(\nabla_{i,j}^{q}\right)^2} \tag{7}$$

where W and H are image width and height, respectively. $\nabla_{i,j}^{p}$ and $\nabla_{i,j}^{q}$ represent the gradient of pixel $x_{i,j}$ along the horizontal and vertical sides, respectively, and $\nabla_{i,j}^{p}$ and $\nabla_{i,j}^{q}$ satisfy the following equations:

$$\nabla_{i,j}^{p} = x_{i,j+1} - x_{i,j}, \quad \nabla_{i,j}^{q} = x_{i+1,j} - x_{i,j} \tag{8}$$

Because the total-variation model considers only the gradient of pixels on the two sides, it is easy for the ladder effect to occur. Therefore, combined with the diagonal information of the pixels, the diagonal total-variation model is proposed as follows:

$$\begin{aligned} DT(x) &= \sum_{i=1}^{H}\sum_{j=1}^{W}\left|\nabla_{i,j}x\right| \\ &= \sum_{i=1}^{H}\sum_{j=1}^{W}\sqrt{\left(\nabla_{i,j}^{p}\right)^2 + \left(\nabla_{i,j}^{q}\right)^2 + \left(\nabla_{i,j}^{r}\right)^2 + \left(\nabla_{i,j}^{k}\right)^2} \end{aligned} \tag{9}$$

where $\nabla_{i,j}^{r}$ and $\nabla_{i,j}^{k}$ represent the gradient of pixel $x_{i,j}$ in the two diagonal directions, respectively, and $\nabla_{i,j}^{r}$ and $\nabla_{i,j}^{k}$ satisfy the following equations:

$$\nabla_{i,j}^{r} = x_{i+1,j} - x_{i,j+1}, \nabla_{i,j}^{k} = x_{i+1,j+1} - x_{i,j} \tag{10}$$

The diagonal total-variation model balances the influence on all sides of the pixel points, and can overcome the step effect and protect details of the edges in the image while smoothing it. To enable the spatial loss to control the intensity of constraints on the diagonal total-variation at the pixel points, the indicator of difference curvature [34], [35] is used to distinguish the edge region from the flat region of the image. Spatial loss can thus be adjusted according to the structure of the image. The difference curvature $C_{i,j}$ is defined as follows:

$$C_{i,j} = \left|\left|u_{\eta\eta}\right| - \left|u_{\varepsilon\varepsilon}\right|\right| \tag{11}$$

$$u_{\eta\eta} = \frac{\mu_x^2\mu_{xx} + 2\mu_x\mu_y\mu_{xy} + \mu_y^2\mu_{yy}}{\mu_x^2 + \mu_y^2} \tag{12}$$

$$u_{\varepsilon\varepsilon} = \frac{\mu_y^2\mu_{xx} - 2\mu_x\mu_y\mu_{xy} + \mu_x^2\mu_{yy}}{\mu_x^2 + \mu_y^2} \tag{13}$$

where $u_{\eta\eta}$ and $u_{\varepsilon\varepsilon}$ represent the second derivatives along the directions of the gradient and the vertical gradient, respectively. $|.|$ represents the absolute value, and $\mu_x, \mu_y, \mu_{xy}, \mu_{xx}, \mu_{yy}$ represent the first and second derivatives of the gradient information of pixel points. For the edge region, $\left|u_{\eta\eta}\right|$ is large and $|u_{\varepsilon\varepsilon}|$ is small. $C_{i,j}$ is thus large in the edge region. For the flat region, $\left|u_{\eta\eta}\right|$ and $|u_{\varepsilon\varepsilon}|$ are small. $C_{i,j}$ is thus small in the flat region. The value of $C_{i,j}$ can distinguish regions occupied by edges from flat regions. The difference curvature can be used to build the weight of spatial information $W_{i,j}$ that can be adjusted adaptively and dynamically. $W_{i,j}$ is defined as follows:

$$W_{i,j} = \frac{1}{1 + \beta C_{i,j}} \tag{14}$$

where β is a constant. In areas occupied by edges, the value of $C_{i,j}$ is large, and the adaptive value of the weight of spatial information $W_{i,j}$ is small. Thus, to better maintain details of edges of the image, in flat areas, the value of $C_{i,j}$ is small and the adaptive value of the weight of spatial information $W_{i,j}$ is large. This ensures that the generated HR-image and the real sample do not exhibit large deviations in detail. The adaptive diagonal total-variation model and spatial loss are as follows:

$$\begin{aligned} &ADT(x) \\ &= \sum_{i=1}^{H}\sum_{j=1}^{W}\left|\nabla_{i,j}x\right| \\ &= \sum_{i=1}^{H}\sum_{j=1}^{W} W_{i,j}\sqrt{\left(\nabla_{i,j}^{p}\right)^2 + \left(\nabla_{i,j}^{q}\right)^2 + \left(\nabla_{i,j}^{r}\right)^2 + \left(\nabla_{i,j}^{k}\right)^2} \end{aligned} \tag{15}$$

$$\begin{aligned} &L_{Spa}^{SR} \\ &= ADT(G_{\theta_G}\left(I_{x_{i,j}}^{LR}\right)) \end{aligned} \tag{16}$$

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. TRAINING DETAILS

The experiment uses the framework of TensorFlow GPU. The experimental hardware environment included a Xeon E5-2600 2.1 GHz six-core processor with 32 GB of memory, and an NVIDIA Tesla P4 (8G) graphics card. The experimental software environment included a Ubuntu 16.04 Operating System and the CUDA 8.0 Development Kit. The DIV2K training set [19] is used in this experiment, and consisted of 1,000 images, with 80% used for training, 10% for validation, and the remainder for testing. All these images contain at least 2,040 pixels on the horizontal or the vertical axis. Data enhancement solutions are also provided for the training process, rotated images of the training set clockwise at 0°,

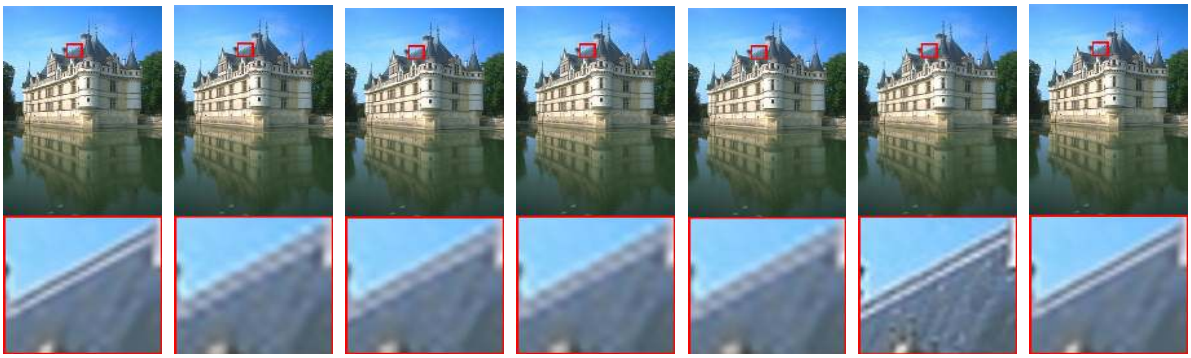


FIGURE 2. PSNR and SSIM of castle.png Reconstructed by different methods (3× upscaling). (a) SRCNN/25.93/0.8363; (b) VDSR/26.96/0.8680; (c) LapSRN/26.95/0.8673; (d) GuideAE/26.98/0.8685; (e) SRGAN/26.91/0.8593; (f) ADTV-SRGAN/26.96/0.8690.

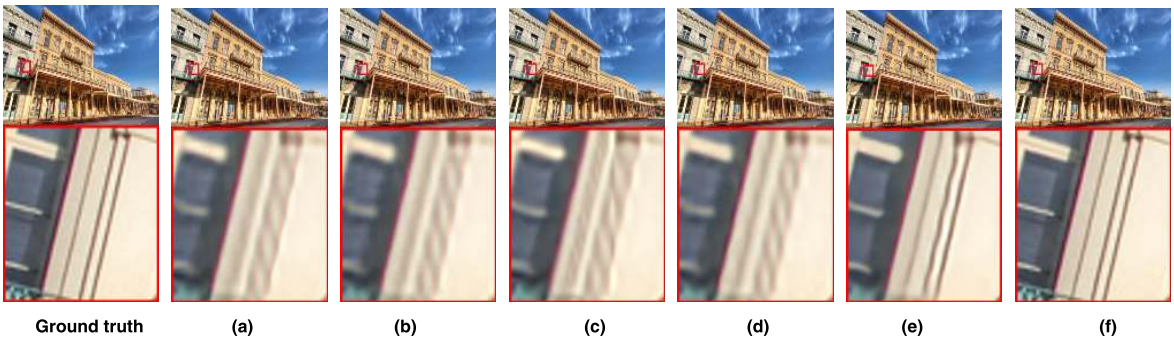


FIGURE 3. PSNR and SSIM of building.png reconstructed by different methods (4× upscaling). (a) SRCNN/25.00/0.7448; (b) VDSR/25.26/0.7568; (c) LapSRN/25.28/0.7571; (d) GuideAE/25.30/0.7574; (e) SRGAN/25.24/0.7570; (f)ADTV-SRGAN/25.27/0.7576.

90°, 180°, and 270° and flipped them horizontally. This enables to increase the number of training images to eight times more than the original. Network training is performed using the Adam optimizer [36], at a learning-rate 0.0001 and 20,000 iterations.

In equation (14), the parameter β is used to control the weight of spatial information W. As shown in Table 1, the effect of β on the super-resolution performance is compared at a magnification factor of $\times 4$ on the set5 dataset.. Different high and low values of β such as 0.01, 0.05, 0.2, 0,1, 0.5, 1 and 2 are used. It is found that a better super-resolution performance in the peak signal-to-noise ratio (PSNR) and structural similarity index measurement (SSIM) can be obtained when β is set at 0.1. Therefore, β is empirically set to be 0.1 in the paper. The PSNR and SSIM are objective indicators of image quality. The PSNR reflects error in the corresponding pixel points between images, where a higher PSNR indicates less distortion. The SSIM reflects the similarity between images, where a higher value indicates that the SR-image is more similar to the HR-image. Methods to calculate the PSNR and SSIM are provided in [37], [38].

TABLE 1. Comparison of the effect of parameter β on the super-resolution performance.

β	0.01	0.05	0.1	0.2	0.5	1	2
PSNR	28.85	30.42	**31.58**	30.92	30.21	29.08	28.21
SSIM	0.8658	0.8795	**0.8921**	0.8893	0.8814	0.8739	0.8607

B. COMPARISON WITH OTHER METHODS

To verify performance, benchmark sets Set5, Set14, BSD100, and Urban100 are used, and the results are compared with many methods to obtain super-resolution images, like the SRCNN [7], VDSR [13], LapSRN [14], GuideAE [15], and SRGAN [20]. The subjective perception of the visual effects are used to judge the performance of the proposed method by examining the characteristics of textural details of different sample images subjected to different methods to reconstruct super-resolution images. Figures 2–4 show a comparison of the effect of image super-resolution on sample images with magnification factors of 3×, 4×, and 8×, respectively. From left to right are the ground HR-image, and images reconstructed by the SRCNN, VDSR, LapSRN, GuideAE, SRGAN, and ADTV-SRGAN. We partially enlarge the roof of the castle, the window of the building, and the frame of the curtain wall in the images. The SRCNN, VDSR, LapSRN, and GuideAE methods based on the CNN are

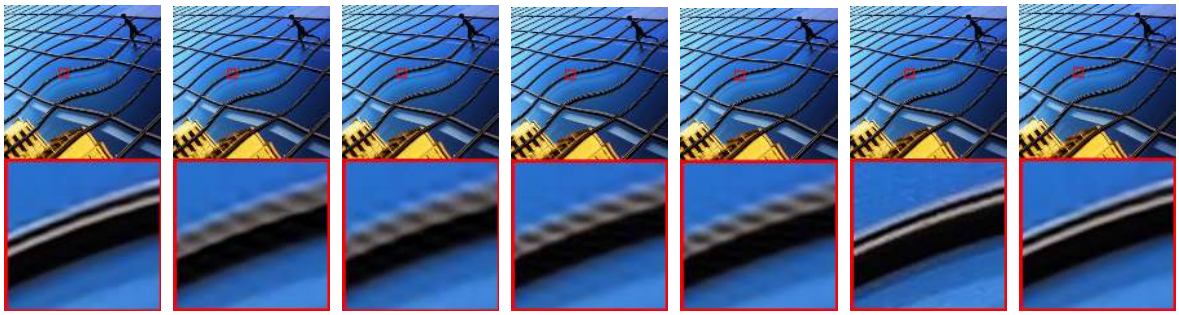


FIGURE 4. PSNR and SSIM of curtain.png reconstructed by different methods (8× upscaling). (a) SRCNN/21.20/0.5432; (b) VDSR/21.68/0.5632; (c) LapSRN/21.78/0.5711; (d) GuideAE/21.81/0.5813; (e) SRGAN/21.70/0.5716; (f) ADTV-SRGAN/21.76/0.5815.

TABLE 2. Comparison of methods in terms of PSNR/SSIM.

Dataset	Scale	Bicubic[1]	A+ [2]	SRCNN[7]	VDSR [13]	LapSRN[14]	GuideAE[15]	SRGAN[20]	ADTV-SRGAN
Set5	3	30.40/0.8686	32.58/0.9088	32.75/0.9090	33.67/0.9210	33.82/0.9227	**33.82**/0.9227	33.73/0.9102	33.80/**0.9229**
	4	28.43/0.8109	30.28/0.8603	30.48/0.8628	31.35/0.8830	31.54/0.8866	31.54/0.8850	30.52/0.8791	**31.58**/**0.8921**
	8	24.39/0.6572	25.52/0.6923	25.33/0.6892	25.93/0.7240	**26.14**/0.7382	26.12/0.7384	25.88/0.7069	26.12/**0.7385**
Set14	3	27.54/0.7741	29.13/0.8188	29.28/0.8209	29.78/0.8320	**29.79**/0.8320	29.87/**0.8321**	29.58/0.8215	29.82/0.8318
	4	26.00/0.7023	27.32/0.7491	27.49/0.7503	28.02/0.7680	28.09/0.7694	**28.19**/0.7720	27.83/0.7603	28.17/**0.7812**
	8	23.19/0.5681	23.98/0.5974	23.85/0.5931	24.26/0.6140	**24.44**/0.6230	24.42/**0.6233**	24.08/0.6015	24.39/0.6228
B100	3	27.21/0.7389	28.29/0.7835	28.41/0.7863	28.83/0.7990	28.82/0.7973	**28.84**/**0.7996**	28.68/0.7896	28.82/**0.8014**
	4	25.96/0.6678	26.82/0.7087	26.90/0.7101	27.29/0.0726	27.32/0.7264	27.32/0.7270	27.10/0.7233	**27.53**/**0.7352**
	8	23.67/0.5470	24.20/0.5681	24.13/0.5652	24.49/0.5830	**24.54**/**0.5861**	24.52/0.5860	24.43/0.5785	24.50/0.5860
Urban 100	3	24.46/0.7349	26.03/0.7973	26.24/0.7989	27.14/0.8279	27.07/0.8271	**27.16**/0.8278	27.04/0.8270	27.08/**0.8281**
	4	23.14/0.6574	24.32/0.7183	24.52/0.7221	25.18/0.7540	25.21/0.7553	**25.21**/0.7560	25.06/0.7508	25.18/**0.7562**
	8	20.74/0.5152	21.37/0.5453	21.29/0.5432	21.70/0.5710	21.81/0.5813	**21.81**/0.5816	21.62/0.5618	21.75/**0.5820**

generated smoother images in terms of visual effect than the image reconstructed by methods based on the GAN, where the edges exhibit a certain step effect, especially at high magnification. For example, at a magnification factor of 3×, edges are still visible in images generated using the SRCNN, VDSR, LapSRN, and GuideAE. At a 4× magnification, edges are barely visible in images reconstructed by these methods, and become blurred at a magnification of 8×. However, the SRGAN and ADTV-SRGAN of GAN-based super-resolution methods are able to reconstruct higher-frequency details in the images than other methods. The edges are clearer and complete in these images, especially at a high magnification factor, and the improvement in the quality of the reconstructed image is obvious. Compared with the SRGAN, the proposed model ADTV-SRGAN uses global perception and local structure adaptation, which enable the loss function to adapt to the structure of the image space, and lead to a more realistic visual effect than the SRGAN. Visual deviation from the original sample is minimal, and better than that attained by the SRGAN.

In the aspect of objective evaluation, as shown in Table 2, a comparison in terms of PSNR and SSIM is made between the ADTV-SRGAN, and the Bicubic [1], A+ [2], SRCNN [7], VDSR [13], LapSRN [14], GuideAE [15], and SRGAN [20]. The parameters of these methods remain unchanged and can be referred in the corresponding study. The average PSNR and SSIM are obtained at magnification factors 3×, 4×, and 8×. The values of both indicators are higher for deep learning-based methods than traditional methods, such as Bicubic and A+, and the PSNR values of the VDSR, LapSRN, and GuideAE are generally higher that of the SRCNN. Although the SRGAN is more consistent in terms of subjective effects with perceptions of human visual system, it lags behind methods of super-resolution imaging based on the CNN in terms of PSNR and SSIM. The proposed ADTV-SRGAN method exhibits some advantages over the

TABLE 3. Comparison of methods in terms of PI.

Dataset	Bicubic[1]	SRCNN[7]	VDSR [13]	SRGAN[20]	EnhanceNet[21]	PESRGAN[22]	ADTV-SRGAN
Set5	7.32	6.79	6.45	3.18	2.93	3.42	**2.93**
Set14	6.97	6.03	5.77	2.80	3.02	2.66	**2.62**
B100	6.94	6.04	5.70	2.59	2.91	2.25	**2.23**
Urban 100	6.88	5.94	5.54	3.30	3.47	3.41	**3.25**
Average	7.03	6.20	5.87	2.97	3.08	2.94	**2.76**

TABLE 4. Design of different combinations of loss functions.

Model design	Loss function combination	Description
PA	$L_{Pix}^{SR} + L_{Adv}^{SR}$	pixel loss and adversarial loss
PVA	$L_{Pix}^{SR} + L_{Vgg}^{SR} + L_{Adv}^{SR}$	pixel loss, characteristic loss and adversarial loss
PVA-TV	$L_{Pix}^{SR} + L_{Vgg}^{SR} + L_{Adv}^{SR} + L_{Spa-TV}^{SR}$	pixel loss, characteristic loss, adversarial loss and spatial loss based on the total-variation model
PVA-ADTV	$L_{Pix}^{SR} + L_{Vgg}^{SR} + L_{Adv}^{SR} + L_{Spa-ADTV}^{SR}$	pixel loss, characteristic loss, adversarial loss and spatial loss based on the adaptive diagonal total-variation model

SRGAN in terms of subjective visual effect and the objective evaluation indexes. Although its scores of PSNR and SSIM are not the highest on all datasets, it yields good results in relation to the similarity indicators, and yields PSNR values comparable to those of the VDSR, LapSRN, and GuideAE. The subjective visual effects of its reconstructed images are superior in terms of delicacy and naturalness. Because GAN tends to focus on the perceptual quality, evaluation index PSNR and SSIM cannot accurately measure its visual effect [19]. It has been proved by [18] that the perception index(PI) is more suitable for GAN-based super-resolution assessments, where a lower PI indicates better perceptual quality. Table 3 shows a comparison at magnification factor 4× in terms of PI is made between the ADTV-SRGAN,

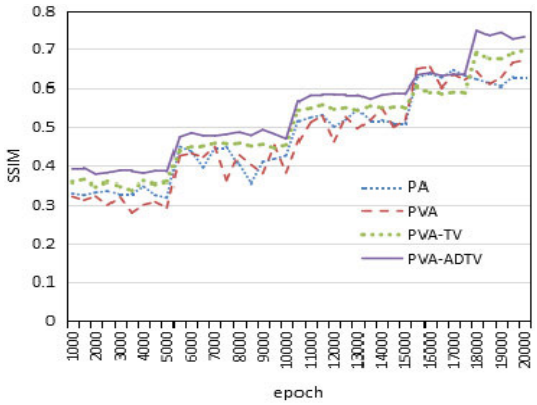


FIGURE 5. Comparison in terms of SSIM of different combinations of loss functions.

and the Bicubic [1], SRCNN [7], VDSR [13], SRGAN [20], EnhanceNet [21], and PESRGAN [22]. Method to calculate the PI is provided in [18]. From Table 3, ADTV-SRGAN obtains the best quantitative results than other methods in terms of PI, which the average value is 3.11, 0.21 and 0.32 less than that of VDSR, SRGAN and EnhanceNet separately. The results demonstrate that ADTV-SRGAN significantly improves the perceptual quality of super-resolution images.

C. COMPARISON OF DIFFERENT LOSS COMBINATIONS

To assess the loss function, verification models of different loss functions are designed. As shown in Table 4, pixel loss, characteristic loss, adversarial loss, and spatial loss are combined in different ways. To assess spatial loss based on the adaptive diagonal total-variation model, L_{Spa-TV}^{SR} is specially designed in the experiment.

The results are shown in Figure 5. The PVA-ADTV model combined with pixel loss, characteristic loss, adversarial loss, and spatial loss, and based on the adaptive diagonal total-variation model obtained the highest SSIM in the training epochs, followed by spatial loss based on the total-variation model. This shows that the PVA-ADTV model can be adjusted adaptively to the spatial features to improve reconstruction capability. The loss convergence of the ADTV-SRGAN is shown in Figure 7. It can be seen that the loss of ADTV-SRGAN is decreasing steadily until

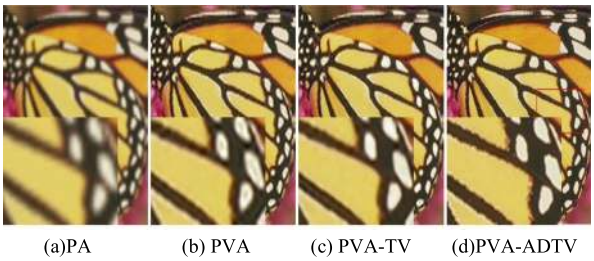


FIGURE 6. Comparison of visual quality of different combinations of loss functions.

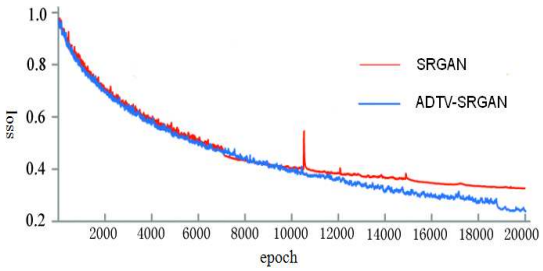


FIGURE 7. The loss convergence of the ADTV-SRGAN.

convergence is finished and provides better convergence property than SRGAN.

To further compare the capability of different combinations of loss function, Figure 6 shows a comparison of their visual quality. It is clear that the PA model lacked a sufficient number of high-frequency features, and the edge contours of reconstructed image are the most fuzzy. The PVA model uses characteristic loss and retains the high-frequency features of the edge contours, but its overall visual effect is not clear. The PVA-TV model uses spatial loss based on the total-variation model to preserve details of the images. The PVA-ADTV model applies spatial loss based on the adaptive diagonal total-variation model, and yields the clearest and most natural high-frequency features to yield a more realistic subjective visual effect.

D. COMPARISON OF COMPUTATIONAL COMPLEXITY

The computational complexity [40] of the proposed ADTV-SRGAN model is compared with the SRCNN [7], VDSR [13], LapSRN [14], GuideAE [15], and SRGAN [20]. The theoretical complexity of ADTV-SRGAN is $O\left(\sum_{l=1}^{D} M_l K_l C_{l-1} C_l\right)$, where D is the depth of the proposed model, l is the l-th layer, M is size of feature map and K is the kernel size, C is the number of feature maps. Table 5 shows the number of network layers, parameters, run time, and experimental results at a magnification factor of $\times 4$ on the set5 dataset. The run time is the average time for the super-resolution reconstruction for all images in set5. Table 2 shows that the SRCNN method is the fastest because it has the simplest network architecture and the fewest network layers, whereas the VDSR, LapSRN, and GuideAE contained deeper network convolution layers, which slow down computation. Because the SRGAN and ADTV-SRGAN use the network architecture of the GAN, they are slightly slower than GuideAE. However, the proposed method yields higher values of the PSNR and SSIM than the SRGAN. Thus, the ADTV-SRGAN can significantly improve the quality of the reconstructed image without excessive loss of speed, which verifies its feasibility.

TABLE 5. Comparison of computational complexit.

Methods	SRCNN	VDSR	LapSRN	GuideAE	SRGAN	ADTV-SRGAN
Layers	3	20	24	32	33	33
Param(k)	57	665	812	1,482	1,500	1,562
Time(s)	0.197	0.298	0.311	0.489	0.512	0.513
PSNR	30.48	31.35	31.54	31.54	30.52	**31.58**
SSIM	0.8628	0.8830	0.8866	0.8850	0.8791	**0.8921**

V. CONCLUSION

To solve the problems of an excessively smooth image with insufficiently precise details of the edges and texture, and distorted visual effects at high magnification in reconstructing super-resolution images, a model based on the adaptive diagonal total-variation generative adversarial network is proposed by combining the GAN and the total-variation model. Spatial loss based on the diagonal total-variation model is introduced to adjusted the loss function according to the spatial features, and pixel loss and characteristic loss are used for the fusion optimization of total loss. Through comparisons of various aspects of performance in experiments, it is clear that the proposed method can fully restore the textural features of images, maintain high-frequency details while improving the image quality, and can better reconstruct super-resolution images.

REFERENCES

[1] F. Zhou, W. Yang, and Q. Liao, "Interpolation-based image super-resolution using multisurface fitting," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3312–3318, Jul. 2012.
[2] S. Khan, D.-H. Lee, M. A. Khan, A. R. Gilal, and G. Mujtaba, "Efficient edge-based image interpolation method using neighboring slope information," *IEEE Access*, vol. 7, pp. 133539–133548, 2019.
[3] A. Giachetti and N. Asuni, "Real-time artifact-free image upscaling," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2760–2768, Oct. 2011.
[4] J. Liu, S. Dai, Z. Guo, and D. Zhang, "An improved POCS super-resolution infrared image reconstruction algorithm based on visual mechanism," *Infr. Phys. Technol.*, vol. 78, pp. 92–98, Sep. 2016.
[5] R. Timofte, R. Rothe, and L. V. Gool, "Seven ways to improve example-based single image super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1865–1873.
[6] T. Yao, Y. Luo, Y. Chen, D. Yang, and L. Zhao, "Single-image super-resolution: A survey," in *Proc. Int. Conf. Commun., Signal Process., Syst.* Singapore: Springer, 2018, pp. 119–125.

[7] 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.

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

[9] V. K. Ha, J. Ren, X. Xu, S. Zhao, G. Xie, and V. M. Vargas, "Deep learning based single image super-resolution: A survey," in *Proc. Int. Conf. Brain Inspired Cogn. Syst.* Cham, Switzerland: Springer, 2018, pp. 106–119.

[10] X. Li, Y. Wu, W. Zhang, R. Wang, and F. Hou, "Deep learning methods in real-time image super-resolution: A survey," *J. Real-Time Image Process.*, vol. 3, pp. 1–25, Nov. 2019.

[11] I. J. Goodfellow, J. Pouget-Abadie, and M. Mirza, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 3, 2014, pp. 2672–2680.

[12] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "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.

[13] 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.

[14] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 624–632.

[15] R. Chen, Y. Qu, C. Li, K. Zeng, Y. Xie, and C. Li, "Single-image super-resolution via joint statistic models-guided deep auto-encoder network," *Neural Comput. Appl.*, vol. 5, pp. 1–12, Nov. 2018.

[16] Y. Wang, F. Perazzi, B. McWilliams, A. Sorkine-Hornung, O. Sorkine-Hornung, and C. Schroers, "A fully progressive approach to single-image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 864–873.

[17] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 694–711.

[18] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6228–6237.

[19] M. S. Rad, B. Bozorgtabar, U.-V. Marti, M. Basler, H. K. Ekenel, and J.-P. Thiran, "SROBB: Targeted perceptual loss for single image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2710–2719.

[20] 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.

[21] 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.

[22] T. Vu, T. M. Luu, and C. D. Yoo, "Perception-enhanced image super-resolution via relativistic generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 98–113.

[23] V. Chudasama and K. Upla, "ISRGAN: Improved super-resolution using generative adversarial networks," in *Proc. Sci. Inf. Conf.* Cham, Switzerland: Springer, 2019, pp. 109–127.

[24] G. Zhang, J. Wang, X. Zhang, H. Fei, and B. Tu, "Adaptive total variation-based spectral-spatial feature extraction of hyperspectral image," *J. Vis. Commun. Image Represent.*, vol. 56, pp. 150–159, Oct. 2018.

[25] Y. Zhao, J. G. Liu, B. Zhang, W. Hong, and Y.-R. Wu, "Adaptive total variation regularization based SAR image despeckling and despeckling evaluation index," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2765–2774, May 2015.

[26] L. Deng, D. Mi, P. He, P. Feng, P. Yu, M. Chen, Z. Li, J. Wang, and B. Wei, "A CT reconstruction approach from sparse projection with adaptive-weighted diagonal total-variation in biomedical application," *Bio-Medical Mater. Eng.*, vol. 26, no. s1, pp. S1685–S1693, Aug. 2015.

[27] L.-Z. Deng, P. He, S.-H. Jiang, M.-Y. Chen, B. Wei, and P. Feng, "Hybrid reconstruction algorithm for computed tomography based on diagonal total variation," *Nucl. Sci. Techn.*, vol. 29, no. 3, p. 45, Mar. 2018.

[28] G. Seif and D. Androutsos, "Edge-based loss function for single image super-resolution," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 1468–1472.

[29] J. T. Barron, "A general and adaptive robust loss function," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4331–4339.

[30] 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)*, 2018, pp. 1–17.

[31] R. Duan, D. W. Zhou, and L. J. Zhao, "Image super-resolution reconstruction based on multi-scale feature mapping network," *J. Zhejiang Univ.*, vol. 53, no. 7, pp. 1331–1339, 2019.

[32] A. Langer and F. Gaspoz, "Overlapping domain decomposition methods for total variation denoising," *SIAM J. Numer. Anal.*, vol. 57, no. 3, pp. 1411–1444, Jan. 2019.

[33] J. Liu, T.-Z. Huang, I. W. Selesnick, X.-G. Lv, and P.-Y. Chen, "Image restoration using total variation with overlapping group sparsity," *Inf. Sci.*, vol. 295, pp. 232–246, Feb. 2015.

[34] L. Chen, Q. Kou, D. Cheng, and J. Yao, "Content-guided deep residual network for single image super-resolution," *Optik*, vol. 202, Feb. 2020, Art. no. 163678.

[35] M. Wang, X. Zheng, J. Pan, and B. Wang, "Unidirectional total variation destriping using difference curvature in MODIS emissive bands," *Infr. Phys. Technol.*, vol. 75, pp. 1–11, Mar. 2016.

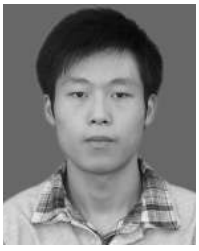
[36] I. K. M. Jais, A. R. Ismail, and S. Q. Nisa, "Adam optimization algorithm for wide and deep neural network," *Knowl. Eng. Data Sci.*, vol. 2, no. 1, pp. 41–46, 2019.

[37] W. Wan, J. Wu, G. Shi, Y. Li, and W. Dong, "Super-resolution quality assessment: Subjective evaluation database and quality index based on perceptual structure measurement," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2018, pp. 1–6.

[38] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, 2010, pp. 2366–2369.

[39] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 1–21.

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



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

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

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

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

MA LU was born in Anhui, China, in 1979. He received the master's degree in computer application technology from the Computer Department, Nanjing University. He is currently a Visiting Scholar with the Computer Department, Nanjing University. His research interests include image processing and video compression.