Abstract—The super-resolution (SR) reconstruction method based on deep learning can significantly improve the spatial SR of remote sensing images. However, the current methods make insufficient use of the remote context information and channel information in shallow feature extraction, resulting in the limited effect of SR reconstruction. This article proposed a new SR reconstruction model, SIEGAN, which uses generative adversarial network with shallow information enhancement to improve the effect of SR reconstruction of remote sensing images. Similar to other generative adversarial models, SIEGAN is composed of generator and discriminator. But SIEGAN enhances the generator’s ability to extract shallow information by using three different scale convolution operations. Specifically, a depthwise convolution is used to extract the local context information of each band of the image. A depthwise dilation convolution is used to capture the remote context information in the image. Finally, a 1 × 1 convolution is used to extract the correlation features between different channels in remote sensing images. In addition, SIEGAN uses U-Net network as its discriminator to provide detailed feedback per pixel to the generator to improve the model’s ability to identify image details. And the spectral–spatial total variation loss function is introduced to ensure the spectral–spatial reliability of the reconstructed images. The experimental results on Gaofen-1 data proved that compared with the state-of-the-art models, SIEGAN has achieved better SR reconstruction performance. Furthermore, the reconstructed images by SIEGAN demonstrate better performance in land cover classification.

Index Terms—Generative adversarial network (GAN), multiscale shallow information, remote sensing images, super-resolution (SR) reconstruction.

I. INTRODUCTION

SATellite remote sensing images provide important and effective information for earth surface monitoring, and are widely used in target detection [1], land cover classification [2], disaster early warning [3], urban economic level assessment [4], ocean exploration [5], single modality [6], and multimodal classification [7], [8], [9]. However, due to the limitations of environment, imaging equipment, and other factors, the spatial resolution of the acquired remote sensing images is limited, which affects the application effect of remote sensing image in practical tasks. Moreover, hyperspectral images collected from airborne or satellite sources inevitably suffer from spectral variability issues [10], resulting in low-quality images that cannot meet actual needs. Image super-resolution (SR) technology can improve the resolution of the original images without changing the hardware equipment of the imaging system. By generating clearer high-resolution reconstructed images, it can meet the needs of various applications for high-resolution remote sensing images.

Image SR reconstruction technology refers to the process of improving the spatial resolution of images, which was first proposed by Harris et al. [11] in the 1960s. Through the process of SR reconstruction, the pixel density and detail information in the images can be significantly improved. In recent years, image SR reconstruction technology combined with deep learning methods has become the mainstream. Dong et al. [12] proposed an SR reconstruction convolutional model of SRCNN. Compared with the traditional image SR reconstruction models, SRCNN has higher peak signal-to-noise ratio (PSNR). Subsequently, more models based on deep learning have been proposed, such as FSRCNN [13], VDSR [14], EDSR [15], and FormResNet [16]. However, these models use the mean square error (MSE) as the loss function, which makes the SR reconstructed images lose a lot of high-frequency details. Ledig et al. [17] proposed SRGAN model based on the generative adversarial network (GAN), which uses the pretrained VGG network to calculate the perceptual loss, so as to generate more realistic SR reconstructed images. Based on SRGAN, Wang et al. [18] proposed ESRGAN model, which uses several residual-in-residual dense blocks (RRDBs) as its basic structure. In addition, Liang et al. proposed SwinIR model [19] to reconstruct SR images based on transformer mechanism.

Recently, the models based on deep learning have also been extended to the field of SR reconstruction of satellite remote sensing images. However, different from natural images, satellite remote sensing images contain richer texture information, such as mountains, buildings, and rivers, which makes the
mapping relationship between high and low resolution extracted from natural images unable to be directly used for SR reconstruction of remote sensing images. Liebe et al. [20] retrained the SRCNN model with Sentinel2 data to learn the mapping relationship between high and low resolution of remote sensing images, and proposed msisSRCNN model to reconstruct the remote sensing images. Haut et al. [21] proposed the RSRCAN model for reconstructing the SR remote sensing image, which combines the attention mechanism with the residual network. Dong et al. [22] designed the dense sampling super-resolution model to broaden the feature channel in the SR reconstruction of remote sensing images to extract more feature information. Xiong et al. [23] improved the SRGAN model to improve the model’s generalization ability for SR reconstruction of remote sensing images. Dong et al. [24] proposed a coupling network combining CNN and hyperspectral unmixing to improve the spatial resolution of hyperspectral images in an unsupervised manner. Jiang et al. [25] proposed EGGAN model, which uses the Laplace operator to enhance the edge information of the image. Luis et al. [26] further improved the ESRGAN model to increase the resolution of Sentinel2 images to 2 m. In the SR model based on deep learning, the shallow feature extraction is mainly used to recover the structural information of the images, while the deep feature extraction is mainly used to recover the detail information of the images, which complement each other. However, the current SR reconstruction models only use single-scale convolution operation to extract the shallow feature of the images, resulting in insufficient extraction of the shallow information of images, which will affect the effect of image SR reconstruction of remote sensing images.

To address the above problems, a new SR reconstruction model, named SIEGAN, is proposed to improve the effect of SR reconstruction of remote sensing images. SIEGAN adopts the GAN as the backbone structure and improves its ability to extract shallow information from remote sensing images by introducing the idea of large kernel convolution. By decomposing the large kernel convolution into depthwise convolution, depthwise dilation convolution and 1 × 1 convolution, SIEGAN can extract the shallow feature of remote sensing image at three different scales, which will provide enough initial information for the subsequent SR reconstruction process.

The contributions of the current article are as follows:

1) A model of SIEGAN is proposed, which can significantly improve the quality of SR reconstruction of remote sensing images.
2) A strategy of extracting shallow feature by using three different scale convolution operations is proposed to effectively enhance the shallow feature representation in the process of image SR reconstruction.
3) U-Net is used as the discriminator of SIEGAN, which enables SIEGAN to identify the authenticity of each pixel, thus improving the model’s ability to recognize image details.
4) The spectral–spatial total variation loss function is introduced to improve the spectral–spatial reliability of the SR reconstructed remote sensing images.

The rest of this article is organized as follows. Section II discusses the framework of SIEGAN, and its key modules. Section III gives the experimental results and Section IV concludes this article.

II. METHODOLOGY

In this article, a model of SIEGAN is proposed to solve the problem of insufficient extraction of shallow feature in the process of SR reconstruction of remote sensing image. Similar with other models based on GAN, SIEGAN model is also composed of two components of generator and discriminator. The generator is used to extract shallow and deep features from the original low-resolution remote sensing images and generate high-resolution remote sensing images based on the fusion of these features. The discriminator is used to distinguish the images generated by the generator from the original high-resolution remote sensing images. The generator network and discriminator network are trained and updated iteratively to ensure that the model converges to convincing results. In this process, the full extraction of image features and the effective selection of discriminators will affect the effect of image SR reconstruction.

A. Generator of SIEGAN Model

Fig. 1 shows the structure of the generator network in SIEGAN. The generator network is composed of shallow feature extraction module, deep feature extraction module, and high-resolution image reconstruction module. In the shallow feature extraction module, three different convolution operations are used to extract image shallow features from multiple perspectives. RRDBs are used in the deep feature extraction module to extract the deep features of images. Through the long-distance skip connection, the shallow and deep features are aggregated and input into the high-resolution image reconstruction module to generate high-resolution remote sensing image.

1) Shallow Feature Extraction: When extracting the shallow features of remote sensing image, most of the existing models focus on extracting the local feature of the image, while ignoring the long-distance correlation feature and channel feature of the image. The limitation of shallow feature extraction will affect the quality of image SR reconstruction. Inspired by MobileNet [27] and VAN [28], this article introduces the idea of large kernel convolution to extract shallow features from remote sensing image. The large kernel convolution is decomposed into three convolutions with different scales. Each convolution can extract shallow feature from a specific scale. Specifically, given a $K \times K$ large kernel convolution, it can be decomposed into a $(2d−1) \times (2d−1)$ depthwise convolution, a $K dl \times K dl$ depthwise dilation convolution, and a $1 \times 1$ convolution, where $d$ denotes the dilation rate. The depthwise convolution is used to extract local contextual information in each band of the remote sensing images. The depthwise dilation convolution has a large receptive field and focuses on capture the long-distance relationship in the image. The $1 \times 1$ convolution is used to aggregate the information between each band of the remote
sensing images. These convolutions with different scales help the generator to extract shallow feature from different receptive fields, thus providing more abundant shallow feature for the SR reconstruction process. Fig. 2 shows the structure of the shallow feature extraction module, where the gray grid represents the position of the centroid and the blue grid represents the position of the convolution kernel.

Given a low-resolution image $I_{LR}$, the process of extracting shallow feature $F_{SF} \in \mathbb{R}^{H \times W \times C_{LR}}$ is shown as

$$F_{SF} = \text{Conv}_{1 \times 1} (\text{Conv}_{DWD} (\text{Conv}_{DW} (I_{LR})))$$

where $C$ is the feature channel number, $\text{Conv}_{1 \times 1}$ represents channel convolution, $\text{Conv}_{DW}$ represents depthwise convolution, and $\text{Conv}_{DWD}$ represents depthwise dilation convolution.

2) Deep Feature Extraction: In the deep feature extraction module, the deep feature of remote sensing images is extracted based on the obtained shallow feature. The core part of the deep feature extraction module is several RRDBs, which consists of multilevel residual networks and dense connections. Fig. 1(a) shows the structure of RRDB. RRDB uses dense blocks as its main component and multiplies the residual information by a constant $\beta$ to enhance the stability of the network [29]. This dense connection can alleviate the problem of gradient vanishing, enhance the reuse of features, and reduce the number of parameters, so as to extract the deep feature of the image more effectively.

Based on the extracted shallow feature $F_{SF}$, the deep feature $F_{DF} \in \mathbb{R}^{H \times W \times C}$ is obtained by the deep feature extraction module, as follows:

$$F_{DF} = H_{DF} (F_{SF})$$

where $H_{DF}$ denotes the deep feature extraction module, and it contains a $3 \times 3$ convolutional layer and $K$ RRDBs. Intermediate feature $F_i$, $m$ and the output deep feature $F_{DF}$ are extracted as

$$F_i = H_{RRDB_i} (F_{i-1}), i = 1, 2, \ldots, K$$

$$F_{DF} = H_{Conv} (F_K)$$

where $F_0 = F_{SF}$ denotes the shallow features, $H_{RRDB_i}$ denotes the $i$ th RRDB, and $H_{Conv}$ is a $3 \times 3$ convolutional layer.

3) High-Resolution Image Reconstruction: The fusion of shallow and deep features is helpful to improve the effect of SR reconstruction of remote sensing image. In the high-resolution image reconstruction module, a long-distance skip connection operation is used to aggregate the acquired shallow and deep features at the pixel level. The aggregated features are convoluted twice and mapped into HR space to obtain high-resolution reconstructed remote sensing images. The process is shown as follows:

$$I_{SR} = H_{HRG} (F_{SF} + F_{DF})$$

where $H_{HRG}$ denotes the high-resolution image reconstruction module, which consists of two $3 \times 3$ convolutions and an activation function. $I_{SR}$ denotes the generated SR image and + denotes the skip connection.
B. Discriminator of SIEGAN Model

Remote sensing images contain rich texture and detail information. Making full use of these texture and detail information to distinguish the authenticity of the reconstructed high-resolution remote sensing image will help to improve the quality of the reconstructed images. This article uses U-Net [31] network as the discriminator of SIEGAN model. U-net can output the realness value of each pixel in the image, which will improve the ability of SIEGAN to distinguish image details.

Fig. 3 shows the structure of the discriminator in SIEGAN model. The U-Net has a contracting path on the left and an expansive path on the right. The contracting path consists of several identical 3 × 3 convolutions, each followed by a spectral normalization (spectral norm) and a downsampling operation. The spectral norm can help alleviate the annoying artifacts caused by GAN training. The features obtained from the contracting path are fed into the expansive path. Each convolutional layer in the expansive path is also followed by a spectral norm, and every step consists of an upsampling of the feature map, and the resulting features are connected to the feature map in the contracting path and fed into a 3 × 3 convolutional layer at the end of the network. Finally, a 1 × 1 convolution maps the feature vector to the feature space we need for determining the authenticity of each pixel in the image. At the same time, in order to avoid the problem that convolution operation is easy to lose image boundary information, we add paddings in the convolution layer to keep the scale of feature mapping unchanged. This will further ensure the discriminator’s ability to distinguish image details.

C. Loss Functions

In this article, $L_1$ loss, perceptual loss, adversarial loss, and spectral–spatial total variation loss are used to train the model.

1) $L_1$ Loss: $L_1$ loss, also known as mean absolute error loss, is calculated as the mean of the sum of the absolute differences between the generated output $G(x_i)$ and the target $y_i$ for each $i$th image. $L_1$ loss is calculated as follows:

$$L_1(G) = E_{x,y} ||G(x_i) - y_i||_1.$$  

2) Perceptual Loss: The commonly used perceptual loss function is the MSE, which can achieve PSNR. However, MSE usually ignores high-frequency information, which will cause the texture of the generated image to be too smooth. To solve this problem, Johnson et al. [32] introduced a loss function $L_\text{percep}$ that is closer to perceptual similarity, which uses a pretrained VGG-19 [33] network as a feature extractor to compute the average error between two feature maps.

In this article, the pretrained VGG model is fine-tuned to calculate the perceptual loss in the four bands of remotely sensed images, which is calculated as

$$L_\text{percep} = \frac{1}{C_j H_j W_j} \left\| \phi_i(G(x_i) - \phi_i(y_i)) \right\|_1$$  

where $\phi_i$ denotes the pretrained VGG network, and $C_j$, $H_j$, and $W_j$ denote the size of the $j$th layer feature map.

3) Spectral–Spatial Total Variation Loss: The spectral–spatial total variation loss function [34] can consider both the spectral and spatial characteristics of satellite remote sensing images to ensure the spatial and spectral confidence of the reconstructed images, which is calculated as

$$L_{\text{sstv}} = \frac{1}{HWC} \sum_{i,j,k} \sqrt{(I_{x+1,y,k} - I_{x,y,k}) + (I_{x,y,k+1} - I_{x,y,k}) + (I_{x,y,k+1} - I_{x,y,k})}.$$  

4) Adversarial Loss: The adversarial loss of the generator is as follows, where $x_r$ denotes the true HR image and $x_f$ denotes the HR image generated by the generator:

$$L_\text{adv} = -E_{x_r} \left[ \log (1 - D(x_r, x_f)) \right] - E_{x_f} \left[ \log (D(x_f, x_r)) \right].$$  

The adversarial loss of the discriminator is calculated as follows:

$$L_D = -E_{x_r} \left[ \log (D(x_r, x_f)) \right] - E_{x_f} \left[ \log (1 - D(x_f, x_r)) \right].$$
To create a training sample set containing pairs of high-resolution and low-resolution remote sensing images, two sensors are required to image common ground targets at the same position and the same spectral band. The GF1 satellite images in Shandong Province are selected as the image sample of our experiment. The red, green, blue, and near-infrared data in PMS images and WFV images are extracted, respectively. Shandong Province is rich in surface and ecological types, including different forms of woodland, shrub, grassland, farmland, buildings, desert, ice and snow, water, wetland, and other geomorphic types. The selected remote sensing images are preprocessed, including radiometric correction, atmospheric correction, and orthophoto correction. In addition, WFV image is resampled to 8-m resolution to achieve registration with PMS image. At the same time, all images are cropped to 128 × 128 pixels and randomly assigned to the training set and the test set. Fig. 4 shows the true color display of some preprocessed remote sensing images. These photos were taken in May 2019. It can be seen that in each pair of remote sensing images, the HR image is clearer than the LR image and contains more detailed information.

### B. Metrics

The evaluation indexes of PSNR, structural similarity index (SSIM), spectral angle mapper (SAM), and local standard deviation (LSD) are used to evaluate the performance of SR reconstruction of remote sensing images.

1) PSNR: PSNR is the most common evaluation metric for image SR reconstruction. This index evaluates the image quality based on the error between the corresponding pixels. The larger the PSNR value, the smaller the distortion of the reconstructed image and the better the reconstruction image.

PSNR (dB) is calculated as

\[
\text{PSNR}(x, y) = 10 \cdot \log \left( \frac{\text{MaxVal}^2}{\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (x - y)^2} \right)
\]

where MaxVal denotes the maximum pixel value of the image and \(a \times b\) is the image size.

2) SSIM: SSIM measures the similarity between images in terms of brightness, structure, and contrast, which can better reflect the subjective perception of human eyes. The calculation of SSIM is

\[
\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\mu_x^2 + \mu_y^2 + C_1(\sigma_x^2 + \sigma_y^2 + C_2)}
\]

where the constants \(C_1 = (K_1L)^2\) and \(C_2 = (K_2L)^2\) depends on the dynamic range of the pixel values \(L\). The default is \(K_1 = 0.01\) and \(K_2 = 0.03\). The value range of SSIM is \([0, 1]\). The closer it is to 1, the better the image quality.

3) SAM: SAM regards the spectrum of each image element as a high-dimensional vector and measures the similarity between the spectra by calculating the angle between the two vectors. The smaller the angle, the more similar the two spectra are. SAM is calculated as

\[
\text{SAM}(x, y) = \arccos \left( \frac{x \cdot y}{\|x\|_2 \|y\|_2} \right).
\]
Fig. 4. Example of paired remote sensing images. First row shows the examples of low-resolution remote sensing images and second row shows the examples of high-resolution remote sensing images.

4) LSD: LSD reflects the degree of gray variation in local areas of an image. In remote sensing images, the same ground objects have similar spectral characteristics and have similar gray values in the same band. The boundary is used to segment different ground objects. In the boundary of the image, there is usually a large LSD due to the large gray variation between different objects. Therefore, the index LSD can also be used to measure the restoration effect of the SR reconstructed image on the details of the original image.

The LSD value of pixel \((i, j)\) is calculated as

\[
\text{LSD}(i,j) = \sqrt{\frac{1}{K^2} \sum_{t=1}^{K^2} (f(i,j) - \text{LM})^2}
\]  

(15)

where \(f(i,j)\) is the gray value of pixel \((i,j)\); and \(\text{LM}\) represents the average value of the gray value in the \(K \times K\) window \((K = 3\) in this article).  

C. Parameter Setting

The WFV data were resampled to 8-m resolution and coaligned with the PMS images. The images are cropped to a size of 128 \(\times\) 128 pixels, and the overlapping pixel size is 10 to avoid too obvious splicing traces. A total of 7800 images are obtained through the above operations, which are randomly divided into the training samples and the test samples. Finally, a training set containing 5000 remote sensing images and a test set containing 2800 remote sensing images are generated.

Adam optimizer is used in the training process of SIEGAN, and the learning rate is initialized to \(2 \times 10^{-4}\), and the input mini-patch is 8. SIEGAN is trained by alternatively updating the discriminator and generator. Specifically, our model is trained in two steps. In the first step, we train the generator by minimizing the \(L_1\) loss for 100k iterations, which allows the subsequent adversarial training to have a better initialization. Then, train our model based on (10) and (11), where \(\lambda_1 = 1, \lambda_2 = 1e^{-3}, \lambda_3 = 1e^{-1}, \lambda_4 = 5e^{-3}\), learning rate \(1 \times 10^{-4}\), and 200K iterations.

D. Experimental Results and Analysis

1) Evaluation on Super-Resolution: Several deep-learning-based SR reconstruction models are selected as the benchmark comparison models. These models include the convolution-based model of SRCNN, the residual-network-based models of EDSR, RCAN [35], and VDSR; the GAN-based models of ESRGAN and Real_ESRGAN [36], and the transformer-based model of SwinIR. For fair comparison, these models are retrained on the remote sensing image dataset constructed in this article. Table III and Fig. 5 show the performance of SR reconstruction of different models under different metrics. In Table III, the best results are indicated in bold.

It can be seen from Table III and Fig. 5 that SIEGAN performs well under all evaluation indicators. In particular, SIEGAN achieved the best performance under SSIM and SAM indicators. The good performance of SIEGAN in three indexes shows that it is feasible and effective to use SIEGAN to SR reconstruct of remote sensing images. Besides the SIEGAN model, SRCNN and Real_ESRGAN

### TABLE III

| Model         | PSNR | SSIM  | SAM  |
|---------------|------|-------|------|
| SRCNN         | 50.862 | 0.982 | 2.573 |
| EDSR          | 43.909 | 0.944 | 8.156 |
| RCAN          | **51.150** | 0.990 | 6.732 |
| VDSR          | 46.449 | 0.974 | 10.560 |
| ESRGAN        | 46.616 | 0.977 | 4.615 |
| Real_ESRGAN   | 50.765 | 0.982 | 2.859 |
| SwinIR        | 50.469 | 0.987 | 4.579 |
| SIEGAN        | 51.102 | **0.991** | **2.487** |

An NVIDIA Tesla V100-SXM2 GPU with 32 GB VRAM is used for training and testing. SIEGAN model is implemented on Pytorch with Python3 under CUDA 10.1 system.
Fig. 5. Performance of super-resolution reconstruction under different indexes.

Group (1)

LR  SRCNN  EDSR  RCAN  VDSR

ESRGAN  Real_ESRGAN  SwinIR  SIEGAN  HR

Group (2)

LR  SRCNN  EDSR  RCAN  VDSR

ESRGAN  Real_ESRGAN  SwinIR  SIEGAN  HR

Fig. 6. Performance of super-resolution reconstruction of remote sensing images under different models.
Fig. 7. Results of local standard deviation.

also achieved good performance in three indicators. However, the performance of EDSR and VDSR is relatively poor under the three indicators.

Fig. 6 shows the visual effects of remote sensing image SR reconstruction under different models. Group (1) in Fig. 6 shows the remote sensing images containing more flat areas, and there is relatively little detailed information in the images. Group (2) shows the remote sensing images relatively more detailed information in the images. From the visual effect of SR reconstruction of two groups of remote sensing images, the SIEGAN model proposed in this article has achieved a good visual effect of image SR reconstruction, and the image reconstructed by SIEGAN model is more real. In the images reconstructed by SIEGAN, the ground object information is restored more clearly, the contour is defined more clearly, and there are fewer artifacts. Therefore, combined with the quantitative results given in Table III and the qualitative visual effects given in Fig. 6, SIEGAN model proposed in this article achieves the best SR reconstruction performance compared with other SR reconstruction models.

In addition, models of RCAN, and SRCNN also perform well under the metrics, but the visual effect of reconstructed images under the two models is not as good as that of SIEGAN. The main reason is that the SR images reconstructed by the two models cannot produce rich details, and some even produce artificial artifacts. It is worth mentioning that the SR reconstructed images obtained by SwinIR model cannot achieve a good quantitative performance value, but it has achieved a good visual effect. This shows that the quantitative performance index of SR
reconstruction is not enough to fully characterize the effect of image SR reconstruction. And in the task of image SR reconstruction, it is still a difficult problem to find a better quantitative index that conforms to the effect of human visual recognition.

Fig. 7 shows the effect of image SR reconstruction measured by LSD index. The boundary between different ground objects in the image can be displayed according to the change of the LSD value. It can be seen that the image reconstructed by model SIEGAN generate more complete and clear boundaries, and these boundaries are closer to the boundaries in the original HR image. Other models are inferior to SIEGAN model in reconstructing the remote sensing images. The images reconstructed by other models have different degrees of boundary loss, especially SRCNN, EDSR, RCAN, VDSR, and ESRGAN. Although Real_ESRGAN and SwinIR can generate relatively clear boundaries, the restoration effect of the boundary information is worse than SIEGAN.

Each spectral band of satellite remote sensing image is very important. The red band can measure the health status of green plants. The green band can distinguish man-made features and evaluate crop growth trends. The blue band is used to reflect soil and vegetation information. The main function of the near-infrared band is to distinguish vegetation from nonvegetation. SIEGAN model considers four spectral features of the WFV images when reconstructing SR images. We decompose the reconstructed remote sensing images into four bands: red, green, blue, and near-infrared, and compare the reconstruction effect of the SIEGAN model on the four bands with that in the original low-resolution remote sensing image.

Fig. 8 shows the effect of SR reconstruction of SIEGAN on each band. It can be seen that the image reconstruction by SIEGAN has improved the clarity of all bands and the contours of the ground features have become clear. This shows that the SIEGAN model cannot only improve the spatial resolution of
Fig. 10. Results of ablation experiments.

TABLE IV
CLASSIFICATION RESULTS OF IMAGES BEFORE AND AFTER SUPER-RESOLUTION

|               | LR  | SRCNN | EDSR | VDSR | RCAN | ESRGAN | Real_ESRAN | SwinIR | SIEGAN |
|---------------|-----|-------|------|------|------|--------|------------|--------|--------|
| OA            | 73.6% | 75.8% | 77.2% | 78.6% | 79.2% | 80.5%  | 81.3%      | 82.1%  | 82.7%  |
| Kappa         | 0.6753 | 0.6865 | 0.6896 | 0.6937 | 0.7142 | 0.7326 | 0.7403     | 0.7486 | 0.7588 |

2) Evaluation on Super-Resolution With Land Cover Classification: Land cover classification is an important research work in the field of remote sensing applications, especially the monitoring of land cover dynamics in large-scale areas is the basis of scientific research on global change and sustainable development. In this article, the remote sensing images before and after SR are applied to the task of land cover classification to further explore the reconstruction performance of the SR model. The remote sensing images used for land cover classification in this article are selected from Shandong Province. The land cover classification labels refer to the labels given in the 10-m resolution land cover data released by ESA in 2020. The data includes 11 types of labels: woodland, shrub, grassland, cultivated land, architecture, desert, snow (ice), waterbody, wetland, mangrove, and moss. The remote sensing images selected in this experiment contain eight kinds of ground objects. Segnet network [37] is used to realize the task of land cover classification. Classification overall accuracy and Kappa coefficient are used to evaluate the effect of classification. The Kappa coefficient considers both missing and wrong pixels. The higher the Kappa coefficient, the better the classification effect of the model.

Table IV shows the land cover classification results of remote sensing images. It can be seen that the classification results of the images reconstructed by the SR model are improved compared with the original low-resolution images. Among them, SIEGAN improves the overall accuracy of the image before reconstruction by 10%, the Kappa coefficient increases by 0.08, and the classification results of the images generated by SIEGAN are better than other methods. It can also be seen from Fig. 9 that the classification effect of the SIEGAN reconstructed image is better than the original low-resolution image and other advanced models in the recognition of image boundaries and scattered objects. Combining qualitative and quantitative results, the SIEGAN method proposed in this article performs the best...
TABLE V

| Comparison Results of the Ablation Analysis |
|-------------------------------------------|
| DW-Conv | DW-D-Conv | 1x1 Conv | PSNR | SSIM | SAM |
|---------|-----------|----------|------|------|-----|
| x       | ✓        | ✓        | 50.851 | 0.990 | 2.674 |
| ✓       | x        | ✓        | 50.913 | 0.990 | 2.603 |
| ✓       | ✓        | x        | 50.872 | 0.990 | 2.739 |
| ✓       | ✓        | ✓        | 50.964 | 0.991 | 2.496 |

and can significantly improve the performance of land cover classification, which also verifies the effectiveness of the model in improving the spatial resolution of remote sensing images.

E. Ablation Study

In order to verify the necessity of extracting shallow feature information by convolution operation at different scales in SIEGAN model, ablation experiments were conducted in this article. On the basis of the SIEGAN model, the convolution operation of a specific scale is removed, and the SR reconstruction of remote sensing images is completed on the basis of the simplified model. Table V shows the results of ablation experiments, and Fig. 10 shows the results of visualization of ablation experiments. It can be seen that removing convolution operation at any specific scale will affect the effect of SR reconstruction of the model. This shows that the shallow information extracted by the three convolution operations plays an important role in improving the SR reconstruction effect of the image. That is, the idea of extracting the shallow information of an image from three different receptive fields proposed in this article is reasonable. From the visual effect shown in Fig. 10, the complete SIEGAN model generates a clearer SR reconstructed image than the simplified model with a specific convolution operation removed. This further verifies the necessity of performing three convolution operations in the SIEGAN model.

IV. CONCLUSION

Aiming at the problem that the existing SR reconstruction models have limited ability to extract the shallow feature of remote sensing images, an SR reconstruction model based on GAN is proposed, named SIEGAN. SIEGAN designed three convolution operations with different scales in the generator to extract shallow information. By capturing more abundant shallow feature, it can enhance the shallow feature in the process of SR reconstruction. SIEGAN uses U-Net as its discriminator, which can output the pixel realness of the image to improve the reconstruction ability of the model to image details. At the same time, SIEGAN introduces the spectral–spatial total variation loss function to ensure the spectral spatial reliability of remote sensing images. The experimental results show that compared with the baseline model, the SR reconstruction images based on SIEGAN achieve better overall visual effect and the contour details of ground objects in the image are clearer. The images before and after SR reconstruction are applied to land cover classification. The results show that the images reconstructed by SIEGAN have better classification accuracy, which further verifies the effect of SIEGAN for SR reconstruction of remote sensing images.

In future article, we will further explore the semi-supervised or unsupervised SR reconstruction models to broaden the research scope of image SR reconstruction. Furthermore, this article only discusses the application of SR reconstruction remote sensing images in land classification. It is also an interesting and important work to explore a suitable SR reconstruction model in a wider range of practical applications, such as target detection and instance segmentation.

REFERENCES

[1] Z. Zou and Z. Shi, “Random access memories: A new paradigm for target detection in high resolution aerial remote sensing images,” IEEE Trans. Image Process., vol. 27, no. 3, pp. 1100–1111, Mar. 2018.
[2] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, “SuperPCA: A sparsely penalized PCA approach for hyperspectral feature extraction,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 8, pp. 4581–4593, Aug. 2018.
[3] S. Bredemeyer et al., “Radar path delay effects in volcanic gas plumes: The case of lásca volcano, Northern Chile,” Remote Sens., vol. 10, no. 10, 2018, Art. no. 1514.
[4] C. Li et al., “Hyperspectral unmixing with robust collaborative sparse regression,” Remote Sens., vol. 8, no. 7, 2016, Art. no. 588.
[5] L. Mu et al., “Arctic-wide sea ice thickness estimates from combining satellite remote sensing data and a dynamic ice-ocean model with data assimilation during the cryosat-2 period,” J. Geophys. Res.: Oceans, vol. 123, no. 11, pp. 7763–7780, 2018.
[6] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, “Graph convolutional networks for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 7, pp. 5966–5978, Jul. 2021.
[7] D. Hong et al., “More diverse means better: Multimodal deep learning meets remote-sensing imagery classification,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 5, pp. 4340–4354, May 2021.
[8] X. Wu, D. Hong, and J. Chanussot, “Convolutional neural networks for multimodal remote sensing data classification,” IEEE Trans. Geosci. Remote Sens., vol. 60, 2022, Art. no. 5517010.
[9] X. Liu, D. Hong, J. Chanussot, B. Zhao, and P. Ghamisi, “Modality translation in remote sensing time series,” IEEE Trans. Geosci. Remote Sens., vol. 60, 2022, Art. no. 5401614.
[10] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, “An augmented linear mixing model to address spectral variability for hyperspectral unmixing,” IEEE Trans. Image Process., vol. 28, no. 4, pp. 1923–1938, Apr. 2019.
[11] J. L. Harris, “Diffraction and resolving power,” JOSA, vol. 54, no. 7, pp. 931–936, 1964.
[12] C. Dong et al., “Learning a deep convolutional network for image super-resolution,” in Proc. Eur. Conf. Comput. Vis., 2014, pp. 184–199.
[13] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in Proc. Eur. Conf. Comput. Vis., 2016, pp. 391–407.
[14] 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., 2016, pp. 1646–1654.
[15] B. Lim, S. Soo, 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, 2017, pp. 136–144.
[16] J. Jiao, W.-C. Tu, S. He, and R. W. H. Lau, “Formresnet: Formatted residual learning for image restoration,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2017, pp. 1034–1042.
[17] C. Ledig et al., “Photo-realistic single image super-resolution using a generative adversarial network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2017, pp. 4681–4690.
[18] X. Wang et al., “Esrgan: Enhanced super-resolution generative adversarial networks,” in Proc. Eur. Conf. Comput. Vis., 2018, pp. 63–79.
[19] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, “Swinir: Image restoration using swin transformer,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops, 2021, pp. 1833–1844.
[20] L. Liebel and M. Kümmer, “Single-image super resolution for multispectral remote sensing data using convolutional neural networks,” ISPRS-Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci., vol. 41, pp. 883–890, 2016.
[21] J. M. Haut, R. Fernandez-Beltran, M. E. Paoletti, J. Plaza, and A. Plaza, “Remote sensing image superresolution using deep residual channel attention,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 11, pp. 9277–9289, Nov. 2019.

[22] X. Dong, X. Sun, X. Jia, Z. Xi, L. Gao, and B. Zhang, “Remote sensing image super-resolution using novel dense-sampling networks,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 2, pp. 1618–1633, Feb. 2021.

[23] Y. Xiong et al., “Improved SRGAN for remote sensing image super-resolution across locations and sensors,” Remote Sens., vol. 12, no. 8, 2020, Art. no. 1263.

[24] J. Yao et al., “Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution,” in Proc. 16th Eur. Conf. Comput. Vis., 2020, pp. 208–224.

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

[26] L. Salgueiro Romero, J. Marcello, and V. Vilaplana, “Super-resolution of sentinel-2 imagery using generative adversarial networks,” Remote Sens., vol. 12, no. 15, 2020, Art. no. 2424.

[27] A. G. Howard et al., “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” 2017, arXiv:1704.04861.

[28] M.-H. Guo et al., “Visual attention network,” 2022, arXiv:2202.09741.

[29] C. Szegedy et al., “Inception-v4, inception-resnet and the impact of residual connections on learning,” in Proc. 31st AAAI Conf. Artif. Intell., 2017, pp. 4278–4284.

[30] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 2261–2269.

[31] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv., 2015, pp. 234–241.

[32] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in Proc. Eur. Conf. Comput. Vis., 2016, pp. 694–711.

[33] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. Int. Conf. Learn. Represent., 2015, pp. 1–14.

[34] J. Jiang, H. Sun, X. Liu, and J. Ma, “Learning spatial-spectral prior for super-resolution of hyperspectral imagery,” IEEE Trans. Comput. Imag., vol. 6, pp. 1082–1096, May 2020.

[35] Y. Zhang et al., “Image super-resolution using very deep residual channel attention networks,” in Proc. Eur. Conf. Comput. Vis., 2018, pp. 294–310.

[36] X. Wang, L. Xie, C. Dong, and Y. Shan, “Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops, Oct. 2021, pp. 1905–1914.

[37] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A deep convolutional encoder-decoder architecture for image segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

Yujia Fu received the B.S. degree in information management and information system from the Northeast Forestry University, Harbin, China, in 2020. She is currently working toward the Ph.D. degree in forestry engineering with the College of Information and Computer Engineering, Northeast Forestry University, Harbin, China, under the supervision of Prof. M. Wang. Her research interests include remote sensing image processing and deep learning.

Xiangrong Zhang received the Ph.D. degree in management science and engineering from the Harbin Institute of Technology, Harbin, China, in 2016. She is currently an Associate Professor with the School of Management, Heilongjiang Institute of Technology, Harbin, China. Her research interests include data mining, natural language processing, and deep learning.

Mingyang Wang received the Ph.D. degree in management science and engineering from the School of Economics and Management, Harbin Institute of Technology, Harbin, China, in 2011. She is currently a Professor with the College of Information and Computer Engineering, Northeast Forestry University, Harbin, China. She has authored/coauthored a number of publications in international journals and conferences. Her research interests include image processing, deep learning, and data mining.

Prof. Wang is a reviewer of several international journals.