IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2019.DOI

An Unsupervised Remote Sensing Single-Image Super-Resolution Method Based on Generative Adversarial Network

NING ZHANG[1,2], YONGCHENG WANG[1], XIN ZHANG[1,2], DONGDONG XU[1,2], AND XIAODONG WANG[1]

[1]Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
[2]School of Optoelectronics, University of Chinese Academy of Sciences, Beijing 100049, China

Corresponding author: Yongcheng Wang (e-mail: wangyc@ciomp.ac.cn).

This work was supported by the National Natural Science Foundation of China under Grant 11703027.

ABSTRACT Image super-resolution (SR) technique can improve the spatial resolution of images without upgrading the imaging system. As a result, SR promotes the development of high resolution (HR) remote sensing image applications. Many remote sensing image SR algorithms based on deep learning have been proposed recently, which can effectively improve the spatial resolution under the constraints of HR images. However, images acquired by remote sensing imaging devices typically have lower resolution. Hence, an insufficient number of HR remote sensing images are available for training deep neural networks. In view of this problem, we propose an unsupervised SR method that does not require HR remote sensing images. The proposed method introduces a generative adversarial network (GAN) that obtains SR images through the generator; then, the SR images are downsampled to train the discriminator with low resolution (LR) images. Our method outperformed several methods in terms of the quality of the obtained SR images as measured by 6 evaluation metrics, which proves the satisfactory performance of the proposed unsupervised method for improving the spatial resolution of remote sensing images.

INDEX TERMS Image super-resolution, unsupervised learning, remote sensing, generative adversarial network

I. INTRODUCTION

HIGH resolution remote sensing images play an important role in resource exploration, environmental monitoring and military reconnaissance. However, overcoming the limitations of imaging sensors is time-consuming and extremely expensive, thereby rendering image super-resolution (SR) technology a feasible and economical approach for improving the resolution of remote sensing images. Image super-resolution refers to the estimation of high resolution (HR) image from one or more low resolution (LR) observations of the same scene, for which digital image processing techniques are typically employed. The first SR strategy was motivated by the requirement to improve the resolution of Landsat remote sensing images [1]. Since then, remote sensing image super-resolution has become one of the most important applications of SR technology. SR methods can be distinguished by the number of input images as single-image super-resolution (SISR) methods and multiple-image or multi-frame super-resolution methods [2]. Both single-image super-resolution and multi-frame super-resolution can be regarded as ill-posed problems since we may reconstruct more than one SR image that is similar to the original HR image. Only a few traditional SISR methods are available because minimal information could be used from a single LR image to infer its high-resolution counterpart. In contrast, multiple-image super-resolution typically utilizes several LR observations from various angles, and these LR images contain more information, namely, they are regarded as constraints that approximate the real solution. However, due to the difficulty of obtaining images that satisfy the requirements for SR in remote sensing, single-image super-resolution is usually adopted.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2972300, IEEE Access

The rapid development of deep learning has had a profound influence in image processing and has made SISR a hot issue, especially when involving convolutional neural networks (CNNs) and generative adversarial networks (GANs). Since Chao [3] introduced CNN into image super-resolution, many researchers have focused on deep-learning-based methods and have proposed excellent networks, such as VDSR [4], RED [5], and RCAN [6]. These models were designed to reconstruct natural images from LR-HR image pairs and yielded outstanding results. These models usually require large datasets for training the complex network and overcoming the overfitting problem. However, in contrast to natural images, sufficient HR remote sensing images are not available, especially if the cost of upgrading imaging sensors is extremely high. Moreover, it is not feasible to directly reconstruct remote sensing images using models that were pre-trained by natural images. As shown in Fig. 1, we use natural images to train SRResNet [7] and to test the model with remote sensing images. Using natural images for training can lead to unexpected distortion due to the differences in degradation between natural images and remote sensing images. Although [8] [9] [10] fine-tuned the pre-trained network with small datasets to reconstruct SR images based on transfer learning, they still required HR images. These challenges render highly difficult in the reconstruction of remote sensing images by using a deep neural network for supervised learning.

FIGURE 1. Distortion results of remote sensing images that were super-resolved by SRResNet after it was trained with natural images.

The lack of labeled data has led to the emergence of unsupervised learning methods such as autoencoders [11] [12], deep belief networks [13] and generative adversarial networks. The original GAN can generate images with inputs of random noises. Ledig [7] successfully applied GAN to SR via a supervised approach (SRGAN). [14] used deep CNN to capture a great deal of low-level image statistics prior for image reconstruction tasks. It first came up with the idea of downsampling SR images. And followed [14], Haut [15] developed an unsupervised strategy for the generation of SR images without HR images. Therefore, inspired by [7], [14], and [15], we propose an unsupervised GAN network for reconstructing remote sensing images without HR labels. Our method constructs an encoder-decoder network as a generator to super-resolve remote sensing images, and downsamples the SR images via average pooling to obtain LR-size images for training the discriminator. Compared with SRGAN, the proposed unsupervised neural network does not require HR remote sensing images in training. In contrast to [14], we take the downsampling strategy but use pooling layers; and the model is trained with external dataset rather than exploiting the self-similarity in one image. When compared to [15], it simplifies the downsampling to render the model more adaptive and can handle various degradations; it also uses a discriminator to super-resolve finer texture details. Our unsupervised GAN model can divide the SR process into training and testing. Once the training process has been completed, it can be put into practical use. However, [15] requires tens of thousands of iterations for the reconstruction of an image.

In summarize, the main contributions of this work are as follows:

- An unsupervised model that is based on GAN was proposed and it realized a new state-of-the-art performance for unsupervised remote sensing image SR.
- The encoder-decoder structure of the generator extracts and encodes the features of LR images step by step for the generation of SR images. This structure ensures that more information can be reserved during the downsampling process for unsupervised learning.
- Using average pooling, the downsampling process was simplified to adapt various degradations of remote sensing images and to improve the generalization performance of the model.
- By introducing a discriminator and optimizing the network with a new loss function, this method can construct more texture details and obtain more precise results.

The remainder of this paper is organized as follows: Section II describes related works on SR background, GAN, and unsupervised learning. Section III describes the methodology and network structure of our proposed method, including downsampling and the loss function. Section IV presents the implementation details and the experimental results with a discussion, and concluding remarks are provided in Section V.

II. RELATED WORK

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2972300, IEEE Access

A. IMAGE SUPER-RESOLUTION BACKGROUND

In 1960s, Harris [16] and Goodman [17] introduced the concept of "super-resolution" by solving the diffraction problem in optical systems and, therefore, established the foundation of SR. However, this technique was not successfully implemented until 1984, when Tsai and Huang [18] obtained an HR image from several LR images in the Fourier domain. This technique attracted substantial attention from researchers and initiated practical applications of SR. Initially, many researchers [19] [20] followed [18] and super-resolved an image in the transform domain due to the simple relationship between HR and LR images and the computational efficiency. However, these methods require global motion in LR observations and cannot introduce prior information. As a result, spatial-domain methods became the main trend. Due to the shortage of multiple images in real scenarios, SISR methods are generally preferred. Fig 2 illustrates the taxonomy of SR techniques according to the number of input images and the processing domain.

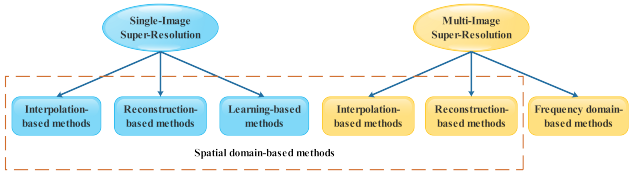


FIGURE 2. Taxonomy of SR techniques.

There are three types of SISR algorithms: interpolation-based, reconstruction-based and learning-based methods. Interpolation-based algorithms are the most basic SISR methods, which include the bilinear, bicubic and Lancozs [21]. Interpolation of images via these algorithms is typically conducted as pre-processing for reconstruction-based methods and learning-based methods. Reconstruction-based algorithms usually exploit prior information and form image models as constraints for the reconstruction of HR images, such as iterative back-projection (IBP) [22], gradient profile prior [23], deconvolution [24] [25] [26], and regularization [27] [28]. Learning-based methods use external or internal information to locally estimate the HR details. Internal similarities [29] [30] [31] [32] [33] within the same image can be regarded as several multi-scale LR observations and contain more information. But SR techniques that use external information often require datasets for mapping the relationship between LR and HR images. Via sparse coding [34] [35] [36], an image signal can be well-represented as a sparse linear combination of elements from a suitable over-complete dictionary. Compressive sensing [33] [37] [38] methods aim at identifying a dictionary that represents HR image patches sparsely. Deep-learning-based methods [3] [39] [40] [41] [42] [43] [44] [45] [46] [47] [48] [49] [50] [51] [52] utilize massive images and image priors to learn the map between LR and HR images and have yielded the state-of-the-art results in recent years. Various methods such as [53] [54] [55] [56], use LR images for training and can be regarded as hybrids of reconstruction-based and learning-based methods.

Deep-learning-based SR methods have been under development for several years and have been employed in many application scenarios; one of the most important applications is in remote sensing. Ducournau [8], Luo [9] and Ma [10] used the transfer learning strategy, namely, they fine-tuned the available models, such as SRCNN, VDSR, and SRGAN, to reconstruct optical remote sensing images. Lei [57] designed a multi-fork structure for the extraction of both local details and global environmental priors for remote sensing image SR. Liu [58] realized image SR and image colorization synchronously with a multi-task learning deep neural network. Pan [59] applies a residual backprojection block, which utilizes residual learning and backprojection, to super-resolve remote sensing images. Ma [60] realized remote sensing image super-resolution in the spatial domain by incorporating recursive Res-Net and wavelet transform. In hyperspectral image super-resolution, Li [61] and Hu [62] combined CNN with traditional methods for the super-resolution of image and spectrum, respectively. Mei [63] improved the spatial resolution of hyperspectral images with 3D-CNN and Hao [64] even utilized SR images for image classification. Overall, deep-learning-based methods have become the mainstream methods of SR.

B. GENERATIVE ADVERSARIAL NETWORKS

Generative adversarial network (GAN) was proposed by Ian Goodfellow et al. [65] in 2014 for generating images from random noises. GANs are based on a minimax two-player game in which the generator network must compete against an adversary. The training process drives the discriminator to learn to correctly classify samples as real or fake. Meanwhile, the generator attempts to fool the discriminator into believing that its outputs are real. The proposal of GAN and its variants has had a substantial influence on image processing and deep learning. Many networks use GANs for image generation [66], image translation [67], image style transfer [68], and image SR [7] [69] [70] [71] [72] in either supervised ways or unsupervised ways.

C. UNSUPERVISED LEARNING

The distinction between supervised learning and unsupervised learning is not formally or rigidly defined because there is no objective test for distinguishing whether a value is a feature or a target that is provided by a supervisor [73]. In image SR, unsupervised learning can be informally defined by the specific method. The "Zero-Shot" SR (ZSSR) model [74] can exploit the internal recurrence of information within a single image. Without any prior image examples or prior training, ZSSR is referred to as an unsupervised SR method. Zhao [72] used bi-cycle network DNSR to train both the degradation and SR reconstruction. Via this approach, LR images can be obtained by the model and for the super-resolution of HR images, making DNSR an unsupervised SR model. Yuan [71] proposed a Cycle-in-Cycle GAN. First, a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2972300, IEEE Access

GAN with two discriminators was employed for image denoising; then another GAN was used for image SR. This process does not require LR and HR image pairs; and therefore it is an unsupervised SR method. Another instance is [15]: this model was proposed for remote sensing image SR. It calculated the loss between downsampled SR images and LR images instead of SR and HR images. In this paper, we adopt the dawnsampling strategy and use a GAN to super-resolve remote sensing images. First, in remote sensing, where paired data are unavailable, it is essential to develop a method that does not require HR images. The downsampling strategy can address this problem and introduces external information for training CNNs. Second, a GAN can restrain the SR results with discriminators to realize promising performance. It comes out that our method outperforms state-of-the-art unsupervised remote sensing image SR.

III. METHOD

A. METHODOLOGY

Traditionally, SR models received LR images as input, calculated the loss between SR and HR images, and used the loss to update the network parameters, as illustrated in (1):

$$\min Loss(I^{SR}, I^{HR}) \tag{1}$$

These supervised methods typically require many HR images as training reference, while in remote sensing, there are no adequate HR images. The use of a small amount of data may lead to overfitting when training deep neural networks. It remains difficult to meet the requirement as the number of HR remote sensing images that are available is small, even though transfer learning can address this problem to some extent. Therefore, our proposed method aims at super-resolving LR images without HR labels. As illustrated in Fig. 3, an interpolated LR image I^{ILR} with HR-size is fed into the generator, which can recover an SR image I^{SR}. Before inputting the SR image into the discriminator, it was downsampled to LR-size to simulate the downsampling procedure with pooling layers. To accommodate the design of this approach, the loss function can be re-formulated as (2):

$$\min Loss(I^{SR}, I^{HR}) \rightarrow \min Loss(I^{SR'}, I^{LR}) \tag{2}$$

The loss function consists of the output of the discriminator and other indices (see Section III-C). The training process does not require HR images, thereby realizing the objective of unsupervised learning. LR images and HR images can have C color channels, and with scale factor r, they are represented as real-valued tensors of sizes $H \times W \times C$ and $rH \times rW \times C$, respectively. The corresponding $I^{SR'}$ and I^{SR} are of sizes $H \times W \times C$ and $rH \times rW \times C$.

B. STRUCTURE OF THE NETWORK

Following SRGAN, we take the structure of a generator and a discriminator. And we use an improved encoder-decoder structure [5] as the generator and also fine-tune the discriminator to adapt to the remote sensing image characteristics.

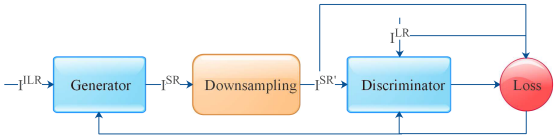


FIGURE 3. Proposed unsupervised SR method.

The largest advantage is that we introduce pooling layers for downsampling the SR images for unsupervised learning.

1) Generator

The residual encoder-decoder network [5] is designed for supervised SR, and there is only one convolutional layer for feature extraction. However, the first convolutional layer, which has a small kernel size (k=3), cannot extract global features and the large stride (s=2) may cause information loss. Thus, it cannot be directly applied to unsupervised SR since unsupervised SR requires higher information preservation. We modified this model as our generator for generating the S-R images. As illustrated in Fig. 4, we employed convolutional layers with larger kernel sizes (k=7, 5, 3) followed by ReLU [75]. This module can extract global features from LR images and preserve more information than the module that consists of filters with a smaller kernel size. Then, the next layers constitute the encoder, which has a small stride of 1. The small stride is also designed to preserve more information for reconstructing SR images without HR labels. The decoder is composed of deconvolutional layers [76], which are also called transpose convolution layers, for recovering the image details. To maintain the symmetric structure, three additional deconvolutional layers were added for reconstructing SR images. The most important part is residual learning because the subtle details of the image contents may be lost during convolution and the residual is able to transmit the initial message to subsequent layers. Since this network is not highly deep, only one skip connection is added before the convolutional layer and after its corresponding mirrored deconvolutional layer. Through this generator, an SR image can be obtained.

2) Downsampling

To realize unsupervised SR, downsampling is indispensable. Convolutional layers are used to extract high-frequency features, such as edges, instead of preserving most of the image information. Max-pooling cannot maintain the information consistency between an SR image and its corresponding downsampled counterpart. Hence, we used average-pooling to map the SR image to its LR-size because the output of the average-pooling is determined by the kernel size and contains the information of neighbors. Therefore, average-pooling can preserve more information of the image to be downsampled. Besides, average-pooling can be considered as a mean filter,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2972300, IEEE Access

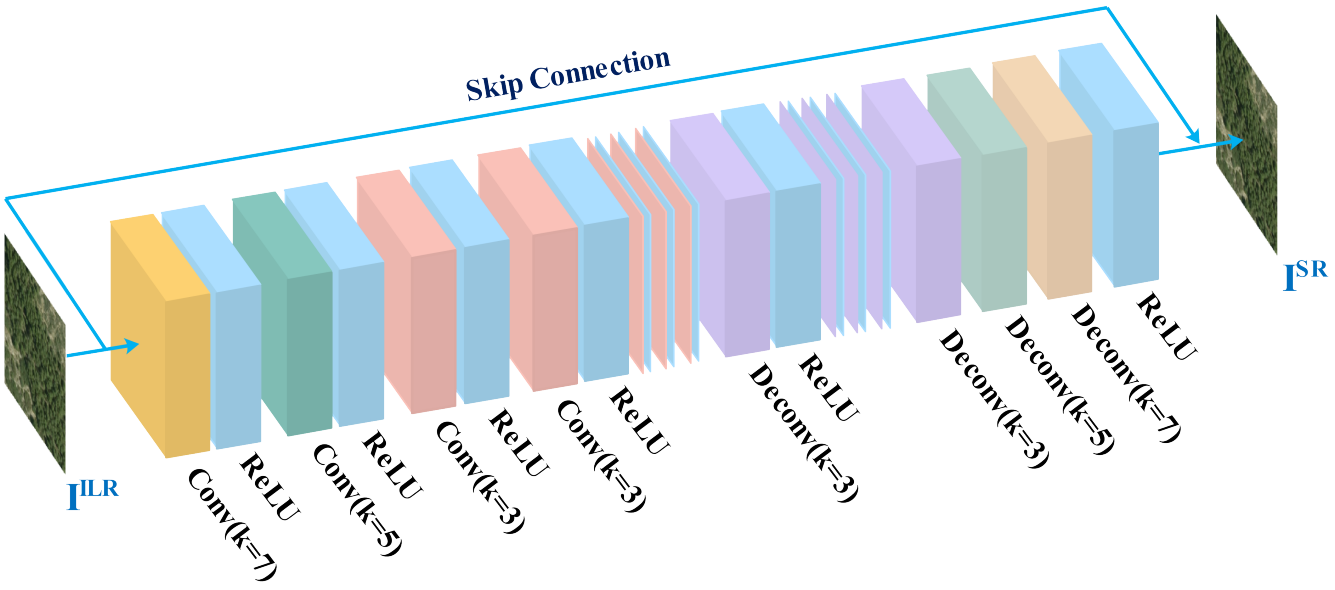


FIGURE 4. Structure of the generator with symmetric convolutional and deconvolutional layers.

which has a stride that equals to the window size. The mean filter is typically regarded as a low-pass filter and it filters out the high-frequency details of the HR image, which is similar to the imaging procedure. Meanwhile, downsampling was realized with the stride of average-pooling. Moreover, in remote sensing, the degradation of LR images is typically complicated and cannot be easily formulated, which forces the model to improve its generalization by adopting average-pooling during the training process regardless of the type of degradation.

3) Discriminator

In contrast to SRGAN, our discriminator attempts to discriminate downsampled SR images from LR images, and the images to be processed are remote sensing data. Therefore, we have to fine-tune the discriminator to adapt to their characteristics. Remote sensing images usually contain many objects since they were captured from a distance. In addition, the average gray values of images differ substantially among scenes. We found that the discriminator of SRGAN cannot distinguish images with high average gray values because the sigmoid function saturates across most of their domains [73], as saturation occurs when most of the pixel values are too large. To overcome this problem, a batch normalization (BN) layer was added before the sigmoid, as illustrated in Fig. 5.

C. LOSS FUNCTION

Instead of minimizing the mean square errors or L_1 loss between I^{LR} and downsampled $I^{SR'}$, we use a robust loss function that is based on perceptual loss function from [77]. This loss was formulated by the weighted sum of the content loss, adversarial loss and TV loss components, as expressed in (3):

$$L_G = L_{image} + L_{perception} + L_{Adv} + 2 \times 10^{-8} \cdot L_{TV} \quad (3)$$

These components are described in detail in the following.

1) Content loss

The content loss consists of the image loss and the perceptual loss. Previous works [6] [44] [48] [49] [51] [52] showed that the L_1 loss outperforms the L_2 loss. Therefore, the image loss was calculated using the L_1 norm as (4):

$$L_{image} = \frac{1}{r^2WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} ||I_{x,y}^{LR} - I_{x,y}^{SR'}||_1 \quad (4)$$

However, according to the VGG loss [77], [78], [79] that SRGAN employed, the features of SR and HR image are closer in terms of perceptual similarity. In addition, to learn pleasing images, the loss should also be perceptually motivated by SSIM [80]. Therefore, the losses of VGG and SSIM constitute the perception loss. The VGG loss was formulated with the feature map of pre-trained VGG16 [81] [82] as (5):

$$L_{VGG} = \frac{1}{W'H'} \sum_{x=1}^{W'} \sum_{y=1}^{H'} ||\Phi(I^{LR})_{x,y} - \Phi(I^{SR'})_{x,y}||_1 \quad (5)$$

where $\Phi(\cdot)$ is the feature map that is obtained by the last convolution (after ReLU) and before the last max-pooling layer within the VGG16 network, and W' and H' denote the size of the feature map Φ. SSIM for an image patch with central pixel p is defined as (6):

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2972300, IEEE Access

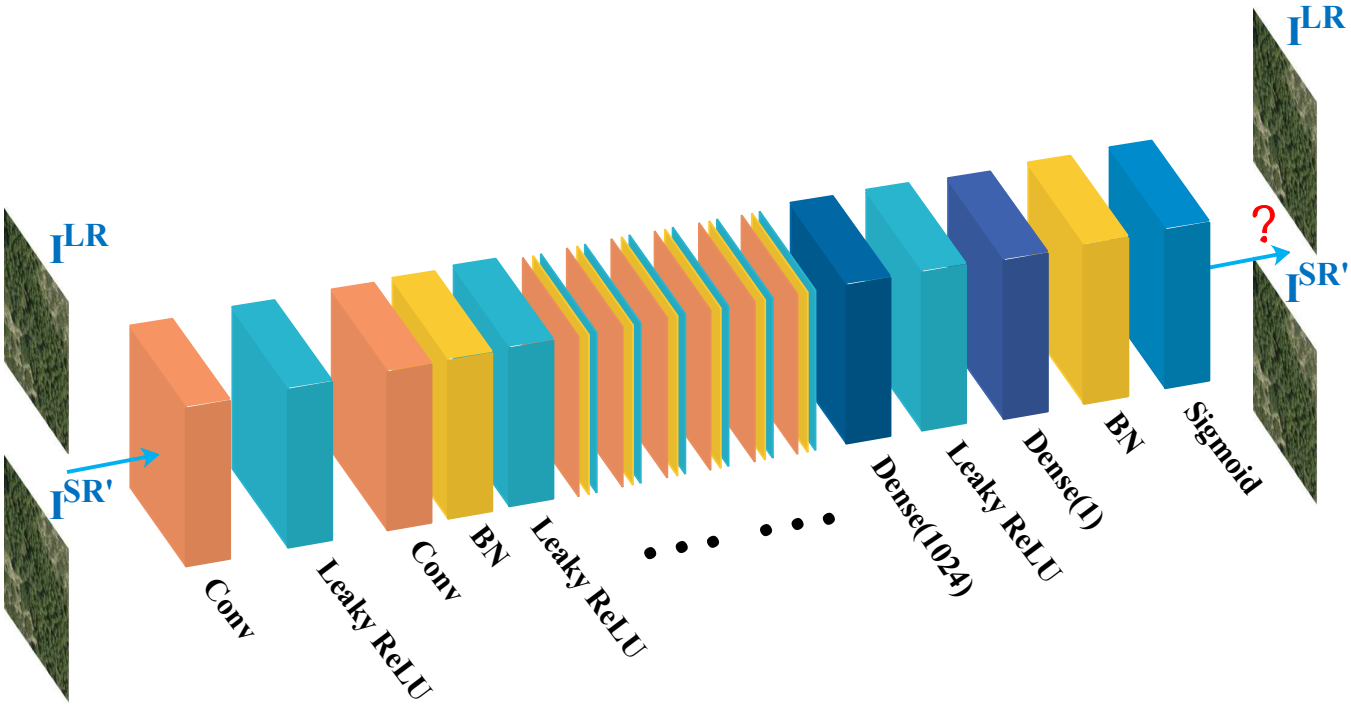


FIGURE 5. Structure of the discriminator, which has a BN layer before sigmoid.

$$\mathrm{SSIM}(p) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \cdot \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \tag{6}$$

Thus, the loss function for SSIM can be then expressed as (7):

$$L_{SSIM} = 1 - \mathrm{SSIM}(I_{x,y}^{SR'}, I_{x,y}^{LR}) \tag{7}$$

Finally, the perception loss was formulated as (8):

$$L_{perception} = 0.003 \times L_{VGG} + 0.1 \times L_{SSIM} \tag{8}$$

2) Adveisarial loss

The score that is obtained from discriminator can facilitate the discrimination of $I^{SR'}$ and I^{LR} as part of the adversarial loss. In addition, the generator loss should try to optimize the generator and fool the discriminator simultaneously. As a result, our adversarial loss is designed for the discriminator and the generator. For better gradient behavior, the difference between $D(I^{SR'})$ and $D(I^{LR})$ was calculated. The final adversarial loss was formulated as the sum of cross-entropies between $D(I^{SR'})$ and the true labels and between $D(I^{SR'})$ and $D(I^{LR})$ as (9):

$$\begin{aligned} L_{Adv} = -\frac{1}{n}\sum_{i=1}^{n}(w_1 \cdot \log(D(I^{SR'})_i) \\ + w_2 \cdot (\log(D(I^{LR})_i) + \log(1 - D(I^{SR'})_i))) \end{aligned} \tag{9}$$

Here, $D(I^{SR'})$ and $D(I^{LR})$ are scores (outputs) of $I^{SR'}$ and I^{LR}, respectively, and w_1 and w_2 are weights for balancing the adversarial loss.

The discriminator loss was calculated by the cross-entropies of $D(I^{SR'})$ and the fake label, and of $D(I^{LR})$ and the true label, as expressed in (10):

$$L_D = -\frac{1}{2n}\sum_{i=1}^{n}(\log(1 - D(I^{SR'})_i) + \log(D(I^{LR})_i)) \tag{10}$$

3) Total variation loss

Rudin [83] et al. observed that the total variation (TV) of noisy images was significantly larger than that of noiseless images. Hence, by minimizing the total variation loss, it is possible to remove the noise from images and to preserve the edges. Following [77], we introduced the TV loss as a regularization term, as expressed in (11):

$$L_{TV} = \sum_{i,j}((I_{x,y-1} - I_{x,y})^2 + (I_{x+1,y} - I_{x,y})^2)^{\frac{\beta}{2}} \tag{11}$$

Here, β is a small scalar to ensure differentiability. The first term and the second term are the differences in the horizontal and vertical directions, respectively.

IV. EXPERIMENT

A. DATASET

We perform experiments on three widely used remote sensing benchmark datasets, UC Merced dataset [84], NWPU-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2972300, IEEE Access

RESIS45 [85] and WHU-RS19 [86]. The employed training repositories are described in the following.

UC Merced dataset contains 21 land-use scenes, agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis courts images. Each class is composed of 100 images with the size of 256×256 pixels and the spatial resolution of about 30 cm.

NWPU-RESIS45 is a large dataset created by Northwestern Polytechnical University. It consists of 45 kinds of land class with 700 images per class. And the size of each image is also 256×256 pixels.

WHU-RS19 was collected from Google Earth by the remote sensing group of Wuhan University. It is composed of 19 scenes with a large size of 600×600 pixels each image. The total number is 950 images, 50 to 61 images each category.

B. IMPLEMENTATION DETAILS

For training, random patches of size 128×128 from the training set were downsampled via bicubic resampling. Considering the noise-free schemes that have been presented in other approaches, such as [3], [4], [7], and [44], no additional noise is added in our approach. As HR-sized inputs are needed, I^{LR} was interpolated via bicubic interpolation. We train our model with the ADAM optimizer by setting $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The minibatch size was set as 64 and the learning rate was initialized as 5×10^{-4}. We implemented our networks in the PyTorch framework and trained them using an NVIDIA Titan V GPU for about 4 days. The model was tested on the same testing set as [15]. Note that these images were not used during training. Self-ensemble was used during testing, and the tested and evaluation methods were obtained from the MATLAB toolbox [87] [88].

The evaluation results of this comparison are presented in Section IV-C, and Table 1 presents the metrics of SR image evaluation that are generally used. Note that RMSE is just for representing the fomulas in ERGAS and PSNR and not used for image evaluation because it has the same trend with PSNR.

C. RESULTS

As mentioned, the testing dataset is sampled from the training set and consists of 12 images, including agricultural, agricultural2, airplane, baseball, bridge, circular-farmland, harbor, industry, intersection, parking, residential and road (see Fig. 6).

The proposed method was compared with reconstruction-based, learning-based and a hybrid of these two methods as mentioned in section II, including supervised and unsupervised methods. Table 2, Table 3 and Table 4 present the quantitative results. Two scale factors($\times 2$ and $\times 4$) are considered. The qualitative results are presented in Fig. 7, Fig. 8 and Fig. 9.

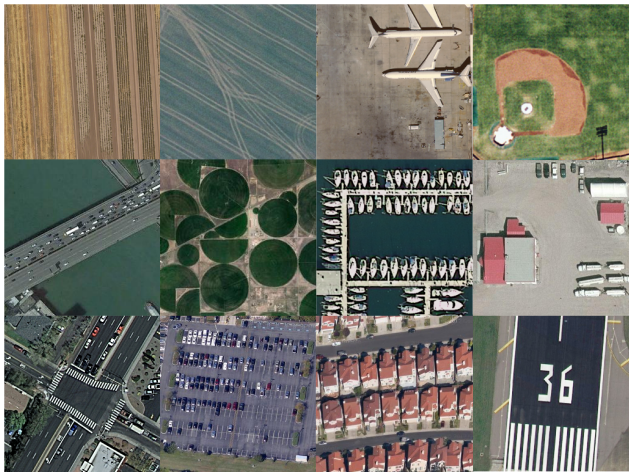


FIGURE 6. Testing dataset.

1) Comparison with reconstruction-based methods

Table 2 presents the average assessment results of six reconstruction-based methods. The best results are highlighted.

According to Table 2, the proposed method, together with IBP, DLU and DRE, outperformed FSR. GPP provided the second-best PSNR when considering scale factor 4 but did not outperform IBP, DLU and DRE with a scale factor of 2, possibly because these methods cannot better handle a larger scale than GPP with the global gradient. However, GPP requires many computations and a long runtime on large-sized images, whereas the proposed method can super-resolve an image of any size once the network has finished training and yields the best results in terms of various metrics.

2) Comparison with learning-based methods

Most of the learning-based methods require HR images as references. In Table 3, quantitative results are presented for nine learning-based methods. Since we were not able to obtain the source code or SR images of deep generative network for unsupervised remote sensing single-image super resolution (UGN) [15], the published numerical results were cited.

Compared with the learning-based methods, our method did not achieve all the best results; however it performed more consistently, because it obtained the best average results on the $Q2^n$ and SSIM metrics, and the second-best value on the ERGAS, and PSNR metrics. It should be noted that SAM performed worse (second best) when a scale factor of 4 was used, and the sCC and $Q2^n$ metric were not considered in [15]. Our method outperformed other learning-based methods: sparse coding (SIP, SDS, MDL), neighborhood embedding (ANR, GLR), and deep convolutional neural network (CNN). Limited by the requirement for supplemental neighborhood information in neighborhood embedding, the boundaries of these SR images are missing. Therefore, the ERGAS and PSNR metrics were affected to some extent.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2972300, IEEE Access

TABLE 1. Metrics for Image Evaluation

Metrics	Formula	Description
SAM	$\mathrm{SAM}(I^{SR}, I^{HR}) = \frac{1}{N}\sum_{i=1}^{N}\arccos\frac{I_i^{SR}\cdot I_i^{HR}}{\|\|I_i^{SR}\|\|\cdot\|\|I_i^{HR}\|\|}$	Spectral angle mapper is used for spectral assessment. It considers each band as a coordinate axis and computes the average angle between the pixels of I^{HR} and I^{SR}. the ideal value of SAM is 0.
RMSE	$\mathrm{RMSE}(I^{SR}, I^{HR}) = \sqrt{\frac{1}{C\cdot N}\sum_{j=1}^{C}\sum_{i=1}^{N}((I_i^{HR})_j - (I_i^{SR})_j)^2}$	Root mean square error measures the distance between the data predicted by a model I^{SR} and the original data I^{HR}, which is an error measure.
ERGAS	$\mathrm{ERGAS}(I^{SR}, I^{HR}) = \frac{100}{r}\sqrt{\frac{1}{C}\sum_{j=1}^{C}\frac{\mathrm{RMSE}((I_i^{HR})_j,(I_i^{SR})_j)}{(\bar{I}_i^{HR})_j}}$	Erreur relative globale adimensionnelle de synthese measures the quality of obtained SR taking into account the scaling factor to evaluate I^{SR}. The best ERGAS index value means less distortion.
CC	$\mathrm{CC}(I^{SR}, I^{HR}) = \frac{\sigma_{I^{HR},I^{SR}}}{\sigma_{I^{HR}}\cdot\sigma_{I^{SR}}}$	Correlation coefficient measure the linear relationship between two images normalizing their range values. The best CC value is 1, implying that both images are linearly correlated. In remote sensing, it is common to compute the spatial correlation coefficient over the edges detected by Canny or Sobel (this paper adopts Sobel) in images.
$Q2^n$	$Q2^n(I^{SR}, I^{HR}) = \frac{\sigma_{I^{HR},I^{SR}}}{\sigma_{I^{HR}}\cdot\sigma_{I^{SR}}}\cdot\frac{2\bar{I}^{HR}\cdot\bar{I}^{SR}}{(\bar{I}^{HR})^2(\bar{I}^{SR})^2}\cdot\frac{2\sigma_{I^{HR}}\cdot\sigma_{I^{SR}}}{(\sigma_{I^{HR}})^2(\sigma_{I^{SR}})^2}$	Universal Image Quality Index gathers three different properties in the image evaluation: correlation, luminance and contrast. $Q2^n$ [89] extends the Universal Image Quality Index for monochrome images to multispectral and hyperspectral images through hypercomplex numbers.
SSIM	$\mathrm{SSIM}(I^{SR}, I^{HR}) = \frac{(2\bar{I}^{HR}\cdot\bar{I}^{SR}+c_1)(2\sigma_{I^{HR}}\cdot\sigma_{I^{SR}}+c_2)}{((\bar{I}^{HR})^2+(\bar{I}^{SR})^2+c_1)\cdot((\sigma_{I^{HR}})^2+(\sigma_{I^{SR}})^2+c_2)}$	Structural similarity is an extension of the Q-index in order to avoid around null values. The range of its values is [-1,1].
PSNR	$\mathrm{PSNR}(I^{SR}, I^{HR}) = 20\cdot\log_{10}\frac{255}{\mathrm{RMSE}}$	A higher peak signal-to noise ratio value represents a better image quality and its domain is all positive real numbers greater than 0.

TABLE 2. Average SR Results Obtained by Reconstruction-based Methods

Scale	Method	SAM	ERGAS	sCC	$Q2^n$	SSIM	PSNR
× 2	IBP [22]	1.024	5.514	0.9247	0.9115	0.8994	28.76
	GPP [23]	1.055	5.994	0.9079	0.8960	0.8798	28.19
	DLU [25]	1.026	5.373	0.9292	0.9113	0.8982	29.00
	DRE [28]	1.026	5.372	0.9292	0.9112	0.8981	29.00
	FSR [90]	1.084	7.002	0.8819	0.8627	0.8518	26.65
	ours	**0.914**	**5.219**	**0.9341**	**0.9168**	**0.9050**	**29.26**
× 4	IBP [22]	1.428	4.892	0.6941	0.7244	0.6835	23.66
	GPP [23]	1.384	4.761	0.6959	0.7236	0.7016	23.92
	DLU [25]	1.436	4.766	0.7054	0.7398	0.6936	23.90
	DRE [28]	1.437	4.766	0.7054	0.7398	0.6935	23.90
	FSR [90]	1.561	6.136	0.5758	0.5601	0.5748	21.63
	ours	**1.356**	**4.626**	**0.7187**	**0.7648**	**0.7136**	**24.20**

TABLE 3. Average SR Results Obtained by Learning-based Methods

Scale	Method	SAM	ERGAS	sCC	$Q2^n$	SSIM	PSNR
× 2	ANR [29]	1.164	10.82	0.6264	0.7458	0.9043	21.41
	GLR [29]	1.261	10.99	0.6478	0.7471	0.8946	21.28
	GPR [30]	1.137	8.254	0.8112	0.8037	0.7761	25.25
	SIP [34]	1.485	7.956	0.8724	0.8282	0.8479	24.60
	SDS [37]	1.497	10.16	0.8046	0.8216	0.7969	23.58
	MDL [38]	1.104	5.767	0.9209	0.9035	0.8933	28.02
	SRCNN [3]	1.083	5.871	0.9131	0.8970	0.8852	28.15
	UGN [15]	0.934	**4.366**	–	–	0.8836	**30.57**
	ours	**0.914**	5.219	**0.9341**	**0.9168**	**0.9050**	29.26
× 4	ANR [29]	1.743	8.493	0.3693	0.5940	0.7041	17.90
	GLR [29]	1.882	8.587	0.3908	0.6005	0.6936	17.81
	GPR [30]	1.454	5.218	0.9774	0.6844	0.6525	23.02
	SIP [34]	1.545	6.161	0.4960	0.6081	0.6572	20.85
	SDS [37]	1.568	6.075	0.6038	0.6518	0.6171	21.70
	MDL [38]	1.443	4.722	0.7133	0.7498	0.7089	23.97
	SRCNN [3]	1.545	5.078	0.6769	0.7301	0.6787	23.48
	UGN [15]	**1.352**	**4.193**	–	–	0.6776	**25.21**
	ours	1.356	4.626	**0.7187**	**0.7648**	**0.7136**	24.20

However, they still realized competitive results on $Q2^n$ and SSIM.

Compared with UGN, the average PSNR values of our proposed method are slightly lower, possibly because the differences in methodology between these two methods. First, UGN uses the L_2 norm as the loss function. As we all know, minimiziation of the L_2 loss is generally preferred since it maximizes the PSNR [44]. However, it also leads to overly smooth results and is not closely related to other properties of the image. In our proposed method, a comprehensive loss function was adopted. Our proposed method can realize not only high ERGAS and PSNR but also higher SAM and SSIM. Moreover, in UGN, an image generator was constructed for generating images from random noises. During this procedure, the generated image of the current iteration was used as the input of the next iteration. After tens of thousands of epochs of training, an SR image can be generated. As the number of iterations increases, the amount of information from image labels that is utilized increases.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2972300, IEEE Access

While our method introduced a massive amount of information from an external image dataset, and the training process only required hundreds of epochs. Once the training has finished, the generator can be used to super-resolve remote sensing LR images. Although UGN realized higher PSNR, it could only reconstruct one remote sensing image after the entire training procedure. The high costs in terms of time and computing resources make it impossible to be applied in practice. In contrast, after training, our method can super-resolve multiple images, and the average time consumption is approximately 0.04s for a 256×256 pixel image.

3) Comparison with hybrid methods

Some methods combine the advantages of reconstruction-based and learning-based methods. According to the average SR results in Table 4, SRSI and TSE obtained the best and the second best results (except for SAM with a scale factor of 4) because remote sensing images usually contain repeated and various scales of the same object, and these two methods exploit structural self-similarity in one image to obtain subpixel misalignments. View from qualitative results (see Fig. 8), LSE using local self-similarity super-resolved clear edges but the shape of the car is indistinguishable. This result is unfavorable for subsequent applications, such as object detection. Besides, the details in the upper left corner and upper right corner have obvious distortion. Maybe that is why the numerical results are lower. Fig. 9 also presents the visual results of SRSI, TSE and the proposed method. Although SRSI and TSE obtained sharper edges, the texture and details are smoother and the object such as cars and trees, were distorted. The numerical results demonstrate the limitations of the proposed method compared with hybrid methods, but they also reveal that we can improve our method by considering structural self-similarity in remote sensing image SR in the future.

TABLE 4. Average SR Results Obtained by Hybrid Methods

Scale	Method	SAM	ERGAS	sCC	$Q2^n$	SSIM	PSNR
× 2	SRSI [53]	0.969	**4.462**	0.9502	0.9257	0.9145	**30.49**
	LSE [54]	1.080	6.197	0.9050	0.8658	0.8875	27.02
	BDB [56]	1.308	10.36	0.7351	0.7838	0.7600	23.11
	TSE [55]	**0.866**	4.551	**0.9514**	**0.9290**	**0.9186**	30.33
	ours	0.914	5.219	0.9341	0.9168	0.9050	29.26
× 4	SRSI [53]	1.388	**4.213**	**0.7696**	**0.7943**	**0.7499**	**25.08**
	LSE [54]	1.773	6.629	0.5252	0.6126	0.6441	20.61
	BDB [56]	1.504	5.438	0.6369	0.7150	0.6635	22.62
	TSE [55]	1.378	4.353	0.7515	0.7884	0.7415	24.84
	ours	**1.356**	4.626	0.7187	0.7648	0.7136	24.20

D. MODEL ANALYSES

To evaluate the performance of our proposed model, we conduct several experiments with various numbers and kernal sizes of convolutional layers, patch sizes of input images, and downsamplinging strategy.

The RED used one convolution layer, which can extract shallow features. However, to realize unsupervised learning SR, more features need to be gathered from LR images. Therefore in our proposed model, gradually decreasing kernel sizes (7, 5, and 3) were used to extract features in various receptive fields. We further examine networks with different convolutional layers: (i) 7-5-3, (ii) 5-3 and (iii) 3. The convergence curves in Fig. 10 show that using 7-5-3 convolution layers to extract features could significantly improve the performance. To be specific, the average PSNR values that were realized by 7-5-3 are higher than those by the other two models by large margins. These results suggest that gradual utilization of the neighborhood information and the construction of a deeper network are beneficial for feature extraction.

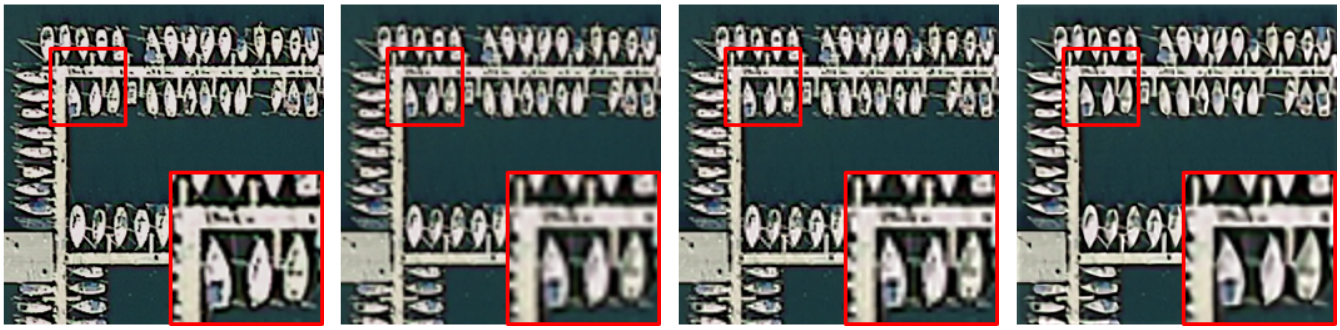
Furthermore, it is found that different patch size can affect the network performance. Random cropping of patches from LR images can be regarded as a method of data augmentation, since every epoch may obtain a different part of each original image. Theoretically, a small patch size can yield diverse training data. However, experiments show that a small patch size does not always result in better performance; see Fig. 11. Specifically, when we reduce the crop size to 96 and 64, the performances fail to surpass that with patch size 128 in terms of PSNR. Smaller patches may not contain global features, thereby leading to model degradation.

To determine whether average-pooling can outperform max-pooling, another experiment was conducted, in which max-pooling was used for downsampling instead of average-pooling. According to Fig. 12, the PSNR values that were obtained by average-pooling are higher than those obtained via max-pooling by a large margin. The two curves are not even in the same coordinate range. The red curve also shows the instability of max-pooling as well. As discussed previously, max-pooling preserves the maximum value in the neighborhood while average-pooling calculates the all the pixels in the neighborhood. As a result, more information can be preserved by average-pooling for further processing.

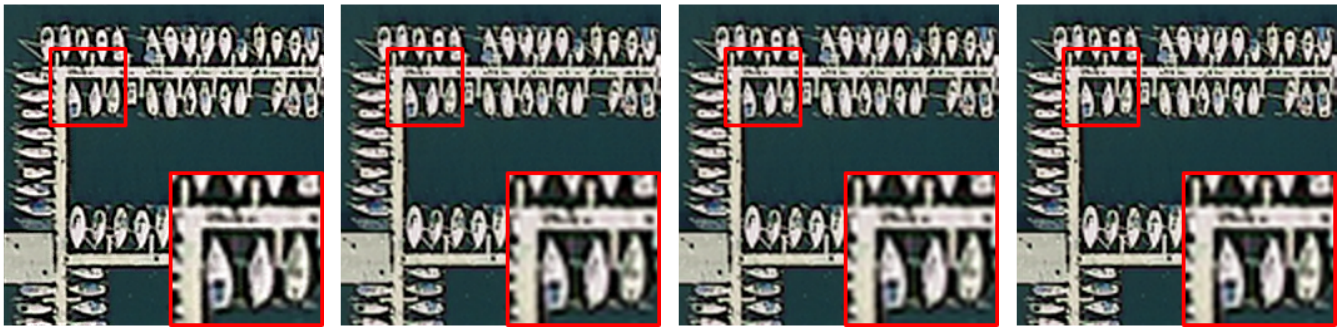
V. CONCLUSION

In this paper, an unsupervised SR method based on GAN is proposed to super-resolve single remote sensing image without HR labels. In summary, we introduced downsampling and trained a discriminator with downsampled SR images and LR images, thereby realizing unsupervised learning. The average-pooling operation downsampled the SR images without aiming at a specified level of degradation, which forces the generator to learn the relationship between the LR and HR images by training. In terms of model construction, convolution layers with gradually decreasing kernel sizes were used to extract various scales of features and to preserve more information for unsupervised SR. In addition, the loss function was improved by calculating the L_1 loss of each image and its high-level features, the SSIM loss, the cross-entropy of the discriminator output and the TV loss. Experimental results demonstrate the satisfactory performance of the proposed approach in terms of 6 metrics compared with several SR methods with two scaling factors (2 and 4), and prove the effectiveness of GAN in dealing with unsupervised

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2972300, IEEE Access



(a) HR (b) BCI (21.76dB/0.8924) (c) IBP (22.91dB/0.9179) (d) LSE (21.63dB/0.9118)



(e) MDL (22.88dB/0.9195) (f) DLU (23.42dB/0.9221) (g) DRE (23.43dB/0.9220) (h) Ours (23.83dB/0.9315)

FIGURE 7. Qualitative comparison of several SR methods with our method at a scaling factor of 2.(PSNR/SSIM)

(a) HR (b) BCI (27.52dB/0.8901) (c) IBP (27.56dB/0.8893) (d) LSE (24.73dB/0.8843)

(e) MDL (28.09dB/0.9044) (f) DLU (27.85dB/0.8955) (g) DRE (27.86dB/0.0.8955) (h) Ours (28.25dB/0.9069)

FIGURE 8. Qualitative comparison of several SR methods with our method at a scaling factor of 4.(PSNR/SSIM)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2972300, IEEE Access

(a) HR (b) SRSI (22.18dB) (c) TSE (22.03dB) (d) ours (22.04dB)

FIGURE 9. Visual results of parking lot that were obtained by SRSI, TSE and ours at a scaling factor of 4.

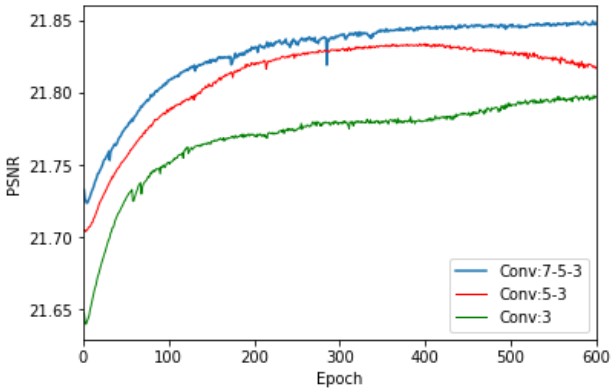


FIGURE 10. Average PSNR values of SR images with various convolutional layers.

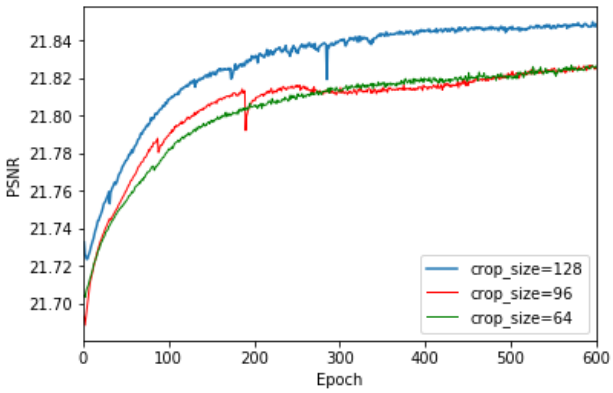


FIGURE 11. Average PSNR values of SR Images with different crop sizes.

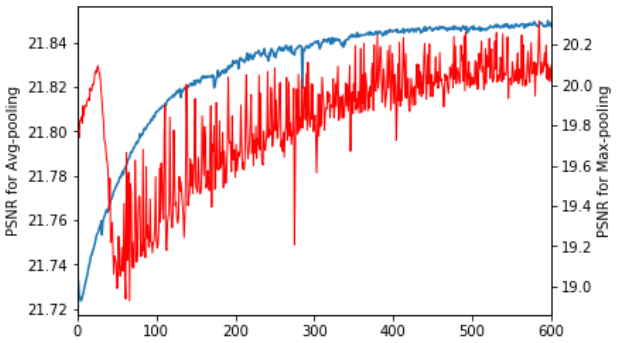


FIGURE 12. Average PSNR values of SR images with average-pooling and max-pooling.

learning problems in image super-resolution.

REFERENCES

[1] L. Yue, H. Shen, L. Jie, Q. Yuan, H. Zhang, and L. Zhang, "Image super-resolution: The techniques, applications, and future," Signal Processing, vol. 128, pp. 389–408, 2016.

[2] K. Hayat, "Super-resolution via deep learning," arXiv: 1706.09077, 2017.

[3] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," IEEE Trans Pattern Anal Mach Intell, vol. 38, no. 2, pp. 295–307, 2016.

[4] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1646–1654.

[5] X. J. Mao, C. Shen, and Y. B. Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," arXiv: 1606.08921, 2016.

[6] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in The European Conference on Computer Vision, 2018, pp. 294–310.

[7] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4681–4690.

[8] A. Ducournau and R. Fablet, "Deep learning for ocean remote sensing: an application of convolutional neural networks for super-resolution on satellite-derived sst data," in 2016 9th IAPR Workshop on Pattern Recogniton in Remote Sensing, 2016, pp. 1–6.

[9] Y. Luo, L. Zhou, W. Shu, and Z. Wang, "Video satellite imagery super resolution via convolutional neural networks," IEEE Geoscience & Remote Sensing Letters, vol. 14, no. 12, pp. 2398–2402, 2017.

[10] W. Ma, Z. Pan, J. Guo, and B. Lei, "Super-resolution of remote sensing images based on transferred generative adversarial network," in 2018 IEEE International Geoscience and Remote Sensing Symposium, 2018, pp. 1148–1151.

[11] O. Firat and F. T. Y. Vural, "Representation learning with convolutional sparse autoencoders for remote sensing," in 2013 21st Signal Processing and Communications Applications Conference, 2013, pp. 1–4.

[12] W. Cui and Z. Zhou, Q.and Zheng, "Application of a hybrid model based on a convolutional auto-encoder and convolutional neural network in objectoriented remote sensing classification," Algorithms, vol. 11, no. 1, p. 9, 2018.

[13] R. Tanase, M. Datcu, and R. Dan, "A convolutional deep belief network

for polarimetric sar data feature extraction," in 2016 IEEE International Geoscience and Remote Sensing Symposium, 2016, pp. 7545–7548.

[14] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9446–9454.

[15] J. M. Haut, R. Fernandez-Beltran, M. E. Paoletti, J. Plaza, and F. Pla, "A new deep generative network for unsupervised remote sensing single-image super-resolution," IEEE Transactions on Geoscience and Remote Sensing, vol. PP, no. 99, pp. 1–19, 2018.

[16] J. L. Harris, "Diffraction and resolving power," Journal of the Optical Society of America (1917-1983), vol. 54, no. 7, pp. 931–933, 1964.

[17] J. Goodman, Introduction To Fourier Optics. Mc Graw-Hill, 1968.

[18] R. Y. Tsai and T. S. Huang, Multiple Frame Image Restoration and Registration. Greenwich, CT, England: JAI Press, 1984.

[19] S. Rhee and M. G. Kang, "Discrete cosine transform based regularized high-resolution image reconstruction algorithm," Optical Engineering, vol. 38, no. 8, pp. 1348–1356, 1999.

[20] N. Nguyen and P. Milanfar, "A wavelet-based interpolation-restoration method for superresolution (wavelet superresolution)," Circuits Systems and Signal Processing, vol. 19, no. 4, pp. 321–338, 2000.

[21] K. Turkowski, "Filters for common resampling tasks," Graphics gems, pp. 147–165, 1990.

[22] M. Irani and S. Peleg, "Improving resolution by image registration," Cvgip Graphical Models & Image Processing, vol. 53, no. 3, pp. 231–239, 1991.

[23] J. Sun, Z. Xu, and H. Shum, "Image super-resolution using gradient profile prior," in 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[24] Q. Shan, Z. Li, J. Jia, and C. Tang, "Fast image/video upsampling," international conference on computer graphics and interactive techniques, vol. 27, no. 5, p. 153, 2008.

[25] L. B. Lucy, "An iterative technique for the rectification of observed distributions," The Astronomical Journal, vol. 79, pp. 745–754, 1974.

[26] N. Zhao, Q. Wer, A. Basarab, D. Kouame, and J. Tourneret, "Single image super-resolution of medical ultrasound images using a fast algorithm," in IEEE 13th International Symposium on Biomedical Imaging, 2016, pp. 473–476.

[27] H. Aly and E. Dubois, "Image up-sampling using total-variation regularization with a new observation model," IEEE Transactions on Image Processing, vol. 14, no. 10, pp. 1647–1659, 2005.

[28] R. C. Gonzalez and R. E. Woods, Digital Image Processing (3rd Edition), 2007.

[29] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in Proceedings of the 2013 IEEE International Conference on Computer Vision, 2013, pp. 1920–1927.

[30] H. He and W. C. Siu, "Single image super-resolution using gaussian process regression," in The 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 20–25.

[31] T. M. Chan, J. Zhang, J. Pu, and H. Huang, "Neighbor embedding based super-resolution algorithm through edge detection and feature selection," Pattern Recognition Letters, vol. 30, no. 5, pp. 494–502, 2009.

[32] H. Shen, H. Biao, W. Zaidao, and J. Licheng, "Structural-correlated self-examples based superresolution of single remote sensing image," IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, vol. 11, no. 9, pp. 3209–3223, 2018.

[33] Z. Pan, J. Yu, H. Huang, S. Hu, A. Zhang, H. Ma, and W. Sun, "Super-resolution based on compressive sensing and structural self-similarity for remote sensing images," IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 9, pp. 4864–4876, 2013.

[34] Y. Jianchao, W. John, H. Thomas, and M. Yi, "Image super-resolution via sparse representation," IEEE Transactions on Image Processing, vol. 19, no. 11, pp. 2861–2873, 2010.

[35] B. Hou, K. Zhou, and L. Jiao, "Adaptive super-resolution for remote sensing images based on sparse representation with global joint dictionary model," IEEE Transactions on Geoscience and Remote Sensing, vol. PP, no. 99, pp. 1–16, 2017.

[36] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," IEEE Transactions on Image Processing, vol. 20, no. 7, pp. 1838–1857, 2010.

[37] S. J. Sreeja and M. Wilscy, "Single image super-resolution based on compressive sensing and tv minimization sparse recovery for remote sensing images," in IEEE Recent Advances in Intelligent Computational Systems, 2013, pp. 215–220.

[38] P. Purkait and B. Chanda, "Image upscaling using multiple dictionaries of natural image patches," in Proceedings of the 11th Asian conference on Computer Vision - Volume Part III, 2012, pp. 284–295.

[39] D. Chao, C. L. Chen, and X. Tang, "Accelerating the super-resolution convolutional neural network," in The European Conference on Computer Vision, 2016, pp. 391–407.

[40] W. Shi, J. Caballero, F. Huszar, J. Totz, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1874–1883.

[41] K. M. He, X. Y. Zhang, and S. Q. Ren, "Deep residual learning for image recognition," in IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[42] J. W. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1637–1645.

[43] Y. S. Tai, J. X. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2790–2798.

[44] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 136–144.

[45] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in IEEE International Conference on Computer Vision, 2017, pp. 4799–4807.

[46] W. Lai, J. Huang, N. Ahuja, and M. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 624–632.

[47] Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration," in IEEE International Conference on Computer Vision, 2017, pp. 4539–4547.

[48] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2472–2481.

[49] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network." in IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 723–731.

[50] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1664–1673.

[51] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in The European Conference on Computer Vision, 2018, pp. 517–532.

[52] N. Ahn, B. Kang, and K. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in The European Conference on Computer Vision, 2018, pp. 517–532.

[53] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in IEEE 12th International Conference on Computer Vision, 2009, pp. 349–356.

[54] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," ACM Transactions on Graphics, vol. 30, no. 2, p. 12, 2011.

[55] J. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5197–5206.

[56] T. Michaeli and M. Irani, "Blind deblurring using internal patch recurrence," in The European Conference on Computer Vision, 2014, pp. 783–798.

[57] S. Lei, Z. Shi, and Z. Zou, "Super-resolution for remote sensing images via local-global combined network," IEEE Geoscience & Remote Sensing Letters, vol. PP, no. 99, pp. 1–5, 2017.

[58] H. Liu, Z. Fu, J. Han, S. Ling, and H. Liu, "Single satellite imagery simultaneous super-resolution and colorization using multi-task deep neural networks," Journal of Visual Communication & Image Representation, vol. 53, pp. 20–30, 2018.

[59] Z. Pan, W. Ma, J. Guo, and B. Lei, "Super-resolution of single remote sensing image based on residual dense backprojection networks," IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 10, pp. 7918–7933, 2019.

[60] W. Ma, Z. Pan, J. Guo, and B. Lei, "Achieving super-resolution remote sensing images via the wavelet transform combined with the recursive res-net," IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 6, pp. 1–16, 2019.

[61] Y. Li, H. Jing, Z. Xi, W. Xie, and J. J. Li, "Hyperspectral image super-resolution using deep convolutional neural network," Neurocomputing, vol. 266, pp. 29–41, 2017.

[62] J. Hu, Y. Li, X. Zhao, and W. Xie, "A spatial constraint and deep learning based hyperspectral image super-resolution method," in IEEE International Geoscience and Remote Sensing Symposium, 2017, pp. 5129–5132.

[63] S. Mei, X. Yuan, J. Ji, S. Wan, J. Hou, and Q. Du, "Hyperspectral image super-resolution via convolutional neural network," in IEEE International Conference on Image Processing, Sep. 2017, pp. 4297–4301.

[64] S. Hao, W. Wei, Y. Ye, E. Li, and L. Bruzzone, "A deep network architecture for super-resolution-aided hyperspectral image classification with classwise loss," IEEE Transactions on Geoscience & Remote Sensing, vol. 56, no. 8, pp. 1–14, 2018.

[65] I. Goodfellow, J. Pouget-Abadie, B. W.-F. D. Mirza, M.and Xu, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in The 27th International Conference on Neural Information Processing Systems, 2016, pp. 2672–2680.

[66] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv: 1511.06434, 2015.

[67] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.

[68] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 4401–4410.

[69] S. J. Park, H. Son, S. Cho, K. Hong, and S. Lee, "Srfeat: Single image super-resolution with feature discrimination," in The European Conference on Computer Vision, 2018, pp. 439–455.

[70] A. Bulat, J. Yang, and G. Tzimiropoulos, "To learn image super-resolution, use a gan to learn how to do image degradation first," in The European Conference on Computer Vision, 2018, pp. 185–200.

[71] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, and L. Lin, "Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks." in IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 701–710.

[72] T. Zhao, C. Zhang, W. Ren, D. Ren, and Q. Hu, "Unsupervised degradation learning for single image super-resolution," arXiv: 1812.04240, 2018.

[73] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016, http://www.deeplearningbook.org.

[74] A. Shocher, N. Cohen, and M. Irani, ""zero-shot" super-resolution using deep internal learning," in IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3118–3126.

[75] O. Shahar, A. Faktor, and M. Irani, "Space-time super-resolution from a single video," in IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 3353–3360.

[76] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," IEEE Transactions on Image Processing, vol. 13, no. 10, pp. 1327–1344, 2004.

[77] W. Shi, J. Caballero, C. Ledig, X. Zhuang, W. Bai, K. K. Bhatia, A. De Marvao, T. Dawes, D. P. O. Regan, and D. Rueckert, "Cardiac image super-resolution with global correspondence using multi-atlas patchmatch," in Medical Image Computing and Computer-Assisted Intervention, vol. 16, 2013, pp. 9–16.

[78] S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networks for human action recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 221–231, 2013.

[79] D. Dai, R. Timofte, and L. Van Gool, "Jointly optimized regressors for image super-resolution," Computer Graphics Forum, vol. 34, no. 2, pp. 95–104, 2015.

[80] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," IEEE Transactions on Computational Imaging, vol. 3, no. 1, pp. 47–57, 2017.

[81] D. Chen, L. Yuan, J. Liao, N. Yu, and G. Hua, "Stylebank: An explicit representation for neural image style transfer," in IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1897–1906.

[82] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[83] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," international symposium on physical design, vol. 60, pp. 259–268, 1992.

[84] Y. Yang and S. D. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," advances in geographic information systems, pp. 270–279, 2010.

[85] C. Gong, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," Proceedings of the IEEE, vol. 10, no. 105, pp. 1865–1883, 2017.

[86] D. Dai and W. Yang, "Satellite image classification via two-layer sparse coding with biased image representation," IEEE Geoscience and Remote Sensing Letters, vol. 8, no. 1, pp. 173–176, 2011.

[87] R. Fernandezbeltran, P. Latorrecarmona, and F. Pla, "Single-frame super-resolution in remote sensing: a practical overview," International Journal of Remote Sensing, vol. 38, no. 1, pp. 314–354, 2017.

[88] G. Vivone, L. Alparone, J. Chanussot, M. D. Mura, A. Garzelli, G. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 5, pp. 2565–2586, 2015.

[89] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi/hyperspectral images," IEEE Geoscience and Remote Sensing Letters, vol. 6, no. 4, pp. 662–665, 2009.

[90] N. Zhao, Q. Wei, A. Basarab, N. Dobigeon, D. Kouame, and J. Tourneret, "Fast single image super-resolution using a new analytical solution for ℓ_2-ℓ_2 problems," IEEE Transactions on Image Processing, vol. 25, no. 8, pp. 3683–3697, 2016.

NING ZHANG received her bachelor's degree from Northeastern University, Qinhuangdao, China, in 2017. She is currently a PhD student at Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China. Her research interests cover image processing, deep learning and remote sensing image super-resolution.

YONGCHENG WANG received his bachelor's degree from Jilin University in 2003 and Ph. D. degree from Chinese Academy of Sciences in 2010. He is a researcher of Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China. His research interests include artificial intelligence, image engineering, and embedded system of space payload.

XIN ZHANG received her bachelor's degree from Northeastern University, Qinhuangdao, China, in 2016. She is now a Ph. D. candidate at Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China. Her research interests cover deep learning and hyperspectral image classification.

DONGDONG XU received his bachelor's degree from Shandong University in 2013 and master's degree from Harbin Institute of Technology in 2015. And he is currently a Ph. D. student and a research assistant at Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China. His research interests include deep learning, image fusion and embedded system.

XIAODONG WANG received his bachelor's degree from Changchun University of Science and Technology in 1992 and Ph. D. degree from Chinese Academy of Sciences in 2003. He is now a researcher of Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China. His research interests cover imaging technology and information processing of space optical remote sensing devices.