Starting from the Sampling Imaging System, A Comprehensive Review on the Remote Sensing Image Super-Resolution Technology

Lin Lan and Chunling Lu(✉)

DFH Satellite CO., LTD., Beijing 100094, China
chunll501@sina.com

Abstract. High resolution is the eternal pursuit of remote sensing satellites. At present, the highest resolution of American lock-eye satellite KH-12 has reached 10 cm, but due to the limitation of the diffraction limit of the remote sensing satellite optical system, the development of a higher resolution remote sensing satellite has encountered a bottleneck. The satellites in orbit, such as Pleiades-NEO, have found another way to super-resolve 30 cm images to 15 cm image products using super-resolution technology based on a convolutional neural network. This method of restoring low-resolution images to high-resolution images through image super-resolution techniques has attracted a lot of attention since the middle and late 20th century. The physical continuous graphic signal is sampled and quantized into discrete digital arrays by CCD or CMOS camera, while the diffraction limit of the optical system and many other factors of degradation also exacerbated this resolution degradation phenomenon. This paper writes from the optical sampling imaging system to induce the model of remote sensing image degradation and analyze the causes of resolution degradation from the source; Further, this paper investigates the image super-resolution techniques of about the last two decades and categorizes them into two categories: traditional algorithm-based, and learning-based. This paper analyzes in detail the key algorithms in the history of super-resolution, and focuses on today's deep learning-based algorithms, clarifying the problems targeted by each type of algorithm, analyzing their design ideas and implementation principles, and how these algorithms can be adapted for super-resolution algorithms for remote sensing images. Then this paper compares the basic features of the main remote sensing image super-resolution algorithms and their advantages and disadvantages. It also introduces the current more widely used super-resolution effect evaluation metrics. Finally, this paper lists and analyzes the latest and most complete various remote sensing super-resolution datasets publicly available on the Internet, looks forward to the possible future development trends, and points out that joined with imaging systems on satellites, unsupervised learning, and multi-source remote sensing image fusion are the development directions of future remote sensing image super-resolution technologies.

Keywords: Remote Sensing Image · Super-Resolution · Degradation Model · Deep Learning

H. P. Urbach and H. Jiang (Eds.): ISSOIA 2022, SPPHY 295, pp. 207–228, 2023.
https://doi.org/10.1007/978-981-99-4098-1_19

1 Introduction

Since the 21st century, high-resolution remote sensing satellites have become a hot spot for competition among space powers, and the current highest resolution of the U.S. lock-eye satellite KH-12 has reached 10 cm [1]. However, the cost and risk of increasing the resolution by reducing the orbital altitude and increasing the focal length of the optical system are very huge. And large remote sensor equipment cannot be equipped in the current popular nanosatellites, because of its small size and small weight limitation, which means it is a luxury for remote sensing nanosatellites to obtain high-resolution images directly. Therefore, a super-resolution (SR) image processing technique, which restores low-resolution (LR) images into high-resolution images (HR) at the software level, has gradually come into view. We counted the number of papers on remote sensing images super-resolution on four bibliometrics (ScienceDirect, IEEE Xplore, CNKI database, and Wanfang database) in the last 20 years, as shown in Fig. 1. It can be seen that this software-level algorithm has received more and more attention and research. Harris [2] is the first to propose the concept of super-resolution in the 1960s, and most of the early techniques used simple interpolation, such as nearest-neighbor interpolation, bilinear interpolation, and bicubic interpolation [3]. Into the 1990s, pioneering methods such as multi-frame super-resolution based on sub-pixel shifts [4] and multi-sensor image fusion [5] were also proposed, and then super-resolution techniques began to develop rapidly. Some super-resolution models based on probability theory, transform domain, machine learning, and artificial neural networks were proposed and applied.

After 2008, super-resolution methods based on sparse coding became popular. Then 2012, because of the significant increase in hardware computing power, methods based on deep artificial neural networks gained widespread attention because of their powerful multi-layer feature learning and representation capabilities, And super-resolution techniques are evolved into models such as SRCNN [6] based on convolutional neural networks, SRGAN [7] based on generative adversarial networks, attention mechanisms, and unsupervised models. Early super-resolution techniques were applied to natural images before remote sensing image super-resolution (RSISR) techniques were developed. For single frame image super-resolution (SISR), the techniques for natural and remotely sensed images are very similar, with the possible difference that remote sensing images have more noise, as well as lower resolution and less texture information. For the super-resolution of multi/hyperspectral images with multiple frames, there are more changes to the SISR technique. In this paper, we will start from the sampling imaging of optical remote sensing images, introduce the degradation model. Then we use the degradation model as traction, discuss the mainstream SISR algorithms, and make some comparisons, as well as the implementation of SR on remote sensing images. Finally, we introduce the metrics to measure the processing results and the latest and most comprehensive remote sensing datasets, summarize the SR technology for remote sensing images and make an outlook.

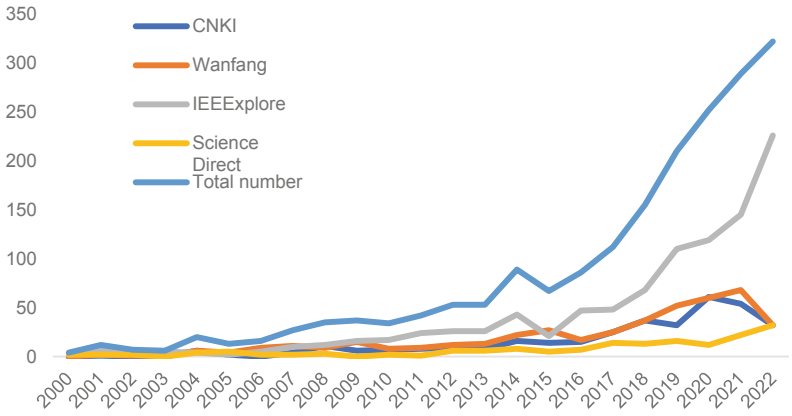

Fig. 1. RSISR literature quantity trend chart

2 Sampling Imaging System and Degradation Model

Super-resolution originated from the study of natural images and was later extended to remote sensing images. However, remote sensing images contain rich types of features, many degradation factors such as sampling, deformation, degradation clarity and noise on the imaging link as well as ground artifacts caused by cloud cover, terrain undulation, haze and other lighting changes, which make the semantic information of remote sensing images much more complex than natural images, and thus the super-resolution of remote sensing images is more difficult.

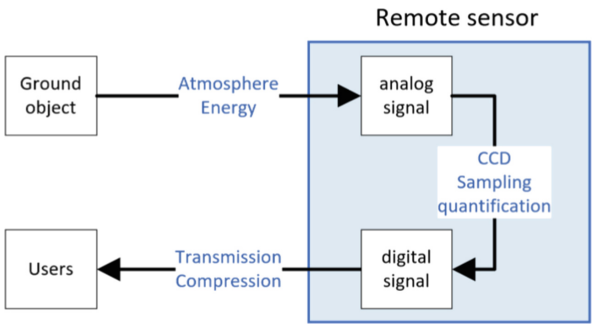

Fig. 2. Remote sensing image imaging link and degradation process

The remote sensing imaging link can be briefly summarized as follows: the target irradiance is captured by the remote sensor (CCD camera) in the sky after being disturbed by atmospheric and energy factors, then transmitted to the ground by the remote sensor through compression, and finally displayed and image processed on the ground for other applications. Among them, the continuous graphic signal of nature is quantified into discrete digital array by CCD camera sampling, this process directly degrades the continuous clear image into an image composed of one pixel, if the GSD of remote

sensing satellite is 1 m, it represents the degradation of a square meter area of the ground into the gray value of one pixel point, which greatly reduces the resolution, coupled with the atmospheric interference of the earth, energy attenuation, noise, etc. impact, further causing blurring of the image, making the image degraded from high resolution to low resolution. This process is shown in Fig. 2.

In the actual image super-resolution process, we do not reconstruct the digital signal into an analog signal, but generate more pixels to make the image clearer, that is, to enlarge (recover) a low-resolution image **LR** into a high-resolution image **HR**, the common magnification is 2, 3, 4, 8, etc. The formula can be expressed as $\boldsymbol{Y} = DB\boldsymbol{X} + N$, where $\boldsymbol{Y}$ is the low-resolution image we observe, $\boldsymbol{X}$ is the high-resolution image we want to obtain, D is a down-sampling operation, B is a blurring filter, N is the noise. After integrating D, B and N into $\boldsymbol{D_m}$, the formula can be further abstracted to $\boldsymbol{Y} = \boldsymbol{D_m X}$, where $\boldsymbol{D_m}$ represents the degradation model. The essence of the super-resolution algorithm is to find a suitable degenerate model $\boldsymbol{D_m}$ to ensure that the recovered $\boldsymbol{X}$ is consistent with the input $\boldsymbol{Y}$.

3 Traditional Algorithms

The traditional implementation idea of super-resolution is to reconstruct HR images based on a custom degradation model with a priori constraints. In this paper, according to the traditional algorithm development lineage and highlights, we focus on analyzing only four types of these algorithms: interpolation method (with bicubic interpolation [3] as an example), iterative back-projection method (IBP) [8], convex set projection method (POCS) [9] and maximum a posteriori probability method (MAP) [10].

3.1 Bicubic Interpolation (BC)

The bicubic interpolation (BC) method proposed by Keys in 1981 [3] is the most commonly used interpolation method in two-dimensional space. The core idea is that the pixel value of a certain interpolated point P is obtained by weighting the pixel values of the surrounding 16 sampled points, and the 16 nearest neighboring points of point P are selected by their relative positions as in Fig. 3. Keys constructed a bicubic function to calculate the weights of the surrounding 16 points as follows.

$$W(x) = \begin{cases} x = (a+2)|x|^3 - (a+3)|x|^2 + 1 & for\ |x| \le 1 \\ y = a|x|^3 - 5a|x|^2 + 8a|x| - 4a & for\ 1 < |x| < 2 \\ z = 0 & otherwise \end{cases} \tag{1}$$

Here a is generally taken as -0.5. After getting to the weights, we just need to weight up the pixel values of these 16 points, and the formula for interpolation is as follows.

$$f(x, y) = \sum_{i=0}^{3} \sum_{j=0}^{3} f(x_i, y_j)\, W(x - x_i)\, W(y - y_j) \tag{2}$$

It is important to note that the pixel values obtained by weighting are restricted to 0 to 255. Since the bicubic interpolation uses 16 points and a smoother cubic function, the computation is more complicated than the previous nearest-neighbor interpolation and bilinear interpolation, but the generated images are smoother and have better details.

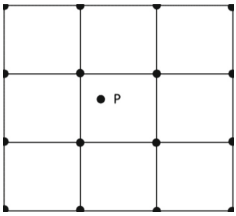

Fig. 3. Diagram of the bicubic interpolation method

3.2 Iterative Back-Projection (IBP)

Iterative back-projection (IBP) [8] was proposed by Irani and Peleg in 1991, and the core idea is to iteratively eliminate the residuals between the observed LR image and the simulated LR image to recover multiple LR images into one HR image. Which customizes the degradation model H and the inverse degradation model H^{BP}. The iterative formula is as follows

$$\hat{f}^{n+1} = \hat{f}^{n} - \lambda \sum_{i=1}^{p} H^{BP}\left(\hat{y}_i^n - y_i\right) \tag{3}$$

where: n is the number of iterations. $\hat{f}^{n+1}$ and $\hat{f}^{n}$ denote the super-resolution images obtained from the $(\mathrm{n}+1)_{\mathrm{th}}$ and n_{th} iterations. H^{BP} is the inverse projection operator, which can be obtained by synthesizing the matrix of translation, rotation, lucidity reduction, down-sampling and noise. $\hat{y}_i^n$ is the simulated LR image, and $\hat{y}_i^n = H\hat{f}^n$; p is the number of frames of the LR image used for reconstruction. λ is the projection relaxation factor.

The advantage of IBP is that the algorithm is simple and intuitively easy to understand, but it is very sensitive to high frequency noise. The method has no unique solution because the problem is inherently ill-posed. The method has some difficulties in choosing H^{BP} H^{BP}, and it is more difficult to add a priori constraints.

3.3 Projections onto Convex Sets (POCS)

In order to solve the drawback of the IBP algorithm which is difficult to use a priori information, in 1989, Stark et al. [9] applied the projections onto convex sets (POCS) method to super-resolution. In this theory, constraints or prior knowledge (such as positivity, energy boundedness, observation consistency, and smoothness) can be defined as convex sets in vector space (convex sets), and the solution space of the super-resolution reconstruction problem is formed by intersecting these convex sets.

For m prior knowledge, there will be m corresponding closed convex sets C_i , $i = 1, 2, \cdots m$. Stark and Oskoui define a convex set projection operator P_i for each convex set C_i. Then the high-resolution image is $f \in C_0 = \bigcap_{i=1}^{m} C_i P_i$, where C_0 is a nonempty closed convex set, and the iterative formula for the high-resolution image f_{k+1} is as follows.

$$f_{k+1} = T_m T_{m-1} \cdots T_1 f_k \ , \ k = 1, 2 \cdots\cdots \tag{4}$$

where $T_i = (1-\lambda_i)I + \lambda_i P_i$, and $0 < \lambda_i < 2$ is the relaxed projection operator, meaning f_{k+1} weakly converges to a feasible solution at the C_0. Any element in the intersection set C_0 is one that satisfies all the prior knowledge or constraints, so the feasible solution obtained by the POCS method is generally not unique.

The POCS algorithm has remarkable features such as simple and direct and powerful image prior knowledge embedding capability, but it also has some disadvantages as follows: 1) non-uniqueness of the solution, 2) dependence on the initial estimate, order of each projection operator, 3) higher computational complexity, but less than MAP algorithm.

3.4 The Maximum a Posteriori Probability (MAP)

The maximum a posteriori probability (MAP) method belongs to statistical estimation methods. The MAP algorithm in super-resolution reconstruction [10] refers to the maximization of the posterior probability of the occurrence of a high-resolution image given a known sequence of low-resolution images. The ideal high-resolution image is A and the observed low-resolution image is B. According to Bayes' probability theorem, the posterior probability of the synthesized high-resolution image is $P(A|B) = P(A) \cdot P(B|A) \big/ P(B)$. Where $P(B|A)$ is the conditional probability of a sequence of low-resolution images given an ideal high-resolution image; $P(A)$ is the prior probability of an ideal high-resolution image. In comparing the posterior probability $P(A|B)$, $P(\mathrm{B})$ is the same and can be deleted, so the optimization equation is $A_{MAP} = \arg\max_A \big[\log P(A) + \log P(B|A)\big]$. The conditional probability $P(B|A)$ generally uses a Gaussian model. The prior probability $P(A)$ uses different models in different algorithms, such as the Markov random field model and the Gibbs random field model.

4 Learning-Based Algorithms

The learning-based method is a data-driven approach, where the learned data is divided into labeled data (supervised learning) and unlabeled data (unsupervised learning). Ideally, the larger the number of dataset and model, the more information can be learned, and the better the final result can be. Of course, the actual learning, due to hardware and time constraints, the experimenter should choose the right amount of data, as well as optimize the model size. Learning-based super-resolution algorithms are essentially based on big data, learning the mapping of low-resolution images to high-resolution images, or learning an inverse degradation model. In this section, the learning-based single-frame image super-resolution algorithms are divided into four methods based on sparse coding, based on CNN, based on GAN and based on attention, of which the former belongs to traditional machine algorithms and the latter three belong to deep learning algorithms. In addition, since unsupervised is the development trend of deep learning, we also analyze the current single-frame image super-resolution algorithms based on unsupervised deep learning.

4.1 Based on Sparse Coding

Sparse Coding (SC). The sparse coding (called sparse representation also) wants to express most or all of the original signal with a linear combination of fewer basic signals, just as we can remember a person's face based on a small number of features. Where the original signal is an N*1-dimensional vector that can be flattened from a 2D image patch which in general is dense, i.e., most elements are not 0. These basic signals are called atoms, one atom is a vector (N*1), K atoms form an under-determined dictionary D (N*K), and the under-determined dictionary means K $>>$ N. In general the bases of the under-determined bases are redundant, making the image representation under the under-determined basis more sparse than the determined orthogonal basis, and the clean part of the image can be linearly represented using a small number of non-zero sparse representation coefficients.

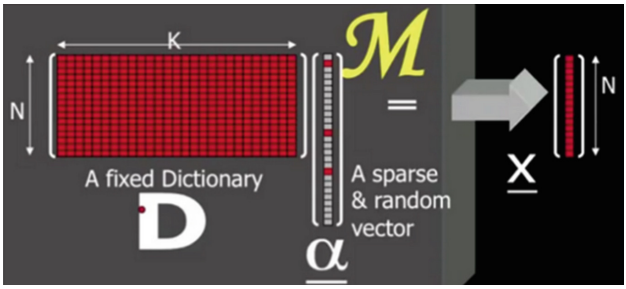

Fig. 4. Sparse representation

A sparse coding means finding a coefficient vector α_i (K*1) and a dictionary matrix D (N*K) such that $D\alpha_i$ as close as possible to x_i and α_i is as sparse as possible, α_i is the sparse representation of x_i, and the formula is $D\alpha_i = x_i$, as shown in Fig. 4. Expressed as an optimization problem, the simplest form of the sparse coding is

$$\min_{D,\alpha_i} \sum_{i=1}^{m} \|x_i - D\alpha_i\|_2 + \lambda \sum_{i=1}^{m} \|\alpha_i\|_2 \tag{5}$$

where x_i is the i_{th} sample, D is the dictionary matrix, α_i is the sparse representation of x_i, and λ is a penalty factor greater than 0 (used to make α_i more sparse). The specific methods for finding b and solving the dictionary D are not expanded in detail due to space limitations.

Application of Sparse Coding in Image Super-Resolution. The super-resolution technique based on sparse coding can be understood from the perspective of degradation model, let X represent **HR** image and Y represent **LR** image, both X and Y can be used for sparse coding, corresponding to under-determined dictionaries as D_H and D_L, then $D_H\alpha_X = X$, $D_L\alpha_Y = Y$; if $\alpha_X = \alpha_Y$, then $X = (D_H)^{-1}D_LY = D_mY$, $(D_H)^{-1}$ is the pseudo-inverse, D_m is the degradation model we are looking for. Yang et al. [11] did just that, and Yang first proposed a super-resolution reconstruction method based on sparse

coding in 2008. Yang establishes the underdetermined dictionaries D_L and D_H for high- and low-resolution image patches, respectively, while assuming that the sparse representation α of the low-resolution image patches is the same as the sparse representation α of the corresponding high-resolution image patches, and further synthesizes D_L and D_H as D, transforming the problem into a solution Eq. 6 For better results, in addition to the standard steps mentioned above, subsequent global constraints, etc., are often required. In the testing stage, we synthesize HR image using D_H and the sparse representation α of LR image, as shown in Fig. 5.

Zheng et al. [12]. First applied the sparse coding to the super-resolution of remote sensing images. For the problem that remote sensing images contain more noise, he proposed that the previous super-resolution technique of early denoising may bring more interference to the later super-resolution, so he used K-SVD [13] and OMP [14] to solve the sparse coefficients while suppressing the noise. Further, Dong et al. [15] generated multiple sub-dictionaries by clustering image patches so that a given patch of images could select sub-dictionaries adaptively, and imposed adaptive regularization constraints on the optimization equations to make the super-resolution reconstruction more accurate; Zhang et al. [16] used the initial sparse dictionary and the residual sparse dictionary to significantly improve the resolution of remote sensing images, with the former dictionary reconstructing LR images as the initial HR image, and the latter dictionary repairs the detail information lost in the initial HR image.

Fig. 5. Sparse coding for super-resolution reconstruction

4.2 Based on CNN

In order to solve the problem of losing spatial information by expanding images into vectors for processing in previous fully connected neural networks, convolutional neural networks use 2-D convolution operators to process images in 2-D directly (Fig. 6), which not only retains spatial information but also reduces the parameters of the network, and at the same time, the randomness of rounding off information also makes CNNs enter the overfitting state more slowly and improves the generalization performance of CNNs. CNNs containing 2-D convolution operator, pooling, dropout, and batch-normalization processes have opened a new era of deep learning image processing.

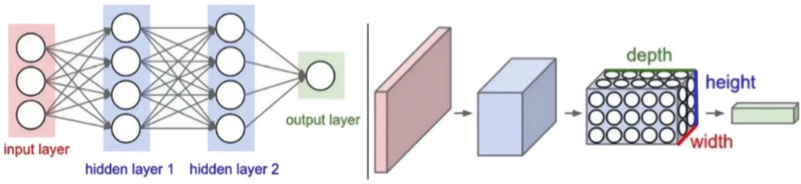

Fig. 6. Fully connected neural networks vs. convolutional neural networks

SRCNN. Dong et al. [5] proposed SRCNN (Super Resolution Convolutional Neural Network) in 2015, which kicked off the convolutional neural network for super-resolution. The network structure of SRCNN is shown in Fig. 7, where a LR_image (1*256*256) is input, which passes through convolution_kernel_1 (1,9,9,64) to get feature_map_1 (64*248*248), then passes through convolution_kernel_2 (64*1*1*32) to get feature_map_2 (32*248*248), and finally passes through convolution_kernel_3 (32*5*5 *1) to output the HR_image (1*244*244). When input is a color image, the input image is the Y channel in the YCbCr color space; and the input low-resolution image is pre-interpolated by bicubic interpolation and scaled up to a high-resolution size. The experiment shows that the high-resolution image recovered by CNN is significantly better than generated by bicubic interpolation.

The whole process from data to learning is: original HR image $\mathbf{X}$ → bicubic interpolation down-sampling → LR $\mathbf{Y_1}$ → bicubic interpolation up-sampling → immature HR $\mathbf{Y_2}$ → CNN → generated HR $\mathbf{Y}$. Where the CNN network learns the best convolutional kernel parameters in CNN by optimizing the mean square error of $\mathbf{X}$ and $\mathbf{Y}$ by SGD method. In fact, CNN has a great similarity with the idea of sparse coding, as shown in Fig. 7. Convolution_kernel_1 is like a dictionary of low-resolution image patches, transforming each low-resolution image patch into a representation on the feature map, although not sparse. Convolution_kernel_2, on the other hand, represents a nonlinear mapping that maps the low-resolution representation into a representation of the high-resolution image patch. Convolution_kernel_3 is a dictionary of high-resolution image patches, which eventually reconstructs the representations of the high-resolution patches into a high-resolution image. However, these are implicitly implemented through the convolution kernel.

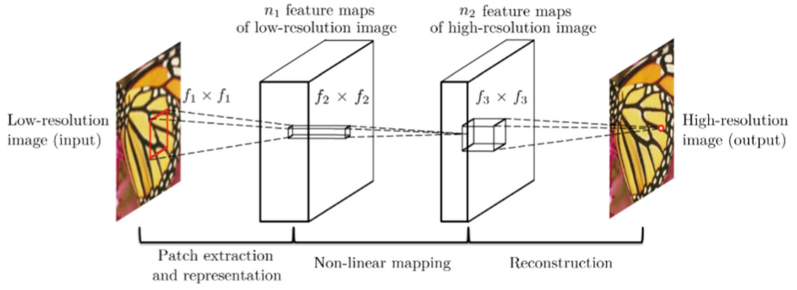

Fig. 7. The framework of SRCNN [5]

super-resolution is only applied on the luminance channel (Y channel in YCbCr color space). so c = 1 in the first/last layer, and performance (e.g., PSNR and SSIM) is evaluated on the Y channel.

Remote Sensing Images Super-Resolution on CNN. Since remote sensing images are missing local detail information, and the common deep CNN super-resolution methods using large perceptual fields tend to ignore local information, Lei et al. [17] designed a Local-Global Combined-Network (**LGCnet**) to address this problem. The network is designed to learn the multi-scale representation of remote sensing images by combining the convolution results of different layers to better perform RSISR.

Pan et al. [18] proposed a method based on residual dense networks (Residual DBPN, **RDBPN**) inspired by the Deep Back-Projection Networks (DBPN) proposed by Haris et al. [19]. The method adds dense skip connections to the DBPN projection units to construct global and local residuals, and provides information for high magnification through feature reuse, thus making it show better performance at high magnification.

4.3 Based on GAN

Deep learning models can be divided into discriminative models and generative models. Due to the invention of algorithms such as back propagation (BP), discriminative models (e.g., BP networks, CNNs, RNNs, etc.) have been developed rapidly. However, the development of generative models was slow until 2014, when Goodfellow et al. [20] proposed the most successful generative model, the Generative Adversarial Network (GAN), and generative models exploded.

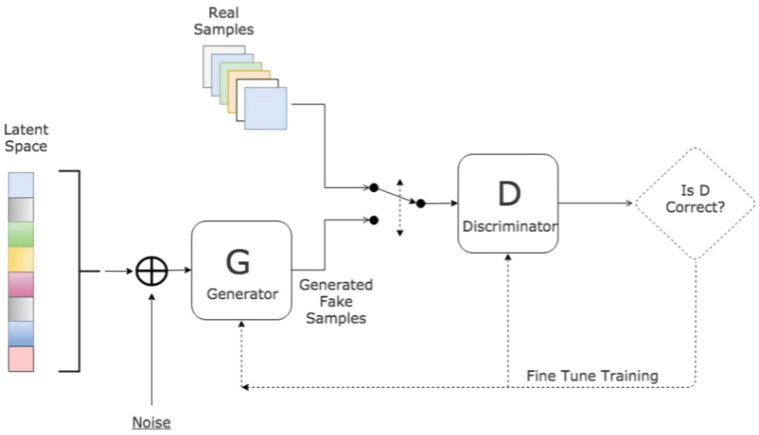

Fig. 8. The framework of GAN network

GAN consists of a generator and a discriminator, as shown in Fig. 8. Latent Space is a random hidden variable of fixed size, often obeying Gaussian distribution, and the **G**enerator tries to generate fake samples from the hidden variable. The **D**iscriminator learns the real samples to obtain the ability to identify the real and fake samples. In

this way, through the continuous learning and confrontation between the **D**iscriminator and the **G**enerator, the **G**enerator is finally able to generate fake samples that are highly similar to the real samples, causing the **D**iscriminator to be unable to judge, which means that the **G**enerator has successfully modeled the process of producing real samples.

SRGAN. GAN network is applied by **SRGAN** [7] for super-resolution algorithm reconstruction for the first time. SRGAN added ResNet to the Generator, which enables to train deeper network and improve the network accuracy. SRGAN proposed perceptual loss function, which can generate images more in line with human visual perception and the loss is an effective guarantee for the algorithm. The perceptual loss can be split into content loss and adversarial loss, and the content loss is a great innovation. The previous MSE loss can improve the PSNR, but it also loses some high-frequency information, resulting in blurred images. The content loss is calculated on the VGG's feature map by passing the true and false samples through the VGG network, which can greatly improve the blurring problem caused by MSE loss.

Remote Sensing Images Super-Resolution on GAN. The degradation process of remote sensing images often contains more noise, and the original GAN-based method is more sensitive to noise, which generates high-frequency noise independent of the input image. In response, Jiang et al. [21] proposed **EEGAN** (Edge-Enhanced GAN) from the perspective of edge enhancement. The Edge-Enhanced Sub-Network (EESN), constructed by Laplacian operator, fuses the SR reference images and their edges to generate HR remote sensing images with clear edges, so as to alleviate the problem of blurred edges in remote sensing images.

It is found through statistical analysis that there are more low-frequency components (flat regions) in the remotely sensed images than in the natural images. When using GAN for RSISR, it is difficult for the Discriminator in the network to determine whether these low-frequency regions are generated from the real HR remote sensing images, which leads to the quality of the generated HR images to be affected. In this regard, Lei et al. [22] designed a Coupled-Discriminate GAN network (**CDGAN**). The two-channel network in the coupled discriminator joins the features extracted from the real HR image and the generated HR image and input them into the subsequent layers, and a dedicated coupled loss function is constructed to update the network parameters. The model improves the GAN-based image SR method in processing low-frequency image regions with blurred resolution.

4.4 Remote Sensing Images Super-Resolution on Attention

Although CNNs have made significant achievements in deep learning, there are some problems, such as sliding weight windows applied to all spatial channels or spectral channels of the feature map equally when performing convolutional computation, and this uniform computation makes it difficult to extract the part of features that need special attention. As a result, attention mechanisms came into the vision of researchers, Hu et al. [23] proposed the Squeeze-and-Excitation network SENet based on channel attention module (**CAM**), and Woo et al. [24] proposed the spatial attention module (**SAM**)-based convolutional block attention module (CBAM), like Fig. 9. On the whole, CAM means

generating a weight vector MC with the same number of channels as the input feature map, and the size of the weights corresponds to whether a channel in the feature map is more important. SAM means generating a weight matrix MS with the same height and width dimensions as the input feature map, and the size of the weights corresponds to whether a pixel in the feature map is more important. These two attention models can be used independently or mixed, and more attention models have been developed later. Also the attention mechanism can be used not only in CNNs but also in other deep learning networks, and later it has been developed into the transformer [25] module which is very popular nowadays.

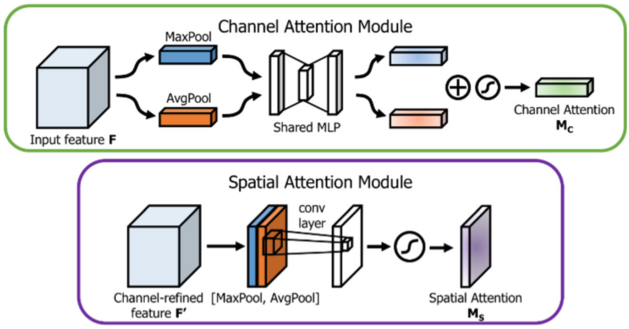

Fig. 9. Channel Attention Module and Spatial Attention Module [24]

The Attention module was first introduced into image super-resolution processing by Zhang et al. [26], who proposed a deep residual channel attention network (**RCAN**). Zhang utilized the channel attention module to adaptively adjust the channel features. Later, based on the idea of RCAN [26], Huat et al. [27] proposed **RSRCAN** network to introduce channel attention module into remote sensing image super-resolution. Huat pointed out that the current deep learning super-resolution model is difficult to train due to its own complexity and the lack of important training data, which becomes an important limitation for satellite remote sensing image super-resolution. Moreover, most CNN-based super-resolution algorithms for remote sensing images default to equal importance of all features extracted from LR input images, which may lead to a lack of flexibility in analyzing the presence of different types of features in remote sensing images. Moreover, remotely sensed image data suffers some degradation during its acquisition, and this usually introduces a lot of noise and variability in the data (in addition to the rich low-frequency information), so Huat uses the channel attention module to focus on the surface features that require finer HR detail by enhancing the high-frequency information of the image and suppressing the low-frequency information, thus allowing the model's to learn more about the mapping relationships between the high-frequency components. Since texture information varies greatly with different remote sensing images, but most SR methods use the same learning model for all scenes, the existing SR methods have poor generalization capability. To be able to adaptively adjust the network according to the input images, Jia et al. [28] proposed a multi-attentive remote sensing super-resolution reconstruction network (**MA-GAN**). The main body of MA-GAN [28] contains three

modules: a pyramidal convolutional residual dense (PCRD) block, an attention-based up-sampling (AUP) block, and an attention-based fusion (AF) block. The PCRD block combines multi-scale convolution and channel attention to automatically learn and scale the residuals for better representation, the AUP block uses pixel attention to perform arbitrary scale up-sampling, and the AF block uses branch attention to integrate the up-sampled low-resolution images with high-level features.

4.5 Unsupervised Algorithms

Supervised deep neural networks often require HR-LR image pairs during training, and the LR images used are generally obtained from HR images by custom downsampling algorithms (e.g., bicubic interpolation). However, in reality, high-resolution remote sensing images are usually difficult to obtain, and there are still differences between the images obtained by custom degradation and the actual low-resolution remote sensing images.

Haut et al. [29] constructed a CNN-based deep generative network (**ANDGN**) for RSISR from an unsupervised perspective. The method first uses a CNN to expand the random noise to the target HR dimension, then downsamples [30] the generated HR results to obtain the generated LR image. The loss is calculated by generated LR image and original LR image and minimized through iterations until the final desired HR remote sensing image is generated. The unsupervised aspect of ANDGN are reflected in the fact that only LR images are used. Wang et al. [31] proposed an unsupervised learning network **CycleCNN** for SR of remote sensing images based on CycleGAN [32], which contains a cyclic network composed of image degradation network and image super-resolution network. The HR image selected in the paper is a panchromatic image with GSD of 1 m/pixel in GaoFen-2, and the HR image is single-frame of multispectral image with GSD of 4 m/pixel in GaoFen-2. The LR image undergoes the cycle process of super-resolution and degradation in the network, and the HR image undergoes the cycle process of super-resolution and degradation in the network. Wang then constructs cycle loss and identity loss functions to enable the degraded network to degrade more realistic low-resolution images and improve the performance of the super-resolution network. The unsupervised nature of CycleCNN is reflected in the fact that no HR-LR image pairs are used. Considering that high-resolution remote sensing images are difficult to obtain, Zhang et al. [33] in 2022 proposed an unsupervised visible image-guided remote sensing image super-resolution network (**UVRSR**) by guiding low-resolution (LR) remote sensing images through HR visible natural images, which was successfully established.

5 Comparison and Discussion

5.1 Comparison of Algorithms

The comparison of traditional algorithms is shown in Table 1, and it can be seen that since the bicubic interpolation method is a typical mathematical violence interpolation method, which is relatively fixed and not good for improvement, although it is simple

and intuitive. The later appearing super-resolution algorithms (IBP, POCS, MAP, etc.) have more data, theoretical support and more priori information making them much superior to the bicubic interpolation method in terms of image super-resolution, which is now basically used to appear as a comparison method in the paper. Specifically, the IBP algorithm obtains the inverse projection operator based on a custom degradation model, which is used for iterative optimization to obtain super-resolved images. However, this prevents the introduction of a priori information such as, for example, positivity, energy boundedness, observation consistency, and smoothness, which is compensated by the MAP and POCS methods, where the MAP method goes further and uses the statistical laws of a large amount of data to hyper-segment the image, and also lays the foundation for later data-driven models based on them.

Table 2 shows the improvement of learning-based super-resolution techniques for remote sensing images compared to bicubic interpolation on PSNR and SSIM, using data from the original papers of the relevant methods. SC in Table 2 is the traditional machine learning algorithm, and the others are deep learning algorithms. Overall, different learning-based algorithms have slight advantages at different magnifications or different evaluation metrics, for example, RDBPN, MA-GAN and UVRSR still have good results at 8 times. The latter emerged methods have better image quality compared to the former usually, but it also tends to mean that their networks also become more complex and require more hardware resources. Specifically for deep learning-based super-resolution methods, the network in the algorithm only targets a specific magnification and the size of the input image. If the image size is larger than the set network input size, it is often super-resolved by cropping and then stitched into a large super-resolved image. If the image size is smaller than the set network input size, it is generally not recommended to select this network for super-resolution, or to up-sample the magnification to the network input size first. In the case of end-to-end pure convolutional networks, in principle, images of various sizes can be input, but in practice, it is limited by hardware resources as well as image quality requirements. Different networks need to be trained for different magnifications, which means that deep learning-based methods have high hardware requirements and are currently difficult to be used directly for Onboard processing.

It shows the performance of the previously mentioned methods on the road test set of UC Merced with a super-resolution factor of 4 and an objective evaluation metric of PSNR, with data and images from the papers of ANDGN [27] and RSRCAN [29] in Fig. 10. ANDGN is the unsupervised method and the remaining six are supervised methods. From the figure, we can roughly see that the super-resolution techniques have been developed over the years and the results that can be presented are getting better and better. For the current public dataset, the supervised methods may yield somewhat better results than the unsupervised ones, but unsupervised methods are still the future direction of development.

Table 1. Comparison of traditional super-resolution algorithms

method	bicubic interpolation	IBP	MAP	POCS
Applicable theory	Single	Limited	Lots of	Limited
priori information	no	no	Prior probability density function	Convex sets, simple and efficient
Super-resolution optimal solution	Existing and unique	Not unique, cannot be constrained a priori	Existence and uniqueness, MAP optimal estimation	Not unique, within the intersection of constrained sets
Optimization method	No iteration	Iterative projection method	Standard iterative algorithm	Iterative projection method
Convergence speed	Fast	Slow	Slow	Slow
Convergence stability	High	Relatively low	High	Relatively low
Calculation volume	Relatively low	Relatively high	High	Relatively high
Complexity	Medium	Definition of the projection operator	Optimization under non-convex a priori information	Definition of the projection operator
Noise reduction capability	Relatively low	Relatively low	High	Relatively low
Image smoothing control capability	No	Relaxation factor coefficient	A priori model, regularization factor	Relaxation factor coefficient
Edge preservation capability	Medium	Relaxation factor coefficient	Medium	High
Number of pictures needed / piece	1	$>=1$	$>=1$	$>=1$

5.2 Evaluation Metrics

There is not yet a unified metric for evaluating super-resolution-generated images, because it is not like object detection, image segmentation, etc., where only two options of right or wrong are evaluated on pixels that have already existed. Super-resolution is achieved by adding pixels, and whether the added pixels satisfy the requirements is often subjective and objective criteria are difficult to unify. Therefore, new evaluation metrics have also been continuously proposed by researchers, and the commonly used metrics are as: 1) Mean Square Error (**MSE**). MSE is the average of the difference between the

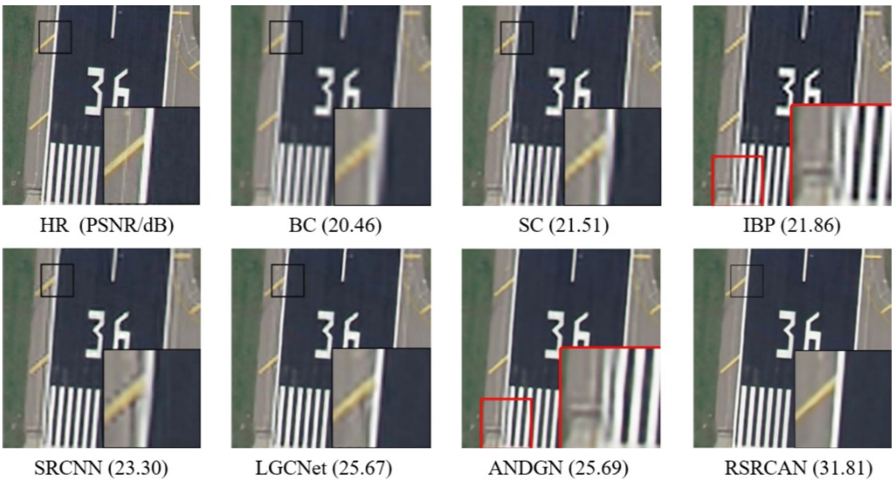

Fig. 10. PSNR of the UC Merced road test image considering a 4 × scaling factor. [27, 29]

pixel values of two images, the smaller the value, the smaller the difference between the two images, but it focuses too much on the comparison of individual pixels, ignoring the overall visual perception of the image. 2) Normalized mean square error (**PSNR**). PSNR is one of the most commonly used image quality evaluation metrics, which combines MSE with the dynamic range of the whole image, and has a more holistic sense compared to MSE. The larger the PSNR value, the more similar the image is. 3) Structural similarity (**SSIM**) [35], another popular objective assessment metric, is a global perceptual model using luminance, contrast, and structure. The higher the SSIM value, the more similar the two signals are. 4) Natural Image Quality Evaluator (**NIQE**) [36], NIQE neither seeks a priori information about distorted images nor relies on any human opinion scores, and a smaller NIQE value indicates better visual quality. 5) Perceptual Index (**PI**) [37], pi value represents the subjective perceptual quality of an image and is also a popular evaluation metric. Often, the lower the pi value, the better the perceptual quality of the image, which is the opposite of the PSNR value. 6) Learned Perceptual Image Patch Similarity (**LPIPS**) [38], LPIPS is more consistent with human perception than traditional methods (than PSNR, SSIM). a lower value of LPIPS indicates that the two images are more similar, and vice versa, the greater the difference. In addition, other metrics, which will not be continued in this paper.

5.3 Datasets

The common datasets of deep learning-based methods for super-resolution reconstruction of remote sensing images are as follows.

1) UC Merced [39]. It is a dataset used to study land use, with a total of 2,100 images containing 21 categories of scenes, 100 images per category, and the pixel size of each image is 256 × 256.
2) RSCNN7 [40]. A total of 2800 remote sensing images from 7 different scene categories, each category contains 400 images.

Table 2. Learning-based methods relative to bicubic linear interpolation for PSNR and SSIM enhancement

method	dataset	metrics	enlarge factor			
			x2	x3	x4	x8
SC [34]	Google Earth	PSNR	–	3.99%	–	–
		SSIM	–	3.80%	–	–
LGCNet [17]	UC Merced	PSNR	8.84%	6.63%	14.15%	–
		SSIM	5.19%	7.95%	22.50%	–
RDBPN [18]	UC Merced	PSNR	–	–	9.64%	5.15%
		SSIM	–	–	16.94%	16.10%
EEGAN [21]	UC Merced	PSNR	14.14%	14.15%	13.38%	–
		SSIM	3.73%	7.59%	11.14%	–
CDGAN [22]	UC Merced	PSNR	–	–	1.95%	–
		SSIM	–	–	–	–
RSRCAN [27]	UC Merced	PSNR	11.74%	10.20%	8.69%	–
		SSIM	5.77%	11.48%	14.60%	–
ANDGN [29]	UC Merced RSCNN7 WHU-RS19	PSNR	8.75%	–	6.87%	–
		SSIM	5.12%	–	11.69%	–
CycleCNN [31]	GaoFen-2	PSNR	–	–	4.66%	–
		SSIM	–	–	1.27%	–
MA-GAN [28]	NWPU-RESISC45	PSNR	11.59%	–	12.16%	10.80%
		SSIM	5.31%	–	16.64%	14.73%
UVRSR [33]	UC Merced	PSNR	7.08%	–	4.74%	18.17%
		SSIM	–	–	–	–

3) NWPU-RESISC45 [41]. From Northwestern Polytechnic University, it contains a total of 31,500 images divided into 45 scene categories with 700 images in each category, and the pixel size of each image is 256 × 256.
4) Kaggle open source dataset [42]. This dataset consists of more than 1000 VHR aerial photographs collected in southern California, USA. It contains 350 images for training and 1370 images for testing.
5) Sentinel-2 dataset, which is one of the Sentinel series, the data is free of charge, and the main payload is a multispectral imager with 13 bands in the spectrum from 0.4 to 2.4 m, covering visible, near-infrared and short-wave infrared, and this dataset is increasingly used as a complement to Landsat in the field of Earth observation.
6) AID [43]. This dataset was released by Wuhan University in 2012. The data source is Google Earth and includes 30 types of remotely sensed scenes such as parks, airports, mountains, and churches. Each type has 200 to 400 images, and all images are 600 × 600 pixels in size. The spatial resolution is 0.58 m/pixel

7) PatternNet. This dataset was published by Wuhan University in 2018. The data source is Google Maps and includes 38 types of remotely sensed scenes, such as forests, highways, railroads, shipyards, and soccer fields. There are 800 images for each category. The size of all images is 256 × 256 pixels with a spatial resolution of 0.064.7m/pixel.
8) WHU-RS19 [44]. WHU-RS19 contains 19 different land classes. Fifty VHR images for each category were obtained from Google Earth with a size of 600 × 600 pixels.
9) IEEE Data Fusion Contest (DFC) 2019 [45]. The DFC (2019) dataset consists of 2783 multi-date satellite images captured by the WV-3 satellite for training and 50 images for testing. The size of the sample image blocks is 1024 × 1024 pixels.
10) SpaceNet dataset (AWS2018) [46]. SpaceNet is another large-scale satellite image dataset, acquired exclusively from the VHR WV-3 satellite.
11) GeoEye-1 dataset. GeoEye-1 satellite is a commercial satellite launched by the United States on September 6, 2008 from Vandenberg Air Force Base, California, to acquire quad-band (blue, green, red, and near-infrared) multispectral images with a spatial resolution of 0.41 m and a spatial resolution of 1.65 m.
12) SPOT-6 dataset. SPOT-6 satellite was launched on September 9, 2012, with a spatial resolution of 1.5 m for panchromatic images and 6 m for multispectral images, including blue, green, red and near infrared.
13) Gaofen2 dataset. Launched on August 19, 2014, the Gaofen 2 satellite has a spatial resolution better than 1 m. It is equipped with two cameras, HR for 1 m panchromatic imaging and 4 m multispectral imaging.
14) DOTA [47]. DOTA is a large-scale benchmark dataset based on aerial imagery generated for target detection tasks. The dataset contains 2806 images collected from Google Earth, GF-2 and JL-1 satellites and aerial images provided by CycloMedia B.V.
15) Multi-sensor Remote Sensing Dataset (MSRSD) [48]. This dataset consists mainly of VHR satellite images acquired by Pleiades 1A/1B, GeoEye-1, QuickBird2, WV-2, WV-3, and DEIMOS satellites and most of them are publicly available. MSRSD includes satellite images from seven different satellites, from different geographic locations and various landscape conditions to simplify the transferability of the model globally and universality, it is a common and rich dataset.

6 Conclusion

Super-resolution processing has been a hot topic in the field of image processing, especially in the field of remote sensing, which has very important application value. This paper reviews image super-resolution processing techniques in remote sensing, categorizing them into traditional methods and learning-based methods, among which, deep learning-based methods are the hot topic nowadays, and more pages are spent on that part in this paper. To facilitate the introduction of RSISR techniques, many methods are categorized in this paper, but there are actually many methods that do not belong to one of these categories, and they integrate many classes of methods as a sub-method to form one method that works better. For example, SWCGAN [49], which integrates CNN, GAN, and Swin Transformer [50] together. For single image super-resolution (SISR), the techniques based on natural images are easy to migrate to remote sensing

images, but because the resolution of remote sensing images is much lower than that of natural images, thus causing much less detailed texture information, these images also contain more noise, so blindly performing remote sensing image super-resolution may result in a large amount of false information, which instead reduces the image quality. Therefore, in order to improve the quality of remote sensing image super-resolution, it is suggested that the original resolution of single frame satellite remote sensing image utilized by super-resolution technique is often above 2 m, and preferably higher than 1 m. In addition, during this literature research, the author considers the following main development directions of optical remote sensing image super-resolution technology.

1) Research on the evaluation metric of super-resolution algorithm. As mentioned in Sect. 5.1, super-resolution is intuitively expressed as an increase in the number of pixels, while whether the image meets the requirements after the increase of pixels often needs to be evaluated by the human eye, and objective evaluation metrics cannot be formed yet. Therefore, the image quality evaluation metrics applicable to super-resolution processing still need further research.
2) Research on unsupervised learning of super-resolution reconstruction method for remote sensing images. It is difficult to obtain remote sensing images with different resolutions in the same scene, so there is a lack of training samples, and the low-resolution samples in current supervised algorithms are usually obtained by artificial down-sampling, which cannot simulate the image degradation model well and can hardly cope with the super-resolution reconstruction tasks of some actual scenes, so unsupervised super-resolution reconstruction models of remote sensing images have important research values.
3) Research on the special neural network structure for remote sensing image characteristics. Although the existing methods of super-resolution reconstruction of remote sensing images based on deep learning have achieved good results, remote sensing images are characterized by scale diversity, viewpoint specificity, multi-direction and high background complexity, so it is still a question worth exploring how to use existing technologies (such as attention mechanism, multi-scale feature fusion, reinforcement learning, etc.) to build a more efficient and super-resolution network adapted to the characteristics of remote sensing images problem.
4) Jointly with the imaging system on the satellite. For example, the algorithm proposed by CNES to enhance the resolution of PAN SPOT5 using a quincunx sampling mode, which uses two CCD detectors shifted in the focal plane in order to acquire two images with a resolution of 5 m, resulting in a synthetic resolution of about 2.5 m images [53]

References

1. Lin, C.: Analysis of Electronic intelligence safeguards during the gulf war. National Air Intelligence Center Wright-Patterson AFB OH (1996)
2. Harris, J.L.: Diffraction and resolving power. JOSA **54**, 931–936 (1964)
3. Keys, R.: Cubic convolution interpolation for digital image processing. IEEE Trans. Acoust. Speech Sig. Process. **29**, 1153–1160 (1981). https://doi.org/10.1109/TASSP.1981.1163711

4. Peleg, S., Keren, D., Schweitzer, L.: Improving image resolution using subpixel motion. Pattern Recogn. Lett. **5**, 223–226 (1987)
5. Nguyen, N., Milanfar, P., Golub, G.: A computationally efficient super resolution image reconstruction algorithm. IEEE Trans. Image Process. **10**, 573–583 (2001)
6. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. http://arxiv.org/abs/1501.00092 (2015)
7. Ledig, C., et al.: Photo-realistic single image super-resolution using a generative adversarial network. http://arxiv.org/abs/1609.04802 (2017)
8. Irani, M., Peleg, S.: Improving resolution by image registration. CVGIP: Graph. Model. Image Process. **53**, 231–239 (1991). https://doi.org/10.1016/1049-9652(91)90045-L
9. Stark, H., Oskoui, P.: High-resolution image recovery from image-plane arrays, using convex projections. J. Opt. Soc. Am. A. **6**, 1715 (1989). https://doi.org/10.1364/JOSAA.6.001715
10. Schultz, R.R., Stevenson, R.L.: Extraction of high-resolution frames from video sequences. IEEE Trans. Image Process. **5**, 996–1011 (1996). https://doi.org/10.1109/83.503915
11. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution as sparse representation of raw image patches. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE, Anchorage, AK, USA (2008). https://doi.org/10.1109/CVPR.2008.4587647
12. Liebel, L., Körner, M.: Single-image super resolution for multispectral remote sensing data using convolutional neural networks. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. **XLI-B3**, 883–890 (2016). https://doi.org/10.5194/isprsarchives-XLI-B3-883-2016
13. Aharon, M., Elad, M., Bruckstein, A.: $rm K$-SVD: an algorithm for designing over complete dictionaries for sparse representation. IEEE Trans. Signal Process. **54**, 4311–4322 (2006). https://doi.org/10.1109/TSP.2006.881199
14. Tropp, J.A., Gilbert, A.C.: Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans. Inform. Theory. **53**, 4655–4666 (2007). https://doi.org/10.1109/TIT.2007.909108
15. Dong, W., Zhang, L., Shi, G., Xiaolin, W.: Image Deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. IEEE Trans. Image Process. **20**, 1838–1857 (2011). https://doi.org/10.1109/TIP.2011.2108306
16. Zhang, Y., Wu, W., Dai, Y., Yang, X., Yan, B., Lu, W.: Remote sensing images super-resolution based on sparse dictionaries and residual dictionaries. In: 2013 IEEE 11th International Conference on Dependable, Autonomic and Secure Computing, pp. 318–323. IEEE, Chengdu, China (2013). https://doi.org/10.1109/DASC.2013.82
17. Lei, S., Shi, Z., Zou, Z.: Super-resolution for remote sensing images via local-global combined network. IEEE Geosci. Remote Sensing Lett. **14**, 1243–1247 (2017). https://doi.org/10.1109/LGRS.2017.2704122
18. Pan, Z., Ma, W., Guo, J., Lei, B.: Super-resolution of single remote sensing image based on residual dense back projection networks. IEEE Trans. Geosci. Remote Sensing. **57**, 7918–7933 (2019). https://doi.org/10.1109/TGRS.2019.2917427
19. Haris, M., Shakhnarovich, G., Ukita, N.: deep back-projection networks for super-resolution. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1664–1673. IEEE, Salt Lake City, UT, USA (2018). https://doi.org/10.1109/CVPR.2018.00179
20. Goodfellow, I.J., et al.: Generative adversarial networks. http://arxiv.org/abs/1406.2661 (2014)
21. Jiang, K., Wang, Z., Yi, P., Wang, G., Lu, T., Jiang, J.: Edge-enhanced GAN for remote sensing image super resolution. IEEE Trans. Geosci. Remote Sens. **57**, 5799–5812 (2019). https://doi.org/10.1109/TGRS.2019.2902431
22. Lei, S., Shi, Z., Zou, Z.: Coupled adversarial training for remote sensing image super-resolution. IEEE Trans. Geosci. Remote Sens. **58**, 3633–3643 (2020). https://doi.org/10.1109/TGRS.2019.2959020

23. Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E.: Squeeze-and-excitation networks. http://arxiv.org/abs/1709.01507 (2019)
24. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: convolutional block attention module. http://arxiv.org/abs/1807.06521 (2018)
25. Vaswani, A., et al.: Attention is all you need. http://arxiv.org/abs/1706.03762 (2017)
26. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. http://arxiv.org/abs/1807.02758 (2018)
27. Haut, J.M., Fernandez-Beltran, R., Paoletti, M.E., Plaza, J., Plaza, A.: Remote sensing image superresolution using deep residual channel attention. IEEE Trans. Geosci. Remote Sens. **57**, 9277–9289 (2019). https://doi.org/10.1109/TGRS.2019.2924818
28. Jia, S., Wang, Z., Li, Q., Jia, X., Xu, M.: Multiattention generative adversarial net-work for remote sensing image super-resolution. IEEE Trans. Geosci. Remote Sens. **60**, 1–15 (2022). https://doi.org/10.1109/TGRS.2022.3180068
29. Haut, J.M., Fernandez-Beltran, R., Paoletti, M.E., Plaza, J., Plaza, A., Pla, F.: A new deep generative network for unsupervised remote sensing single-image super-resolution. IEEE Trans. Geosci. Remote Sens. **56**, 6792–6810 (2018). https://doi.org/10.1109/TGRS.2018.2843525
30. Turkowski, K.: Filters for common resampling tasks. In: Graphics Gems, pp. 147–165. Elsevier (1990). https://doi.org/10.1016/B978-0-08-050753-8.50042-5
31. Wang, P., Zhang, H., Zhou, F., Jiang, Z.: Unsupervised remote sensing image super-resolution using cycle CNN. In: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 3117–3120. IEEE, Yokohama, Japan (2019). https://doi.org/10.1109/IGARSS.2019.8898648
32. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2242–2251. IEEE, Venice (2017). https://doi.org/10.1109/ICCV.2017.244
33. Zhang, Z., Tian, Y., Li, J., Xu, Y.: Unsupervised remote sensing image super-resolution guided by visible images. Remote Sens. **14**, 1513 (2022). https://doi.org/10.3390/rs14061513
34. Zhihui, Z., Bo, W., Kang, S.: Single remote sensing image super-resolution and de-noising via sparse representation. In: 2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping, pp. 1–5. IEEE, Xiamen, China (2011). https://doi.org/10.1109/M2RSM.2011.5697420
35. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**, 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861
36. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Sig. Process. Lett. **20**, 209–212 (2013). https://doi.org/10.1109/LSP.2012.2227726
37. Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., Zelnik-Manor, L.: The 2018 PIRM challenge on perceptual image super-resolution (2019). https://doi.org/10.48550/arXiv.1809.07517. http://arxiv.org/abs/1809.07517
38. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric (2018). https://doi.org/10.48550/arXiv.1801.03924. http://arxiv.org/abs/1801.03924
39. Yang, Y., Newsam, S.: Bag-of-visual-words and spatial extensions for land-use classi-fication. In: Proceedings of the 18th SIGSPATIAL International Conference on Ad-vances in Geographic Information Systems - GIS 2010, p. 270. ACM Press, San Jose, California (2010). https://doi.org/10.1145/1869790.1869829
40. Zou, Q., Ni, L., Zhang, T., Wang, Q.: Deep learning based feature selection for remote sensing scene classification. IEEE Geosci. Remote Sens. Lett. **12**, 2321–2325 (2015). https://doi.org/10.1109/LGRS.2015.2475299

41. Cheng, G., Han, J., Lu, X.: Remote sensing image scene classification: benchmark and state of the art. Proc. IEEE. **105**, 1865–1883 (2017). https://doi.org/10.1109/JPROC.2017.2675998
42. Kaggle: Kaggle Open Source Dataset. https://www.kaggle.com/c/draper-satellite-image-chronology/data
43. Xia, G.-S., et al.: AID: a benchmark data set for performance evaluation of aerial scene classification. IEEE Trans. Geosci. Remote Sens. **55**, 3965–3981 (2017). https://doi.org/10.1109/TGRS.2017.2685945
44. Sheng, G., Yang, W., Xu, T., Sun, H.: High-resolution satellite scene classification using a sparse coding based multiple feature combination. Int. J. Remote Sens. **33**, 2395–2412 (2012). https://doi.org/10.1080/01431161.2011.608740
45. Bertrand: DATA FUSION CONTEST 2019. https://ieee-dataport.org/open-access/data-fusion-contest-2019-dfc2019
46. SpaceNet: The SpaceNet Datasets. https://spacenet.ai/datasets/
47. Xia, G.-S., et al.: DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3974–3983. IEEE, Salt Lake City, UT (2018). https://doi.org/10.1109/CVPR.2018.00418
48. Wang, P., Bayram, B., Sertel, E.: A comprehensive review on deep learning based remote sensing image super-resolution methods. Earth Sci. Rev. **232**, 104110 (2022). https://doi.org/10.1016/j.earscirev.2022.104110
49. Tu, J., Mei, G., Ma, Z., Piccialli, F.: SWCGAN: generative adversarial network combining swin transformer and CNN for remote sensing image super-resolution. IEEE J. Sel. Top. Appl. Earth Ob. Remote Sens. **15**, 5662–5673 (2022). https://doi.org/10.1109/JSTARS.2022.3190322
50. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin Trans-former: Hierarchical Vision Transformer using Shifted Windows (2021). https://doi.org/10.48550/arXiv.2103.14030. http://arxiv.org/abs/2103.14030