Single Satellite Imagery Superresolution Based on Hybrid Nonlocal Similarity Constrained Convolution Sparse Coding

Nan Chen, *Student Member, IEEE*, Lichun Sui, Biao Zhang, Hongjie He, José Marcato Junior, *Member, IEEE*, and Jonathan Li, *Senior Member, IEEE*

***Abstract*—The traditional superresolution methods based on image patches often ignore the consistency between the overlapped patches, causing block effects in produced images. The convolutional sparse coding based superresolution method uses the translation invariance of the convolution filter to directly encode the entire image, maintaining consistency and good performance. In this article, we propose a novel approach to single-image superresolution reconstruction based on hybrid nonlocal similarity constrained convolution sparse coding. We first decompose the input image into a smooth part and a texture part. The Bayesian nonparametric model can use more prior information of the original image, so we replace the previous bicubic interpolation with this method to better reconstruct the residual high-frequency information in the smooth part. When reconstructing the texture part, this article proposes a nonlocal similarity constrained convolutional sparse coding method, which transforms the reconstruction of the texture part to minimize the convolution sparse coding noise of the feature maps and classifies the image patches in the search space by using the correlation coefficients as the structural information, avoiding unnecessary weight calculation. Several methods were tested on satellite images extensively. Both visual inspection and quantitative analysis results demonstrate that our method outperforms other state-of-the-art methods and increases noise immunity effectively.**

***Index Terms*—Bayesian nonparametric model, convolution sparse coding, correlation coefficient, nonlocal (NL) similarity structure, superresolution (SR) reconstruction.**

NOMENCLATURE

Y	Input LR images for testing.
Y^*	Final reconstructed HR image.
Y_l, Y_h	Smooth/texture part of Y.
f_i^l, f_i^h	LR/HR filters.
W	Mapping function of feature maps.
f^s	Low-pass filter.
f^h, f^v	Horizontal and vertical gradient operators.
$D^{(l)}$ $D^{(h)}$	LR/HR dictionary.
$x^{(l)}$ $x^{(h)}$	LR/HR sample images.
K	Dictionary size.
s_{ik}	Sparse-valued weights.
z_{ik}	Binary-valued.
A	Sparse coefficient.
π_k	Probability of using dictionary element.
γ_s	Precision of the sparse weights.
γ_ϵ	Observation noise.
Z^l, Z^h	Smooth/texture feature map.
$\bar{Z}^h$	Real feature map.
$\tilde{Z}^h$	Reasonable estimation for real feature map.
$v_z = Z^h - \bar{Z}^h$	Sparse coding noise.
$Z^{(h)*}$	Optimal feature map.
$\hat{Z}_i^h$	Initial HR feature map.
Ω_i	Index position information set.
ϵ_i	Observation noise.
$S_\tau(\cdot)$	Soft threshold operation.
$\tilde{\mathrm{x}}_{i(-k)}$	Reconstruction error.
λ_i γ λ^l	Regularization parameters.
$c_0, d_0, e_0, f_0, \tau_0, \eta_0$	Hyperparameters.
h, c η. Q U_i d	Auxiliary parameter.

Manuscript received June 22, 2020; revised August 19, 2020 and September 20, 2020; accepted September 23, 2020. Date of publication October 7, 2020; date of current version August 4, 2021. This work was supported by the National Natural Scientific Foundation of China under Grant 41601345, Grant 41571346, Grant 41871380, and Grant 4141379. *(Corresponding authors: Lichun Sui; Jonathan Li.)*

Nan Chen and Lichun Sui are with the College of Geological Engineering and Geomatics, Chang'an University, Xi'an 710054, China (e-mail: chcdut@126.com; sui1011@chd.edu.cn).

Biao Zhang is with the Geovis Spatial Technology Co.,Ltd, Xi'an 710199, China (e-mail: 240531197@qq.com).

José Marcato Junior is with the Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande 79070-900, Brazil (e-mail: jose.marcato@ufms.br).

Hongjie He and Jonathan Li are with the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: h69he@uwaterloo.ca; junli@uwaterloo.ca).

Digital Object Identifier 10.1109/JSTARS.2020.3028774

I. INTRODUCTION

TO OVERCOME the limitations of the remote sensors and optics manufacturing technology, superresolution (SR) image reconstruction has proven to be very promising in obtaining high-resolution (HR) image from the observed multiple low-resolution (LR) images [1], [2]. The SR image reconstruction process is a morbid inversion process, which usually incorporates prior knowledge or uses a regularization method to constrain its solution space to solve the ill-posed problem. Single-image SR (SISR) aims to obtain an HR output from one of its LR versions, which attempts to improve the resolution by the prior information of one image. Due to the influence of

atmospheric refraction, motion blur, random noise, and insufficient sampling of optical sensors, the quality of satellite images obtained is relatively poor compared with natural images, and the restoration of high-frequency details of satellite images is more difficult than that of natural images. Therefore, the SR problem of single satellite imagery has become a hot research topic in the field of remote sensing image data processing [3].

Several studies suggested that most advanced SR reconstruction methods nowadays consider an example-based reconstruction strategy [4]–[12]. Among them, the sparse coding-based methods are more popular, which uses image training sets to build the relationship between the LR and HR patch pairs [13], [14]. Image SR via sparse representation (SCSR) method proposed in [15] was such a pioneering algorithm. An online dictionary learning algorithm was proposed in [16] to improve the efficiency of dictionary learning, which greatly reduced the computational complexity by processing samples in batches. A novel global joint dictionary model (GJDM) was proposed in [17] to explore the local and global characteristics of images, which effectively improved image reconstruction results. In the above-mentioned reconstruction methods, the dictionary size is an unknown parameter and needs to be set artificially in advance. The Bayesian nonparametric method was used in [18] for SR image reconstruction, which approximated the posterior distribution of the model parameters based on Markov chain Monte Carlo or variable inference, and derived a complete dictionary and sparse coefficient through probability statistics.

The patch-based image reconstruction methods mentioned above obtain the final reconstruction result by averaging the overlapping portions of adjacent image patches. Since the pixels in overlapping areas are often inconsistent, a simple averaging strategy will destroy the spatial structure of the image. To use the consistency prior, an SR reconstruction scheme based on convolutional sparse coding (CSC-SR) was proposed in [19], which takes the entire image as an input and uses convolution filters and feature maps to estimate the target image. Nonlinear convolution operations were used as the mapping function between LR and HR features in a semicoupled convolutional sparse learning method (SCCSL) [20], which reduce the coupling between LR and HR representations. Using the adaptive CSC and convolutional neural network (CNN) to reconstruct the texture part and the smooth part of the input image, respectively, a hybrid adaptive convolutional sparse coding-based SR method was proposed [21]. However, the prior information of the image itself is not introduced into the CSC framework in these methods to further improve the reconstruction effect, and the obtained reconstruction effect is suboptimal.

The above-mentioned SR methods mainly realize the improvement of the spatial resolution of natural images. Compared with natural images, satellite images have the characteristics of wide-range imaging and complex features of ground features. CSC is a global sparse coding method [22], [23]. It directly uses the global correlation of the whole image to reconstruct the local structure of the image robustly, which can ensure the integrity of the feature structure. Currently, a suitable CSC-SR model for satellite images has not been proposed. Satellite images have more NL similar structures as compared with natural images. In this article, we make full use of this favorable prior information to constrain the NL similarity of all feature maps and propose an improved CSC-SR model suitable for satellite imagery.

There are many NL redundant structures on the natural or satellite images, and this critical prior information plays an important role in image restoration [24]–[26]. Some studies prove that combining local sparsity and NL redundant structures can provide more accurate reconstruction models, for example, the concept of sparse coding noise (SCN) presented in [27] and [28]. Due to image damage, blurring, or incompleteness, the sparse coding coefficients used for reconstruction are not the same as the actual sparse coding coefficients. Image patches are nonlocally correlated, and there is a certain correlation between sparse coding coefficients, so using the NL similarity of the image can reduce SCN, thereby improving the quality of the estimated image. Inspired by this, we propose a hybrid NL similarity convolution sparse coding SR (HNLSCSC-SR) method.

The CSC-SR method divides the LR image into a smooth part and a texture part and uses bicubic interpolation and CSC to reconstruct them, respectively. The residual high-frequency information in the smooth part can hardly recover well by the bicubic interpolation and the feature learning of the texture part also lacks local details, resulting in the estimated HR images that are not satisfactory in terms of edges and details. The proposed HNLSCSC-SR method combines the advantages of the Bayesian nonparametric method and the CSC method. On the one hand, the Bayesian nonparametric method uses the external training set composed of the smooth part of a large number of sample images to accurately train a dictionary that fully reflects the structural characteristics of the smooth part of the satellite image, which can better restore the high-frequency information remaining in the smooth image. On the other hand, the CSC method convolves the entire texture image, which can overcome the blocking phenomenon and maintain the spatial structure of the image. Besides, NL self-similarity prior constraints are imposed on all feature maps of the texture part, and the final texture feature maps obtained by solving the constraint model through iterative optimization are closer to the real value, thereby better reconstructing the high-frequency information of the texture part. The NLM algorithm has been approved very effective in sharpening edges and preserving image details [25], [26]. These methods are complementary to each other, making the proposed method highly effective in preserving the edges and details of the image while reducing unexpected noises and artifacts.

Three main contributions of this study can be summarized as follows.

First, this article innovatively introduces the prior information of the NL similarity of the satellite image itself in the CSC framework and proposes an HNLSCSC-SR model suitable for single satellite imagery, which can integrate the advantages of the Bayesian nonparametric method and the CSC method to a certain extent.

Second, the Bayesian nonparametric method is used to better explore the residual high-frequency information of the smooth part by taking advantage of more prior information of the original image than the use of the bicubic interpolation. The texture reconstruction is transformed by considering the convolutional

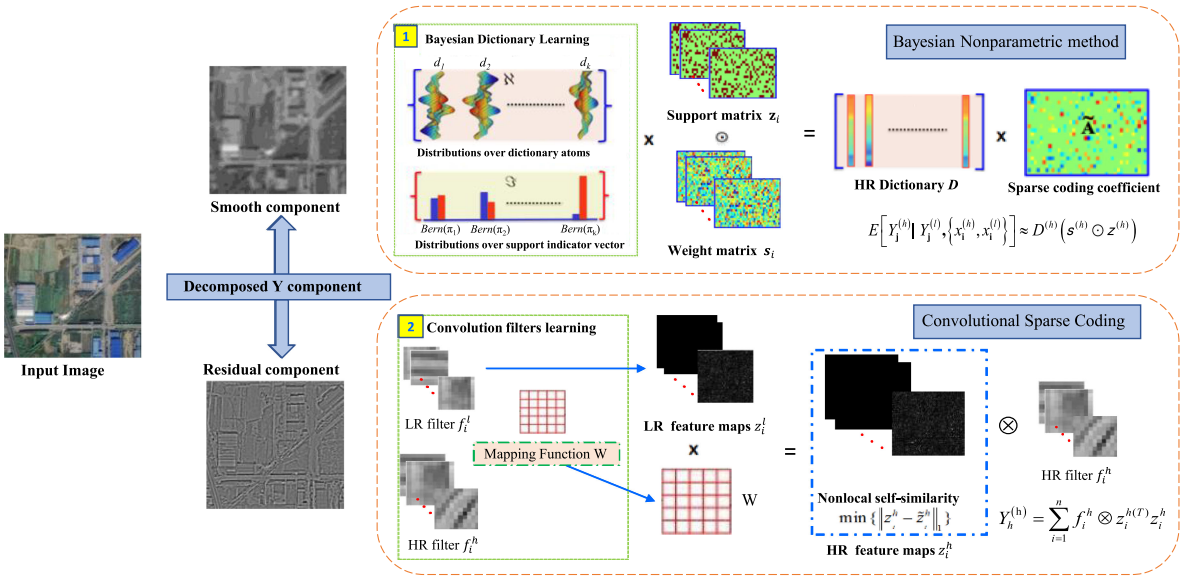

Fig. 1. Reconstruction process of the proposed HNLSCSC-SR method.

sparse coding noise (CSCN) model of the feature maps as an embedded constraint, which minimizes the CSCN.

Third, in the texture reconstruction stage, the correlation coefficients are used as structural information to preselect the most relevant subset of image patches in the search space, which avoids unnecessary weight calculation and reduces the computational burden of the NL mean algorithms.

The rest of this article is organized as follows. Section II describes our proposed method. Section III describes the experimental setup and presents experimental results. Section IV discusses the results and Section V concludes this article.

II. Method

Fig. 1 shows the flowchart of our proposed HNLSCSC-SR method. The LR input image is divided into a smooth part and a texture part. The high-frequency information contained in the smooth part is not obvious and difficult to extract. We use the Bayesian nonparametric method to train and derive the posterior distribution of the model parameters and then reconstruct the smooth part HR image. Besides, we decompose the texture part into the summation of convolutions of the sparse LR feature maps and the corresponding learned LR filters and use the mapping function to project the sparse LR feature maps into the sparse HR feature maps. Then, we make the sparse HR feature maps closer to the real values by using NL similarity constraints. Next, the texture part HR image is reconstructed by the summation of convolutions of the learned HR filters and the optimized sparse HR feature maps. Finally, the smooth part HR image and the texture part HR image are added to obtain the reconstructed HR image.

A. Decomposition of Input LR Image Y

To obtain feature maps with sparser edge textures and details, our method first divides the input LR images Y into a smooth part Y_l and a texture part Y_h. Similar to [19], we use a filtering method to extract the smooth feature map Z_l by solving the following optimization problem:

$$Z_l = \min_{z_l} \|Y - f^s \otimes Z_l\|_F^2 + \gamma \left\|f^h \otimes Z_l\right\|_F^2 + \gamma \|f^v \otimes Z_l\|_F^2 \tag{1}$$

where f^s is a 3×3 low-pass filter and its all entries are $1/9$, and $f^h = [-1, 1]$ and $f^v = [-1, 1]^\top$ are the horizontal and vertical gradient operators. From the smooth feature map Z_l, we can calculate the smooth components Y_l and the corresponding texture components Y_h of the original image Y by the following equation:

$$Y_l = f^s \otimes Z_l, \quad Y_h = Y - Y_l. \tag{2}$$

B. SR for Smooth Components Y_l Based on Bayesian Model

The Bayesian nonparametric model transforms the ill-posed problem into a well-posed problem by introducing prior knowledge. This method can make full use of the potential spatial structure information and more prior knowledge of the image itself to derive the model parameters. The posterior distribution of the model parameters can be adaptively learned directly from the sample dataset, which is more flexible and expandable and can adapt to various uncertain problems [29]–[31]. This method can well recover the residual high-frequency information of Y_l.

The HR dictionary $D^{(h)}$ with the dictionary elements $d_k^{(h)}$ and LR dictionary $D^{(l)}$ with the dictionary elements $d_k^{(l)}$ are learned from a set of training samples composed of HR images $x_i^{(h)}$ and the corresponding LR images $x_i^{(l)}$. $P^{(h)}$ and $P^{(l)}$ represent the dictionary dimensions corresponding to HR and LR dictionaries. I_P represents a $P \times P$ identity matrix. A hierarchical model based on the Beta-Bernoulli process for dictionary learning was proposed in [29]. The probability distribution model of all

hidden variables used in dictionary learning is as follows:

$$\begin{array}{ll} x_i^{(l)} = D^{(l)}(s_i \odot z_i) + \epsilon_i^{(l)} & x_i^{(h)} = D^{(h)}(s_i \odot z_i) + \epsilon_i^{(h)} \\ d_k^{(l)} \sim \mathcal{N}(0, {P^{(l)}}^{-1} I_{P^{(l)}}) & d_k^{(h)} \sim \mathcal{N}(0, {P^{(h)}}^{-1} I_{P^{(h)}}) \\ s_{ik}^{(l)} \sim \mathcal{N}(0, 1/\gamma_{s^{(l)}}) & s_{ik}^{(h)} \sim \mathcal{N}(0, 1/\gamma_{s^{(h)}}) \\ z_{ik} \sim \prod_{k=1}^{K} \text{Bernoulli}\,(\pi_k) & \pi_k \sim \text{Beta}(\tau_0\eta_0, \tau_0(1-\eta_0)) \\ \gamma_s \sim \text{Gamma}(e_0, f_0) & \gamma_\varepsilon \sim \text{Gamma}(c_0, d_0) \\ \epsilon_i^{(l)} \sim \mathcal{N}(0, {\gamma_\epsilon}^{-1} I_{p^{(l)}}) & \epsilon_i^{(h)} \sim \mathcal{N}(0, {\gamma_\epsilon}^{-1} I_{p^{(h)}}) \end{array} \tag{3}$$

where $\odot$ denotes the elementwise product, K is the dictionary size, s_{ik} are the sparse valued weights, and z_{ik} are the binary-valued to record whether the dictionary element is activated. π_k denotes the probability of using element d_k, γ_s and $\gamma_\epsilon$represent the precision of the sparse weights and observation noise, c_0, d_0, e_0, f_0, τ_0, and η_0 denote the hyperparameters, and ϵ_i is the observation noise.

At the stage of dictionary training, the posterior distribution of the coupling dictionary $\{D^{(h)}, D^{(l)}\}$ and hidden variables π, z, s, γ_s, and γ_ϵ can be calculated by the Gibbs sampling method. In the reconstruction stage, these posterior distributions obtained are used as the prior information for SR reconstruction. Specifically, the smooth component Y_l obtained in Section II-A is taken as the observation, and the obtained posterior distributions of the parameters π, γ_s, and γ_ϵ in the dictionary training stage are used as the initial iteration values of the Gibbs sampling process in the reconstruction stage. The dictionary obtained in the dictionary training stage does not change and only uses the formulas (6) and (7) in the Gibbs sampling formula to resample the sparse weight components z_{ik} and s_{ik} associated with Y_l. Finally, the HR image corresponding to Y_l can be reconstructed by

$$E\left[\mathrm{Y}_{\mathrm{j}}^{(h)} | Y_{\mathrm{j}}^{(l)}, \left\{x_{\mathrm{i}}^{(h)}, x_{\mathrm{i}}^{(l)}\right\}\right] \approx \hat{D}^{(h)}\left(s_{jk}^{(h)} \odot z_{jk}^{(h)}\right) = \hat{D}^{(h)} A_{jk}^{(h)} \tag{4}$$

where $\hat{D}^{(h)}$ is the mean of the posterior distribution of the HR dictionaries, and $A^{(h)} = s^{(h)} \odot z^{(h)}$ is the sparse coefficient obtained in the reconstruction phase.

The prior distribution of all hidden variables belongs to the conjugate exponential family. We use the Gibbs sampling approach for posterior reasoning. It is necessary to establish a Markov chain to simulate the sampling of unknown variables. By running it long enough to make the sampling distribution reach a steady state, the stationary distribution is the required posterior distribution. The corresponding sampling formula is as follows [32]–[35]:

Sample d_k: from $\mathcal{N}(\tilde{\mu}_k, \widetilde{\sum}_k)$, where

$$\widetilde{\sum}_k = \left(2P\mathrm{I}_{2P} + \gamma_\epsilon \sum_{i=1}^{N} (s_{ik} z_{ik})^2\right)^{-1} \quad \tilde{\mu}_k = \gamma_\epsilon \widetilde{\sum}_k \sum_{i=1}^{N} s_{ik} \tilde{\mathrm{x}}_{i(-k)}) \tag{5}$$

Sample z_{ik}: from $\text{Bern}(\frac{\varsigma \pi_{k_0}}{1-\pi_{k_0}+\varsigma\pi_{k_0}})$, where

$$\varsigma = \exp\left(-\gamma_\epsilon \left(s_{ik}^2 d_k^{\mathrm{T}} d_k - 2 s_{ik} d_k^{\mathrm{T}} \tilde{\mathrm{x}}_{i(-k)}\right)/2\right) \tag{6}$$

Sample s_{ik}: from $\mathcal{N}(\mu_{ik}, \widehat{\sum}_{ik})$, where

$$\hat{\sum}_{ik} = \left(\gamma_s + \gamma_\epsilon d_k^{\mathrm{T}} d_k\right)^{-1}, \quad \mu_{ik} = \gamma_\epsilon \hat{\sum}_{ik} d_k^{\mathrm{T}} \tilde{\mathrm{x}}_{i(-k)} \tag{7}$$

Sample π_k: from $\text{Beta}(a, b)$, where

$$a = \tau_0\eta_0 + \sum_{i=1}^{N} z_{ik},\ b = N - \sum_{i=1}^{N} z_{ik} + \tau_0\,(1-\eta_0) \tag{8}$$

Sample γ_ϵ: from $\text{Gamma}(c, d)$, where

$$c = c_0 + NP, \quad d = d_0 + \frac{1}{2}\sum_{i=1}^{N} \|\mathrm{x}_i - D_i\,(s_i \odot z_i)\|_2^2 \tag{9}$$

Sample γ_s: from $\text{Gamma}(e, f)$, where

$$e = e_0 + NK/2,\ f = f_0 + \frac{1}{2}\sum_{i=1}^{N} \|s_i\|_2^2 \tag{10}$$

where $\tilde{\mathrm{x}}_{i(-k)} = \mathrm{x}_i - D_i(s_i \odot z_i) + d_k(s_{ik} \odot z_{ik})$ is the reconstruction error using all except the *k*th dictionary element.

C. SR for Texture Components Y_l Based on NL Similarity Constrained Convolutional Sparse Coding (NLSCSC) Method

In this section, we propose an NLSCSC-SR method to reconstruct the texture part. The convolution filters have translation invariant, and for the same features of different positions in the image, the convolution filter can accurately represent them.

1) Obtaining HR and LR Filters and Mapping Functions Between Feature Maps: Given a set of training samples composed of HR images $x_i^{(h)}$ and the corresponding LR images $x_i^{(l)}$. The LR filters $\{f_i^l\}_{i=1}^n$ can be obtained by optimizing the following problem:

$$\begin{aligned} &\min_{\{f_i\}_{i=1}^n} \left\| X^{(l)} - \sum_{i=1}^{n} f_i^l \otimes Z_i^l \right\|_F^2, \\ &\text{s.t.} \quad \left\|f_i^l\right\|_F^2 \leqslant e_n, i = 1, 2, \ldots, n \end{aligned} \tag{11}$$

where e_n is a scalar to constraint the energy of the LR filter.

To save computer memory, the stochastic alternating direction method of multipliers algorithm (SA-ADMM) in [36] was used to solve it. We introduce the augmented variable $s = f^l$ and the Lagrange variable d. Let $Z^l = [(Z_1^l)^{\mathrm{T}}, (Z_2^l)^{\mathrm{T}}, \ldots, (Z_n^l)^{\mathrm{T}}]$ and Q is the upper bound of the eigenvalue of $(Z^l)^{\mathrm{T}} Z^l$, then the iterative solutions of LR filters in (11) are

$$\begin{aligned} \left(f^l\right)^{(t+1)} &= \left(\frac{Q}{n}\sum_{i=1}^{n} \left(f^l\right)^{(t)} - \rho\left(d^{(t)} - s^{(t)}\right)\right. \\ &\quad \left. - \frac{1}{n}\sum_{i=1}^{n} \left(Z_i^l\right)^T \left(Z_i^l \left(f^l\right)^{(t)} - X^{(l)}\right)\right) \Big/ (\rho + Q) \\ s^{(t+1)} &= \arg\min_s \frac{\rho}{2}\left\|\left(f^l\right)^{(t+1)} + d^{(t)} - s^{(t)}\right\|_F^2 \\ d^{(t+1)} &= d^{(t)} + \left(f^l\right)^{(t+1)} - s^{(t+1)}. \end{aligned} \tag{12}$$

With the obtained LR filters $\{f_i^l\}_{i=1}^n$ from (11), we will learn the LR feature maps $\{Z_i^l\}_{i=1}^n$ by the following optimization:

$$\begin{aligned} &\min_{\{z_i\}_{i=1}^n, x} \left\| X^{(l)} - \sum_{i=1}^{n} f_i^l \otimes Z_i^l \right\|_F^2 + \lambda^l \sum_{i=1}^{n} \left\|Z_i^l\right\|_1, \\ &\text{s.t.} \quad \left\|f_i^l\right\|_F^2 \leqslant 1 \end{aligned} \tag{13}$$

we solve (13) by using ADMM [37]. Introducing the auxiliary variable $M_i = Z_i^l$ and the dual variable U_i, the iterative solutions of LR feature maps in (13) are

$$\begin{aligned}\left(Z_i^l\right)^{(k+1)} = & \arg\min_{z^l}\left(\left\|X^{(l)} - \sum_{i=1}^{n} f_i^l \otimes Z_i^l\right\|_F^2\right. \\ & \left. + \frac{\rho}{2}\left\|Z_i^l - M_i^{(k)} + U_i^{(k)}\right\|_F^2\right) \\ M_i^{(k+1)} = & \arg\min_{M_i}\left(\lambda^l \sum_{i=1}^{n}\|M_i\|_1\right. \\ & \left. + \frac{\rho}{2}\left\|\left(Z_i^l\right)^{(k+1)} - M_i + U_i^{(k)}\right\|_F^2\right) \\ U_i^{(k+1)} = & U_i^{(k)} + \left(Z_i^l\right)^{(k+1)} - M_i^{(k+1)}. \end{aligned} \tag{14}$$

Assume that there is a mapping function W that can scale the LR feature maps to the HR feature maps in terms of space size and the number of feature maps, that is [19]

$$\begin{aligned} & Z_j^h(kx, ky) = g\left(Z_{:}^l(x, y); w_j\right) = w_j^{\top} Z_{:}^l(x, y), \ \text{s.t.} \\ & \quad w_j \succcurlyeq 0, |w_j|_1 = 1 \end{aligned} \tag{15}$$

where $Z_{:}^l(x, y)$ is a vector composed of all the coefficients of N LR feature maps in point (x, y), and w_j is a transformation vector for the HR feature map Z_j^h. The joint optimization training model of HR filters $\{f_i^h\}_{i=1}^n$ and the mapping function W is constructed as follows:

$$\begin{aligned} & \min_{\{w_j\}_{i=1}^n, \{f_j^h\}_{j=1}^n} \sum_{k=1}^{K}\left\|X^{(h)} - \sum_{j=1}^{n} f_j^h \otimes g\left(U_{k,:}^l; w_j\right)\right\|_F^2 \\ & \text{s.t.} \quad \left\|f_j^h\right\|_F^2 \leqslant e_n \\ & w_j \succcurlyeq 0, |w_j|_1 = 1, j = 1, 2, \ldots, m \end{aligned} \tag{16}$$

where e_n is a scalar to constraint the energy of the HR filters. The HR filters and the mapping function W are alternately solved by SA-ADMM.

2) NLSCSC-SR Method Constrained With SCN: Using the learned LR filter $\{f_i^l\}_{i=1}^n$, HR filter $\{f_i^h\}_{i=1}^n$, and the mapping function W, We take the texture part Y_h obtained in Section II-A as the observations to perform NLSCSC-SR reconstruction. First, the LR filters $\{f_i^l\}_{i=1}^n$ are used to perform convolution decomposition Y_h to obtain the LR feature maps $\{Z_i^l\}_{i=1}^n$, and then the mapping function W is used to project the LR feature maps $\{Z_i^l\}_{i=1}^n$ into the HR feature maps $\{Z_i^h\}_{i=1}^n$.

Due to the influence of blur and noise in the process of original image acquisition, there is a deviation between Z^h obtained and the real feature map $\bar{Z}^h$. Some studies prove that the SCN can achieve good SR reconstruction results [27], [28]. Inspired by this, we introduce the SCN based on feature maps in the CSC model. Its basic form is $v_z = Z^h - \bar{Z}^h$. Since $\bar{Z}^h$ is unknown, we use the NL similarity prior of the image to make a reasonable estimation $\tilde{Z}^h$ for $\bar{Z}^h$. We take $v_z \approx Z^h - \tilde{Z}^h$ as the regularization constraint and iteratively solve (17) to make the SCN get the minimum value, to obtain the optimal feature map $Z^{(h)*}$ closest to the true value $\bar{Z}^h$. Due to the sparsity of $v_z \approx Z^h - \tilde{Z}^h$, the SR frame can be simplified as follows:

$$Z^h = \arg\min_{z_i^h}\left\{\left\|y - DB\sum_{i=1}^{n} f_i^h \otimes z_i^h\right\|_F^2 + \sum_{i=1}^{n}\lambda_i\left\|z_i^h - \tilde{z}_i^h\right\|_1\right\} \tag{17}$$

where D is a downsampling operator, B is a blurring operator, and λ_i denotes the regularization parameter.

First, the initial HR feature map $\{\hat{Z}_i^h\}_{i=1}^n$ is obtained by projecting the LR feature map $\{Z_i^l\}_{i=1}^n$ directly using the projection function W. We use a convolution operation to obtain a less accurate initial reconstructed image $\hat{Y}_h$ of the texture part Y_h. Divide $\hat{Y}_h$ into nonoverlapping image patch sets $Y' = \{y_i\}_{i=1}^b$ with patch size $a \times a$, where b is the number of image patches. Previous NL means algorithms used Euclidean distance to measure the self-similarity between image patches [27], [28]. The calculation of Euclidian distance is just a time-consuming process and the additional information contained in image patches with a similar structure cannot be expressed by this distance. When two image patches have a linear relationship, they have also been proved to be similar [38]. We preselect the most relevant patch subset in the search space to remove irrelevant patches and speedup the weight calculation process in NL means algorithms.

We use the correlation coefficient between any two image patches as structural information to classify the search space, extract the image patches with the most relevant structural information, only calculate the Euclidean distance between these related image patches, arrange these Euclidean distances in descending order, and select the first L image patches for weighted calculation. Assuming that x_1 and $x_2$are the vectors corresponding to two image patches, the expression of the correlation coefficient between them is

$$\rho\left(x_1, x_2\right) = \operatorname{cov}\left(x_1, x_2\right) / \sigma_{x_1}\sigma_{x_2} \tag{18}$$

where σ_{x_1} and σ_{x_2} represent the standard deviations of x_1 andx_2, respectively, and $\operatorname{cov}(x_1, x_2) = \mathrm{E}((\mathrm{x}_1 - \mu_{\mathrm{x}_1})(\mathrm{x}_2 - \mu_{\mathrm{x}_2}))$represents the covariance between two image patches. We calculate the correlation coefficient between each image patches $y_i (i = 1, 2, \ldots, b)$ and other image patches $Y' = \{y_i\}_{i=1}^b$. By setting the threshold conditions of structural information, the classification calculation of the necessary weights can be performed well

$$S_w(y_i, y_j) = \begin{cases} \exp\left(-\left\|y_i - y_j\right\|_2^2 / h\right) & \rho \geq \eta \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

where η is a predefined value. Then, we take the image patches with the largest L Euclidean distance in the most relevant patch subset as the most similar image patches to the given image patch and record the index position information of the L image patches to form a set Ω_i.

Second, in the convolution operation, the convolution filter extracts the global features of the entire image. The feature map corresponding to each filter should have the same structure information corresponding to the global features. Based on this characteristic, we perform the same NL means algorithms on the initial value of each HR feature map $\{\hat{Z}_i^h\}_{i=1}^n$. Each HR feature map is divided into nonoverlapping image patch sets $\hat{z}_i^{(t)} = \{\hat{z}_{ij}^{(t)} = R_j \hat{z}_i^{(t)}\}_{j=1}^b$ with patches size $a \times a$, where R_j is the patch extraction matrix for the patch j. We can notice that the image patches in the same location of each feature map have the same similar patch index location information. According to the index position information set Ω_i of image similar patches of global features obtained in the previous step, the approximate value $\tilde{Z}_i^h$ of each HR feature map added with NL similar information is calculated by using NL means algorithm [39], that is

$$\tilde{z}_i^{(t+1)} = \sum_{j=1}^{b} R_j^T \sum_{p \in \Omega_j} S_{w_{ij,p}} \hat{z}_{ip}^{(t)}$$
$$S_{w_{ij,p}} = \frac{1}{S_{W_{ij}}} \exp\left(-\left\|\hat{z}_{ij}^{(t)} - \hat{z}_{ip}^{(t)}\right\|_2^2 / h\right), S_{W_{ij}} = \sum_{p \in \Omega_j} S_{w_{ij,p}} \tag{20}$$

where t is the number of iterations and h is a scalar. $S_{W_{ij,p}}$ is the weight, and R_j^T is the inverse process of R_j.

Third, the approximate value $\tilde{Z}^h$ of each HR feature map in the texture part can be calculated by using the NL method. To minimize the SCN $v_z \approx Z^h - \tilde{Z}^h$ of the feature maps, $\tilde{Z}^h$ needs to be substituted into the sparse reconstruction model in (17) for joint optimization, and iterative update value $Z^{h(t+1)}$ of each HR feature map that is closer to the true value is solved. The optimization process can use an iterative threshold algorithm [40]

$$z_i^{h(t+1)} = S_\tau\left(v_i^{(t)} - \tilde{z}_i^{(t+1)}\right) + \tilde{z}_i^{(t+1)} \tag{21}$$

where $v_i^{(t)} = f_i' \otimes ((DB)^T \times (y - DB\sum_{i=1}^n f_i \otimes z_i^{(t)}))/c + z_i^{(t)}$, f_i' is the corresponding value after f_i is rotated 180°, and $\tau = \lambda_i / c$, where c is an auxiliary parameter that guarantees the convexity of the function. $S_\tau(\cdot)$ is a soft threshold operation, $\lambda_i = 2\sqrt{2}/\sigma_i$, and $\sigma_i^2 = \frac{1}{n_1 \times n_2}\sum_{d=1}^{n_i}\sum_{e=1}^{n_2}\left(z_{de}^{(i)} - \tilde{z}_{de}^{(i)}\right)^2$, where n_1 and $n_2$are the numbers of rows and columns of the feature map. $z_{de}^{(t)}$ and $\tilde{z}_{de}^{(t)}$ represent the element values of $z_i^{(t)}$ and $\tilde{z}_i^{(t)}$, respectively.

Finally, using the learned HR filter $\{f_i^h\}_{i=1}^n$ and the updated HR feature map $\{Z_i^{h(t+1)}\}_{i=1}^n$ for convolution operation, the HR image $Y_h{}^{(t+1)} = \sum_{i=1}^n f_i^h \otimes z_i^{h(t+1)}$ under the current iteration can be obtained. We continue to repeat steps (1)–(3) until the iteration termination condition is reached (i.e., $t = T$, T is the set total number of iterations) and obtain the optimal value of the feature map $Z^{(h)*} = \{z_i^{h(T)}\}_{i=1}^n$. Finally, the texture part of the input image is reconstructed by $Y_h^{(\mathrm{h})} = \sum_{i=1}^n f_i^h \otimes z_i^{h(T)}$. Therefore, the final HR image is: $Y^* = Y_l^{(h)} + Y_h^{(h)}$. Finally, we also use back projection [41] to obtain more robust reconstruction results. We summarize the proposed HNLSCSC-SR method in Algorithm 1.

Algorithm 1: The Proposed Hybrid NL Similarity Convolution Sparse Coding SR (HNLSCSC-SR) Method.

Input: the input image Y, two training image sets $\{x^{(h)}, x^{(l)}\}$ and zooming factor k, iteration times T, auxiliary parameter M, c, η.
Step 1: Decompose the LR image Y into the smooth component Y_l and the texture component Y_h.
Step 2: Reconstruct the HR image $Y_l^{(\mathrm{h})}$ of the smooth part $Y_l$based on the Bayesian nonparametric method as (3)–(10).
Step 3: Reconstruct the HR image part $Y_h^{(\mathrm{h})}$of the texture part $Y_h$by the proposed NLSCSC-SR method.
(1) Learn the LR filters $\{f_i^l\}_{i=1}^n$, HR filters $\{f_i^h\}_{i=1}^n$, and mapping function W base on the training set $\{x^{(h)}, x^{(l)}\}$ by the SA-ADMM algorithm [32].
(2) Get the LR feature maps $\{Z_i^l\}_{i=1}^n$ by solving ADMM [46], and combine W to get the initial value of HR feature maps$\{\hat{Z}_i^h\}_{i=1}^n$.
(3) Construct the SCN of the feature maps, use structure information to classify the search space and perform effective NL means, and get the optimal HR feature maps $\{Z_i^{h(T)}\}_{i=1}^n$ of the texture part Y_h by using (17)–(21).
(4) Obtain the HR reconstruction image $Y_h^{(\mathrm{h})}$ of the texture part Y_h.
Step 4: Combine $Y_l^{(\mathrm{h})}$ and $Y_h^{(\mathrm{h})}$to get the HR reconstructed image Y^* of the input image.
Output: The corresponding HR image Y^*.

III. Experiments and Results

In this study, extensive SR reconstruction experiments were carried out to assess the performance of the proposed method. Considering satellite images obtained from GF-1, ZY-3, WV-2, and TM-5 as test images, with spatial resolutions of 2, 2.1, 0.5, and 30 m, respectively. We compare our method with some other state-of-the-art methods, including Bicubic, SCSR [15], NPDDR [42], GJDM [17], LapSRN [43], RCAN [44], CSC [19], and HCSC [21]. All the experimental parameters of the comparison algorithms are referenced to the original literature. The brightness component (Y) of the LR image is reconstructed, and the chroma channel (CB, Cr) is only performed by bicubic interpolation. We use peak signal-to-noise ratio (PSNR) and feature similarity (FSIM) [45] to evaluate the quality of the estimated image quantitatively. The absolute error map is used for visual evaluation. The brighter the points are on the error map, the larger the difference exists, and the darker the points are, the smaller the difference exists. The LapSRN developed on the MatConvNet package [46], which is a convolutional neural network framework for MATLAB. The RCAN was implemented in the Tensorflow. Except for RCAN, all other experiments were

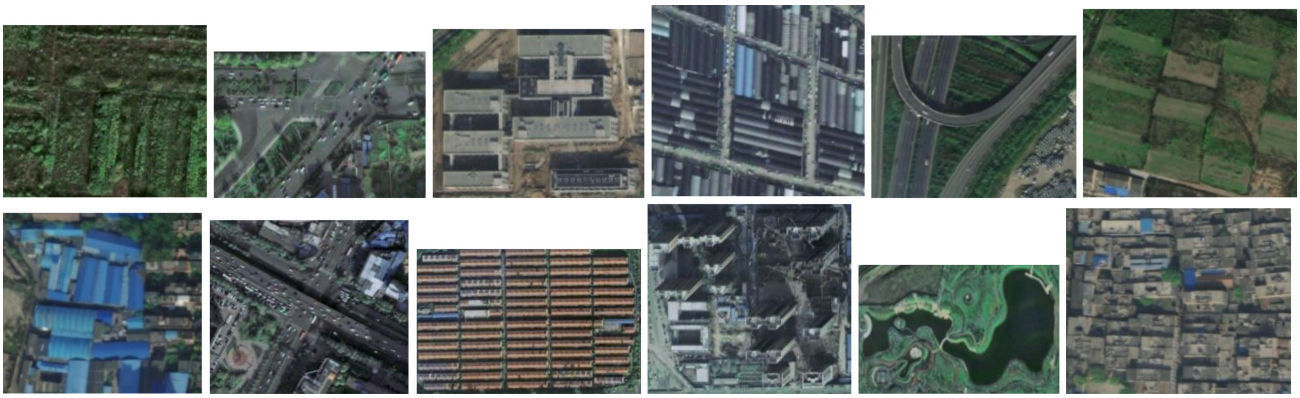

Fig. 2. Several training samples used to construct a dictionary or filter sets in our method.

performed on MATLAB R2018b in a computer with Intel (R) Xeon (R) CPU E5-2630 V3.

This article uses satellite images of different sizes, different sensors, and different noise types as experimental data. The Bayesian nonparametric method is used to reconstruct the smooth part, and the NLSCSC-SR method is used to reconstruct the texture part. Both need to use the training sample set to learn dictionary elements or filters in the early stage. Based on the principle of selecting as many feature types as possible, 100 GF-1 orthoimages of different sizes with the spatial resolution of 2 m were intercepted to form the training sample set. Fig. 2 shows several training samples used in our method to construct a dictionary or filter sets.

Different from Zhou's method [18], we use the Bayesian nonparametric method to reconstruct the smooth part of the test image. Therefore, in dictionary training, we first separate all the sample images into smooth and texture parts. Then, we only train the smooth part of all sample images to get the joint dictionary of high frequency and low frequency, which can reflect the structural characteristics of the smooth image. To make a fair comparison, in the LR and HR filters training stage for the CSC and HCSC, we randomly cropped 1000 72×72 images from 100 GF-1 orthoimages to train LR/HR filters. For other comparative methods, their models are also retrained on this sample training set.

When the Bayesian nonparametric method is used for the smooth part, the parameter settings are as follows: in the training stage, the 100 GF-1 orthoimages and corresponding LR images are divided into 8×8 patches with the maximum overlap; 10^5 HR and LR image patch pairs are randomly selected to form the training set for dictionary learning; the initial values of dictionary elements and other model parameters are initialized through the singular value decomposition algorithm; and the hyperparameters are $c_0 = d_0 = e_0 = f_0 = 10^{-6}$, $\tau_0 = 2$, and $\eta_0 = 0.5$. We used 5000 Gibbs samples for our model, where the burn-in is 4500 samples and the rest 500 samples are used to approximate the posterior distribution of dictionary elements and other parameters. The nonparametric derivation of dictionary size in this study shows that the best dictionary sizes for SR reconstruction corresponding to the zooming factors 2, 3, and 4 are 630, 619, and 564, respectively.

When using the proposed NLSCSC-SR method to reconstruct the texture part, the parameter settings are as follows: the regularization parameter γ is 30, λ^l is 0.02, and the energy limiting parameter e_n with the zooming factors 2, 3, 4, 9, and 12. The size of the LR filter is 5×5, and its number is 800. The corresponding HR filter is $5 \times k$, where k is the magnification, and its number is 1200. In the reconstruction process, each feature map is divided into nonoverlapping patches of size 6×6. The position index set Ω_i of similar patches of each feature map includes the $L = 9$ most similar image patches. We use the target image patch as the center to search and match similar image patches within the range of 60×60 and set h as 75, c as 1, and η as 0.7.

A. SR Reconstruction of Satellite Images With different Sizes

In this experiment, five ZY3 orthoimages with different sizes and without any noise are used as test images. The test images are all square and the side lengths are 240, 360, 480, 600, and 720 pixels. Considering the NL similar structure, the test images with buildings as the main feature are selected. Fig. 3 shows the LR input images obtained from the original HR reference images under the zooming factor 3 (45% of the original scale in the figure due to the large size of the original image).

To compare the performance of different reconstruction methods, this experiment performed the SR reconstruction for the satellite images with different pixel sizes under the zooming factors 2, 3, and 4. All reconstruction results are visually and quantitatively compared. Table I shows the PSNR and FSIM comparison results of the proposed method with other state-of-the-art methods, where the zooming factor is 3. Fig. 4 shows the PSNR comparison histograms of the reconstructed images under the zooming factors 2 and 4. It can be seen that as the size of the test image increases, the quantitative evaluation value of the SR reconstruction based on the learning method is far superior to the interpolation method. Because the Bayesian nonparametric method is used to extract the high-frequency information remaining in the smooth part, and the convolution sparse noise of nonlocally similar structures is iteratively updated to the minimum value, which fully considered in the texture part, the proposed method is far superior to other methods, and its performance is the best in PSNR and FSIM evaluation indicators.

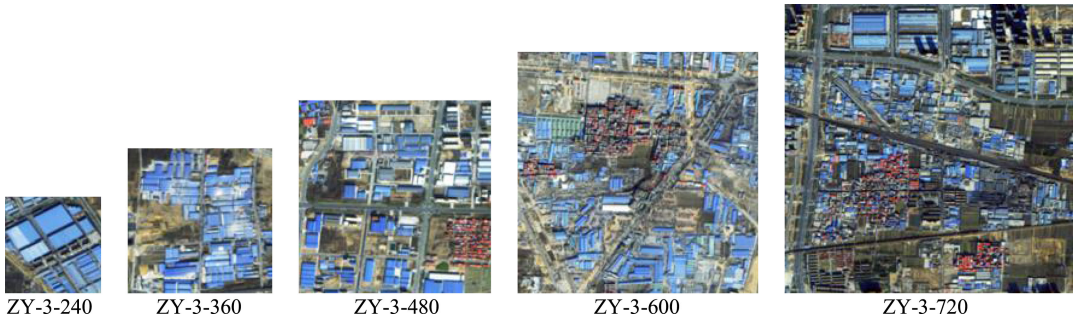

Fig. 3. LR input images with different pixel sizes obtained from the corresponding original HR images under the zooming factor 3.

TABLE I
PSNR AND FSIM COMPARISON OF DIFFERENT RECONSTRUCTION METHODS ON DIFFERENT SIZE IMAGES UNDER THE ZOOMING FACTOR 3

Measures	ZY3_240		ZY3_360		ZY3_480		ZY3_600		ZY3_720	
	PSNR	FSIM	PSNR	FSIM	PSNR	FSIM	PSNR	FSIM	PSNR	FSIM
Bicubic	36.3970	0.9415	35.0184	0.9579	34.3162	0.9705	33.6053	0.9711	32.8904	0.9848
SCSR[15]	37.0565	0.9495	35.5900	0.9642	34.7672	0.9864	34.3086	0.9877	33.4158	0.9948
NPDDR[42]	37.9491	0.9628	35.9262	0.9770	35.1408	0.9892	34.6483	0.9876	33.6541	0.9952
GJDM[17]	38.0256	0.9602	36.1158	0.9689	35.4625	0.9890	34.6529	0.9895	33.7221	0.9955
LapSRN[43]	38.4483	0.9606	36.3412	0.9760	35.5898	0.9902	34.7573	0.9894	33.8684	0.9955
RCAN[44]	38.6229	0.9638	36.7046	0.9793	35.8775	0.9910	34.9462	0.9896	33.9690	0.9957
CSC[19]	38.1141	0.9572	36.2715	0.9745	35.5476	0.9902	34.7874	0.9894	33.8891	0.9957
HCSC[21]	38.2138	0.9578	36.3198	0.9756	35.5958	0.9906	34.8550	0.9896	33.9428	0.9957
Ours	**38.6676**	**0.9656**	**36.7143**	**0.9815**	**35.9241**	**0.9915**	**35.0473**	**0.9902**	**34.0563**	**0.9958**

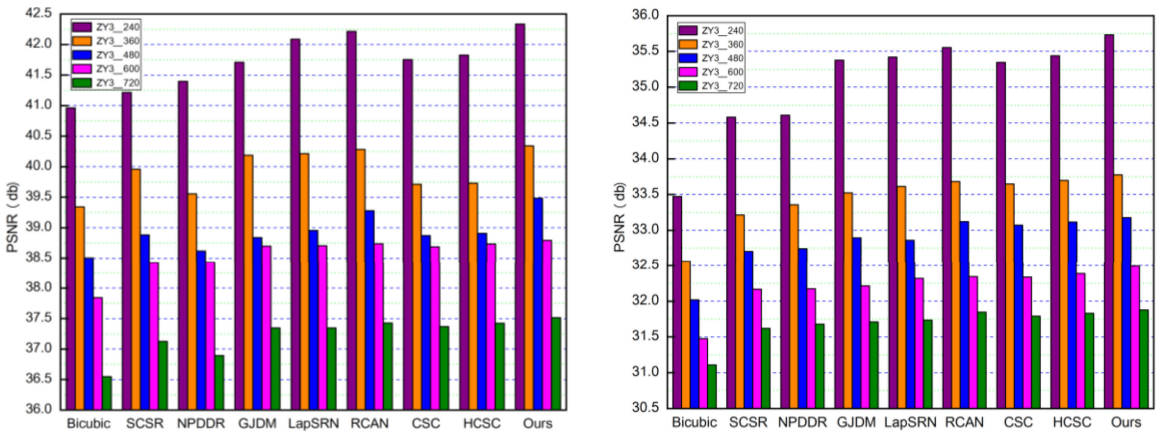

Fig. 4. PSNR comparison of several reconstruction methods on different image sizes under different zooming factors (left: factor = 2; right: factor = 4).

Taking the reconstruction result of the test image with side length 480 under the zooming factor 3 as an example, the PSNR value of the proposed method is 1.1569 dB higher than that of SCSR, 0.3765 dB higher than that of CSC, and the FSIM value is 0.9915 for the proposed method, which is also the best than other methods. It shows that the proposed method can better recover image quality and maintain the structural characteristics of the original image.

The method strategy adopted in this article significantly improves the performance of the CSC-SR method. Image size is an important factor for the self-similarity of image features. The change in image size causes a change in the proportion of similar structures in the image, which has a significant impact on the results of SR reconstruction. As shown in Table I, with the increase of image size, on the one hand, the PSNR value of reconstruction results of all methods gradually decreases, and the PSNR value of our method always decreases the slowest. On the other hand, because the proportion of self-similar structure contained in each image is different, the FSIM value of the reconstruction image of all methods fluctuates with the image

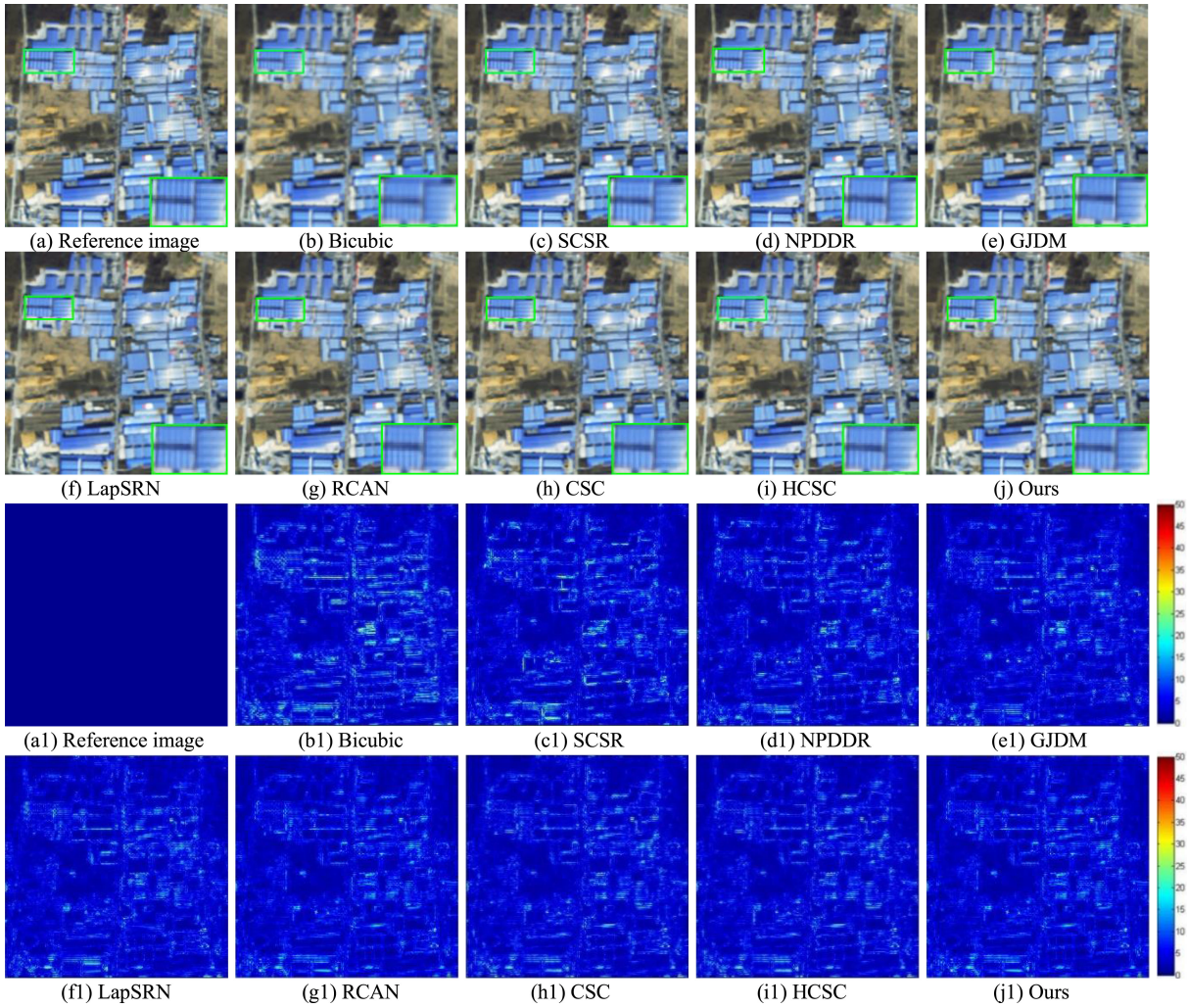

Fig. 5. SR results and absolute error images with respect to ZY3-360 under the zooming factor 3. (a) Ground truth HR image. (b)–(h) SR results from bicubic, SCSR [15], NPDDR [42], GJDM [17], LapSRN [43], RCAN [44], CSC [19], HCSC [21], and our method. (b1)–(h1) Absolute error images with respect to the ground truth from bicubic, SCSR [15], NPDDR [42], GJDM [17], LapSRN [43], RCAN [44], CSC [19], HCSC [21], and our method.

size, generally showing an increasing trend, and the FSIM value of our method is also the best than the others.

We take the reconstruction result of the test image with side length 360 under the zooming factor 3 as an example. We compare and analyze the SR reconstruction effect of different methods and the corresponding absolute error map. As shown in Fig. 5, the image reconstructed by the bicubic interpolation method is excessively smooth and fuzzy. The reconstruction effect of SCSR and BPJDL is improved, but as shown in the magnified part of the test image, their reconstructed image shows an obvious block effect because of the patch strategy. The SRCNN algorithm uses a CNN model to greatly improve the visual effect of the image, but from the magnified part of the image, the reconstruction image still has aliasing and ghosting. The CSC and HCSC perform SR reconstruction of the entire image, effectively avoiding blocky images, and the contours and textures of the reconstructed image are relatively clear. Our method regularizes the reconstruction model by using the NL self-similarity of the feature maps, iteratively optimizes the minimum value of the convolution sparse noise, so that the feature map is closer to its real value, so the edge contour and texture structure of the reconstructed image obtained by our method are more obvious. Besides, the Bayesian nonparametric method is used to perform the residual high-frequency information based on the prior information of the smooth part to retain more details.

B. SR Reconstruction of Satellite Images of Different Sensors

This experiment aims to study the SR reconstruction of satellite images with different resolutions, which are obtained from the different sensors. According to Experiment A, the SR reconstruction effect of satellite images with different pixel sizes is different, so this experiment selected four types of remote sensing sensors with the same pixel size (400×400) for the experiments. In this experiment, eight images (as shown in Fig. 6) without any noise are reconstructed under zooming factors 2, 3, and 4. Landsat TM-5 images have the lowest spatial resolution,

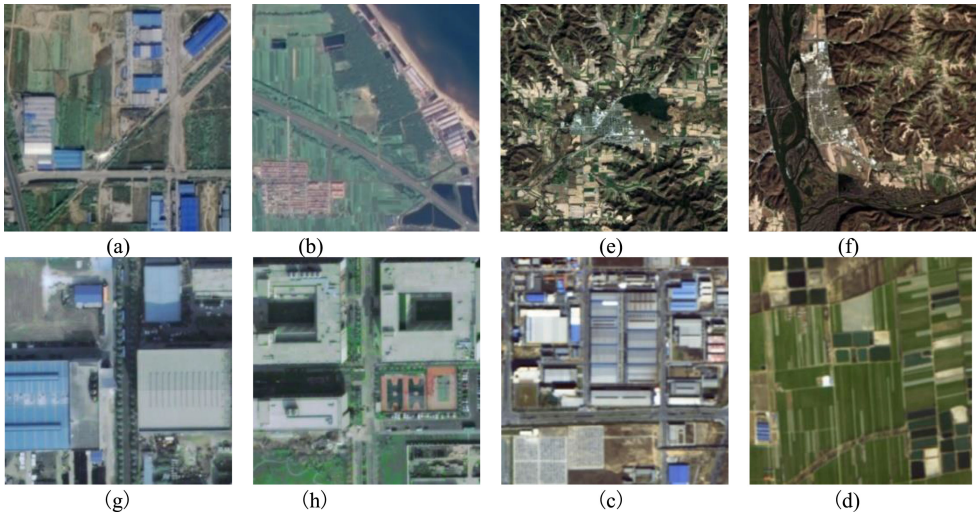

Fig. 6. Test images of different sensors selected in this experiment. (a) GF1_1. (b) GF1_2. (c) ZY3_1. (d) ZY3_2. (e) TM5_1. (f) TM5_2. (g) WV2_1. (h) WV2_1.

TABLE II
PSNR AND FSIM COMPARISON OF DIFFERENT RECONSTRUCTION METHODS ON DIFFERENT SENSOR IMAGES UNDER THE ZOOMING FACTOR 2

	Bicubic	SCSR[15]	NPDDR[42]	GJDM[17]	LapSRN[43]	RCA[44]	CSC[19]	HCSC[21]	Ours
GF1_1	40.2692	42.8467	42.7753	42.8059	43.1341	43.2772	42.9119	43.0391	**43.4403**
	0.9965	0.9992	0.9995	0.9992	0.9995	0.9995	0.9995	0.9995	**0.9996**
GF1_2	42.5759	45.8324	45.7725	45.9552	46.1710	46.6490	45.9070	46.1257	**46.7212**
	0.9973	0.9996	0.9997	0.9996	0.9997	0.9997	**0.9998**	**0.9998**	**0.9998**
TM5_1	31.3681	31.9964	32.1466	32.1699	32.1789	**32.2062**	32.1344	32.1666	32.1570
	0.9907	0.9974	0.9974	0.9973	**0.9976**	**0.9976**	0.9975	0.9975	0.9975
TM5_2	31.9644	32.6037	32.7814	32.7612	32.8031	**32.8440**	32.7443	32.7834	32.7833
	0.9907	0.9975	0.9975	0.9974	0.9976	**0.9977**	0.9975	0.9976	0.9976
WV2_1	39.6485	41.4743	41.6905	41.8526	41.9097	42.0226	41.8759	41.9478	**42.0382**
	0.9970	0.9993	0.9994	0.9994	0.9994	0.9994	**0.9995**	**0.9995**	**0.9995**
WV2_2	38.1214	39.7690	39.7712	39.9084	39.9636	40.0199	39.9297	40.0079	**40.0595**
	0.9966	0.9993	0.9993	0.9993	**0.9994**	0.9993	**0.9994**	**0.9994**	**0.9994**
ZY3_1	39.2581	39.5603	39.5916	39.9381	39.9722	40.1472	39.7287	39.7471	**40.1590**
	0.9974	0.9994	0.9994	0.9981	0.9993	0.9994	0.9994	0.9994	**0.9995**
ZY3_2	43.0719	43.1127	43.1122	43.1669	43.4140	43.4372	43.1342	43.1684	**43.5324**
	0.9986	0.9995	0.9996	0.9990	0.9996	0.9996	0.9996	0.9996	**0.9997**

and the features and boundaries in the images are blurred. The other three kinds of images with higher spatial resolution have clear features and boundaries. To verify the superiority of the proposed method, the test images contain different proportions of nonlocally similar structures, such as buildings and cultivated land. The quantitative comparison of the reconstruction results in three cases is presented in Tables II –IV.

We can see that in most reconstruction experiments, our proposed method shows good PSNR and FSIM values, and the higher the spatial resolution, the better the image reconstruction effect. Simultaneously, the richer the nonlocally similar structure of the image acquired by the same sensor, the better the reconstruction effect. For example, the ground features in GF1_1 are mainly buildings and cultivated land, and the ground features in GF1_2 are mainly cultivated land and waters. The image comparison revealed that the NL similarity structure in the GF1_2 was significantly superior to the GF1_1. As shown in Tables II–IV, the PSNR and FSIM of the reconstruction results of GF1_2 under the zooming factors 2, 3, and 4 are always better than GF1_1.

It is worth noting that the SR reconstruction results of CSC and HCSC for the ZY3_2 image are inferior to the results of SRCNN under the zooming factors 2 and 3. However, the proposed method has achieved excellent reconstruction results, and the PSNR value is 0.047 and 0.204 dB higher than that by SRCNN. It shows that our method improves the performance of CSC and HCSC very well. After the above analysis, the proposed method has a better geometrical structure than other methods and improves the reconstruction quality of most test images. Obviously, the higher the resolution and the more abundant the NL similar structure the image has, the better the reconstruction effect our method achieves.

In satellite images, besides buildings, cultivated land is also one of the typical features of nonlocally similar structures. The cultivated land is often distributed in large areas, and the shape of the cultivated land is relatively regular and similar, which

TABLE III
PSNR AND FSIM COMPARISON OF DIFFERENT RECONSTRUCTION METHODS ON DIFFERENT SENSOR IMAGES UNDER THE ZOOMING FACTOR 3

	Bicubic	SCSR[15]	NPDDR[42]	GJDM[17]	LapSRN[43]	RCA[44]	CSC[19]	HCSC[21]	Ours
GF1_1	35.8749	36.8939	37.6245	37.6815	37.8162	37.8609	37.7239	37.8085	**37.9178**
	0.9663	0.9837	0.9851	0.9855	0.9861	0.9863	0.9863	**0.9864**	**0.9864**
GF1_2	37.7212	38.8399	39.2369	39.3940	39.4049	39.4362	39.4144	39.4996	**39.6249**
	0.9710	0.9871	0.9876	0.9893	0.9893	0.9894	0.9894	**0.9895**	**0.9895**
TM5_1	30.0886	30.4025	30.4893	30.5257	30.5715	**30.6111**	30.5327	30.5505	30.5645
	0.9422	0.9673	0.9682	0.9578	0.9713	**0.9717**	0.9694	0.9697	0.9698
TM5_2	30.5709	30.9061	30.9841	31.0621	31.1035	**31.1503**	31.0551	31.0772	31.0740
	0.9411	0.9676	0.9685	0.9723	0.9734	**0.9735**	0.9707	0.9708	0.9708
WV2_1	36.3848	36.9417	37.5170	37.5392	37.5672	37.6430	37.7030	37.7248	**37.7269**
	0.9751	0.9859	0.9855	0.9863	0.9808	0.9884	**0.9893**	**0.9893**	**0.9893**
WV2_2	34.9761	35.5834	35.9821	36.0342	36.2344	36.2650	36.2576	36.2962	**36.3025**
	0.9732	0.9851	0.9844	0.9859	0.9830	0.9839	0.9879	**0.9882**	**0.9882**
ZY3_1	35.0656	35.5375	36.0686	35.8691	35.9070	36.4558	36.3724	36.4407	**36.7182**
	0.9713	0.9866	0.9855	0.9878	0.9875	0.9910	0.9905	0.9909	**0.9913**
ZY3_2	39.2212	39.4663	39.9888	40.0717	39.9611	40.2161	39.9129	39.9549	**40.4266**
	0.9841	0.9935	0.9921	0.9947	0.9953	0.9953	0.9946	0.9947	**0.9958**

TABLE IV
PSNR AND FSIM COMPARISON OF DIFFERENT RECONSTRUCTION METHODS ON DIFFERENT SENSOR IMAGES UNDER THE ZOOMING FACTOR 4

	Bicubic	SCSR[15]	NPDDR[42]	GJDM[17]	LapSRN[43]	RCA[44]	CSC[19]	HCSC[21]	Ours
GF1_1	33.8187	34.5591	34.7070	34.8532	34.9248	35.0607	34.9577	35.0258	**35.0786**
	0.9185	0.9436	0.9414	0.9479	0.9435	0.9479	0.9505	0.9507	**0.9511**
GF1_2	35.5928	36.4493	36.4146	36.6804	36.6730	36.7056	36.6927	36.7744	**36.7934**
	0.9264	0.9441	0.9457	0.9483	0.9405	0.9497	0.9568	0.9573	**0.9574**
TM5_1	29.4222	29.7042	29.7076	29.7279	29.7519	**29.7981**	29.7383	29.7412	29.7577
	0.8804	0.9142	0.9088	0.9118	0.9193	**0.9198**	0.9172	0.9175	0.9178
TM5_2	29.8482	30.1664	30.1987	30.2093	30.2333	**30.2791**	30.2202	30.2402	30.2324
	0.8766	0.9147	0.9072	0.9203	0.9231	**0.9232**	0.9203	0.9210	0.9209
WV2_1	34.6654	35.2804	35.3565	35.5979	35.6916	35.8527	35.8150	35.8572	**35.8707**
	0.9400	0.9582	0.9554	0.9630	0.9562	0.9584	0.9663	0.9664	**0.9667**
WV2_2	33.4286	33.9610	34.0225	34.0992	34.2561	34.3060	34.2786	34.3005	**34.3199**
	0.9389	0.9554	0.9531	0.9567	0.9556	0.9586	0.9605	0.9605	0.9611
ZY3_1	32.7192	33.4555	33.4910	33.4615	33.4932	33.8581	33.7873	33.8398	**33.8829**
	0.9175	0.9414	0.9420	0.9460	0.9429	0.9506	0.9504	**0.9514**	0.9507
ZY3_2	36.4017	37.1352	37.2095	37.6016	37.5260	38.0238	37.7460	37.8232	**38.0466**
	0.9482	0.9681	0.9674	**0.9780**	0.9773	0.9768	0.9751	0.9755	0.9769

provides a richer NL structure and can better reflect the superiority of the proposed algorithm. Next, we take the reconstruction result of image GF1_2 with the zooming factor 3 as an example to compare the reconstruction effect of various algorithms on different sensor images and the corresponding absolute error map. As shown in Fig. 7, the texture features of the cultivated land reconstructed by the proposed method are more obvious, there are fewer bright spots on the absolute error map, and the reconstruction effect is better. Besides, the ZY3_2 image has a larger proportion of cultivated land than the GF1_2 image. As shown in Tables II–IV, the reconstruction effect of the proposed methods in the image ZY3_2 is greater than that in the GF1_2 image. It has been proven once again that our method has a better reconstruction effect on images with more NL similarity structures.

C. SR Reconstruction of Satellite Images With Different Noise Types

Satellite sensors are often disturbed by noise during image acquisition and processing, so whether the algorithm has good noise immunity will directly affect the image reconstruction effect. The purpose of this experiment is to verify the antinoise performance of our method. First, we added salt and pepper noise with a noise level of 0.003 to the eight satellite images adopted in Experiment B. Fig. 8 shows the reconstruction results and the corresponding absolute error map of image WV2_1 with salt and pepper noise by our method and some other state-of-the-art methods under the zooming factor 3. From the absolute error map in Fig. 8, we can see that the noise spots on the roof of the buildings in the reconstruction image obtained by our method are the least. Table V shows that the proposed method is very robust under the influence of low-density noise and still shows the best reconstruction effect for most of the test images. Since the NL mean algorithm itself can effectively reduce noise, the proposed method is relatively effective for SR reconstruction of noisy images. Compared with other methods, the image reconstructed by our proposed method is clearer and can effectively reduce the influence of noise.

Besides, we add a small amount of Gaussian noise ($\sigma =$ 0, 2, 4, 6) to the eight satellite images and then perform SR reconstruction for these images with Gaussian noise under the zooming factors 3 and 4. Fig. 9 shows the comparison results of PSNR and FSIM between our method and some other state-of-the-art methods for GF1_1 at different levels of Gaussian noise. As shown in Fig. 9(a) and (b), various methods perform

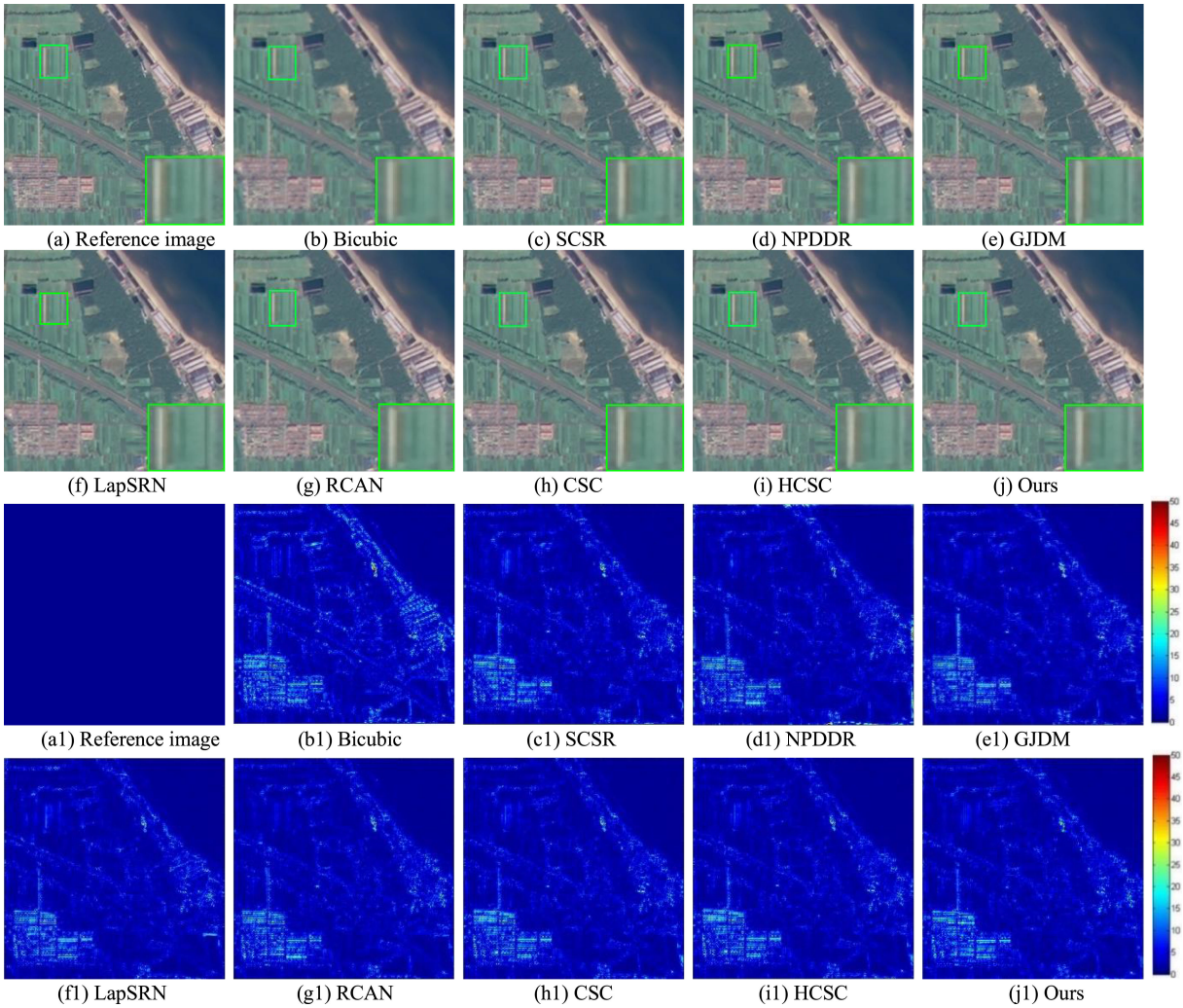

Fig. 7. SR reconstruction results and absolute error images with respect to GF1_2 under the zooming factor 3. (a) Ground truth HR image. (b)–(h) SR results from bicubic, SCSR [15], NPDDR [42], GJDM [17], LapSRN [43], RCAN [44], CSC [19], HCSC [21], and our method. (b1)–(h1) Absolute error images with respect to the ground truth from bicubic, SCSR [15], NPDDR [42], GJDM [17], LapSRN [43], RCAN [44], CSC [19], HCSC[21], and our method.

TABLE V
PSNR AND FSIM COMPARISON OF DIFFERENT RECONSTRUCTION METHODS ON IMAGES WITH SALT AND PEPPER NOISE UNDER THE ZOOMING FACTOR 3

	Bicubic	SCSR[15]	NPDDR[42]	GJDM[17]	LapSRN[43]	RCA[44]	CSC[19]	HCSC[21]	Ours
GF1_1	35.4581	36.4880	36.7587	36.8360	36.8648	36.9653	36.8778	36.9442	**37.0127**
	0.9633	0.9808	0.9796	0.9819	0.9832	**0.9839**	0.9832	0.9831	**0.9839**
GF1_2	37.0716	37.8975	38.1163	38.1574	38.1807	38.2695	38.1937	38.2606	**38.3818**
	0.9676	0.9840	0.9833	0.9809	0.9860	0.9860	0.9860	**0.9862**	**0.9862**
TM5_1	30.0204	30.3534	30.3899	30.4216	30.4766	**30.4967**	30.4414	30.4532	30.4580
	0.9417	0.9663	0.9649	0.9688	0.9707	**0.9710**	0.9687	0.9685	0.9688
TM5_2	30.4469	30.7761	30.7952	30.9010	30.9127	**31.0546**	30.8934	30.9111	30.8935
	0.9404	0.9660	0.9640	0.9695	0.9723	**0.9729**	0.9695	0.9695	0.9691
WV2_1	35.9678	36.5665	36.7773	36.8332	36.8644	36.9553	36.9587	36.9847	**36.9962**
	0.9719	0.9837	0.9821	0.9823	0.9780	0.9860	0.9860	**0.9861**	0.9860
WV2_2	34.6984	35.3646	35.5073	35.5959	35.7298	35.7596	35.7379	35.7696	**35.7774**
	0.9693	0.9819	0.9807	0.9822	0.9799	0.9840	0.9842	0.9842	**0.9844**
ZY3_1	34.7686	35.6806	35.5808	35.4546	35.8862	36.1206	35.8707	35.9288	**36.1831**
	0.9695	0.9863	0.9829	0.9855	0.9880	0.9886	0.9883	0.9886	**0.9891**
ZY3_2	38.0850	38.3862	38.4230	38.4321	38.4851	38.6718	38.3724	38.4179	**38.7562**
	0.9804	0.9908	0.9882	0.9900	0.9920	0.9920	0.9911	0.9911	**0.9921**

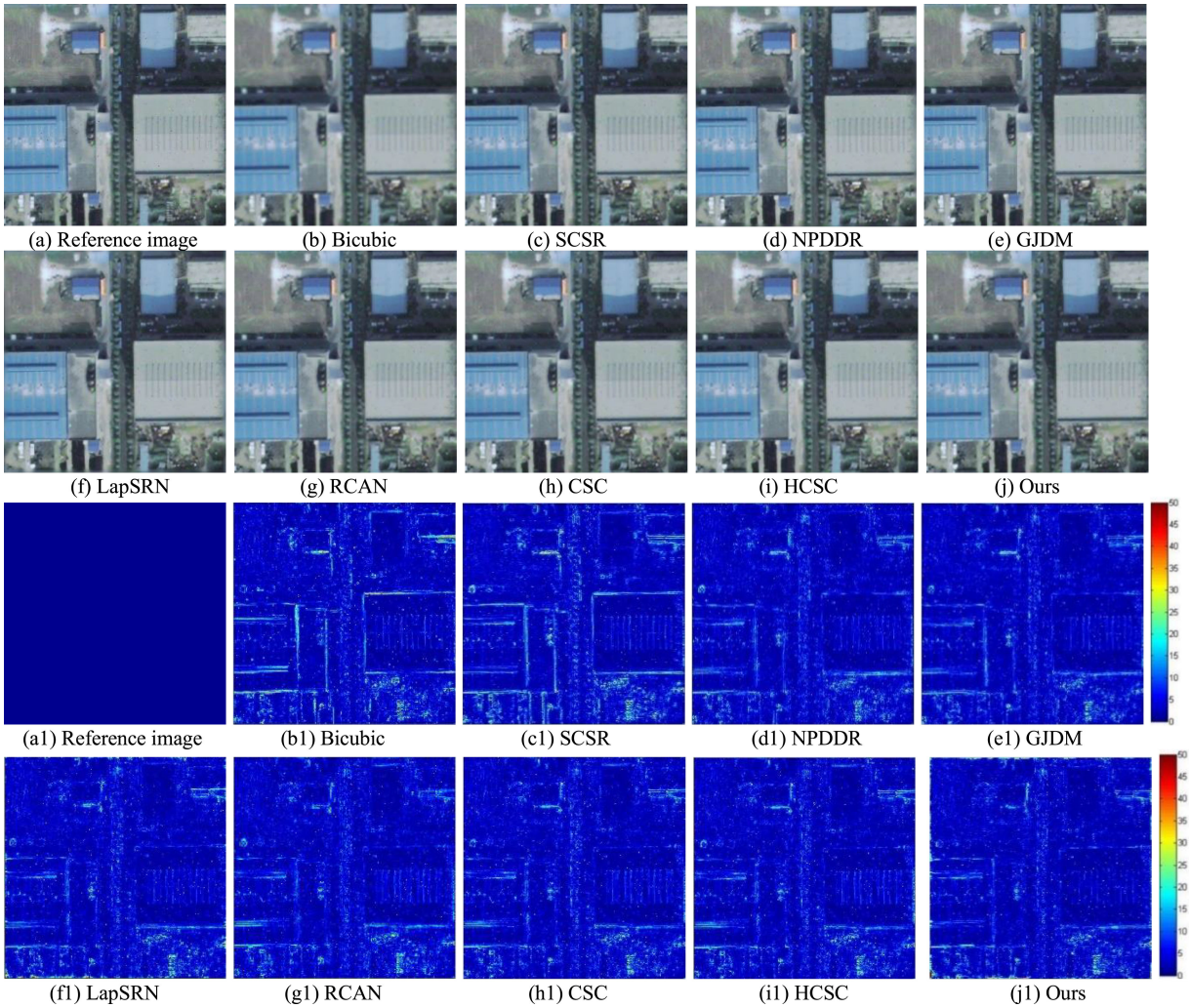

Fig. 8. SR results and absolute error images with respect to WV2_1 with salt and pepper noise under the zooming factor 3. (a) Ground truth HR image. (b)–(h) SR results from bicubic, SCSR [15], NPDDR [42], GJDM [17], LapSRN [43], RCAN [44], CSC [19], HCSC [21], and our method. (b1)–(h1) Absolute error images with respect to the ground truth from bicubic, SCSR [15], NPDDR [42], GJDM [17], LapSRN [43], RCAN [44], CSC [19], HCSC [21], and our method.

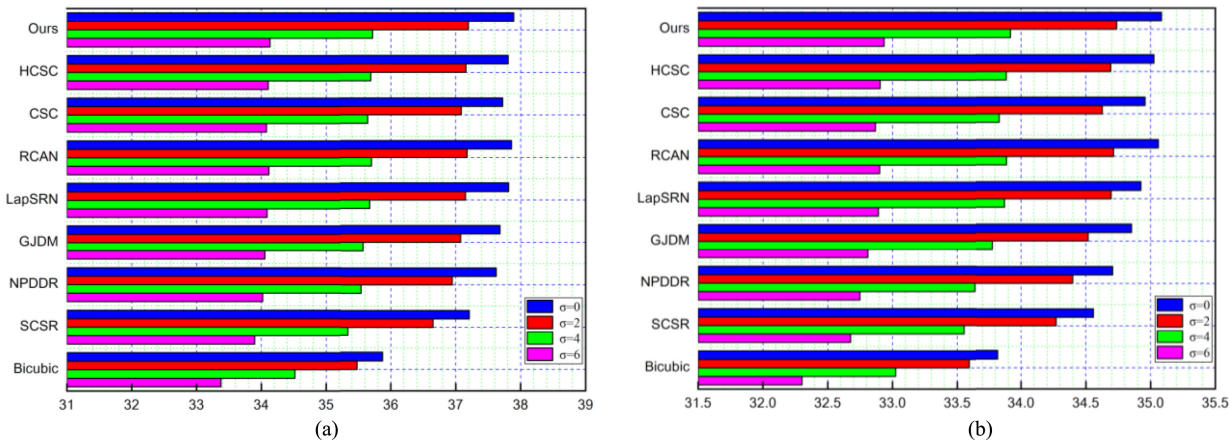

Fig. 9. PSNR comparison of our method with other methods for GF1-1 at different levels of Gaussian noise. (a) factor = 3. (b) factor = 4.

TABLE VI
IMPROVEMENT OF PSNR OF RECONSTRUCTION RESULTS OF IMAGE GF1_1_400 OBTAINED BY OUR METHOD (DB)

Image	Methods	Factor=2	Factor=3	Factor=4
Smooth part	CSC	0.0249↑	0.0361↑	0.0387↑
	HCSC	0.0143↑	0.0304↑	0.0299↑
Texture part	CSC	0.5035↑	0.1578↑	0.0822↑
	HCSC	0.3763↑	0.075↑	0.0141↑

SR reconstruction on GF1_1, and the obtained PSNR and FSIM values both continuously decrease as the noise level increases. Besides, as the noise level increases, the reconstruction effects of all methods are getting closer. Through the statistical analysis, we can see that the PSNR and FSIM values obtained by the proposed method are always the best than others. These results show that for images with different Gaussian noise levels, the proposed method can not only obtain high-quality reconstruction images but also maintain the spatial structure of the original image. It can be proved that our method has certain robustness to noise.

IV. DISCUSSION

A. Model Validity

The proposed HNLSCSC-SR method intends to apply CSC theory to single satellite image SR reconstruction, aiming to improve the performance of the CSC-SR method. The CSC-SR method uses bicubic interpolation to reconstruct the smooth part and uses the CSC method to reconstruct the texture part. The HCSC-SR method is also an improved method based on CSC-SR, which uses the convolutional neural network (SRCNN) to reconstruct the smooth part, and uses the adaptive CSC algorithm to reconstruct the texture part. To further reflect the superiority and effectiveness of our method, we take the GF1_1_400 image as an example to test the contribution of the three methods to the reconstruction of the smooth and texture parts of the image. Table VI shows the improvement of the quantitative evaluation index PSNR of the proposed HNLSCSC method compared with the reconstruction results of the CSC and HCSC methods under the zooming factors 2, 3, and 4.

After a comparative analysis of ablation studies, it is found that our method has improved the reconstruction results of the smooth part and the texture part of the image. Compared with bicubic interpolation and SRCNN, the Bayesian nonparametric method can be used to train a structured dictionary, which is more suitable for the smoothing part, which is conducive to restore the residual high-frequency information in the smooth part. Besides, we introduce the NL similarity prior information of the feature map into the CSC framework to reconstruct the texture part, thereby improving the reconstruction performance of the CSC method.

B. Number of Similar Patches

The proposed method focuses on the impact of nonlocally similar structures on the SR reconstruction results. The selection

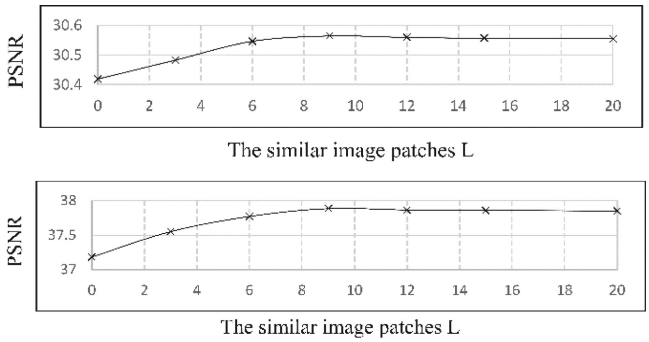

Fig. 10. Relationship between PSNR and the number of similar image blocks (upper: TM5_1; lower: GF1_1).

TABLE VII
IMPROVEMENT OF PSNR OF RECONSTRUCTION RESULTS OF IMAGE GF1_1_400 OBTAINED BY OUR METHOD (DB)

Patch size	Dictionary size	ΔPSNR
6 × 6	510	0.0298↑
7 × 7	551	0.0310↑
8 × 8	619	0.0316↑
9 × 9	623	0.0314↑
10 × 10	631	0.0309↑

of the number of similar patches L is an important factor. First, the image patches set that participates in weight calculation is extracted by using structure information. Then, the Euclidean distance between image patches in the image set is calculated and arranged in descending order. The image patches corresponding to the largest L distances are extracted as the most similar image patches. There are experiments performed on images GF1_1 and TM5_1 to analyze the relationship between the number of similar patches L and the quality of the reconstructed image. As shown in Fig. 10, when $L \leq 9$, the PSNR increases with the increase in L, indicating that the reconstruction quality improves with the increase in the number of similar patches. For $L \geq 9$, the PSNR decreases slowly with the increase in L. When L is larger, the set of similar patches will contain some image patches that are not similar to the target patches to participate in the estimation, so that the quality of reconstruction results will be worse. In this article, to obtain the optimal reconstruction results in all experiments, we adopt $L = 9$ as the number of similar patches.

C. Patches Size

Intuitively, a patch size that is too large or too small tends to produce slow or smooth training. To find the optimal patch size, we trained the corresponding dictionary to reconstruct the smooth part of the image GF1_1_400 by setting different patch sizes in the Bayesian nonparametric method and compared with the reconstruction results of the CSC under the zooming factor 3. The impact of the patch size is evaluated and listed in Table VII.

A fixed-size dictionary will result in poor matching accuracy and long calculation time. The nonparametric derivation of the dictionary is another advantage of the Bayesian nonparametric

TABLE VIII
FIDELITY ANALYSIS OF RECONSTRUCTION RESULTS OF IMAGE GF1_1_400 OBTAINED BY THE NINE ALGORITHMS

Measures	IFC	VIF
Bicubic	3.722	0.4878
SCSR[15]	4.848	0.5907
NPDDR[42]	4.956	0.6212
GJDM[17]	4.504	0.6156
LapSRN[43]	5.033	0.6307
RCAN[44]	5.213	0.6348
CSC[19]	5.181	0.6249
HCSC[21]	5.205	0.6279
Ours	5.216	0.6363

method. In Table VII, we give the initialization dictionary size of 1024, and the dictionary size obtained by nonparametric derivation is different, which is the optimal dictionary size for the corresponding experiment. Since there is less high-frequency information remaining in the smooth part, the difference in the results of the smooth image reconstruction using the dictionary trained with different patches sizes is not very obvious. The results in Table VII imply that setting the patch size to 8×8 has the best reconstruction effect.

D. Fidelity Analysis

Information fidelity criteria (IFC) [47] and visual information fidelity criteria (VIF) [48] are considered to be better for evaluating SR images than the widely used PSNR and SSIM index [49]. They measure the quality of the image to be evaluated by calculating the mutual information between the image to be evaluated and the reference image. For the SR task of GF1_1_400 under the zoom factor 3, we calculated the image quality scores of the nine algorithms used in this article. In Table VIII, we show the fidelity analysis results of the nine algorithms in this article for GF1_1_400 reconstructed images.

The minimum IFC value is 0. The higher the IFC value, the better the visual quality of the image based on contrast and mutual information. The VIF index value ranges from 0 to 1. The higher the value is, the better the image reconstruction result will be, which is 1 in the ideal state. It can be inferred from Table VIII that our proposed method also shows better results in fidelity and visual perception than other methods.

E. Limitation

Since dictionary learning and reconstruction are needed for the smooth part of the image, and CSC reconstruction needs to be performed after NL similarity constraint optimization on each feature map in the texture part, the method in this article takes longer inference time than in CSC and HCSC. The average SR reconstruction times of CSC, HCSC, and HNLSCSC under the zooming factor 2 are 64.39, 580.46, and 749.88 s, respectively. However, the method proposed in this article has not been optimized, so the calculation efficiency is relatively low. Our model is relatively independent of the smooth part and texture part reconstruction process, and the NL similarity optimization constraint of each feature map during the texture part reconstruction is also independent. The parallelization techniques are a potential solution to accelerate the inference time of our model. In the future, we will study parallel algorithms and try to extend HNLSCSC to other applications, such as scene recognition and semantic image segmentation and other high-level tasks.

It is worth noting that through the comparison of all experimental results, it is found that the HNLSCSC method proposed in this article is only slightly better than the current advanced deep learning SR reconstruction network RCAN in the reconstruction result of most test images. Because the spatial resolution of TM images is low and the nonlocal similar structures are not obvious, the RCAN method performs better in the reconstruction of TM images. However, the HNLSCSC method is developed based on CSC theory. Compared with the deep learning framework, the principle of the algorithm proposed in this article is easier for readers to understand, avoiding a large number of parameter settings, and it is easier to introduce more prior information of the original image to further improve the accuracy of the algorithm. Besides, our next research work is to introduce our method into the python environment for further updates and optimizations, and then compare it with more recent deep network SR methods.

V. CONCLUSION

In this article, we proposed a hybrid NL similarity constrained convolution sparse coding method for SISR reconstruction (HNLSCSC-SR). We used the Bayesian nonparametric method and the NL similarity CSC method to reconstruct the smooth part and the texture part, respectively. In the SR reconstruction of the texture part, we classified the search space based on structural information and found the most relevant image patches set to reduce unnecessary weight calculation. Our study also indicated that the feature map corresponding to each filter has the same structure information corresponding to the global features. Also, different feature maps share the same index position information of the similar patches. Then, the iterative threshold optimization is required to minimize the CSCN of the feature maps, thereby obtaining the optimal feature maps. The results obtained from three SR reconstruction experiments on satellite images demonstrated that our method can solve the block effect of the image and has a relatively better performance on PSNR and FSIM, when compared with the state of the art. Also, our method was able to make the texture and edge structure of the reconstructed image clearer and showed good noise immunity. Besides, the proposed method is more effective for image reconstruction with higher resolution and richer NL similar structure.

ACKNOWLEDGMENT

The authors would like to thank Prof. M. Zhou for sharing the codes of [18], Prof. S. Gu for sharing the codes of [19], Prof. J. Yang for sharing the codes of [15], Prof. W. Dong for sharing the codes of [28], Prof. J. Zhao for sharing the codes of [21], Prof. S. Gou for sharing the codes of [42], Prof. C. Dong for sharing the codes of [43], Prof. B. Hou for sharing the codes of [17], and Prof. Y. Zhang for sharing the codes of [44]. They would also like to thank the anonymous Reviewers whose comments and suggestions helped improve and clarify this manuscript.

REFERENCES

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

[2] Z. Pan, W. Ma, J. Guo, and B. Lei, "Super-resolution of single remote sensing image based on residual dense backprojection networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7918–7933, Oct. 2019.

[3] Z. Shao, L. Wang, Z. Wang, and J. Deng, "Remote sensing image super-resolution using sparse representation and coupled sparse autoencoder," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 2663–2674, Aug. 2019.

[4] C. Kim, K. Choi, and J.B. Ra, "Example-based super-resolution via structure analysis of patches," *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 407–410, 2013.

[5] M. Bevilacqua, A. Roumy, C. Guillemot, and M.A. Morel, "Single-image super-resolution via linear mapping of interpolated self-examples," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5334–5347, Dec. 2014.

[6] J. Yu, X. Gao, D. Tao, X. Li, and K. Zhang, "A unified learning framework for single image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 4, pp. 780–792, Apr. 2014.

[7] Z. Zhu, F. Guo, H. Yu, and C. Chen, "Fast single image super-resolution via self-example learning and sparse representation," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2178–2190, Dec. 2014.

[8] D. Liu, Z. Wang, B. Wen, J. Yang, W. Han, and T. S. Huang, "Robust single image super-resolution via deep networks with sparse prior," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3194–3207, Jul. 2016.

[9] F.-J. Lin and J.-H. Chuang, "Image super-resolution by estimating the enhancement weight of self example and external missing patches," *Multimed. Tools Appl.*, vol. 77, no. 15, pp. 19071–19087, Aug. 2018.

[10] P. Liu, Y. Hong, and Y. Liu, "Deep differential convolutional network for single image super-resolution," *IEEE Access*, vol. 7, pp. 37555–37564, 2019.

[11] Z. Jiang, Y. Huang, and L. Hu, "Single image super-resolution: Depthwise separable convolution super-resolution generative adversarial network," *Appl. Sci.*, vol. 10, Jan. 2020, Art. no. 375.

[12] P. Song, X. Deng, J.F.C. Mota, N. Deligiannis, P.L. Dragotti, and M. R. D. Rodrigues, "Multimodal image super-resolution via joint sparse representations induced by coupled dictionaries," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 57–72, 2020.

[13] J. Zhao, T. Sun, and F. Cao, "Image super-resolution via adaptive sparse representation and self-learning," *IET Comput. Vis.*, vol. 12, no. 5, pp. 753–761, 2018.

[14] Y.-Q. Yang, "Research on the single image super-resolution method based on sparse Bayesian estimation," *Cluster Comput.*, vol. 22, pp. 1505–1513, Jan. 2019.

[15] J. Yang, J. Wright, T.S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[16] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 791–804, Apr. 2012.

[17] B. Hou, K. Zhou, and L. Jiao, "Adaptive super-resolution for remote sensing images based on sparse representation with global joint dictionary model," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2321–2327, Apr. 2018.

[18] G. Polatkan, M. Zhou, L. Carin, D. Blei, and I. Daubechies, "A Bayesian nonparametric approach to image super-Resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 2, pp. 346–358, Feb. 2015.

[19] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, "Convolutional sparse coding for image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1823–1831.

[20] L. Li, S. Zhang, F. Liu, S. Yang, and X. Tang, "Semi-coupled convolutional sparse learning for image super-resolution," *Remote Sens*., vol. 11, Nov. 2019, Art. no. 2593.

[21] J. Zhao, C. Chen, Z. Zhou, and F. Cao, "Single image super-resolution based on adaptive convolutional sparse coding and convolutional neural networks," *J. Vis. Commun. Image Represent.*, vol. 58, pp. 651–661, Jan. 2019.

[22] L. Yang *et al.*, "Image reconstruction via manifold constrained convolutional sparse coding for image sets," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 7, pp. 1072–1081, Oct. 2017.

[23] K. Zhang, M. Wang, S. Yang, and L. Jiao, "Convolution structure sparse coding for fusion of panchromatic and multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1117–1130, Feb. 2019.

[24] G. Wu, X. Zhu, Q. Wang, and D. Shen, "Image super-resolution by supervised adaption of patchwise self-similarity from high-resolution image," in *Patch-Based Techniques in Medical Imaging*, vol. 9467. Cham, Switzerland: Springer, pp. 10–18, 2015.

[25] Y. Chen *et al.*, "Single-image super-resolution algorithm based on structural self-similarity and deformation block features," *IEEE Access*, vol. 7, pp. 58791–58801, 2019.

[26] C. Ren, X. He, Y. Pu, and T. Q. Nguyen, "Enhanced non-local total variation model and multi-directional feature prediction prior for single image super resolution," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 3778–3793, Aug. 2019.

[27] W. Dong, L. Zhang, and G. Shi, "Centralized sparse representation for image restoration," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 1259–1266.

[28] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.

[29] M. Zhou, H. Chen, J. Paisley, L. Ren, G. Sapiro, and L. Carin, "Non-parametric Bayesian dictionary learning for sparse image representations," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 2295–2303.

[30] J. Sodjo, A. Giremus, N. Dobigeon, and F. Caron, "A Bayesian non-parametric model for unsupervised joint segmentation of a collection of images," *IEEE Access*, vol. 7, pp. 120176–120188, 2019.

[31] L. Sui, L. Li, J. Li, N. Chen, and Y. Jiao, "Fusion of hyperspectral and multispectral images based on a Bayesian nonparametric approach," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1205–1218, Apr. 2019.

[32] C. Robert and G. Casella, "Monte Carlo statistical method," *Technometrics*, vol. 42, pp. 430–431, Nov. 2000.

[33] F. A. Quintana and P. Müller, "Nonparametric Bayesian data analysis," *Statist. Sci.*, vol. 19, no. 1, pp. 95–110, 2004.

[34] A. Rodríguez and D. B. Dunson, "Nonparametric Bayesian models through probit stick-breaking processes," *Bayesian Anal.*, vol. 6, no. 1, pp. 145–178, 2011.

[35] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2015, pp. 3631–3640.

[36] Z. Wenliang and K. James, "Fast stochastic alternating direction method of multipliers," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 46–54.

[37] B. Wohlberg, "Efficient convolutional sparse coding," in *Proc. ICASSP*, 2014, pp. 7173–7177.

[38] Z. Pan *et al.*, "Super-resolution based on compressive sensing and structural self-similarity for remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4864–4876, Sep. 2013.

[39] A. Buades, B. Coll, and J.M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, Jan. 2005.

[40] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, Jan. 2008.

[41] Y.-W. Tai, S. Liu, M. Brown, and S. Lin, "Super resolution using edge prior and single Image detail synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2010, pp. 2400–2407.

[42] S. Gou, S. Liu, S. Yang, and L. Jiao, "Remote sensing image super-resolution reconstruction based on nonlocal pairwise dictionaries and double regularization," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4784–4792, Dec. 2014.

[43] W. Lai, J. Huang, N. Ahuja, and M. Yang, "Deep Laplacian pyramid networks for fast and accurate super-Resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2017, pp. 5835–5843.

[44] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Pro. Eur. Conf. Comput. Vis.*, 2018, pp. 294–310.

[45] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[46] A. Vedaldi, K. Lenc, and A. Gupta, "MatConvNet—Convolutional neural networks for MATLAB," in *Proc. 23rd ACM Int. Conf. Multimed.*, 2014, pp. 689–692.

[47] H. R. Sheikh, A. C. Bovik, and G. D. Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[48] H.R. Sheikh and A.C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[49] C.-Y. Yang, C. Ma, and M.-H. Yang, "Single-image super-resolution: A benchmark," in *Pro. Eur. Conf. Comput. Vis.*, 2014, pp. 372–386.

Nan Chen (Student Member, IEEE) received the B.Eng. degree in surveying and mapping engineering and the M.Sc. degree in photogrammetry and remote sensing from the Chengdu University of Technology, Chengdu, China, in 2009 and 2012, respectively. She is currently working toward the Ph.D. degree in photogrammetry and remote sensing with Chang'an University, Xi'an, China.

Her current research interests include image superresolution and image fusion.

Lichun Sui received the B.Eng. and M.Sc. degrees in photogrammetry from Wuhan University, Wuhan, China, in 1983 and 1989, respectively, and the Dr.-Ing. degree in photogrammetry and remote sensing from the Technische Universität Berlin, Berlin, Germany, in 1998.

He is currently a Full Professor with the College of Geological Engineering and Geomatics, Chang'an University, Xi'an, China. He has authored more than 100 papers in refereed journals and conferences, including the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. His main research interests comprise LiDAR data processing, photogrammetry and remote sensing applications, remote sensing information extraction and application, image analysis, InSAR applications, and 3-D laser scanning for geohazard monitoring.

Biao Zhang received the B.Eng. degree in geographic information science and the Ph.D. degree in cartography and geographic information engineering from Chang'an University, Xi'an, China, in 2017 and 2020, respectively.

He is currently an Engineer with the Geovis Spatial Technology Co.,Ltd, Xi'an, China. His current research interests include spatial data modeling, geospatial data processing, and analysis.

Hongjie He received the B.Eng. degree in surveying and mapping engineering from China University of Petroleum, Beijing, China, in 2016, and the M.Sc. degree in cartography and geographic information system from Lanzhou University, Lanzhou, China, in 2019. He is currently working toward the Ph.D. degree in geomatics engineering with the Geospatial Sensing and Data Intelligence Lab, Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON, Canada.

His current research interests include deep learning, semantic segmentation, object detection, image classification, image superresolution, and instance segmentation in remote sensing.

José Marcato Junior (Member, IEEE) received the Ph.D. degree in cartographic science from Sao Paulo State University, Sao Paulo, Brazil, in 2014.

He is currently a Professor with the Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande, Brazil. He has authored or coauthored more than 30 papers in refereed journals and more than 70 papers in conferences, including papers published in *ISPRS Journal of Photogrammetry and Remote Sensing*, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS, and *Remote Sensing*, *Sensors*, and *The Photogrammetric Record*. His current research interests include UAV photogrammetry and deep neural networks for object detection, classification, and segmentation.

Jonathan Li (Senior Member, IEEE) received the Ph.D. degree in geomatics engineering from the University of Cape Town, Cape Town, South Africa, in 2000.

He is currently a Professor with both the Department of Geography and Environmental Management and the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada. He has coauthored more than 400 publications, including more than 240 refereed journal papers. His main research interests include image and point cloud analytics, mobile mapping, and AI-empowered information extraction from LiDAR point clouds and earth observation images.

Dr. Li is currently an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, and CANADIAN JOURNAL OF REMOTE SENSING. He was the recipient of the Outstanding Achievement in Mobile Mapping Technology Award, in 2019 for his pioneering contributions in developing and promoting mobile mapping technology and the ISPRS Samuel Gamble Award, in 2020 for his significant contributions to the development, organization or professional activities of the photogrammetry, remote sensing, and spatial information sciences at national and international levels.