Thick Cloud Removal in Multitemporal Remote Sensing Images via Low-Rank Regularized Self-Supervised Network

Yong Chen, Maolin Chen, Wei He, *Senior Member, IEEE*, Jinshan Zeng, *Member, IEEE*, Min Huang, and Yu-Bang Zheng, *Member, IEEE*

Abstract— The existence of thick clouds covers the comprehensive Earth observation of optical remote sensing images (RSIs). Cloud removal is an effective and economical preprocessing step to improve the subsequent applications of RSIs. Deep learning (DL)-based methods have attracted much attention and achieved state-of-the-art results. However, most of these methods suffer from the following issues: 1) ignore the physical characteristics of RSIs; 2) require paired images with/without cloud or extra auxiliary images; and 3) demand the cloud mask. These issues might have limited the flexibility of existing networks. In this article, we propose a novel low-rank regularized self-supervised network (LRRSSN) that couples model-driven and data-driven methods to remove the thick cloud from multitemporal RSIs (MRSIs). First, motivated by the equal importance of image and cloud components as well as their intrinsic characteristics, we decompose the observed image into low-rank image and structural sparse cloud components. In this way, we obtain a model-driven thick cloud removal method where the spectral–temporal low-rank correlation of the image component and the spectral structural sparsity of the cloud component are effectively exploited. Second, to capture the complex nonlinear features of different scenarios, the data-driven self-supervised network that does not require external training datasets is designed to explore the deep prior of the image component. Third, the coupled model-driven and data-driven LRRSSN is optimized by an efficient half quadratic splitting (HQS) algorithm. Finally, without knowing the exact cloud mask, we estimate the cloud mask to preserve information in cloud-free areas as much as possible. Experiments conducted in synthetic and real-world scenarios demonstrate the effectiveness of the proposed approach.

Index Terms— Low rank, multitemporal remote images, self-supervised network, thick cloud removal.

I. INTRODUCTION

WITH the development of remote sensing technology and hardware equipment, optical remote sensing images (RSIs) with higher spatial, spectral, and temporal resolution are now available. Benefit from abundant spatial, spectral, and temporal information, RSIs have been widely employed in Earth observation applications [1], such as land cover classification [2], object detection [3], and environmental monitoring [4]. However, due to the influence of imaging devices and external environments, the collected RSIs are inevitably degraded by thick clouds [5], which seriously affects the downstream application. Analysis indicates that clouds cover 35% of the Earth's surface annually [6], making it challenging to obtain high-quality RSIs, especially for multitemporal RSIs (MRSIs) [7]. Consequently, cloud removal becomes an urgent problem in optical remote sensing imaging.

Numerous approaches for thick cloud removal have been put forward, and they can be broadly classified into two categories: traditional methods and deep learning (DL) methods [8]. Traditional methods achieve thick cloud removal using a multistep process or model-driven optimization model, which contains spatial approaches, spectral approaches, and temporal approaches depending on how each method uses auxiliary information type. Spatial approaches are mainly designed for thick cloud removal in a single RSI, and they transform the task of thick cloud removal into image inpainting [9], [10], [11], [12]. They reconstruct information in the cloud-covered area by leveraging spatial details from adjacent, cloud-free patches within the same image. Maalouf et al. [9] reconstructed the cloud region by propagating the geometrical flow curves inside that zone. While effective for removing small cloud regions with simple image details, these methods may struggle with larger cloud regions with complex details.

Spectral approaches reconstruct the thick cloud bands by introducing extra auxiliary cloud-free bands [13], [14], [15], [16], [17]. These auxiliary bands can be collected from

different temporal acquisitions or different sensors. As the strong correlation of Moderate Resolution Imaging Spectroradiometer (MODIS) bands 6 and 7, Rakwatin et al. [13] and Shen et al. [14] employed the auxiliary clean band 7 to reconstruct the missing information of band 6. Additionally, Xu et al. [16] developed a cloud removal method by using signal transmission and spectral mixture analysis. Expanding on this, Xu et al. [17] further proposed a noise-adjusted principal component transform model for cloud removal. Overall, spectral methods can reconstruct large cloud regions to some extent. Nonetheless, these methods become inapplicable when the spectral bands exhibit a lack of correlation due to the absence of auxiliary information.

Since optical satellites can perform periodic imaging in the same scene, temporal approaches reconstruct the thick cloud regions by introducing correspondence images at different times (i.e., MRSIs) [18], [19], [20], [21]. Lin et al. [21] proposed an information cloning-based multistage method, which reconstructs the information of cloud regions utilizing temporal correlation of MRSIs. Given the high correlation of MRSIs in the spectral–temporal dimension, model-driven techniques regard thick cloud regions as missing information and design low-rank completion models to reconstruct the original information [22]. Ji et al. [23] reconstructed the missing information of cloud regions in MRSIs using a low-rank tensor completion model. He et al. [24] proposed a tensor ring completion model utilizing the low-rank property of MRSIs of different dimensions. However, the performance of these completion methods is highly dependent on the accuracy of the cloud mask. In MRSIs, model-driven blind thick cloud removal techniques have been progressively developed to mitigate the impact of cloud masks. These methods regard the task of thick cloud removal as a denoising inverse problem and explore the statistical prior of the image and cloud components [25], [26], [27]. By decomposing the observed image into the clean image and cloud components, Chen et al. [25] first proposed a blind cloud removal method by decomposing the cloud-contaminated image into a low-rank clean image component and sparse cloud component and detecting the cloud mask from the estimated cloud component. Following the perspective of image decomposition [28], [29], [30], [31], Lin et al. [29] suggested a coupled tensor factorization to explore the relationship between the representation coefficients of the MRSIs in the same scene. Zheng et al. [30] regarded the thick cloud-contaminated image as the combination of low-rank tensor component and group sparse component. While the model-driven approach attempts to tackle the critical issue of cloud masks in MRSIs, it relies heavily on the effectiveness of handcrafted priors. However, these handcrafted priors may not be suitable for every data with different resolutions and different scenarios.

Recently, the powerful nonlinear fitting capabilities of neural networks have made DL a successful tool in image processing [32]. As a result, DL-based methods are widely used in thick cloud removal, learning a nonlinear end-to-end mapping between the thick cloud-contaminated image and the paired cloud-free image [33], [34], [35], [36]. The convolutional neural network (CNN) is usually designed to reconstruct the thick cloud-contaminated regions [37], [38], [39], [40], [41], [42], [43]. Zhang et al. [37] designed a progressive spatiotemporal patch group learning framework for thick cloud removal in MRSIs. Ji et al. [38] proposed a self-trained multiscale full convolutional network for cloud removal in MRSIs. Ebel et al. [43] designed a multitemporal 3-D CNN that predicts a cloud-free image from a sequence of synthetic aperture radar (SAR) images. To improve the performance of CNN, the generative adversarial networks (GANs) are designed for thick cloud removal [8], [44], [45], [46], [47], [48]. Xu et al. [8] introduced the attention mechanism-based GAN for cloud removal in Landsat images. Since DL-based methods effectively learn the multiscale features of clouds, they can achieve better results than traditional methods with sufficient paired training data. However, existing DL methods focus on thick cloud removal for single or multispectral images (RGB images) and ignore the multitemporal intrinsic characteristics of thick cloud-contaminated images. Moreover, cloud masks need to be given accurately so that cloud detection is a key preprocessing step and greatly affects the results. Furthermore, these networks are often designed for specific data and cloud conditions, which means that they may not apply to test data that differ from the training data in resolution, number of spectral bands, or number of temporal nodes.

The combination of data-driven and model-driven methods can effectively utilize the statistical priors from the data and the potential deep image priors [49], [50], [51], [52], [53]. Zhao et al. [50] introduced a novel tensor completion framework that simultaneously utilizes complementary global, local, and nonlocal priors. Luo et al. [51] propose a nonlinear multilayer neural network to learn a nonlinear transform by solely using the observed tensor in a self-supervised manner. Wang et al. [52] presented an unsupervised domain factorization network for thick cloud removal, enabling effective and efficient utilization of the rich spatial–spectral–temporal information of MRSIs. However, there has been relatively little research on combining data-driven and model-driven approaches specifically for cloud removal tasks, and the potential of such hybrid methods has not been fully explored.

Considering the advantages and obstacles of the traditional and DL methods, in this article, we propose a novel low-rank regularized self-supervised network (LRRSSN) for thick cloud removal in MRSIs, which has the ability to address the limitations of both types of approaches and inherit their strengths. First, inspired by the traditional blind thick cloud removal framework, we give a detailed analysis of both the intrinsic physical characteristics of the image and cloud components. We regard the thick cloud removal problem as an image decomposition issue. In this context, the model-driven low-rank prior is introduced to capture the global spectral–temporal correlation of the image component, while the cloud component is leveraged by model-driven structural sparsity. Within this framework, the thick cloud-contaminated image is decomposed into the image and cloud components. These two components are treated equally and completely decoupled. Second, due to the complex features of MRSIs, model-driven priors may not be applicable to every data. Aiming at utilizing the powerful nonlinear feature extraction
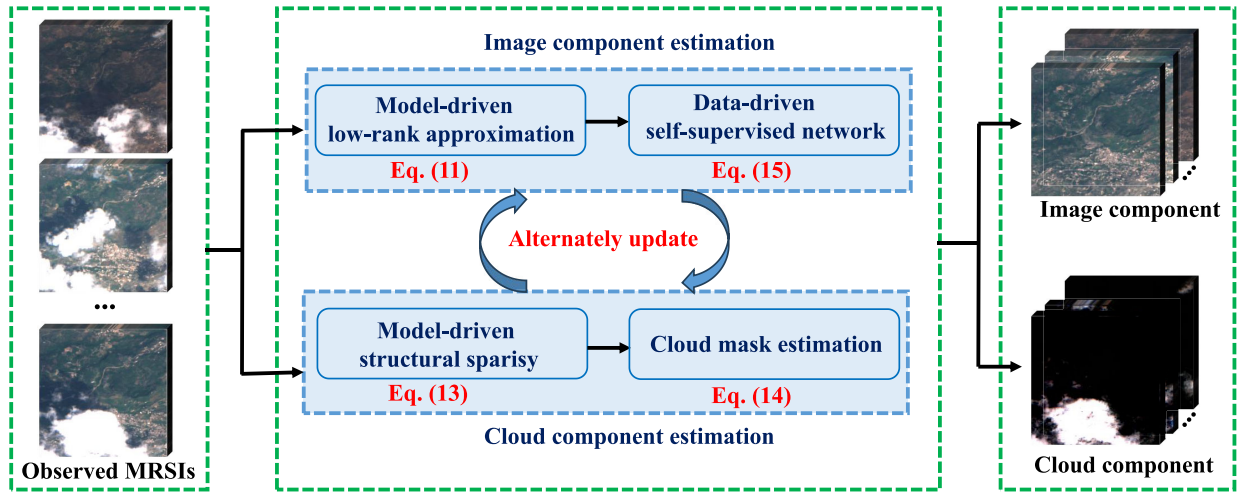
Fig. 1. Illustration of our proposed LRRSSN for thick cloud removal in MRSIs.

capability of DL-based methods, we integrate the deep prior captured by a self-supervised network into our model-driven optimization model to improve the data representation abilities. Third, using the HQS framework, an efficient algorithm has been designed to solve the model-driven and data-driven image decomposition model. In this decomposition framework, the image component and cloud component can be solved alternatively. This iterative estimation strategy synergistically enhances both components. Finally, we design a simple threshold method to estimate the cloud mask based on the cloud component extracted during the iterative process, and the pixels of the cloud-free regions are projected onto the image component to avoid information loss in the final reconstructed image. The illustration of the proposed LRRSSN for thick cloud removal in MRSIs is presented in Fig. 1. Overall, the main contributions of this article are summarized as follows.

1) We propose a novel method LRRSSN for thick cloud removal in MRSIs via coupling the superiorities of model-driven and data-driven methods, which can simultaneously reconstruct the cloud coverage information and estimate the cloud mask. Compared with traditional methods, the incorporation of data-driven deep prior improves the ability of image reconstruction. Meanwhile, model-driven priors explore the intrinsic physical characteristics of image and cloud components, which can improve the data representation capability based on DL methods.

2) We jointly employ the model-driven low-rank prior and data-driven deep prior to explore the intrinsic global spectral–temporal correlation and complex spatial–spectral features of the image component. In particular, the data-driven deep prior is characterized by a self-supervised network, which does not require any training data and external auxiliary data.

3) We design an effective HQS algorithm to iteratively optimize the model-driven regularization and data-driven self-supervised network, which can benefit each other. The comprehensive experiments conducted on both

simulated and real datasets verify the superiority of our LRRSSN over traditional and DL-based methods.

The remainder of this article is organized as follows. Section II introduces the problem formulation and the proposed method. Section III reports the experimental results on both simulated and real data, ablation study, and model discussion. Finally, Section IV concludes this article.

II. METHODOLOGY

A. Problem Formulation

Let $\mathcal{Y} \in \mathbb{R}^{m \times n \times b \times t}$ be the thick cloud-contaminated MRSIs, where m and n are the spatial resolutions, and b and t are the number of spectral bands and time nodes, respectively. Assuming that the observed MRSIs $\mathcal{Y}$ can be decomposed into the clean image component and the cloud component, then the degradation model can be mathematically represented as follows [25], [29]:

$$\mathcal{Y} = \mathcal{M} \odot \mathcal{X} + \mathcal{C} \tag{1}$$

where $\mathcal{X} \in \mathbb{R}^{m \times n \times b \times t}$ is the clean image component, $\mathcal{C} \in \mathbb{R}^{m \times n \times b \times t}$ is the cloud component, $\mathcal{M} \in \mathbb{R}^{m \times n \times b \times t}$ is a binary mask with the zeros representing the thick cloud-contaminated regions and ones representing the cloud-free regions, and $\odot$ denotes the Hadamard product.

The problem of estimating the clean image $\mathcal{X}$ from the thick cloud-contaminated image $\mathcal{Y}$ is an ill-posed issue. Therefore, suitable priors are needed to regularize them for stable solutions. According to the maximum a posteriori (MAP) probability, the thick cloud removal problem can be transferred to minimize the following optimization function:

$$\min_{\mathcal{X}, \mathcal{C}, \mathcal{M}} \frac{1}{2} \|\mathcal{Y} - \mathcal{M} \odot \mathcal{X} - \mathcal{C}\|_F^2 + \lambda_1 R(\mathcal{X}) + \lambda_2 R(\mathcal{C}) \tag{2}$$

where the first term represents the data fidelity term, λ_1 and λ_2 are the regularization parameters, and $R(\mathcal{X})$ and $R(\mathcal{C})$ are the regularizers to describe the intrinsic physical characteristics of $\mathcal{X}$ and $\mathcal{C}$, respectively.
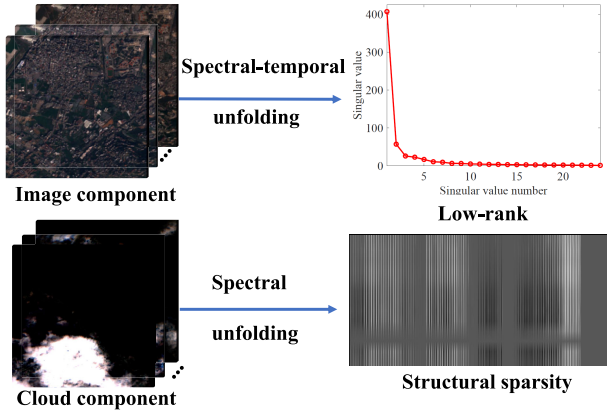
Fig. 2. Low-rank property and structural sparse property of the image component and cloud component, respectively.

B. Model-Driven Image Decomposition Framework

We present a detailed analysis on both the intrinsic physical characteristics of the image and the cloud components and then design the corresponding regularizers.

1) Low-Rank Characteristic of Image Component: Since the MRSIs are collected from the same region at different time nodes, we assume that ground object features within this region exhibit minimal variation over a short period. Therefore, there is a noticeable high correlation along the spectral–temporal dimension. To demonstrate this characteristic, we quantitatively analyze the statistical distribution of the image component in MRSIs. Fig. 2 illustrates the distribution of the singular values of the MRSIs in the spectral–temporal dimension, i.e., reshape the $\mathcal{X} \in \mathbb{R}^{m \times n \times b \times t}$ as $\mathbf{X} \in \mathbb{R}^{mn \times bt}$. We can clearly observe a declining trend in the singular values of the image component, indicating the strong correlation of the image component. To effectively capture this correlation, we introduce a model-driven low-rank prior to regularize the image component. Thus, the regularizer $R(\mathcal{X})$ is designed as follows:

$$R(\mathcal{X}) = \|\mathcal{X}\|_* = \sum_i \sigma_i(\mathcal{R}(\mathcal{X})) \tag{3}$$

where $\mathcal{R}$ is a reshape operator, i.e., $\mathcal{R}: \mathcal{X} \in \mathbb{R}^{m \times n \times b \times t} \to \mathbf{X} \in \mathbb{R}^{mn \times bt}$, and $\sigma_i(\mathbf{X})$ is the ith singular value of $\mathbf{X}$.

2) Structural Sparse Characteristic of Cloud Component: In general, the positions of the thick cloud in the imaging region change over time. We can assume that a certain location is obscured by thick cloud at some time nodes, while it remains cloud-free at other time nodes. Thick cloud obscures only a fraction of the regions in the MRSIs, leaving the majority cloud-free, implying that most of the pixel values in the cloud component are zero or close to zero, with only a few nonzero pixels representing the thick cloud. Naturally, the sparse prior can be introduced to describe the characteristics of the cloud component. However, sparsity only considers the random distribution of sparse elements and neglects the intrinsic structure between spectral bands. Assuming that the thick cloud contaminates all bands in the MRSIs, the cloud component possesses a distinctive sparse structure along the spectral dimension at each time node [27]. To better understand this characteristic, we visually present the spectral

unfolding matrix of the cloud component for a time node in Fig. 2. It is evident that the unfolding matrix has a significant structural characteristic, and the elements of each column approximately tend to be all zero or all nonzero. Since the proportion of cloud regions is much smaller than that of cloud-free regions, the number of zero columns is much larger than that of the nonzero columns. Especially, the location of a column similar to a stripe is where the thick cloud is located. Inspired by this observation, the $l_{2,1}$-norm is introduced to characterize the structural sparse characteristic of the cloud component along the spectral dimension, i.e., the regularizer $R(\mathcal{C})$ is designed as follows:

$$R(\mathcal{C}) = \|\mathcal{C}\|_{2,1} = \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{q=1}^{t} \|\mathcal{C}_{i,j,:,q}\|_2 \tag{4}$$

where $\mathcal{C}_{i,j,:,q}$ is the tube of $\mathcal{C}$.

Based on the above analysis, the model-driven low-rank and structural sparse decomposition framework can be formulated as follows:

$$\min_{\mathcal{X},\mathcal{C},\mathcal{M}} \frac{1}{2}\|\mathcal{Y} - \mathcal{M} \odot \mathcal{X} - \mathcal{C}\|_F^2 + \lambda_1\|\mathcal{X}\|_* + \lambda_2\|\mathcal{C}\|_{2,1}. \tag{5}$$

C. Proposed LRRSSN

Model-driven priors effectively explore the intrinsic physical characteristics of the image and cloud components and limit the solution space to a reasonable range, but this is not sufficient to capture the complicated features of the MRSIs. Recent works have demonstrated that data-driven DL methods are capable of learning complicated structures in the data [54], [55], [56]. To leverage this advantage, we integrate the data-driven deep prior captured by neural networks into the image decomposition framework to improve the representation abilities of the thick cloud removal model. Many neural networks can be used to adaptively learn the implicit deep prior of the image component, such as FFDNet [57] and DIP [58]. However, since there may be cloud-free time nodes in MRSIs, these networks fail to fully utilize this information to guide accurate image reconstruction. Recently, Uezato et al. [59] presented a guided deep decoder (GDD) network to learn the deep image prior. GDD learns multiscale features from the guidance image and then incorporates the multiscale features into the deep decoder network to guide the image reconstruction. Moreover, GDD does not require external training data and employs a self-supervised approach to optimize network parameters. The network structure of GDD is shown in Fig. 3, and more details on GDD can be referred to [59]. Inspired by the superiority of GDD and the special data characteristics of MRSIs, we utilize GDD to learn implicit deep priors for the image component. The deep prior assumes that the image component $\mathcal{X}$ is the output of a neural network, i.e.,

$$\mathcal{X} = \mathcal{T}_\Theta(\mathcal{E}, \mathcal{G}) \tag{6}$$

where $\mathcal{T}$ represents the GDD network, Θ denotes the network parameters, $\mathcal{E}$ represents the random tensor input to the network, and $\mathcal{G}$ denotes the guidance image.
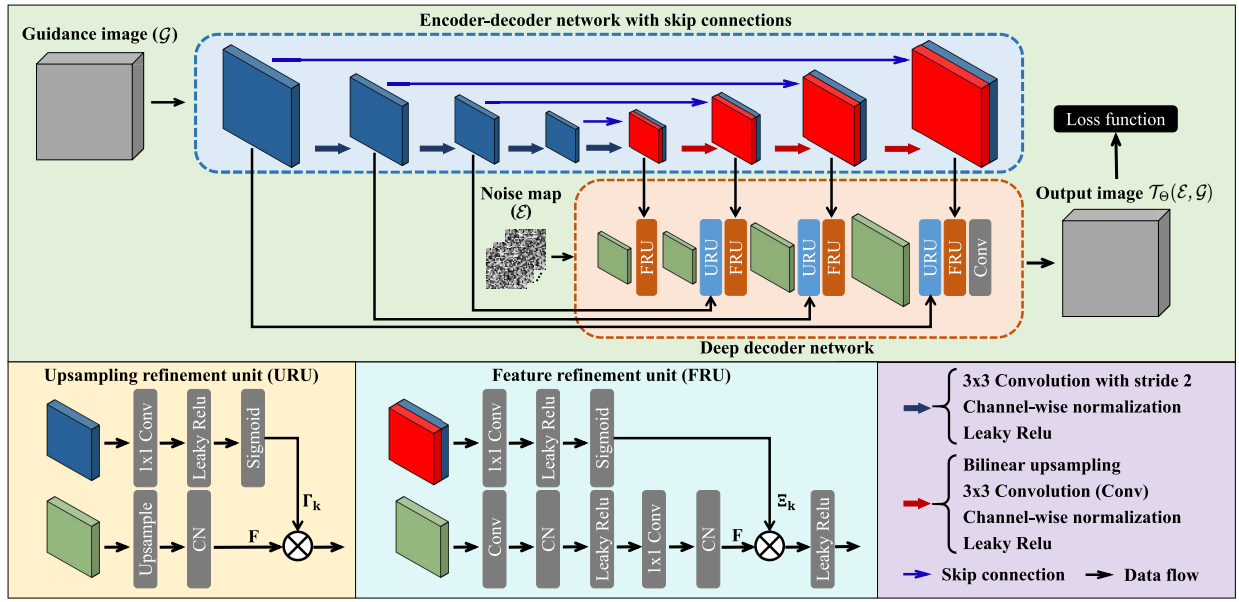
Fig. 3. Illustration of the GDD network. The guidance image $\mathcal{G}$ is first input into an encoder–decoder network to extract its multiscale semantic features. The semantic features are then employed to guide the parameter estimation and image generation in the deep decode by two units upsampling refinement unit (URU) and feature refinement unit (FRU), where URU integrates an attention gate to assign weights to features obtained after upsampling and channel-wise normalization (CN) within the deep decoder, and FRU leverages the high-level semantic features of the guidance image to promote semantic alignment.

Combining (5) and (6), we can obtain the proposed LRRSSN model as follows:

$$\min_{\mathcal{C},\Theta,\mathcal{M}} \frac{1}{2}\|\mathcal{Y} - \mathcal{M} \odot \mathcal{T}_\Theta(\mathcal{E},\mathcal{G}) - \mathcal{C}\|_F^2 + \lambda_1 \|\mathcal{T}_\Theta(\mathcal{E},\mathcal{G})\|_* + \lambda_2\|\mathcal{C}\|_{2,1}. \quad (7)$$

It is worth noting that in [59], GDD is directly employed as an end-to-end trained network for image fusion. This is very different from our proposed model. The proposed model integrates the advantages of the model-driven and data-driven methods while circumventing their disadvantages for MRSIs thick cloud removal, and GDD is embedded in an iterative optimization framework.

D. Optimization

By introducing the auxiliary variable $\mathcal{W}$, the problem can be rewritten as follows:

$$\min_{\mathcal{C},\Theta,\mathcal{M},\mathcal{W}} \frac{1}{2}\|\mathcal{Y} - \mathcal{M} \odot \mathcal{T}_\Theta(\mathcal{E},\mathcal{G}) - \mathcal{C}\|_F^2 + \lambda_1\|\mathcal{W}\|_* + \lambda_2\|\mathcal{C}\|_{2,1}$$
$$\text{s.t.} \quad \mathcal{W} = \mathcal{T}_\Theta(\mathcal{E},\mathcal{G}). \quad (8)$$

Based on the HQS algorithm, the quadratic penalty function of problem (8) is given as follows:

$$\min_{\mathcal{C},\Theta,\mathcal{M},\mathcal{W}} \frac{1}{2}\|\mathcal{Y} - \mathcal{M} \odot \mathcal{T}_\Theta(\mathcal{E},\mathcal{G}) - \mathcal{C}\|_F^2 + \lambda_1\|\mathcal{W}\|_* + \lambda_2\|\mathcal{C}\|_{2,1} + \frac{\rho}{2}\|\mathcal{T}_\Theta(\mathcal{E},\mathcal{G}) - \mathcal{W}\|_F^2 \quad (9)$$

where ρ is a penalty parameter. The optimization of (9) can be decomposed into the following subproblems.

Updating $\mathcal{W}$: The $\mathcal{W}$-subproblem is given as follows:

$$\arg\min_{\mathcal{W}} \lambda_1\|\mathcal{W}\|_* + \frac{\rho}{2}\|\mathcal{T}_\Theta(\mathcal{E},\mathcal{G}) - \mathcal{W}\|_F^2. \quad (10)$$

The closed-form solution of $\mathcal{W}$ can be obtained by the singular value thresholding operator

$$\mathcal{W} = \mathcal{R}^{-1}(U\Sigma_+ V^T) \quad (11)$$

where $\Sigma_+ = \max(\Sigma - \lambda_1/\rho, 0)$, and $U\Sigma V^T = \mathcal{R}(\mathcal{T}_\Theta(\mathcal{E},\mathcal{G}))$.

Updating $\mathcal{C}$: The $\mathcal{C}$-subproblem is given as follows:

$$\arg\min_{\mathcal{C}} \frac{1}{2}\|\mathcal{Y} - \mathcal{M} \odot \mathcal{T}_\Theta(\mathcal{E},\mathcal{G}) - \mathcal{C}\|_F^2 + \lambda_2\|\mathcal{C}\|_{2,1}. \quad (12)$$

The optimization is solved by the soft shrinkage operator

$$\mathcal{C}_{i,j,:,q} = \begin{cases} \dfrac{\|\mathcal{B}_{i,j,:,q}\|_2 - \lambda_2}{\|\mathcal{B}_{i,j,:,q}\|_2}\mathcal{B}_{i,j,:,q}, & \text{if } \|\mathcal{B}_{i,j,:,q}\|_2 > \lambda_2 \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

where $\mathcal{B} = \mathcal{Y} - \mathcal{M} \odot \mathcal{T}_\Theta(\mathcal{E},\mathcal{G})$.

Updating $\mathcal{M}$: When the cloud component is extracted from the observed MRSIs, a simple threshold method can be used to estimate the cloud mask. The idea is derived from that the location of the thick cloud has a relatively large value in the cloud component. Therefore, the cloud mask $\mathcal{M}$ is generated by the following threshold operator:

$$\mathcal{M}_{i,j,:,q} = \begin{cases} 1, & \text{if } \dfrac{1}{b}\sum_{p=1}^{b}|\mathcal{C}_{i,j,p,q}| < \epsilon \\ 0, & \text{else} \end{cases} \quad (14)$$

where $\epsilon > 0$ denotes a given threshold value.

Updating Θ: The Θ-subproblem is given as follows:

$$\arg\min_{\Theta} \frac{1}{2}\|\mathcal{Y} - \mathcal{M} \odot \mathcal{T}_\Theta(\mathcal{E},\mathcal{G}) - \mathcal{C}\|_F^2 + \frac{\rho}{2}\|\mathcal{T}_\Theta(\mathcal{E},\mathcal{G}) - \mathcal{W}\|_F^2. \quad (15)$$

The Θ-subproblem is a quadratic minimization problem, and many gradient descent methods can be employed to solve

it. To effectively optimize the Θ-subproblem, the Adam algorithm [60] is chosen as the optimizer in the GDD network. It is important to note that the guidance image is obtained from the weighted average of the last iteration of the network, rather than external auxiliary data, i.e.,

$$\mathcal{G}^{k+1} = \sum_{q=1}^{t} \mathbf{w}_q \mathcal{T}_\Theta(\mathcal{E}, \mathcal{G}^k)(:, :, :, q) \tag{16}$$

where $\mathbf{w}_q$ denotes the weight of the qth time node, and we empirically set weights of the cloud-free time node and cloud time node as $\mathbf{w}_q = 1$ and $\mathbf{w}_q = 0.1$, respectively.

Finally, to preserve the information of cloud-free regions in MRSIs, the thick cloud removal result (image component) is achieved as follows:

$$\mathcal{X} = \mathcal{M} \odot \mathcal{Y} + (1 - \mathcal{M}) \odot \mathcal{T}_\Theta(\mathcal{E}, \mathcal{G}). \tag{17}$$

We summarize the whole thick cloud removal procedure in Algorithm 1. We implement the GDD network in the PyTorch framework and minimize the loss function using the Adam optimizer ($\beta_1 = 0.9$ and $\beta_2 = 0.999$), and the learning rate is set as 0.002. All experiments are run on a platform with Intel i9-12900 and NVIDIA GeForce RTX 3090.

Algorithm 1 The Optimization Process for LRRSSN Solver

Input: Observed cloud-contaminated MRSIs $\mathcal{Y}$, regularization parameters λ_1 and λ_2, threshold value ϵ, and weight $\mathbf{w}$.

1: Initialize: Let $\mathcal{X} = \mathcal{Y}$, $\mathcal{C} = \mathcal{O}$, $\mathcal{M} = \mathcal{O}$, $\rho = 0.001$, $\alpha = 1.02$, $K = 25$, and tolerance for stopping criterion $tol = 10^{-3}$.
2: **for** $k = 1 : K$ **do**
3:     Update $\mathcal{W}$ via (11);
4:     Update $\mathcal{C}$ via (13);
5:     Update $\mathcal{M}$ via (14);
6:     Update guidance image $\mathcal{G}$ via (16);
7:     Update Θ via GDD in (15);
8:     Update the penalty parameter $\rho = \alpha * \rho$;
9:     Achieve the image component $\mathcal{X}$ via (17);
10:     Check the convergence condition: $\frac{\|\mathcal{X}^{(k)} - \mathcal{X}^{(k-1)}\|_F}{\|\mathcal{X}^{(k-1)}\|_F} \leq tol.$
11: **end for**

Output: Image component $\mathcal{X}$ and cloud component $\mathcal{C}$.

III. EXPERIMENTS

To validate the effectiveness of the proposed method for thick cloud removal in MRSIs, we conduct experiments on both synthetic and real datasets. Our method is compared with six state-of-the-art thick cloud removal methods, including model-driven methods with high-accuracy low-rank tensor completion (HaLRTC) [22], augmented Lagrangian method with inexact proximal gradients (ALM-IPGs) [61], total variation regularized low-rank sparsity decomposition with information compensation method (TVLRDC) [25], robust thick cloud removal method (RTCR) [29], self-supervised method MT [42], and inpainting method Mosaicing [43],

TABLE I
THREE DIFFERENT MRSIS IN THE SIMULATED EXPERIMENT

Dataset	Spatial size	Spectral number	Time node
Mountain	256×256	4 (B2, B3, B4, B8)	6
City	256×256	4 (B2, B3, B4, B8)	12
Farmland	256×256	12 (B1 to B12)	6

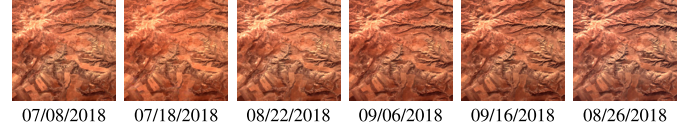07/08/2018   07/18/2018   08/22/2018   09/06/2018   09/16/2018   08/26/2018

Fig. 4. Pseudo color images (R: B4, G: B3, and B: B2) of the Mountain dataset ("MM/DD/YYYY" means the acquired time of the image).
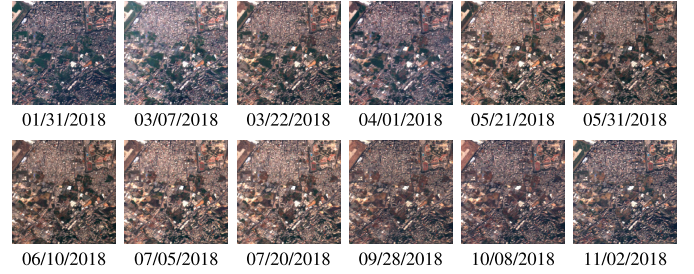

01/31/2018   03/07/2018   03/22/2018   04/01/2018   05/21/2018   05/31/2018

06/10/2018   07/05/2018   07/20/2018   09/28/2018   10/08/2018   11/02/2018

Fig. 5. Pseudo color images (R: B4, G: B3, and B: B2) of the City dataset ("MM/DD/YYYY" means the acquired time of the image).

05/31/2018   06/10/2018   06/25/2018   07/20/2018   07/30/2018   08/09/2018
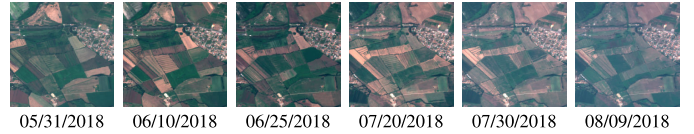
Fig. 6. Pseudo color images (R: B4, G: B3, and B: B2) of the Farmland dataset ("MM/DD/YYYY" means the acquired time of the image).

where HaLRTC and ALM-IPG are nonblind completion methods, RTCR is a semiblind optimization method, TVLRDC is a blind optimization method, MT is a nonblind DL method, and Mosaicing is a traditional nonblind inpainting method. The reproduction codes of comparison methods are released publicly or provided by the authors, and we try our best to choose the optimal parameters so as to achieve the best results. For nonblind methods, we provide them with an accurate mask.

A. Simulated Experimental Setting

To thoroughly evaluate the performance of the proposed methods and demonstrate the flexibility for thick cloud removal, we choose different scenes in MRSIs and consider different thick cloud-contaminated cases.

1) Datasets: We select three different ground object scenes from the SEM12MS-CR-TS dataset[1] as simulation data, termed the Mountain dataset, City dataset, and Farmland dataset, respectively. The image sizes of the three datasets are listed in Table I, and the pseudo color images and acquired times are presented in Figs. 4–6.

---

[1]https://github.com/PatrickTUM/SEN12MS-CR-TS

TABLE II
QUANTITATIVE COMPARISON OF DIFFERENT METHODS UNDER DIFFERENT CLOUD COVERAGE ON THE MOUNTAIN DATASET

| Cloud coverage	Index	Observed	HaLRTC	ALM-IPG	TVLRDC	RTCR	MT	Mosaicing	LRRSSN
15% ∼ 30%	PSNR	10.237	42.047	42.134	38.888	41.493	35.264	37.474	**45.152**
	SSIM	0.786	0.989	0.989	0.980	0.988	0.833	0.979	**0.994**
	RMSE	0.589	0.026	0.026	0.035	0.028	0.084	0.044	**0.017**
30% ∼ 45%	PSNR	7.835	36.511	36.992	32.863	37.187	32.002	34.168	**39.951**
	SSIM	0.646	0.972	0.968	0.94	0.974	0.925	0.957	**0.985**
	RMSE	0.645	0.049	0.046	0.068	0.044	0.079	0.067	**0.031**
45% ∼ 60%	PSNR	6.356	33.133	33.691	27.897	33.608	30.937	32.635	**36.238**
	SSIM	0.561	0.950	0.941	0.899	0.95	0.91	0.937	**0.968**
	RMSE	0.676	0.074	0.069	0.172	0.069	0.100	0.078	**0.048**
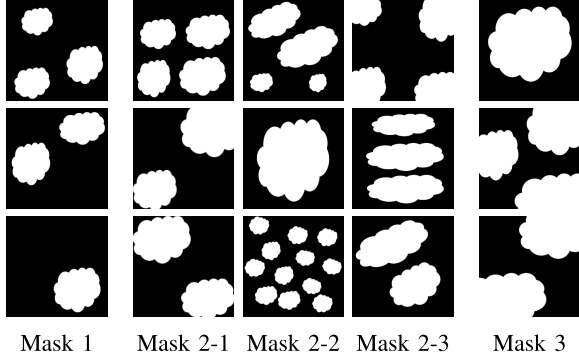
Fig. 7. Different shapes and coverages of cloud masks designed in the simulated experiment. The visualization is shown by $1-\mathcal{M}$, i.e., the white and black are thick cloud locations and cloud-free locations, respectively. Mask 1: Cloud coverage ranges from 15% to 30%. Mask 2-(1-3): Cloud coverage ranges from 30% to 45%. Mask 3: Cloud coverage ranges from 45% to 60%.

Mask 1   Mask 2-1   Mask 2-2   Mask 2-3   Mask 3

2) Cloud Mask and Thick Cloud Simulation Setting: Real-world MRSIs are usually contaminated by different types of thick clouds, and the shape and coverage of clouds in MRSIs change over time. To simulate these real thick cloud scenarios as far as possible, we design different shapes and coverage of cloud masks in the simulated experiment, which are shown in Fig. 7. Given the cloud mask $\mathcal{M}$, we simulate the thick cloud-contaminated MRSIs according to the following formula:

$$\mathcal{Y}_{i,j,p,q} = \begin{cases} 1 & \text{if } \mathcal{M}_{i,j,p,q} = 0 \\ \mathcal{X}_{i,j,p,q} & \text{otherwise} \end{cases}$$

where $\mathcal{X}$ is the ground truth image.

3) Simulated Datasets: To demonstrate the flexibility of our approach to the thick cloud, we consider the impact of different factors on the results of thick cloud removal. The influence factors of cloud coverage, the number of time nodes, and the number of spectral bands are designed in our experiments. Next, we illustrate in detail how to simulate the cloud-contaminated MRSIs.

Mountain Dataset: We conduct different cloud coverage on the Mountain dataset to evaluate the performance of different methods. Three different cloud coverages are considered in the experiment, including $15\%-30\%$, $30\%-45\%$, and $45\%-60\%$. Since there will be cloud-free time nodes in the real scene, we select the first three time nodes to add a thick cloud, and the remaining time nodes are cloud-free data. The cloud masks with different cloud coverage shown in Fig. 7 are employed to generate the contaminated image with a thick cloud.

City Dataset: In the City dataset, we examine the impact of the number of time nodes. We change the time nodes from the set of $N = \{3, 6, 9, 12\}$ and add the cloud coverage within the range of 30%–45%. To simulate the different cloud-contaminated images, we choose $2N/3$ of the time nodes and select the corresponding number of cloud masks from the Mask 2 group to add the thick cloud.

Farmland Dataset: We verify the robustness of the proposed method with regard to different numbers of spectral bands in each time node on the Farmland dataset. These experiments are conducted under the condition of cloud coverage with 30%–45%. The number of spectral bands varies from 4 to 10 with a step of 2. Analogously, the first three-time nodes are selected to add a thick cloud, and Mask 2-1 is employed as cloud masks.

4) Quantitative Metrics: To evaluate the reconstruction quality of different thick cloud removal methods, three quantitative metrics, including peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [62], and root-mean-square error (RMSE), are employed. We calculate the mean values of recovered images for cloud-contaminated temporal nodes to evaluate different methods. Generally, the higher values of PSNR and SSIM and the smaller results of RMSE typically indicate a better thick cloud removal result.

B. Simulate Experimental Results

1) Experimental Results on Mountain Dataset: We present the quantitative comparison of different methods under different cloud coverages on the Mountain dataset in Table II, with the best results highlighted in bold. We can clearly observe from Table II that the proposed LRRSSN outperforms other comparison methods based on all quantitative metrics. The quality of all thick cloud removal methods decreases as cloud coverage increases, indicating that large cloud coverage is relatively more difficult. In particular, the results of the proposed LRRSSN method are significantly improved compared to RTCR, indicating the effectiveness of a data-driven self-supervised network. Moreover, the proposed self-supervised method is obviously superior to the self-supervised learning MT method, which is mainly due to the combination of model-driven and data-driven methods.

To further illustrate the reconstruction quality of different methods, we present the pseudo color images obtained by different thick cloud removal methods on the Mountain dataset with cloud coverage from 30% to 45%, as shown in Fig. 8.
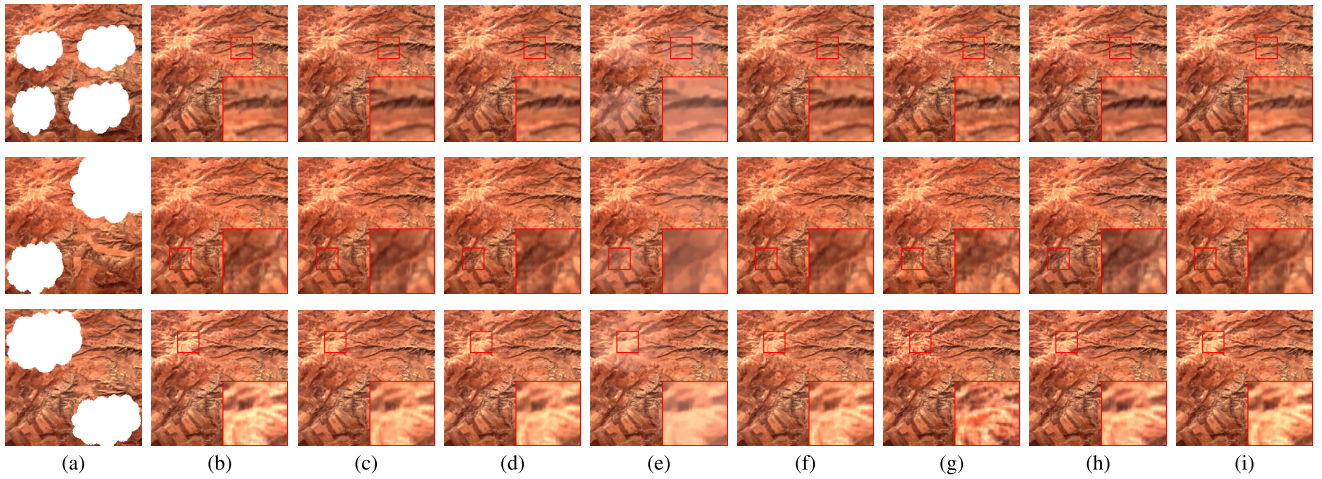
Fig. 8. Recovered results of all comparison methods on the Mountain dataset with cloud coverage from 30% to 45%. From top to bottom: recovered results of time nodes 1, 2, and 3. (a) Observed. (b) Ground truth. (c) HaLRTC. (d) ALM-IPG. (e) TVLRDC. (f) RTCR. (g) MT. (h) Mosaicing. (i) LRRSSN.

TABLE III
QUANTITATIVE COMPARISON OF DIFFERENT METHODS UNDER DIFFERENT TIME NODES ON THE CITY DATASET

| Time node	Index	Observed	HaLRTC	ALM-IPG	TVLRDC	RTCR	MT	Mosaicing	LRRSSN
3	PSNR	6.882	29.794	30.191	26.735	29.558	28.167	28.722	**32.314**
	SSIM	0.606	0.936	**0.943**	0.848	0.935	0.888	0.941	0.940
	RMSE	0.71	0.144	0.137	0.185	0.146	0.157	0.156	**0.107**
6	PSNR	7.098	32.426	32.587	26.944	32.651	26.642	29.56	**34.869**
	SSIM	0.622	0.946	0.947	0.83	0.948	0.826	0.910	**0.966**
	RMSE	0.705	0.102	0.100	0.201	0.098	0.182	0.136	**0.079**
9	PSNR	7.048	33.162	33.089	28.606	32.945	28.12	29.814	**34.904**
	SSIM	0.611	0.947	0.947	0.880	0.946	0.857	0.911	**0.961**
	RMSE	0.709	0.094	0.095	0.171	0.096	0.164	0.134	**0.075**
12	PSNR	6.902	33.924	33.815	28.925	33.506	27.69	29.99	**35.083**
	SSIM	0.606	0.955	0.954	0.884	0.952	0.853	0.914	**0.962**
	RMSE	0.713	0.086	0.087	0.159	0.088	0.161	0.135	**0.073**
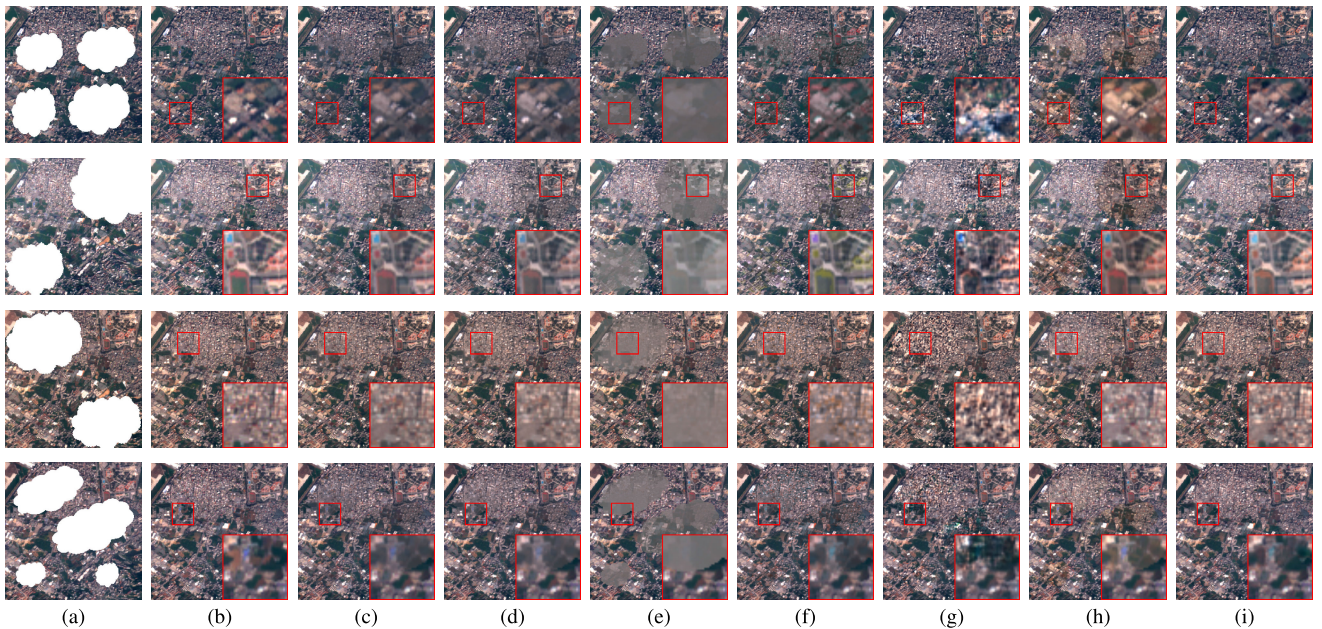
Fig. 9. Recovered results of all comparison methods on the City dataset with six-time nodes. From top to bottom: recovered results of time nodes 1, 2, 3, and 4. (a) Observed. (b) Ground truth. (c) HaLRTC. (d) ALM-IPG. (e) TVLRDC. (f) RTCR. (g) MT. (h) Mosaicing. (i) LRRSSN.

To better highlight the difference between different results, we magnify a representative subregion in the image. We can see that TVLRDC blurs the image details, while HaLRTC, ALM-IPG, RTCR, and MT suffer from spectral distortion in the first time node. The Mosaicing method obtains undesirable ringing artifacts. In contrast, the proposed LRRSSN effectively suppresses the thick clouds while reconstructing fine image structures and details.

TABLE IV
QUANTITATIVE COMPARISON OF DIFFERENT METHODS UNDER DIFFERENT SPECTRAL BANDS ON THE FARMLAND DATASET

| Spectral band	Index	Observed	HalRTC	ALM-IPG	TVLRDC	RTCR	MT	Mosaicing	LRRSSN
4	PSNR	6.678	32.435	33.045	29.645	32.328	30.639	31.416	**34.152**
	SSIM	0.673	0.924	0.931	0.882	0.922	0.875	0.919	**0.939**
	RMSE	0.744	0.148	0.132	0.176	0.146	0.161	0.146	**0.112**
6	PSNR	7.068	32.303	32.364	29.279	31.587	29.537	29.403	**33.991**
	SSIM	0.666	0.921	0.924	0.861	0.910	0.854	0.905	**0.932**
	RMSE	0.718	0.120	0.116	0.232	0.122	0.165	0.132	**0.088**
8	PSNR	7.377	31.886	32.048	28.859	31.298	30.425	30.471	**33.899**
	SSIM	0.660	0.917	0.918	0.870	0.908	0.842	0.908	**0.932**
	RMSE	0.692	0.123	0.110	0.196	0.113	0.143	0.124	**0.086**
10	PSNR	7.397	32.770	32.975	29.431	32.182	30.761	30.265	**34.269**
	SSIM	0.662	0.925	0.928	0.881	0.916	0.835	0.910	**0.934**
	RMSE	0.692	0.116	0.106	0.164	0.105	0.214	0.120	**0.082**
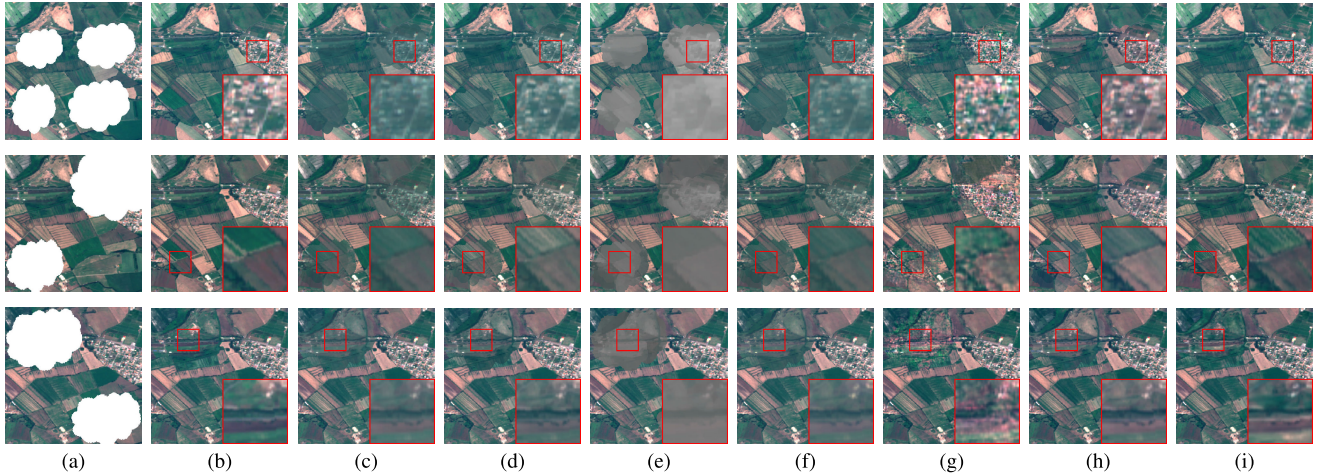
Fig. 10. Recovered results of all comparison methods on the Farmland dataset with four spectral bands. From top to bottom: recovered results of time nodes 1, 2, and 3. (a) Observed. (b) Ground truth. (c) HaLRTC. (d) ALM-IPG. (e) TVLRDC. (f) RTCR. (g) MT. (h) Mosaicing. (i) LRRSSN.

2) Experimental Results on City Dataset: Table III presents the quantitative comparison of different methods under different time nodes on the City dataset. With the increase in time nodes, the results of using low-rank regularization methods are gradually improved, as the spectral–temporal correlation becomes stronger due to the increased auxiliary information. However, the quantitative results of the self-supervised DL method do not improve with the increase in auxiliary information, indicating that the lack of model-driven prior makes them to not make full use of the auxiliary information. From Table III, we can find that our LRRSSN model achieves the best performance among all the comparison methods in different cases, further indicating the advantage of coupling model-driven and data-driven methods.

The image recovery of different methods on the City dataset in the case of six-time nodes is shown in Fig. 9. It is easy to see that the proposed LRRSSN method can achieve better thick cloud removal performance, i.e., more faithfully maintaining the image's spectral color and recovering the image details. It should be noted that the comparison methods either cause spectral distortion or destroy the spatial features of the image. The proposed method better balances the model-driven low-rank prior and data-driven deep prior of HSI data to get good thick cloud removal results.

3) Experimental Results on Farmland Dataset: The experimental results on the Farmland dataset are listed in Table IV. As the number of spectral bands gradually increases, the

performance of each method does not show a significant improvement. The reason is that the thick cloud contaminates all spectral bands, and an increase in the number of spectral bands does not contribute additional information about the cloudy region. It can be observed that the proposed LRRSSN achieves better quantitative results than all comparison methods in all cases. Furthermore, the proposed LRRSSN achieves a result nearly 5 dB higher than that of TVLRDC, demonstrating the advantage of the self-supervised deep network.

In Fig. 10, we show the visual recovery performance of all comparison methods in the case of four spectral bands. This figure clearly shows that all methods cannot perfectly recover the information of the original image, and the comparison methods have obviously damaged the color fidelity of pseudo color images and lost spatial details. Comparatively, the proposed LRRSSN model still maintains a relatively better performance, maintaining better texture details and achieving better color fidelity, which is consistent with the experimental results of the other two datasets.

C. Real Experiment

To further demonstrate the performance of the proposed method, we conduct to remove the real thick clouds. The real dataset is derived from the 112th region of interest in the SEM12MS-CR-TS dataset. This data is of the size $256 \times 256 \times 13 \times 5$, and four of the time nodes are heavily polluted
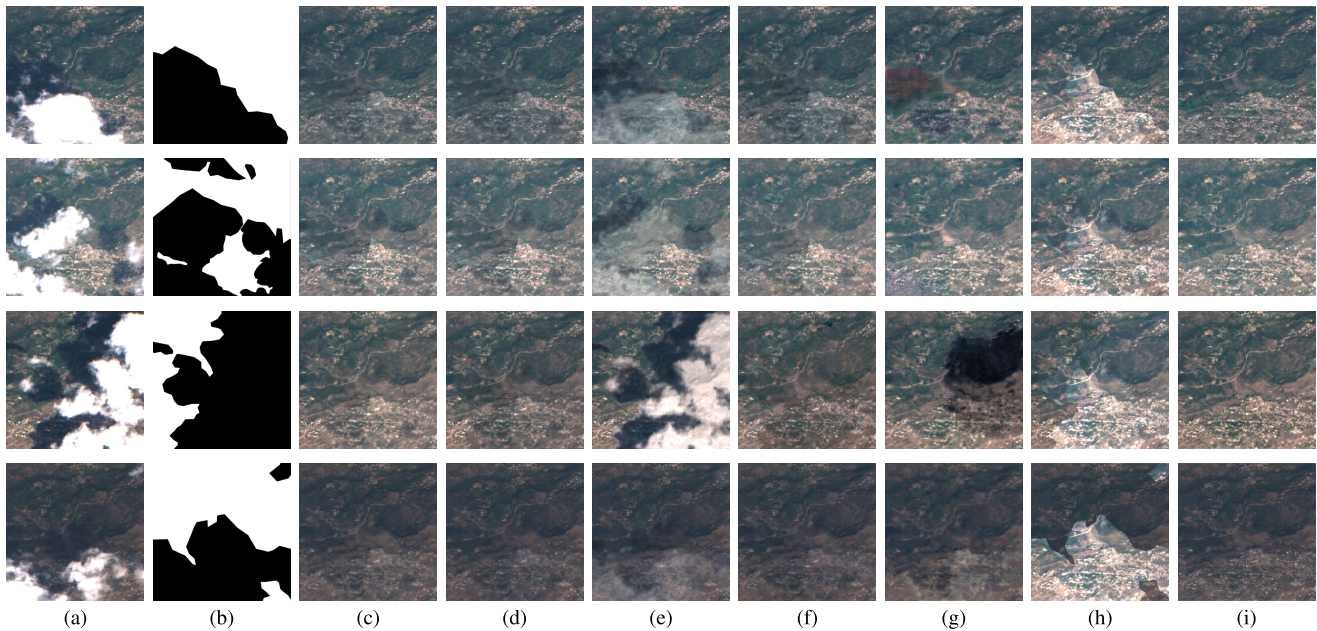
Fig. 11. Recovered results of all comparison methods on the real dataset. From top to bottom: recovered results of time nodes 1, 2, 3, and 5. (a) Observed. (b) Mask. (c) HaLRTC. (d) ALM-IPG. (e) TVLRDC. (f) RTCR. (g) MT. (h) Mosaicing. (i) LRRSSN.

by thick clouds and shadows shown in Fig. 11(a). Since the comparison methods require cloud masks, we provide them with a mask that covers all clouds and shadow areas shown in Fig. 11(b).

Fig. 11 displays the recovered performance of all methods on the real dataset. TVLRDC reduces the thickness of the clouds but fails to remove the thick clouds and restore the original image. MT can handle small-scale thick clouds, but it is difficult to solve the time node of large-scale cloud coverage. Mosaicing mainly employs the information of the cloud-free time node to remove the thick clouds, so that the pseudo information can be restored. Given an accurate mask, methods such as HaLRTC, ALM-IPG and RTCR can remove thick clouds. However, they introduce significant ringing artifacts in the transition region between dense clouds and clear regions. In contrast, the proposed LRRSSN not only recovers fine image details but also eliminates the visual artifacts effectively. Moreover, the information reconstructed by the proposed method in thick cloud areas maintains a consistent brightness gradient with cloud-free areas, making the reconstruction effect more reasonable.

D. Ablation Study and Discussion

1) Ablation Study: We conduct several ablation studies to verify the effectiveness of coupling the model-driven low-rank prior and data-driven deep prior. First, to investigate the necessity of incorporating the data-driven deep prior, we disable the deep regularizer (i.e., GDD network) of the image component, which is denoted as LRRSSN w/o DP. Moreover, we show the results of disabling the model-driven low-rank prior, which is represented as LRRSSN w/o LR. Table V lists the recovered PSNRs obtained by the original LRRSSN and its variants on all simulated datasets. It is evident that LRRSSN outperforms its variants, and these two

regularizers contribute significantly to the performance of the proposed LRRSSN. Specifically, in the case of large cloud coverage areas (such as the case of 45%–60% in the Mountain dataset), the introduced data-driven deep prior can significantly improve the ability of thick cloud removal. When there are more time nodes in MRSIs (such as the case of the City dataset with twelve-time nodes), model-driven low-rank prior can capture global spectral–temporal correlation and make the reconstruction results better. In summary, the ablation studies demonstrate that the model-driven low-rank and data-driven deep prior are complementary in the proposed method.

2) Analysis of Regularization Parameter λ_1: Parameter λ_1 is employed to control the contribution of the model-driven low-rank regularizer. In our model, we set the parameter $\lambda_1 = c\rho$, where c is a tuning parameter. Fig. 12(a) shows the change in PSNR values with respect to the parameter c in the City dataset with six-time nodes. With the increase in parameter c, the PSNR values increase significantly. This outcome is reasonable since the enhancement of low-rank constraints should lead to higher reconstruction accuracy. However, choosing a larger parameter results in loss of information, leading to lower results. From the result, the best result is achieved when c is set as 25, thus we set $\lambda_1 = 25\rho$ in all experiments.

3) Analysis of Regularization Parameter λ_2: Parameter λ_2 is employed to control the structural sparsity of the cloud component. Fig. 12(b) shows the change in the PSNR values with the change in the parameter λ_2. It is shown that the thick cloud removal result depends on the choice of λ_2. Since the degree of structural sparsity of cloud components varies in different scenarios, we empirically select parameter λ_2 in the interval [0.1, 0.7].

4) Convergence Analysis: Since the solved algorithm involves the optimization of the deep neural network (DNN), it is difficult to give a convergence theoretically. Instead,

TABLE V
PSNR RESULTS OF ABLATION STUDY ON THE SIMULATED DATASETS

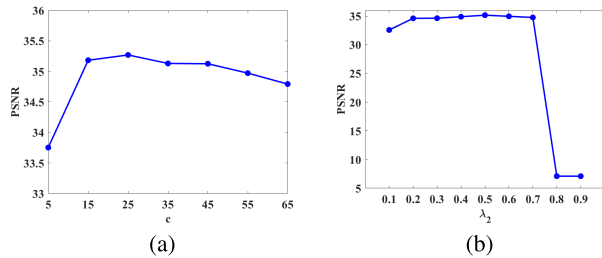| model	Mountain dataset (cloud coverage)			City dataset (time node)				Farmland dataset (spectral band)			
	$15\% \sim 30\%$	$30\% \sim 45\%$	$45\% \sim 60\%$	3	6	9	12	4	6	8	10
LRRSSN w/o DP	42.009	35.715	33.645	31.170	32.509	33.240	34.007	31.980	32.081	31.891	32.704
LRRSSN w/o LR	40.451	38.346	35.524	29.339	33.163	32.234	32.460	33.077	32.589	32.773	33.962
LRRSSN	**45.152**	**39.951**	**36.238**	**32.314**	**34.869**	**34.904**	**35.083**	**34.152**	**33.991**	**33.899**	**34.269**

Fig. 12. Sensitivity analysis of the (a) parameters λ_1 and (b) λ_2 under the case of City dataset with six-time nodes.
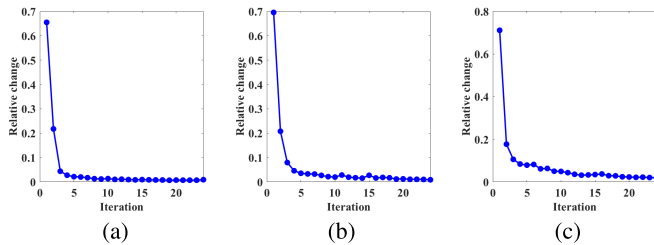

Fig. 13. RelCha versus iteration number in three different cases. (a) Mountain dataset with cloud coverage ranging from 30% to 45%. (b) City dataset with six-time nodes. (c) Farmland dataset with four spectral bands.

we numerically demonstrate the convergence of the proposed method. Fig. 13 presents the relative change (RelCha) versus the iteration number on three different thick cloud removal cases. It is evident that the RelCha values achieve stability and converge to nearly zero as the iteration number increases, indicating the strong convergence of the proposed HQS algorithm to solve the LRRSSN model.

IV. CONCLUSION

In this article, we have proposed a novel thick cloud removal method for MRSIs, named LRRSSN, by coupling model-driven optimization approaches and data-driven DL techniques. To make the recovery results not affected by cloud masks, we regard the thick cloud removal problem as an image restoration problem and decompose the observed image into the equally important image component and cloud component. In this decomposition framework, the model-driven low-rank prior effectively discriminates the physical characteristics of the image component. Meanwhile, the structural sparsity applied to the cloud component aids in extracting the thick cloud image distinctly. Moreover, the introduced data-driven deep prior inherits the powerful nonlinear fitting capability of DL-based approaches for the effective exploitation of local and nonlocal multiscale features in the image. In general, the proposed method does not require cloud masks and external training datasets, so it can be flexibly applied to thick cloud

removal problems for different types of data. A series of simulated and real data experiments have been conducted to demonstrate the effectiveness of the proposed method over some model-driven or data-driven methods in terms of both the quantitative metrics and visual results. In the future, we will consider using the neural network to learn the regularization parameters in the model adaptively.

REFERENCES

[1] D. Hong et al., "Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks," *Remote Sens. Environ.*, vol. 299, Dec. 2023, Art. no. 113856.

[2] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, Aug. 2020.

[3] N. Yokoya and A. Iwasaki, "Object detection based on sparse representation and Hough voting for optical remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 2053–2062, May 2015.

[4] J. Peng, S. Chen, H. Lü, Y. Liu, and J. Wu, "Spatiotemporal patterns of remotely sensed $PM_{2.5}$ concentration in China from 1999 to 2011," *Remote Sens. Environ.*, vol. 174, pp. 109–121, Mar. 2016.

[5] Q. Weng, "Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends," *ISPRS J. Photogramm. Remote Sens.*, vol. 64, no. 4, pp. 335–344, Jul. 2009.

[6] J. Ju and D. P. Roy, "The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally," *Remote Sens. Environ.*, vol. 112, no. 3, pp. 1196–1211, Mar. 2008.

[7] L. Wang et al., "Satellite-based assessment of the long-term efficacy of $PM_{2.5}$ pollution control policies across the Taiwan strait," *Remote Sens. Environ.*, vol. 251, Dec. 2020, Art. no. 112067.

[8] M. Xu, F. Deng, S. Jia, X. Jia, and A. J. Plaza, "Attention mechanism-based generative adversarial networks for cloud removal in Landsat images," *Remote Sens. Environ.*, vol. 271, Mar. 2022, Art. no. 112902.

[9] A. Maalouf, P. Carre, B. Augereau, and C. Fernandez-Maloigne, "A bandelet-based inpainting technique for clouds removal from remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2363–2371, Jul. 2009.

[10] L. Lorenzi, F. Melgani, and G. Mercier, "Inpainting strategies for reconstruction of missing data in VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 914–918, Sep. 2011.

[11] H. Shen, H. Li, Y. Qian, L. Zhang, and Q. Yuan, "An effective thin cloud removal procedure for visible remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 96, pp. 224–235, Oct. 2014.

[12] H. Xu, X. Tang, B. Ai, X. Gao, F. Yang, and Z. Wen, "Missing data reconstruction in VHR images based on progressive structure prediction and texture generation," *ISPRS J. Photogramm. Remote Sens.*, vol. 171, pp. 266–277, Jan. 2021.

[13] P. Rakwatin, W. Takeuchi, and Y. Yasuoka, "Restoration of Aqua MODIS band 6 using histogram matching and local least squares fitting," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 2, pp. 613–627, Feb. 2009.

[14] H. Shen, C. Zeng, and L. Zhang, "Recovering reflectance of Aqua MODIS band 6 based on within-class local fitting," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 185–192, Mar. 2011.

[15] C. Zhang, W. Li, and D. J. Travis, "Restoration of clouded pixels in multispectral remotely sensed imagery with cokriging," *Int. J. Remote Sens.*, vol. 30, no. 9, pp. 2173–2195, May 2009.

[16] M. Xu, M. Pickering, A. J. Plaza, and X. Jia, "Thin cloud removal based on signal transmission principles and spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1659–1669, Mar. 2016.

[17] M. Xu, X. Jia, M. Pickering, and S. Jia, "Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform," *ISPRS J. Photogramm. Remote Sens.*, vol. 149, pp. 215–225, Mar. 2019.

[18] X. Li, H. Shen, L. Zhang, and H. Li, "Sparse-based reconstruction of missing information in remote sensing images from spectral/temporal complementary information," *ISPRS J. Photogramm. Remote Sens.*, vol. 106, pp. 1–15, Aug. 2015.

[19] X. Li, H. Shen, H. Li, and L. Zhang, "Patch matching-based multitemporal group sparse representation for the missing information reconstruction of remote-sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 8, pp. 3629–3641, Aug. 2016.

[20] M. K. Ng, Q. Yuan, L. Yan, and J. Sun, "An adaptive weighted tensor completion method for the recovery of remote sensing images with missing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3367–3381, Jun. 2017.

[21] C.-H. Lin, P.-H. Tsai, K.-H. Lai, and J.-Y. Chen, "Cloud removal from multitemporal satellite images using information cloning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 232–241, Jan. 2013.

[22] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.

[23] T.-Y. Ji, N. Yokoya, X. X. Zhu, and T.-Z. Huang, "Nonlocal tensor completion for multitemporal remotely sensed images' inpainting," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3047–3061, Jun. 2018.

[24] W. He, N. Yokoya, L. Yuan, and Q. Zhao, "Remote sensing image reconstruction using tensor ring completion and total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8998–9009, Nov. 2019.

[25] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, "Blind cloud and cloud shadow removal of multitemporal images based on total variation regularized low-rank sparsity decomposition," *ISPRS J. Photogramm. Remote Sens.*, vol. 157, pp. 93–107, Nov. 2019.

[26] Y. Zhang, F. Wen, Z. Gao, and X. Ling, "A coarse-to-fine framework for cloud removal in remote sensing image sequence," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5963–5974, Aug. 2019.

[27] T.-Y. Ji, D. Chu, X.-L. Zhao, and D. Hong, "A unified framework of cloud detection and removal based on low-rank and group sparse regularizations for multitemporal multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022, Art. no. 5303015.

[28] Y. Chen, W. He, N. Yokoya, and T. Huang, "Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3556–3570, Aug. 2020.

[29] J. Lin, T.-Z. Huang, X.-L. Zhao, Y. Chen, Q. Zhang, and Q. Yuan, "Robust thick cloud removal for multitemporal remote sensing images using coupled tensor factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022, Art. no. 5406916.

[30] W.-J. Zheng, X.-L. Zhao, Y.-B. Zheng, J. Lin, L. Zhuang, and T.-Z. Huang, "Spatial–spectral–temporal connective tensor network decomposition for thick cloud removal," *ISPRS J. Photogramm. Remote Sens.*, vol. 199, pp. 182–194, May 2023.

[31] Y. Luo, X. Zhao, Z. Li, M. K. Ng, and D. Meng, "Low-rank tensor function representation for multi-dimensional data recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–18, 2023, doi: 10.1109/TPAMI.2023.3341688.

[32] D. Hong et al., "SpectralGPT: Spectral foundation model," 2023, arXiv:2311.07113.

[33] Z. Xu, K. Wu, L. Huang, Q. Wang, and P. Ren, "Cloudy image arithmetic: A cloudy scene synthesis paradigm with an application to deep-learning-based thin cloud removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022, Art. no. 5612616.

[34] Y. Li, F. Wei, Y. Zhang, W. Chen, and J. Ma, "HS2P: Hierarchical spectral and structure-preserving fusion network for multimodal remote sensing image cloud and shadow removal," *Inf. Fusion*, vol. 94, pp. 215–228, Jun. 2023.

[35] C. Tao, S. Fu, J. Qi, and H. Li, "Thick cloud removal in optical remote sensing images using a texture complexity guided self-paced learning method," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5619612.

[36] Y. Chen, Q. Weng, L. Tang, X. Zhang, M. Bilal, and Q. Li, "Thick clouds removing from multitemporal Landsat images using spatiotemporal neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022, Art. no. 4400214.

[37] Q. Zhang, Q. Yuan, J. Li, Z. Li, H. Shen, and L. Zhang, "Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 148–160, Apr. 2020.

[38] S. Ji, P. Dai, M. Lu, and Y. Zhang, "Simultaneous cloud detection and removal from bitemporal remote sensing images using cascade convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 732–748, Jan. 2021.

[39] Q. Zhang, Q. Yuan, Z. Li, F. Sun, and L. Zhang, "Combined deep prior with low-rank tensor SVD for thick cloud removal in multitemporal images," *ISPRS J. Photogramm. Remote Sens.*, vol. 177, pp. 161–173, Jul. 2021.

[40] B. Jiang et al., "A deep-learning reconstruction method for remote sensing images with large thick cloud cover," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 115, Dec. 2022, Art. no. 103079.

[41] C. Lu, M. Xia, M. Qian, and B. Chen, "Dual-branch network for cloud and cloud shadow segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022, Art. no. 5410012.

[42] M. Czerkawski et al., "Deep internal learning for inpainting of cloud-affected regions in satellite imagery," *Remote Sens.*, vol. 14, no. 6, p. 1342, Mar. 2022.

[43] P. Ebel, Y. Xu, M. Schmitt, and X. X. Zhu, "SEN12MS-CR-TS: A remote-sensing data set for multimodal multitemporal cloud removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022, Art. no. 5222414.

[44] J. Li et al., "Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 373–389, Aug. 2020.

[45] W. He and N. Yokoya, "Multi-temporal Sentinel-1 and -2 data fusion for optical image simulation," *ISPRS Int. J. Geo-Information*, vol. 7, no. 10, p. 389, Sep. 2018.

[46] J. Gao, Q. Yuan, J. Li, H. Zhang, and X. Su, "Cloud removal with fusion of high resolution optical and SAR images using generative adversarial networks," *Remote Sens.*, vol. 12, no. 1, p. 191, Jan. 2020.

[47] A. Meraner, P. Ebel, X. X. Zhu, and M. Schmitt, "Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 333–346, Aug. 2020.

[48] F. N. Darbaghshahi, M. R. Mohammadi, and M. Soryani, "Cloud removal in remote sensing images using generative adversarial networks and SAR-to-optical image translation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–9, 2022, Art. no. 4105309.

[49] B.-Z. Li, X.-L. Zhao, X. Zhang, T.-Y. Ji, X. Chen, and M. K. Ng, "A learnable group-tube transform induced tensor nuclear norm and its application for tensor completion," *SIAM J. Imag. Sci.*, vol. 16, no. 3, pp. 1370–1397, Sep. 2023.

[50] X.-L. Zhao, J.-H. Yang, T.-H. Ma, T.-X. Jiang, M. K. Ng, and T.-Z. Huang, "Tensor completion via complementary global, local, and nonlocal priors," *IEEE Trans. Image Process.*, vol. 31, pp. 984–999, 2022.

[51] Y. Luo, X. Zhao, T. Jiang, Y. Chang, M. K. Ng, and C. Li, "Self-supervised nonlinear transform-based tensor nuclear norm for multi-dimensional image recovery," *IEEE Trans. Image Process.*, vol. 31, pp. 3793–3808, 2022.

[52] J.-L. Wang, X.-L. Zhao, H.-C. Li, K.-X. Cao, J. Miao, and T.-Z. Huang, "Unsupervised domain factorization network for thick cloud removal of multitemporal remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–12, 2023, Art. no. 5405912.

[53] Y. Chen, W. Lai, W. He, X.-L. Zhao, and J. Zeng, "Hyperspectral compressive snapshot reconstruction via coupled low-rank subspace representation and self-supervised deep network," *IEEE Trans. Image Process.*, vol. 33, pp. 926–941, 2024, doi: 10.1109/TIP.2024.3354127.

[54] C. Li, B. Zhang, D. Hong, J. Yao, and J. Chanussot, "LRR-Net: An interpretable deep unfolding network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023.

[55] D. Hong, J. Yao, C. Li, D. Meng, N. Yokoya, and J. Chanussot, "Decoupled-and-coupled networks: Self-supervised hyperspectral image super-resolution with subpixel fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–12, 2023, Art. no. 5527812.

[56] Y. Chen, X. Gui, J. Zeng, X.-L. Zhao, and W. He, "Combining low-rank and deep plug-and-play priors for snapshot compressive imaging," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–13, 2023, doi: 10.1109/TNNLS.2023.3294262.

[57] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[58] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.

[59] T. Uezato, D. Hong, N. Yokoya, and W. He, "Guided deep decoder: Unsupervised image pair fusion," in *Proc. Eur. Conf. Comput. Vis. (ECCV).* Cham, Switzerland: Springer, 2020, pp. 87–102.

[60] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980.*

[61] J. Wang, P. A. Olsen, A. R. Conn, and A. C. Lozano, "Removing clouds and recovering ground observations in satellite image sequences via temporally contiguous robust matrix completion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2754–2763.

[62] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

Jinshan Zeng (Member, IEEE) received the Ph.D. degree in mathematics from Xi'an Jiaotong University, Xi'an, China, in 2015.

He is currently a Professor with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang, China, where he has been serving as the Associate Dean since July 2022. He has authored over 60 paperss in high-impact journals and conferences such as *Journal of Machine Learning Research* (JMLR), IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), IEEE TRANSACTIONS ON SIGNAL PROCESSING (TSP), IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP), IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS (TNNLS), IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS), ICML, and AAAI. He has had two papers coauthored with collaborators that received the International Consortium of Chinese Mathematicians (ICCM) Best Paper Award in 2018 and 2020. His research interests include nonconvex optimization, machine learning, remote sensing, and computer vision.

Yong Chen received the B.S. degree from the School of Science, East China University of Technology, Nanchang, China, in 2015, and the Ph.D. degree from the School of Mathematical Sciences, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2020.

From 2018 to 2019, he was a Research Intern with the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan. He is currently working with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang. His research interests include hyperspectral image processing, low-rank matrix/tensor representation, and model-driven deep learning.

Maolin Chen received the B.S. degree from Jiangxi Normal University, Nanchang, China, in 2022, where he is currently pursuing the M.S. degree with the School of Computer and Information Engineering.

His research interests include cloud removal and deep learning.

Min Huang received the B.Sc. degree in remote sensing science and technology and the Ph.D. degree in cartography and geographical information system from Wuhan University, Wuhan, China, in 2015 and 2021, respectively.

He is currently a Lecturer with the School of Geography and Environment, Jiangxi Normal University, Nanchang, China. His research interests include urban remote sensing, smart cities, land use/land cover classification, surface parameter inversion, multisourced data fusion, and geospatial big data analysis.

Wei He (Senior Member, IEEE) received the B.S. degree from the School of Mathematics and Statistics, Wuhan University, Wuhan, China, in 2012, and the Ph.D. degree from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, in 2017.

From 2018 to 2020, he was a Researcher at the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan, where he was a Research Scientist from 2020 to 2021. He is currently a Full Professor with LIESMARS, Wuhan University. His research interests include image quality improvement, remote sensing image processing and low-rank representation, and deep learning.

Yu-Bang Zheng (Member, IEEE) received the B.S. degree from the Institute of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, China, in 2017, and the Ph.D. degree from the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China, in 2022.

From 2021 to 2022, he was a Student Trainee with the Tensor Learning Team, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan. He is currently working with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu. His research interests include tensor modeling and computing, tensor learning, and high-dimensional data processing.