ISPRS Journal of Photogrammetry and Remote Sensing 162 (2020) 148–160

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning

Qiang Zhang[a], Qiangqiang Yuan[b,c,*], Jie Li[b], Zhiwei Li[d], Huanfeng Shen[d], Liangpei Zhang[a,c,*]

[a] State Key Laboratory of Information Engineering, Survey Mapping and Remote Sensing, Wuhan University, Wuhan, China
[b] School of Geodesy and Geomatics, Wuhan University, Wuhan, China
[c] Collaborative Innovation Center of Geospatial Technology, Wuhan, China
[d] School of Resource and Environmental Science, Wuhan University, Wuhan, China

ARTICLE INFO

Keywords:
Thick cloud and cloud shadow
Spatio-temporal
Gap-filling
Patch group
Global-local CNN
Progressive iteration

ABSTRACT

Thick cloud and its shadow severely reduce the data usability of optical satellite remote sensing data. Although many approaches have been presented for cloud and cloud shadow removal, most of these approaches are still inadequate in terms of dealing with the following three issues: (1) thick cloud cover with large-scale areas, (2) all the temporal images included cloud or shadow, and (3) deficient utilization of only single temporal images. A novel spatio-temporal patch group deep learning framework for gap-filling through multiple temporal cloudy images is proposed to overcome these issues. The global-local loss function is presented to optimize the training model through cloud-covered and free regions, considering both the global consistency and local particularity. In addition, weighted aggregation and progressive iteration are utilized for reconstructing the holistic results. A series of simulated and real experiments are then performed to validate the effectiveness of the proposed method. Especially on Sentinel-2 MSI and Landsat-8 OLI with single/multitemporal images, under small/large scale regions, respectively.

1. Introduction

Satellite remote sensing has gradually become important in earth observation tasks in recent years. Such as land-use change (Toure et al., 2018), vegetation mapping and estimation (Erinjery et al., 2018), land surface temperature retrieval (Weng and Fu, 2014), and environmental pollution monitoring (Peng et al., 2016). However, the imagery acquired by satellite sensors is inevitably contaminated by thick cloud and its shadow (Cheng et al., 2014). This greatly reduces the data usability (Qiu et al., 2017; Chen et al., 2019a). Clouds are generated by vast floating water droplets, and cloud shadows are formed by optical linear propagation covering. Therefore, authentic reflectivity information is destroyed within the covered areas. Thus, reconstructing cloud-free and shadow-free images through cloud and cloud shadow removal methods is essential (Zhu and Woodcock, 2012b; Li et al., 2015, 2019c).

A variety of cloud removal algorithms have been presented over the last two decades. Given that cloud and cloud shadow removal is a process of missing information reconstruction (Wang et al., 2019), cloud and cloud shadow removal approaches can be separated into four main types, according to the type of dependent information (Shen et al., 2015). (1) Spatial information based methods; (2) spectral information based methods; (3) temporal information based methods; and (4) hybrid methods. Details of these methods are provided in the following.

1) Spatial information based methods: This type of method is the most fundamental strategy for cloud and cloud shadow removal. Regions without cloud or cloud shadow have similar contextual information between their adjacent cloudy regions (Guillemot and Olivier, 2014). Meanwhile, the spatial relationship between the local and non-local regions can also be incorporated in the cloud and cloud shadow removal procedure (Van der Meer, 2012). The spatial information based methods include interpolation-based methods (Rossi et al., 1994), exemplar-based methods (Criminisi el al., 2004; He and Sun, 2014), spreading or diffusion based methods (Bertalmio et al., 2003), and total variational methods (Chan, 2001). However, the application ranges of these approaches are often severely limited by the similar contextual structure or the small size of cloud areas. Therefore, the spatial information based approaches are typically just suitable for removing small clouds with regular texture (Ng et al., 2017). For large-scale cloud cover or complicated regions, the reconstruction accuracy and reliability are usually not satisfactory.

---

* Corresponding authors at: School of Geodesy and Geomatics, Wuhan University, Wuhan, China. State Key Laboratory of Information Engineering, Survey Mapping and Remote Sensing, Wuhan University, Wuhan, China.
E-mail addresses: yqiang86@gmail.com (Q. Yuan), zlp62@whu.edu.cn (L. Zhang).

https://doi.org/10.1016/j.isprsjprs.2020.02.008
Received 29 August 2019; Received in revised form 13 February 2020; Accepted 13 February 2020

2) Spectral information based methods: Different from the spatial information based methods, the spectral information based methods are addressing a different issue such as sensor failures or thin clouds (Lv et al., 2016). For multispectral or hyperspectral data, the close relevance and similarity within the various spectral bands can be fully utilized (Xu et al., 2019). It provides the possibility of recovering degraded regions by relying on the available complementary information. For example, the Aqua MODIS band 6 contaminated by deadlines, can be reconstructed through the closely correlated spectral bands (Rakwatin et al., 2009). The thin cloud in multispectral data can also be removed through spectral information (Shen et al., 2014). However, spectral information-based methods can't solve the shortcomings of spatial-based methods for thick cloud.

3) Temporal information based methods: This type methods have gradually been put forward for thick cloud removal in recent years (Chen et al., 2019b). Because satellites can revisit the same location at different times, the obtained multi-temporal data includes complementary information. This can be employed to reconstruct thick cloud covered areas (Xu et al., 2016; Zhang, et al., 2019b). The temporal based methods assume that the category and geometric position of ground objects rarely show changes over short time intervals (Gao and Gu, 2017). From this point of view, an effective missing data reconstruction algorithm named the neighborhood similar pixel interpolator (NSPI) (Chen et al., 2011) was put forward for the Landsat Enhanced Thematic Mapper Plus (ETM+) scan-line corrector (SLC) malfunction problem. Subsequently, Zeng et al. (2013) presented a weighted linear regression (WLR) model for filling the gaps of incomplete ETM+ data. Chen et al. (2017) combined the spatio-temporal dimension characteristics to remove cloud in Landsat-8 OLI data. In conclusion, despite the favorable performance on thick cloud removal, the large-scale area reconstruction problem is still a major obstacle. Furthermore, the temporal differences between multitemporal images may introduce negative effects for gap-filling, such as spectral reflectance change (Zhu et al., 2012a; Wang et al., 2016; Li et al., 2019a).

4) Hybrid based methods: These methods utilize both spatial, temporal, and spectral information for cloud removal. For instance, Li et al. (2014) developed a temporal group sparse learning algorithm with multi-temporal and multi-spectral images. Ji et al. (2018) considered multiple temporal images as a four-order tensor with spatial, temporal, and spectral dimensions. Then they reconstructed cloud covered areas using the low-rank property and non-local complementary information.

Although many methods have been widely applied to thick cloud removal, the following limitations still exist in actual scenes.

(1) This inverse problem has the seriously ill-posed degree, where large-scale thick cloud and cloud shadow cover cannot be well reconstructed by most methods.
(2) Almost all temporal images are unavoidably influenced by cloud and cloud shadow. However, most cloud removal methods usually assume that complementary data must be cloud-free.
(3) Theoretically, redundant temporal information (information from images acquired at the same areas but in different dates) leads to more complementary information. (Di Mauro et al., 2017; Pelletier et al., 2019). However, most existing methods can only utilize single temporal images rather than multitemporal images.

Recently, on account of the powerful feature extraction and expression ability, deep learning (LeCun et al., 2015) has been applied for many fields. For remote sensing data quality improvement tasks, several solutions have been provided through a data-driven learning framework (Zhang et al., 2017). Such as SAR image despeckling (Zhang et al., 2018a), hyperspectral image denoising (Zhang et al., 2019a), and pansharpening (Wei et al., 2017; Yuan et al., 2018; Shen et al., 2019). These methods can achieve state-of-the-art reconstruction effects.

In terms of missing information reconstruction such as thick cloud removal, we developed an integrated spatial-temporal-spectral (STS) feature learning model to solve missing data reconstruction (Zhang et al., 2018a,b). However, STS cannot effectively reconstruct large-scale areas due to several reasons. First, STS only deals with global information and does not consider the local particularity between missing regions and their neighboring regions. Second, STS can only utilize single temporal images. Lastly, STS can only employ the cloud-free image as complementary data. However, most temporal images are still influenced by cloud. These disadvantages reduce the effectiveness of STS approach in cloud and cloud shadow removal.

To overcome these limitations, we propose a novel framework for thick cloud and shadow removal. Based on STS, the proposed framework combines the global-local spatio-temporal information in remote sensing imagery with the non-linear learning capability of deep neural network. The main contributions are as follows.

(1) A progressive spatio-temporal patch group learning framework for thick cloud and cloud shadow removal is proposed. Arbitrary numbers of temporal images can be employed to remove cloud and cloud shadow through this framework. Furthermore, the integrity of the temporal images, referring to the non-missing value need not to be guaranteed, compared with STS and other methods.
(2) A global-local DCNN is presented to recover cloud and cloud shadow-covered regions. Considering the global consistency and local particularity, the proposed method applies a global-local loss function in the supervised learning procedure. In addition, weighted aggregation and progressive iteration are utilized to reconstruct the final cloud removal result in the proposed framework.
(3) A series of experiments are performed to validate the applicability of the proposed method with Sentinel-2 MSI and Landsat 8 OLI using single/multi-temporal images of small/large-scale scenes. The quantitative evaluation and visual effects of the proposed method are verified in both simulated and real experiments.

The rest of this paper is arranged as follows. Section 2 describes the methodology for removing cloud and cloud shadow using spatio-temporal information. Section 3 provides the experimental results of Sentinel-2 MSI and Landsat 8 OLI data. Section 4 summarizes the conclusions and expectations of the proposed work.

2. Methodology

The flowchart of the proposed method is displayed in Fig. 1. First, cloud and shadow detection of the multitemporal data are executed to obtain the mask data. Second, the spatial patch and its corresponding multitemporal patches are stacked and sorted in a patch group fashion. Third, a spatiotemporal patch group recovery model is developed to reconstruct the information of the cloud-covered areas. Finally, weighted aggregation is applied to all the reconstructed patches included overlapping regions. Then the spatial data and cloud mask are regenerated through progressive iteration until all the cloud-covered areas are recovered. The details of each part are provided in Sections 2.1–2.4.

2.1. Multi-temporal imagery cloud and cloud shadow detection

Before the reconstruction procedure, obtaining accurate location information for the cloud and shadow is a necessary operation (Li et al., 2017; Baetens et al., 2019; Zhong et al., 2019). Therefore, we apply the Fmask (Zhu et al., 2015) and MSCFF (Li et al., 2019b) method to obtain cloud and shadow masks. In the generated binary mask, 0 and 1 respectively represent the undamaged pixel, and the cloud or shadow

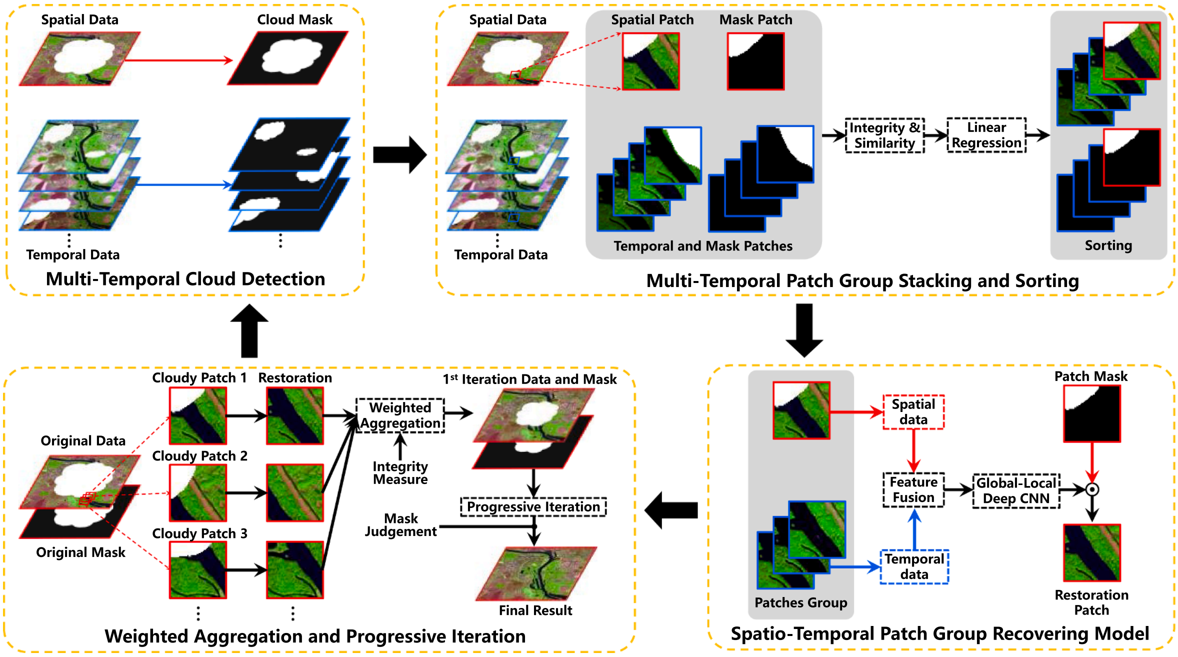


Fig. 1. Flowchart of the proposed spatio-temporal cloud and cloud shadow removal framework.

pixel. In addition, cloud and shadow edge dilation of two pixels is implemented to decrease the cloud and shadow detection errors. The final masks of the multi-temporal data are then used to the next stage of the workflow, as described in Section 2.2.

2.2. *Multi-temporal patch group stacking and sorting*

The purpose of this section is to establish multi-temporal patch group. These patch groups include complementary information in series images, as shown in Fig. 2. Notably, clouds and cloud shadows are non-uniformly distributed in global areas. These complementary regions have the same or similar texture information, which can be employed for reconstructing cloud regions.

Therefore, the patch containing cloud or shadow is extracted through image traversal. If the currently selected patch is cloud-free, the window with the fixed stride is sequentially shifted until the end. Then, temporal patch sorting is executed to find the most similar information. This operation can introduce reliable temporal information rather than avoid polluted and invalid information. It can also reduce model complexity for rapid and improved optimization and learning.

After obtaining the masks of multitemporal data, the spatial cloudy patch $\mathbf{P}_S$ with size $W \times H$ is selected as the current reconstruction object through image traversal. $\mathbf{P}_S(w, h)$ stands for the reflectance value of pixel at location (w, h). Notably, cloudy and intact regions simultaneously exist in the selected spatial patch. Therefore, the integrity ratio corresponding to $\mathbf{P}_S$ is defined as I_S in the proposed model. It is calculated by the corresponding cloud masks. I_T^t and similarity correlation measurements C_{S-T}^t for the corresponding multi-temporal patches $\mathbf{P}_T^t$ are also evaluated. t stands for the *t-th* temporal patch. I_T^t is the undamaged area ratio of the *t-th* temporal patch through its mask $\mathbf{M}_T^t$. C_{S-T}^t and I_T^t are calculated below:

$$C_{S-T}^t = \frac{\sum_{w=1}^{W}\sum_{h=1}^{H}(\mathbf{V}_S - mean(\mathbf{V}_S))\cdot(\mathbf{V}_T^t - mean(\mathbf{V}_T^t))}{\sqrt{\sum_{w=1}^{W}\sum_{h=1}^{H}(\mathbf{V}_S - mean(\mathbf{V}_S))^2\cdot\sum_{w=1}^{W}\sum_{h=1}^{H}(\mathbf{V}_T^t - mean(\mathbf{V}_T^t))^2}} \tag{1}$$

$$I_T^t = 1 - \frac{\sum \mathbf{M}_T^t}{W \times H} \tag{2}$$

where $\mathbf{V}_S$ and $\mathbf{V}_T^t$ are the same cloud-free areas in $\mathbf{P}_S$ and $\mathbf{P}_T^t$, respectively.

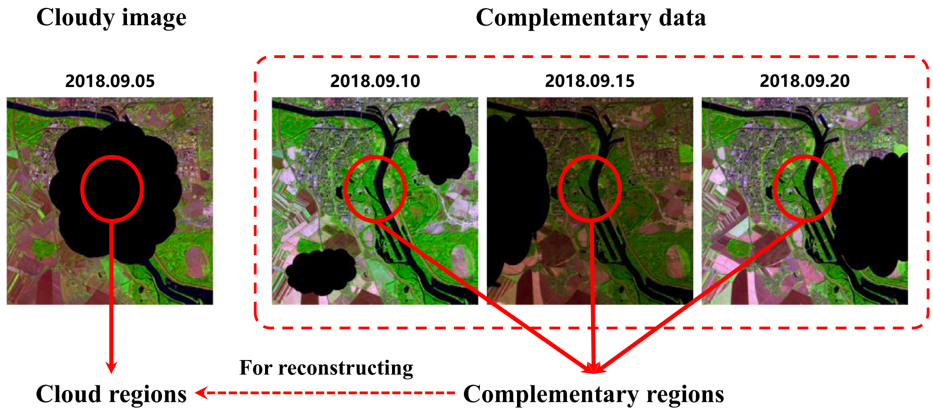


Fig. 2. Schematic diagram of complementary information.

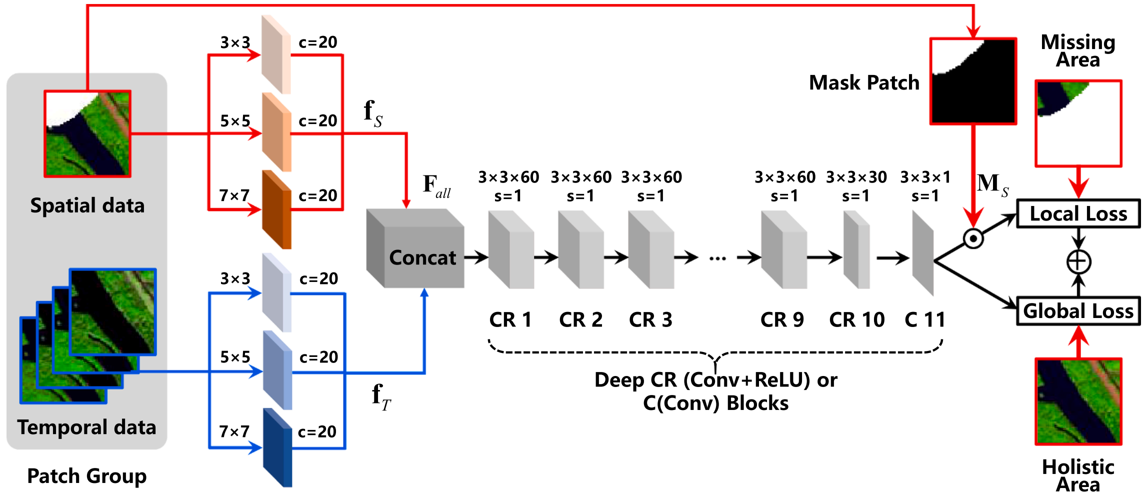


Fig. 3. The spatio-temporal patch group recovery model.

This step only focuses on the mutual undamaged regions between the spatial and temporal patches to estimate the similarity for the subsequent processing. The re-ranked sequence of the patch group $\mathbf{P}_T^t$ is then sorted by from high to low. Furthermore, the proposed method removes these temporal patches with low I_T^t values, because part of the multitemporal patches is potentially contaminated by cloud. The selection number is set as a fixed value. And the top temporal patches are selected while the other temporal patches with low integrity ratio values are eliminated. The sensitivity analysis of the hyperparameters is discussed in Section 3.4.

After patch group ranking and elimination, the sorted temporal patches are selected as the complementary data for the following recovery model. All the sorted temporal patches are transformed through linear regression. This operation can further reduce model complexity and achieve effective optimization and learning.

$$\mathbf{P}_T^{t'} = a\cdot\mathbf{P}_T^t + b \tag{3}$$

Parameters a and b are determined by least-squares approximation.

$$a = \frac{\mathbf{V}_S\bar{\mathbf{V}}_T^t - \bar{\mathbf{V}}_S\cdot\bar{\mathbf{V}}_T^t}{(\bar{\mathbf{V}}_T^t)^2 - \bar{\mathbf{V}}_T^{t\,2}} \tag{4}$$

$$b = \bar{\mathbf{V}}_S - a\cdot\bar{\mathbf{V}}_T^t \tag{5}$$

2.3. Spatio-temporal patch group recovery model

According to STS method (Zhang et al., 2018a,b), a novel learning model for cloud and cloud shadow removal is developed. The entire structure of the recovery model is depicted in Fig. 3. The patch group in Section 2.2 is simultaneously set as the input data for the spatial and corresponding temporal patches. The structure and modules of the proposed model are also displayed in Fig. 3.

2.3.1. Spatiotemporal multi-scale feature extraction and fusion

A crucial characteristic representation usually counts on multi-scale semantic and context to reconstruct cloudy areas. From this intrinsic perspective, a spatiotemporal multiscale feature block is incorporated in the reconstruction network. This model excavates deep characteristics to provide multi-context relevance. Furthermore, multiscale convolutional filters can obtain diverse receptive field scales, especially for the scenarios including cloud-covered areas. As portrayed in Fig. 3, the multiscale convolutional block comprises multiple convolutional processing layers with a size of 3, 5, and 7 kernel filters. They respectively dispose of the spatial and temporal data as follows:

$$\mathbf{f}_S^k = \mathbf{W}_S^k * \mathbf{P}_S + \mathbf{b}_S^k \tag{6}$$

$$\mathbf{f}_T^k = \mathbf{W}_T^k * \mathbf{P}_T' + \mathbf{b}_T^k \tag{7}$$

where $\mathbf{W}_S^k$, $\mathbf{W}_T^k$, $\mathbf{b}_S^k$, and $\mathbf{b}_T^k$ are the learnable parameters in the convolutional filters for the spatial and temporal data, and k stands for the kernel filter sizes 3, 5, and 7. These convolutional filters are then collaboratively generated as various independent results. The three spatial results of $\mathbf{f}_S^3$, $\mathbf{f}_S^5$, $\mathbf{f}_S^7$, and the three temporal results of $\mathbf{f}_T^3$, $\mathbf{f}_T^5$, $\mathbf{f}_T^7$ are jointly fused as an individual multi-channel characteristic result $\mathbf{F}_{all}$ as follows:

$$\mathbf{F}_{all} = Concat(\mathbf{f}_S^3, \mathbf{f}_S^5, \mathbf{f}_S^7, \mathbf{f}_T^3, \mathbf{f}_T^5, \mathbf{f}_T^7) \tag{8}$$

where Concat is utilized to aggregate the feature maps of the spatial and temporal data with different convolutional kernel sizes. The spatial and temporal features are integrated in union for the sequential operations in the proposed recovery model.

2.3.2. Global-local deep convolutional neural network

After extracting the fused spatiotemporal features, a global-local DCNN is used to recover the cloud and cloud shadow areas in the spatial patch. The proposed model is constructed with 11 layers to mine nonlinear characteristics, as displayed in Fig. 3. All the layers employ 3 × 3 convolutional kernels with a stride equal to 1. These layers comprise convolutional (Conv) and rectified linear unit (ReLU) layers in the 10 front layers and only a convolutional block in the last layer. The numbers of feature maps within the different layers and the block parameters are shown in Fig. 3. In particular, the last convolutional layer in the proposed model outputs a single feature map for the following operations.

In terms of the loss function for model optimization, the Euclidean loss function is utilized in the traditional image restoration tasks, such as super-resolution (Dong et al., 2016) and image denoising (Yuan et al., 2019). However, these methods only consider the integral information for data restoration. They ignore the particularity of the local information, especially for the cloud and cloud shadow removal task. Considering the global consistency and local particularity, the proposed model utilizes a global-local loss function in the supervised learning procedure as follows:

$$\xi(\Theta) = \lambda\cdot\xi_{Local} + \xi_{Global} \tag{9}$$

where Θ stands for the model learning parameters. These parameters are updated in the training procedure through Eq. (9) with a back-

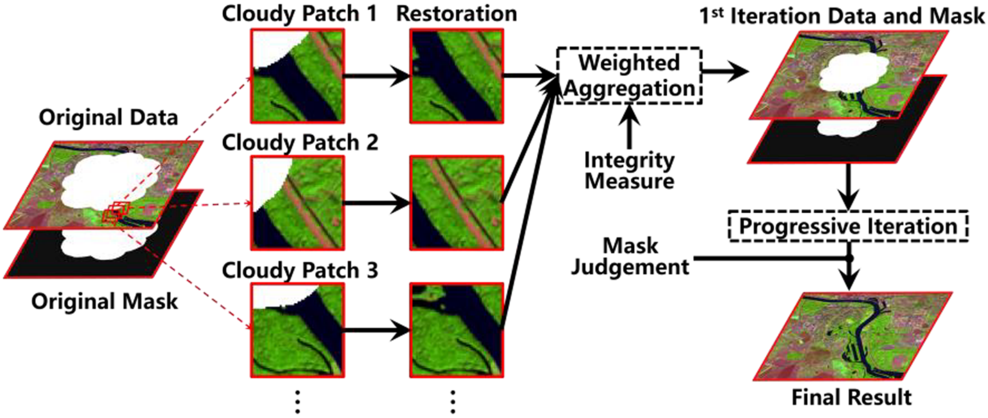


Fig. 4. Weighted aggregation and progressive iteration.

propagation algorithm. λ is the balancing parameter between the local loss ξ_{Local} and global loss ξ_{Global}. They can be respectively defined as follows:

$$\xi_{Local} = ||\mathbf{M}_S \odot \Phi - \mathbf{M}_S \odot \mathbf{P}_S||_F \tag{10}$$

$$\xi_{Global} = ||\Phi - \mathbf{P}_S||_F \tag{11}$$

where $\mathbf{M}_S$ is the corresponding patch mask of $\mathbf{P}_S$. $\odot$ stands for the point multiplication operation between two matrices. Φ represents the feature map of the last layer. The final restored patch $\mathbf{R}_S$ is generated as follows:

$$\mathbf{R}_S = \mathbf{M}_S \odot \Phi + (1 - \mathbf{M}_S) \odot \mathbf{P}_S \tag{12}$$

2.4. Weighted aggregation and progressive iteration

After traversing all available patches of spatial data, overlapping regions always exist in the multiple restored patches. Therefore, weighted aggregation is utilized to reconstruct the final result through these patches, as shown in Fig. 4.

From the aspect of the recovery reliability, the larger the spatial integrity ratio I_S is, the more believable the complementary information is. The operation of weighted aggregation for the first iterative global result $\mathbf{A}_S^1$ is executed as follows:

$$\mathbf{A}_S^1 = \frac{\sum_{i=1}^{\mu} \frac{1}{1 - I_S^i} \cdot \mathbf{R}_S}{\sum_{i=1}^{\mu} \frac{1}{1 - I_S^i}} \tag{13}$$

where μ is the number of the current restored spatial patches with same cloud and cloud shadow regions. The new global spatial mask $\mathbf{A}_{\mathbf{M}_S}^1$ is regenerated after the following aggregation procedure:

$$\mathbf{A}_{\mathbf{M}_S}^1 = \mathbf{A}_{\mathbf{M}_S}^0 \cup \mathbf{M}_S^1 \cup \mathbf{M}_S^2 \cup \cdots \cup \mathbf{M}_S^\eta \tag{14}$$

where η represents the mask patch number. $\mathbf{M}_S^\eta$ stands for the updated mask patch after reconstructing the corresponding spatial patch.

In practice, weighted aggregation cannot remove all the clouds and shadows in a single iteration because of the limited areas. Therefore, a progressive iteration strategy is applied to recover the integrated imagery in the proposed framework.

The original spatial cloudy image $\mathbf{A}_S^0$, the original spatial mask $\mathbf{A}_{\mathbf{M}_S}^0$, the temporal images $\mathbf{A}_T^t$, and the masks $\mathbf{A}_{\mathbf{M}_T}^t$ are integrally set as the input data. By traversing $\mathbf{A}_{\mathbf{M}_S}^{iter}$ under patch mode, the patch group is sorted according to Section 2.2. The current spatial patch is then reconstructed using the model presented in Section 2.3. By progressively reconstructing the spatial data in Eq. (13), the new spatial mask is iteratively regenerated according to Eq. (14) until the final result is obtained. The workflow of the progressive iteration algorithm is shown below.

Progressive iteration algorithm for cloud and cloud shadow removal

```
Input: Cloudy data $\mathbf{A}_S^0$, mask $\mathbf{A}_{\mathbf{M}_S}^0$, temporal data $\mathbf{A}_T^t$, masks $\mathbf{A}_{\mathbf{M}_T}^t$, threshold $I_S'$
Initialization: $iter = 0$, $\mathbf{A}_S^{iter} = \mathbf{A}_S^0$, $\mathbf{A}_{\mathbf{M}_S}^{iter} = \mathbf{A}_{\mathbf{M}_S}^0$
Iteration: until $\mathbf{A}_{\mathbf{M}_S}^{iter}$ is cloud-free
  Traversal $\mathbf{A}_{\mathbf{M}_S}^{iter}$ through global window traversal
    If $I_S = 1$ or $I_S < I_S'$ Continue;
    Else
      Sorting the spatio-temporal patch group in Section 2.2
      Reconstructing the current spatial patch in Section 2.3
    End spatial patch reconstruction
  End Traversal
  $iter = iter + 1$
  Update $\mathbf{A}_S^{iter}$ and $\mathbf{A}_{\mathbf{M}_S}^{iter}$ through Eqs. (13) and (14)
End Iteration
Output: The final cloud-free result $\mathbf{A}_S^{iter}$
```

3. Experimental results

The simulated and real experiments were presented to verify the effectiveness of the proposed framework. Sentinel-2 MSI (20 m spatial-resolution, five-day temporal resolution, L1C level product) and Landsat 8 OLI data (30 m spatial resolution, 16-day temporal resolution, L1 level product) were employed in this work. Small/large-scale scenarios and single/multitemporal images were also employed in the experiments. The source code of this work is available at https://github.com/WHUQZhang/PSTCR.

3.1. Model training and parameter setting

We organize the detailed information of the proposed method: (1) Training data descriptions; (2) Hyperparameter settings; and (3) Operating environment.

1) Training data descriptions: For the training data of the proposed model, a sequence of seven different cloud-free Landsat Thematic

Table 1
Comparison of the spectral central wavelengths (nm) of Sentinel-2 and Landsat-8.

Data	Blue	Green	Red	NIR
Sentinel-2	490	560	665	842
Landsat-8	480	560	655	865

Mapper (TM) images from October 25, 2004 to February 14, 2005 (image size 2720 × 3200, blue, green, red, and NIR bands, 16-day revisit period) was used. These data organized into temporal groups. The training data for the seven temporal Landsat images mainly cover the area of Darlington Point in Australia (34°36́–35°3́S, 145°40́–146°2́E). Band normalization (min–max scaling per band) was applied for the training and test data. The training data included different spectra from the blue, green, red, and NIR bands. This can ensure the spectral generalization capability of the proposed method. The comparison of the spectral central wavelengths of Sentinel-2 and Landsat-8 are listed in Table 1.

Need for training in deep learning-based methods (STS-CNN and the proposed method) is that time-series images are cloud-free. Then the synthetic cloud masks were added into the original cloud-free data as the cloudy data for calculating training loss. Besides, the need for testing in state-of-the-art methods must obtain at least one cloud-free image; And the need for testing in the proposed method can utilize multiple time-series images.

Temporal patch groups were established by traversing the temporal groups with the 21,760 patches, size of 40 × 40. Using the augmentation data strategy (multi-angle data spin and multiscale resolution), 174,080 patches are finally generated as the training samples. The point multiplication operation was performed for generating training data. The simulated mask data were generated in the Landsat 8 OLI data with cloud and cloud shadow using Fmask 4. Some parts were used as the training patch mask.

2) Hyperparameter settings: The related hyperparameters in the proposed method are given in the following. The similarity correlation measurement C_{S-T}^{t} was also set as 0.85 for sorting the patch group. The integrity threshold I_S^{t} is defined as 0.3 for all the spatial patches. The balancing parameter λ was determined as 0.15 for the global-local loss function. The temporal number t was set as 4 for the training and test procedures. The number of trainable parameters in the deep learning network is about 1488 K. The proposed framework was adaptively trained through the Adam optimization approach (Kingma and Ba, 2014), with fixed hyperparameters of 0.9, 0.999, and 10^{-8}. In addition, the learning rate was varied. It started from 0.001 with an attenuation coefficient of 0.8 at intervals of 20 epochs. The entire training stage of the proposed model was set as 200 epochs.

3) Operating environment: The Caffe structure (Jia et al., 2014) was employed to generate the final training model on a Dell T7910 workstation. The proposed method executed in Windows 10 under GPU acceleration mode. The time taken to train the proposed method was about 21 h 20 min.

3.2. Simulated data experiments

In the simulated experiments, the practicability of the proposed framework under various conditions (single/multitemporal and large/small scales of missing areas) was verified. Special scenarios were simulated by the three cases below:

(1) ***Case 1 (large missing areas, with single-temporal data)***: The spatial image was corrupted by large missing areas. This case employed a single-temporal image as the complementary information, as displayed in Fig. 5(a)–(c). The original data and temporal data (20 m spatial-resolution, 10 × 10 km) were acquired by the Sentinel-2 MSI sensor on July 19 and July 29, 2018, respectively, over Wuhan in China.

(2) ***Case 2 (multiple small missing areas, with single-temporal data):*** The spatial image was corrupted by multiple small missing areas. This case employed a single-temporal image as the complementary information, as displayed in Fig. 6(a)–(c). The original data and temporal data (30 m spatial-resolution, 15 × 15 km) were acquired by the Landsat 8 OLI sensor on April 11 and April 27, 2014, in Colorado, U.S.

(3) ***Case 3 (large/small missing areas, with multi-temporal data):*** Sequential remote sensing images were corrupted by various types of missing areas. This case employed multiple temporal images as the complementary information, as displayed in Fig. 8(a)–(e). Five sequential images (20 m spatial-resolution, 10 × 10 km) were acquired by the Sentinel-2 MSI on September 5, 15, 20, October 5, and 15, 2018, in Belgium.

The proposed framework was evaluated with mNSPI (Zhu et al., 2012a), WLR (Zeng et al., 2013), AWTC (Ng, et al., 2017), and STS (Zhang et al., 2018b). The links to source codes for the competitors are also displayed below:

(1) mNSPI: https://xiaolinzhu.weebly.com/.
(2) WLR: http://sendimage.whu.edu.cn/send-resource-download/.
(3) AWTC: https://sites.google.com/site/tengyuji90/.
(4) STS: https://github.com/WHUQZhang/STS-CNN.

The quantitative evaluation indexes of the correlation coefficient (CC), structural similarity index measure (SSIM), and root mean square error (RMSE), were all used to assess the recovery effects of the different algorithms. The evaluation indexes for Cases 1 and 2 are listed in Table 2. Fig. 5(d)–(h) show the results of Case 1 in the five methods for Sentinel-2 MSI data. Fig. 6(d)–(h) show the results of Case 2 in the five methods for Landsat 8 OLI data. Fig. 7 displays the scatter plots of the reconstructed regions for the different algorithms. Fig. 8 presents the results of Case 3 for the multiple sequence data.

In Cases 1 and 2, the regions reconstructed by mNSPI, AWTC, STS, and WLR contain some spectral distortion, as indicated in Figs. 5 and 6. Furthermore, the restored areas within the red circles through WLR exhibit a lack of contextual continuity under Case 2. The main reason may be possibly the intricate and non-linear temporal relations. Thus, these relations cannot be clearly expressed using a linear regression strategy. mNSPI and STS demonstrated poor performance under Case 1. Both approaches exhibited vague details and content differences around the marginal regions of the reconstruction results under Case 2. These results indicate that most approaches cannot fit the complex relations for the reconstruction of large cloud and cloud shadow-covered areas. However, in both cases, the proposed framework showed satisfactory performance in preserving spectral consistency, as displayed in the reconstructed images in Fig. 5(h) and the red circles in Fig. 6(h). The CC, SSIM, and RMSE values listed in Table 2 also suggest that the proposed method outperforms mNSPI, WLR, and STS. The scatter diagrams and their CCs in both cases also imply the reconstruction precision of the proposed methods, as presented in Fig. 7.

In Case 3, we consider that the multitemporal sequence of remote sensing images is almost totally disturbed by cloud and its shadow. The situation is simulated to further test the effectiveness in a challenging scenario (top row of Fig. 8). Unlike Cases 1 and 2, Case 3 involved five temporal images as the reconstruction objects. All of them are contaminated with cloud and cloud shadow. A distinct advantage of the proposed method is that the reconstruction framework can completely utilize the available and highly correlated spatiotemporal information. By contrast, mNSPI, STS, WLR, and most other cloud removal methods can only be employed for single temporal imagery. This condition is rarely the case for actual optical Earth observation systems. Multitemporal sequential images were simulated with cloud cover (top row of Fig. 8) to deal with this actual scenario. The proposed method can simultaneously recover multiple incomplete images. It can also preserve the semantic context and acquire credible results (middle row of Fig. 8). In addition, the scatter diagrams and their CCs verify the reconstruction accuracy of the proposed framework (bottom row of Fig. 8).

3.3. Real-data experiments

Three cloudy data sets presented in Figs. 9, 10, and 12 were

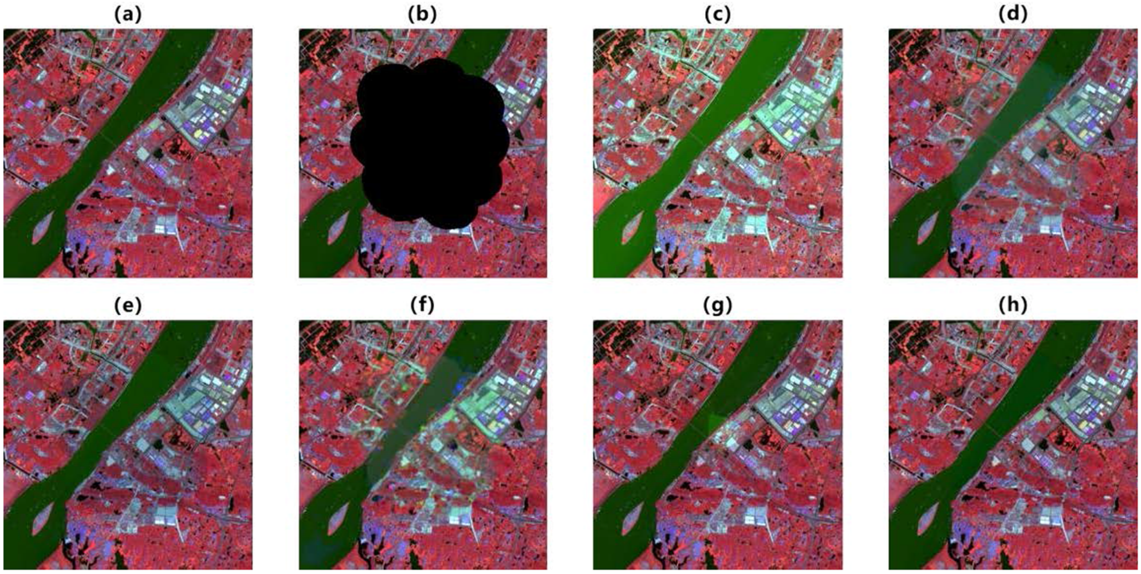


Fig. 5. Case 1. (20 m spatial-resolution; 10 × 10 km; single-temporal data; pseudo-color: B05, B8A, B12 in Sentinel-2 MSI; from Wuhan, China.) (a) Original data from July 19, 2018. (b) Simulated cloud-covered area. (c) Data from July 29, 2018. (d)–(h) Recovery results of mNSPI, AWTC, STS, WLR, and the proposed method.

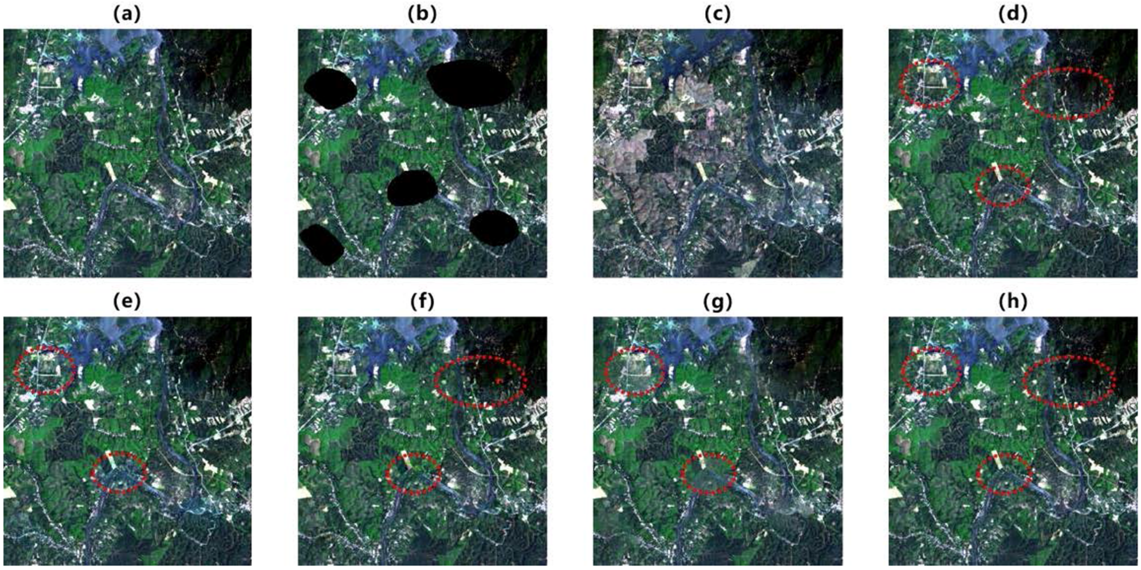


Fig. 6. Case 2. (30 m spatial-resolution; 15 × 15 km; single-temporal data; true-color: red, green, and blue bands in Landsat 8 OLI; from Colorado, U.S.) (a) Original data from April 11, 2014. (b) Simulated cloud cover. (c) Data from April 27, 2014. (d)–(g) Results of mNSPI, AWTC, STS, WLR, and the proposed method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Quantitative evaluation indexes in the simulated experiments of Case 1 and 2.

Index	Band		Case 1: Sentinel-2 MSI in Fig. 5					Case 2: Landsat-8 OLI in Fig. 6				
		mNSPI	AWTC	WLR	STS	Ours		mNSPI	AWTC	WLR	STS	Ours
CC	B05	0.937	0.947	0.958	0.915	**0.981**	Blue	0.973	0.988	0.985	0.986	**0.993**
	B8A	0.964	0.965	0.963	0.962	**0.991**	Green	0.988	0.990	0.991	0.986	**0.996**
	B12	0.968	0.967	0.962	0.940	**0.991**	Red	0.974	0.987	0.981	0.989	**0.991**
	Mean	0.957	0.960	0.961	0.939	**0.988**	NIR	0.965	0.969	0.974	0.977	**0.982**
SSIM	B05	0.826	0.889	0.905	0.887	**0.959**	Blue	0.935	0.967	0.959	0.963	**0.979**
	B8A	0.815	0.874	0.918	0.856	**0.933**	Green	0.948	0.975	0.974	0.983	**0.988**
	B12	0.798	0.836	0.900	0.805	**0.930**	Red	0.954	0.962	0.952	0.958	**0.976**
	Mean	0.813	0.866	0.908	0.849	**0.941**	NIR	0.947	0.958	0.956	0.961	**0.968**
RMSE	B05	0.067	0.059	0.057	0.072	**0.028**	Blue	0.044	0.029	0.032	0.031	**0.022**
	B8A	0.059	0.054	0.047	0.065	**0.032**	Green	0.032	0.022	0.025	0.021	**0.020**
	B12	0.066	0.061	0.047	0.078	**0.029**	Red	0.038	0.033	0.037	0.037	**0.021**
	Mean	0.064	0.058	0.051	0.072	**0.029**	NIR	0.042	0.041	0.039	0.036	**0.031**

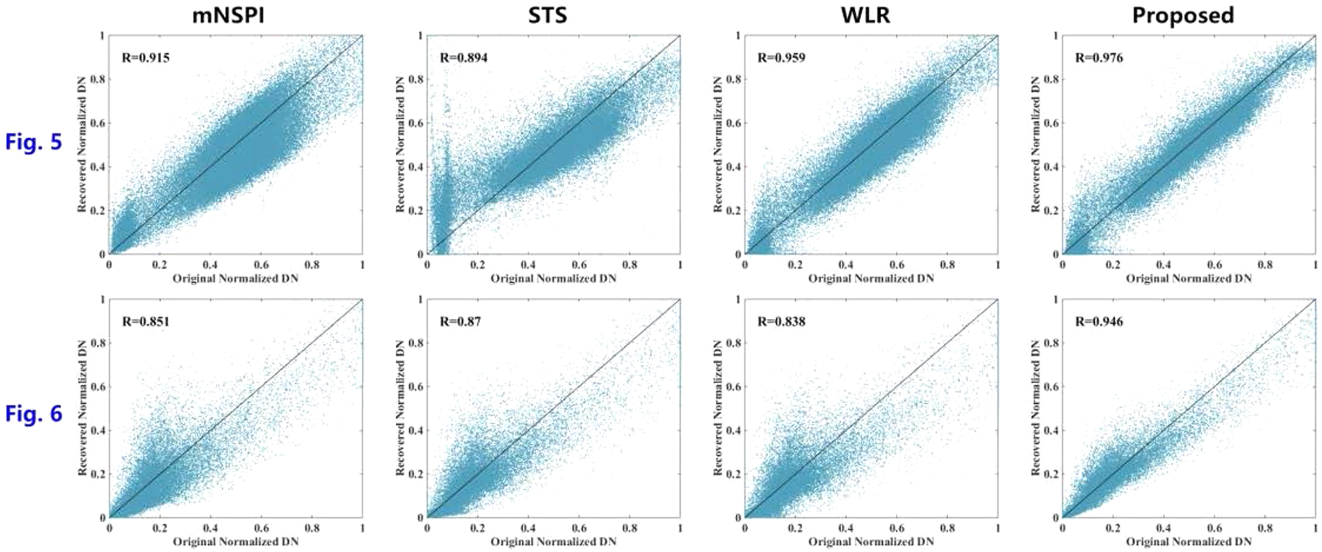


Fig. 7. Scatter diagrams in reconstructed regions of mNSPI, STS, WLR, and the proposed method.

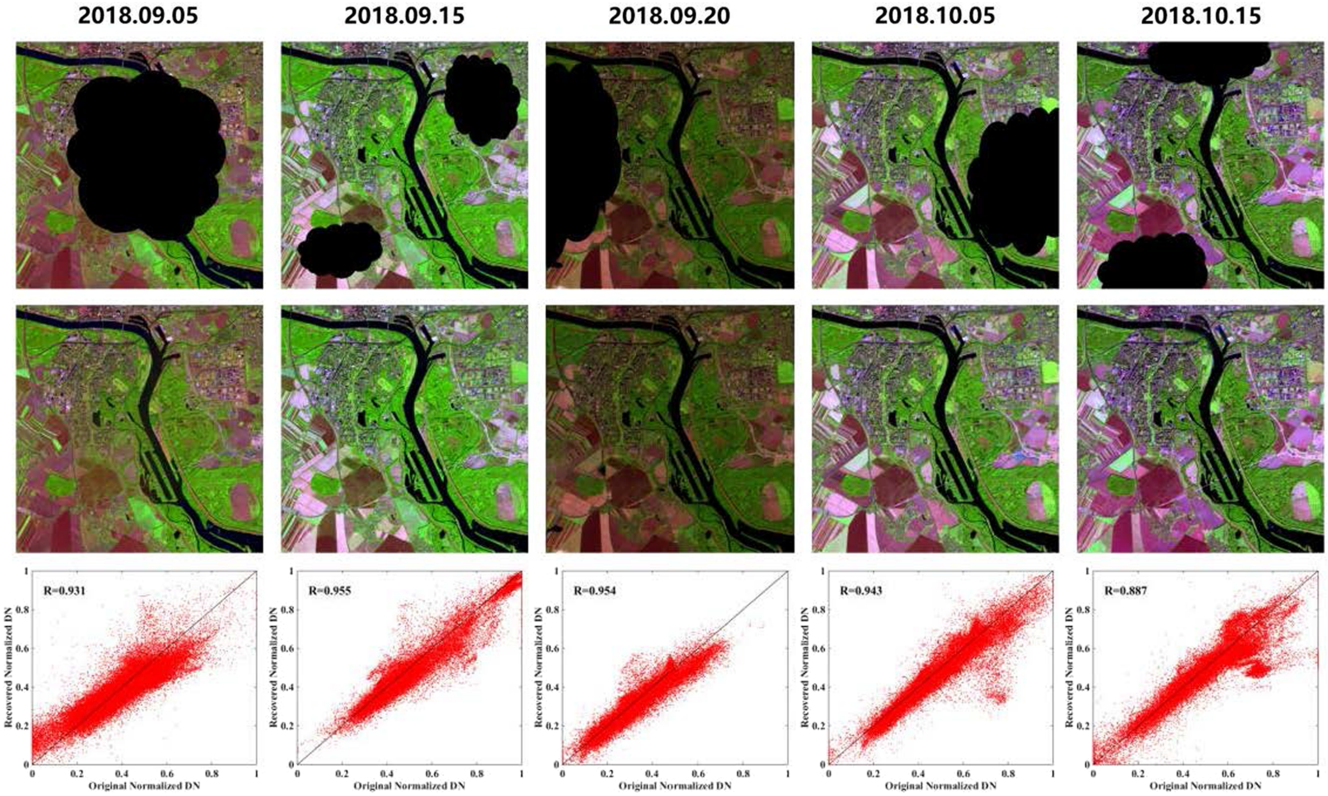


Fig. 8. Case 3: Sentinel-2 MSI multi-temporal data sequence cloud removal experiments (20 m spatial-resolution; 10 × 10 km; five temporal images) over Mechelen in Belgium. Top row: simulated data; middle row: reconstruction results; bottom row: scatter plots.

employed in the real-data experiments. The Sentinel-2 MSI and Landsat 8 OLI cloud removal results through single corresponding temporal image are respectively shown in Figs. 9 and 10. Fig. 12 shows the multitemporal image sequence cloud removal results for the large-scale full Sentinel-2 scenario.

1) Sentinel-2 MSI/Landsat 8 OLI cloud and shadow removal: Fig. 9 depicts the Sentinel-2 real-data cloud and shadow removal experiments (10 × 10 km, 20 m spatial resolution, single temporal data) conducted over Wuhan, China. Fig. 10 depicts the Landsat 8 real-data cloud and shadow removal experiments (24 × 24 km, 30 m spatial resolution, single temporal data) conducted over Colorado, U.S. Temporal difference inevitably exists in Figs. 9 and 10 due to the scene changes and radiation differences. In terms of the LR, mNSPI, and STS methods, spectral differences and blurry details are observed in Figs. 9 and 10. This condition possibly neglects the local particularity. WLR shows fragmentation in Figs. 9(g) and 10(g) due to the intricate and non-linear temporal relations. By contrast, the proposed approach outperforms the other methods. It can suppress spectral differences and ensure the continuity of the texture patterns (such as the rivers in Figs. 9(h) and 10(h)). In Fig. 9, urban areas and rivers are blurry and mixed by WLR, mNSPI, and STS. While the proposed method obtains sharp results in these areas. In Fig. 10(f), the reconstruction results of

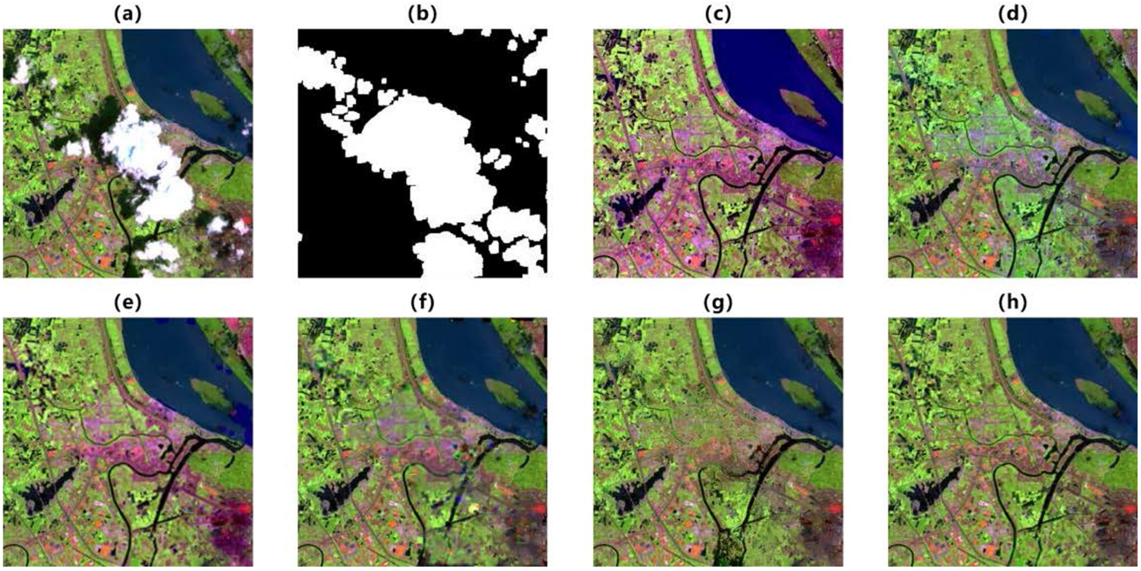


Fig. 9. Sentinel-2 MSI real-data experiments (20 m spatial-resolution; 10 × 10 km, single-temporal data) over Wuhan, China. (a) Original data from July 19, 2018. (b) Cloud and cloud shadow mask of (a). (c) Temporal data from July 19, 2018. (d) LR. (e) mNSPI. (f) STS. (g) WLR. (h) Proposed.

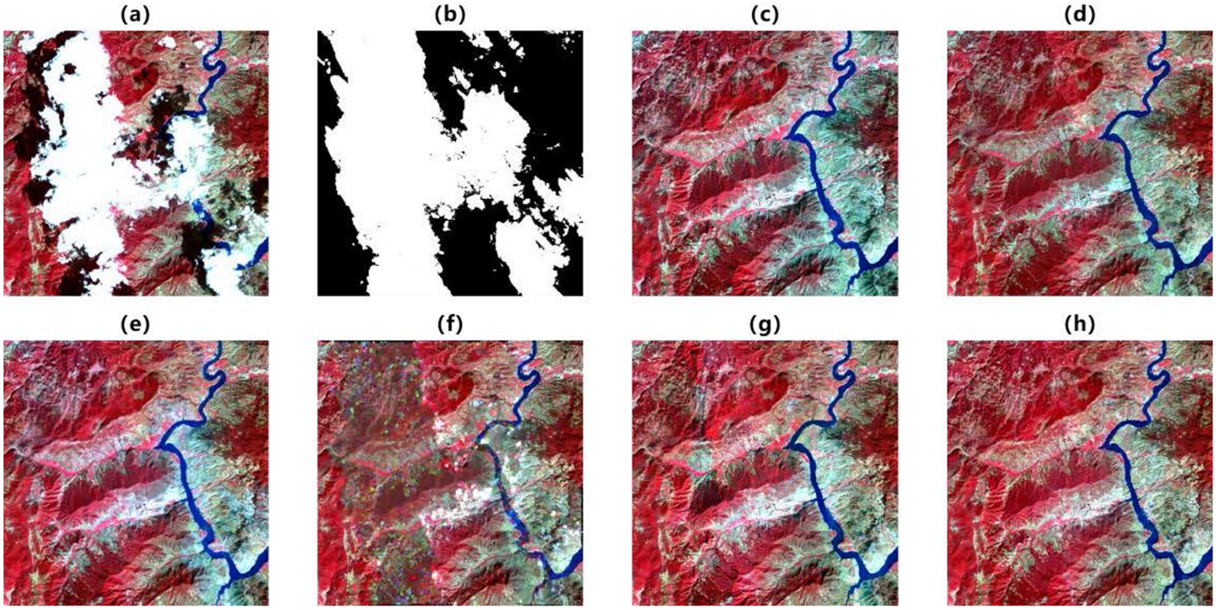


Fig. 10. Landsat-8 OLI real-data experiments (30 m spatial-resolution; 24 × 24 km, single-temporal data) over Colorado, U.S. (a) Original data from Jan 24, 2015. (b) Cloud and cloud shadow Mask of (a). (c) Temporal data from Feb 9, 2015. (d) LR. (e) mNSPI. (f) STS. (g) WLR. (h) Proposed.

STS in the mountainous areas show serious ambiguity and noise distortion. In Fig. 10(e) and (g), the mountain cloud shadow areas contain radiation changes in mNSPI and WLR. The main reason is the complexity and heterogeneity of this mixed scenario.

Runtime comparisons of the different methods for the real-data experiments is shown in Figs. 9 and 10. They are also provided in Table 3 (Operating system: Windows 10, Language: MATLAB R2018a, CPU: Intel E5-2609@1.9 GHz&6Cores, GPU: NVIDA TITAN X, RAM: 16-GB). The proposed method performs the high efficiency due to the superiority of data-driven strategy.

To further verify the effectiveness of results, land-cover classification results in Figs. 9 and 10 were produced for mNSPI, STS, WLR, and the proposed method. The land-cover classification results were acquired with the maximum likelihood supervised classifier (Otukei and Blaschke, 2010). The classification elements comprised the following categories: residential, vegetation, water, bare soil, and rock. The sample regions were manually chosen from the cloud-free regions. The classification results are displayed in Fig. 11. Table 4 lists the overall accuracy (OA) and kappa coefficient values. For the proposed method, the values of OA and kappa are both superior to other algorithms. The reconstructed roads and rivers are clear and continuous in Fig. 11. In comparison, the road classes by WLR are fragmentary. And the reconstructed results of mNSPI and STS contain some misclassified categories. These results verify the importance of the proposed framework

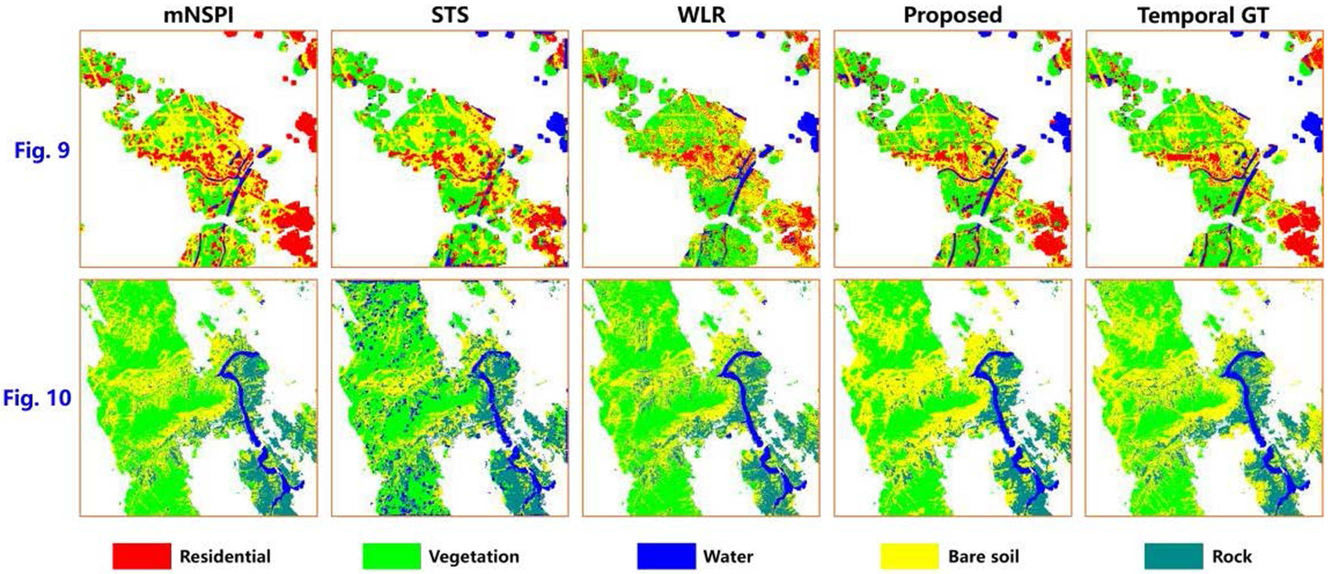


Fig. 11. Classification results of the different methods in the cloud and cloud shadow covered regions.

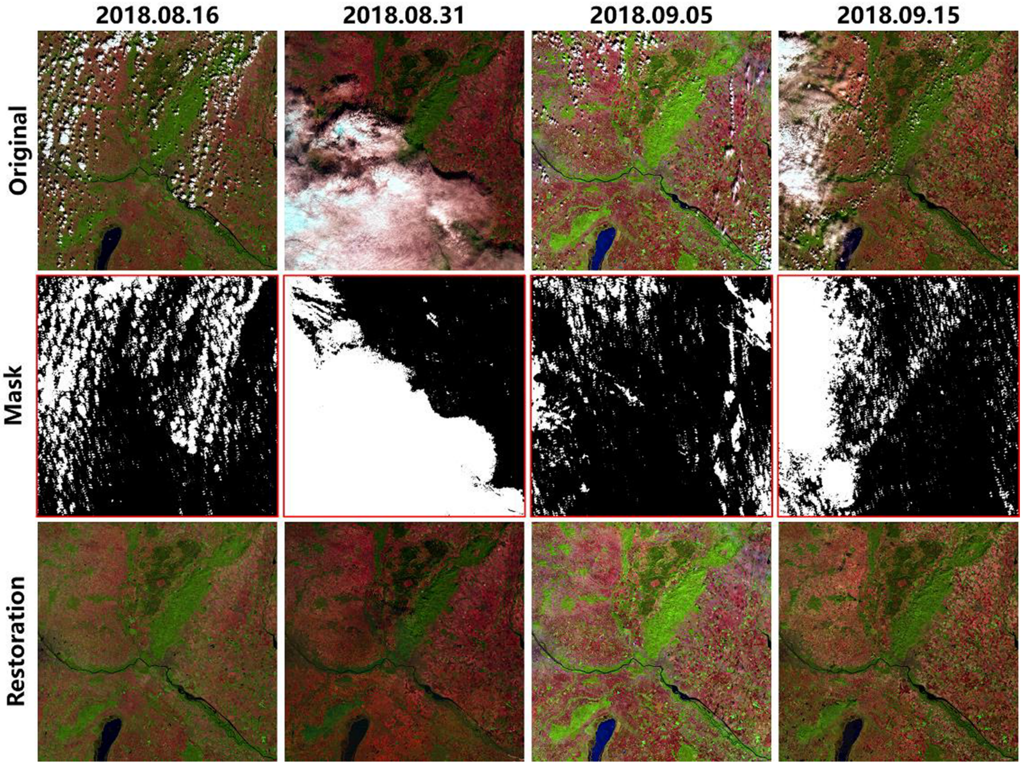


Fig. 12. Sentinel-2 MSI large-scale real-data experiments (20 m spatial-resolution; 5490 × 5490; 110 × 110 km; pseudo-color: B05, B07, B11; four temporal images).

Table 3
Runtime comparison (Second) of the real experiments (Testing/Training time).

Data	LR	mNSPI	STS	WLR	Proposed
Sentinel-2	0.25/-	1.36/-	0.28/23 h	11.73/-	1.16/21 h
Landsat-8	0.41/-	3.27/-	0.49/23 h	25.62/-	6.48/21 h

Table 4
Quantitative classification indexes (OA/Kappa) for Fig. 10.

	mNSPI	STS	WLR	Proposed
Fig. 9	74.4%/0.628	69.5%/0.587	76.8%/0.646	**81.3%/0.695**
Fig. 10	76.8%/0.647	64.5%/0.549	77.3%/0.652	**79.4%/0.683**

for subsequent applications.

2) Multi-temporal sequence and large-scale full scenarios of Sentinel-2 MSI: Four Sentinel-2 MSI images were contaminated by thick/thin cloud and shadow in different degrees, as shown in Fig. 12. The selected sequence images were large-scale with full scenarios. Details of these data are provided in Table 5.

The original images are shown in the top row of Fig. 12. The corresponding masks generated by MSCFF (Li et al., 2019b) are shown in

Table 5
Details of the multi-temporal sequence images used in the real-data experiments.

Sensing date	Satellite sensor	Row/col number	Field size	Cloud cover ratio
2018-08-16	S2B	N0206, R079	110 × 110 km(5490 × 5490)	**14.76%**
2018-08-31	S2A			**51.83%**
2018-09-05	S2B			**10.97%**
2018-09-15	S2B			**27.35%**

the middle row of Fig. 12. The cloud and shadow removal results are displayed in the bottom row of Fig. 12. From the overall visual perspective, the proposed method can synchronously protect semantic contextual information. In addition, the large missing areas caused by cloud and shadow can be well recovered. The global and two local zoomed results of the original and reconstructed results for September 15, 2018, are provided in Fig. 13. These results verify the effectiveness of the proposed cloud removal framework under large-scale, multi-temporal, and large-area reconstruction scenarios.

3.4. Further discussion

3.4.1. Parameter sensitivity analysis

The simulated and real-data experiments verified the cloud and cloud shadow removal capability of the proposed framework. The parameter sensitivity of this model is discussed in this part.

The influence of parameter λ on the reconstruction result is first discussed. Section 2.3 shows that λ plays a key role in the global–local loss function. The proposed method was applied for Cases 1 and 2 described in Section 3.1. For a quantitative assessment, the RMSE values for different values of λ are plotted together in Fig. 14(a). As λ increases, the RMSE initially declines and then increases later in Cases 1 and 2. The RMSE reaches the lowest value when λ is equal to 0.15. This represents the optimal reconstruction accuracy. Fig. 14(a) shows that λ plays an important role in cloud and shadow area reconstruction.

The sensitivity of the temporal parameter t was also analyzed, as shown in Fig. 14(b). In test Case A, the temporal series of images are all cloudless data. By contrast, the temporal series of images are all cloudy data of different degrees in test Case B. As the temporal threshold t increases, the RMSE first slightly increases and then decreases in Case A. While the RMSE quickly decreases and then remains stable in Case B. Considering both cases, the temporal number threshold t as equal to 4 is recommended in the reconstruction procedure.

3.4.2. Linear regression validation

In Section 2.2, we employ linear regression transformation to further reduce model complexity. Therefore, the effectiveness of with/without this operation is then discussed in the proposed framework. In the training and testing procedure, the proposed model can also learn without linear regression transformation. Models with linear regression transformation perform better and more stable in the training and testing procedures, as depicted in Fig. 15. Due to the temporal atmosphere variation, the radiation difference is observed in in Fig. 12. Global radiation difference may introduce negative effects for optimizing the training model. Therefore, linear regression transformation can reduce this negative effect, which performs better than models without linear regression transformation.

4. Conclusions

This study proposed a progressive spatio-temporal patch group learning framework for cloud and cloud shadow removal. The global-local loss function is presented to optimize the training model through cloud-covered and free regions, considering both the global consistency and local particularity. In addition, weighted aggregation and progressive iteration are utilized for reconstructing the holistic results. It overcomes some limitations of the existing methods (i.e., large-scale cloud cover, incomplete temporal information, and temporal-series images). The experiments on Sentinel-2 MSI and Landsat 8 OLI data verified the effectiveness of the proposed framework through single/multitemporal data under small/large-scale scenarios.

Although the proposed method can achieve a satisfactory reconstruction effect in thick cloud and shadow removal, several issues must be considered. First, the cloud and cloud shadow-covered areas are regarded as invalid information. It is wasteful for the reconstruction because many regions may still have useful ground information. Second, the proposed method cannot reconstruct cloudy areas without complementary temporal information. In future research, thin cloud, thick cloud and cloud shadow regions will be separately reconstructed

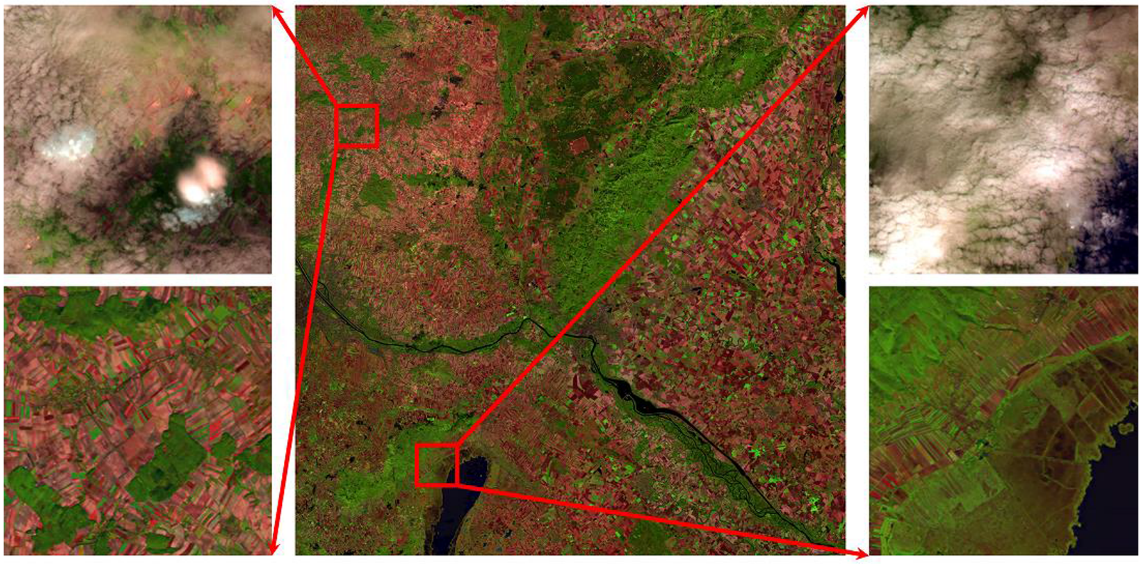
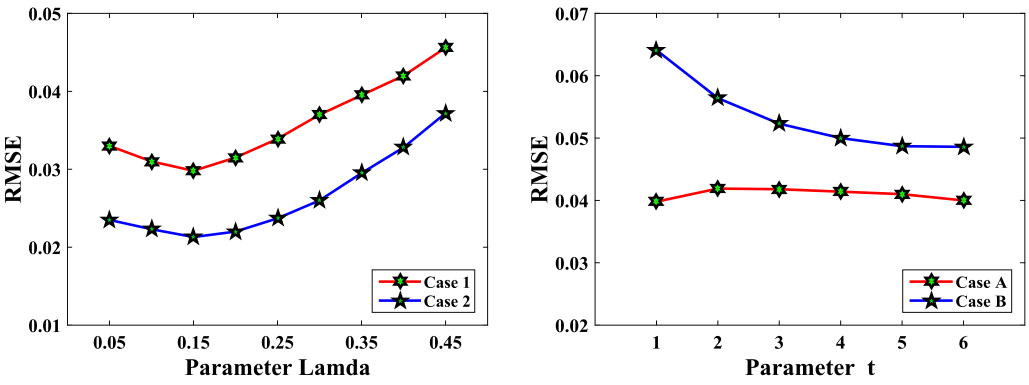
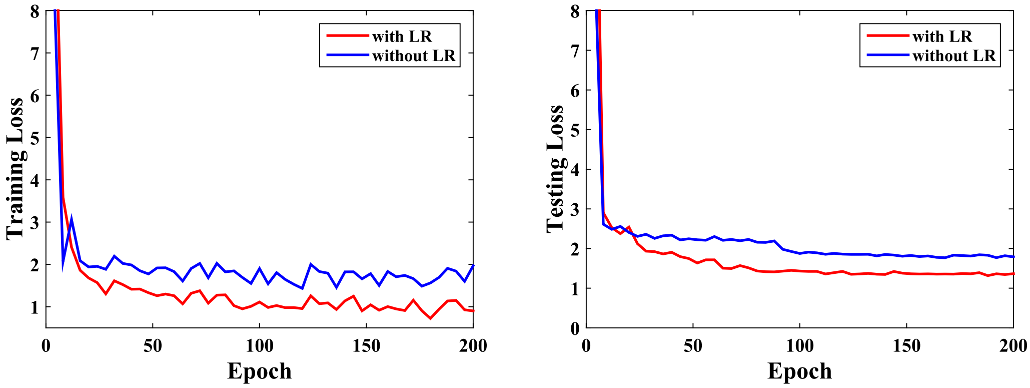


Fig. 13. Global and local zoomed results (original and reconstruction results for 2018.09.15 of Fig. 12).



(a) (b)

Fig. 14. Parameter analysis results. (a) Threshold parameter λ in the local loss function. (b) Temporal number threshold parameter *t*.



(a) (b)

Fig. 15. Training loss with or without linear regression (LR) transformation. (a) with/without LR in the training procedure. (b) with/without LR in the testing procedure.

to overcome the limitation. Besides, cloudy regions without any cloud-free temporal information must be considered through spatial autocorrelation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We gratefully acknowledge the support from the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA19090104), and the National Natural Science Foundation of China (Grant Nos. 41922008 and 61971319). Besides, we would like to gratefully thank Dr. Z. Zhu and S. Qiu, for them providing the cloud detection software Fmask 4.0.

References

Baetens, L., Desjardins, C., Hagolle, O., 2019. Validation of copernicus sentinel-2 cloud masks obtained from MAJA, Sen2Cor, and Fmask processors using reference cloud masks generated with a supervised active learning procedure. Remote Sens. 11 (4), 433.

Bertalmio, M., Vese, L., Sapiro, G., Osher, S., 2003. Simultaneous structure and texture image inpainting. IEEE Trans. Image Process. 12 (8), 882–889.

Chan, T., 2001. Local inpainting models and TV inpainting. SIAM J. Appl. Math. 62 (3), 1019–1043.

Chen, B., Huang, B., Chen, L., Xu, B., 2017. Spatially and temporally weighted regression: a novel method to produce continuous cloud-free Landsat imagery. IEEE Trans. Geosci. Remote Sens. 55 (1), 27–37.

Chen, B., Jin, Y., Brown, P., 2019a. Automatic mapping of planting year for tree crops with Landsat satellite time series stacks. ISPRS J. Photogramm. Remote Sens. 151, 176–188.

Chen, Y., He, W., Yokoya, N., Huang, T., 2019b. Blind cloud and cloud shadow removal of multitemporal images based on total variation regularized low-rank sparsity decomposition. ISPRS J. Photogramm. Remote Sens. 157, 93–107.

Chen, J., Zhu, X., Vogelmann, J.E., Gao, F., Jin, S., 2011. A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. Remote Sens. Environ. 115 (4), 1053–1064.

Cheng, Q., Shen, H., Zhang, L., Yuan, Q., Zeng, C., 2014. Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal MRF model. ISPRS J. Photogramm. Remote Sens. 92, 54–68.

Criminisi, A., Pérez, P., Toyama, K., 2004. Region filling and object removal by exemplar-based image inpainting. IEEE Trans. Image Process. 13 (9), 1200–1212.

Di Mauro, N., Vergari, A., Basile, T.M.A., Ventola, F.G., Esposito, F., 2017. End-to-end learning of deep spatio-temporal representations for satellite image time series classification. In DC@ PKDD/ECML.

Dong, C., Loy, C.C., He, K., Tang, X., 2016. Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. 38 (2), 295–307.

Erinjery, J.J., Singh, M., Kent, R., 2018. Mapping and assessment of vegetation types in the tropical rainforests of the Western Ghats using multispectral Sentinel-2 and SAR Sentinel-1 satellite imagery. Remote Sens. Environ. 216, 345–354.

Gao, G., Gu, Y., 2017. Multitemporal Landsat missing data recovery based on tempo-spectral angle model. IEEE Trans. Geosci. Remote Sens. 55 (7), 3656–3668.

Guillemot, C., Olivier, M., 2014. Image inpainting: Overview and recent advances. IEEE Signal Process Mag. 31 (1), 127–144.

He, K., Sun, J., 2014. Image completion approaches using the statistics of similar patches. IEEE Trans. Pattern Anal. Mach. Intell. 36 (12), 2423–2435.

Ji, T.Y., Yokoya, N., Zhu, X.X., Huang, T.Z., 2018. Nonlocal tensor completion for multitemporal remotely sensed images' inpainting. IEEE Trans. Geosci. Remote Sens. 56 (6), 3047–3061.

Jia, Y., Shelhamer, E., Donahue, et al., 2014. Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia. ACM, pp. 675–678.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436.

Li, X., Shen, H., Zhang, L., Li, H., 2015. Sparse-based reconstruction of missing information in remote sensing images from spectral/temporal complementary information. ISPRS J. Photogramm. Remote Sens. 106, 1–15.

Li, X., Shen, H., Zhang, L., Zhang, H., Yuan, Q., Yang, G., 2014. Recovering quantitative remote sensing products contaminated by thick clouds and shadows using multitemporal dictionary learning. IEEE Trans. Geosci. Remote Sens. 52 (11), 7086–7098.

Li, X., Wang, L., Cheng, Q., Wu, P., Gan, W., Fang, L., 2019a. Cloud removal in remote sensing images using nonnegative matrix factorization and error correction. ISPRS J. Photogramm. Remote Sens. 148, 103–113.

Li, Z., Shen, H., Cheng, Q., Liu, Y., You, S., He, Z., 2019b. Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. ISPRS J. Photogramm. Remote Sens. 150, 197–212.

Li, Z., Shen, H., Cheng, Q., Li, W., Zhang, L., 2019c. Thick cloud removal in high-resolution satellite images using stepwise radiometric adjustment and residual correction. Remote Sens. 11 (16), 1925.

Li, Z., Shen, H., Li, H., Xia, G., Gamba, P., Zhang, L., 2017. Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery. Remote Sens. Environ. 191, 342–358.

Lv, H., Wang, Y., Shen, Y., 2016. An empirical and radiative transfer model based algorithm to remove thin clouds in visible bands. Remote Sens. Environ. 179, 183–195.

Ng, M.K.P., Yuan, Q., Yan, L., Sun, J., 2017. An adaptive weighted tensor completion method for the recovery of remote sensing images with missing data. IEEE Trans. Geosci. Remote Sens. 55 (6), 3367–3381.

Otukei, J., Blaschke, T., 2010. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. Int. J. Appl. Earth Obs. Geoinf. 12, 27–31.

Pelletier, C., Webb, G.I., Petitjean, F., 2019. Temporal convolutional neural network for the classification of satellite image time series. Remote Sens. 11 (5), 523.

Peng, J., Chen, S., Lü, H., Liu, Y., Wu, J., 2016. Spatiotemporal patterns of remotely sensed PM2. 5 concentration in China from 1999 to 2011. Remote Sens. Environ. 174, 109–121.

Rakwatin, P., Takeuchi, W., Yasuoka, Y., 2009. Restoration of Aqua MODIS band 6 using histogram matching and local least squares fitting. IEEE Trans. Geosci. Remote Sens. 47 (2), 613–627.

Rossi, R.E., Dungan, J.L., Beck, L.R., 1994. Kriging in the shadows: geostatistical interpolation for remote sensing. Remote Sens. Environ. 49 (1), 32–40.

Shen, H., Jiang, M., Li, J., Yuan, Q., Wei, Y., Zhang, L., 2019. Spatial-spectral fusion by combining deep learning and variation model. IEEE Trans. Geosci. Remote Sens. 57, 1–13.

Shen, H., Li, H., Qian, Y., Zhang, L., Yuan, Q., 2014. An effective thin cloud removal procedure for visible remote sensing images. ISPRS J. Photogramm. Remote Sens. 96, 224–235.

Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H., Zhang, L., 2015. Missing information reconstruction of remote sensing data: A technical review. IEEE Geosci. Remote Sens. Mag. 3 (3), 61–85.

Toure, S.I., Stow, D.A., Shih, H.C., Weeks, J., Lopez-Carr, D., 2018. Land cover and land use change analysis using multi-spatial resolution data and object-based image analysis. Remote Sens. Environ. 210, 259–268.

Qiu, S., He, B., Zhu, Z., Liao, Z., Quan, X., 2017. Improving Fmask cloud and cloud shadow detection in mountainous area for Landsats 4–8 images. Remote Sens. Environ. 199, 107–119.

Van der Meer, F., 2012. Remote-sensing image analysis and geostatistics. Int. J. Remote Sens. 33 (18), 5644–5676.

Wang, J., Olsen, P.A., Conn, A.R., Lozano, A.C., 2016. Removing clouds and recovering ground observations in satellite image sequences via temporally contiguous robust matrix completion. CVPR 2754–2763.

Wang, Y., Yuan, Q., Li, T., Shen, H., Zheng, L., Zhang, L., 2019. Large-scale MODIS AOD products recovery: Spatial-temporal hybrid fusion considering aerosol variation mitigation. ISPRS J. Photogramm. Remote Sens. 157, 1–12.

Wei, Y., Yuan, Q., Shen, H., Zhang, L., 2017. Boosting the accuracy of multispectral image pansharpening by learning a deep residual network. IEEE Geosci. Remote Sens. Lett. 14 (10), 1795–1799.

Weng, Q., Fu, P., 2014. Modeling annual parameters of clear-sky land surface temperature variations and evaluating the impact of cloud cover using time series of Landsat TIR data. Remote Sens. Environ. 140, 267–278.

Xu, M., Jia, X., Pickering, M., Plaza, A.J., 2016. Cloud removal based on sparse representation via multitemporal dictionary learning. IEEE Trans. Geosci. Remote Sens. 54 (5), 2998–3006.

Xu, M., Jia, X., Pickering, M., Jia, S., 2019. Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform. ISPRS J. Photogramm. Remote Sens. 149, 215–225.

Yuan, Q., Wei, Y., Meng, X., Shen, H., Zhang, L., 2018. A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (3), 978–989.

Yuan, Q., Zhang, Q., Li, J., Shen, H., Zhang, L., 2019. Hyperspectral Image denoising employing a spatial-spectral deep residual convolutional neural network. IEEE Trans. Geosci. Remote Sens. 57 (2), 1205–1218.

Zeng, C., Shen, H., Zhang, L., 2013. Recovering missing pixels for Landsat ETM+ SLC-off imagery using multi-temporal regression analysis and a regularization method. Remote Sens. Environ. 131, 182–194.

Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L., 2017. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Trans. Image Process. 26 (7), 3142–3155.

Zhang, Q., Yuan, Q., Li, J., Liu, X., Shen, H., Zhang, L., 2019a. Hybrid noise removal in hyperspectral imagery with a spatial-spectral gradient network. IEEE Trans. Geosci. Remote Sens. 57 (10), 7317–7329.

Zhang, Q., Yuan, Q., Li, J., Yang, Z., Ma, X., 2018a. Learning a dilated residual network for SAR image despeckling. Remote Sens. 10 (2), 196.

Zhang, Q., Yuan, Q., Zeng, C., Li, X., Wei, Y., 2018b. Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network. IEEE Trans. Geosci. Remote Sens. 56 (8), 4274–4288.

Zhang, Y., Wen, F., Gao, Z., Ling, X., 2019b. A coarse-to-fine framework for cloud removal in remote sensing image sequence. IEEE Trans. Geosci. Remote Sens. 1–12.

Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote Sens. Environ. 221, 430–443.

Zhu, X., Gao, F., Liu, D., Chen, J., 2012. A modified neighborhood similar pixel interpolator approach for removing thick clouds in Landsat images. IEEE Geosci. Remote Sens. Lett. 9 (3), 521–525.

Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sens. Environ. 118, 83–94.

Zhu, Z., Wang, S., Woodcock, C.E., 2015. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. Remote Sens. Environ. 159, 269–277.