

Geo-spatial Information Science

ISSN: 1009-5020 (Print) 1993-5153 (Online) Journal homepage: www.tandfonline.com/journals/tgsi20

A joint thin and thick cloud removal method for high-resolution optical satellite images

Shaocong Zhu, Tao Jiang, Zhiwei Li, Dekun Lin, Jingan Wu & Huanfeng Shen

To cite this article: Shaocong Zhu, Tao Jiang, Zhiwei Li, Dekun Lin, Jingan Wu & Huanfeng Shen (06 May 2026): A joint thin and thick cloud removal method for high-resolution optical satellite images, Geo-spatial Information Science, DOI: 10.1080/10095020.2026.2664312

To link to this article: https://doi.org/10.1080/10095020.2026.2664312

© 2026 Wuhan University. Published by Informa UK Limited, trading as Taylor & Francis Group.

View supplementary material

Published online: 06 May 2026.

Submit your article to this journal

Article views: 17

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tgsi20

GEO-SPATIAL INFORMATION SCIENCE
https://doi.org/10.1080/10095020.2026.2664312

Taylor & Francis
by informa

OPEN ACCESS Check for updates

A joint thin and thick cloud removal method for high-resolution optical satellite images

Shaocong Zhu [a], Tao Jiang [a], Zhiwei Li [a,b], Dekun Lin [a], Jingan Wu [c] and Huanfeng Shen [a,d,e]

[a]School of Resource and Environmental Sciences, Wuhan University, Wuhan, China; [b]Department of Civil, Urban, Earth, and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea; [c]School of Geospatial Engineering and Science, Sun Yat-Sen University, Zhuhai, China; [d]Key Laboratory of Geographic Information System, Ministry of Education, Wuhan, China; [e]Key Laboratory of Digital Cartography and Land Information Application, Ministry of Natural Resources, Wuhan, China

ABSTRACT

Thick clouds, thin clouds, and cloud shadows inevitably degrade high spatial resolution optical remote sensing images (HRIs), thereby greatly restricting their subsequent application. Thin and thick clouds often coexist in HRIs, and the existing methods rarely achieve integrated removal. In this paper, we propose a joint thin–thick cloud removal (JTTR) framework to effectively remove both types of cloud contamination. The JTTR framework employs a multi-scale cloud detection approach to automatically select clean samples for the estimation of a thin cloud thickness map. In addition, an information cloning model based on radiometric adjustment and boundary interpolation (ICM-RABI) is incorporated. For the first time, the ICM-RABI model was applied to thin cloud removal, enabling joint removal of both thin and thick clouds. The proposed JTTR framework achieved a satisfactory performance across multiple sensors and complex scenes in both real and simulated experiments using Gaofen satellite data, demonstrating its strong generalization ability and great potential for practical applications.

ARTICLE HISTORY
Received 5 November 2025
Accepted 20 April 2026

KEYWORDS
Cloud removal; cloud detection; Gaofen; information cloning model

1. Introduction

With the rapid advancement of satellite technology, high spatial resolution optical remote sensing images (HRIs) have been widely applied in various fields (Chen et al. 2022; Li et al. 2024; Xie et al. 2024; Zheng et al. 2021). However, due to the inherent limitations of optical imaging systems (Li et al. 2017), HRIs are inevitably affected by clouds and cloud shadows, which significantly reduce the accuracy of the subsequent application (Li, Xu, and Weng 2024; Wu et al. 2021; Zhou et al. 2019). To date, most studies (Huang and Dewitt 2025; Li et al. 2024; Shao et al. 2019) have classified clouds into two categories – thin and thick clouds – based on their level of opacity. Thin clouds attenuate surface-reflected radiance, leading to information degradation (Xu et al. 2019), whereas thick clouds completely block surface-reflected radiance, resulting in complete information loss (Shen et al. 2015). Consequently, cloud detection and removal methods to accurately identify and remove cloud-contaminated regions are essential preprocessing steps for HRIs.

Cloud detection methods can be divided into two categories: mono-temporal methods (Wright et al. 2024, 2025; Zhu, Li, and Shen 2024; Zhu, Wang, and Woodcock 2015) and multi-temporal methods (Lee et al. 2025; Wang et al. 2024; Zhang et al. 2021). The mono-temporal methods exploit the spectral or morphological differences between clouds and cloud-free surfaces using physical rules or machine learning approaches, whereas the multi-temporal methods identify clouds by detecting abrupt changes across temporal image sequences. Benefiting from the advances in deep learning for image segmentation (Kotaridis and Lazaridou 2021) and classification (Chen et al. 2021), mono-temporal detection methods using deep learning have achieved notable success and been widely applied to HRIs. However, the current cloud detection methods for HRIs struggle to accurately detect haze and thin clouds, due to their weak spectral characteristics. The omission of haze and thin clouds significantly degrades the performance of the cloud detection methods and limits their applicability in downstream tasks.

The current thin cloud removal algorithms primarily employ image transformation techniques (Li et al. 2024) and can be categorized into three main methods: physical-based methods (Bi et al. 2022; Chen et al. 2016; Lv, Wang, and Shen 2016), band assistance-based methods (Jiang et al. 2024; Xu et al. 2019;

CONTACT Huanfeng Shen shenhf@whu.edu.cn
Supplemental data for this article can be accessed online at https://doi.org/10.1080/10095020.2026.2664312
© 2026 Wuhan University. Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Zhang, Li, and Shen 2021), and deep learning based methods (Ding et al. 2024; Wu et al. 2023; Xu et al. 2024; Xu et al. 2022). The physics-based methods leverage prior knowledge of thin clouds or specific image features to achieve thin cloud removal. The band assistance-based methods rely on specific spectral bands – such as shortwave infrared or cirrus bands – as auxiliary information to perform thin cloud removal. The deep learning-based methods train networks to learn the features of cloud-contaminated and cloud-free images from datasets. However, the applicability of thin cloud removal methods for HRIs is constrained by the limited spectral bands (blue, green, red, and near-infrared), resulting in spectral distortion over bright surfaces and water regions.

Thick cloud removal in HRIs is generally treated as a process of reconstructing the missing information (Shen et al. 2015). The existing methods can be divided into spatial-based methods (Chen et al. 2023; Tao et al. 2022) and temporal-based methods (Chen et al. 2024; Yun, Jung, and Han 2024; Zhang et al. 2025; Zhang et al. 2021). The spatial-based methods exploit auxiliary information from cloud-free regions within the same image to reconstruct the cloud-contaminated areas, whereas the temporal-based methods leverage complementary information from multi-temporal images. Although most thick cloud removal methods can effectively remove the thick clouds in HRIs, their performance in practical applications is affected by haze and thin clouds, making it difficult to obtain visually satisfactory results that are consistent with those under ideal conditions.

In summary, although numerous cloud removal methods for HRIs have been proposed, thick and thin clouds often coexist and interact, making it difficult to obtain satisfactory results using these methods individually. Some studies (Li et al. 2024) have attempted joint cloud removal by introducing synthetic aperture radar (SAR) images. However, the difficulty of acquiring temporally close SAR images with a comparable resolution, along with the radiometric discrepancies between SAR and optical images (Xu et al. 2025), significantly limits their applicability and generalization. Other studies (Hu et al. 2014; Li et al. 2020) have combined thin and thick cloud removal methods in a sequential manner to achieve joint cloud removal in HRIs. Nevertheless, such methods neglect the complementarity between the two processes, failing to address the spectral distortion in thin cloud removal and the instability in relatively thicker thin clouds. Moreover, they often require substantial manual intervention to obtain satisfactory joint cloud removal results.

To address these limitations, we propose a joint thin–thick cloud removal (JTTR) framework to remove thin and thick clouds in multi-sensor HRIs. The proposed JTTR framework first employs cloud detection techniques to automatically estimate a thin cloud thickness map (TCTM) and identify the thick clouds. The TCTM is then utilized for global thin cloud removal. Based on an information cloning model (ICM), the original spectral characteristics of the ground surface are utilized to eliminate the spectral distortion in the global thin cloud removal results, while complementary information from multi-temporal images is introduced to reconstruct the identified thick cloud regions, thereby achieving JTTR.

2. Methods

The workflow of the proposed JTTR framework for multi-sensor HRIs is shown in Figure 1. It should be noted that this study classifies the cloud-contaminated regions in HRIs into four categories: thick clouds, relatively thick thin clouds, haze/thin clouds, and cloud shadows, as shown in Figure 2. Given the unstable performance of thin cloud removal methods on relatively thick thin clouds (Jiang et al. 2025; Shen et al. 2020), as illustrated in Figure S1, the proposed framework treats thick clouds, relatively thick thin clouds, and cloud shadows as regions of complete information loss and reconstructs them using a thick cloud removal technique.

In addition, by deliberately diverting relatively thick thin cloud regions – that could otherwise be removed by thin cloud removal technique – into the reconstruction technique, a buffer is established between thin cloud and thick cloud removal processes, as shown in Figure 3. This strategy reduces the dependence of JTTR on the accuracy of cloud masks and on the stability of thin cloud removal methods in relatively thick thin cloud regions.

2.1. Cloud detection method based on multi-scale networks

Under clear-sky conditions, a strong correlation exists between the top-of-atmosphere (TOA) reflectance of the blue band I_b and the red band I_r, which can be defined as a clear-sky line (Zhang, Guindon, and Cihlar 2002). A robust linear fitting technique (Holland and Welsch 1977) is then applied to estimate the weight coefficient k between I_b and I_r in the clear-sky regions. Accordingly, the TCTM (Jiang et al. 2025) can be estimated as follows:

$$\mathrm{TCTM} = \min_{\Omega(3)}\left(I_b + \frac{k}{\sqrt{k^2+1}} \cdot I_b - \frac{1}{\sqrt{k^2+1}} \cdot I_r\right) \quad (1)$$

where $\min_{\Omega(3)}$ is a minimum filter with a window size of 3.

Traditional TCTM estimation requires manual selection of reliable cloud-free regions. To avoid manual intervention and provide a reliable binary cloud

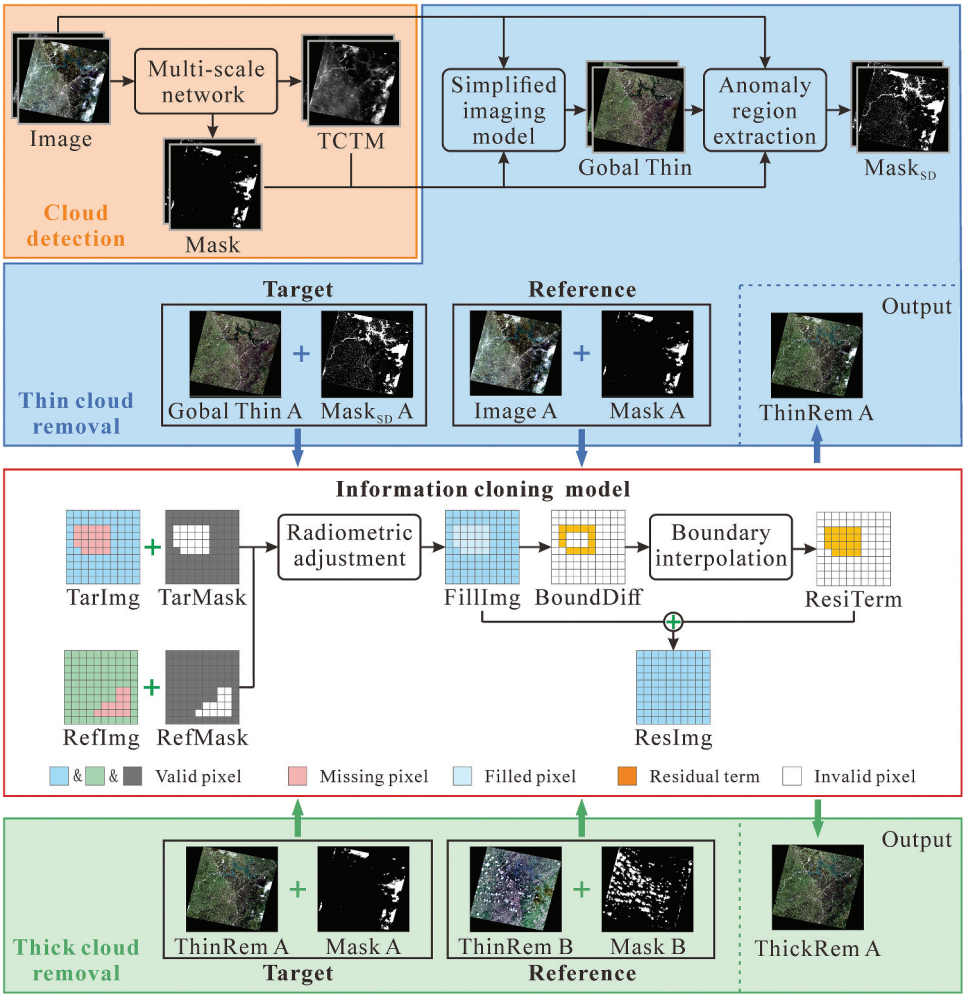


Figure 1. Workflow of the proposed JTTR framework.

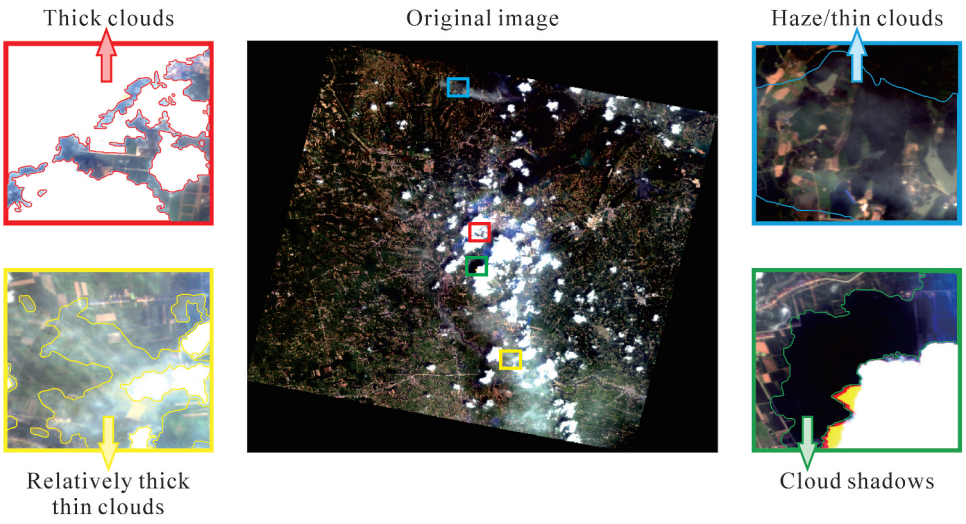


Figure 2. Examples of typical cloud-contaminated regions in HRIs.

mask for the subsequent thick cloud removal, a cloud detection method based on multi-scale networks is proposed, as shown in Figure 4. First, a pixel-level cloud detection network and three scene-level networks at different scales (256×256, 128×128, and 64×64) are trained using a pixel-level dataset. Subsequently, the scene-level networks are employed to automatically extract a 256×256 cloud-free patch with the highest confidence, which is utilized as the clear-sky region for TCTM estimation.

Owing to the distinct boundary characteristics of thick clouds (Huang, Wang, and Chen 2018; Wu et al. 2022; Zhu et al. 2026), the gradient of the TCTM exhibits significantly higher values at thick cloud

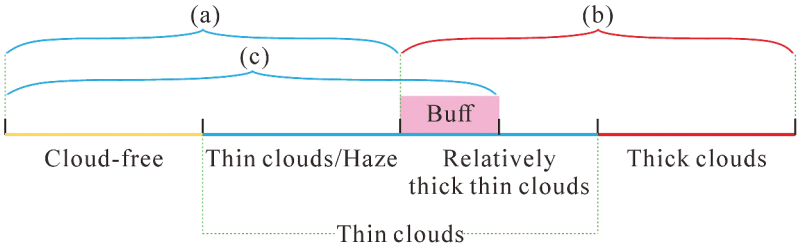


Figure 3. Cloud removal strategies for different cloud types of the JTTR framework. (a) Regions removed by thin cloud removal technique. (b) Regions removed by thick cloud removal technique. (c) Regions that can be effectively removed by thin cloud removal technique.

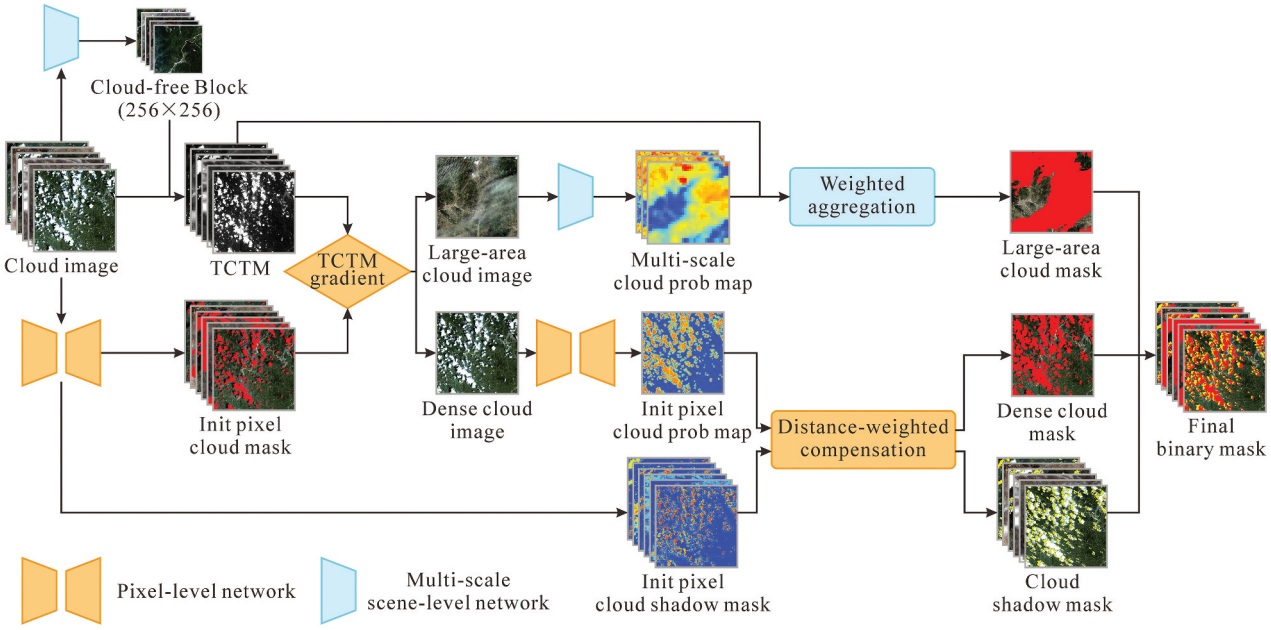


Figure 4. Framework of the cloud detection method based on multi-scale networks.

boundaries. Leveraging this characteristic, images are classified into dense and large-area cloud categories by calculating the mean gradient values along the boundaries of each cloud block in the initial pixel-level cloud mask generated by the pixel-level network.

For dense cloud images, an initial pixel-level cloud probability map $\mathcal{P}_{\mathrm{INIT}}(x)$ is generated using the pixel-level network. Adaptive distance-weighted compensation is applied to $\mathcal{P}_{\mathrm{INIT}}(x)$, and the compensated probability map $\mathcal{P}_{\mathrm{OPT}}(x)$ is segmented using the initial threshold to obtain a dense cloud mask. $\mathcal{P}_{\mathrm{OPT}}(x)$ is calculated as follows:

$$\mathcal{P}_{\mathrm{OPT}}(x) = \mathcal{P}_{\mathrm{INIT}}(x) + \frac{50-\mathrm{Dist}(x)}{50} \cdot P \qquad (2)$$

where $\mathrm{Dist}(x)$ denotes the distance from the current pixel to the nearest cloud block, and P is the compensation constant, calculated as the mean value of $\mathcal{P}_{\mathrm{INIT}}(x)$ within the initial pixel cloud mask.

For large-area cloud images, we adopt the multi-scale scene-level network strategy based on the SpecMCD method to generate a binary cloud mask, while replacing the cloud thickness map with the TCTM, which is more suitable for thin clouds. However, due to the imaging angle limitations of remote sensing sensors, large-area cloud shadow coverage is rare in HRIs, making it difficult to construct a suitable scene-level dataset for cloud shadows. Therefore, only pixel-level cloud shadow labels are utilized to train a pixel-level cloud shadow detection network, and the binary cloud shadow mask is obtained using the same optimization method as for the dense cloud mask.

In the proposed approach, cloud masks and cloud shadow masks are generated independently. To ensure spatial consistency in the final binary mask, a cloud mask prioritization strategy is applied to overlapping regions during the fusion process.

2.2. Joint thin–thick cloud removal via an information cloning model

We propose a JTTR framework via an ICM. The conventional thin cloud removal methods are unable to remove thick clouds and often lead to color distortion over complex surface regions. Moreover, the presence of thin clouds tends to interfere with the performance

of the thick cloud removal methods. To address these issues, the JTTR framework leverages the ICM to effectively transfer the local texture and spectral information between images. Specifically, the original surface spectral features are employed to correct the spectral distortion in the thin cloud removal results, while complementary information from multitemporal images is incorporated to reconstruct the regions affected by thick clouds, thereby achieving JTTR.

2.2.1. Information cloning model

The major challenge of applying the ICM in HRIs lies in the large radiometric differences between images. To achieve seamless information cloning, an information cloning model based on radiometric adjustment and boundary interpolation (ICM-RABI) is proposed.

First, based on the valid pixels between the images, the values of the regions to be filled are adjusted as follows:

$$I_A^*(x') = \frac{\sigma_A}{\sigma_B} \cdot I_B(x') + \mu_A - \frac{\sigma_A}{\sigma_B} \cdot \mu_B \tag{3}$$

where I_A^* represents the values to be filled into image A after radiometric adjustment; σ_A and σ_B are the standard deviations; and μ_A and μ_B are the mean values of the valid pixels within a local window of images A and B, respectively. The local window size is set to half of the smaller dimension of the minimum enclosing rectangle of the target filling region. To improve the computational efficiency, box filtering (Pires, Singh, and Moura 2011) is employed to compute the mean and standard deviation within each local window.

Subsequently, to eliminate the subtle radiometric differences between the filling region and the original clear region, a convolutional pyramid-based boundary interpolation method (Farbman, Fattal, and Lischinski 2011; Shepard 1968) is employed for residual correction. Each pixel in the filling region can be calculated as follows:

$$\overline{I_A}(x') = I_A^*(x') + \tilde{I}_A(x') \tag{4}$$

$$\tilde{I}_A(x') = \frac{\sum_k \omega_k(x') \cdot b(x_k)}{\sum_k \omega_k(x')} \tag{5}$$

where $\overline{I_A}$ represents the values after residual correction; $\tilde{I}_A$ is the residual term; $b(x_k)$ is the value difference between the filling region and the original clear region at the boundary; and $\omega_k(x')$ is a weight function, which can be empirically defined as follows:

$$\omega_k(x') = \frac{1}{d(x', x_k)^3} \tag{6}$$

where $d(x', x_k)$ is the distance between x' and x_k. To improve the efficiency of the image fusion, $b(x_k)$ is extended to the entire filling region:

$$\hat{r}(x') = \begin{cases} b(x_k), if x' = x_k \\ 0, otherwise \end{cases} \tag{7}$$

Equation (5) can then be rewritten as a ratio of convolutions as follows:

$$\tilde{I}_A(x') = \frac{\sum_{i=0}^{n} \omega(x', x_i) \cdot \hat{r}(x_i)}{\sum_{k=0}^{n} \omega(x', x_i) \cdot \chi_{\hat{r}}(x_i)} = \frac{\omega \cdot \hat{r}}{\omega \cdot \chi_{\hat{r}}} \tag{8}$$

where $\chi_{\hat{r}}$ is the characteristic function corresponding to $\hat{r}$, which equals 1 where $\hat{r}$ is non-zero and 0 otherwise. Equation (8) is efficiently evaluated using a convolution pyramid (Farbman, Fattal, and Lischinski 2011) with a set of separable filters: two 5×5 kernels $\boldsymbol{SF}_{5\times5}^1$ and $\boldsymbol{SF}_{5\times5}^2$, and one 3×3 kernel $\boldsymbol{SF}_{3\times3}^1$. In this study, the separable filters are configured as follows:

$$k = [0.1507, 0.6836, 1.0334, 0.0270, 0.0312, 0.7753] \tag{9}$$

$$KS = \begin{cases} \boldsymbol{SF}_{5\times5}^1 = \boldsymbol{K}_1^{\mathrm{T}} \cdot \boldsymbol{K}_1, \boldsymbol{K}_1 = [k_1, k_2, k_3, k_2, k_1] \\ \boldsymbol{SF}_{5\times5}^2 = \boldsymbol{SF}_{5\times5}^1 \cdot k_4 \\ \boldsymbol{SF}_{3\times3}^1 = \boldsymbol{K}_2^{\mathrm{T}} \cdot \boldsymbol{K}_2, \boldsymbol{K}_1 = [k_5, k_6, k_5] \end{cases} \tag{10}$$

Finally, residual correction is applied to the filling regions using Equation (4) to achieve seamless information cloning.

2.2.2. Joint cloud removal

As thin clouds affect surface reflectance and lead to the degradation of surface information, their effect on the image radiance (Zhang, Li, and Shen 2021) can be expressed as follows:

$$I_i = I_i^s + I_i^c, i \in \{b, g, r, nir\} \tag{11}$$

where I is the TOA reflectance of the images with thin clouds, I^s is the TOA reflectance of the clear-sky surfaces, and I^c is the reflectance of the thin cloud. To remove thin clouds, the TCTM derived from Equation (1) is utilized to estimate I^c (Jiang et al. 2025), and the global thin cloud removal result is then obtained using Equation (11).

Constrained by the accuracy of the I^c estimation and the application of the minimum filter, the global thin cloud removal results may exhibit significant spectral distortion over bright surfaces, water regions, and image boundaries (Jiang et al. 2025; Makarau et al. 2014; Shen et al. 2020). In the proposed approach, these regions in the global thin cloud removal results are regarded as areas with completely missing information. Subsequently, the regions covered by the cloud and cloud shadow masks, as well as the spectrally distorted areas, are treated as the target regions to be filled in the global thin cloud removal results, while the remaining regions are treated as valid pixels. As shown in Figure 5, the radiometric information from the original image is then filled into the global thin cloud removal results via the ICM, yielding the final thin cloud removal results.

In the process of thin cloud removal, spectral distortion regions are identified by leveraging the TCTM, the normalized difference water index (NDWI) (McFeeters 1996), and the radiometric differences between the original image and the global thin cloud removal results, which can be calculated as follows:

$$\Omega_{\text{Bound}}\left(x'_{\text{ED}}\right) = I\left(x'_{\text{ED}}\right) - I^{s}\left(x'_{\text{ED}}\right), x'_{\text{ED}} \in \{x | 0 < d(x, \partial\Omega_{\text{null}}) \leq 20\} \quad (12)$$

$$M_{\text{SD}}(x) = \begin{cases} 1, \text{TCTM}(x) > \tau_1 \cup^{\text{NDWI}}(x) > \tau_2 \cup^{\Omega_{\text{Bound}}}\left(x'_{\text{ED}}\right) = 0 \\ 0, \text{else} \end{cases} \quad (13)$$

where x'_{ED} represents pixels at the image boundary, defined as a region extending 20 pixels inward from the invalid data $\partial\Omega_{\text{null}}$; Ω_{Bound} is the image boundary; and τ_1 is the segmentation threshold for bright surfaces, with values of 0.18, 0.15, and 0.16 for Gaofen-1 multispectral (GF1) images, Gaofen-2 multispectral (GF2) images, and Gaofen-7 multispectral (GF7) images, respectively. τ_2 is the segmentation threshold for water regions, with values of 0, 0.15, and 0.2 for GF1, GF2, and GF7 images, respectively.

After thin cloud removal, haze and thin clouds in HRI are effectively removed, whereas contamination from thick clouds, relatively thick thin clouds, and cloud shadows still remains. To reconstruct the surface information in these regions, a reference image from another temporal phase is introduced to provide complementary information. However, HRIs are highly sensitive to subtle surface changes (Wen et al. 2021), and abrupt radiometric differences between multi-temporal images may lead to anomalous calculations of correction coefficients in the ICM, as shown in Figure 6. Therefore, before applying thick cloud removal via the information cloning method, it is essential to first identify abrupt change regions between images from different temporal phases. The abrupt change regions are extracted as follows:

$$I_{\text{dif}}(x) = \max\left|\hat{I}_i^A(x) - \hat{I}_i^B(x)\right|, i \in \{b, g, r, nir\} \quad (14)$$

$$M_{\text{AC}}(x) = \begin{cases} 1, I_{\text{dif}}(x) > 0.2 \\ 0, otherwise \end{cases} \quad (15)$$

where I_{dif} represents the reflectance differences between the images from different temporal phases, $\hat{I}$

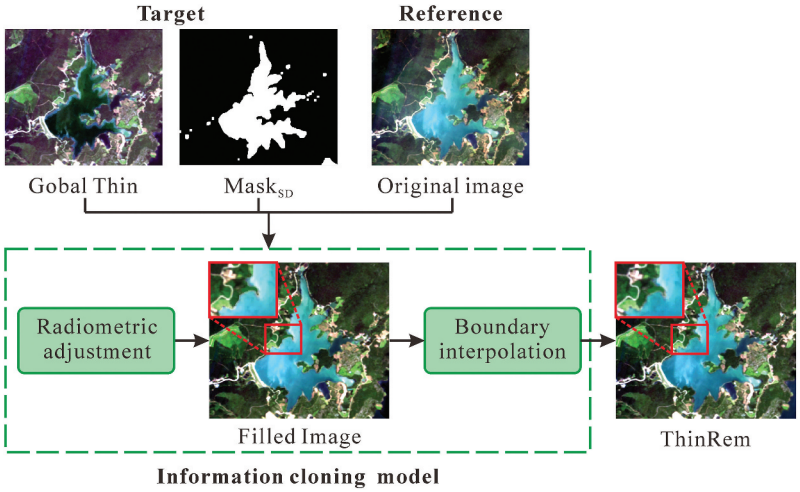


Figure 5. Framework of thin cloud removal via the ICM.

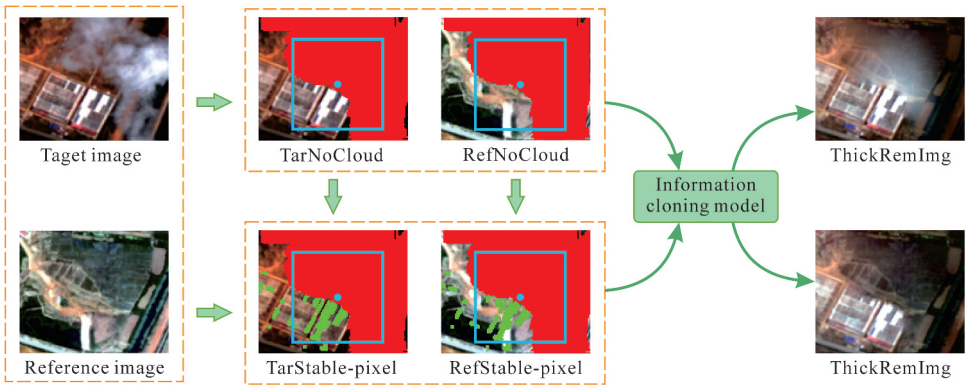


Figure 6. Visual example of thick cloud removal using non-cloud pixels and stable pixels (red: cloud-covered regions; green: abrupt change regions; blue box: local window for radiometric adjustment).

denotes the normalized reflectance, and M_{AC} is the binary mask of the abrupt change regions.

The missing regions in the target image are then filled using the corresponding regions from the reference image through the ICM, where the valid regions are defined as the areas not covered by M_{AC} or the cloud/cloud shadow masks.

3. Experimental data and results

3.1. Experimental data and settings

As shown in Table 1, we adopted the pixel-level cloud detection generation method developed by Zhu, Li, and Shen (2024) to construct the pixel-level GF-Thick-WHU dataset and the multi-scale scene-level GF-Scene-WHU dataset for 8-m GF1 images, 3.2-m GF2 images, and 3.2-m GF7 images. Given the scarcity of large-scale cloud shadow datasets for HRIs, we constructed the pixel-level cloud shadow GF-Shadow-WHU dataset by combining the pixel-level labels from the GF1-WHU dataset (Li et al. 2017) with manual annotation. To validate the effectiveness of different cloud and cloud shadow detection methods, we divided the manually optimized dataset named GF1-Thin-WHU, containing both thin cloud, thick cloud, and cloud shadow labels, into training and validation sets at a 7:3 ratio. The validation images are shown in Figure S2.

To evaluate the effectiveness of the proposed JTTR framework for thin, thick, and mixed cloud removal, one thin cloud-contaminated image and its corresponding temporally adjacent cloud-free image were selected from each of the GF1, GF2, and GF7 images to assess the thin cloud removal accuracy. In addition, three pairs of cloud-free images from these sensors were utilized to conduct thick cloud simulation experiments for a comparative evaluation of the thick cloud removal. Furthermore, eight GF1 images, seven GF2 images, and nine GF7 images affected by both thin and thick clouds were employed as the target images to comprehensively validate the effectiveness of the JTTR framework. For each target image, a temporally adjacent reference image providing complementary information was selected for thick cloud removal, as shown in Figure S3. All the images were acquired between January 2020 and December 2021. The bandwidth information of the images utilized in this study was: blue (0.45–0.52 μm), green (0.52–0.59 μm), red (0.63–0.69 μm), and near-infrared (0.77–0.89 μm).

The cloud detection datasets used in this study were annotated in an automated manner (Zhu, Li, and Shen 2024), which has been demonstrated to be well-suited for the RegNetY-4.0GF network (Radosavovic et al. 2020). Therefore, the RegNetY-4.0GF network with ImageNet-pretrained weights (Deng et al. 2009) was selected as the cloud and cloud shadow detection network. Both the pixel-level and scene-level networks were trained with a cross-entropy loss function and the Adam optimizer. The parameters of all the comparison methods were configured according to their original publications. All the experiments were conducted on a desktop computer with a Windows operating system, Intel Core i7-10700 CPU @ 2.90 GHz, 32 GB RAM, and an NVIDIA GeForce RTX 3070Ti GPU with 8 GB of memory.

3.2. Evaluation of the thin cloud removal performance

The effectiveness of the proposed JTTR framework for thin cloud removal (JTTR-Thin) was assessed by comparing it with the state-of-the-art thin cloud removal methods. Two single-temporal thin cloud removal methods applicable to the visible and near-infrared bands – SSADCP (Shen et al. 2020) and DPASC (Jiang et al. 2025) – were selected as comparison methods. Quantitative evaluation was performed using the correlation coefficient (CC), root-mean-square error (RMSE), and spectral angle (SA) between the regions not covered by cloud masks in the thin cloud-removed images and their corresponding temporally adjacent cloud-free images. As shown in Table 2, the SSADCP method exhibits a lower accuracy across all the evaluation metrics, compared with DPASC and JTTR-Thin. Although the JTTR-Thin method performs slightly worse than DPASC in RMSE for the GF2 experiment, it outperforms DPASC in all the other metrics.

To further validate the performance of the JTTR framework for thin cloud removal, we performed a visual comparison, as shown in Figure 7. The results demonstrate that all three methods are effective in removing thin clouds. However, the SSADCP method exhibits significant spectral distortion, particularly in cloud-free regions, and introduces noticeable radiometric anomalies along image boundaries. Both

Table 1. Summary of the cloud and cloud shadow detection datasets utilized in this study.

Dataset	Image size	Number	Label type	Label source
GF-Thick-WHU	256 × 256	>330K	Pixel-level	Automated annotation
GF-Scene-WHU	256 × 256	>330K	Scene-level	GF-Thick-WHU
	128 × 128	>1.3 M		
	64 × 64	>5 M		
GF-Shadow-WHU	256 × 256	>70K	Pixel-level	Manual annotation
GF-Thin-WHU	>4500 × 4500	60	Pixel-level	Manual annotation

Table 2. Quantitative evaluation of the state-of-the-art thin cloud removal methods on GF1, GF2, and GF7 images.

Image source	Image date	Method	CC (↑)	RMSE (↓)	SA (↓)
8-m GF1	2020.12.22/	SSADCP	0.7890	0.0787	1.5451
(5479 × 4742 × 4)	2020.11.11	DPASC	0.8404	0.0338	1.5215
		JTTR-Thin	**0.8678**	**0.0267**	**1.5175**
3.2-m GF2	2021.09.18/	SSADCP	0.7211	0.0759	1.5436
(6787 × 5958 × 4)	2021.09.23	DPASC	0.9122	**0.0205**	1.5093
		JTTR-Thin	**0.9169**	0.0218	**1.5077**
3.2-m GF7	2021.11.28/	SSADCP	0.5224	0.1847	1.4378
(6462 × 6669 × 4)	2021.12.03	DPASC	0.7790	0.0468	1.2692
		JTTR-Thin	**0.7989**	**0.0447**	**1.2686**

DPASC and JTTR-Thin achieve a satisfactory overall performance, but a detailed analysis reveals that DPASC causes color distortion over bright surfaces and water regions, whereas JTTR-Thin effectively removes thin clouds, while better preserving the original spectral features of the ground surface.

3.3. Evaluation of the thick cloud removal performance

Based on cloud-free images from the GF1, GF2, and GF7 sensors, simulation experiments were conducted to evaluate the performance of thick cloud removal with the proposed JTTR framework. In the simulation experiments, the ground truth within the cloud masks of the target images was removed and reconstructed using complementary information from reference images. For comparison, we selected the dual-temporal optical image reconstruction methods of WLR (Zeng, Shen, and Zhang 2013), FRARC (Zhu et al. 2023), and PGFCTS (Wu et al. 2024). The performance of the different methods was then evaluated based on the CC, RMSE, SA, and structural similarity index (SSIM) between the reconstructed regions and the corresponding ground truth. Table 3 presents the

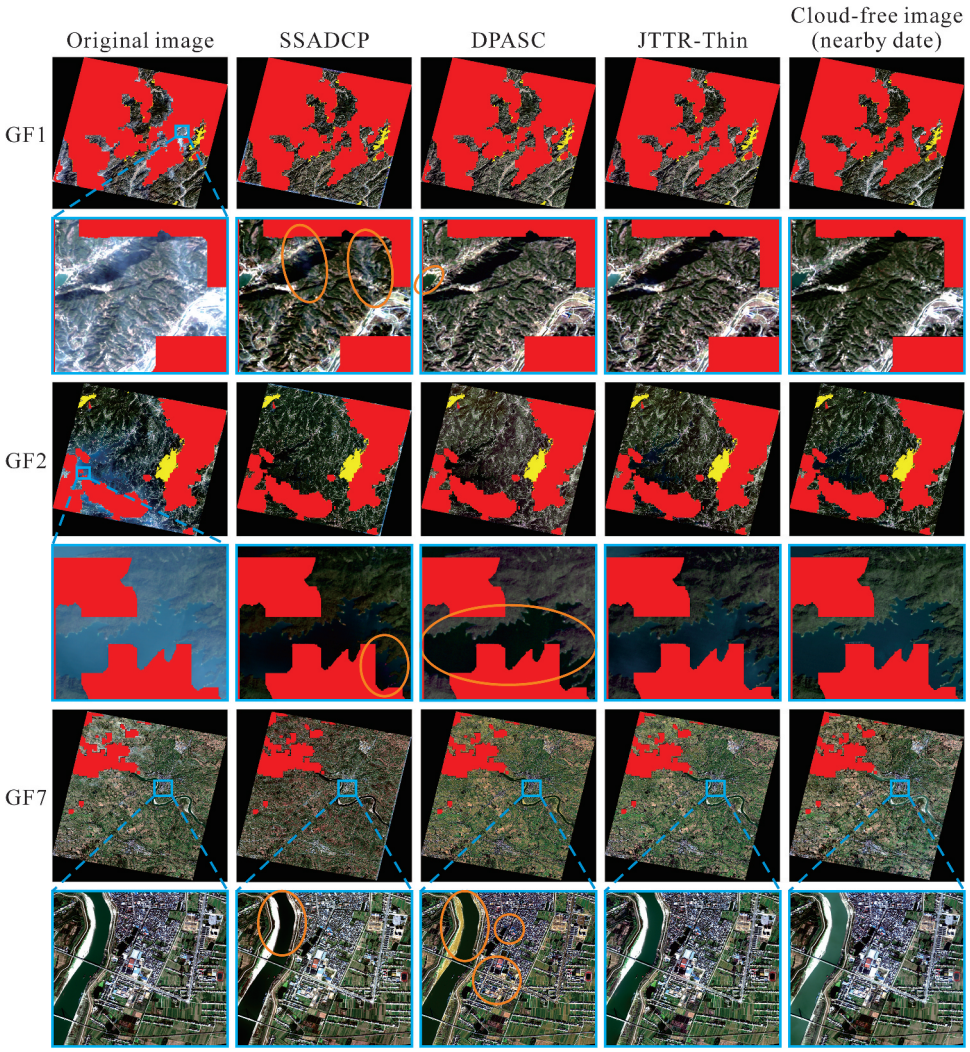


Figure 7. Thin cloud removal results of the different thin cloud removal methods (the red and yellow regions denote the cloud and cloud shadow masks generated by the proposed JTTR framework).

quantitative evaluation results of the different thick cloud removal methods, showing that FRARC, PGFCTS, and JTTR-Thick achieve a superior accuracy across all the simulation experiments. Although PGFCTS yields the lowest RMSE values, its performance in SA is inferior to that of JTTR-Thick, suggesting that PGFCTS better preserves the radiometric intensity, while JTTR-Thick more effectively maintains the spectral characteristics of the ground surface. Furthermore, as the image resolution increases, JTTR-Thick achieves the highest CC and SSIM, demonstrating its superior ability to preserve structural details and its effectiveness in reconstructing HRIs with complex textures.

Subsequently, we compared the spectral profiles of the reconstructed regions obtained using the different thick cloud removal methods, as shown in Figure 8. The WLR method consistently underestimates the radiometric intensity in the GF2 and GF7 experiments, leading to a poor reconstruction performance. In contrast, PGFCTS tends to overestimate the visible-band values in the GF1 and GF7 images, while underestimating in the GF2 imagery. FRARC and JTTR-Thick show partial overestimation in the visible bands but achieve a better overall radiometric reconstruction. In the near-infrared band, the reconstruction is strongly affected by the abrupt surface changes between multi-temporal images. The similar pixel-based PGFCTS method is able to reconstruct the abrupt reflectance changes within the reconstructed regions. The ICM-based FRARC and JTTR-Thick methods tend to preserve the surface features of the reference images within the reconstructed regions, resulting in a lower near-infrared reconstruction accuracy, compared with PGFCTS.

The visual results of the different thick cloud removal methods are presented in Figure 9. WLR and PGFCTS achieve seamless reconstruction in the small-area missing regions, but exhibit notable radiometric inconsistencies in the large-area missing regions. The analysis of the enlarged regions reveals noticeable color distortions and artifacts in the complex-texture areas. In contrast, FRARC and Thick-ICM obtain satisfactory reconstruction results across cloud-covered regions of various sizes. Nonetheless, FRARC still shows slight color discrepancies in water regions. Considering the quantitative accuracy and visual quality, the proposed JTTR framework demonstrates a superior performance and is suitable for thick cloud removal in HRIs.

3.4. Evaluation of the mixed thin–thick cloud removal performance

To comprehensively assess the effectiveness of the proposed JTTR framework, we conducted a comparative analysis on mixed thin–thick cloud images with the state-of-the-art thin and thick cloud removal methods, as illustrated in Figure 10. First, the zoomed-in sub-regions of the DPASC results on GF1 and GF7 images reveal that relying solely on the advanced thin cloud removal method is incapable of effectively eliminating thick clouds. This limitation leads to pronounced radiometric anomalies in thick cloud areas, manifested as abnormal red or green appearances, and introduces artifacts in relatively thick thin cloud regions along thick-cloud boundaries, where residual cloud textures remain visible. In the GF2 results, DPASC exhibits severe spectral distortions over certain land-cover types, where originally red-toned urban areas are incorrectly transformed into purplish hues after thin cloud removal.

For the PGFCTS method, although the detected thick clouds and cloud shadow regions are effectively reconstructed, the zoomed results from GF7 images indicate substantial artifacts and texture loss in urban areas with complex spatial details. Moreover, both global and local results on GF1 and GF2 images show that the reconstructed regions obtained by PGFCTS tend to exhibit abnormally elevated reflectance, resembling thin cloud or haze, indicating that undetected thin clouds can still adversely affect the reconstruction process and lead to an overall overestimation of surface reflectance.

Compared to the other methods, the proposed JTTR framework not only removes the hard-to-detect thin clouds but also effectively reconstructs the detected

Table 3. Quantitative evaluation of the state-of-the-art thick cloud removal methods on GF1, GF2, and GF7 images.

Image source	Cloud percentage	Image date	Method	CC (↑)	RMSE (↓)	SA (↓)	SSIM (↑)
8-m GF1 (5482 × 4707 × 4)	31.8%	2020.11.11/ 2021.11.15	WLR	0.8537	0.0224	1.4433	0.7793
			FRARC	0.8352	0.0246	1.4434	0.7883
			PGFCTS	**0.8613**	**0.0217**	1.4428	**0.8026**
			JTTR-Thick	0.8549	0.0232	**1.4423**	0.7947
3.2-m GF2 (7100 × 5981 × 4)	70.6%	2020.05.23/ 2020.03.05	WLR	0.6325	0.0555	1.2649	0.6761
			FRARC	0.6644	0.0550	1.2564	0.7266
			PGFCTS	0.6715	**0.0516**	1.2553	0.7215
			JTTR-Thick	**0.6784**	0.0542	**1.2542**	**0.7294**
3.2-m GF7 (6010 × 6828 × 4)	52.1%	2021.09.21/ 2021.01.13	WLR	0.9068	0.0317	1.0640	0.8506
			FRARC	0.9300	0.0283	1.0616	0.8581
			PGFCTS	0.9351	**0.0268**	1.0630	0.8482
			JTTR-Thick	**0.9354**	0.0274	**1.0614**	**0.8620**

Note: Bold numbers in the table indicate the best-performing method within each experiment.

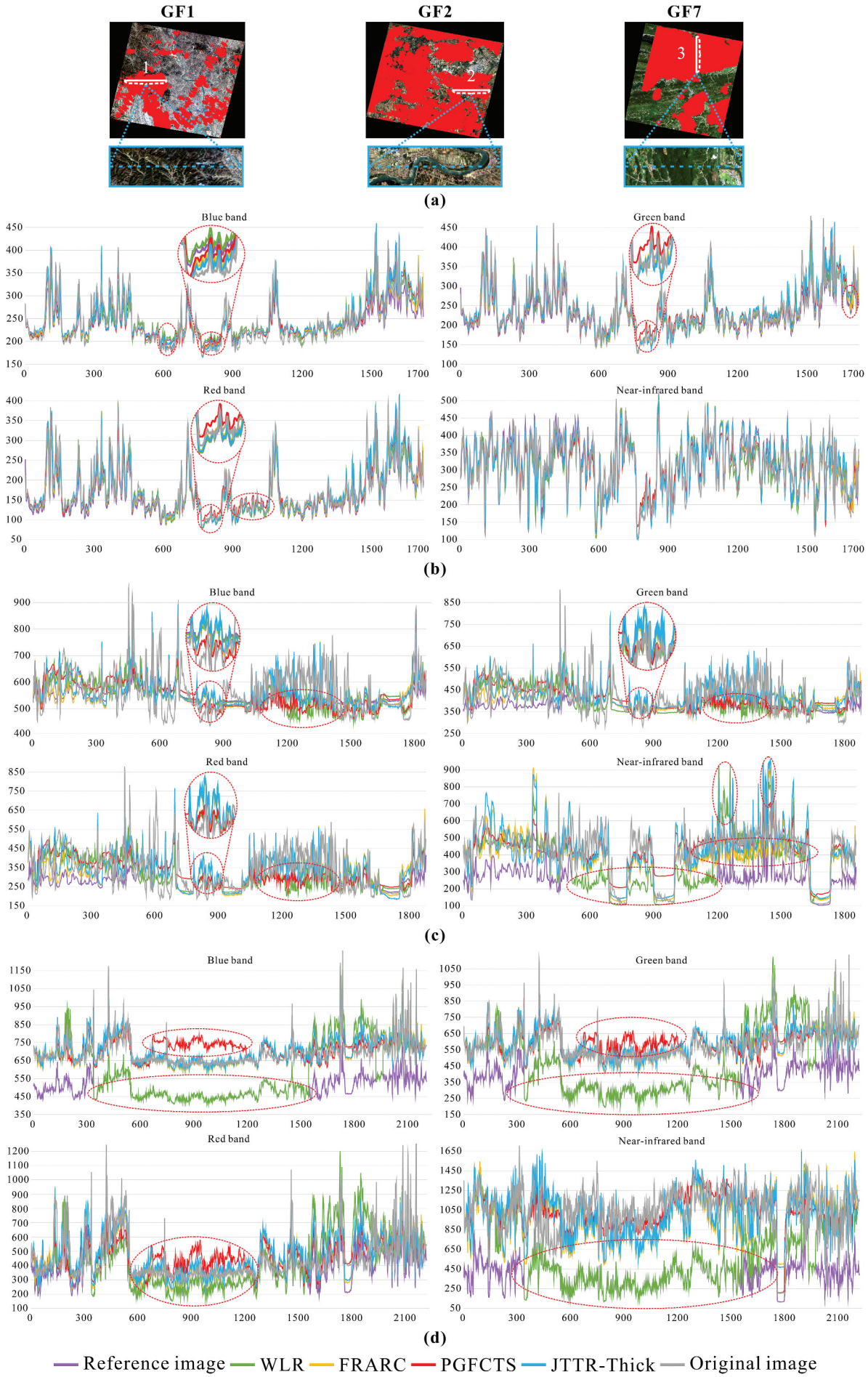


Figure 8. Spectral profile comparisons of the different thick cloud removal methods in the GF1, GF2, and GF7 experiments. (a) Example of the spectral profile regions in the simulated thick cloud removal experiment. (b) Spectral profile of region 1 in (a). (c) Spectral profile of region 2 in (a). (d) Spectral profile of region 3 in (a).

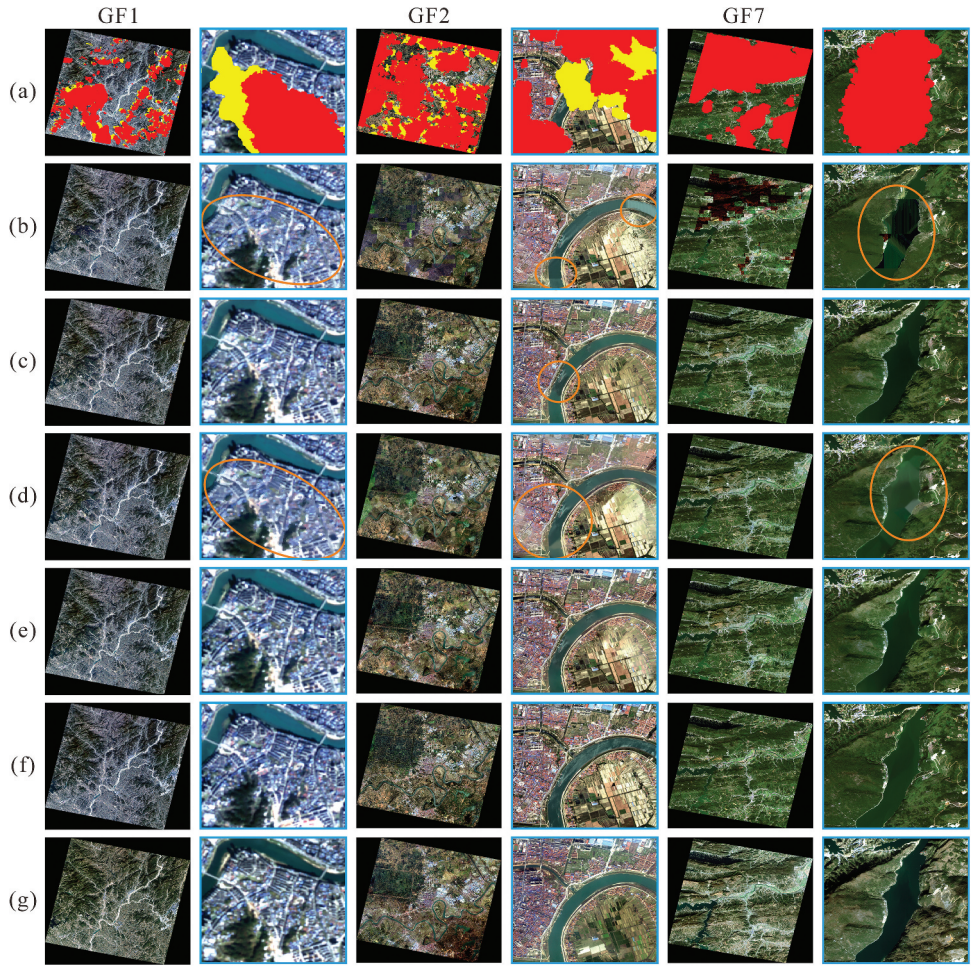


Figure 9. Thick cloud removal results of the different thick cloud removal methods (the red and yellow regions denote the simulated cloud and cloud shadow masks). (a) Target image. (b) The WLR method. (c) The FRARC method. (d) The PGFCTS method. (e) The JTTR-Thick method. (f) Ground truth. (g) Reference image.

cloud-contaminated regions. It demonstrates a superior performance in preserving the spectral-spatial fidelity and obtaining radiometrically consistent, spatially continuous cloud-free images at the target temporal phase.

Subsequently, the effectiveness of the JTTR framework was comprehensively validated through real-data experiments involving both thin and thick clouds. Given the restriction of the manuscript length, Figure 11 only presents parts of the results, whereas the results for all the experiments of each key step are shown in Figures S4–S9. The results demonstrate that the proposed JTTR framework effectively integrates cloud detection, thin cloud removal, and thick cloud removal, achieving a satisfactory performance across multi-sensor images in complex scenarios. Benefiting from the buffer strategy established between the thin cloud and thick cloud removal processes, JTTR achieves high-fidelity and radiometric cloud removal in both large-area clouds and dense clouds images. In addition, the introduction of the boundary-interpolation step in the ICM ensures strong spatial continuity between reconstructed regions detected as clouds and regions corrected by thin cloud removal technique. This enables effective cloud removal regardless of whether clouds are continuously distributed or exhibit distinct boundaries. Moreover, the detailed visual results further confirm that the JTTR framework is effective across diverse landscapes, including urban areas, farmland, and water–land transitions.

4. Discussion

4.1. Accuracy assessment of the cloud and cloud shadow detection

In the proposed JTTR framework, thick clouds, relatively thick thin clouds, and cloud shadows are regarded as regions with completely missing information. Therefore, these three categories of cloud contamination were merged into a single class for the accuracy assessment. As the ultimate goal of JTTR is to achieve automated cloud removal, this study focused on the omission errors, so as to minimize manual intervention caused by missed detections. For the quantitative evaluation, the recall, F1-score, and F2-score were employed as the metrics, which can be calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{16}$$

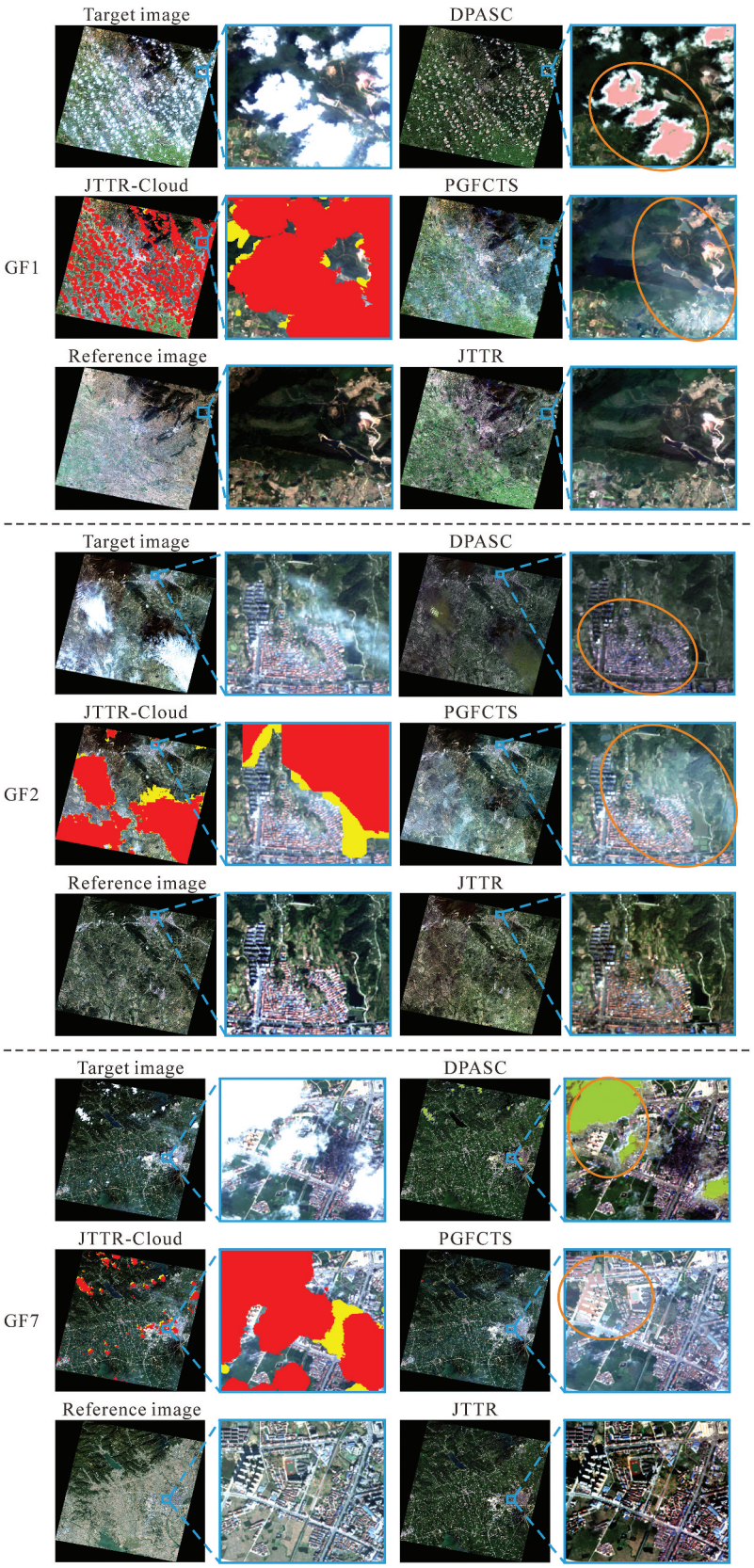


Figure 10. Cloud removal results of the different methods on mixed thin–thick cloud images (the red-colored regions denote the cloud regions, and the yellow-colored regions denote the cloud shadow regions).

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{17}$$

$$\text{F1} - \text{score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{18}$$

$$\text{F2} - \text{score} = \frac{5\times\text{Precision}\times\text{Recall}}{4\times\text{Precision}+\text{Recall}} \tag{19}$$

where TP denotes the number of pixels correctly predicted as cloud, FP indicates the number of pixels

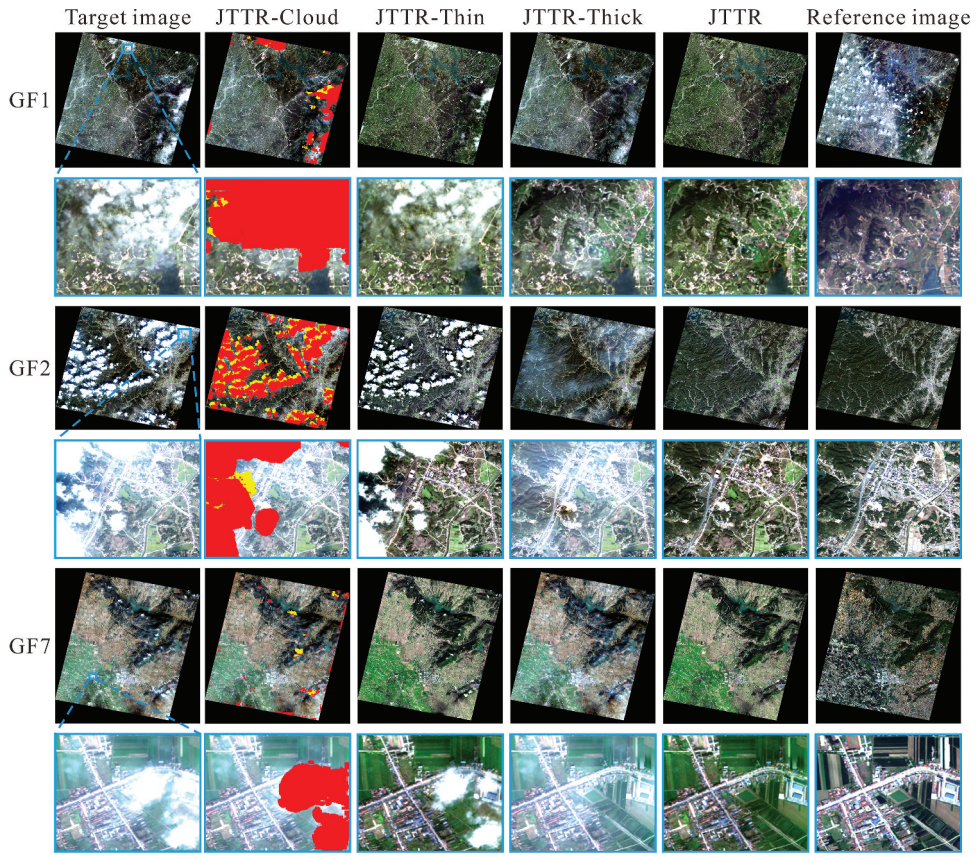


Figure 11. Results of the proposed JTTR framework for cloud detection (JTTR-Cloud), independent thin cloud removal (JTTR-Thin), independent thick cloud removal (JTTR-Thick), and joint cloud removal (JTTR) on mixed thin–thick cloud images.

incorrectly predicted as cloud, and FN stands for the number of pixels incorrectly predicted as non-cloud.

For a comparative analysis, two fully supervised cloud detection networks – RegNetY (Radosavovic et al. 2020) and BoundaryNet (Wu et al. 2022) – and two weakly supervised cloud detection methods – WDCD (Li et al. 2020) and SpecMCD (Zhu et al. 2026) – were selected as fully supervised comparison methods. Table 4 presents the quantitative evaluation of the binary masks obtained by the different cloud detection methods. RegNetY and BoundaryNet, trained solely on thick cloud pixel-level annotations, show poor thin cloud detection, with F1-score and F2-score values below 0.7. Although incorporating pixel-level thin cloud samples significantly improves the overall performance, the recall still remains lower than that of SpecMCD and JTTR-Cloud. Compared to the other methods, the proposed JTTR framework achieves the highest scores across all the metrics, demonstrating its superior capability in thin cloud detection, without introducing excessive false positives.

Due to the space limitations, the visual results of the binary masks generated by the different cloud detection methods on the validation dataset are shown in Figure S10. RegNetY and BoundaryNet perform well in thick cloud detection but fail on the thin clouds. In contrast, RegNetY_Thin and BoundaryNet_Thin, trained with thin cloud pixel-level labels, produce satisfactory results for the dense cloud images but still suffer from notable omission errors in the large-area cloud images. The proposed JTTR-Cloud achieves a high accuracy in both dense and large-area cloud detection, demonstrating a superior overall performance.

Table 4. Quantitative evaluation of the state-of-the-art cloud and cloud shadow detection methods on the GF1-Thin-WHU dataset.

Method	Training dataset	Recall	F1-score	F2-score
RegNetY	GF-Thick-WHU	0.5495	0.6773	0.5923
BoundaryNets		0.5245	0.6376	0.5624
RegNetY_Thin	GF-Thin-WHU	0.8876	0.9088	0.8943
BoundaryNets_Thin		0.8540	0.8807	0.8626
WDCD	GF-Scene-WHU	0.8599	0.7847	0.8185
SpecMCD		0.9309	0.8945	0.9147
JTTR-Cloud	GF-Thick-WHU + GF-Scene-WHU	**0.9417**	**0.9109**	**0.9285**

4.2. Effectiveness of the information cloning model in thin cloud removal

Since the proposed JTTR method employs an ICM to correct the spectral distorted regions in the global thin cloud removal results and adopts a threshold-based strategy to extract the corresponding regions, we analyze the effectiveness of the ICM in thin cloud removal as well as its sensitivity to the spectral distortion mask.

First, we compared the global thin cloud removal results with those adjusted using the ICM in the spectral distorted regions, as shown in Figure 12. The results indicate that the global thin cloud removal results exhibit notable spectral distortions in water regions, bright surfaces, and image boundaries. In contrast, correction with the ICM effectively restores the original spectral characteristics of the land cover, enabling high-fidelity thin cloud removal.

To assess the sensitivity of the ICM to the spectral distortion region mask, we applied 20-pixel morphological erosion and dilation to the extracted masks. We then evaluated the thin cloud removal accuracy for three image scenes using these modified masks and compared the results with those obtained using the original mask, as summarized in Table 5. The quantitative results demonstrate that the performance variations caused by the eroded and dilated masks are minimal across all experiments. Specifically, the original mask consistently yields the highest thin cloud removal accuracy. In contrast, both the 20-pixel erosion and dilation introduce marginal degradations, indicating that the proposed method is robust to perturbations in the mask.

Subsequently, a visual comparison of the optimized results obtained using different masks was conducted, as shown in Figure S11. The results indicate that applying morphological erosion to the mask prevents the complete correction of spectral distortion; only the pixels within the eroded mask are corrected, leaving portions of the distorted regions unaddressed. By comparison, the dilated mask enables more comprehensive correction of the spectral distortion. Although its quantitative performance is slightly inferior to that achieved using the original mask, the visual quality remains essentially consistent. These results collectively demonstrate that the proposed method exhibits strong robustness to moderate mask perturbations, particularly in cases where the mask overestimates the distorted region, which has a negligible impact on the visual quality. Therefore, when an accurate spectral-distortion mask is difficult to obtain, applying an appropriate dilation can be an effective strategy to achieve optimal visual performance.

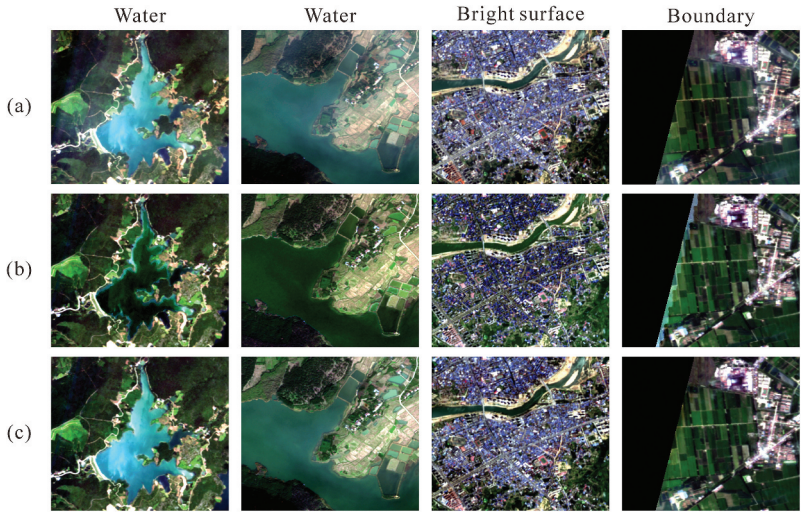


Figure 12. Examples of thin cloud removal results before and after spectral correction using the ICM. (a) Original image. (b) Global thin cloud removal result. (c) JTTR-Thin.

Table 5. Quantitative evaluation of thin cloud removal result using eroded, original, and dilated masks.

Image source	Mask	CC (↑)	RMSE (↓)	SA (↓)
GF-1	20-Pixel erosion	0.8464	0.0276	1.5177
	Original mask	**0.8678**	**0.0267**	**1.5175**
	20-Pixel dilation	0.8468	0.0277	1.5177
GF-2	20-Pixel erosion	0.9152	0.0220	1.5078
	Original mask	**0.9169**	**0.0218**	1.5077
	20-Pixel dilation	0.9150	0.0223	**1.5076**
GF-7	20-Pixel erosion	0.7862	0.0461	1.2688
	Original mask	**0.7989**	**0.0447**	**1.2686**
	20-Pixel dilation	0.7957	0.0486	1.2757

4.3. Impact of reference image quality on JTTR framework

Due to the fact that completely cloud-free reference images are often difficult to obtain, we further analyze the impact of reference image quality on the cloud removal performance of the proposed JTTR framework. As illustrated in Figure 13, although the SpecMCD cloud detection method achieves relatively comprehensive detection of both thin and thick clouds, it inevitably causes the loss of a large amount of available surface information, resulting in numerous cloud-covered regions that cannot be effectively reconstructed. Moreover, the SpecMCD method is still unable to fully cover all thin clouds and haze, which results in locally radiometric overestimation in the SpecMCD-based reconstruction results. Moreover,

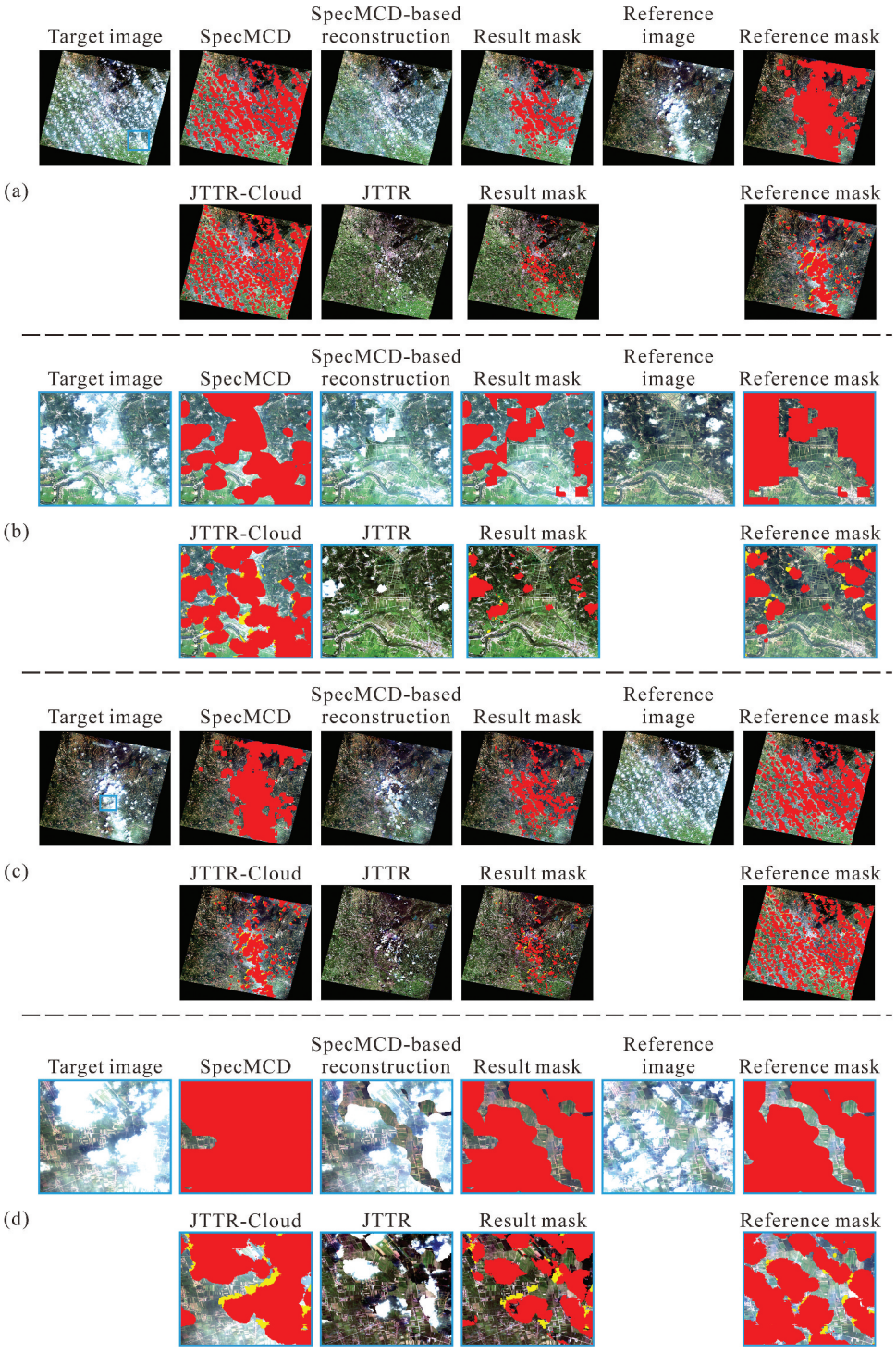


Figure 13. Examples of different cloud removal frameworks using auxiliary images with incomplete complementary information. (b) and (d) show enlarged views of the subregions indicated by the blue boxes in (a) and (c), respectively.

since the SpecMCD method is still unable to fully cover all thin clouds and haze, the corresponding reconstruction suffers from local radiometric overestimation.

In contrast, by integrating a thin cloud removal strategy, the proposed JTTR framework does not require the cloud mask to completely cover thin clouds or haze, thereby preserving a larger amount of available information. Consequently, even when the reference images also contain cloud-contaminated regions – particularly in scenarios where sufficient complementary information is unavailable – the proposed JTTR framework is able to remove thicker clouds and generate cloud removal results with higher overall visual quality and improved radiometric fidelity.

4.4. Time efficiency of the dual-temporal joint thin–thick cloud removal framework

The proposed JTTR framework for HRIs is capable of performing three distinct tasks: cloud and cloud-shadow detection, thin cloud removal, and thick cloud removal. To verify its practical applicability, we conducted multi-sensor experiments to evaluate the computational efficiency of each step and compared the JTTR framework with the state-of-the-art methods for the corresponding tasks, as shown in Table 6. It can be observed that the proposed JTTR framework shows the highest computational efficiency across all the tasks. While the processing time increases with the image size, the relationship is not strictly linear, due to influencing factors such as land surface complexity and cloud coverage. Overall, the proposed JTTR framework requires approximately 500 s for an image of size 6500×6500 pixels, which is sufficient to meet the requirements of practical applications.

4.5. Limitations and future perspectives

Most of the existing JTTR methods generally rely on manual adjustment of a cloud mask (Li et al. 2020) or the incorporation of auxiliary data, such as SAR data (Li et al. 2024), to achieve a satisfactory performance, which greatly limits their practical applicability. In contrast, the proposed JTTR framework can effectively remove both thin and thick clouds using only dual-temporal optical images, without the need for additional data or manual intervention.

Despite its effectiveness, the JTTR framework still has several limitations. First, since this study focused on thin and thick cloud removal, cloud shadows were treated as missing information and removed using thick cloud removal methods. The simple shadow detection method adopted here can detect thick cloud shadows but struggles with thin cloud shadows, thereby degrading the visual quality of the final results, as shown in Figure S12(a). Future work will leverage the radiometric differences in dual-temporal images to further refine the cloud shadow mask or incorporate cloud shadow removal methods, thereby improving the visual quality of the reconstructed imagery.

Furthermore, although the JTTR framework can effectively suppress reflectance overestimation in urban regions under thin clouds, its ability to reconstruct complex textures remains limited, as illustrated in Figure S12(b). To address the loss of texture details in urban areas, future work will explore the recovery of these features using guided filtering or a generative adversarial network.

While the proposed JTTR framework demonstrates strong cloud removal capability, it does not account for radiometric anomalies in the original images. Consequently, the method is relatively sensitive to these anomalous regions, which can exacerbate radiometric inconsistencies in the final results, as illustrated in Figure S12(c). To address this limitation, future work will incorporate a deep learning-based radiometric anomaly detection approach to identify anomalous regions and reconstruct the affected areas.

To validate the cross-sensor cloud removal capability of the proposed JTTR framework, we evaluated its performance on Landsat imagery using cloud masks generated by a dilated Fmask (Qiu, Zhu, and He 2019), as shown in Figure S13. The results indicate that, when reliable cloud masks are provided, the JTTR framework can achieve satisfactory cloud removal performance on Landsat images. However, the cloud detection component of JTTR relies on deep networks, and currently no

Table 6. Time efficiency of the proposed JTTR framework compared with the state-of-the-art methods(s).

		Image (source, size, cloud percentage)		
Task	Method	GF-1, $5342 \times 4679 \times 4$, 44.52%	GF-2, $6761 \times 5969 \times 4$, 45.84%	GF-7, $7060 \times 6696 \times 4$, 5.99%
Cloud detection	SpecMCD	156.3	284.5	293.1
	JTTR-Cloud	100.2	89.2	117.5
Thin cloud removal	DPASC	616.8	1471.7	1073.2
	JTTR-Thin	157.3	327.0	453.6
Thick cloud removal	PGFCTS	156.0	798.0	145.0
	JTTR-Thick	38.1	48.0	39.1
Joint cloud removal	JTTR	295.6	464.2	610.2

cloud detection method can be directly generalized to all sensors without sensor-specific adaptation. Therefore, to fully extend the JTTR framework to other sensors, the cloud detection network must be retrained using data from the target sensor. Future work will focus on developing automated multi-sensor cloud detection strategies that reduce the need for extensive sensor-specific adjustments, thereby enabling the JTTR framework to be more widely applied in multi-sensor images.

5. Conclusions

In this study, we proposed a joint thin and thick cloud removal framework to remove thin and thick clouds in multi-sensor HRIs. By integrating multi-scale cloud detection with an ICM, JTTR enables high-fidelity thin cloud removal from single-temporal images and efficient thick cloud reconstruction using multi-temporal images. Quantitative evaluations on Gaofen satellite data demonstrate that JTTR achieves an average improvement of 1.4% in CC over the state-of-the-art DPASC method for thin cloud removal, and an average improvement of 1.1% in SA over the advanced PGFCTS method for thick cloud removal. In addition, JTTR significantly enhances computational efficiency, delivering more than a 2.4× speedup for thin cloud removal and a 3.7× speedup for thick cloud removal compared with other state-of-the-art approaches. Experimental results from 24 real-data experiments of different sensors further confirm that the proposed framework can effectively achieve joint removal of thin and thick clouds under complex mixed-cloud conditions, highlighting its stability and robustness.

Nevertheless, the framework remains limited in cloud shadow detection and in recovering complex surface textures obscured by thin clouds. Future work will focus on refining cloud shadow detection and thin cloud removal techniques to further improve joint cloud removal performance.

Author contributions

CRediT: **Shaocong Zhu:** Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft; **Tao Jiang:** Investigation, Software, Supervision, Validation, Writing – review & editing; **Zhiwei Li:** Funding acquisition, Resources, Supervision, Writing – review & editing; **Dekun Lin:** Data curation, Formal analysis, Software; **Jingan Wu:** Investigation, Software, Validation; **Huanfeng Shen:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by the National Natural Science Foundation of China [Grant numbers 42130108 and 42101357].

ORCID

Shaocong Zhu http://orcid.org/0009-0003-4334-8373
Tao Jiang http://orcid.org/0000-0001-6893-0982
Zhiwei Li http://orcid.org/0000-0001-5635-8499
Dekun Lin http://orcid.org/0009-0004-1800-287X
Jingan Wu http://orcid.org/0000-0002-4701-937X
Huanfeng Shen http://orcid.org/0000-0002-4140-1869

Data availability statement

The data that support the findings of this study are available upon request, and the executable file of the program is available at https://github.com/whu-ZSC/Executable-package-of-the-JTTR-framework.git.

References

Bi, G., G. Si, Y. Zhao, B. Qi, and H. Lv. 2022. "Haze Removal for a Single Remote Sensing Image Using Low-Rank and Sparse Prior." *IEEE Transactions on Geoscience & Remote Sensing* 60:5615513. https://doi.org/10.1109/TGRS.2021.3135975.

Chen, L., H. Letu, M. Fan, H. Shang, J. Tao, L. Wu, Y. Zhang, et al. 2022. "An Introduction to the Chinese High-Resolution Earth Observation System: Gaofen-1~7 Civilian Satellites." *Journal of Remote Sensing* 2022:9769536. https://doi.org/10.34133/2022/9769536.

Chen, L., S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao. 2021. "Review of Image Classification Algorithms Based on Convolutional Neural Networks." *Remote Sensing* 13 (22): 4712. https://doi.org/10.3390/rs13224712.

Chen, S., X. Chen, J. Chen, P. Jia, X. Cao, and C. Liu. 2016. "An Iterative Haze Optimized Transformation for Automatic Cloud/Haze Detection of Landsat Imagery." *IEEE Transactions on Geoscience & Remote Sensing* 54 (5): 2682–2694. https://doi.org/10.1109/TGRS.2015.2504369.

Chen, Y., Z. Cai, J. Yuan, and L. Wu. 2023. "A Novel Dense-Attention Network for Thick Cloud Removal by Reconstructing Semantic Information." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 16:2339–2351. https://doi.org/10.1109/JSTARS.2023.3236384.

Chen, Y., M. Chen, W. He, J. Zeng, M. Huang, and Y. Zheng. 2024. "Thick Cloud Removal in Multitemporal Remote Sensing Images via Low-Rank Regularized Self-Supervised Network." *IEEE Transactions on Geoscience & Remote Sensing* 62:5506613. https://doi.org/10.1109/TGRS.2024.3358493 .

Deng, J., W. Dong, R. Socher, L. Li, K. Li, and F. Li. 2009. "ImageNet: A Large-Scale Hierarchical Image Database." *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. Miami, FL, USA: IEEE. https://doi.org/10.1109/cvpr.2009.5206848.

Ding, H., F. Xie, L. Qiu, X. Zhang, and Z. Shi. 2024. "Robust Haze and Thin Cloud Removal via Conditional Variational Autoencoders." *IEEE Transactions on*

Geoscience & Remote Sensing 62:5604616. https://doi.org/10.1109/TGRS.2024.3349779.

Farbman, Z., R. Fattal, and D. Lischinski. 2011. "Convolution Pyramids." *Proceedings of the 2011 SIGGRAPH Asia Conference*. Vol. 175, 1–8. Hong Kong, China: New York Association for Computing Machinery. https://doi.org/10.1145/2024156.2024209.

Holland, P. W., and R. E. Welsch. 1977. "Robust Regression Using Iteratively Reweighted Least-Squares." *Communications in Statistics - Theory and Methods* 6 (9): 813–827. https://doi.org/10.1080/03610927708827533.

Hu, G., X. Sun, D. Liang, and Y. Sun. 2014. "Cloud Removal of Remote Sensing Image Based on Multi-Output Support Vector Regression." *Journal of Systems Engineering and Electronics* 25 (6): 1082–1088. https://doi.org/10.1109/JSEE.2014.00124.

Huang, W., Y. Wang, and X. Chen. 2018. "Cloud Detection for High-Resolution Remote-Sensing Images of Urban Areas Using Colour and Edge Features Based on Dual-Colour Models." *International Journal of Remote Sensing* 39 (20): 6657–6675. https://doi.org/10.1080/01431161.2018.1466069.

Huang, Y., and B. Dewitt. 2025. "Differentiating Thick Clouds from Thin Clouds by Using Intensity Inhomogeneity." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 18:5782–5792. https://doi.org/10.1109/JSTARS.2025.3539322.

Jiang, T., H. Shen, H. Li, and L. Xu. 2025. "Combining Dark Pixels and Spectral Characteristics for Thin Cloud Removal in High-Resolution Remote Sensing Images." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 18:20499–20512. https://doi.org/10.1109/JSTARS.2025.3596135.

Jiang, T., H. Shen, H. Li, C. Zhang, L. Xu, and D. Lin. 2024. "A Fast and Robust Cirrus Removal Method for Landsat 8/9 Images." *International Journal of Applied Earth Observation and Geoinformation* 128:103691. https://doi.org/10.1016/j.jag.2024.103691.

Kotaridis, I., and M. Lazaridou. 2021. "Remote Sensing Image Segmentation Advances: A Meta-Analysis." *ISPRS Journal of Photogrammetry & Remote Sensing* 173:309–322. https://doi.org/10.1016/j.isprsjprs.2021.01.020 .

Lee, Y., S. Min, J. Yoon, J. Ha, S. Jeong, S. Ryu, and M. Ahn. 2025. "Application of Deep Learning in Cloud Cover Prediction Using Geostationary Satellite Images." *GIScience and Remote Sensing* 62 (1): 2440506. https://doi.org/10.1080/15481603.2024.2440506.

Li, X., Z. Li, R. Feng, S. Luo, C. Zhang, M. Jiang, and H. Shen. 2020. "Generating High-Quality and High-Resolution Seamless Satellite Imagery for Large-Scale Urban Regions." *Remote Sensing* 12 (1): 81. https://doi.org/10.3390/rs12010081.

Li, X., X. Zhao, F. Wang, and P. Ren. 2024. "HF-T2CR: High-Fidelity Thin and Thick Cloud Removal in Optical Satellite Images Through SAR Fusion." *IEEE Transactions on Geoscience & Remote Sensing* 62:4105813. https://doi.org/10.1109/TGRS.2024.3416128 .

Li, Y., W. Chen, Y. Zhang, C. Tao, R. Xiao, and Y. Tan. 2020. "Accurate Cloud Detection in High-Resolution Remote Sensing Imagery by Weakly Supervised Deep Learning." *Remote Sensing of Environment* 250:112045. https://doi.org/10.1016/j.rse.2020.112045.

Li, Z., H. Shen, H. Li, G. Xia, P. Gamba, and L. Zhang. 2017. "Multi-Feature Combined Cloud and Cloud Shadow Detection in GaoFen-1 Wide Field of View Imagery." *Remote Sensing of Environment* 191:342–358. https://doi.org/10.1016/j.rse.2017.01.026.

Li, Z., Q. Weng, Y. Zhou, P. Dou, and X. Ding. 2024. "Learning Spectral-Indices-Fused Deep Models for Time-Series Land Use and Land Cover Mapping in Cloud-Prone Areas: The Case of Pearl River Delta." *Remote Sensing of Environment* 308:114190. https://doi.org/10.1016/j.rse.2024.114190.

Li, Z., S. Xu, and Q. Weng. 2024. "Beyond Clouds: Seamless Flood Mapping Using Harmonized Landsat and Sentinel-2 Time Series Imagery and Water Occurrence Data." *ISPRS Journal of Photogrammetry & Remote Sensing* 216:185–199. https://doi.org/10.1016/j.isprsjprs.2024.07.022.

Lv, H., Y. Wang, and Y. Shen. 2016. "An Empirical and Radiative Transfer Model Based Algorithm to Remove Thin Clouds in Visible Bands." *Remote Sensing of Environment* 179:183–195. https://doi.org/10.1016/j.rse.2016.03.034.

Makarau, A., R. Richter, R. Müller, and P. Reinartz. 2014. "Haze Detection and Removal in Remotely Sensed Multispectral Imagery." *IEEE Transactions on Geoscience & Remote Sensing* 52 (9): 5895–5905. https://doi.org/10.1109/TGRS.2013.2293662.

McFeeters, S. 1996. "The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features." *International Journal of Remote Sensing*. 17:1425–1432. https://doi.org/10.1080/01431169608948714.

Pires, B., K. Singh, and J. Moura. 2011. "Approximating Image Filters with Box Filters." *2011 IEEE International Conference on Image Processing*, 85–88. Brussels, Belgium: IEEE. https://doi.org/10.1109/ICIP.2011.6116693.

Qiu, S., Z. Zhu, and B. He. 2019. "Fmask 4.0: Improved Cloud and Cloud Shadow Detection in Landsats 4-8 and Sentinel-2 Imagery." *Remote Sensing of Environment* 231:111205. https://doi.org/10.1016/j.rse.2019.05.024.

Radosavovic, I., R. P. Kosaraju, R. Girshick, K. He, and P. Dollár. 2020. "Designing Network Design Spaces." *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10425–10433. Seattle, WA, USA: IEEE. https://doi.org/10.1109/CVPR42600.2020.01044.

Shao, Z., Y. Pan, C. Diao, and J. Cai. 2019. "Cloud Detection in Remote Sensing Images Based on Multiscale Features-Convolutional Neural Network." *IEEE Transactions on Geoscience & Remote Sensing* 57 (6): 4062–4076. https://doi.org/10.1109/TGRS.2018.2889677.

Shen, H., X. Li, Q. Cheng, C. Zeng, G. Yang, H. Li, and L. Zhang. 2015. "Missing Information Reconstruction of Remote Sensing Data: A Technical Review." *IEEE Geoscience and Remote Sensing Magazine* 3 (3): 61–85. https://doi.org/10.1109/MGRS.2015.2441912.

Shen, H., C. Zhang, H. Li, Q. Yuan, and L. Zhang. 2020. "A Spatial-Spectral Adaptive Haze Removal Method for Visible Remote Sensing Images." *IEEE Transactions on Geoscience & Remote Sensing* 58 (9): 6168–6180. https://doi.org/10.1109/TGRS.2020.2974807.

Shepard, D. 1968. "A Two-Dimensional Interpolation Function for Irregularly-Spaced Data." In *Proceedings of the 1968 23rd ACM National Conference*, edited by R. B. Blue and A. M. Rosenberg, 517–524. New York: Association for Computing Machinery. https://doi.org/10.1145/800186.810616.

Tao, C., S. Fu, J. Qi, and H. Li. 2022. "Thick Cloud Removal in Optical Remote Sensing Images Using a Texture Complexity Guided Self-Paced Learning Method." *IEEE*

Transactions on Geoscience & Remote Sensing 60:5619612. https://doi.org/10.1109/TGRS.2022.3157917 .

Wang, Q., J. Li, X. Tong, and P. Atkinson. 2024. "TSI-Siamnet : A Siamese Network for Cloud and Shadow Detection Based on Time-Series Cloudy Images." *ISPRS Journal of Photogrammetry & Remote Sensing* 213:107–123. https://doi.org/10.1016/j.isprsjprs.2024.05.022.

Wen, D., X. Huang, F. Bovolo, J. Li, X. Ke, A. Zhang, and J. Benediktsson. 2021. "Change Detection from Very-High-Spatial-Resolution Optical Remote Sensing Images: Methods, Applications, and Future Directions." *IEEE Geoscience and Remote Sensing Magazine* 9:68–101. https://doi.org/10.1109/MGRS.2021.3063465.

Wright, N., J. Duncan, J. Callow, S. Thompson, and R. George. 2024. "CloudS2Mask: A Novel Deep Learning Approach for Improved Cloud and Cloud Shadow Masking in Sentinel-2 Imagery." *Remote Sensing of Environment* 306:114122. https://doi.org/10.1016/j.rse.2024.114122.

Wright, N., J. Duncan, J. Callow, S. Thompson, and R. George. 2025. "Training Sensor-Agnostic Deep Learning Models for Remote Sensing: Achieving State-of-the-Art Cloud and Cloud Shadow Identification with OmniCloudMask." *Remote Sensing of Environment* 322:114694. https://doi.org/10.1016/j.rse.2025.114694.

Wu, J., T. Li, L. Lin, and C. Zeng. 2024. "Progressive Gap-Filling in Optical Remote Sensing Imagery Through a Cascade of Temporal and Spatial Reconstruction Models." *Remote Sensing of Environment* 311:114245. https://doi.org/10.1016/j.rse.2024.114245.

Wu, K., Z. Xu, X. Lyu, and P. Ren. 2022. "Cloud Detection with Boundary Nets." *ISPRS Journal of Photogrammetry & Remote Sensing* 186:218–231. https://doi.org/10.1016/j.isprsjprs.2022.02.010.

Wu, K., Z. Xu, X. Lyu, and P. Ren. 2023. "Cross-Supervised Learning for Cloud Detection." *GIScience and Remote Sensing* 60 (1): 2147298. https://doi.org/10.1080/15481603.2022.2147298.

Wu, P., Z. Yin, C. Zeng, S. Duan, F. Gottsche, X. Ma, X. Li, H. Yang, and H. Shen. 2021. "Spatially Continuous and High-Resolution Land Surface Temperature Product Generation: A Review of Reconstruction and Spatiotemporal Fusion Techniques." *IEEE Geoscience and Remote Sensing Magazine* 9 (3): 112–137. https://doi.org/10.1109/MGRS.2021.3050782.

Xie, Z., Y. Zhao, R. Jiang, M. Zhang, G. Hammer, S. Chapman, J. Brider, and A. B. Potgieter. 2024. "Seasonal Dynamics of Fallow and Cropping Lands in the Broadacre Cropping Region of Australia." *Remote Sensing of Environment* 305:114070. https://doi.org/10.1016/j.rse.2024.114070.

Xu, K., A. Wang, Z. Zhu, W. Ju, and W. Huang. 2025. "Supervised Alignment Perception and Modality Difference Selection Enhanced SAR-Optical Fusion for Cloud Removal." *IEEE Transactions on Geoscience & Remote Sensing* 63:4706015. https://doi.org/10.1109/TGRS.2025.3594507.

Xu, L., H. Li, H. Shen, C. Zhang, and L. Zhang. 2024. "A Thin Cloud Blind Correction Method Coupling a Physical Model with Unsupervised Deep Learning for Remote Sensing Imagery." *ISPRS Journal of Photogrammetry & Remote Sensing* 218:246–259. https://doi.org/10.1016/j.isprsjprs.2024.09.008.

Xu, M., F. Deng, S. Jia, X. Jia, and A. Plaza. 2022. "Attention Mechanism-Based Generative Adversarial Networks for Cloud Removal in Landsat Images." *Remote Sensing of Environment* 271:112902. https://doi.org/10.1016/j.rse.2022.112902.

Xu, M., X. Jia, M. Pickering, and S. Jia. 2019. "Thin Cloud Removal from Optical Remote Sensing Images Using the Noise Adjusted Principal Components Transform." *ISPRS Journal of Photogrammetry & Remote Sensing* 149:215–225. https://doi.org/10.1016/j.isprsjprs.2019.01.025.

Yun, Y., J. Jung, and Y. Han. 2024. "Cloud Restoration of Optical Satellite Imagery Using Time-Series Spectral Similarity Group." *GIScience and Remote Sensing* 61 (1): 2324553. https://doi.org/10.1080/15481603.2024.2324553.

Zeng, C., H. Shen, and L. Zhang. 2013. "Recovering Missing Pixels for Landsat ETM Plus SLC-Off Imagery Using Multi-Temporal Regression Analysis and a Regularization Method." *Remote Sensing of Environment* 131:182–194. https://doi.org/10.1016/j.rse.2012.12.012.

Zhang, C., H. Li, and H. Shen. 2021. "A Scattering Law Based Cirrus Correction Method for Landsat 8 OLI Visible and Near-Infrared Images." *Remote Sensing of Environment* 253:112202. https://doi.org/10.1016/j.rse.2020.112202.

Zhang, H., Q. Huang, H. Zhai, and L. Zhang. 2021. "Multi-Temporal Cloud Detection Based on Robust PCA for Optical Remote Sensing Imagery." *Computers and Electronics in Agriculture* 188:106342. https://doi.org/10.1016/j.compag.2021.106342.

Zhang, K., H. Nie, W. Li, J. Wang, B. Tang, L. Wang, and Z. Fu. 2025. "A Multitemporal Remote Sensing Thick Cloud Removal Network Based on Implicit Reconstruction." *International Journal of Remote Sensing* 46 (4): 1574–1593. https://doi.org/10.1080/01431161.2024.2433755.

Zhang, Q., Q. Yuan, Z. Li, F. Sun, and L. Zhang. 2021. "Combined Deep Prior with Low-Rank Tensor SVD for Thick Cloud Removal in Multitemporal Images." *ISPRS Journal of Photogrammetry & Remote Sensing* 177:161–173. https://doi.org/10.1016/j.isprsjprs.2021.04.021.

Zhang, Y., B. Guindon, and J. Cihlar. 2002. "An Image Transform to Characterize and Compensate for Spatial Variations in Thin Cloud Contamination of Landsat Images." *Remote Sensing of Environment* 82 (2–3): 173–187. https://doi.org/10.1016/S0034-4257(02)00034-2.

Zheng, Z., Y. Zhong, J. Wang, A. Ma, and L. Zhang. 2021. "Building Damage Assessment for Rapid Disaster Response with a Deep Object-Cased Semantic Change Detection Framework: From Natural Disasters to Man-Made Disasters." *Remote Sensing of Environment* 265:112636. https://doi.org/10.1016/j.rse.2021.112636.

Zhou, D., J. Xiao, S. Bonafoni, C. Berger, K. Deilami, Y. Zhou, S. Frolking, R. Yao, Z. Qiao, and J. A. Sobrino. 2019. "Satellite Remote Sensing of Surface Urban Heat Islands: Progress, Challenges, and Perspectives." *Remote Sensing* 11 (1): 48. https://doi.org/10.3390/rs11010048.

Zhu, S., Z. Li, X. Li, and H. Shen. 2026. "Weakly Supervised Cloud Detection Combining Spectral Features and Multi-Scale Deep Network." *GIScience and Remote Sensing* 63 (1): 2626022. https://doi.org/10.1080/15481603.2026.2626022.

Zhu, S., Z. Li, and H. Shen. 2024. "Transferring Deep Models for Cloud Detection in Multisensor Images via Weakly Supervised Learning." *IEEE Transactions on*

Geoscience & Remote Sensing 62:5609518. https://doi.org/10.1109/TGRS.2024.3358824.

Zhu, S., Z. Li, H. Shen, and D. Lin. 2023. "A Fast Two-Step Algorithm for Large-Area Thick Cloud Removal in High-Resolution Images." *Remote Sensing Letters* 14 (1): 1–9. https://doi.org/10.1080/2150704X.2022.2152753.

Zhu, Z., S. Wang, and C. Woodcock. 2015. "Improvement and Expansion of the Fmask Algorithm: Cloud, Cloud Shadow, and Snow Detection for Landsats 4-7, 8, and Sentinel 2 Images." *Remote Sensing of Environment* 159:269–277. https://doi.org/10.1016/j.rse.2014.12.014.