Hindawi
International Journal of Intelligent Systems
Volume 2023, Article ID 9953198, 17 pages
https://doi.org/10.1155/2023/9953198

Research Article

Bishift Networks for Thick Cloud Removal with Multitemporal Remote Sensing Images

Chaojun Long,[1] Xinghua Li ,[1,2] Yinghong Jing,[3] and Huanfeng Shen[3]

[1]*School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China*
[2]*Hubei Luojia Laboratory, Wuhan, China*
[3]*School of Resource and Environmental Sciences, Wuhan University, Wuhan, China*

Correspondence should be addressed to Xinghua Li; lixinghua5540@whu.edu.cn

Received 14 September 2022; Accepted 22 October 2022; Published 21 February 2023

Academic Editor: Gennaro Vessio

Because of the presence of clouds, the available information in optical remote sensing images is greatly reduced. These temporal-based methods are widely used for cloud removal. However, the temporal differences in multitemporal images have consistently been a challenge for these types of methods. Towards this end, a bishift network (BSN) model is proposed to remove thick clouds from optical remote sensing images. As its name implies, BSN is combined of two dependent shifts. Moment matching (MM) and deep style transfer (DST) are the first shift to preliminarily eliminate temporal differences in multitemporal images. In the second shift, an improved shift net is proposed to reconstruct missing information under cloud covers. It introduces multiscale feature connectivity with shift connections and depthwise separable convolution (DSC), which can capture local details and global semantics effectively. Through experiments with Sentinel-2 images, it has been demonstrated that the proposed BSN has great advantages over traditional methods and state-of-the-art methods in cloud removal.

1. Introduction

Optical remote sensing image is an important data source for large-area research and application. However, clouds and shadows make it difficult to obtain high-quality optical images. Mostly, clouds are detrimental to the practical applications of remote sensing images. When the ground is covered by thin clouds, the sensor captures a mixture of thin clouds and ground objects. When the ground is covered by thick clouds, thick clouds and shadows completely obstruct the ground, and the optical sensor usually cannot capture ground information. Clouds (especially thick clouds) and their shadows have long been considered a difficult problem in remote sensing image processing and applications.

In the past decades, great efforts have been made to remove clouds from remote sensing images. Depending on the type of data source, a wide variety of cloud removal methods are classified into four categories: spatial-based, spectral-based, temporal-based, and hybrid methods [1]. Spatial-based methods are not suitable for large-size and complex cloud-covered images. Spectral-based methods work well with thin clouds, but not with thick clouds. Therefore, the most significant and widely used methods are temporal-based methods and hybrid methods.

Remote sensing platforms usually have a fixed visit period and can acquire images of the same area at different time intervals. Thus, they provide a reliable reference data source for cloud removal with multitemporal remote sensing images. Temporal-based methods introduce additional observations from multitemporal images to reconstruct cloud-covered regions rather than using only the cloudy image itself. They can alleviate temporal differences caused by observational conditions and regular changes in geographic features (e.g., phenological changes). The representative methods include temporal replacement methods [2], temporal filter methods [3, 4], and temporal learning methods [5–8]. Hybrid methods attempt to make better use of correlations among spatial, spectral, and temporal domains using the same or different sensor data. They can take full consideration of the advantages of the above three methods

to achieve better cloud removal results. The hybrid methods include joint spatiotemporal methods [9] and joint spatio-spectral methods [10]. Additionally, multisource data (SAR and optical images) [11, 12] are also used as auxiliary images to improve the effects of cloud removal. Generally, hybrid methods have difficulty in multisource data acquisition. Relatively speaking, temporal-based methods are more popular and available for cloud removal from optical images.

1.1. Related Work. During the past decades, researchers have developed a number of temporal-based methods. Chen et al. proposed a Savitzky–Golay filter to remove noise from images [3] but failed to remove thick clouds. Based on the idea of replacement, Lin et al. proposed a multitemporal information cloning method for removing thick clouds [2]. An automatic cloud removal method based on Poisson blending using temporal similarity of multitemporal images is applied [13]. Li et al. reconstructed cloud-covered regions from remote sensing images within a framework of sparse representation (PM-MTGSR) [14]. It can effectively exploit nonlocal correlations to reconstruct missing information in optical remote sensing images [5]. An improved Bayesian dictionary-learning algorithm based on compressed sensing was proposed to restore remote sensing images [15]. Then, multitemporal dictionary learning is expanded into the recovery of quantitative data contaminated by thick clouds and shadows [7]. To recover the original information covered by clouds and accompanying shadows, a non-negative matrix factorization error correction method was proposed [8]. Although researchers have made significant developments in multitemporal-based methods, traditional methods have some limitations. For example, they cannot take advantage of the deep correlation of the image to remove clouds, which is especially important for remote sensing images. Traditional methods often underperform when dealing with clouds and feature boundaries, and reconstructed features are not sufficiently accurate. Furthermore, temporal-based methods are extremely dependent on the quality of multitemporal reference images, and if reference temporal images are contaminated by clouds, the results of cloud removal are significantly influenced.

Recently, owing to its powerful nonlinear representation capability [16, 17], deep learning has attracted more and more attention in missing information reconstruction in remote sensing images. Deep learning-based methods can exploit deep correlations between multitemporal images compared to traditional methods. Sandhan and Choi successfully used a generative model to specifically remove extremely thin high-altitude clouds [18]. However, it cannot effectively process more heavily obscured images. Thus, researchers moved to thick cloud removal based on convolutional neural networks (CNNs). For example, Zhang et al. proposed a CNN-based spatial-temporal-spectrum (STS) framework [19] to generate accurate reconstruction results. They also improved the STS framework to combine the global-local spatiotemporal information for cloud removal [20]. Considering the specificity of different types of information, a traditional CNN-based joint content, texture, and spectrum generation network was proposed for cloud removal [21]. Based on the idea of integration, Ji et al. proposed an integrated cloud detection and removal method with cascaded CNNs [22]. Moreover, a novel gated convolutional network (GCN) has stronger differentiation ability than general convolutional networks for cloud removal [23]. Generative adversarial networks (GANs) [24] are also used for cloud removal. Yu et al. [25] proposed a GAN with a contextual attention method (GAN-CA) for reconstructing information from cloud-obscured images. It can explicitly focus on relevant feature blocks at distant spatial locations, but the ability to process high-resolution tasks is insufficient. A trainable Spatio-Temporal Generator Adversarial Network (STGAN) [26] casts cloud removal as a conditional image synthesis. Gao et al. proposed the SAR-opt-GAN method, which joins SAR and optical data to facilitate cloud removal [27]. The image translation approach has been adopted by researchers as a recent idea for SAR-assisted cloud removal [28]. Deep learning-based methods are robust and stable and therefore less susceptible to the influence of the dataset quality than traditional methods. Deep learning-based methods have a large number of model parameters and therefore have a relatively high accuracy due to pretraining requirements. The current cloud removal methods based on deep learning have made great advancements, but there are still some problems. For example, due to the large size of remote sensing datasets, deep learning methods cannot process them directly and researchers have to crop them to smaller images. Moreover, if the ancillary images in the datasets have large temporal differences, the results of cloud removal will be unsatisfactory. Therefore, the datasets need a high correlation of spatial, spectral, and temporal aspects with target images. However, deep learning-based methods have rarely focused on temporal differences among multitemporal images. This is a great obstacle to utilizing the temporal correlation of multitemporal images.

1.2. Contributions. In order to restrain the temporal differences in multitemporal images and cooperate with the advantages of deep learning, a novel bishift network (BSN) model with double shifts is proposed in this paper. BSN can improve the correlation between the target image and datasets somewhat, facilitating the model to learn more effective information required for cloud removal.

The first shift includes moment matching (MM) and deep style transfer (DST). Multitemporal images (reference images) are statistically normalized to cloud-covered images by traditional MM and then processed by DST to further eliminate temporal differences. MM utilizes the mean and variance in optical remote sensing images to match features from datasets to target images. In the stage of DST, the transferred images are constrained to be represented by locally affine color transformations to prevent distortions.

The second shift takes full advantage of a proposed reconstruction network to reduce the temporal differences of images once again for better cloud removal. It is an improved version of Shift-Net [29] with shift connections and

depthwise separable convolution (DSC). Shift connections reduce information loss during reconstruction and effectively improve the accuracy of cloud removal. DSC can partially reduce the number of parameters of the model and improve training efficiency. The proposed reconstruction network can better capture the local details and global semantics of images. By two successive shift operations, the temporal differences in the multitemporal images can be suppressed effectively. Eventually, high-quality cloud-removed images can be obtained.

The rest of this paper is organized as follows: The proposed BSN is introduced in Section 2. The effectiveness of BSN is tested by simulated experiments and real experiments in Section 3. Finally, Section 4 summarizes the article.

2. Proposed Method

BSN is a further improvement on our previous research [30]. It consists of two shifts, as shown in Figure 1. The first shift is to preprocess multitemporal images (reference images) with MM and DST to obtain reliable preliminary results. In the second shift, the reconstruction network from improved Shift-Net will reconstruct the target image covered by clouds and shadows to generate accurate cloud-free images. BSN requires at least one temporal reference image to ensure the capability of the network. It is also capable of dealing with many multitemporal reference images. This section will introduce BSN in detail.

2.1. First Shift. It is a challenge that multitemporal remote sensing images have different temporal characteristics. To this end, the first shift of the proposed BSN is used to normalize multitemporal reference images to cloudy images (target images). The first shift contains statistical MM and DST for reducing temporal differences in reference images.

2.1.1. Moment Matching. MM is a mathematical statistical method, commonly used in remote sensing image denoising [31] and difference elimination [32]. In BSN, MM is used to normalize multitemporal reference images. It should be noted that, because parts of the information of the cloudy image are covered by clouds, the statistical values of target images are from cloudless regions. Thus, the cloud mask is used to distinguish between the cloud and cloud-free pixels. The MM formula is as follows:

$$I_o(x, y) = \frac{\sigma_T}{\sigma_R} I_R(x, y) - \frac{\sigma_T}{\sigma_R}\mu_R + \mu_T, \tag{1}$$

where $I(x, y)$ is the pixel at the position (x, y) and μ and σ represent the mean and variance of the entire image, respectively. The abbreviations O, R, T represent the output image, the reference image, and the target image, respectively.

Before MM, there are feature differences between multitemporal reference images due to temporal, weather, and light intensity. For example, there are differences in the greenness degree of vegetation and the intensity of light reflections. MM can reduce these differences to a certain extent and contribute to the efficiency of style transfer.

2.1.2. Deep Style Transfer. Style transfer is an evolving research field. Gatys et al. first proposed a style transfer algorithm based on neural networks [33, 34]. In neural style transfer, two input images include a "content image" and a "style image." The content of the image is defined as the feature response from the pretrained CNN, while the style is a summary feature statistic. The task of style transfer is to convert the image to an artistic style by changing the style. However, preserving the required semantic content is a key challenge, which is achieved by generally changing the weights of pretrained CNN models [35]. In this paper, DST is applied to multitemporal reference images for normalizing temporal differences.

The traditional style transfer method is effective for simple styles, such as global color changes and tone curves. It generates the output image by using the reference style image on the content image. The general objective function is

$$L_{\text{total}} = \sum_{l=1}^{N} \alpha_l L_C^l + \gamma \sum_{l=1}^{N} \beta_l L_S^l, \tag{2}$$

where L_{total} is the total loss in deep CNN, α_l and β_l are the loss weights of the content image C and the style image S in the l-th convolutional layer of the total N-layer, respectively, γ is the weight of all style images, and L_C^l and L_C^l represent the content loss and style loss in the style transfer process, respectively.

Before BSN, general fast style transfer had been used for temporal difference elimination [30]. However, regular style transfer methods are not suitable for realistic style and complex transfers. Style transfer is particularly difficult for remote sensing images with complex features, large scale, and high resolution. To solve this problem, the style transformation from input to output is constrained to be a local affine projection in RGB color space. It has been demonstrated that the local style transfer algorithm based on spatial color mapping is more expressive [36]. Therefore, DST is adopted [37].

DST can ensure the preservation of image structure, semantic accuracy, and transfer faithfulness, which is beneficial to complex remote sensing images. In optimization, a realistic regularization term L_m is proposed in the objective function. The reconstructed image is constrained to be represented by locally affine color transformations of the input to prevent distortions. An optional guidance is introduced to the style transfer process based on semantic segmentation of reference images. Styles are only transferred between features of reference images and similar features of target images. The style transfer process minimizes the content-mismatch problem, which greatly improves the photorealism of output images. This indicates that the multiple classes of features and complex textures of remote sensing images will not be lost in the process of style transfer. The method expects no cloud information in the transferred

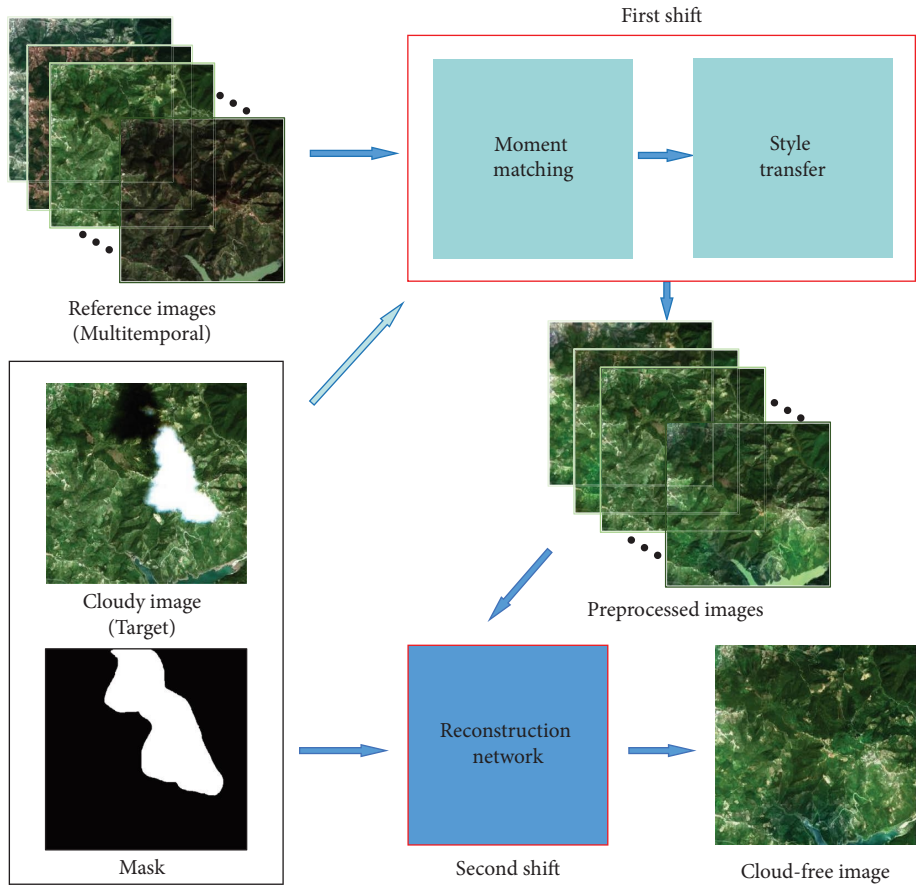


Figure 1: The framework of the proposed BSN.

images. Therefore, regions covered by the cloud are useless. The mask is added to the input image as an additional channel to distinguish between cloud-covered and cloud-free regions. Then, the neural style algorithm is enhanced by concatenating segmentation channels and updating the augmented style loss L_{S+}. More details can be found in [37].

The improved objective function of DST is

$$L_{\text{total}} = \sum_{l=1}^{L} \alpha_l L_C^l + \gamma \sum_{l=1}^{L} \beta_l L_{S+}^l + \delta L_m, \tag{3}$$

where δ indicates the corresponding weight of the realistic regularization term and L_C^l and L_{S+} represent the content loss and the augmented style loss, respectively.

The traditional fast style transfer method requires a decision on the content and style of the image. Therefore, the original content clarity of the image cannot be preserved, while the style is completely transferred. Fast style transfer can only achieve rough style transfer, with little change in its content. This leads to the fact that the transferred image is not close enough to the style image (cloudy target image). For example, in Figure 2(d), vegetation and bareland are in a gradual transition in the cloudy image, but the difference is obvious in the reference image. Although the images generated by fast style transfer are similar in style to cloudy images, vegetation and bareland are clearly distinguished. It is obviously inappropriate as a preliminary result of cloud removal. However, DST can realize transfer between similar features, which makes the transfer process more accurate. As a result of DST, the features of rivers, vegetation, and bareland are highly consistent with cloudy target images in Figure 2(e).

In the first shift, MM and DST are applied to process multitemporal reference remote sensing images, gradually removing temporal differences among images so that they can contain more information similar to the target image. The preprocessed images provide more available information for the subsequent work and are therefore reliable. In the subsequent experimental sections, ablation experiments are conducted to demonstrate the effects of preprocessed measures.

2.2. Second Shift. The first shift preliminarily solves the temporal difference problem of multitemporal images. To obtain more accurate cloud-removed images, the second shift recovers cloud-covered information by the proposed reconstruction network. This network is performed on the feature domain of the deep encoder learned end-to-end from the training data, which is different from the traditional exemplar-based method [38] to fill pixels or patches.

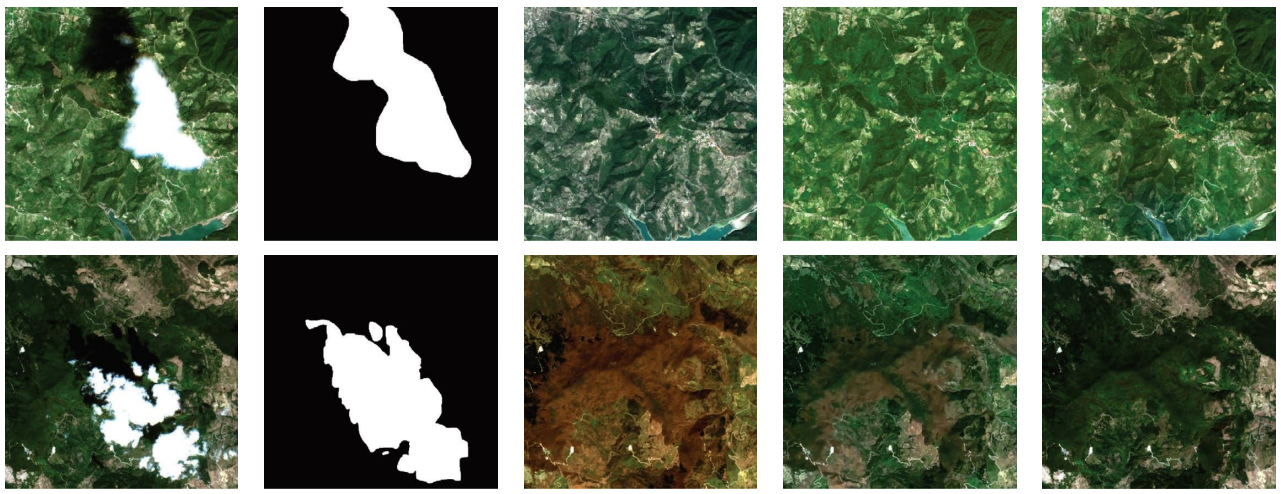


Figure 2: The results of style transfer. (a) Cloudy image. (b) Mask. (c) Reference image. (d) Fast style transfer. (e) Deep style transfer.

2.2.1. Reconstruction Network. Adversarial learning has been adopted in low-level vision [39], image generation [40, 41], and image inpainting [25] and exhibits its superiority in restoring fine details and photo-realistic textures. In the stage of the second shift, the reconstruction network architecture is based on GAN, which includes a generator and a discriminator, as shown in Figure 3. The generator learns distribution of data and improves the ability to remove clouds from images by adversarial learning of the generator and discriminator. After training, the generator can transform the input cloud-covered image into a cloud-free image.

The reconstruction network in BSN is inspired by Shift-Net [29]. As shown in Figure 4, its generator includes eight convolutional modules and corresponding deconvolutional modules. Convolutional modules are the encoders of the network, and deconvolutional modules are decoders. Encoders and corresponding decoders are associated with skip connections [42]. Skip connections fuse different scales and different levels of features, which can effectively reduce gradient disappearance and network degradation problems. Furthermore, skip connections can facilitate the use of information of convolution and deconvolution layers, which is valuable for capturing local visual details of cloud removal tasks of remote sensing images. In order to further enhance the generator's ability to capture local details, shift connections are introduced. Shift connections demonstrate even greater advantages through deep feature rearrangement. The appropriate placement of the shift connection layer ensures both the computation time and the reconstruction performance of the network.

Based on Shift-Net, BSN introduces DSC to improve the network's capabilities. For the first and second modules of the encoder, a combination of DSC [43], batch normalization (BN) layer, and leaky ReLU activation function is adopted. DSC splits the convolution operation into depthwise convolution and 1×1 pointwise convolution operation. DSC of multichannel feature maps of the previous layer is performed by first splitting them all into single-channel feature maps. Then, separate single-channel convolutions are performed and restacked. DSC reduces the computation in convolution. Therefore, for smaller models, the ability of the model may be significantly reduced if 2D convolution is replaced by DSC. As a result, it may be suboptimal. However, if used properly, DSC can help achieve efficiency gains without degrading the model performance. In processing remote sensing images, it is beneficial to minimize network parameters to improve training efficiency due to the large size of the remote sensing image, and it has been found experimentally that DSC can improve the accuracy of cloud removal, as is shown later in the experimental section. Experience shows that the DSC layers in the first and second layers not only reduce the burden of the network but also improve the effectiveness of feature extraction. In addition, it can effectively generate clear, detailed, and photo-realistic images.

The discriminator judges whether input images are real images in the dataset or generated fake cloud-free images. The structure of the discriminator is shown in Figure 5. The discriminator of GAN for cloud removal consists of five convolutional modules. The first module contains a convolutional layer and a leaky ReLU layer. The second to fourth modules are a combination of convolutional, leaky ReLU, and instance batch norm (IBN) layers. The last module is a single convolutional layer. The input image can be discriminated by the multilayer convolution module of the discriminator. The discriminator determines whether the input image is a false cloudless image or a real image generated by using the generator to evaluate the cloud removal ability of the generator. More information on the parameters of the generator and the discriminator can be found in Table 1.

2.2.2. Objective Function. For remote sensing images with complex features, a suitable compound loss function is

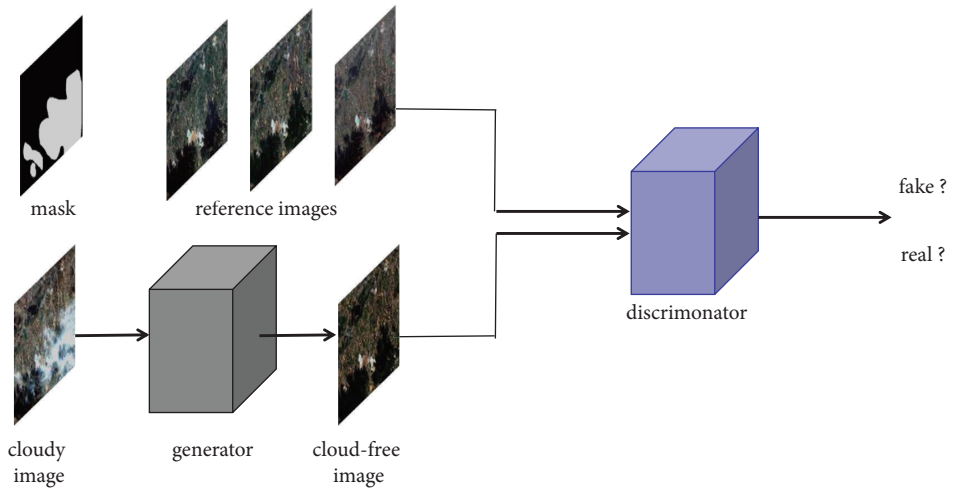


Figure 3: Architecture of the reconstruction network.

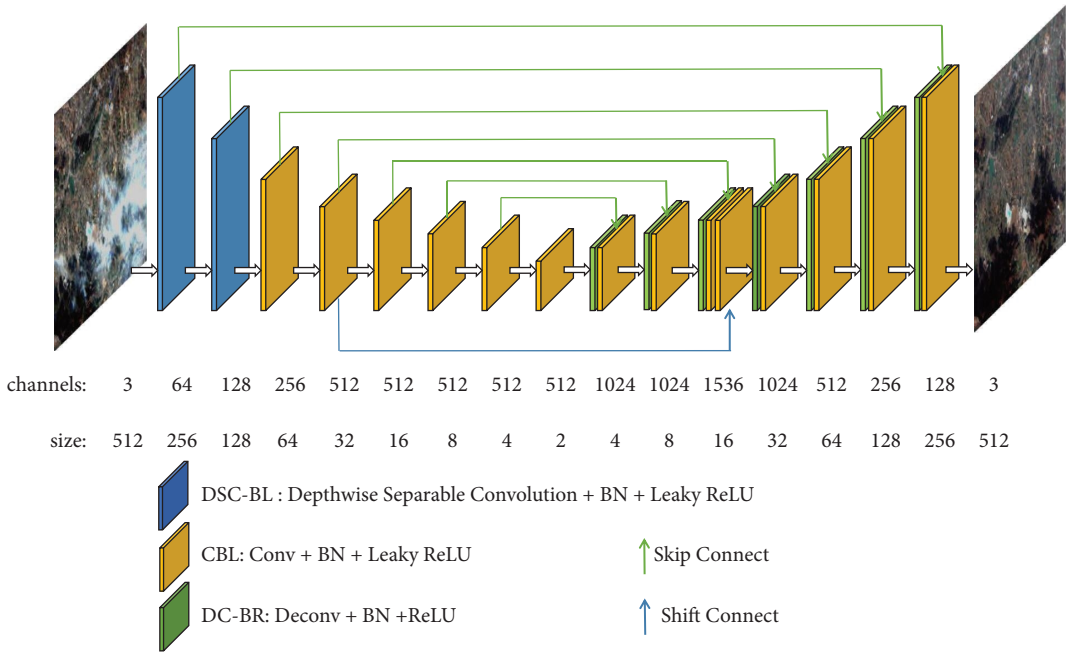


Figure 4: Generator of the reconstruction network.

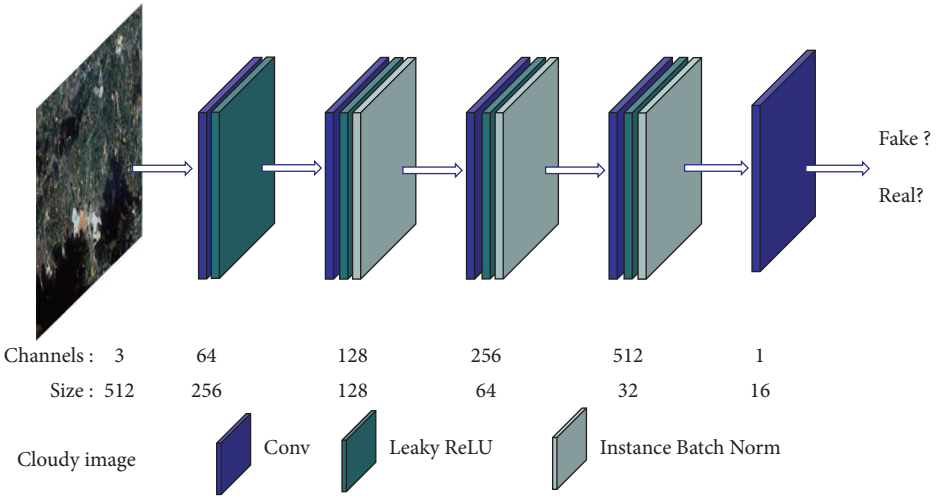


Figure 5: Discriminator of the reconstruction network.

Table 1: Parameters of the generator and the discriminator.

	Kernel size	Stride	Padding	Negative slope
Conv	4	2	1	
Deconv	4	2	1	
DSC (depthwise conv)	3	1	1	
DSC (pointwise conv)	1	1	0	
Leaky ReLU				0.2

required to ensure a positive training process. After experimental attempts, the weighted combination of the generative adversarial loss L_g, the l_1 norm loss L_{l1}, and the shift loss L_{shift} is applicable to the training of remote sensing images. The objective function is as follows:

$$\text{Loss} = L_g + \lambda_{l1} L_{l1} + \lambda_{\text{shift}} L_{\text{shift}}, \tag{4}$$

where λ_{l1} and λ_{shift} are the hyperparameters of weights of L_{l1}, and L_{shift}, respectively.

The generative adversarial loss is used to guide the optimization of the generator and the discriminator [40], as shown in (5). $E(*)$ denotes the expected value of the distribution function. $P_{\text{cloudfree}}(x)$ and $P_{\text{cloud}}(z)$ are the values of pixel distribution in the cloud-covered region and the cloud-free region.

$$L_g = \min_G \max_D V(D,G) = E_{x\sim p_{\text{cloudfree}}(x)}[\log D(x)] + E_{z\sim p_{\text{cloud}}(z)}[\log(1 - D(G(z)))]. \tag{5}$$

The l_1 norm loss is the absolute difference between ground truth I^{gt} and the estimated value I, which is expressed as follows:

$$L_{l_1} = I - I^{gt}. \tag{6}$$

In the reconstruction network of the encoder-decoder architecture, the shift loss L_{shift} is the constraint between encoder features Φ_{L-l} and corresponding decoder features Φ_l, which is expressed as follows:

$$L_{\text{shift}} = \Phi_{L-l} - \Phi_l^2. \tag{7}$$

After the first shift, the temporal differences in reference images are preliminarily eliminated. The normalized multitemporal images are then used for the training of the reconstruction network in the second shift. Finally, the reconstruction network recovers the information of the cloud-covered region.

3. Experimental Results

In order to test the proposed BSN method, it was compared with traditional cloud removal methods and deep learning methods in real and simulated experiments. The selected traditional methods include exemplar-based methods [38] and information cloning methods [2], which represent spatial-based methods and temporal-based traditional methods, respectively. The compared deep learning methods include style transfer [37], GAN-CA [25], U-Net [42], ShiftNet [29], and SAR-opt-GAN [27]. Then, the ablation experiments are conducted to demonstrate the role of double shifts and loss function. The training data of all methods are the same.

3.1. Experimental Settings. For different datasets, different experimental settings should be used to achieve good results. Details of our experimental settings are given.

3.1.1. Datasets. Our dataset consists of high-resolution Level-1C Sentinel-2 images between 2019 and 2021 (downloaded from https://www.copernicus.eu/). Only visible bands (B2, B3, and B4) with a spatial resolution of 10 meters were selected, as shown in Table 2. In the simulation experiments, four datasets containing different feature types such as mountains, rivers, urban buildings, lakes, and oceans were used to validate the effects of cloud removal. In real experiments, another four datasets testified the cloud removal capability for different types and locations of cloud occlusions.

In the simulated and real experiments, the whole scene images were cropped as samples with a size of $512 \times 512 \times 3$. The datasets of the simulated experiments are all cloudless images, and simulated clouds are artificially added. The real experimental datasets include cloud-covered images and cloudless multitemporal images. It is worth noting that it is helpful for the result of cloud removal with as many temporal images as possible.

3.1.2. Parameter Settings. Reconstruction network training adopts adaptive moment estimation (Adam) as a gradient descent optimization algorithm [44]. Adam uses gradient first-order moment estimation and second-order moment estimation to dynamically adjust the learning rate. The learning rate is initialized to 0.002, and the number of iterations is set to 1000 epochs.

DST employed pretrained VGG-19 [45] as a feature extractor. The derivative of the photorealism regularization term is implemented in CUDA for gradient-based optimization. The learning rate of the training is initialized to 0.1 and iterated 1000 epochs.

3.2. Simulated Experiment. In the simulated experiments, comparison experiments and ablation experiments are performed. In the comparison experiments, BSN is

Table 2: Sentinel-2 multitemporal image datasets.

	Jan.	Feb.	Mar.	Apr.	Mar	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
2019									✓	✓	✓	✓
2020	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2021	✓	✓	✓									

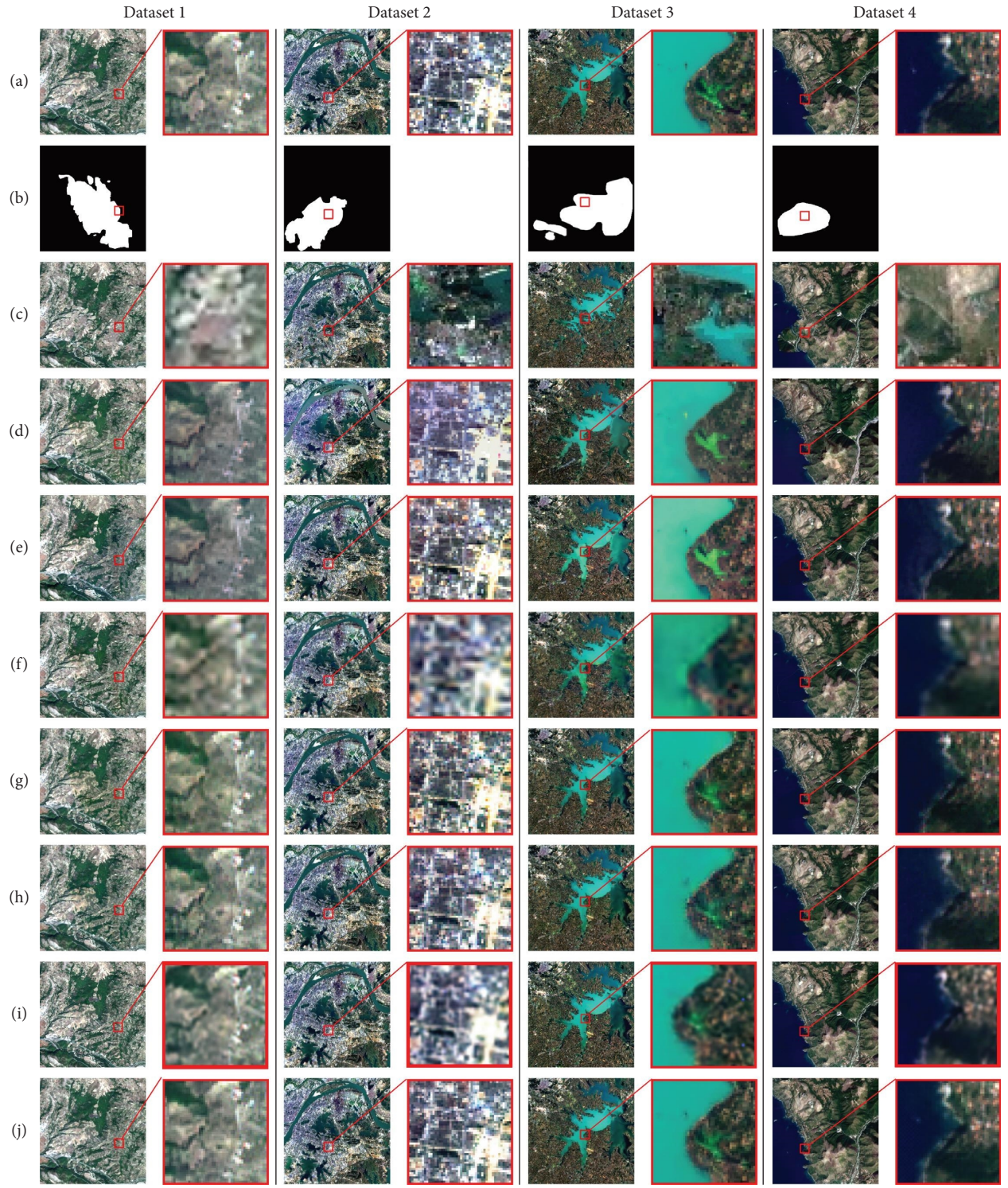


Figure 6: The results of the simulated experiment. (a) Ground truth. (b) Simulated cloud mask. (c) Exemplar-based [38]. (d) Information cloning [2]. (e) Style transfer [37]. (f) GAN-CA [25]. (g) U-Net [42]. (h) Shift-Net [29]. (i) SAR-opt-GAN [27]. (j) Proposed.

compared visually and quantitatively with traditional and deep learning methods. The effects of double shifts and loss function are verified in ablation experiments.

3.2.1. Comparison Experiments. The cloud removal results and detailed enlargements of the simulated experiment are shown in Figure 6. Although the traditional exemplar-based method can reconstruct the feature information of cloud-covered regions, reconstructed features are inconsistent with ground truth images. For example, in the result of dataset 4 in Figure 6(c), the simulated cloud covers the water and land, but it is incorrectly reconstructed as land. The information cloning method cannot overcome temporal differences between images (e.g., color inconsistency), and the result depends on the quality of the temporal image being cloned. For example, in Figure 6(d), datasets 1 and 2 have good results, while datasets 3 and 4 have poor performance. It can be seen from Figure 6(e) that the result of style transfer can only ensure that the style of the reconstructed region is similar to ground truth, but the content information of features is inaccurate. The result of cloud removal is severely affected by the quality of temporal images. GAN-CA, U-Net, Shift-Net, and SAR-opt-GAN are recent deep learning methods, and the results are shown in Figures 6(f)–6(i). The ground objects after being reconstructed are highly consistent with ground truth, which is significantly better than traditional methods. However, the proposed method has the highest clarity of ground objects after reconstruction in cloud-covered regions, which can be further reflected in the following quantitative evaluation.

To better visualize the differences between cloud-removed and ground truth images, an error analysis map is shown in Figure 7. The colors corresponding to errors can be found in the legend. The red scatter on the error map represents the difference between the restoration image and ground truth, with the darker red color representing a larger difference value. The white color in the image indicates no difference in ground truth. Except for style transfer, the error maps of other methods are only present in the simulated cloud-covered area, and the cloud masks are shown in Figure 7(a). In Figure 7(b), the exemplar-based method has the darkest red color, representing the greatest difference between the results and ground truth. The information cloning method sometimes performs well, but it is not stable. For example, in Figure 7(c), the water region of dataset 3 has severe errors, while the other datasets show good results. From Figure 7(d), it can be found that style transfer does not ensure that the content outside the cloud mask remains unchanged, so the entire image is inaccurate by a large margin. The cloud removal results of deep learning methods are shown in Figures 7(e)–7(i), where the proposed method has the best performance.

Furthermore, an error scatter plot was also produced to evaluate the error between the reconstructed image and ground truth. The more dispersed scatter distribution indicates the greater difference between the reconstructed image and ground truth. The results in Figure 8 show that

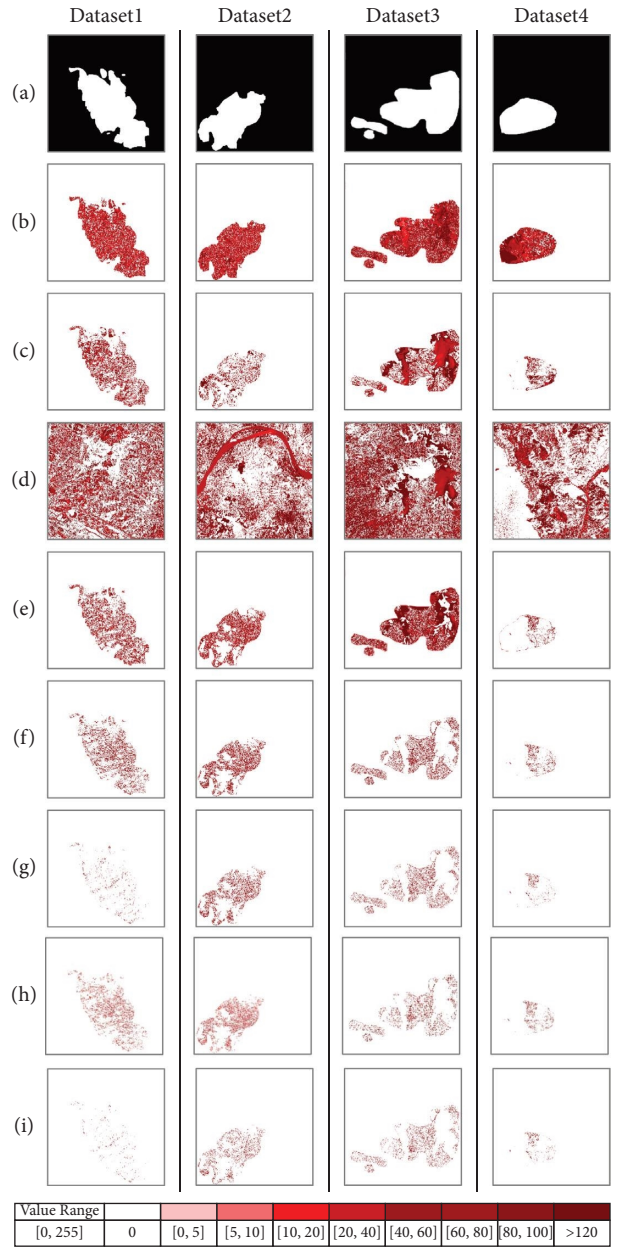


Figure 7: The error analysis maps of ground truth and cloud removal results. (a) Cloud mask. (b) Exemplar-based [38]. (c) Information cloning [2]. (d) Style transfer [37]. (e) GAN-CA [25]. (f) U-Net [42]. (g) Shift-Net [29]. (h) SAR-opt-GAN [27]. (i) Proposed.

the proposed method has the most concentrated error distribution among all methods, and therefore, the results are most similar to ground truth.

Correlation coefficient (CC), structural similarity (SSIM), and peak signal-to-noise ratio (PSNR) are used to evaluate cloud removal results quantitatively in simulated experiments. CC reflects the image correlation between the cloud-free image after cloud removal and the ground truth image, with higher values representing a higher correlation and a maximum of 1. SSIM is a measure of the similarity of two images, and SSIM is equal to 1 when two images are

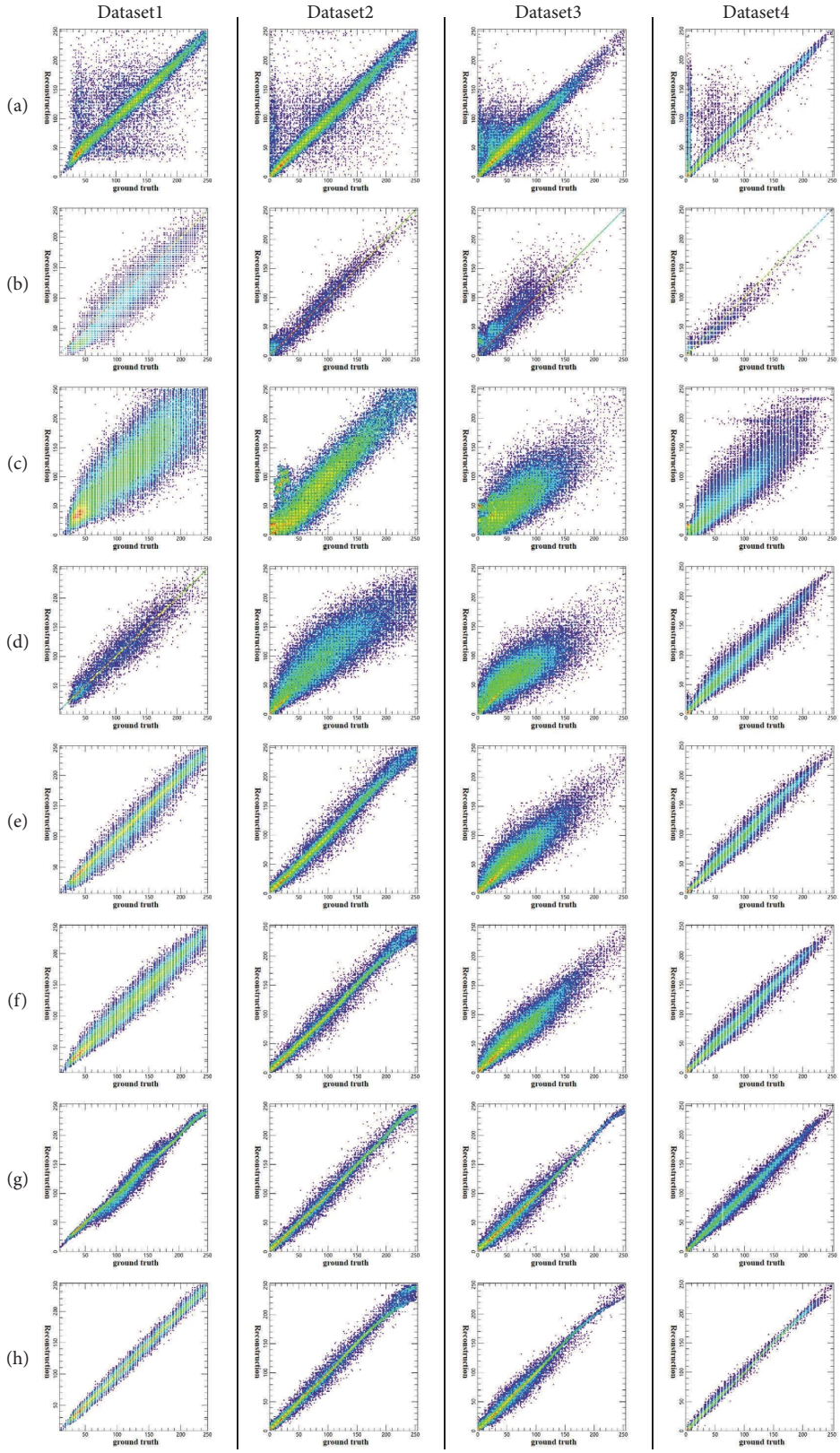


FIGURE 8: The error scatter plot of ground truth and cloud removal results. On each error scatter plot, the X coordinate represents ground truth and the Y coordinate axis represents the reconstructed image (i.e., cloud removal image). (a) Exemplar-based [38]. (b) Information cloning [2]. (c) Style transfer [37]. (d) GAN-CA [25]. (e) U-Net [42]. (f) Shift-Net [29]. (g) SAR-opt-GAN [27]. (h) Proposed.

Table 3: Quantitative evaluations of simulated experiments.

Methods	Dataset	CC↑	SSIM↑	PSNR↑	Parameters
Exemplar-based [38]	1	0.7915	0.6170	17.2868	—
	2	0.8639	0.8142	17.6311	
	3	0.8267	0.7130	19.1175	
	4	0.9072	0.8588	21.9067	
Information cloning [2]	1	0.9675	0.8981	25.2477	—
	2	0.9934	0.9776	30.6375	
	3	0.9435	0.9126	23.3094	
	4	0.9949	0.9749	34.5889	
Style transfer [37]	1	0.8544	0.6201	18.5816	—
	2	0.9281	0.8568	20.1406	
	3	0.8462	0.7097	19.4109	
	4	0.8986	0.7185	21.6804	
GAN-CA [25]	1	0.9485	0.7901	23.2514	9.1 M
	2	0.8887	0.6318	18.5822	
	3	0.8611	0.5795	20.2036	
	4	0.9772	0.8347	28.1387	
U-Net [42]	1	0.9875	0.9495	28.9435	31.0 M
	2	0.9871	0.9046	27.6569	
	3	0.9873	0.9441	30.3411	
	4	0.9964	0.9634	35.9552	
Shift-Net [29]	1	0.9958	0.9756	33.8531	57.7 M
	2	0.9925	0.9746	29.9933	
	3	0.9917	0.9616	32.0589	
	4	0.9973	0.9724	37.3791	
SAR-opt-GAN [27]	1	0.9952	0.9819	33.1435	58.3 M
	2	0.9947	0.9795	31.8443	
	3	0.9862	0.9490	30.2468	
	4	0.9971	0.9810	37.0992	
Proposed	1	**0.9976**	**0.9855**	**36.3430**	60.2 M
	2	**0.9957**	**0.9848**	**32.3386**	
	3	**0.9939**	**0.9712**	**33.4386**	
	4	**0.9984**	**0.9781**	**39.6004**	

Bold values mean the best value.

(a) (b) (c) (d) (e)

Figure 9: The results of ablation experiments in double shifts. (a) Simulated cloudy image. (b) Ground truth image. (c) First shift. (d) Second shift. (e) Proposed.

Table 4: Quantitative evaluation of double shifts.

Method	CC↑	SSIM↑	PSNR↑
First shift	0.7307	0.4428	15.7596
Second shift	0.9269	0.9741	25.3217
Proposed	**0.9976**	**0.9855**	**36.3430**

Bold values mean the best value.

(a) (b) (c) (d) (e) (f) (g)

Figure 10: The results of ablation experiments for the amount of DSC, where (e) is the proposed method with two layers of DSC. (a) Simulated cloudy image. (b) Ground truth image. (c) Without DSC. (d)–(g) 1–4 layer of DSC.

identical. PSNR is the most common and widely used objective measurement to evaluate image quality, and a high PSNR means high image quality after cloud removal. Evaluation metrics for each method are shown in Table 3. In quantitative evaluation, the exemplar-based method and style transfer have the lowest accuracy of cloud removal. This is consistent with the visual results and proves that the image quality after cloud removal is poor. The results of the information cloning method are unstable. For deep learning methods, GAN-CA, U-Net, Shift-Net, SAR-opt-GAN, and the proposed BSN have shown higher accuracies than traditional methods. The proposed method has the highest accuracy in CC, SSIM, and PSNR. This proves that the cloud removal images of the proposed method are most similar to ground truth images and that the proposed method is the most effective.

3.2.2. Ablation Experiments. The effects of double shifts, DSC, loss function, and reference images on cloud removal in remote sensing images are demonstrated in ablation experiments.

(1) The Effects of the Double Shift. First, the effects of double shifts were tested. Ten reference images with large temporal differences in target cloudy images were adopted as datasets. Experiments using only the first shift or the second shift were conducted. It can be seen from Figure 9(c) that when only the first shift was used, the correctness of reconstruction of ground objects could not be guaranteed. When there are large temporal differences between the reference image and the target image, the reconstructed information of the cloud-covered regions usually has temporal differences in the ground truth image. In Figure 9(d), the cloud removal result only using the second shift is visually obviously different from the ground truth image in the cloud-covered region. Meanwhile, the proposed BSN that includes both the first shift and second shift can solve these problems well as shown in Figure 9(e). In Table 4, the quantitative evaluation of the proposed BSN is also significantly better than using the first shift or second shift alone. Therefore, the first shift is helpful in eliminating temporal differences, and the second shift guarantees the accuracy of reconstructed features.

(2) The Effects of Depthwise Separable Convolution. The effects of DSC were tested. The reconstructed network with DSC has stronger detail capturing ability in remote sensing images. To demonstrate this, experiments on changing the number of DSC layers of the proposed BSN were performed. The results are shown in Figure 10, and Figure 10(e) shows the result of the proposed method with the addition of two layers of DSC. Figures 10(c) and 10(d) show that models with no or one layer of DSC result in blurred and discontinuous boundaries. Figures 10(f) and 10(g) show that more than two layers of DSC do not enhance cloud removal but rather lose some important information, as reflected in the inconsistent spectral information of the river. The proposed method with two layers of DSC demonstrates the best visual results, as shown in Figure 10(e). It shows that a proper amount of DSC can improve the accuracy in cloud removal.

Table 5: Quantitative evaluation of loss function.

Loss	CC↑	SSIM↑	PSNR↑
$L_{l1} + L_{\text{shift}}$	0.9951	0.9836	34.4274
$L_g + L_{\text{shift}}$	0.9944	0.9820	31.9331
$L_g + L_{l1}$	0.9973	0.9848	35.6671
Proposed	**0.9976**	**0.9855**	**36.3430**

Bold values mean the best value.

(3) The Effects of the Loss Function. To demonstrate the effects of the loss function on the training process of the reconstructed network, the experiments were conducted for the training process. The loss function of BSN consists of L_g, L_{l1}, and L_{shift}. In the ablation experiment of the loss function, L_g, L_{l1}, and L_{shift} were removed separately and the cloud removal results were evaluated quantitatively. The same training datasets and 1000 training epochs were used in training. Table 5 shows that removing any of L_g, L_{l1}, and L_{shift} will reduce the accuracy of the cloud removal results. The loss function used in the proposed BSN achieves the highest accuracy in CC, SSIM, and PSNR.

(4) The Effects of Reference Images. As a temporal-based method, the cloud removal capability of BSN is affected by the number of reference images. Normally, more reference images usually mean more valid reference information. An experiment on the effect of the number of reference images was conducted. BSN requires at least one reference image to ensure the availability of the network. The effect of reference images with different numbers on the accuracy is shown in Table 6. It can be seen in the experiment that more reference images are helpful for the improvement of accuracy. Therefore, it is desirable to ensure sufficient temporal reference images. If data acquisition is limited, a small number

Table 6: Accuracy of different numbers of reference images.

Number	CC↑	SSIM↑	PSNR↑
1	0.9801	0.9340	27.3401
2	0.9850	0.9451	28.7423
4	0.9900	0.9627	30.2235
6	0.9915	0.9670	31.0447
8	0.9935	0.9705	31.7328
10	**0.9942**	**0.9735**	**32.5419**

Bold values mean the best value.

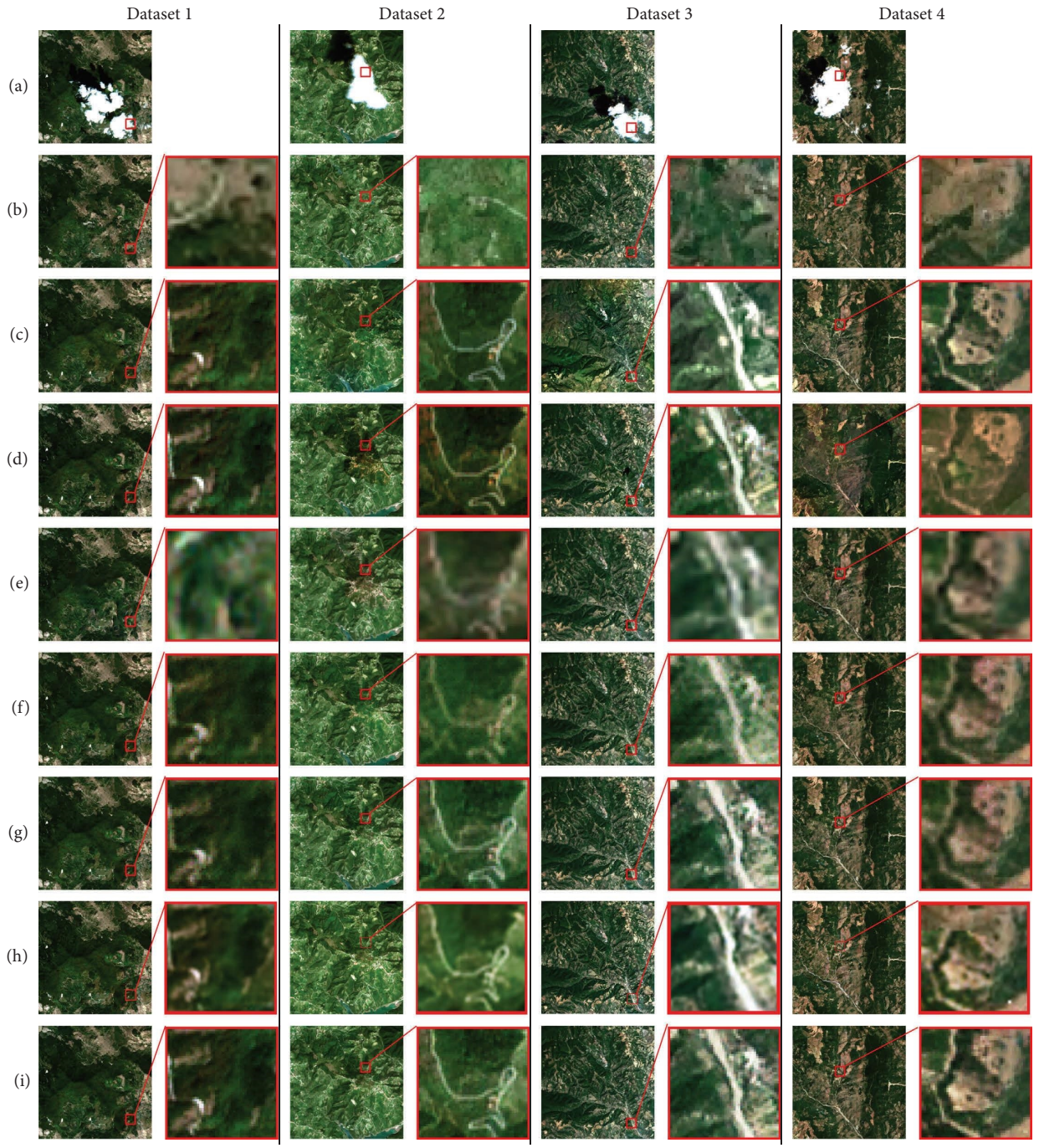
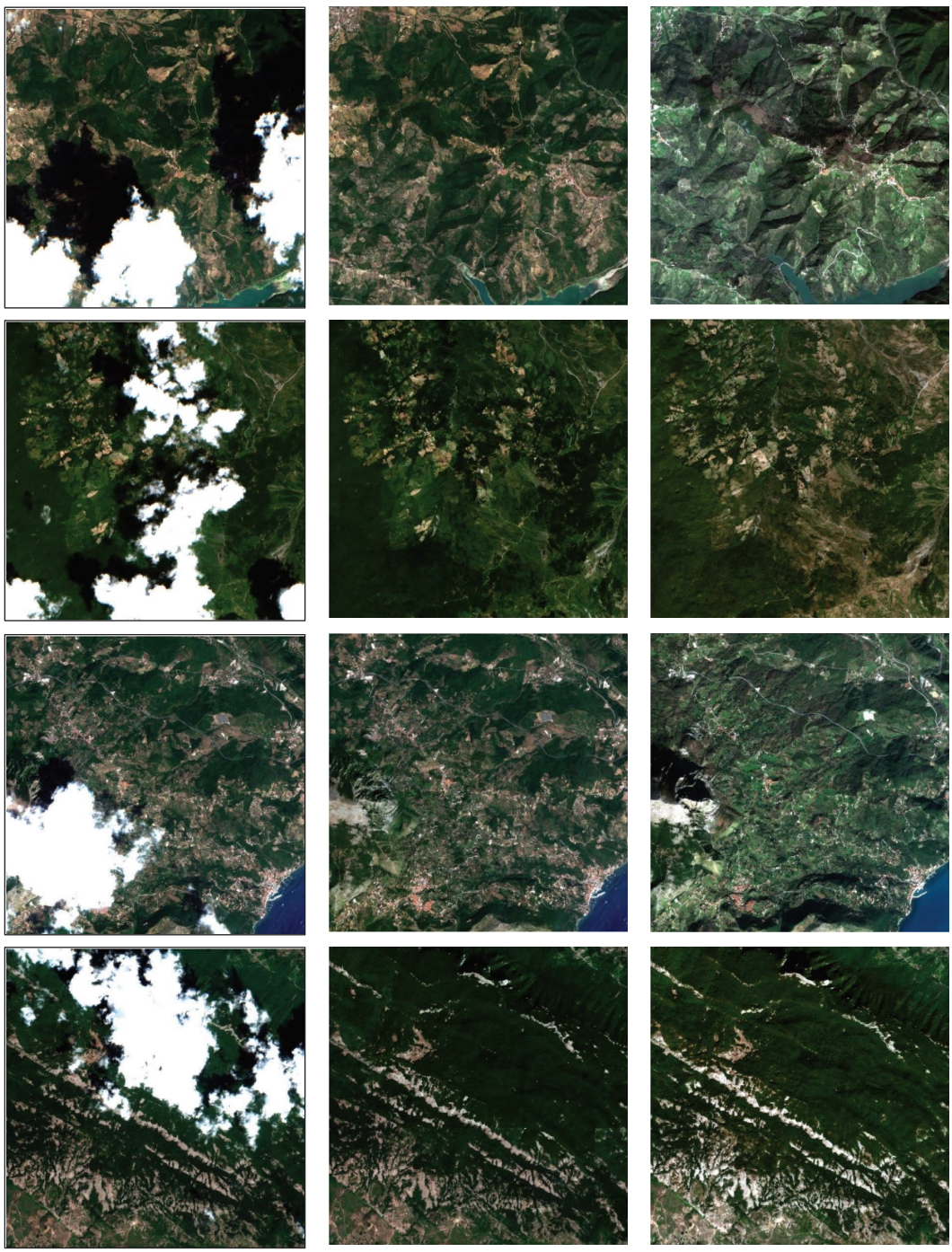


Figure 11: The result of the real experiment. (a) Cloudy image. (b) Exemplar-based [38]. (c) Information cloning [2]. (d) Style transfer [37].(e) GAN-CA [25]. (f) U-Net [42]. (g) Shift-Net [29]. (h) SAR-opt-GAN [27]. (i) Proposed.



(a) (b) (c)

Figure 12: Additional real experiments. (a) Cloudy images. (b) Cloud-removed images. (c) Images in the same region of (a) acquired at other times.

of temporal images can also finish reconstruction. As a trade-off between reconstruction accuracy and the difficulty of data acquisition, ten multitemporal reference images show a good performance in this experiment. Considering that not many reference images are generally used in practical applications and that there may be more adverse effects of sudden changes, we have not conducted more tests.

3.3. Real Experiment. In real experiments, cloudy images are directly reconstructed. It is important to note that the ground truth of cloudy images does not exist. Therefore, the result can only be judged visually. The cloud removal results of the real experiment are shown in Figure 11. The traditional exemplar-based methods can fill cloud-covered image information but not actual features. In Figure 11(b), the exemplar-based method often ignores the roads of images.

In Figure 11(c), information cloning cannot always perform well, and sometimes there are serious spectral errors. As can be seen in Figure 11(d), for the result of style transfer, it is not only that the overall spectral information is not consistent with the target but also that the pixels in noncloud-covered regions are altered. Style transfer and information cloning do not perform well in the elimination of temporal differences, and images show severe chromatic aberrations. GAN-CA cannot accurately capture the local details of the image when removing clouds. It is easy to confuse when dealing with the relationship between different features (such as bareland, vegetation, and buildings). Moreover, the output image of GAN-CA has low definition. It is obvious in the enlarged figure, as shown in Figure 11(e). U-Net has difficulty recovering complex textures, such as vegetation and mountains are interspersed, as shown in Figure 11(f). In Figure 11(g), the cloud removal results of Shift-Net are better than those of the previous methods, and the visual effects are quite good. Although the SAR-opt-GAN method is overall acceptable, the cloud removal results showed temporal differences in dataset 2 and blurred images in dataset 3, as shown in Figure 11(h). BSN has great advantages over other traditional and recent cloud removal methods. For example, the cloud-free images reconstructed by BSN show reasonable global semantics and local details, accurate feature classes, trace-free boundaries, and high-resolution cloud-free images, as shown in Figure 11(i).

To further test and evaluate the effective cloud removal capability of the proposed BSN, the experiments were conducted on four additional real remote sensing images. The results are shown in Figure 12, which includes cloud-contaminated images, cloud-removed images, and multitemporal images of the same region as a reference. Comparing the reconstructed image with the cloudy image, the reconstructed area of the reconstructed image and the noncloud-covered area of the cloud image are consistent spectrally. Spatially, cloud-removed images have the same features and textures as temporal images in reconstructed regions. It can therefore be seen that the proposed BSN is effective in cloud removal and performs well in both global semantics and texture details.

4. Conclusions

In this article, two shift-based BSN was proposed for cloud removal in optical remote sensing images. In the first shift, MM and DST are used. MM can preliminarily normalize multitemporal images. DST can further eliminate temporal differences. The preprocessing of the first shift is conducive to improving the efficiency of the second shift and the accuracy of cloud removal. In the second shift, the reconstruction network is used to remove clouds from cloud-covered images. In order to improve the reconstruction network's ability to extract local details and maintain global consistency, shift connection and DSC are introduced. After simulated experiments and real experiments, the proposed method has obvious advantages over traditional methods and deep learning methods in terms of accuracy and visual effects. The ablation experiments also demonstrate the role of the double shift and loss function. BSN can effectively remove clouds in optical remote sensing images, thereby improving the effective information of optical remote sensing images.

The advantage of the proposed method is that the original spectral information of the image is maintained when clouds are removed, thus providing a good visual effect. Although the proposed BSN has a great effect on removing thick clouds in images, it still has some limitations. For example, it requires cloudless multitemporal images as reference data, and the quality of the results is affected by multitemporal images. Considering the different locations of clouds in cloud-covered images, several images can be processed with simultaneous cloud removal in the future. At present, some researchers have also combined optical and SAR images based on deep learning to improve the ability of cloud removal [11, 12]. For sudden changes in multitemporal remote sensing images, such as new buildings and man-made landscapes, the envisioned future strategy is to use multisource remote sensing images (e.g., SAR images) as auxiliary data to improve the accuracy and reliability.

Data Availability

Our dataset consists of high-resolution Level-1C Sentinel-2 images between 2019 and 2021 (downloaded from https://copernicus.eu/.)

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Chaojun Long was responsible for methodology and writing the manuscript. Xinghua Li was responsible for conceptualization and framework design. Yinghong Jing was responsible for experimental discussion. Huanfeng Shen was responsible for supervision and proofreading.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant no. 42171302 and the Open Fund of Hubei Luojia Laboratory under Grant no. 220100055.

References

[1] H. Shen, X. Li, Q. Cheng et al., "Missing information reconstruction of remote sensing data: a technical review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 61–85, 2015.

[2] C. Lin, P. Tsai, K. Lai, and J. Chen, "Cloud removal from multitemporal satellite images using information cloning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 1, pp. 232–241, 2012.

[3] J. Chen, P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh, "A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay

filter," *Remote Sensing of Environment*, vol. 91, no. 3-4, pp. 332–344, 2004.

[4] G. J. Roerink, M. Menenti, and W. Verhoef, "Reconstructing cloudfree NDVI composites using Fourier analysis of time series," *International Journal of Remote Sensing*, vol. 21, no. 9, pp. 1911–1917, 2000.

[5] X. Li, H. Shen, H. Li, and L. Zhang, "Patch matching-based multitemporal group sparse representation for the missing information reconstruction of remote-sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 8, pp. 3629–3641, 2016.

[6] L. Lorenzi, F. Melgani, and G. Mercier, "Missing-area reconstruction in multispectral images under a compressive sensing perspective," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 7, pp. 3998–4008, 2013.

[7] X. Li, H. Shen, L. Zhang, H. Zhang, Q. Yuan, and G. Yang, "Recovering quantitative remote sensing products contaminated by thick clouds and shadows using multitemporal dictionary learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 7086–7098, 2014.

[8] X. Li, L. Wang, Q. Cheng, P. Wu, W. Gan, and L. Fang, "Cloud removal in remote sensing images using nonnegative matrix factorization and error correction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 148, pp. 103–113, 2019.

[9] Q. Cheng, H. Shen, L. Zhang, Q. Yuan, and C. Zeng, "Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal MRF model," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 92, pp. 54–68, 2014.

[10] S. Benabdelkader and F. Melgani, "Contextual spatiospectral postreconstruction of cloud-contaminated images," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 204–208, 2008.

[11] R. Eckardt, C. Berger, C. Thiel, and C. Schmullius, "Removal of optically thick clouds from multi-spectral satellite images using multi-frequency SAR data," *Remote Sensing*, vol. 5, no. 6, pp. 2973–3006, 2013.

[12] A. Meraner, P. Ebel, X. X. Zhu, and M. Schmitt, "Cloud removal in sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 333–346, 2020.

[13] C. Hu, L. Huo, Z. Zhang, and P. Tang, "Automatic cloud removal from multi-temporal landsat collection 1 data using Poisson blending," in *Proceedings of the IGARSS 2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1661–1664, Yokohama, Japan, 2019.

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

[15] H. Shen, X. Li, L. Zhang, D. Tao, and C. Zeng, "Compressed sensing-based inpainting of aqua moderate resolution imaging spectroradiometer band 6 using adaptive spectrum-weighted sparse bayesian dictionary learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 2, pp. 894–906, 2014.

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[17] Y. Chen, R. Fan, M. Bilal, X. Yang, J. Wang, and W. Li, "Multilevel cloud detection for high-resolution remote sensing imagery using multiple convolutional neural networks," *ISPRS International Journal of Geo-Information*, vol. 7, no. 5, pp. 181–197, 2018.

[18] T. Sandhan and J. Choi, "Simultaneous detection and removal of high altitude clouds from an image," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4789–4798, Venice, Italy, 2017.

[19] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4274–4288, 2018.

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

[21] Y. Chen, L. Tang, X. Yang, R. Fan, M. Bilal, and Q. Li, "Thick clouds removal from multitemporal ZY-3 satellite images using deep learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 143–153, 2020.

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

[23] P. Dai, S. Ji, and Y. Zhang, "Gated convolutional networks for cloud removal from bi-temporal remote sensing images," *Remote Sensing*, vol. 12, no. 20, p. 3427, 2020.

[24] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 179, pp. 14–34, 2021.

[25] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, "Generative image inpainting with contextual attention," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5505–5514, Salt Lake City, UT, USA, 2018.

[26] V. Sarukkai, A. Jain, B. Uzkent, and S. Ermon, "Cloud removal in satellite images using spatiotemporal generative networks," in *Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1785–1794, Snowmass, CO, USA, 2020.

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

[28] F. N. Darbaghshahi, M. R. Mohammadi, and M. Soryani, "Cloud removal in remote sensing images using generative adversarial networks and SAR-to-optical image translation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–9, 2022.

[29] Z. Yan, X. Li, M. Li, W. Zuo, and S. Shan, "Shift-net: image inpainting via deep feature rearrangement," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 1–17, Munich, Germany, 2018.

[30] C. Long, J. Yang, X. Guan, and X. Li, "Thick cloud removal from remote sensing images using double shift networks," in *Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pp. 2687–2690, Brussels, Belgium, 2021.

[31] F. L. Gadallah, F. Csillag, and E. J. M. Smith, "Destriping multisensor imagery with moment matching," *International Journal of Remote Sensing*, vol. 21, no. 12, pp. 2505–2511, 2000.

[32] X. Zhang, R. Feng, X. Li, H. Shen, and Z. Yuan, "Block adjustment-based radiometric normalization by considering global and local differences," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[33] L. Gatys, A. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2414–2423, Las Vegas, NV, USA, 2016.

[34] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. Lempitsky, "Texture networks: feed-forward synthesis of textures and stylized images," in *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 1349–1357, New York, NY, USA, 2016.

[35] E. Risser, P. Wilmot, and C. Barnes, "Stable and controllable neural texture synthesis and style transfer using histogram losses," 2017, http://arxiv.org/abs/1701.08893.

[36] J. Johnson, A. Alahi, and F. Li, "Perceptual losses for realtime style transfer and super-resolution," in *Proceedings of the ECCV European Conference on Computer Vision*, pp. 694–711, Amsterdam, The Netherlands, 2016.

[37] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4990–4998, Honolulu, China, 2017.

[38] A. Criminisi, P. Pérez, and K. Toyama, "Object removal by exemplar-based image inpainting," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.

[39] C. Ledig, L. Theis, and F. Huszar, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4681–4690, Honolulu, China, 2017.

[40] P. Isola, J. Zhu, T. Zhou, and A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134, Honolulu, China, 2017.

[41] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, http://arxiv.org/abs/1511.06434.

[42] O. Ronneberger, P. Fischer, and T. Brox, "Unet: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference Medical Image Computing and Computer Assisted Intervention*, pp. 234–241, Munich, Germany, 2015.

[43] F. Chollet, "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251–1258, Honolulu, China, 2017.

[44] A. A. Hameed, B. Karlik, and M. S. Salman, "Back-propagation algorithm with variable adaptive momentum," *Knowledge-Based Systems*, vol. 114, pp. 79–87, 2016.

[45] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, http://arxiv.org/abs/1409.1556.