Thin Cloud Removal Fusing Full Spectral and Spatial Features for Sentinel-2 Imagery

Jun Li, Yuejie Zhang, Qinghong Sheng, Zhaocong Wu, Bo Wang, Zhongwen Hu, *Member, IEEE*, Guanting Shen, Michael Schmitt, *Senior Member, IEEE*, and Matthieu Molinier, *Member, IEEE*

***Abstract*—Multispectral remote sensing images are widely used for monitoring the globe. Although thin clouds can affect all optical bands, the influences of thin clouds differ with band wavelength. When processing multispectral bands at different resolutions, many methods only remove thin clouds in visible/near-infrared bands or rescale multiresolution bands to the same resolution and then process them together. The former cannot make full use of multispectral information, and in the latter, the rescaling process will introduce noise. In this article, a deep-learning-based thin cloud removal method that fuses full spectral and spatial features in original Sentinel-2 bands is proposed, named CR4S2. A multi-input and output architecture is designed for better fusing information in all bands and reconstructing the background at original resolutions. In addition, two parallel downsampling residual blocks are designed to transfer features extracted from different depths to the bottom of the network. Experiments were conducted on a new globally distributed Sentinel-2 thin cloud removal dataset called WHUS2-CRv. The results show that the best averaged peak signal-to-noise ratio, structural similarity index measurement, normalized root-mean-square error, and spectral angle mapper of the proposed method over 12 bands in all 20 testing images were 39.55, 0.9443, 0.0245, and 2.5676°, respectively. Compared with baseline methods, the proposed CR4S2 method can better restore not only the spatial features but also spectral features. This indicates that the proposed method is very promising for removing thin clouds in multispectral remote sensing images at different resolutions.**

***Index Terms*—Deep learning (DL), multifeature fusion, parallel downsample residual block (PDRB), Sentinel-2, thin cloud removal.**

Manuscript received 22 August 2022; revised 16 September 2022; accepted 28 September 2022. Date of publication 4 October 2022; date of current version 19 October 2022. This work was supported in part by the Natural Science Foundation of Jiangsu Province under Grant BK20220888, in part by the National Key Laboratory Foundation 2021-JCJQ-LB-006 under Grant 6676142411442120, in part by the National Key Laboratory of Science and Technology on Space Microwave under Grant HTKJ2022KL504018, and in part by the Academy of Finland through the Finnish Flagship Programme FCAI: Finnish Center for Artificial Intelligence under Grant 320183. *(Corresponding author: Qinghong Sheng.)*

Jun Li, Yuejie Zhang, Qinghong Sheng, and Bo Wang are with the College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: jun.li@nuaa.edu.cn; zhang_yj@nuaa.edu.cn; qhsheng@nuaa.edu.cn; wangbo_nuaa@nuaa.edu.cn).

Zhaocong Wu and Guanting Shen are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: zcwoo@whu.edu.cn; sgt0620@whu.edu.cn).

Zhongwen Hu is with the MNR Key Laboratory for Geo-Environmental Monitoring of Great Bay Area, Shenzhen University, Shenzhen 518060, China (e-mail: zwhoo@szu.edu.cn).

Michael Schmitt is with the Department of Aerospace Engineering, University of the Bundeswehr Munich, 85577 Neubiberg, Germany (e-mail: michael.schmitt@unibw.de).

Matthieu Molinier is with the VTT Technical Research Centre of Finland, Ltd., 02044 Espoo, Finland (e-mail: matthieu.molinier@vtt.fi).

Digital Object Identifier 10.1109/JSTARS.2022.3211857

I. Introduction

While the development of remote sensing satellite technology, a large number of multispectral images with high spatial resolution have been acquired. Due to the rich information in multispectral remote sensing images, they have been widely used for land use and land cover classification [1], [2], environmental monitoring [3], [4], [5], and urban extraction [6], [7]. However, most of the time, about 67% of the land surface is covered by clouds, which seriously influences the usability of remote sensing images [8]. Clouds can be classified into thick and thin clouds according to their influence on the background signal. Thick clouds completely block the optical signal returned from the Earth surface while thin clouds let at least some signal through. Therefore, in contrast to situations with thin cloud coverage, in the presence of thick clouds, no information about the ground can be retrieved from a single image. For the removal of thick clouds, multitemporal images [9], [10], [11], [12] or auxiliary data such as SAR images are often necessary, regardless of whether the method is traditional, machine learning, or deep learning (DL) based [13], [14], [15], [16].

In order to restore the background information in thin cloud contaminated areas in single-date images, various thin cloud removal methods have been developed [17], [18]. The traditional thin cloud removal methods in remote sensing can be divided into spectral analysis [19], [20] and image filtering methods [21], which are generally used with a simple cloud distortion physical model.

Methods based on spectral analysis assume that there is a high correlation between the bands in the cloud-free multispectral image. The cloud-free image is obtained by fitting the cloudy image to this correlation relationship. Zhang et al. [22] proposed a cloud optimal transformation (HOT), which is based on the statistical analysis of spectral information of a large number of clear pixels. It is assumed that under clear sky conditions, the reflectance of red and blue bands is highly correlated, and a "clear sky line" is used to correct cloud contaminated pixels to clear pixels. Chen et al. [23] proposed an iterative HOT algorithm (IHOT), which detects and removes thin clouds in Landsat images. IHOT uses cloudy and corresponding clear images to solve the spectral confusion between clouds and bright surfaces. He et al. [24] calculated the hot value of each pixel in the image, then adjusted the deviation according to different land cover types, and regarded the HOT image as a digital estimation map (DEM), filled in the low valley and wiped out the peak caused

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

by the highlighted background. Lv et al. [19] assumed that there is a linear relationship between any two visible bands in clear pixels. First, the Fmask algorithm [25], [26] was used to obtain the clear pixels, then the parameters of the constructed model were solved based on clear pixels and the assumption that visible and near-infrared bands are correlated in water areas under thin clouds. The correlation between various bands assumed in methods based on spectral analysis is not applicable to some highlighted ground objects such as snow and buildings [23]. And the spatial information was not fully used in these methods [27].

Image-filtering-based methods treat the background as a high-frequency component and the thin cloud as a low-frequency component, then remove thin clouds by processing the low-frequency and high-frequency components, respectively [28], [29]. A cloud distortion physical model was proposed to describe the radiation transmission process under cloud influence in [21]. This model converted the image to the frequency domain and then filtered the low-frequency component by using homomorphic filtering (HF) to remove thin clouds. Based on the cloud distortion physical model, Liu and Hunt [30] estimated noise in the image with a Kaiser window and then filtered the noise to remove thin clouds. Two different wavelet transforms were adopted to obtain the low-frequency coefficients of the image, and then, HF was used to reduce the low-frequency information for thin cloud removal [31]. Shen et al. [32] proposed an adaptive HF (AHF) method, which uses different truncation frequencies for different bands to filter cloud components, then replaces cloud pixels with the filtered cloud pixels and keeps clear pixels unchanged. Wan and Li [33] decomposed the image into low-frequency and high-frequency components and then set the filter windows with different sizes according to the decomposition level to avoid damaging low-frequency information in the background. In [17], a spherical model was proposed to produce the transmittance map and dark channel prior (DCP) [34] was then used to remove thin clouds. Although image filtering can remove thin clouds, it also filters out some low-frequency components in the background.

In recent years, DL, in particular convolutional neural networks (CNNs), has been widely used for thin cloud removal in remote sensing images and achieved better results than traditional methods [35]. A packet convolution residual network was developed in [36], in which multiple parallel residual subnetworks for processing different bands were used to remove thin clouds in Landsat 8 images. Li et al. [27] proposed a residual symmetrical concatenation network for thin cloud removal. The results of this work indicated that more bands and residual structures are conducive to cloud removal. In [37], a cyclic convolution network was used to extract the potential cloud regions, and then, an automatic encoder was adopted to remove thin clouds in the regions. The guided filtering and multiscale convolution unit were combined to make the network focus on the texture features and obtain a larger receptive field than the original convolution, so as to avoid network degradation by learning the residual between cloudy and cloud-free images [38]. In [39], a gated cloud removal network based on multitemporal images was proposed to take the current cloudy image, recent less cloudy image and their cloud masks as the input, and the total losses of image level, feature level, and change were calculated to obtain a good cloud removal result. Using a U-Net-based network [40] with two input and output branches, Li et al. [41] proved that the information in vegetation red-edge and short wave infrared bands, which is at a lower resolution but less affected by thin clouds, can help remove thin clouds in visible bands. Yu et al. [42] constructed a multiscale cloud removal network (MCRN) with the proposed cloud-aware and feature extraction module as a basic unit. The parameters of cloud distortion model were encoded into trainable parameters in MCRN for cloud removal. Wen et al. [43] introduced channel attention mechanism into residual architecture to suppress thin clouds and enhancing the background details. These CNN-based methods use a synthetic or small dataset for training and testing and ignore some bands when processing multispectral bands at different resolutions.

As a very promising framework, generative adversarial network (GAN) [44] has proved very effective in image generation, synthesizing, and inpainting. GANs have also been applied on thin cloud removal in remote sensing images successfully. Enomoto et al. [45] proposed a conditional GAN (cGAN) for the first time to remove thin clouds in remote sensing images. The cGAN took the cloud-free and corresponding synthesized cloudy visible and near-infrared bands as inputs while only outputting visible bands. Wang et al. [46] also adopted cGAN architecture for thin cloud removal. U-Net was used for thin cloud removal in the first stage of the algorithm proposed in [47]. A cloudy image synthesis paradigm using cloud information in sea scenes was proposed to generate dataset for training the thin cloud removal method in [48]. A Cycle-GAN, which uses unpaired cloudy and cloud-free images for training, was proposed to remove the influence of clouds in Sentinel-2 images both in [49] and [14]. Although the cyclic consistency of Cycle-GAN can retain the texture information of the cloudy image when converting it into a cloud-free image, a large amount of spectral information of the background is lost. In [37], cloud detection was carried out to obtain a cloud probability map that was then put into a generative network with a corresponding cloud-free image. To enhance performance, the perception loss between the cloud removal result and the cloud-free image was established by the VGG network [50]. Xu et al. [51] used cloud masks to adapt different attention to different cloud areas by combining recurrent attention mechanism and GAN, so as to improve the quality of thin cloud removal results. Li et al. [52] combined the cloud distortion model with GAN for thin cloud removal in remote sensing images. Due to the introduction of the cyclic reconstruction process, paired cloudy and cloud-free images are not required in training. In [53], cloud and background components were extracted from a cloud image first. Then, the cloud component was synthesized to a clear image with a physical model. The two steps were applied to the synthesized cloudy image and extracted background, respectively, to construct the cyclic process. However, the training of GAN is usually unstable and results are not very satisfactory when using unpaired images.

Sentinel-2 imagery, with its multispectral bands at different spatial resolutions, has played an important role in Earth observation. Although classical CNN-based methods obtain higher

performance and are more stable than GAN-based methods on thin cloud removal, they usually ignore low spatial resolution bands or rescale all bands to the same resolution when processing multispectral images at different spatial resolutions, which will introduce noise. To solve the limitations of CNN and GAN-based thin cloud removal methods, this article proposes an end-to-end DL-based thin cloud removal method for Sentinel-2 images (CR4S2) by taking all native 13 spectral bands into consideration. The original spectral and spatial features in Sentinel-2 images at three different spatial resolutions are natively extracted and fused in the proposed CR4S2. Since most thin cloud removal datasets either cover small regions [52], [54] or have long time intervals between paired cloudy and cloud-free samples [27], the performances of thin cloud removal methods cannot be evaluated very accurately by these datasets. A new thin cloud removal dataset WHUS2-CRv was collected to solve the limitations in this work. The main contributions are as follows.

1) A novel encoder–decoder network architecture is presented with multiple input and output branches that is tailored for thin cloud removal in Sentinel-2A images with the fusion of all bands. A double-path depthwise separable convolution (DDSC) module is designed to extract and fuse multiscale features with fewer parameters than a normal convolutional layer. Two parallel downsample residual blocks (PDRB-D and PDRB-T) are designed and injected into the encoder to fuse and pass features from different spectral bands to the bottom of the network. A multioptimization loss was introduced to further improve the capability of restoration of edge information and preservation of clear background.
2) A large thin cloud removal validation dataset WHUS2-CRv for Sentinel-2 imagery is presented. This dataset contains 123 paired cloudy and cloud-free images distributed all over the world. The worldwide distribution guarantees the diversity of land cover types. In order to minimize the spectral and spatial difference between cloudy and cloud-free images, the interval time for each paired cloudy and cloud-free images is 10 days, which is the revisit time of Sentinel-2A.

The rest of this article is organized as follows. Section II presents the experimental data. Section III introduces the proposed CR4S2 method. The experimental results are shown in Section V and the discussion is also made in this section. We draw conclusions in Section V.

II. Data

A. Sentinel-2A Multispectral Data

Sentinel-2A is equipped with a high-resolution multispectral imager (MSI) that covers 13 spectral bands (see Table I). The wavelength of Sentinel-2A ranges from 0.443 μm to 2.190 μm. The three vegetation red-edge bands are mainly used for monitoring vegetation. Bands 1 and 9/10 are used to detect the Coastal aerosol and monitor the water vapor/Cirrus, respectively. Bands 1/9/10 are used for detecting and correcting the atmospheric effects that can provide thin cloud information when removing thin clouds in other bands. In this work, all bands in Sentinel-2A imagery are used.

TABLE I
Sentinel-2A Sensor Bands

Band no.	Band name	Central Wavelength (μm)	Bandwidth (nm)	Spatial resolution (m)
Band 1	Coastal aerosol	0.443	27	60
Band 2	Blue	0.490	98	10
Band 3	Green	0.560	45	10
Band 4	Red	0.665	38	10
Band 5	Vegetation Red Edge	0.705	19	20
Band 6	Vegetation Red Edge	0.740	18	20
Band 7	Vegetation Red Edge	0.783	28	20
Band 8	NIR	0.842	145	10
Band 8A	Vegetation Red Edge	0.865	33	20
Band 9	Water Vapor	0.945	26	60
Band 10	SWIR-Cirrus	1.375	75	60
Band 11	SWIR	1.610	143	20
Band 12	SWIR	2.190	242	20

All bands were used in our study.

Fig. 1 shows three examples under the different cloud-contamination condition (odd rows) and the corresponding cloud-free images (even rows). Column 1 shows the true color composited images (T), Columns 2–14 are bands 2/3/4/8 (Visible and Near-Infrared, VNIR), 5/6/7/8A/11/12 (Vegetation Red-Edge NIR Narrow and Short-Wave Infrared, VRE/NIRn/SWIR), 1/9/10 (Coastal Aerosol, Water Vapor, and Cirrus band, Ca/Wv/Cir). All bands are resized to the same size for better visualization. We can see that the influences of thin clouds on these bands are different. In bands 6/7/8/8A/11/12, the clouds are not very noticeable, which means that features in these bands can help restore background information in other bands such as bands 2/3/4. However, the cloud pixels can be easily found on bands Ca/Wv/Cir, which can help locate cloud areas. This article aims to make full use of all spectral bands in Sentinel-2 images for thin cloud removal. We can also see that band 10 contains the least background information among all bands. This is one of the reasons why thin cloud removal took band 10 as input but was not applied to band 10 in this article.

B. Details of the Experimental Dataset

In some related works, the datasets only cover local regions [52], [54], use synthesized cloudy images [55], [56], or include cloudy and cloud-free images with long interval [27], [57]. Although there are some widely distributed datasets, images in these datasets are relatively few, typically less than 40 pairs [41].

Due to the diversity of land cover types on the Earth, a larger and more representative thin cloud removal dataset is needed for comprehensive study. In this article, we present a newly collected thin cloud removal validation dataset, WHUS2-CRv, for Sentinel-2 imagery. Fig. 2 shows the distribution of the sampled regions in WHUS2-CRv. For each region, one cloudy and cloud-free image pair with 10 days interval was selected to avoid reflectance changes between them as much as possible.

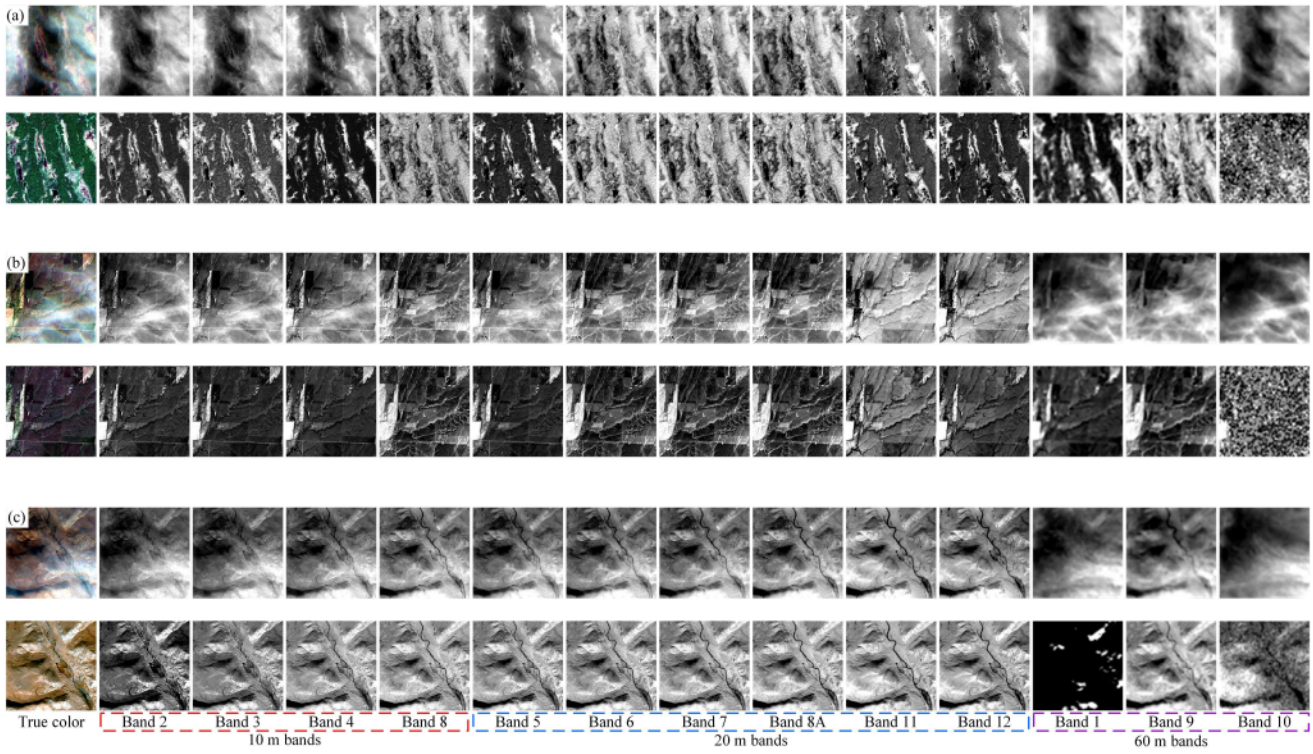


Fig. 1. Three examples of cloud contaminated images (odd rows) corresponding cloud-free images (even rows) (details can be found in Table VII). All bands are resized to the same size for better visualization.

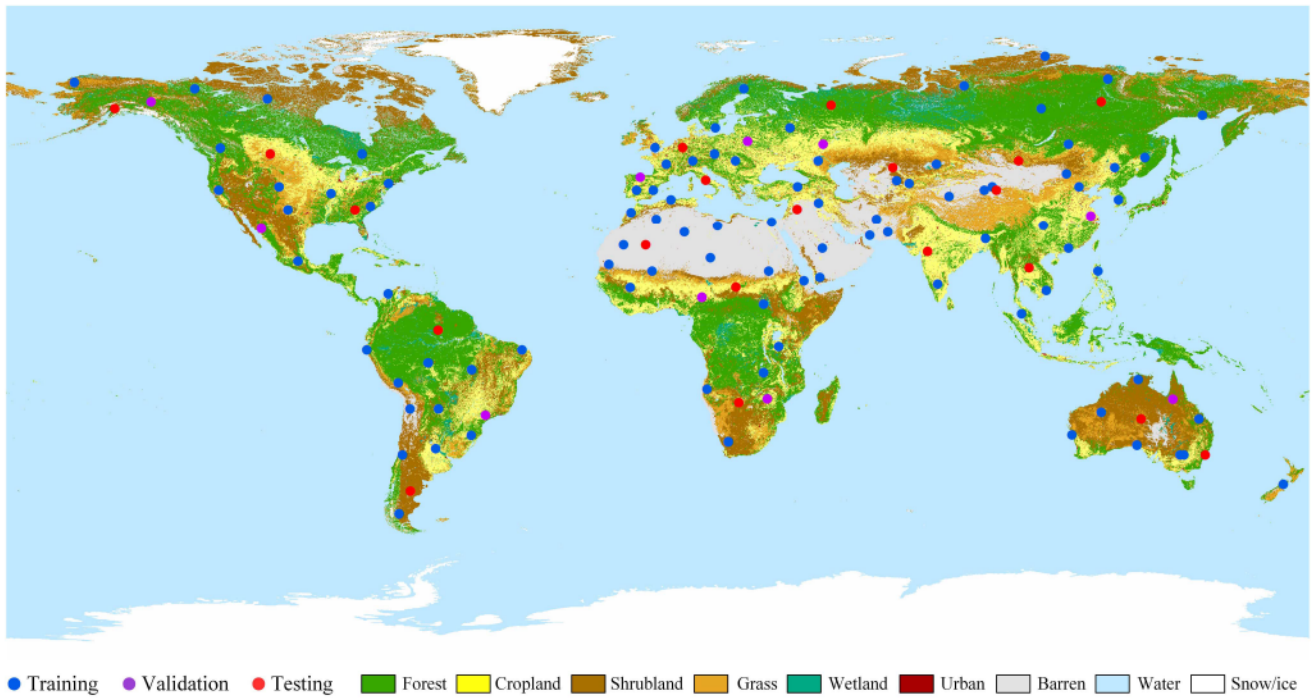


Fig. 2. Distribution of WHUS2-CRv dataset. Training, validation, and testing areas are marked in blue, purple, and red, respectively. The landcover background is derived from 300-m annual global land cover classification map in 2015 (ESA, 2017).

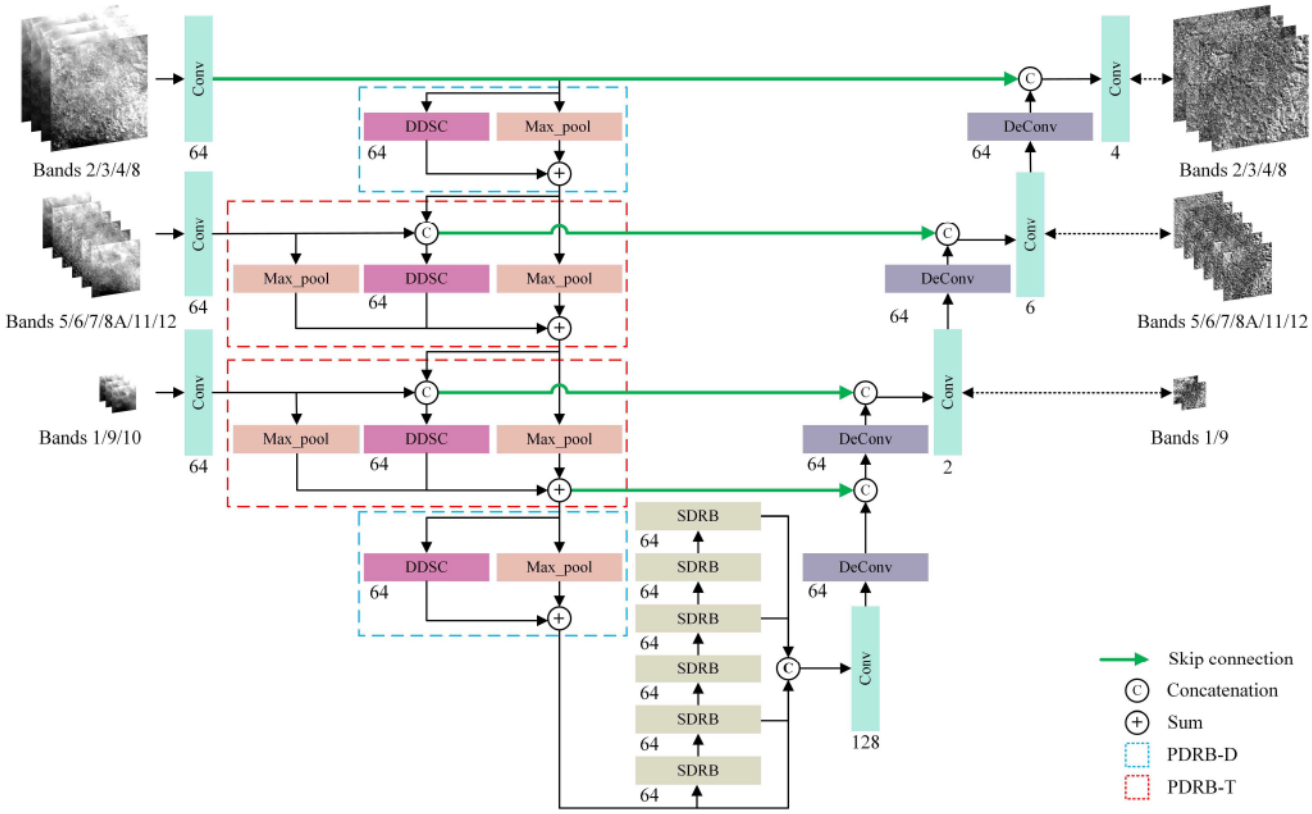


Fig. 3. Architecture of CR4S2. The number under/on each block is the corresponding number of feature maps.

Sampled regions for training and testing areas are spread over all continents; biomes and mainland cover classes to make the dataset representative. Finally, 123 cloudy and cloud-free image pairs covering about 1.47 million km^2 land surface were collected from the Copernicus Open Access Hub website. The acquisition dates of the 123 image pairs range from April 3, 2016 to January 13, 2021 and cover all seasons. From these, 93 image pairs were randomly selected for training, 10 for validation, and 20 for testing. Training, validation, and testing images are evenly distributed (see Fig. 2) to reduce biases in experimental results. All methods can be fully evaluated by WHUS2-CRv, which contain data with short interval, worldwide distribution, large area coverage, and full season coverage.

In order to eliminate atmospheric influence, Sen2Cor was run on all images to produce L2A data (surface reflectance), which was then used as the experimental data. Although the Cir band was not corrected by Sen2Cor, we still combined the L1C band Cir with L2A bands Ca/Wv into a multispectral image. Because memory requirements of CNN-based methods grow with input size, we cropped all experimental images into small patches without overlap. Since the spatial resolution of bands VNIR, VRE/NIRn/SWIR, and bands Ca/Wv/Cir are 10 m, 20 m, and 60 m respectively, the corresponding sliding window sizes and steps were set to 384 × 384, 192 × 192 and 64 × 64 pixels, respectively, which means that there are three multispectral images for each patch. This cropping strategy allows the coverage of 384 × 384 patch at 10 m, 192 × 192 patch at 20 m, and 64 × 64 patch at 60 m in each group to match. In this way, 24 450 patch-triplets were produced from 123 image pairs for cloudy and cloud-free images. The training, validation, and testing datasets contain 18 816, 1888, and 3746 pairs of cloudy and cloud-free patch-triplets generated from 93, 10, and 20 image pairs, respectively. The training samples were augmented by flipping horizontally and vertically and rotating at 90°, 180°, and 270°. Finally, 112 896 pairs of cloudy and cloud-free patch-triplets were obtained for training. The dataset is available on https://github.com/Neooolee/WHUS2-CRv.

III. Methodology

A. Framework of CR4S2

In most of the cloud removal methods, both the input (cloudy) and output (cloud-free) bands are rescaled to the same resolution when processing multiresolution bands. The traditional empirical rescaling operation with constant parameters not only introduces noise to the input but also to the output, because reference clear images contain noise after being rescaled. Therefore, CR4S2 is designed to handle the original Sentinel-2 multispectral bands at their native resolution for cloud removal without using any traditional rescaling strategy. CR4S2 is designed based on the encoder–decoder architecture, which is one of the most widely used architecture in DL and achieves very good performance in image processing.

Fig. 3 shows CR4S2 architecture, including three input branches in encoder and three corresponding output branches in decoder. These branches are used for processing the

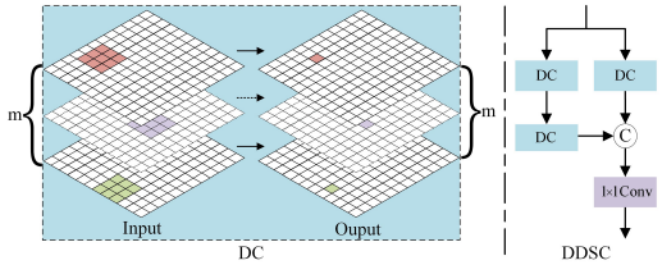


Fig. 4. Details of DDSC block. DC is a depthwise convolutional layer. *K* and *S* are the kernel size and stride, respectively.

patch-triplets, which contain different resolution bands: three input branches to avoid noise introduced by rescaling, and three output branches supervising CR4S2 by the original clear bands. Three CNN-based input branches are used to extract features from input bands. Then, for features from different depths, we designed two PDRBs (PDRB-D/PDRB-T), which can fuse features from current and previous branches and pass original features to the next block in the meantime. In order to reduce the parameters of the proposed method, a DDSC unit was designed to extract and fuse multiscale features from the input. Three output branches were used to produce multiresolution cloud removed bands that can be supervised by original resolution clear bands.

1) *Multiresolution branches:* It can be seen that the three input branches and three output branches are symmetric. The input/output groups include bands VNIR, bands VRE/NIRn/SWIR, and bands Ca/Wv/Cir from top to bottom. It should be noted that the bottom output group only includes bands Ca/Wv, because the atmosphere correction of Sen2Cor is not run on band Cir. However, since the Cirrus information can assist CR4S2 in locating areas affected by clouds when removing clouds in other bands, Cir band is used as one of the input bands.
2) *Top to bottom residual path:* Residual architecture has been proved to be effective in the DL field. In order to better transmit the information through the network, we designed two PDRBs (PDRB-D/T). The shared dilated convolution residual block (SDRB) introduced in [58] that uses shared convolution and residual architecture to solve the grid effect caused by dilated convolution, was adopted to obtain a larger receptive field without downsampling features anymore. As we can see in Fig. 3, the features extracted by the normal convolutional layer in each input branch can be directly passed to the bottom blocks through max-pooling layers in PDRB-D/T. The features input to the SDRB can also be passed to the next SDRB. In this way, features from each input branch can reach the bottom part of the encoder.

B. Components of CR4S2

1) DDSC Block: As shown in Fig. 4, the basic unit of DDSC is depthwise convolutional (DC) layer [59], which has as many convolution kernels and output feature maps as its input feature maps. Three DC layers construct two feature extraction paths in DDSC. The right path is a DC layer with kernel size $= 3 \times 3$ and stride $= 2$. The left path includes two DC layers in which the first DC layer has kernel size $= 3 \times 3$ and stride $= 2$; the second DC layer has kernel size $= 3 \times 3$ and stride $= 1$. The receptive fields of the right and left paths are 3×3 and 7×7, respectively. The outputs of the two paths are concatenated channelwise and then put into a 1×1 convolutional layer for multiscale features fusion. This operation is similar to the last operation of depthwise separable convolution [60]; thus, we call the designed block DDSC. It should be noted that the 1×1 convolutional layer outputs the same number of feature maps as that of the input of DDSC.

2) Parallel Downsample Residual Block: As shown in Fig. 3, there are two versions of PDRB in CR4S2, PDRB-D (in light blue dotted box), and PDRB-T (in red dotted box). PDRB-D contains one DDSC and one max-pooling layer and only has one input. The input of PDRB-D is processed by DDSC and max-pooling layer at the same time and their outputs are summed pixelwise to construct the residual architecture. PDRB-T is specially designed for processing multispectral bands at different spatial resolutions. This makes PDRB-T more suitable for multispectral remote sensing images at different resolutions than other residual modules in the computer vision field. PDRB-T contains one DDSC and two max-pooling layers and has two inputs from previous and current branches. That is why PDRB-T is only used in the second and third input branches. The two inputs are concatenated channelwise and then input into DDSC for feature fusion. One of the max-pooling layers is used for downsampling the feature maps from the current branch and the other max-pooling layer is used for downsampling the feature maps from the previous branch. The outputs of DDSC and two max-pooling layers in PDRB-T are summed pixelwise to construct the residual architecture.

C. Multioptimization Loss

L1 loss is usually used as the loss function in many image restoration networks, because it can prevent blurry images [61]. In CR4S2, the L1 loss between cloud removed images and reference images is calculated pixelwise to restore information at a pixel level. In order to optimize the restoration of the edge information, the L1 loss between the edges of cloud removed and reference images is also calculated. The L1 losses of image pixel and edge for a single band are calculated as follows:

$$L_k(z, x) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} |R(z)_k - x_k| \tag{1}$$

$$L_{\text{edge}-k}(z, x) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} |\nabla(R(z)_k) - \nabla(x_k)| \tag{2}$$

where z is the input image, x_k the kth band in reference image x, $R(z)_k$ the kth band in cloud removed image, ∇ is the edge operator (right pixel minus left pixel, down pixel minus up pixel), and m and n are the width and height of x_k. The pixelwise loss L_k aims to help CR4S2 restore information in the low frequency while $L_{\text{edge}-k}$ aims to restore the information in high frequency. The cloudy and clear remote sensing images are

acquired at different dates, spectral (low-frequency information) may change but edges (high-frequency information) will not change as much generally, except on abrupt changes. Therefore, $L_{\text{edge}-k}$ can provide more accurate supervision on CR4S2.

The values of clear pixels should not be changed when removing clouds, and this constraint can be used to optimize the preservation performance of clear pixels at a pixel level. This idea was proposed independently in [52] and [16]. An optimization process is also designed to preserve the information of clear pixels not only at pixel-level but also on the edges. The reference clear image is put into CR4S2, to make sure the corresponding "cloud removed" image is the same as the input clear image. The L_1 losses for a single band are calculated the same in a clear image as in a cloudy image. By combining the L_1 losses for cloudy and clear images, we give the loss function of a single band as follows:

$$L_{\text{total}-k}\ (c,n) = L_k\ (c,n) + \lambda_1 L_{\text{edge}-k}\,(c,n) + \lambda_2(L_k\,(n,n) + \lambda_1 L_{\text{edge}-k}\,(n,n)) \quad (3)$$

where c and n are cloudy and clear images, respectively. λ_i is the weight balance factors. $\lambda_1 = 0.01$, $\lambda_2 = 0.1$ (term for clear images). For each band, we use (3) to calculate the loss. The total loss of CR4S2 is as follows:

$$L_{\text{total}} = \frac{1}{B}\sum_{k=1}^{B} L_{\text{total}-k}\,(c,n) \quad (4)$$

where B is the number of bands in the cloud removed image. L_{total} is used for optimizing the parameters of CR4S2. After CR4S2 is well trained, it can remove thin clouds in images while preserving the information of clear pixels.

IV. Experiments and Discussion

A. Experimental Setting

1) Baseline Methods: Because DL-based thick cloud removal methods require either multitemporal images or auxiliary data such as SAR, they are not included as baseline methods. In order to evaluate CR4S2 performance, three deep learning-based (DL-based) thin cloud removal methods, RSC-Net [27], FCTF-Net [56], and RSDehazeNet [55] and two traditional methods, DCP [34] and Color Ellipsoid Prior (CEP) [62] were selected for comparison. The DL-based baseline methods were originally proposed for thin cloud removal in remote sensing images and the traditional baseline methods proved very effective for haze removal in natural images. For traditional baseline methods, we kept default parameters and directly run them on testing images. The training and testing details of DL-based methods were described in the following paragraphs.

2) Hardware Environment and Hyperparameters: The training and testing experiments were both conducted on Windows 11 operating system on an 11th Gen Intel (R) Core (TM) i9-11900KF @ 3.50 GHz, with an NVIDIA GeForce GTX 3080Ti with 12-GB memory (7 GB was required for training CR4S2 with batch size = 1). RSC-Net, RSDehazeNet, and CR4S2 are based on the Tensorflow platform with Python 3.8.8, and FCTF-Net is based on PyTorch platform. DCP and CEP do not need training and only use CPU for computation. The codes of all baseline methods are downloaded from their GitHub repositories, except RSC-Net is implemented according to the corresponding article. For parameter optimization of CR4S2, Adam-optimizer [63] is adopted using the following hyperparameters: $\beta_1 = 0.9$, $\beta_2 = 0.999$, initial learning rate = 0.0002, and exponential decay at decay rate = 0.96. For other DL-based baseline methods, the hyperparameters were set as default and the validation dataset was used for setting up early-stop for all DL-based methods.

3) Data Preprocessing: Since our goal is to remove thin cloud in all Sentinel-2 bands except band 10, and baseline methods cannot process the original multispectral bands at different resolutions, bands VRE/NIRn/SWIR and Ca/Wv/Cir were rescaled to the same resolution as that of bands VNIR before being put into DL-based methods. While the original patch-triplets were directly put into CR4S2 due to its network architecture, the surface reflectance values were clipped to [0, 10 000] and then normalized to [0, 1] before being processed by all DL-based methods. Band Cir input was taken from L1C product and ignored in the outputs by all methods. Therefore, the DL-based methods took 13 bands as input, but only output 12 bands. Traditional methods were run on bands VNIR only, because they were designed for RGB natural images.

4) Accuracy Indexes: Structural similarity index measurement (SSIM), peak signal-to-noise ratio (PSNR), mean absolute error (MAE), normalized root-mean-square error (nRMSE), and spectral angle mapper (SAM) were taken as the quantitative evaluation measures. The SSIM evaluates the cloud removal performance in the view of the whole image. The PSNR, MAE, and nRMSE are calculated for pixelwise comparison between cloud removed and clear images. SAM shows the reconstruction ability in the spectral domain.

B. Comparison With DL-Based Methods

1) Effectiveness: Because L2A product does not contain Cirrus band, we only evaluated the cloud removal performances of all DL-based methods on bands VNIR, VRE/NIRn/SWIR, and Ca/Wv. Table II shows the quantitative results of different methods on all testing samples. It can be seen that CR4S2 always obtained the highest PSNR and SSIM among all DL-based methods on 12 bands. The PSNR for CR4S2 ranges from 30.96 (band 8A) to 39.55 (band 1) and SSIM ranges from 0.9002 (band 8) to 0.9443 (band 3), respectively. CR4S2 obtains at least 0.58 higher PSNR, 0.0056 higher SSIM, and 0.0041 higher nRMSE than DL-based baseline methods. RSC-Net performs the worst among all methods, with lowest PSNR and SSIM 29.32 (band 8A/11) and 0.8459 (band 1), respectively.

FCTF-Net obtained the highest nRMSE (1.7052) compared to other methods, indicating lower performance. On the contrary, all DL-based methods obtained the highest PSNR and SSIM on band 1 and band 3. When analyzing the performance of a given model across all bands, both CR4S2 and RSDehazeNet obtained their lowest nRMSE on band 8A. Among all bands, the hardest bands for RSC-Net and for FCTF-Net were band 6 and band 11, respectively. Results in Table II also show that, as wavelength

TABLE II
AVERAGE PSNR, SSIM, AND NRMSE FOR DL-BASED METHODS OVER ALL TESTING IMAGES (3746 TESTING SAMPLES)

Index	Method	B1	B2	B3	B4	B5	B6	B7	B8	B8A	B9	B11	B12
PSNR	RSC-Net	35.06	34.12	33.51	31.24	32.34	30.51	29.9	29.43	29.32	30.07	29.32	29.40
	FCTF-Net	37.19	36.57	35.09	33.77	32.81	30.78	30.25	29.85	29.72	30.50	30.69	32.17
	RSDehazeNet	37.64	37.19	36.08	34.05	33.42	31.48	30.76	29.74	30.20	30.74	30.28	31.50
	CR4S2	**39.55**	**38.17**	**37.05**	**35.55**	**34.37**	**32.15**	**31.40**	**31.00**	**30.96**	**31.32**	**31.47**	**33.31**
SSIM	RSC-Net	0.8459	0.8823	0.9203	0.8919	0.9129	0.8991	0.8930	0.8852	0.8910	0.9010	0.9053	0.8937
	FCTF-Net	0.8807	0.9103	0.9286	0.9122	0.9167	0.9031	0.8982	0.8842	0.8959	0.9046	0.9109	0.9105
	RSDehazeNet	0.8909	0.9203	0.9387	0.9180	0.9221	0.9046	0.9004	0.8893	0.9008	0.9035	0.9145	0.9162
	CR4S2	**0.9185**	**0.9315**	**0.9443**	**0.9302**	**0.9355**	**0.9236**	**0.9195**	**0.9002**	**0.9200**	**0.9252**	**0.9317**	**0.9334**
nRMSE	RSC-Net	1.5915	0.1770	0.0686	0.1058	0.0408	0.0303	0.0317	0.0352	0.0345	0.0370	0.0617	0.1061
	FCTF-Net	1.7052	0.1281	0.0608	0.0633	0.0425	0.0363	0.0356	0.0318	0.0336	0.0336	0.0312	0.0385
	RSDehazeNet	0.9368	0.1039	0.0488	0.0619	0.0405	0.0361	0.0353	0.0370	0.0336	0.0319	0.0369	0.0442
	CR4S2	**0.6362**	**0.0760**	**0.0376**	**0.0426**	**0.0304**	**0.0262**	**0.0255**	**0.0251**	**0.0245**	**0.0264**	**0.0263**	**0.0324**

The best values are marked in bold.

decreases from 0.865 nm (band 8A) to 0.443 nm (band 1), CR4S2, RSC-Net, and FCTF-Net keep on performing well on PSNR. Unlike RSC-Net, the performance of CR4S2, RSC-Net, and FCTF-Net on SSIM improve as wavelength increases from 0.842 nm (band 8) to 2.190 nm (band 12). The nRMSE of CR4S2 decreases as wavelength increases from 0.665 nm (band 4) to 0.865 nm (band 8A) while baseline DL-based methods cannot keep this.

Fig. 5 shows the visual results of all DL-based methods on all bands. CR4S2 obtains more visually similar results on visible bands (see Fig. 5, row 2) to reference images than DL-based baseline methods. RSC-Net produces a lower surface reflectance than the reference image. CR4S2 always achieves the lowest MAE on all bands. For (a), RSC-Net gets the highest MAE on bands 2/3/4/6/11/12/1, FCTF-Net performs the worst on bands band 8/5/9 and RSDehazeNet obtains the highest MAE on bands 7/8A among all methods. It should be noted that there are six regions (marked in red rectangles) in which all methods get very large MAE on all bands, because the land cover types completely changed in these regions. For (b), CR4S2 achieved more acceptable MAE on all bands than other DL-based methods. Unlike in (a), RSDehazeNet performed worse on bands 5/11/12/9, but better on bands 2/3/4/7/8A/1 than RSC-Net in (b). Results in Fig. 5 show that, as wavelength decreases from 0.665 nm (band 4) to 0.443 nm (band 1), the performances of all DL-based methods keep on improving on MAE.

The PSNR and SSIM of DL-based methods on Fig. 5(a) and (b) are shown in Fig. 6. It can be seen that CR4S2 gets the highest PSNR and SSIM among all methods on all bands in (a) and (b). In particular, CR4S2 obtains a significantly higher PSNR value than DL-based baseline methods on bands 5/11 than other bands in (a). CR4S2 outperforms other methods on all bands in (b). The increase of SSIM value with CR4S2 is higher on bands 5/4/11/12 than other bands in (a). CR4S2 also performs significantly better than DL-based baseline methods on bands 2/3/5/6/7/8A/11/12/1/9 than on other bands in (b). Combining quantitative and qualitative results, we can see that CR4S2 can not only restore more spectral information but also texture information than DL-based baseline methods.

2) Efficiency: From Table III, it can be seen that CR4S2 has the most parameters but lowest computational complexity among all DL-based methods. This is because PDRB modules in CR4S2 construct a top-to-down highway for information transfer in its encoder. The computational complexity of FCTF-Net ranks second. This may be because FCTF-Net also has a top-to-down highway in its encoder. RSC-Net has the least parameters, which is because RSC-Net only includes 5 symmetrical convolutional and deconvolutional pairs, in which each convolutional and deconvolutional layer has only 32 feature maps. The number of parameters in RSDehazeNet is only about 18.2% of that in CR4S2; however, RSDehazeNet run about three times slower than CR4S2.

TABLE III
PARAMETERS AND FLOPS OF DL-BASED MODELS WITH 384 × 384×13 AS INPUT SIZE AND 384 × 384×12 AS OUTPUT SIZE

Index	RSC-Net	FCTF-Net	RSDehazeNet	CR4S2
Parameters	**112 332**	170 108	268 828	1485 796
FLOPs	33.1×10^9	23.658×10^9	64.2×10^9	$\mathbf{21.1\times10^9}$

The best values are marked in bold.

C. Influence of Cirrus Band

In order to analyze the improvement on thin cloud removal introduced by Cirrus band, models taking 12 bands (1/2/3/4/5/6/7/8/8A/9/11/12) as input and output (12-bands models) were trained and tested. The PSNR and SSIM of models without considering the Cirrus band as input (models with -12 as suffix) and the corresponding performance improvement δ when adding the Cirrus band are given in Fig. 7. We can see that CR4S2-12 can still obtain higher PSNR and SSIM than RSC-Net-12, FCTF-Net-12, and RSDehazeNet-12. RSC-Net-12 still performs the worst on most bands. From Fig. 7(a), it can be seen that CR4S2 gets higher improvement than RSC-Net, FCTF-Net, and RSDehazeNet on bands 2/3/4/11/12/1/9. RSC-Net even gets lower PSNR than RSC-Net-12 on bands 2/3/4/12/1. The signal preservation performance of FCTF-Net and RSDehazeNet also degrades on bands 2/3/5/1 and bands 4/8/11/12, respectively. Fig. 7(b) shows that the improvement of CR4S2 on SSIM is higher than other methods on bands 2/3/4/8/6/11/12/1/9 but lower than RSDehazeNet on other bands. This is because CR4S2-12 obtains a much higher SSIM than RSC-Net-12, FCTF-Net-12, and RSDehazeNet-12. Therefore, there is less room for improvement for CR4S2 on these bands. The capability of all methods to restore structure has been improved by taking the Cirrus band as input, except that RSC-Net and

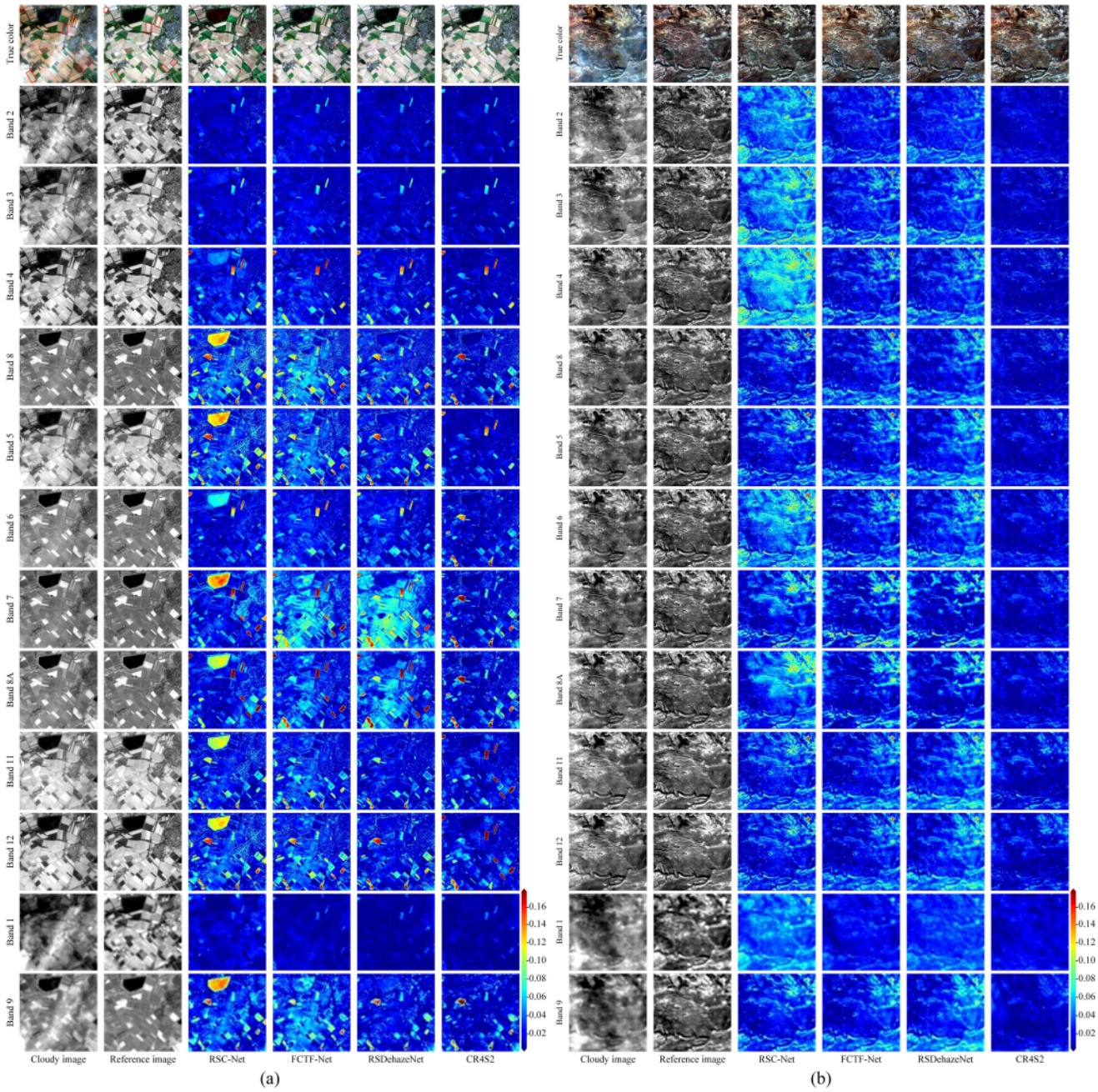


Fig. 5. Visual comparison results of (a) farmland and (b) barren samples (details can be found in Table VII). Columns 1 and 2 are cloudy and cloud-free images, respectively. True color is a true color image, band *n* are MAEs of bands *n*, respectively.

FCTF-Net fail on bands 2/3/4/5/6/12/1 and 2/6/7/8A/11/12/1/9, respectively.

The average nRMSE and δ (performance index of 13-bands models minus the performance index of 12-bands models) of all DL-based methods are shown in Table IV. CR4S2-12 achieves better performance than DL-based methods that take 12 bands as input, except on band 6/7/11/12 and 9 for which FCTF-12 and RSDehazeNet-12 perform marginally better, respectively. When taking the Cirrus band into consideration, CR4S2 gets lower nRMSE than CR4S2-12 on all bands, and the lowest nRMSE overall. However, DL-based baseline methods cannot always reduce nRMSE on all bands, such as bands 11/12 (RSC-Net), bands 2/3/4/5/6/712/1/9 (FCTF-Net), and band 4/8/11/12/9 (RS-DehazeNet). This demonstrates that CR4S2 can make use of the Cirrus band to improve thin cloud removal performance on other bands at pixel-level.

Three different samples with land cover types and cloud thickness are shown in Fig. 8, from which we can see that the cloud effect in vegetation scene (a) cannot be eliminated either by RSC-Net-12, RSC-Net, and FCTF-Net-12. FCTF-Net- and RSDehazeNet-based methods can remove clouds, but the results vary with input bands. The barren region is transferred into red by RSC-Net-12, RSC-Net, and FCTF-Net-12. The region seriously affected by thin cloud is also changed after being

TABLE IV
AVERAGE NRMSE FOR 12-BANDS MODELS, AND δ OF 13-BANDS MODELS OVER 12 BANDS MODELS OVER ALL TESTING IMAGES (3746 TESTING SAMPLES)

Index	Method	B2	B3	B4	B8	B5	B6	B7	B8A	B11	B12	B1	B9
nRMSE	RSC-Net-12	0.2857	0.0781	0.1129	0.0406	0.0467	0.0367	0.0389	0.0413	0.0601	0.1007	4.9248	0.0426
	FCTF-Net-12	0.1207	0.0594	0.0628	0.0354	0.0410	**0.0353**	**0.0346**	0.0340	**0.0319**	**0.0367**	1.4860	0.0314
	RSDehazeNet-12	0.1236	0.0549	0.0614	0.0364	0.0441	0.0426	0.0398	0.0364	0.0333	0.0410	1.5082	**0.0305**
	CR4S2-12	**0.0964**	**0.0502**	**0.0538**	**0.0347**	**0.0398**	0.0359	0.0353	**0.0338**	0.0352	0.0406	**1.4313**	0.0342
δ	RSC-Net	**-0.1087**	-0.0095	-0.0071	-0.0054	-0.0059	-0.0064	-0.0072	-0.0068	0.0016	0.0054	**-3.3333**	-0.0056
	FCTF-Net	0.0074	0.0014	0.0005	-0.0036	0.0015	0.001	0.001	-0.0004	-0.0007	0.0018	0.2192	0.0022
	RSDehazeNet	-0.0197	-0.0061	0.0005	0.0006	-0.0036	-0.0065	-0.0045	-0.0028	0.0036	0.0032	-0.5714	0.0014
	CR4S2	-0.0204	**-0.0126**	**-0.0112**	**-0.0096**	**-0.0094**	**-0.0097**	**-0.0098**	**-0.0093**	**-0.0089**	**-0.0082**	-0.7951	**-0.0078**

The best values are marked in bold.

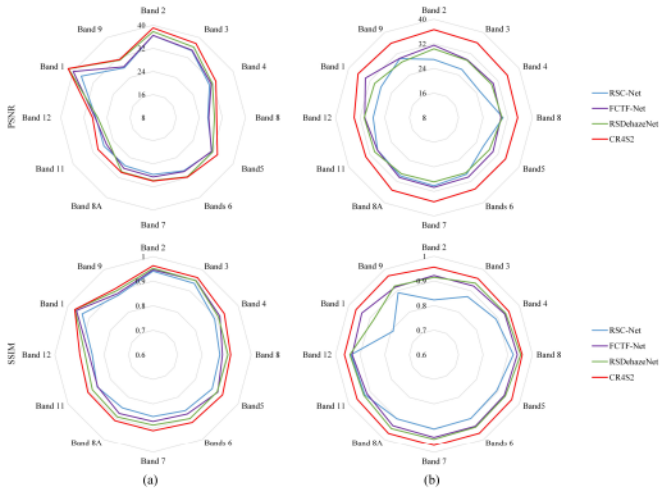


Fig. 6. PSNR and SSIM of different methods on 12 bands corresponding to Fig. 5(a) and (b).

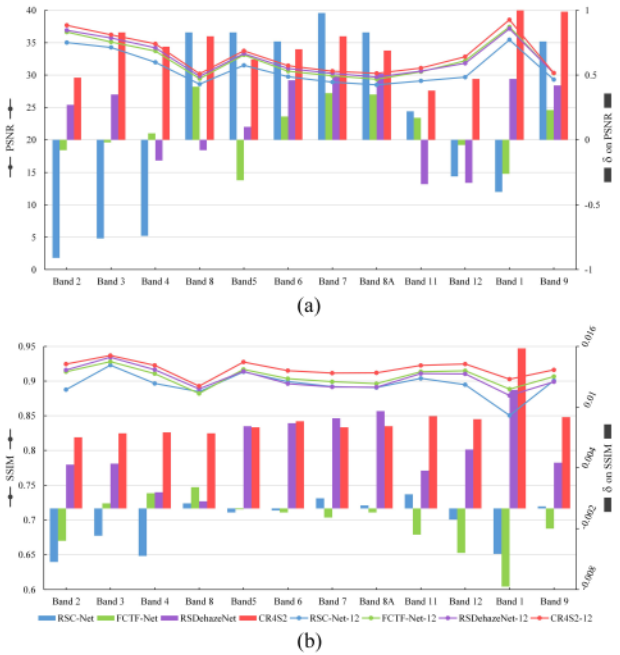


Fig. 7. (a) Average PSNR and (b) SSIM of 12 bands-based methods (marked in line) and corresponding δ on PSNR and SSIM (marked in bar) of 13 bands-based models over 12-bands based models on all testing samples.

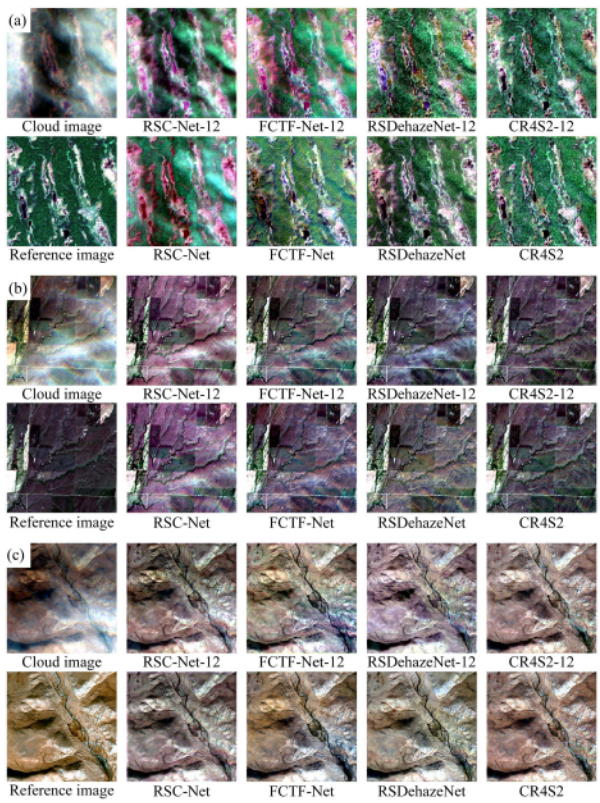


Fig. 8. Visual results of on samples in Fig. 1. (a) Vegetation. (b) Farmland. (c) Barren mountain.

processed by RSDehazeNet-12, RSDehazeNet, and FCTF-Net. However, both the results of CR4S2-12 and CR4S2 retain the spectral features in the region. For the farmland scene (b), Only RSDehazeNet, CR4S2-12, and CR4S2 can remove cloud effects visually. RSC-Net-12, RC-Net, and FCTF-Net-based methods change the spectral features in purple regions. It can also be seen that RSDehazeNet performs much better than RSDehazeNet-12 with the help of the Cirrus band. In the barren mountain region (c), the cloud effect can be removed by all methods except FCTF-Net-12 visually. But the spectral features are changed by RSC-Net-12, RSC-Net, and RSDehazeNet-12 in the left region. From the results in (a), (b), and (c), it can be seen that the cloud is further removed by all methods when taking the Cirrus band as input. Both CR4S2-12 and CR4S2 can not only remove clouds visually but also preserve features in the spectral domain.

TABLE V
AVERAGE PSNR, SSIM, AND NRMSE FOR DIFFERENT METHODS OVER ALL TESTING IMAGES (3746 TESTING SAMPLES)

Index	Method	B2	B3	B4	B8
PSNR	CEP	24.73	23.85	22.18	19.42
	DCP	29.23	30.02	27.04	23.73
	RSC-Net-4	34.23	32.89	31.38	28.40
	FCTF-Net-4	35.73	34.58	32.99	29.46
	RSDehazeNet-4	36.92	35.61	33.89	29.29
	CR4S2-4	**37.30**	**36.11**	**34.69**	**30.51**
SSIM	CEP	0.6517	0.7304	0.6434	0.6669
	DCP	0.8134	0.8723	0.8175	0.8219
	RSC-Net-4	0.8792	0.9136	0.8842	0.8862
	FCTF-Net-4	0.8966	0.9227	0.9040	0.8852
	RSDehazeNet-4	0.9180	0.9352	0.9122	0.8863
	CR4S2-4	**0.3746**	**0.9412**	**0.9278**	**0.9013**
nRMSE	CEP	2.3712	0.8796	1.0126	0.2443
	DCP	0.6979	0.1321	0.2068	0.0937
	RSC-Net-4	0.2515	0.0842	0.1448	0.0393
	FCTF-Net-4	0.1693	0.0677	0.0756	0.0298
	RSDehazeNet-4	0.1206	0.0526	0.0634	0.0355
	CR4S2-4	**0.1056**	**0.0468**	**0.0493**	**0.0246**

The best values are marked in bold.

TABLE VI
AVERAGE SAM (°) FOR DIFFERENT METHODS OVER ALL TESTING IMAGES (3746 TESTING SAMPLES)

Index	Method	VNIR	VRE/NIRn/SWIR	Ca/Wv
SAM	CEP	12.771	/	/
	DCP	7.1273	/	/
	RSC-Net-4	5.5211	/	/
	FCTF-Net-4	4.0164	/	/
	RSDehazeNet-4	3.8266	/	/
	CR4S2-4	**3.3284**	/	/
	RSC-Net-12	5.024	4.1010	4.3627
	FCTF-Net-12	3.6955	3.2118	3.1368
	RSDehazeNet-12	3.6841	3.4584	3.3476
	CR4S2-12	**3.2797**	**3.0387**	**3.0261**
	RSC-Net	5.8411	4.0159	4.4482
	FCTF-Net	3.7620	3.2902	3.2494
	RSDehazeNet	3.5614	3.1610	2.9621
	CR4S2	**3.0454**	**2.9004**	**2.5676**

The best values are marked in bold.

D. Evaluation Using Only VNIR Bands

Since there are many satellites that only acquire bands in the visible and near-infrared part of the spectrum (VNIR), which are the most used for remote sensing applications, the performance of the CR4S2-based model trained on VNIR bands was also evaluated.

We use RSC-Net-4, FCTF-Net-4, RSDehazeNet-4, and CR4S2-4 to present the DL-based methods that only take VNIR bands as input/output. Table V shows that DL-based methods perform much better than traditional methods. CR4S2-4 achieves the best PSNR, SSIM, and nRMSE performance on VNIR bands among all methods. This is because even without the other two input branches, max-pooling in PDRB-T will be removed and PDRB-T becomes PDRB-D, which means CR4S2-4 model can still extract and fuse multiscale features at each level when only inputting VNIR bands. CEP always ranks last on these bands. FCFT-Net-4 always gets better performance than RSC-Net-4 on bands 2/3/4 and RSDehazeNet-4 obtains quite close PSNR, SSIM, and nRMSE to CR4S2-4. This may be because RSDehazeNet-4 also fuses features from different levels for thin cloud removal.

From the true color composited results in Fig. 9, we can see that CR4S2-4 achieves the most similar visual result to the reference image among all methods. Although CEP can remove more clouds than DCP, neither can remove clouds completely. The outputs of CEP, DCP, and RSC-Net-4 have lower pixel values than the reference image. Although the results of FCTF-Net-4 and RSDehazeNet-4 contain no clouds visually, they produced higher surface reflectance than the reference image on bands 2/3/4. The MAE increases from bands 2 to 8 for all methods. DCP performs the worst on VNIR bands among all methods, with a much higher MAE in highlight regions (such as urban and barren) than other methods. DCP and RSC-Net-4 get higher MAE in barren areas than farmland and urban areas on bands 2/3/4. It can also be found that MAE in the farmland region on band 8 is higher than bands 2/3/4 for all methods. This may be because the vegetation in farmland has changed in 10 days. CR4S2-4 obtains the highest PSNR and SSIM on VNIR bands, except that RSC-Net-4 obtains a marginal improvement of SSIM (0.0082) on band 8.

E. Analysis of Spectral Preservation

In order to evaluate the spectral preservation performance of CR4S2-based methods, SAM was calculated pixelwise. SAM values for bands VNIR VER/NIRn/SWIR and Ca/Wv were calculated separately, because the spatial resolutions of these band groups are different. Table VI shows the average SAM for all methods on all testing samples. It can be seen that CR4S2-based methods can always obtain the best SAM when taking different bands as input. For methods taking bands VNIR as input, CEP and DCP perform much worse than DL-based methods. The SAM values of CEP and DCP are almost 3.7 and 2 times larger than that of CR4S2-4. Combining results of methods with 12 bands as inputs, we can see that the spectral preservation ability of DL-based methods is improved when taking bands VRE/NIRn/SWIR and Ca/Wv into consideration. The gain in spectral preservation ability for DL-based methods is further increased when taking the Cirrus band as additional input.

Additionally, the average SAMs for samples in Fig. 1 are also shown in Fig. 10, from which we can see that CEP can barely preserve the spectral information. The SAM of DCP is close to that of DCP. FCTF-Net-4 get much higher SAM than RSC-Net-4, FCTF-Net-4, and CR4S2-4. CR4S2-4 obtains a little better spectral preservation performance than FCTF-Net-4 and RSDehazeNet-4. RSC-Net-12 even get worse results than RSC-Net-4 on bands VNIR. The spectral preservation ability of all models cannot always be improved as the number of input bands increases except CR4S2-based models. CR4S2-based models always achieve a competitive spectral preservation performance on the samples in Fig. 1.

F. Influence of Multioptimization Loss

As mentioned in Section III, the losses for clear image and gradient were introduced to construct the multioptimization loss, which was used to improve the performance of CR4S2. To

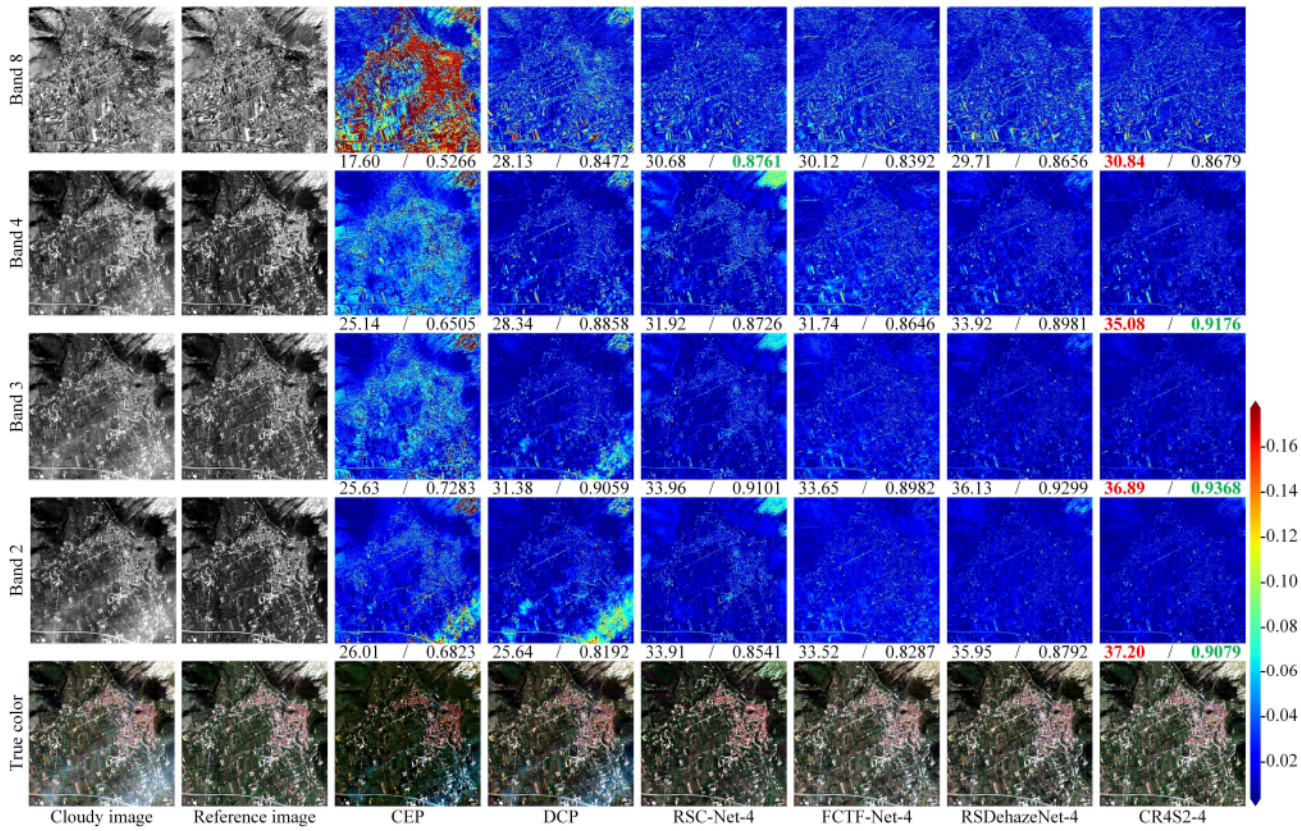


Fig. 9. Results on farmland and urban areas (details can be found in Table VII). Columns 1–8 are cloudy and reference images, results of CEP, DCP, RSC-Net-4, FCTF-Net-4, RSDehazeNet-4, and CR4S2-4, respectively, using a color scale for better visualization of results. True colors are true color composites. Bands 2/3/4/8 are the MAE maps of bands 2/3/4/8. The numbers below each MAE map are PSNR (left) and SSIM (right) for the corresponding band, respectively. The highest values for PSNR and SSIM are marked in red and green, respectively.

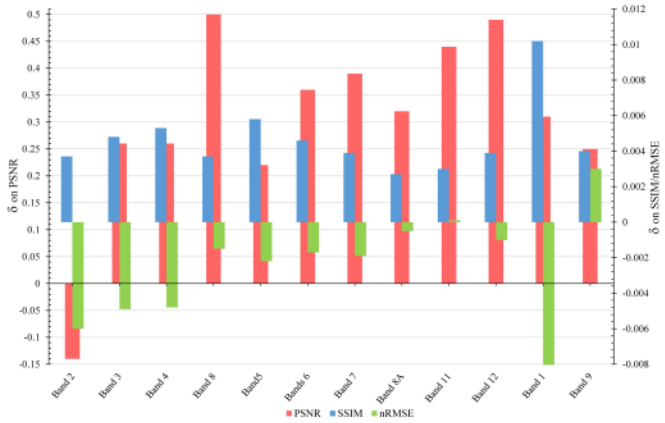


Fig. 10. Average performance improvement introduced by multioptimization loss for PSNR (in red), SSIM (blue), and nRMSE (green) on all test samples. The higher δ on PSNR and SSIM the better, and the lower δ on nRMSE the better.

evaluate the effectiveness of multioptimization loss, a model CR4S2-noopt was trained and tested, with its parameters updated only with cloud images and $L_k(c, n)$. Fig. 11 shows the quantitative comparison results on all test samples between CR4S2 and CR4S2-noopt. We can see that CR4S2 got higher PSNR than CR4S2-noopt on all bands except band 2 (less than 0.14 PSNR). CR4S2 obtained at least 0.21 more in PSNR than CR4S2-noopt on other 11 bands. We can also see that CR4S2 always performs better than CR4S2-noopt in SSIM on all bands. This demonstrates that multioptimization loss can restore more structural features on all bands. The nRMSE of CR4S2-noopt was improved with multioptimization loss on all bands except band 11 (0.0001) and 9 (0.003). The SSIM increases slightly (0.0102) and nRMSE decreases significantly (0.3839) on band 1, which is more affected by thin clouds. This means that multioptimization loss can effectively improve thin cloud removal performance on the coastal band.

In order to analyze the influence of multioptimization loss on spectral preservation, the surface reflectance was calculated pixelwise. Since the spatial resolutions of VNIR, VRE/NIRn/SWIR, and Ca/Wv are different, we selected a central pixel (row 33 of 64 and column 33 of 64) in Ca/Wv bands and corresponding central windows in VNIR (from row 33 6 to 34 6 of 384 and from column 33 6 to 34 6 of 384), and VRE/NIRn/SWIR (from row 33 3 to 34 3 of 192 and from column 33 3 to 34 3 of 192) bands to ensure that the surface reflectances in different bands are from the same location. As shown in Fig. 12, we can see that the result of CR4S2 on vegetation scene (a) is in good agreement with the reference. However, CR4S2-noopt produces higher surface reflectance than reference on bands 6/7/8/8A/9. For farmland scene (b),

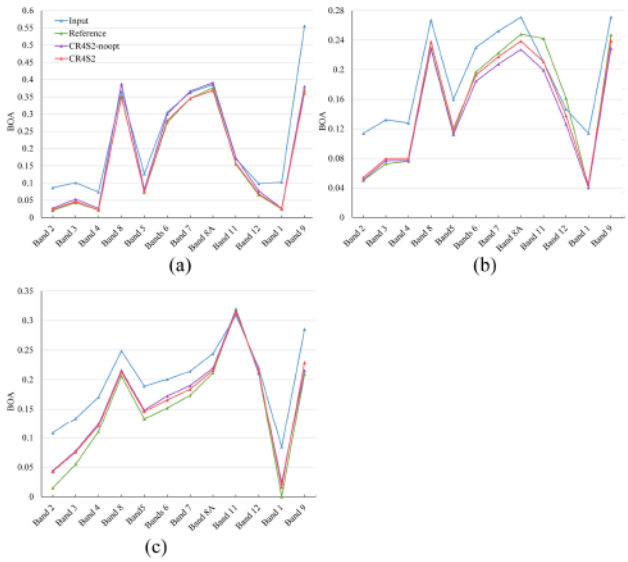


Fig. 11. Average pixel spectra of the central pixels (6× 6 for VNIR bands, 3 × 3 for VRE/NIRn/SWIR bands, and 1 × 1 for Ca/Wv bands) in the respective input, target, and output images of CR4S2 and CR4S2-noopt for samples in Fig. 1. (a) Vegetation. (b) Farmland. (c) Barren mountain.

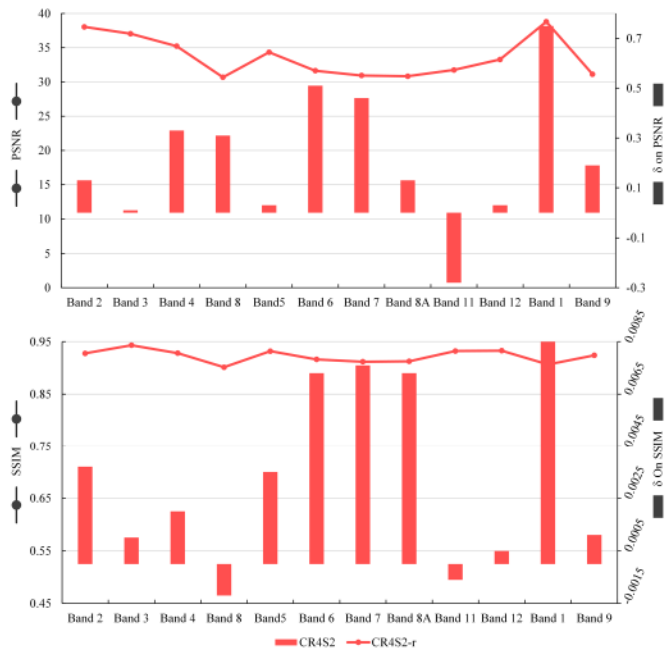


Fig. 12. (a) Average PSNR and (b) SSIM of CR4S2-r (marked in line) and corresponding δ on PSNR and SSIM (marked in bar) of CR4S2 over CR4S2-r on all testing samples.

although both CR4S2 and CR4S2-noopt get lower surface reflectance than reference on bands 8A/11/12/9, the result of CR4S2 is closer to reference than that of CR4S2-noopt. CR4S2 and CR4S2-noopt get higher surface reflectance than reference on bands 2/3/4/5/6/7/1/9 in the barren mountain scene (c), but CR4S2 performs slightly better on bands 6/7. Combining the results on the three samples, it can be seen that the proposed multioptimization loss is effective for thin cloud removal in most bands on the samples in Fig. 1.

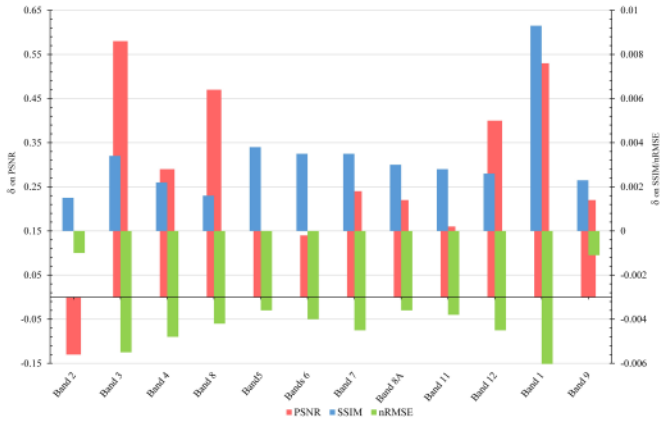


Fig. 13. Average performance improvement introduced by PDRB-D/PDRB-T for PSNR (in red), SSIM (blue), and nRMSE (green) on all test samples. The higher δ on PSNR and SSIM the better, and the lower δ on nRMSE the better.

G. *Effectiveness of Multi-Input/Output Branches*

The key contribution of this work is using multi-input/output branches to process bands at different spatial resolutions, rather than rescaling the bands to the same size by human-designed interpolation algorithms that will introduce noise. To prove the effectiveness of the proposed multi-input/output branches, a CR4S2-r model that uses the rescaled multiresolution bands as the inputs and outputs was trained. Fig. 10 shows the average PSNR and SSIM of CR4S2-r and corresponding δ on PSNR and SSIM of CR4S2 over CR4S2-r on all testing samples. CR4S2 outperforms CR4S2-r in PSNR on all bands except band 11 and in SSIM on all bands except bands 8/11. This demonstrates that the proposed multi-input/output branches can help improve CR4S2 performance both in signal and structure restoration on most bands.

H. *Superiority of PDRB*

In this article, two parallel downsample residual blocks PDRB (PDRB-D/PDRB-T) were designed for multiscale feature fusion. To analyze the superiority of PDRB over normal convolution layer, all PDRB in CR4S2 were replaced with convolution layers (CR4S2-noPDRB). As shown in Fig. 13, CR4S2 performs better than CR4S2-noPDRB in PSNR, SSIM, and nRMSE on all bands, except a little worse in PSNR on band 2. This demonstrated that the proposed PDRB can improve thin cloud removal performance of CR4S2 effectively.

I. *Overall Assessment of CR4S2 Performance*

CR4S2 is still more effective than baseline methods when taking only VNIR bands as input (see Table V). The thin cloud removal performance of CR4S2 method is improved when taking more bands as input. Table IV shows that the Cirrus band, which contains the least land surface information and is the most influenced by clouds, is helpful for CR4S2 to remove thin clouds in other bands. The statistical results in Fig. 11 prove that the proposed multioptimization loss can further improve

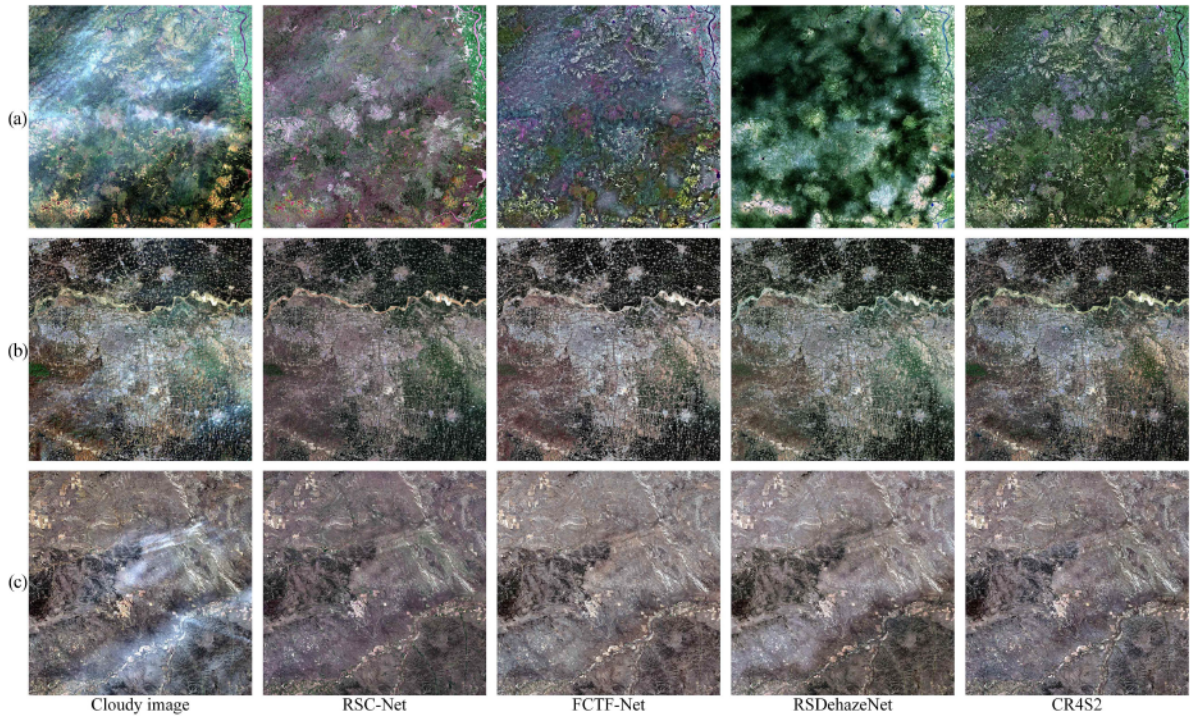


Fig. 14. Visual results of three whole Sentinel-2 scenes. (a) Vegetation (S2A_MSIL1C_20220729T072631_N0400_R049_T41WNM_20220729T074342). (b) Urban (S2B_MSIL1C_20210412T030539_N0300_R075_T49SGU_20210412T045545). (c) Barren (S2A_MSIL1C_20210417T180911_N0300_R084_T12TYS_20210417T221540).

the thin cloud removal ability of spectral and structure restoration on most bands. From the quantitative results in Table II, we can see that CR4S2 can better restore the background information than baseline methods not only at pixel-level but also at structure-level. This is due to the superiority of the structure and multioptimization loss of CR4S2. The spectral preservation results in Table VI demonstrate that the CR4S2 method can restore spectral information more effectively than baseline methods with the same input bands. From Tables II, IV, and V, we can also see that RSC-Net performs the worst among all DL-based methods. This may be because RSC-Net only includes normal convolutional and deconvolutional layers, thus, cannot make full use of the information in the input. Both FCTF-Net and RSDehazeNet inject channel attention operation to their network to learn the weights for different channels, which may be the reason why their performances are not too different on most bands. The visual results of three whole scenes from Sentinel-2 in Fig. 14 show that CR4S2 outperforms baseline DL-based methods on vegetation much more than on urban and barren. Color distortions are seen in the results of baseline DL-based methods on vegetation scenes. In particular, the result of RSC-Net on vegetation scene still contains cloud effect. Overall, the experimental results show that the proposed CR4S2 method is very promising for thin cloud removal in Sentinel-2 imagery.

V. Conclusion

In this work, a novel DL-based method CR4S2 was proposed for thin cloud removal in Sentinel-2 imagery. Three input/output branches were designed for taking original Sentinel-2 images as input/output. In order to extract and fuse multiscale features in different depths, we designed two parallel downsample blocks (PDRB-D and PDRB-T) that are based on a newly proposed DDSC module. A top-to-bottom residual path was constructed by injecting PDRBs into certain branches. Experimental results demonstrate the superiority of CR4S2-based models over baseline methods. CR4S2-based method can restore more spectral information than baseline methods. The influence of the Cirrus band on CR4S2 was also analyzed. The results show that CR4S2 can not only restore more texture information but also can use the cloud information in the Cirrus band to improve its thin cloud removal performance in other bands. The proposed multioptimization loss and multi-input/output branches have also been proved effective for improving thin cloud removal performance in most bands. The encoder of CR4S2 includes several multiscale fusion blocks; however, the decoder only takes normal deconvolutional layers as basic units, which may limit its performance.

In the future, we will consider designing and introducing multiscale feature fusion blocks into the decoder of CR4S2. The application on other multispectral images will also be taken into consideration. The transformer has been successfully applied on image processing; the superiority of the transformer on thin cloud removal will also be explored by injecting it into CR4S2.

Appendix

See Table VII.

TABLE VII
DETAILS OF SAMPLES FOR VISUAL COMPARISON IN THIS ARTICLE

Figures		Product ID	Acquisition Date	Status	Patch number
Fig. 1	(a)	S2A MSIL1C 20190707T213531 N0207 R086 T05VPJ 20190707T231819	2019/07/07	Clear	232
		S2A MSIL1C 20190627T213531 N0207 R086 T05VPJ 20190628T010801	2019/06/28	Cloud	
	(b)	S2A MSIL1C 20190905T180921 N0208 R084 T13UCQ 20190905T214606	2019/09/05	Clear	229
		S2A MSIL1C 20190915T180951 N0208 R084 T13UCQ 20190915T213737	2019/09/15	Cloud	
	(c)	S2A MSIL1C 20191006T041631 N0208 R090 T47TNN 20191006T073338	2019/10/06	Clear	322
		S2A MSIL1C 20190926T041551 N0208 R090 T47TNN 20190926T071222	2019/09/26	Cloud	
Fig. 5	(a)	S2A MSIL1C 20160925T104022 N0204 R008 T32ULB 20160925T104115	2016/09/25	Clear	597
		S2A MSIL1C 20160915T104022 N0204 R008 T32ULB 20160915T104018	2016/09/15	Cloud	
	(b)	S2A MSIL1C 20201004T141741 N0209 R010 T19GEM 20201004T174905	2020/10/04	Clear	181
		S2A MSIL1C 20201014T141741 N0209 R010 T19GEM 20201014T175003	2020/10/14	Cloud	
Fig. 9		S2A MSIL1C 20191018T100031 N0208 R122 T33TUG 20191018T121309	2019/10/18	Clear	301
		S2A MSIL1C 20191028T100121 N0208 R122 T33TUG 20191028T103529	2019/10/28	Cloud	

ACKNOWLEDGMENT

The authors are grateful for the Sentinel-2 data services from the Copernicus Open Access Hub.

REFERENCES

[1] A. Garioud, S. Valero, S. Giordano, and C. Mallet, "Recurrent-based regression of Sentinel time series for continuous vegetation monitoring," *Remote Sens. Environ.*, vol. 263, 2021, Art. no. 112419, doi: 10.1016/j.rse.2021.112419.

[2] L. Ma, M. Schmitt, and X. Zhu, "Uncertainty analysis of object-based land-cover classification using Sentinel-2 time-series data," *Remote Sens.*, vol. 12, no. 22, 2020, Art. no. 3798, doi: 10.3390/rs12223798.

[3] T. Li, H. Shen, Q. Yuan, and L. Zhang, "Geographically and temporally weighted neural networks for satellite-based mapping of ground-level PM2.5," *ISPRS J. Photogramm. Remote Sens.*, vol. 167, pp. 178–188, 2020, doi: 10.1016/j.isprsjprs.2020.06.019.

[4] M. Molinier, J. Miettinen, D. Ienco, S. Qiu, and Z. Zhu, "Optical satellite image time series analysis for environment applications: From classical methods to deep learning and beyond," in *Change Detection and Image Time Series Analysis 2: Supervised Methods*, A. M. Atto, F. Bovolo, and L. Bruzzone, Eds. London, U.K.: ISTE, 2021.

[5] Q. Yuan et al., "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, May 2020, Art. no. 111716, doi: 10.1016/j.rse.2020.111716.

[6] C. Qiu, L. Mou, M. Schmitt, and X. X. Zhu, "Local climate zone-based urban land cover classification from multi-seasonal Sentinel-2 images with a recurrent residual network," *ISPRS J. Photogramm. Remote Sens.*, vol. 154, pp. 151–162, 2019, doi: 10.1016/j.isprsjprs.2019.05.004.

[7] X. X. Zhu et al., "The urban morphology on our planet – Global perspectives from space," *Remote Sens. Environ.*, vol. 269, Feb. 2022, Art. no. 112794, doi: 10.1016/J.RSE.2021.112794.

[8] M. D. King, S. Platnick, W. P. Menzel, S. A. Ackerman, and P. A. Hubanks, "Spatial and temporal distribution of clouds observed by MODIS onboard the terra and aqua satellites," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3826–3852, Jul. 2013, doi: 10.1109/TGRS.2012.2227333.

[9] M. Bayad et al., "Time series of remote sensing and water deficit to predict the occurrence of soil water repellency in New Zealand pastures," *ISPRS J. Photogramm. Remote Sens.*, vol. 169, pp. 292–300, 2020, doi: 10.1016/j.isprsjprs.2020.09.024.

[10] T. Y. Ji, N. Yokoya, X. X. Zhu, and T. Z. Huang, "Nonlocal tensor completion for multitemporal remotely sensed images' inpainting," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3047–3061, Jun. 2018, doi: 10.1109/TGRS.2018.2790262.

[11] Q. Zhang, Q. Yuan, Z. Li, F. Sun, and L. Zhang, "Combined deep prior with low-rank tensor SVD for thick cloud removal in multitemporal images," *ISPRS J. Photogramm. Remote Sens.*, vol. 177, pp. 161–173, 2021, doi: 10.1016/j.isprsjprs.2021.04.021.

[12] 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 J. Photogramm. Remote Sens.*, vol. 148, pp. 103–113, Feb. 2019, doi: 10.1016/j.isprsjprs.2018.12.013.

[13] J. D. Bermudez, P. N. Happ, R. Q. Feitosa, and D. A. B. Oliveira, "Synthesis of multispectral optical images from SAR/optical multitemporal data using conditional generative adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1220–1224, Aug. 2019, doi: 10.1109/LGRS.2019.2894734.

[14] P. Ebel, A. Meraner, M. Schmitt, and X. X. Zhu, "Multisensor data fusion for cloud removal in global and all-season Sentinel-2 imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5866–5878, Jul. 2021, doi: 10.1109/TGRS.2020.3024744.

[15] C. Grohnfeldt, M. Schmitt, and X. Zhu, "A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from Sentinel-2 images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 1726–1729, doi: 10.1109/IGARSS.2018.8519215.

[16] A. Meraner, P. Ebel, X. X. Zhu, and M. Schmitt, "Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 333–346, 2020, doi: 10.1016/j.isprsjprs.2020.05.013.

[17] J. Li, Q. Hu, and M. Ai, "Haze and thin cloud removal via sphere model improved dark channel prior," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 472–476, Mar. 2019, doi: 10.1109/LGRS.2018.2874084.

[18] Y. Zhang, F. Wen, Z. Gao, and X. Ling, "A coarse-to-fine framework for cloud removal in remote sensing image sequence," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5963–5974, Aug. 2019, doi: 10.1109/TGRS.2019.2903594.

[19] H. Lv, Y. Wang, and Y. Shen, "An empirical and radiative transfer model based algorithm to remove thin clouds in visible bands," *Remote Sens. Environ.*, vol. 179, pp. 183–195, 2016, doi: 10.1016/j.rse.2016.03.034.

[20] M. Xu, M. Pickering, A. J. Plaza, and X. Jia, "Thin cloud removal based on signal transmission principles and spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1659–1669, Mar. 2016, doi: 10.1109/TGRS.2015.2486780.

[21] O. R. Mitchell, E. J. Delp, and P. L. Chen, "Filtering to remove cloud cover in satellite imagery," *IEEE Trans. Geosci. Electron.*, vol. 15, no. 3, pp. 137–141, Jul. 1977, doi: 10.1109/tge.1977.6498971.

[22] Y. Zhang, B. Guindon, and J. Cihlar, "An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images," *Remote Sens. Environ.*, vol. 82, no. 2/3, pp. 173–187, 2002, doi: 10.1016/S0034-4257(02)00034-2.

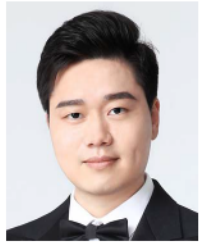
[23] S. Chen, X. Chen, J. Chen, P. Jia, X. Cao, and C. Liu, "An iterative haze optimized transformation for automatic cloud/haze detection of Landsat imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2682–2694, May 2016, doi: 10.1109/TGRS.2015.2504369.

[24] X. Y. He, J. B. Hu, W. Chen, and X. Y. Li, "Haze removal based on advanced haze-optimized transformation (AHOT) for multispectral imagery," *Int. J. Remote Sens.*, vol. 31, no. 20, pp. 5331–5348, 2010, doi: 10.1080/01431160903369600.

[25] Z. Zhu, S. Wang, and C. E. Woodcock, "Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images," *Remote Sens. Environ.*, vol. 159, pp. 269–277, Mar. 2015, doi: 10.1016/j.rse.2014.12.014.

[26] Z. Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in Landsat imagery," *Remote Sens. Environ.*, vol. 118, pp. 83–94, 2012, doi: 10.1016/j.rse.2011.10.028.

[27] W. Li, Y. Li, D. Chen, and J. C. W. Chan, "Thin cloud removal with residual symmetrical concatenation network," *ISPRS J. Photogramm. Remote Sens.*, vol. 153, pp. 137–150, 2019, doi: 10.1016/j.isprsjprs.2019.05.003.
[28] G. Hu, X. Li, and D. Liang, "Thin cloud removal from remote sensing images using multidirectional dual tree complex wavelet transform and transfer least square support vector regression," *J. Appl. Remote Sens.*, vol. 9, no. 1, Sep. 2015, Art. no. 095053, doi: 10.1117/1.jrs.9.095053.
[29] J. Long, Z. Shi, W. Tang, and C. Zhang, "Single remote sensing image dehazing," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 59–63, Jan. 2014, doi: 10.1109/LGRS.2013.2245857.
[30] Z. K. Liu and B. R. Hunt, "A new approach to removing cloud cover from satellite imagery," *Comput. Vis., Graph., Image Process.*, vol. 25, no. 2, pp. 252–256, 1984, doi: 10.1016/0734-189X(84)90107-5.
[31] N. L. Han, C. Liu, L. Zhuang, and W. Zhang, "Removing thin cloud by combining wavelet transforms and homomorphic filter in the CBERS-02B image," *Jilin Univ. (Earth Sci. Ed.)*, vol. 42, no. 1, pp. 275–279, 2012, doi: 10.3969/j.issn.1671-5888.2012.01.035.
[32] H. Shen, H. Li, Y. Qian, L. Zhang, and Q. Yuan, "An effective thin cloud removal procedure for visible remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 96, pp. 224–235, 2014, doi: 10.1016/j.isprsjprs.2014.06.011.
[33] M. Wan and X. Li, "Removing thin cloud on single remote sensing image based on SWF," in *Proc. IEEE Int. Conf. Online Anal. Comput. Sci.*, 2016, pp. 397–400, doi: 10.1109/ICOACS.2016.7563124.
[34] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011, doi: 10.1109/TPAMI.2010.168.
[35] S. Malek, F. Melgani, Y. Bazi, and N. Alajlan, "Reconstructing cloud-contaminated multispectral images with contextualized autoencoder neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2270–2282, Apr. 2018, doi: 10.1109/TGRS.2017.2777886.
[36] M. Qin, F. Xie, W. Li, Z. Shi, and H. Zhang, "Dehazing for multispectral remote sensing images based on a convolutional neural network with the residual architecture," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1645–1655, May 2018, doi: 10.1109/JSTARS.2018.2812726.
[37] L. Sun, Y. Zhang, X. Chang, Y. Wang, and J. Xu, "Cloud-Aware generative network: Removing cloud from optical remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 691–695, Apr. 2020, doi: 10.1109/LGRS.2019.2928840.
[38] Q. Yang, G. Wang, Y. Zhao, X. Zhang, G. Dong, and P. Ren, "Multi-scale deep residual learning for cloud removal," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 4967–4970, doi: 10.1109/IGARSS39084.2020.9323261.
[39] P. Dai, S. Ji, and Y. Zhang, "Gated convolutional networks for cloud removal from bi-temporal remote sensing images," *Remote Sens.*, vol. 12, no. 20, 2020, Art. no. 3427, doi: 10.3390/rs12203427.
[40] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervention*, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.
[41] J. Li, Z. Wu, Z. Hu, Z. Li, Y. Wang, and M. Molinier, "Deep learning based thin cloud removal fusing vegetation red edge and short wave infrared spectral information for Sentinel-2A imagery," *Remote Sens.*, vol. 13, no. 1, Jan. 2021, Art. no. 157, doi: 10.3390/rs13010157.
[42] W. Yu, X. Zhang, M. O. Pun, and M. Liu, "A hybrid model-based and data-driven approach for cloud removal in satellite imagery using multi-scale distortion-aware networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 7160–7163, doi: 10.1109/IGARSS47720.2021.9554963.
[43] X. Wen, Z. Pan, Y. Hu, and J. Liu, "An effective network integrating residual learning and channel attention mechanism for thin cloud removal," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6507605, doi: 10.1109/LGRS.2022.3161062.
[44] I. J. Goodfellow et al., "Generative adversarial nets," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680, doi: 10.5555/2969033.2969125.
[45] K. Enomoto et al., "Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 1533–1541, doi: 10.1109/CVPRW.2017.197.
[46] X. Wang, G. Xu, Y. Wang, D. Lin, P. Li, and X. Lin, "Thin and thick cloud removal on remote sensing image by conditional generative adversarial network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 1426–1429, doi: 10.1109/IGARSS.2019.8897958.
[47] J. Zheng, X. Y. Liu, and X. Wang, "Single image cloud removal using U-Net and generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6371–6385, Aug. 2021, doi: 10.1109/TGRS.2020.3027819.
[48] Z. Xu, K. Wu, L. Huang, Q. Wang, and P. Ren, "Cloudy image arithmetic: A cloudy scene synthesis paradigm with an application to deep-learning-based thin cloud removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5612616, doi: 10.1109/TGRS.2021.3122253.
[49] P. Singh and N. Komodakis, "Cloud-GAN: Cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 1772–1775, doi: 10.1109/IGARSS.2018.8519033.
[50] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[51] M. Xu, F. Deng, S. Jia, X. Jia, and A. J. Plaza, "Attention mechanism-based generative adversarial networks for cloud removal in Landsat images," *Remote Sens. Environ.*, vol. 271, 2022, Art. no. 112902, doi: 10.1016/j.rse.2022.112902.
[52] J. Li et al., "Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 373–389, Aug. 2020, doi: 10.1016/j.isprsjprs.2020.06.021.
[53] Y. Zi, F. Xie, X. Song, Z. Jiang, and H. Zhang, "Thin cloud removal for remote sensing images using a physical model-based CycleGAN with unpaired data," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 1004605, doi: 10.1109/LGRS.2021.3140033.
[54] M. Xu, X. Jia, M. Pickering, and S. Jia, "Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform," *ISPRS J. Photogramm. Remote Sens.*, vol. 149, pp. 215–225, 2019, doi: 10.1016/j.isprsjprs.2019.01.025.
[55] J. Guo, J. Yang, H. Yue, H. Tan, C. Hou, and K. Li, "RSDehazeNet: Dehazing network with channel refinement for multispectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2535–2549, Mar. 2021, doi: 10.1109/TGRS.2020.3004556.
[56] Y. Li and X. Chen, "A coarse-to-fine two-stage attentive network for haze removal of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 10, pp. 1751–1755, Oct. 2021, doi: 10.1109/lgrs.2020.3006533.
[57] P. Ebel, Y. Xu, M. Schmitt, and X. X. Zhu, "SEN12MS-CR-TS: A remote sensing data set for multi-modal multi-temporal cloud removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5222414, doi: 10.1109/tgrs.2022.3146246.
[58] J. Li et al., "A lightweight deep learning-based cloud detection method for Sentinel-2A imagery fusing multiscale spectral and spatial features," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5401219, doi: 10.1109/TGRS.2021.3069641.
[59] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*, doi: 10.48550/arXiv.1704.04861.
[60] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1800–1807, doi: 10.1109/CVPR.2017.195.
[61] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5967–5976, doi: 10.1109/CVPR.2017.632.
[62] T. M. Bui and W. Kim, "Single image dehazing using color ellipsoid prior," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 999–1009, Feb. 2018, doi: 10.1109/TIP.2017.2771158.
[63] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–15, doi: 10.48550/arXiv.1412.6980.



Jun Li received the B.S. degree in remote sensing science and technology, the M.S. degree in geomatics engineering, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2015, 2018, and 2021, respectively.

He is currently an Associate Research Fellow with the College of Astronautics, Nanjing University of Aeronautics & Astronautics, Nanjing, China. His research interests include remote sensing image processing and deep learning.

Yuejie Zhang received the B.S. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2020. He is currently working toward the Ph.D. degree in optical engineering from the Nanjing University of Aeronautics & Astronautics, Nanjing, China.

His research interests include thermal remote sensing and deep learning.

Qinghong Sheng received the B.S. degree in photogrammetry and remote sensing, the M.S. degree in cartography and geography information system, and the Ph.D. degree in photogrammetry and remote sensing techniques from Wuhan University, Wuhan, China, in 2000, 2004, and 2008, respectively.

She is currently a Professor with the College of Astronautics, Nanjing University of Aeronautics & Astronautics, Nanjing, China. Her current research interests include spatial information extraction and SAR target detection.

Zhaocong Wu received the B.S. and Ph.D. degrees in photogrammetry and remote sensing techniques from Wuhan University, Wuhan, China, in 1986 and 2004, respectively.

He is currently a Professor with the School of Remote Sensing and Information Engineering, Wuhan University. His current research interests include high-resolution image processing, image analysis, and pattern recognition.

Bo Wang received the B.S. degree in remote sensing science and technology, the M.S. degree in geomatics engineering, the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2010, 2012, and 2015, respectively.

He is currently an Assistant Professor with the College of Astronautics, Nanjing University of Aeronautics & Astronautics, Nanjing, China. His current research interests include spatial information extraction and remote sensing image processing.

Zhongwen Hu (Member, IEEE) received the B.Sc. degree in remote sensing and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2008 and 2013, respectively.

He is currently an Assistant Professor with the MNR Key Laboratory for Geo-Environmental Monitoring of Great Bay Area & Guangdong Key Laboratory of Urban Informatics & Shenzhen Key Laboratory of Spatial Smart Sensing and Services & Research Institute for Smart Cities, Shenzhen University, Shenzhen, China. His research interests include object-based image analysis and coastal remote sensing.

Guanting Shen received the B.S. degree in remote sensing science and technology in 2021 from Wuhan University, Wuhan, China, where he is currently working toward the M.S. degree in photogrammetry and remote sensing.

His research interests include hyperspectral remote sensing and hyperspectral data processing.

Michael Schmitt (Senior Member, IEEE) received his Dipl.-Ing. (Univ.) degree in geodesy and geoinformation, the Dr.-Ing. degree in remote sensing, and the habilitation in data fusion from the Technical University of Munich (TUM), Munich, Germany, in 2009, 2014, and 2018, respectively.

Since 2021, he has been a Full Professor for Earth Observation with the Department of Aerospace Engineering, University of the Bundeswehr Munich, Neubiberg, Germany. Before that, he was a Professor in applied geodesy and remote sensing with the Department of Geoinformatics, Munich University of Applied Sciences. In 2019, he was additionally appointed as an Adjunct Teaching Professor with the Department of Aerospace and Geodesy, TUM. In 2016, he was a Guest Scientist with the University of Massachusetts, Amherst. From 2015 to 2020, he was a Senior Researcher and Deputy Head with the Professorship for Signal Processing in Earth Observation, TUM. His research interests include technical aspects of Earth observation, in particular image analysis and machine learning applied to the extraction of information from multimodal remote sensing observations. Among his core interests is remote sensing data fusion with a focus on SAR and optical data.

Prof. Schmitt is a Co-Chair of the Working Group "Active Microwave Remote Sensing" of the International Society for Photogrammetry and Remote Sensing and also of the Working Group "Benchmarking" of the IEEE-GRSS Image Analysis and Data Fusion Technical Committee. He frequently serves as a Reviewer for a number of renowned international journals and conferences. He is a recipient of several Best Reviewer awards. He is an Associate Editor for IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Matthieu Molinier (Member, IEEE) received the Engineering degree from École Nationale Supérieure de Physique de Strasbourg (ENSPS), Illkirch-Graffenstaden, France, in 2004, and the M.Sc. degree in image processing from Université Louis Pasteur (ULP), Strasbourg, France, in 2004.

Since then, he has been with VTT, Espoo, Finland, as a Research Scientist in Earth observation, with a main focus on machine learning and change detection for optical satellite images. He is the author of 70 articles in scientific journals and conferences. His research interests include deep learning for multispectral and hyperspectral images, unsupervised change detection, and satellite image time-series analysis, with applications to environment monitoring.