

Article

Deep Learning Based Thin Cloud Removal Fusing Vegetation Red Edge and Short Wave Infrared Spectral Information for Sentinel-2A Imagery

Jun Li ¹, Zhaocong Wu ^{1,*}, Zhongwen Hu ², Zilong Li ¹, Yisong Wang ¹ and Matthieu Molinier ³

¹ School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, Hubei, China; jun_li@whu.edu.cn (J.L.); lizilong@whu.edu.cn (Z.L.); wangys@whu.edu.cn (Y.W.)

² MNR Key Laboratory for Geo-Environmental Monitoring of Great Bay Area, Shenzhen University, Shenzhen 518060, China; zwhoo@szu.edu.cn

³ VTT Technical Research Centre of Finland Ltd., 02044 Espoo, Finland; matthieu.molinier@vtt.fi

* Correspondence: zcwoo@whu.edu.cn

Abstract: Thin clouds seriously affect the availability of optical remote sensing images, especially in visible bands. Short-wave infrared (SWIR) bands are less influenced by thin clouds, but usually have lower spatial resolution than visible (Vis) bands in high spatial resolution remote sensing images (e.g., in Sentinel-2A/B, CBERS04, ZY-1 02D and HJ-1B satellites). Most cloud removal methods do not take advantage of the spectral information available in SWIR bands, which are less affected by clouds, to restore the background information tainted by thin clouds in Vis bands. In this paper, we propose CR-MSS, a novel deep learning-based thin cloud removal method that takes the SWIR and vegetation red edge (VRE) bands as inputs in addition to visible/near infrared (Vis/NIR) bands, in order to improve cloud removal in Sentinel-2 visible bands. Contrary to some traditional and deep learning-based cloud removal methods, which use manually designed rescaling algorithm to handle bands at different resolutions, CR-MSS uses convolutional layers to automatically process bands at different resolution. CR-MSS has two input/output branches that are designed to process Vis/NIR and VRE/SWIR, respectively. Firstly, Vis/NIR cloudy bands are down-sampled by a convolutional layer to low spatial resolution features, which are then concatenated with the corresponding features extracted from VRE/SWIR bands. Secondly, the concatenated features are put into a fusion tunnel to down-sample and fuse the spectral information from Vis/NIR and VRE/SWIR bands. Third, a decomposition tunnel is designed to up-sample and decompose the fused features. Finally, a transpose convolutional layer is used to up-sample the feature maps to the resolution of input Vis/NIR bands. CR-MSS was trained on 28 real Sentinel-2A image pairs over the globe, and tested separately on eight real cloud image pairs and eight simulated cloud image pairs. The average SSIM values (Structural Similarity Index Measurement) for CR-MSS results on Vis/NIR bands over all testing images were 0.69, 0.71, 0.77, and 0.81, respectively, which was on average 1.74% higher than the best baseline method. The visual results on real Sentinel-2 images demonstrate that CR-MSS can produce more realistic cloud and cloud shadow removal results than baseline methods.

Keywords: thin cloud removal; Sentinel-2A imagery; multi-spectral feature fusion; multi-spatial; deep learning

Citation: Li, J.; Wu, Z.; Hu, Z.; Li, Z.; Wang, Y.; Molinier, M. Deep Learning Based Thin Cloud Removal Fusing Vegetation Red Edge and Short Wave Infrared Spectral Information for Sentinel-2A Imagery. *Remote Sens.* **2021**, *13*, 157. <https://doi.org/10.3390/rs13010157>

Received: 2 December 2020

Accepted: 4 January 2021

Published: 5 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Motivation

With the development of optical satellite sensor technology, multispectral and hyperspectral remote sensing images with high spatial resolution (HR) have been much easier to acquire than ever before. Because the spectral characteristics of various landscapes are different, multi- and hyper-spectral images are widely used in land use and land cover

classification [1,2], vegetation monitoring [3], and water resources monitoring [4,5]. However, the global annual mean cloud cover is approximately 66% [6], and optical remote sensing images are easily contaminated by clouds, which greatly reduces the number of pixels effectively available for land cover studies [7,8].

Clouds can be roughly divided into two categories: thick clouds and thin clouds. Thick clouds block most of the electromagnetic signal reflected from the land surface, which makes it impossible to restore the background information using only thick cloud pixels [9]. However, thin clouds can let some electromagnetic signal transmit it, which make it possible to restore the signal with only thin cloud pixels. Therefore, the influence of thin clouds on optical remote sensing images is not only related to their thickness [10], but also to the electromagnetic signal's wavelength. For example, the visible/near infrared (Vis/NIR) bands are much more influenced by thin clouds than short wave infrared (SWIR) bands, which means that SWIR bands preserve more spectral information than visible bands. However, in high spatial resolution remote sensing images, SWIR bands usually have lower spatial resolution than Vis/NIR bands.

Figure 1 shows four examples under the cloud contamination condition (odd rows) and the corresponding cloud-free images (even rows). Column 1 shows the true color composited images (T); columns 2–11 are bands 2/3/4/5/6/7/8/8A/11/12. All bands are zoomed to the same size. We can see that the influences of thin clouds on these bands are different. The transmittances of thin cloud in vegetation red edge (VRE, bands 5/6/7) and short-wave infrared (SWIR, bands 8A/11/12) bands are higher than in Vis bands (2/3/4), causing thin clouds to have a greater effect on Vis bands than on bands VRE/SWIR. Thanks to this, some missing information in Vis bands can be restored from VRE/SWIR bands.

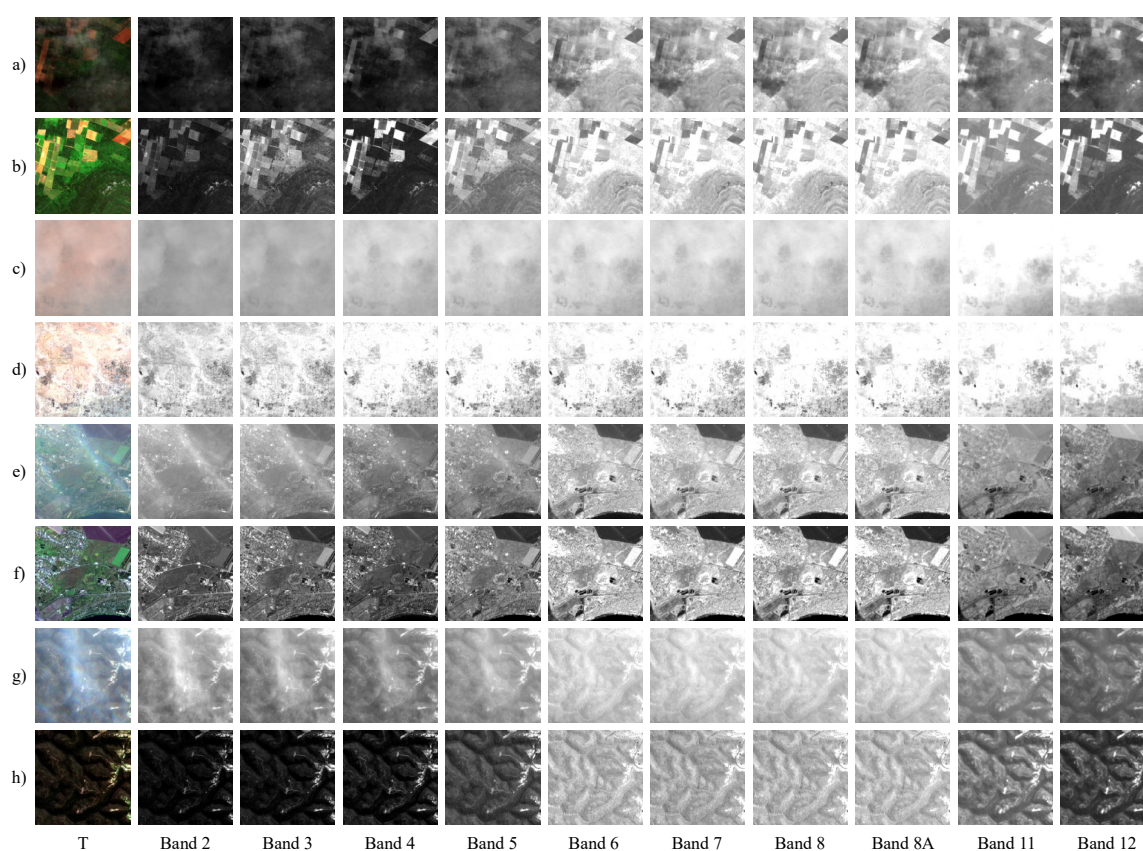


Figure 1. Four different scenes. Rows (a,c,e,g) are cloud contaminated images; rows (b,d,f,h) are corresponding cloud-free images. All bands are resized to the same size for better presentation.

In this paper, we propose a deep learning-based network to handle thin cloud removal in Sentinel-2A images. The network is mainly designed to fuse spectral information of SWIR bands and Vis/NIR bands to better remove thin clouds in Vis/NIR bands. We also take the vegetation red edge (VRE) bands into consideration, because these bands are also less influenced by thin clouds than visible bands. The experiment is conducted both on real and simulated paired cloud and cloud-free dataset covering the globe.

1.2. Related Works

The removal of thick clouds usually requires multitemporal images from the same location [11–13]. In Tseng et al. [14], cloud pixels were directly replaced with clear pixels. Zhang and Wen [15] proposed a discriminative robust principal component analysis (DRPCA) algorithm to recover the background information under thick clouds. Sparse reconstruction was used for cloud removal with time series remote sensing images in Cerra et al. [16]. Image patches clone was adopted to restore background under cloud regions in Lin and Tsai [17]. Chen et al. [18] proposed a spatially and temporally weighted regression model to reconstruct ground pixels occulted by clouds. A nonnegative matrix factorization and error correction algorithm (S-NMF-EC) was proposed in Li and Wang [19] to fuse auxiliary HR and low resolution (LR) cloud-free images to obtain cloud-free HR images. Although thick clouds can be removed by these methods, multitemporal data is required to reconstruct the signal blocked by thick clouds.

The transmittance of clouds increases as thickness of thin clouds decreases [10]. Once the clouds are thin enough, the background signal can transmit through thin clouds and be received by optical satellite sensors, which makes it possible to reconstruct the background signal. Thus, thin clouds removal usually relies on a single cloudy image. Since the response to clouds varies with the wavelength, the features of multispectral bands are very beneficial for the correction of cloud contaminated images. Based on statistics collected on many remote sensing images, the haze optimized transformation (HOT) assumed that ground reflectance in blue and red bands is linearly correlated under clear conditions [20]. An iterative HOT (IHOT) method was proposed in Chen et al. [21] to detect and remove thin clouds in Landsat imagery. Cloudy and corresponding clear images were used in IHOT to solve the spectral confusion between clouds and bright surfaces. Xu et al. [22] proposed a signal transmission and spectral mixture analysis (ST-SMA) algorithm in which the cloud removal model considered the transmission and absorption of clouds. The spectral-based methods can make use of spectral information to remove thin clouds, but they do not take the spatial correlation of neighborhood pixels into consideration. Some other methods for thin cloud removal rely on filtering the cloud component in frequency domain.

There are many methods based on physical model and filtering for cloud/haze removal. In the homomorphic filtering (HF) method [23], Fourier transform was used to separate thin cloud and background components, then a linear filter was adopted to remove the cloud component. Shen et al. [24] proposed an adaptive HF (AHF) method improving HF by treating each spectral band differently and using cloud masks to keep cloud-free areas unchanged. A max–min radiation correction was applied to the result of fast Fourier transform (FFT) and low-pass filter on each band in Liu and Wang [25] to eliminate the influence of transmission and enhance contrast. The methods based on homomorphic filtering can remove cloud components, but low frequency components in the background signal are also removed. Based on the analysis of human visual system, Retinex [26] was proposed to solve the illumination imbalance in images. It was improved in Jonson and Rahman [27], who adopted Gaussian filtering to estimate the incident light. Multi-scale Gaussian filtering in Retinex (MSR) has been proved to be efficient in handling color rendition and dynamic range compression [28]. In order to adjust the color distortion caused by the enhanced contrast in local areas of the image, a color recovery factor was added to MSR (MSRCR) in Jonson and Rahman [29]. MSRCR has been adopted as official image processing tool by NASA and widely used for image dehazing. Although Retinex-

based methods can restore some information in images, the incident light estimated by Gaussian filtering is far from the real incident light.

Deep learning (DL) methods have developed rapidly in recent years due to the great improvement of computing performance and increased availability of labelled data. Convolutional neural network (CNN) is the most effective and widely used in image processing and computer vision fields. CNN has been applied to image classification [30], semantic segmentation [31], image generation [32], and image restoration [33]. Most of these tasks were further improved by adopting Resnet [34] and U-Net [35] architectures. The residual architecture was originally designed to solve the problem of information loss in the training of deep CNN by adding input to output and has been widely used for image segmentation recently. The U-Net architecture introduced skip connection that connects down-sampling and symmetrical up-sampling layers to make full use of the features at low and high levels. Differently from Resnet, skip connections in U-Net concatenate the feature maps of convolution and corresponding deconvolution at feature channel, which is more effective than Resnet in using multi-layer features. Many DL-based methods have been proposed to improve the capacity of automatic data processing and solve the technical problems in the remote sensing field. CNN is widely used for remote sensing image processing [36,37], such as land use and land cover classification [38,39], hyperspectral image classification [40–42], remotely sensed scene classification [43,44], object detection [45,46], and image synthesis [47].

Due to the good performance of CNN in image inpainting, many CNN-based methods have been successfully applied to thick cloud removal in remote sensing images. Li et al. [48] designed a convolutional–mapping–deconvolutional (CMD) network in which optical and SAR images in the same region were transferred into target cloud-free images. Then the cloud pixels in cloudy images were replaced by corresponding cloud-free pixels from the target cloud-free images. In Meraner et al. [49], a DSen2-CR was proposed to fuse SAR image and optical images to remove thick cloud in Sentinel-2 images. DSen2-CR concatenated SAR image and Sentinel-2 image at spectral channel as input and learned the residual between a cloud image and the corresponding cloud-free image. Generative adversarial networks (GANs) were also adopted for thick cloud removal by fusing optical and SAR images in Gao et al. [50]. GANs were also used for thin/thick cloud removal with paired cloudy and cloud-free optical remote sensing images from the same region [51]. Such thick cloud removal methods can achieve good results but require corresponding auxiliary data such as synthetic aperture radar (SAR) images, which are not influenced by cloud coverage. In addition, cloud detection [52–55] and image registration [56] are the necessary steps and important influencing factors for these methods.

Differently from image inpainting, image dehazing aims to remove haze and restore the background information in a single image. The atmospheric degradation model has been widely used in image dehazing. Multi-scale parallel convolution has been proven very useful for producing a transmission map, which is then put into the atmospheric degradation model to get a dehazed image [57]. In Zhang and Patel [58], a densely-connected encoder–decoder and U-Net were used to estimate the transmission map and atmospheric light, respectively. The estimated atmospheric light improved the quality of dehazed image a lot. Although these methods can remove haze in images, the atmospheric degradation model is an empirical model, which cannot accurately simulate the interaction between light and the atmosphere. Therefore, many methods have been recently proposed to directly restore a clear image from a hazy image. In Ren et al. [59], the haze image was driven into three transformed images to handle the influences of atmosphere light, scattering, and attenuation. The model used multi-scale reconstruction losses to learn more details. In Chen et al. [60], the smooth dilated convolution was used to solve the grid artifacts. Differently from previous work, the method designed a gate fusion sub-network to learn the weights of different level features, which were then used for the weight sum of these features. Learning the mapping from hazy image to clear image is very effective, but such methods need paired hazy and clear images, which are challenging to obtain.

Thin clouds in remote sensing images are similar to haze in natural images. Thin clouds let part of the background signal transmit, which makes it possible to restore background information using only cloud-contaminated pixels. Qin et al. [61] designed a multi-scale CNN to remove thin clouds in multispectral images. Instead of merging all bands into one network, individual networks were designed for each band. Then the outputs of each individual network were fused by a multi-scale feature fusion layer. A residual symmetrical concatenation network (RSC-Net) was proposed in Li et al. [62], which took advantage of residual error between cloudy and corresponding cloud-free images to train the cloud removal model. Down-sampling and up-sampling operations that are thought to damage the information in the cloud-free regions were not used in RSC-Net. Cycle-GAN [63] has proven very effective for unpaired image-to-image translation. Inspired by Cycle-GAN, a cloud removal GAN (Cloud-GAN) that transfers a thin cloud image into a cloud-free image was proposed in Singh and Komodakis [64]. Sun et al. [65] proposed a cloud-aware generative network (CAGN) in which convolutional long short-term memory (LSTM) and auto-encoder were combined to detect and remove clouds. The attention mechanism [66] was introduced into CAGN to process cloudy regions differently according to the thickness of clouds. In Li et al. [67], a modified physical model was combined with GANs to remove thin cloud with unpaired images. While producing the clear image, the method also learned the transmission, absorption, and reflection maps from the cloud image. Although these methods have good performances, they only handle bands at a same spatial resolution. This discourages their application to satellite sensors that have multi-spatial resolution bands, especially sensors that have low spatial resolution short wave infrared bands.

Although many deep learning-based methods have been proposed for cloud removal in high spatial resolution remote sensing images and achieved the state-of-the-art performance, most of them only handle the Vis/NIR bands at high spatial resolution and either ignore short wave infrared bands at low spatial resolution that are much less influenced by thin cloud than Vis/NIR bands [64], or rescale low resolution bands to high resolution and then process them with Vis/NIR bands together [49]. Rescaling low resolution bands into high resolution by a manually designed rescaling algorithm can help solve the problem of different spatial resolutions. However, the parameters of the manually designed rescaling algorithm are not optimal, and spectral/spatial information may be lost during the manually designed rescaling process. Convolutional neural networks, which can learn optimal sampling parameters during model training, have been widely used for up/down-sampling features automatically [32,35,68]. Therefore, in order to better preserve/extract the spectral information in low resolution bands, we adopted convolutional layers to handle multi-spectral bands at different spatial resolutions.

There are several high spatial resolution satellites such as Sentinel-2A/B, CBERS-04, ZY-1 02D, and HJ-1B that include low spatial resolution SWIR bands, which makes it desirable to take SWIR bands into consideration when removing thin cloud in Vis/NIR bands. In this paper, we aim to remove thin cloud in Sentinel-2A images, which include four vegetation red edge bands that are also less influenced by thin cloud than visible bands, and propose an end-to-end method based on deep-learning for thin cloud and cloud shadow removal by taking vegetation red edge (VRE) and SWIR bands in Sentinel-2A images into consideration. The main contributions are as follows:

1. We propose an end-to-end network architecture for cloud and cloud shadow removal that is tailored for Sentinel-2A images with the fusion of visible, NIR, VRE, and SWIR bands. The spectral features in VRE/SWIR bands are fully used to recover the cloud contaminated background information in Vis/NIR bands. Convolutional layers were adopted to replace manually designed rescaling algorithm to better preserve and extract spectral information in low resolution VRE/SWIR bands.
2. The experimental data are from different regions of the world. The types of land cover are rich and the acquisition dates of the experimental data cover a long time

period (from 2015 to 2019) and all seasons. Experiments on both real and simulated testing datasets are conducted to analyze the performance of the proposed CR-MSS in different aspects.

3. Three DL-based methods and two traditional methods are compared with CR-MSS. The performance of CR-MSS with/without VRE and SWIR bands as input and output is analyzed. The results show that CR-MSS is very efficient and robust for thin cloud and cloud shadow removal, and it performs the better when taking VRE and SWIR bands into consideration.

The rest of this paper is organized as follows. Section 2 presents the experimental data and study area. Section 3 introduces the proposed CR-MSS method. The experimental results are shown in Section 4, and the discussion is also undertaken in this Section. We conclude in Section 5.

2. Materials and Methods

2.1. Sentinel-2A Multispectral Data

To validate the performance of CR-MSS, Sentinel-2A imagery is selected as the experimental data. Sentinel-2A is a high-resolution multi-spectral imaging satellite that carries a multi-spectral imager (MSI) covering 13 spectral bands in the visible, near infrared, and shortwave infrared (Table 1). The wavelengths of Sentinel-2A image range from 0.443 μm to 2.190 μm and contain four bands in the vegetation red-edge, which is effective for monitoring vegetation health information. Band 1 is used to detect the Coastal aerosol, and bands 9/10 are used to monitor the water vapour/Cirrus, respectively. These bands are used for detecting and correcting the atmospheric effects rather than observing the land surface [49], which means that the atmosphere information in these bands are the most used for practical applications. In Meraner et al. [49], the cloud removal results were worst on 60 m bands even with SAR image as auxiliary input, and atmosphere conditions differ a lot in one day in the same area, let alone in 10 days. Therefore, the introduction of these 60 m bands is not necessary and would not be beneficial to cloud removal in other bands. Therefore, bands 1/9/10 were discarded in our experiments; only bands 2/3/4/8 (Vis/NIR) with size = 10980 $\times$ 10980 pixels and 5/6/7/8A/11/12 (VRE/SWIR) with size = 5490 $\times$ 5490 pixels in Sentinel-2A Level-1C product were selected in our experiments.

Table 1. Sentinel-2A sensor bands. Bands used in our study are marked in bold.

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
---------	------	-------	-----	----

2.2. Selection of Training and Testing Data

In the cited related non-deep learning-based works, the methods are usually empirically based and directly tested on experimental data, e.g., seven Landsat TM images covering 12 different land cover types in Canada were used to test the performance of HOT [20] on cloud removal. In IHOT [21], the method was tested on four pairs of cloudy and cloud-free Landsat, eight images from cropland, urban, snow, and desert land covers. Shen et al. [24] employed five Landsat ETM+ and two GaoFen-1 cloudy images as testing data in AHF. However, in the cited related deep learning-based works, training data is necessary, e.g., eight pairs and two pairs of Landsat, eight images, were selected for training and testing, respectively, in RSC-Net [62]. Singh and Komodakis [64] chose 20 cloudy and 13 cloud-free images to train Cloud-GAN and tested the performance on five synthetic scenes. In Li et al. [67], 16 training sites and four testing sites were chosen in the East coast of United States. The ratio between training and testing images most of these deep learning-based cloud removal methods was 80/20%.

In these related works, the spatial coverage of the study areas only ranges from cities to countries, the datasets are relatively small, and the time period between cloud and corresponding cloud-free images are usually longer than half a month, which means land cover may change greatly. The datasets in non-deep learning-based methods are mainly used for testing, and thus cannot be employed to train deep learning-based methods. Although the experimental data in cited deep learning-based methods includes training and testing data, those datasets were relatively small and were not stratified by land cover types; therefore, those study areas lack diversity.

In this study, cloudy and corresponding cloud-free images from Sentinel-2A satellite are used to train and test all methods. In order to better evaluate the effectiveness of CR-MSS over large scale regions and different land cover types, the training and testing areas are evenly distributed worldwide and according to three main land covers: urban, vegetation, and bare land. Figure 2 shows the 36 locations from all over the world selected as study areas. For training, we chose 28 areas all over the globe including 10 urban areas, 10 vegetation areas, and eight bare land areas. Eight areas were chosen all over the globe for testing, including three urban areas, three vegetation areas, and two bare land areas. Therefore, there is one testing area for three to four training areas, depending on the land cover. From Figure 2, it also can be seen that each testing site is surrounded by three to four training sites. The total ratio between training and testing images in our experiment is 77.8%/22.2%, i.e., slightly more testing data in proportion than the related deep-learning works. The details of 36 pairs cloud and corresponding cloud-free images are shown in Table 2.

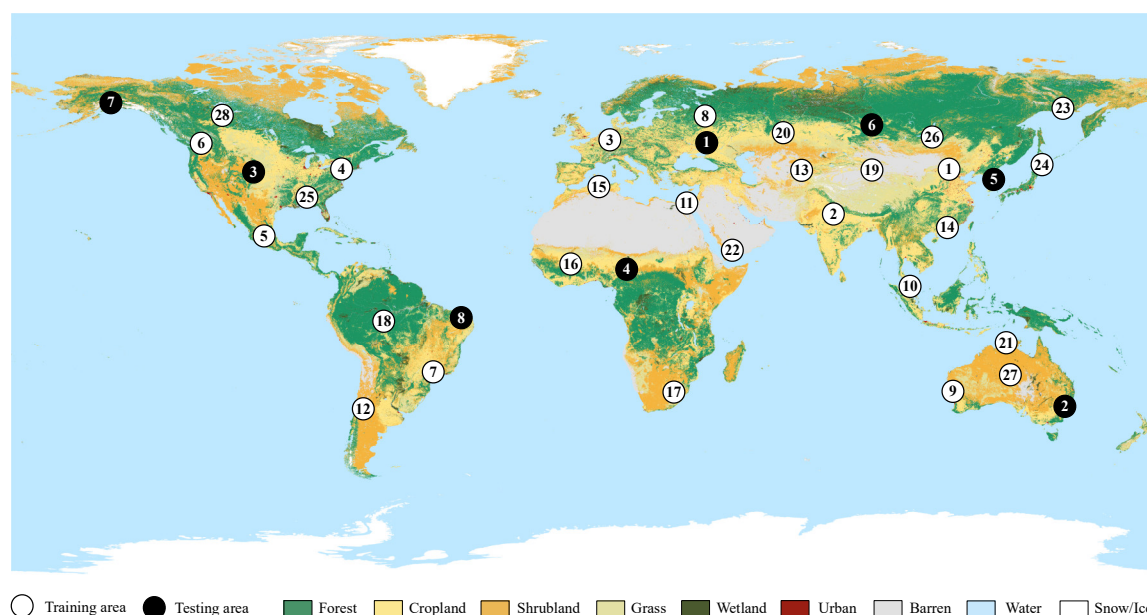


Figure 2. Distributions of training and testing data. Training areas are marked in white; testing areas are marked in black. The number in each area is corresponding to each number in Table 2. The landcover background is derived from ESA-CCI-LC [69].

Table 2. Details of Sentinel-2 training and testing image pairs.

Pair Condition		Product ID	Country/Land Cover	Date
Training	1 Cloud-free	S2A_MSIL1C_20160403T030602_N0201_R075_T50TMK_20160403T031209	China Urban	2016.04.03
	Cloudy	S2A_MSIL1C_20160413T031632_N0201_R075_T50TMK_20160413T031626	Urban	2016.04.13
	2 Cloud-free	S2A_MSIL1C_20181111T053041_N0207_R105_T43RGM_20181111T083104	Indian	2018.11.11
	Cloudy	S2A_MSIL1C_20181121T053121_N0207_R105_T43RGM_20181121T091419	Urban	2018.11.21
	3 Cloud-free	S2A_MSIL1C_20160925T104022_N0204_R008_T32ULB_20160925T104115	Germany	2016.09.25
	Cloudy	S2A_MSIL1C_20160915T104022_N0204_R008_T32ULB_20160915T104018	Urban	2016.09.15
	4 Cloud-free	S2A_MSIL1C_20160528T153912_N0202_R011_T18TWL_20160528T154746	United States	2016.05.28
	Cloudy	S2A_MSIL1C_20160518T155142_N0202_R011_T18TWL_20160518T155138	Urban	2016.05.18
	5 Cloud-free	S2A_MSIL1C_20181208T170701_N0207_R069_T14QMG_20181208T202913	Mexico	2018.12.08
	Cloudy	S2A_MSIL1C_20181218T170711_N0207_R069_T14QMG_20181218T203015	Urban	2018.12.18
	6 Cloud-free	S2A_MSIL1C_20180809T190911_N0206_R056_T10UFB_20180810T002400	Canada	2018.08.10
	Cloudy	S2A_MSIL1C_20180819T190911_N0206_R056_T10UFB_20180820T002955	Vegetation	2018.08.20
	7 Cloud-free	S2A_MSIL1C_20190306T132231_N0207_R038_T22KHV_20190306T164115	Brazil Vegetation	2019.03.06

8	Cloudy	S2A_MSIL1C_20190224T132231_N0207_R038_T22KHV_20190224T164104		2019.02.24
	Cloud-free	S2A_MSIL1C_20181012T084901_N0206_R107_T37VCC_20181012T110218	Russia	2018.10.12
	Cloudy	S2A_MSIL1C_20181022T085011_N0206_R107_T37VCC_20181022T110901	Urban	2018.10.22
9	Cloud-free	S2A_MSIL1C_20190619T023251_N0207_R103_T50JKP_20190619T071925	Australia	2019.06.19
	Cloudy	S2A_MSIL1C_20190629T023251_N0207_R103_T50JKP_20190629T053618	Vegetation	2019.06.29
10	Cloud-free	S2A_MSIL1C_20190901T032541_N0208_R018_T47NPF_20190901T070148	Malaysia	2019.09.01
	Cloudy	S2A_MSIL1C_20190911T032541_N0208_R018_T47NPF_20190911T084555	Urban	2019.09.11
11	Cloud-free	S2A_MSIL1C_20160419T083012_N0201_R021_T36RUU_20160419T083954	Egypt	2016.04.19
	Cloudy	S2A_MSIL1C_20160409T083012_N0201_R021_T36RUU_20160409T084024	Bare land	2016.04.09
12	Cloud-free	S2A_MSIL1C_20190218T143751_N0207_R096_T19HCC_20190218T175945	Chile	2019.02.18
	Cloudy	S2A_MSIL1C_20190208T143751_N0207_R096_T19HCC_20190208T180253	Vegetation	2019.02.08
13	Cloud-free	S2A_MSIL1C_20180609T061631_N0206_R034_T42TWL_20180609T081837	Uzbekistan	2018.06.09
	Cloudy	S2A_MSIL1C_20180530T061631_N0206_R034_T42TWL_20180530T082050	Bare land	2018.05.30
14	Cloud-free	S2A_MSIL1C_20191111T025941_N0208_R032_T49QGF_20191111T055938	China	2019.11.11
	Cloudy	S2A_MSIL1C_20191101T025841_N0208_R032_T49QGF_20191101T054434	Urban	2019.11.01
15	Cloud-free	S2A_MSIL1C_20190818T103031_N0208_R108_T31SEA_20190818T124651	Algeria	2019.08.18
	Cloudy	S2A_MSIL1C_20190808T103031_N0208_R108_T31SEA_20190808T124427	Vegetation	2019.08.08
16	Cloud-free	S2A_MSIL1C_20191202T105421_N0208_R051_T29PPP_20191202T112025	Mali	2019.12.02
	Cloudy	S2A_MSIL1C_20191212T105441_N0208_R051_T29PPP_20191212T111831	Bare land	2019.12.12
17	Cloud-free	S2A_MSIL1C_20190919T074611_N0208_R135_T35JPL_20190919T105208	South Africa	2019.09.19
	Cloudy	S2A_MSIL1C_20190929T074711_N0208_R135_T35JPL_20190929T100745	Bare land	2019.09.29
18	Cloud-free	S2A_MSIL1C_20190725T142801_N0208_R053_T20LMR_20190725T175149	Brazil	2019.07.25
	Cloudy	S2A_MSIL1C_20190804T142801_N0208_R053_T20LMR_20190804T175038	Vegetation	2019.08.04
19	Cloud-free	S2A_MSIL1C_20191101T043931_N0208_R033_T46TDK_20191101T074915	China	2019.11.01
	Cloudy	S2A_MSIL1C_20191022T043831_N0208_R033_T46TDK_20191022T063301	Bare land	2019.10.22

20	Cloud-free	S2A_MSIL1C_20160509T065022_N0202_R020_T41UNV_20160509T065018	Kazakhstan	2016.05.09	
	Cloudy	S2A_MSIL1C_20160519T064632_N0202_R020_T41UNV_20160519T064833	Bare land	2016.05.19	
21	Cloud-free	S2A_MSIL1C_20191019T012631_N0208_R131_T53LKF_20191019T030531	Australia	2019.10.19	
	Cloudy	S2A_MSIL1C_20191029T012721_N0208_R131_T53LKF_20191029T040003	Vegetation	2019.10.29	
22	Cloud-free	S2A_MSIL1C_20190503T071621_N0207_R006_T38PMB_20190503T092340	Yemen	2019.05.03	
	Cloudy	S2A_MSIL1C_20190423T071621_N0207_R006_T38PMB_20190423T093049	Bare land	2019.04.23	
23	Cloud-free	S2A_MSIL1C_20190724T011701_N0208_R031_T56VLM_20190724T031136	Russia	2019.07.24	
	Cloudy	S2A_MSIL1C_20190714T011701_N0208_R031_T56VLM_20190714T031656	Vegetation	2019.07.14	
24	Cloud-free	S2A_MSIL1C_20181020T012651_N0206_R074_T54TXN_20181020T032526	Japan	2018.10.20	
	Cloudy	S2A_MSIL1C_20181010T012651_N0206_R074_T54TXN_20181010T055606	Urban	2018.10.10	
25	Cloud-free	S2A_MSIL1C_20170224T162331_N0204_R040_T16REV_20170224T162512	United States	2017.02.24	
	Cloudy	S2A_MSIL1C_20170214T162351_N0204_R040_T16REV_20170214T163022	Urban	2017.02.14	
26	Cloud-free	S2A_MSIL1C_20190613T032541_N0207_R018_T49UFT_20190613T062257	Russia	2019.06.13	
	Cloudy	S2A_MSIL1C_20190623T032541_N0207_R018_T49UFT_20190623T061953	Vegetation	2019.06.23	
27	Cloud-free	S2A_MSIL1C_20190208T011721_N0207_R088_T53KLP_20190208T024521	Australia	2019.02.08	
	Cloudy	S2A_MSIL1C_20190129T011721_N0207_R088_T53KLP_20190129T024501	Bare land	2019.01.29	
28	Cloud-free	S2A_MSIL1C_20190530T184921_N0207_R113_T12VVN_20190530T222535	Canada	2019.05.30	
	Cloudy	S2A_MSIL1C_20190520T184921_N0207_R113_T12VVN_20190520T222900	Vegetation	2019.05.20	
Testing	1	Cloud-free	S2A_MSIL1C_20150826T084006_N0204_R064_T37UCQ_20150826T084003	Ukraine	2015.08.26
		Cloudy	S2A_MSIL1C_20150905T083736_N0204_R064_T37UCQ_20150905T084002	Urban	2015.09.05
	2	Cloud-free	S2A_MSIL1C_20191101T000241_N0208_R030_T56HLH_20191101T012241	Australia	2019.11.10
		Cloudy	S2A_MSIL1C_20191111T000241_N0208_R030_T56HLH_20191111T012137	Urban	2019.11.11
	3	Cloud-free	S2A_MSIL1C_20190711T174911_N0208_R141_T13TEE_20190711T212846	United States	2019.07.11
		Cloudy	S2A_MSIL1C_20190701T174911_N0207_R141_T13TEE_20190701T212910	Bare land	2019.07.01
	4	Cloud-free	S2A_MSIL1C_20190201T093221_N0207_R136_T32PRR_20190201T113425	Nigeria	2019.02.01

5	Cloudy	S2A_MSIL1C_20190211T093121_N0207_R136_T32PRR_20190211T103706	Bare land	2019.02.11
	Cloud-free	S2A_MSIL1C_20190314T021601_N0207_R003_T52SCF_20190314T055026	South Korea	2019.03.14
6	Cloudy	S2A_MSIL1C_20190304T021601_N0207_R003_T52SCF_20190304T042035	Urban	2019.03.04
	Cloud-free	S2A_MSIL1C_20180804T045701_N0206_R119_T46VDH_20180804T065907	Russia	2018.08.04
7	Cloudy	S2A_MSIL1C_20180725T045701_N0206_R119_T46VDH_20180725T065359	Vegetation	2018.07.25
	Cloud-free	S2A_MSIL1C_20190707T213531_N0207_R086_T05VPJ_20190707T231819	United States	2019.07.07
8	Cloudy	S2A_MSIL1C_20190627T213531_N0207_R086_T05VPJ_20190628T010801	Vegetation	2019.06.28
	Cloud-free	S2A_MSIL1C_20190605T125311_N0207_R052_T24MXV_20190605T160555	Brazil	2019.06.05
	Cloudy	S2A_MSIL1C_20190615T125311_N0207_R052_T24MXV_20190615T142536	Vegetation	2019.06.15

We set the time lag of the acquisition dates of cloudy and corresponding cloud-free images to 10 days, which is the revisit time of Sentinel-2A satellite, to minimize the difference between cloudy and cloud-free images as much as possible. The manual collection of one single pair of cloudy and corresponding cloud-free images with the shortest time lag of data acquisition took about 30 min, because many Sentinel-2A images are covered with clouds, and it was challenging to find a corresponding cloud-free image. To assess the performance of CR-MSS for cloud removal in different seasons, the acquisition dates of training and testing data were from 2015 to 2019, in which all seasons are covered. Twenty-eight training image pairs and eight testing image pairs were collected from the Copernicus Open Access Hub website, under the conditions: 1) widely distributed, 2) covering all seasons, and 3) shortest time lag for data collection. The performance of all methods can be fully evaluated with the all these training and testing data, which have the shortest time lag and different land cover types and covers all seasons. The dataset is shared on <https://github.com/Neooolee/WHUS2-CR>.

2.3. Method

In the proposed CR-MSS method, the paired cloudy and cloud-free multispectral images are used in end-to-end training. Inspired by U-Net architecture, CR-MSS is designed to make better use of the features at different levels. CR-MSS has two input/output branches that are used to handle Vis/NIR bands and VRE/SWIR, respectively. The features of multispectral images are first fused after processed by input branches and then compressed before output branches. The network architecture of CR-MSS is introduced in this section.

We group Vis/NIR bands and VRE/SWIR into two separate multispectral images, that are then inputted to two input branches in CR-MSS. Pooling is the most common operation for down-sampling the feature maps in deep learning, as the cost of losing neighborhood information by this operation. Instead, we use convolutional layers with stride = 2 instead of the pooling layers to down-sample the feature maps in CR-MSS. Convolutional layer can preserve neighborhood information as much as possible by optimizing the parameters. The deconvolution layer with stride = 2 is adopted to up-sample the feature maps. The kernel sizes of convolutional and deconvolution layers in CR-MSS are set to 3×3 and 4×4 , respectively. The details of the convolutional and deconvolution layers are as follows:

1. The *convolution layer* contains multiple convolution kernels and is used to extract features from input data. Each element that constitutes the convolution kernel corresponds to a weight coefficient and a bias. Each neuron in the convolution layer is connected with multiple neurons in the adjacent region from the previous layer, and the size of the region depends on the size of the convolution kernel.
2. The *deconvolution layer* is used to up-sample the input data, by interpolating between the elements of the input matrix, and then, constructing the same connection and operation as a normal convolutional layer, except that it starts from the opposite direction.

As shown in Figure 3, there are two input branches in CR-MSS: Vis/NIR input branch (Vis/NIR-In) and VRE/SWIR input branch (VRE/SWIR-In). A fusion tunnel performs feature fusion, while the compression tunnel performs feature compression. Vis/NIR-In is for processing Vis/NIR bands; VRE/SWIR-In is for VRE/SWIR bands. To extract low-level features of the input images and preserve the original information from input images, the stride in the first convolutional layer in each input branch is set to 1 to keep the size of inputs unchanged. Since Vis/NIR has higher spatial resolution than VRE/SWIR, Vis/NIR-In has one more convolutional layer than VRE/SWIR-In, with stride = 2. This extra convolutional layer in Vis/NIR-In is used to down-sample the input feature maps to the same size as the output of VRE/SWIR-In.

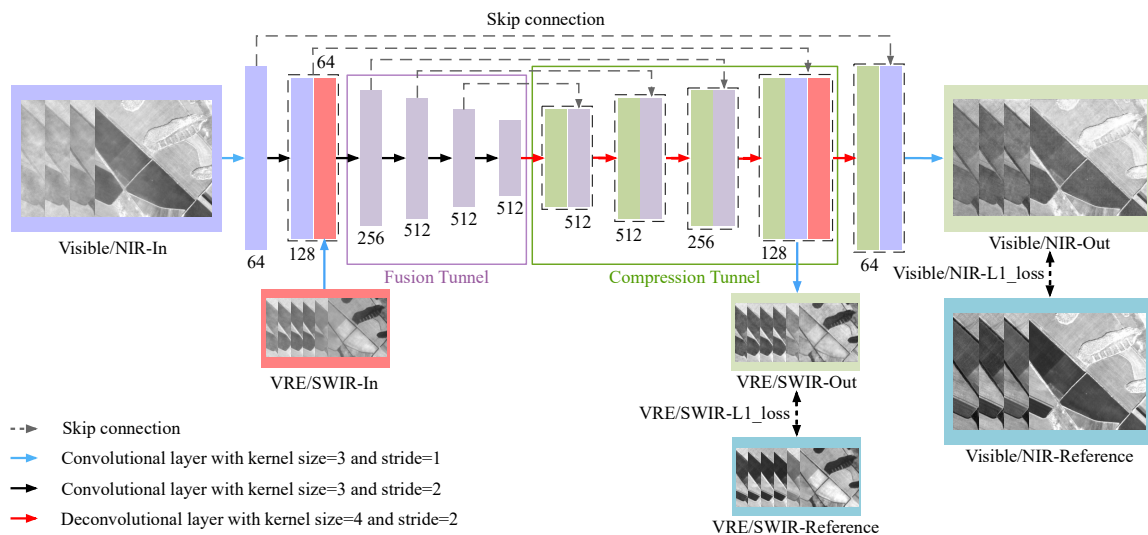


Figure 3. The architecture of CR-MSS. The number under/on each block is the number of feature maps of it. The feature maps in each dotted box are concatenated at feature channels.

As we know, there is a strong correlation among the spectral responses of the same target in different bands. Since SWIR bands are the least affected by clouds among all bands, and VRE bands are less affected than Visible bands, the spectral features of VRE/SWIR bands can be used to restore the missing information in Visible bands. Therefore, the output feature maps of VRE/SWIR-In are concatenated with the output feature maps of Vis/NIR-In at feature (spectral) channel. Then the concatenated feature maps are put into the fusion tunnel to fuse the features from all experimental bands. There are four convolutional layers in the fusion tunnel. The stride of each convolutional layer is set to 2 to down-sample the feature maps and extract high-level features. It can be seen that the number of concatenated feature maps is 192 (128+64); after being processed by the first convolutional layer in the fusion tunnel, the number of feature maps is extended to 256. In this way, the features in Vis/NIR/VRE/SWIR bands are fused and expanded. The next three convolutional layers in the fusion tunnel are used to fuse and extract more features for the restoration of background information.

There are four deconvolution layers with stride = 2 in the compression tunnel that are used to up-sample the feature maps and compress features. The output feature maps of the last convolutional layer in the fusion tunnel are put into the first deconvolutional layer in compression tunnel directly. To make full use of features at different levels and generate high quality cloud-free images, the output of each convolutional layer before the middle convolutional layer is copied and concatenated with the output of each symmetric deconvolution layer. These concatenations can solve the problem of information loss of the original images in the process of down-sample operations and help the network converge faster.

Because there are two input branches to handle Vis/NIR bands and VRE/SWIR bands in CR-MSS, respectively, the CR-MSS also has two output branches to handle Vis/NIR and VRE/SWIR bands. The output branches and the input branches are symmetric in CR-MSS. The output branch that handles Vis/NIR bands (Vis/NIR-Out) has one more deconvolution layer than the branch that handles VRE/SWIR bands (VRE/SWIR-Out). This deconvolution layer in Vis/NIR-Out is used to up-sample the feature maps to the same size as the input of Vis/NIR-In. In order to get the same number of channels as in input images, we add a convolutional layer with stride = 1 and output channels (bands) = 4 and 6 in Vis/NIR-Out and VRE/SWIR-Out branches, respectively. In this way, Vis/NIR-Out outputs a multispectral image with Vis/NIR bands; VRE/SWIR-Out outputs a multispectral image with VRE/SWIR bands.

We adopt L1 loss as the loss function of CR-MSS. For each band in input images, we measure the difference between the output cloud-free and corresponding ground-truth or reference cloud-free band to optimize CR-MSS. The average loss over all bands is taken as the final loss. The parameters of CR-MSS are updated with the loss function L_G as follows:

$$L_G = \frac{1}{k} \sum_{i=1}^k |G(z)_i - x_i| \quad (1)$$

where z is the input image, x_i the i th band in reference images, $G(z)_i$ the i th band in generated cloud-free images, and k the total number of spectral bands of z . CR-MSS will learn the difference between $G(z)$ and x , then optimize the parameters to generate more clear images. After CR-MSS is well-trained, it can restore thin cloud images into cloud-free images.

2.4. Data Pre-processing and Experiment Setting

Three DL-based methods and two traditional methods for cloud removal are compared with CR-MSS in our experiments: RSC-Net [62], U-Net [35], Cloud-GAN [64], AHF [24], and MSRCF [70]. Because CR-MSS is designed for cloud removal in a single image, multi-temporal cloud removal methods are not included in this comparison. RSC-Net and U-Net are end-to-end methods that require paired cloudy and corresponding cloud-free images. Cloud-GAN is a semi-supervised method that requires unpaired cloudy and cloud-free images. AHF is an adaptive homomorphic filtering method for cloud removal and requires cloud masks. MSRCF is based on MRSCR and with cuda parallel acceleration for image enhancement such as image dehazing.

In order to train the CNN-based methods, we cropped all experimental images into small patches without overlapping by slide windows of different sizes. Since the spatial resolution of Vis/NIR bands and VRE/SWIR bands is 10 and 20 m, respectively, we set the corresponding slide window sizes to 256×256 and 128×128 pixels. The area extents of each 256×256 patch at 10 m resolution matches that of each corresponding 128×128 patch at 20 m resolution. In this way, 17,182 pairs of cloudy and corresponding cloud-free multispectral images are produced. Additionally, 13,389 pairs of training samples are generated from 28 pairs of training images, and 3782 pairs of testing samples are generated from eight pairs of testing images. The training dataset was augmented by flipping

training samples horizontally and vertically, and rotating them at 90° , 180° , and 270° angles. In this way, 80,334 pairs of cloudy and cloud-free training patches were obtained. The traditional methods can process the whole image at once to perform cloud removal. Because the AHF method requires cloud masks, we adopt the official Sentinel-2 cloud masks tool Sen2Cor [71] for Sentinel-2 to produce the cloud masks in our experiment.

Because the spectral reflectance values of objects in images will change in 10 days, the cloud-free images in testing data cannot be used to quantitatively evaluate the performance of all methods on spectral information preservation. There are many methods using simulated cloud images to evaluate their performances [12,50,72,73]. Therefore, in order to quantitatively evaluate the performances of all methods on spectral information preservation, we simulated cloudy images with Adobe Photoshop following [65], from cloud-free images of the testing dataset. First, we added a transparent layer for each band with different transparency rate, then we applied the Photoshop command Filter $\rightarrow$ Render $\rightarrow$ Cloud on each transparent layer to produce a cloud layer. Finally, each transparent and band pair were fused together to obtain a cloud band. All methods were applied on the simulated cloud images, in order to remove clouds and quantitatively evaluate the performances of all methods on spectral information preservation.

In the training stage, all deep learning-based methods were trained on real cloud and cloud-free image pairs in training dataset listed in Table 2. In the testing stage, all baseline methods and CR-MSS were tested on real cloud and cloud-free image pairs for PSNR and SSIM values, and on simulated cloud and cloud-free image pairs for NRMSE value, to better control for spectral reflectance changes, as explained in the previous paragraph. Figures 4–13 show the cloud removal results of all methods on real cloud and cloud-free images.

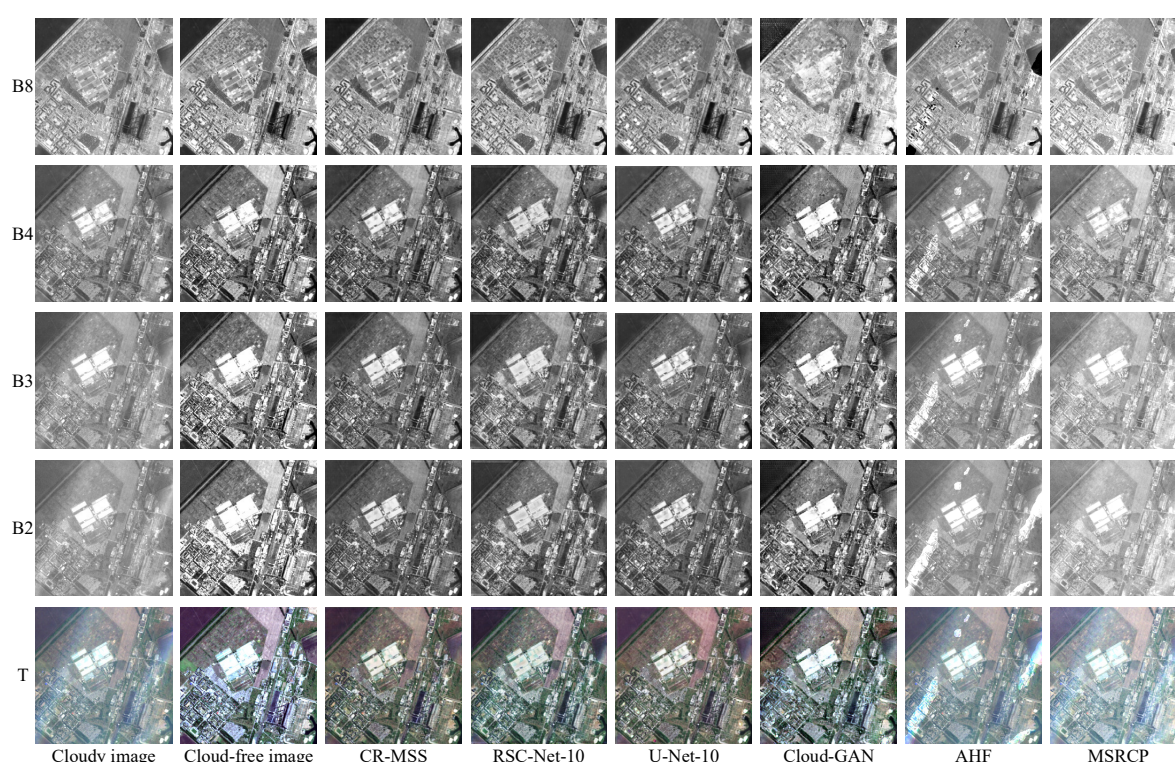


Figure 4. Visual comparison results on urban sample (Ukraine) under thin clouds from first real testing image pair in Table 2. T, B2, B3, B4, and B8 are true color composited image, bands 2, 3, 4, and 8, respectively. Columns 1–8 are cloudy and corresponding cloud-free images, results of CR-MSS, RSC-Net-10, U-Net-10 and Cloud-GAN, AHF, and MSRCP.

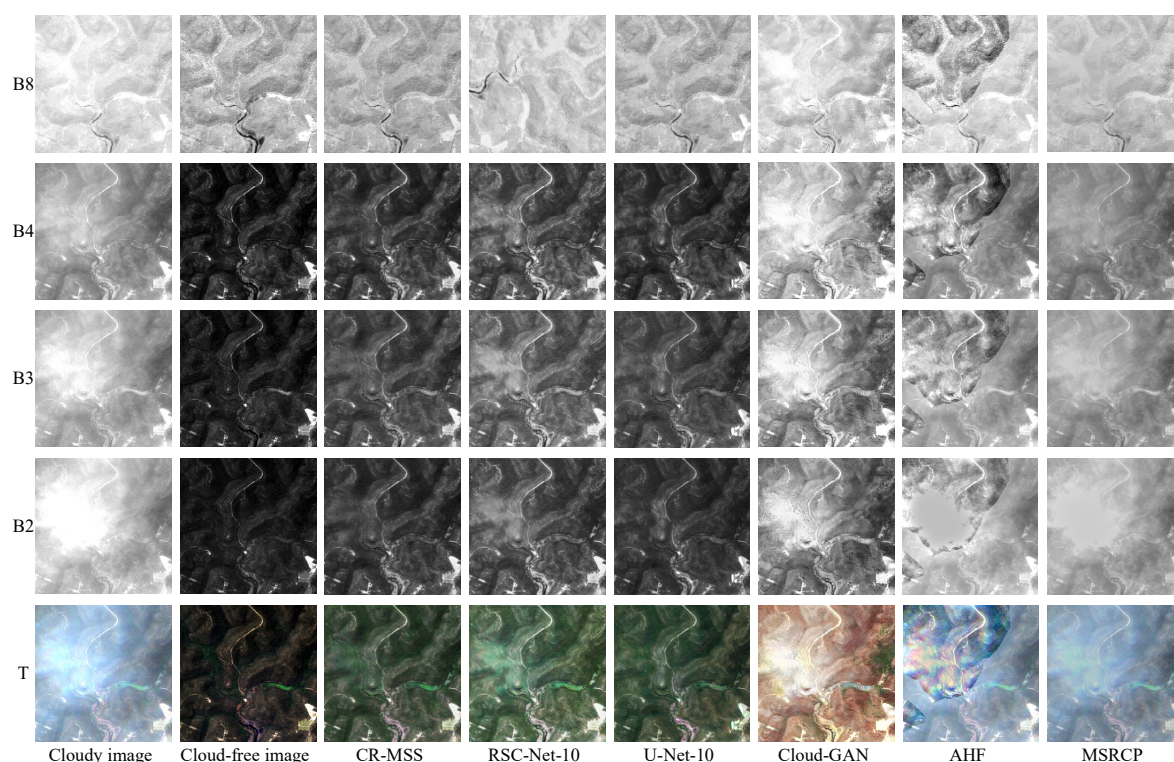


Figure 5. Visual comparison results on vegetation sample (Australia) under thin clouds from second real testing image pair in Table 2. T, B2, B3, B4, and B8 are true color composed image, bands 2, 3, 4, and 8, respectively. Columns 1–8 are cloudy and corresponding cloud-free images, results of CR-MSS, RSC-Net-10, U-Net-10 and Cloud-GAN, AHF, and MSRCP.

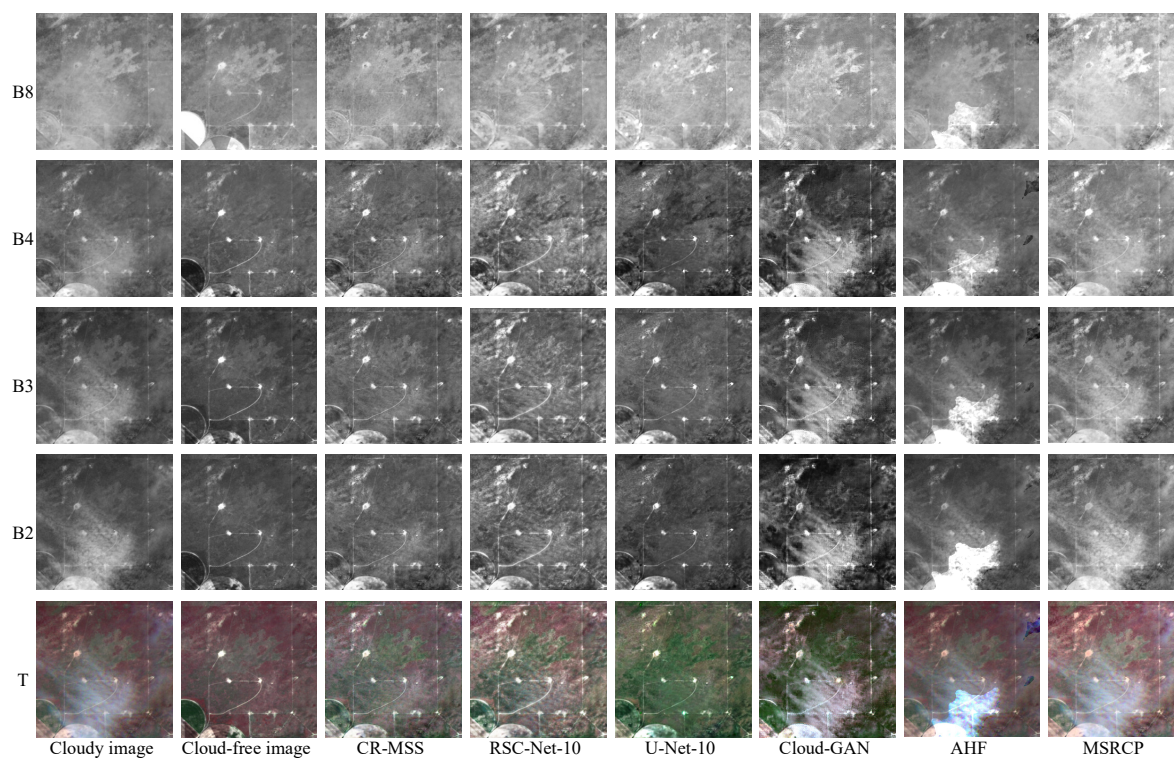


Figure 6. Visual comparison results on bare land sample (United States) under thin clouds from third real testing image pair in Table 2. T, B2, B3, B4, and B8 are true color composed image, bands 2, 3, 4, and 8, respectively. Columns 1–8 are cloudy and corresponding cloud-free images, results of CR-MSS, RSC-Net-10, U-Net-10 and Cloud-GAN, AHF, and MSRCP.

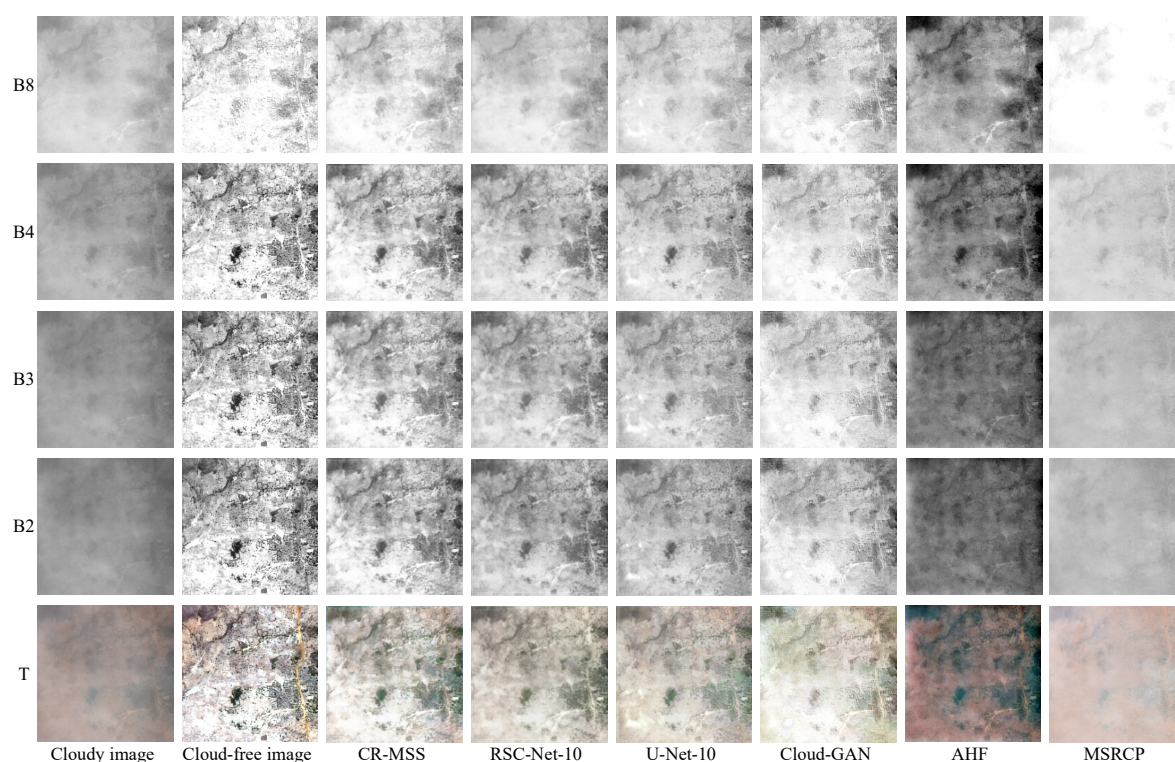


Figure 7. Visual comparison results on bare land sample (Nigeria) under thin clouds from 4th real Table 2. T, B2, B3, B4, and B8 are true color composited image, bands 2, 3, 4, and 8, respectively. Columns 1–8 are cloudy and corresponding cloud-free images, results of CR-MSS, RSC-Net-10, U-Net-10 and Cloud-GAN, AHF, and MSRCP.

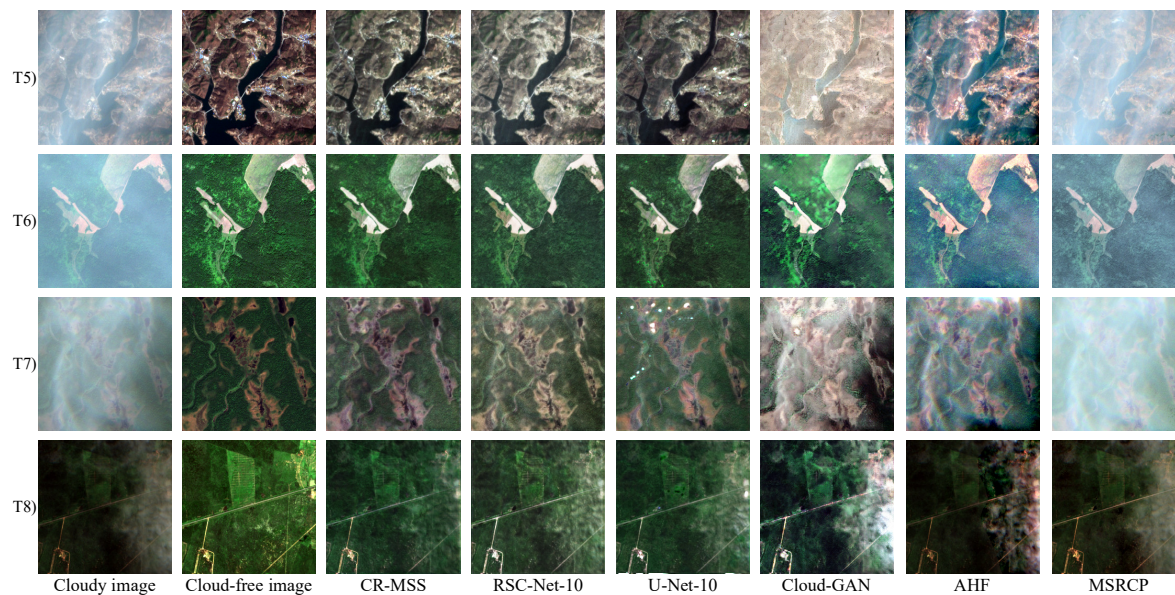


Figure 8. Visual comparison on four samples. (T5), (T6), (T7) and (T8) are true color composites from fifth, sixth, seventh, and eighth real testing image pairs in Table 2. Columns 1–8 are cloudy and corresponding cloud-free images, results of CR-MSS, RSC-Net-10, U-Net-10 and Cloud-GAN, AHF, and MSRCP.

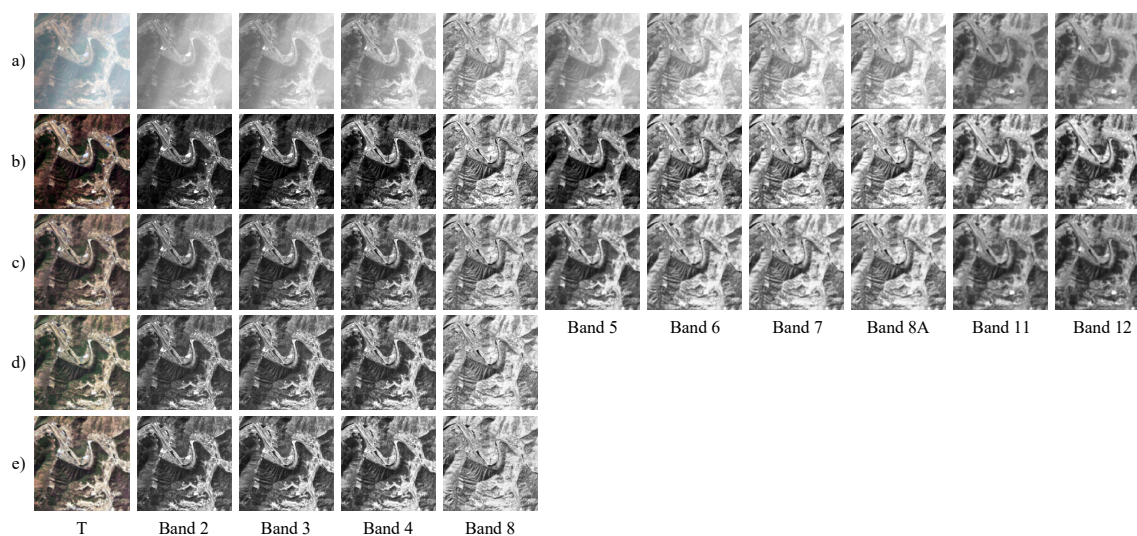


Figure 9. Visual comparison results on mountain sample (South Korea) under thin clouds from fifth real testing image pair in Table 2. By row: (a) cloudy image, (b) corresponding cloud-free image, (c) results of CR-MSS, (d) CR-MSS-10-4, and (e) CR-MSS-4-4.

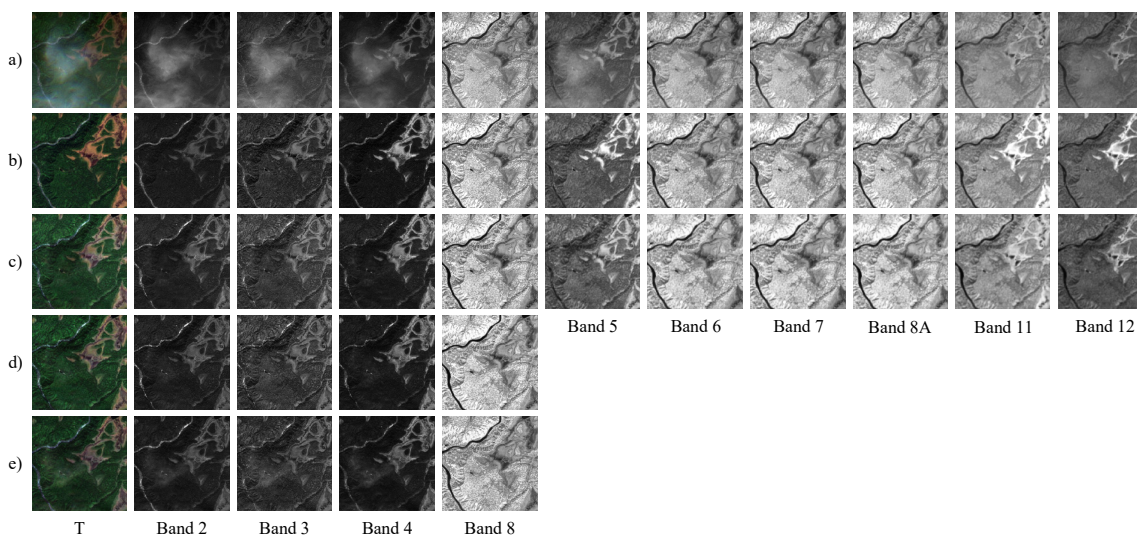


Figure 10. Visual comparison results on grass sample (Russia) under thin clouds from sixth real testing image pair in Table 2. By row: (a) cloudy image, (b) corresponding cloud-free image, (c) results of CR-MSS, (d) CR-MSS-10-4, and (e) CR-MSS-4-4.

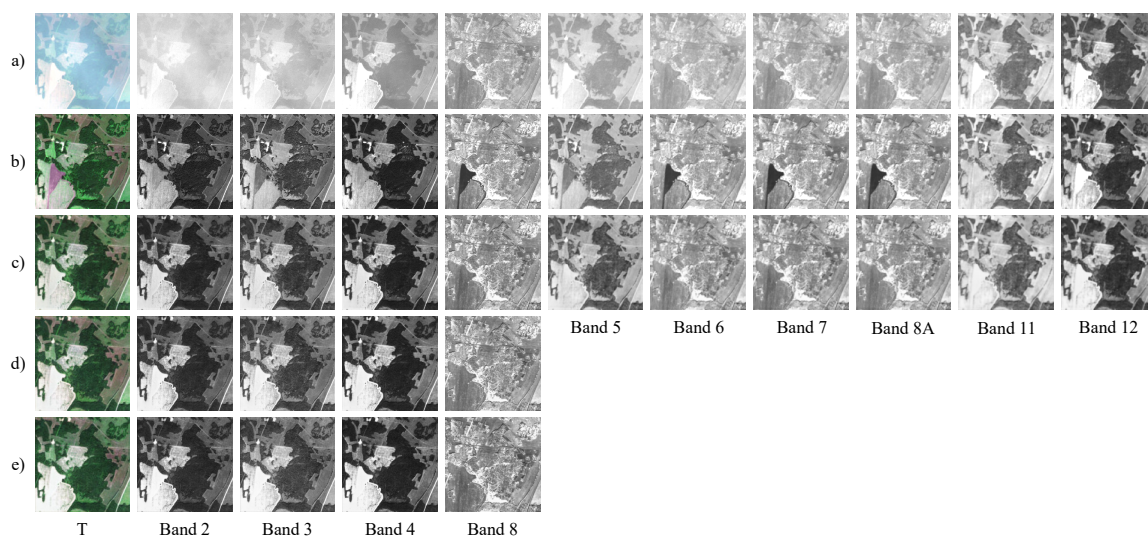


Figure 11. Visual comparison results on forest sample (United States) under thin clouds from seventh real testing image pair in Table 2. By row: (a) cloudy image, (b) corresponding cloud-free image, (c) results of CR-MSS, (d) CR-MSS-10-4, and (e) CR-MSS-4-4.

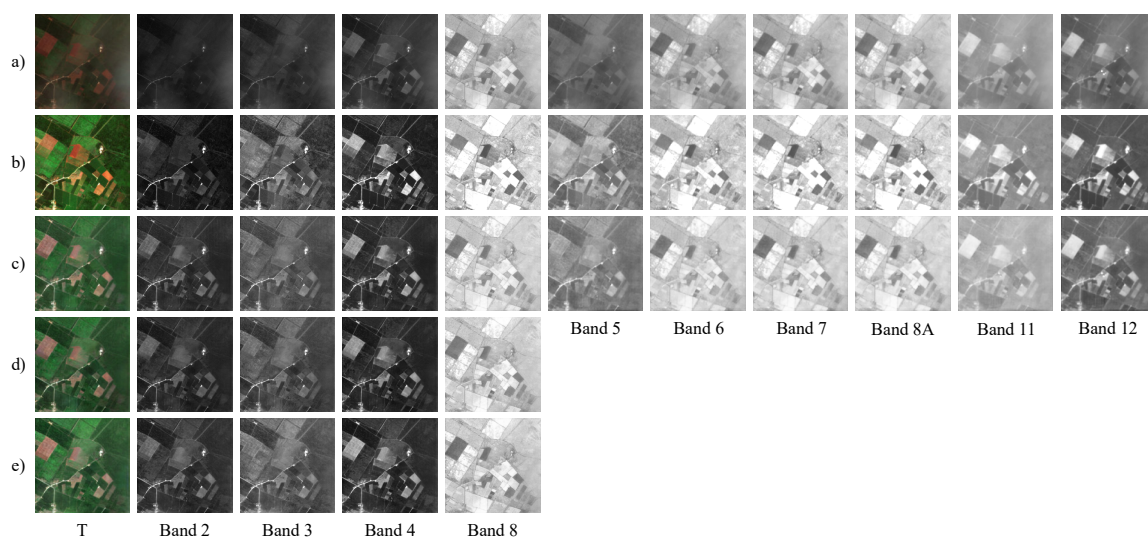


Figure 12. Visual comparison results on farmland sample (Brazil) under thin clouds and cloud shadows from eighth real testing image pair in Table 2. By row: (a) cloudy image, (b) corresponding cloud-free image, (c) results of CR-MSS, (d) CR-MSS-10-4, and (e) CR-MSS-4-4.

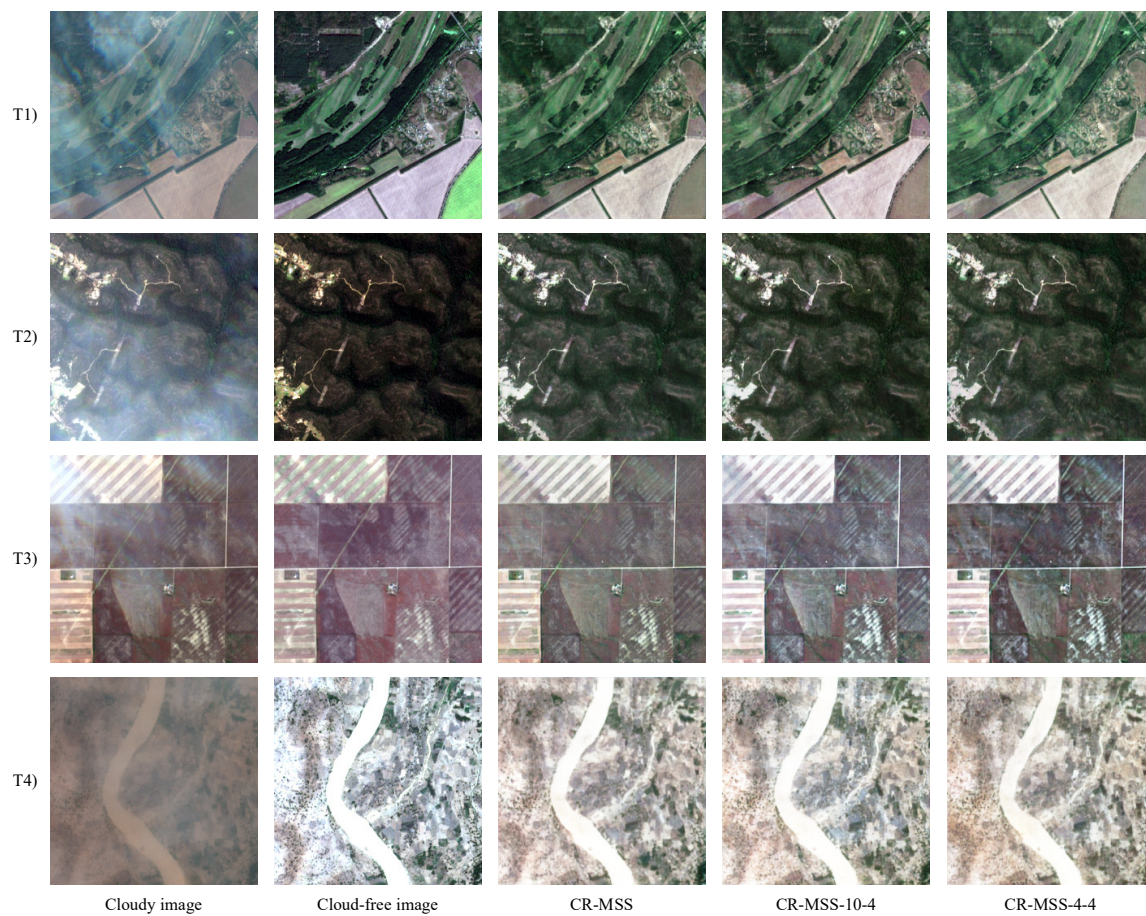


Figure 13. Visual comparison on four samples. (T1), (T2), (T3), and (T4) are true colour composites from first, second, third, and fourth real testing image pairs in Table 2. Columns 1–5 are cloudy and corresponding cloud-free images, results of CR-MSS, RSC-Net-10, U-Net-10, and Cloud-GAN.

The inputs of AHF, MRSCR, and Cloud-GAN are Vis/NIR bands. Since CR-MSS is designed to handle Vis/NIR and VRE/SWIR bands together, the input of CR-MSS is Vis/NIR/VRE/SWIR bands. Because RSC-Net and U-Net are end-to-end deep learning-based methods similar to CR-MSS, in order to compare RSC-Net and U-Net with CR-MSS fairly, VRE/SWIR bands were first rescaled to 10 m resolution by bi-linear interpolation, then concatenated with Vis/NIR bands to train RSC-Net-10 and U-Net-10. In addition, to analyze the influence of the VRE/SWIR bands on CR-MSS performance for thin cloud removal, we conducted ablation experiments, i.e., with or without VRE/SWIR bands as inputs/outputs or not.

The hyper-parameters are set the same for all deep learning-based methods. The batch size is set to 1, and iterations for training are set to 600000. Adam-optimizer [74] is adopted to optimize the parameters of all networks, and the hyper parameters of Adam-optimizer are fixed as: $\beta_1 = 0.9$, $\beta_2 = 0.999$, and the initial learning rate = 0.0002, with exponential decay at decay rate = 0.96. The training and testing experiments are both conducted with the TensorFlow running on Windows 7 operating system on a 16 Intel (R) Xeon CPU E5-2620 v4 @ 2.10 GHz and an NVIDIA GeForce GTX 1080Ti with 11 GB memory.

We take peak signal to noise ratio (PSNR), structural similarity index measurement (SSIM), and normalized root mean square error (NRMSE) as the quantitative evaluation measures. The PSNR and SSIM values of one single testing sample and average of all testing samples in each band are calculated to better compare the performances of all methods. As in Meraner et al. [49], the average NRMSE values in each band over all simulated testing samples were calculated to analyze the spectral preservation in each band.

3. Results

3.1. Comparison of Different Methods

To better present the cloud removal results of different areas worldwide, we selected one sample pair from each testing area. The experimental results of all cloud removal methods on these selected samples are presented in Figures 4–8. Figures 4–7 show the cloud removal results on each band of four samples from the first, second, third, and fourth testing pairs, respectively. The inputs/outputs were original Vis/NIR/VRE/SWIR bands for CR-MSS, original Vis/NIR, and rescaled VRE/SWIR bands for RSC-Net-10 and U-Net-10, and Vis/NIR bands for other baseline methods. Figures 4–7 show results of CR-MSS and baseline methods with Vis/NIR bands.

We can see that the results of CR-MSS are visually better than all compared methods on all bands in these Figs. For example, in Figure 5, the impact of cloud is very severe on background signal. CR-MSS removes most clouds in Vis/NIR bands. While results of RSC-Net-10 and U-Net-10 contain much noise in cloudy regions, Cloud-GAN translates the input cloudy image into another cloud contaminated image, the style of which is completely different from the input cloudy image and makes the result even worse. This is caused by the influence of cycle-consistent of Cloud-GAN. In the training of Cloud-GAN, it first translates a cloudy image into a cloud-free image. Then, the translated cloud-free image is translated back into the input cloudy image, which makes the textures of the cloud-free image and input cloudy image consistent. This results in some cloud information being retained in the translated cloud-free image. Results of both AHF and MSRPC on Figures 5–7 are not good visually, because the cloud component cannot be completely filtered out by Homomorphic filtering in AHF, and the model in MSRPC does not take the reflection of cloud into consideration. However on Figure 7, AHF have a better visual result than MSRPC because the thin clouds are fairly uniform and can be removed by low-pass filtering.

Figure 8 shows the true color composited results of four samples in fifth, sixth, seventh, and eighth testing pairs. Visually, the results of CR-MSS seem more realistic than baseline methods, and MSRPC looked the worst among all methods, while Cloud-GAN performs the worst among all deep learning-based methods. It is worth noticing that the (T8) in Figure 8 contains thick clouds and cloud shadows, and CR-MSS removes cloud shadows and most of the clouds, but results of all baseline methods still include many clouds.

Tables 3 and 4 show the corresponding PSNR and SSIM values of samples in Figures 4–7 and Figure 8, respectively. It can be seen that CR-MSS obtains the best performance on PSNR and SSIM values in most cases. It is worth noticing that AHF performs best on SSIM values on Figure 6 at bands 4/8 and T8 in Figure 8 at band 8. Although the visual results of AHF on Figure 6 and T8 in Figure 9 are worse than CR-MSS, AHF use cloud masks to preserve the cloud-free regions, which improves the SSIM values on average.

Table 3. Corresponding PSNR and SSIM values of samples in Figures 4–7 (all real testing image pairs from 1 to 4, Table 2). Best results are marked in bold.

Image	Index	Method	B2	B3	B4	B8	B5	B6	B7	B8A	B11	B12
T1 (Figure 4)	PSNR	CR-MSS	14.19	16.99	19.38	23.17	22.57	22.94	24.48	25.39	23.68	24.15
		RSC-Net-10	16.32	18.60	19.76	23.63	20.96	24.33	24.57	24.47	24.50	24.90
		U-Net-10	14.21	16.68	19.68	22.34	21.61	21.96	23.99	24.95	23.62	23.23
		Cloud-GAN	14.62	16.04	16.51	15.70	/	/	/	/	/	/
		AHF	13.50	14.26	16.35	20.68	/	/	/	/	/	/
		MRSCP	12.23	12.67	14.42	15.87	/	/	/	/	/	/
	SSIM	CR-MSS	0.67	0.74	0.81	0.86	0.86	0.90	0.91	0.92	0.88	0.89
		RSC-Net-10	0.72	0.76	0.82	0.87	0.85	0.90	0.92	0.92	0.9	0.91

		U-Net-10	0.65	0.7	0.79	0.84	0.83	0.89	0.91	0.91	0.85	0.87
		Cloud-GAN	0.50	0.51	0.52	0.45	/	/	/	/	/	/
		AHF	0.46	0.54	0.67	0.79	/	/	/	/	/	/
		MRSCP	0.50	0.57	0.69	0.78	/	/	/	/	/	/
T2 (Figure 5)	PSNR	CR-MSS	16.57	14.46	17.39	25.35	17.59	27.27	25.41	24.22	23.08	22.91
		RSC-Net-10	12.21	10.01	12.75	19.04	9.93	16.56	18.53	19.88	16.81	15.95
		U-Net-10	17.78	14.97	18.34	25.25	17.11	25.12	26.57	27.74	19.93	23.19
		Cloud-GAN	6.93	5.92	4.55	16.10	/	/	/	/	/	/
		AHF	4.63	5.92	7.15	17.41	/	/	/	/	/	/
		MRSCP	4.92	6.31	8.07	21.48	/	/	/	/	/	/
	SSIM	CR-MSS	0.58	0.54	0.63	0.78	0.80	0.87	0.87	0.88	0.87	0.86
		RSC-Net-10	0.45	0.41	0.52	0.72	0.64	0.82	0.83	0.84	0.83	0.80
		U-Net-10	0.60	0.54	0.63	0.74	0.77	0.86	0.87	0.88	0.84	0.85
		Cloud-GAN	0.23	0.23	0.21	0.41	/	/	/	/	/	/
		AHF	0.21	0.28	0.35	0.78	/	/	/	/	/	/
		MRSCP	0.24	0.31	0.37	0.72	/	/	/	/	/	/
T3 (Figure 6)	PSNR	CR-MSS	18.69	17.98	21.28	19.49	20.67	20.85	19.19	20.01	21.59	19.00
		RSC-Net-10	16.39	15.06	16.87	18.59	15.12	19.65	18.11	18.91	20.02	19.49
		U-Net-10	21.49	19.04	19.86	15.10	19.41	15.95	14.26	16.35	18.18	19.00
		Cloud-GAN	14.29	15.66	15.61	16.94	/	/	/	/	/	/
		AHF	13.01	14.14	16.69	17.99	/	/	/	/	/	/
		MRSCP	12.43	13.13	13.66	12.31	/	/	/	/	/	/
	SSIM	CR-MSS	0.71	0.73	0.74	0.76	0.77	0.78	0.76	0.75	0.81	0.80
		RSC-Net-10	0.60	0.61	0.63	0.77	0.67	0.77	0.77	0.79	0.85	0.82
		U-Net-10	0.70	0.70	0.69	0.72	0.73	0.76	0.74	0.75	0.77	0.78
		Cloud-GAN	0.48	0.57	0.54	0.55	/	/	/	/	/	/
		AHF	0.69	0.73	0.79	0.80	/	/	/	/	/	/
		MRSCP	0.62	0.67	0.72	0.72	/	/	/	/	/	/
T4 (Figure 7)	PSNR	CR-MSS	19.36	19.73	20.64	21.94	22.1	23.61	23.37	21.91	20.84	24.17
		RSC-Net-10	18.43	19.39	19.91	18.9	20.91	21.32	20.68	19.91	22.14	22.9
		U-Net-10	18.19	18.83	19.61	20.37	20.77	22.47	22.09	20.98	22.21	23.88
		Cloud-GAN	16.24	15.85	16.53	18.76	/	/	/	/	/	/
		AHF	9.55	9.63	10.15	9.71	/	/	/	/	/	/
		MRSCP	14.13	14.73	15.44	16.52	/	/	/	/	/	/
	SSIM	CR-MSS	0.62	0.62	0.67	0.66	0.71	0.74	0.75	0.79	0.84	0.85
		RSC-Net-10	0.56	0.55	0.59	0.56	0.61	0.61	0.61	0.67	0.73	0.77
		U-Net-10	0.53	0.54	0.57	0.6	0.62	0.64	0.66	0.71	0.77	0.79
		Cloud-GAN	0.40	0.41	0.42	0.57	/	/	/	/	/	/
		AHF	0.31	0.39	0.52	0.58	/	/	/	/	/	/
		MRSCP	0.18	0.21	0.30	0.40	/	/	/	/	/	/

Table 4. Corresponding PSNR and SSIM values of samples in Figure 8 (all real testing image pairs from 5 to 8, Table 2). Best results are marked in bold.

Image	Index	Method	B2	B3	B4	B8	B5	B6	B7	B8A	B11	B12
T5 (Figure 8)	PSNR	CR-MSS	19.85	18.39	21.17	18.63	20.25	21.14	20.29	17.67	14.98	15.82
		RSC-Net-10	14.42	13.5	16.28	18.73	15.68	17.98	18.74	19.32	17.41	19.36
		U-Net-10	20.02	19.6	20.12	15.94	20.09	17.00	16.21	14.36	12.88	14.10
		Cloud-GAN	7.55	6.71	6.87	10.74	/	/	/	/	/	/
		AHF	10.05	11.18	14.03	15.55	/	/	/	/	/	/

T6 (Figure 8)	SSIM	MRSCP	4.51	5.33	7.25	12.76	/	/	/	/	/
		CR-MSS	0.78	0.77	0.77	0.77	0.84	0.86	0.87	0.86	0.81
		RSC-Net-10	0.68	0.64	0.71	0.71	0.76	0.78	0.8	0.81	0.8
		U-Net-10	0.74	0.74	0.70	0.73	0.79	0.79	0.8	0.79	0.73
		Cloud-GAN	0.32	0.30	0.33	0.38	/	/	/	/	/
		AHF	0.45	0.54	0.70	0.75	/	/	/	/	/
	PSNR	MRSCP	0.17	0.21	0.34	0.59	/	/	/	/	/
		CR-MSS	25.13	22.12	27.54	23.63	23.02	23.56	24.57	24.78	30.29
		RSC-Net-10	22.92	22.46	25.18	19.97	22.24	20.34	20.74	22.95	23.23
		U-Net-10	25.04	19.78	26.58	20.66	19.52	18.81	20.6	22.3	24.33
		Cloud-GAN	16.63	16.53	19.29	14.80	/	/	/	/	/
		AHF	8.87	14.86	14.71	12.49	/	/	/	/	/
T7 (Figure 8)	SSIM	MRSCP	8.44	13.89	13.16	15.71	/	/	/	/	/
		CR-MSS	0.65	0.71	0.80	0.90	0.87	0.91	0.91	0.92	0.96
		RSC-Net-10	0.66	0.7	0.80	0.89	0.87	0.89	0.9	0.91	0.94
		U-Net-10	0.65	0.68	0.78	0.87	0.83	0.87	0.89	0.91	0.92
		Cloud-GAN	0.41	0.46	0.54	0.50	/	/	/	/	/
		AHF	0.28	0.60	0.56	0.67	/	/	/	/	/
	PSNR	MRSCP	0.40	0.58	0.59	0.83	/	/	/	/	/
		CR-MSS	16.73	18.39	16.78	21.57	20.83	23.62	23.54	23.22	22.65
		RSC-Net-10	14.11	13.82	12.69	22.28	16.3	23.16	23.22	23.17	20.96
		U-Net-10	15.01	15.06	15.36	23.01	18.56	22.2	22.84	24.41	21.93
		Cloud-GAN	8.10	8.77	7.57	15.95	/	/	/	/	/
		AHF	9.15	10.88	10.94	15.06	/	/	/	/	/
T8 (Figure 8)	SSIM	MRSCP	3.36	4.35	5.30	13.16	/	/	/	/	/
		CR-MSS	0.61	0.67	0.62	0.83	0.77	0.89	0.90	0.90	0.85
		RSC-Net-10	0.52	0.58	0.52	0.78	0.74	0.87	0.89	0.89	0.77
		U-Net-10	0.53	0.56	0.56	0.8	0.70	0.89	0.90	0.90	0.78
		Cloud-GAN	0.31	0.39	0.33	0.57	/	/	/	/	/
		AHF	0.33	0.51	0.48	0.75	/	/	/	/	/
	PSNR	MRSCP	0.20	0.25	0.26	0.69	/	/	/	/	/
		CR-MSS	22.64	18.66	21.02	20.84	19.16	21.32	21.35	21.55	19.12
		RSC-Net-10	19.06	19.69	17.82	20.79	19.74	19.45	19.56	19.36	19.78
		U-Net-10	18.66	18.69	20.14	19.73	18.17	20.20	20.48	20.29	17.85
		Cloud-GAN	11.26	13.82	13.34	11.33	/	/	/	/	/
		AHF	21.63	13.30	17.91	21.05	/	/	/	/	/
T9 (Figure 8)	SSIM	MRSCP	18.63	16.08	18.27	12.90	/	/	/	/	/
		CR-MSS	0.62	0.57	0.66	0.68	0.59	0.63	0.65	0.65	0.55
		RSC-Net-10	0.56	0.60	0.69	0.67	0.60	0.62	0.63	0.63	0.55
		U-Net-10	0.59	0.56	0.65	0.68	0.58	0.62	0.64	0.64	0.56
		Cloud-GAN	0.34	0.33	0.42	0.31	/	/	/	/	/
		AHF	0.52	0.38	0.59	0.79	/	/	/	/	/
	PSNR	MRSCP	0.52	0.43	0.61	0.40	/	/	/	/	/
		CR-MSS	22.64	18.66	21.02	20.84	19.16	21.32	21.35	21.55	19.12
		RSC-Net-10	19.06	19.69	17.82	20.79	19.74	19.45	19.56	19.36	19.78
		U-Net-10	18.66	18.69	20.14	19.73	18.17	20.20	20.48	20.29	17.85
		Cloud-GAN	11.26	13.82	13.34	11.33	/	/	/	/	/
		AHF	21.63	13.30	17.91	21.05	/	/	/	/	/

Table 5 shows the averaged PSNR and SSIM values on each band over all testing samples. It can be seen that CR-MSS outperforms the compared methods on SSIM values in all bands as well as PSNR values in bands 3/8, and performs almost the same as the best methods on PSNR in bands 2/4 (U-Net-10 and RSC-Net-10, respectively). The SSIM values illustrate that CR-MSS can preserve more structure information than compared methods. Since both U-Net-10 and RSC-Net-10 are end-to-end methods, they achieve good results

that are close to the results of CR-MSS. Still, CR-MSS has the advantage of being able to restore background information in all bands, not just Vis/NIR bands.

Table 5. Average PSNR and SSIM values over all eight real testing image pairs from Table 2 (3782 testing samples). Best results are marked in bold.

Index	Method	B2	B3	B4	B8	B5	B6	B7	B8A	B11	B12
PSNR	CR-MSS	19.53	18.49	20.29	21.09	19.30	20.70	21.03	21.26	21.62	21.98
	RSC-Net-10	18.11	17.97	19.92	20.68	20.09	20.90	21.03	21.32	21.34	21.41
	U-Net-10	18.68	18.09	19.87	20.68	19.39	20.68	20.97	21.11	20.5	20.79
	Cloud-GAN	15.63	15.23	16.22	15.90	/	/	/	/	/	/
	AHF	14.72	14.62	16.90	19.22	/	/	/	/	/	/
	MRSCP	15.13	14.94	16.18	13.15	/	/	/	/	/	/
SSIM	CR-MSS	0.69	0.71	0.77	0.81	0.80	0.82	0.83	0.84	0.85	0.84
	RSC-Net-10	0.66	0.70	0.76	0.80	0.79	0.82	0.83	0.83	0.83	0.84
	U-Net-10	0.67	0.69	0.74	0.79	0.79	0.82	0.83	0.83	0.83	0.82
	Cloud-GAN	0.47	0.49	0.53	0.51	/	/	/	/	/	/
	AHF	0.49	0.54	0.67	0.77	/	/	/	/	/	/
	MRSCP	0.47	0.51	0.62	0.55	/	/	/	/	/	/

U-Net-10 extracts the features at different levels by max/average pooling and combining these features with skip connections. The different level features are fully used by the skip connection, but the neighborhood pixel information is altered in the pooling operation. This is probably the reason why U-Net-10 results are worse than CR-MSS. RSC-Net-10 uses symmetrical concatenations to preserve the information in cloud-free regions. Since it does not down-sample the feature maps, the receptive field of RSC-Net-10 is very small, which means the information from long distance neighborhood pixels cannot be used to restore background information. Unlike U-Net-10 and RSC-Net-10, CR-MSS down-samples the feature maps by convolutional layers with stride = 2, thus preserving neighborhood pixel information and keeping a large receptive field. Although Cloud-GAN does not need paired cloud and cloud-free images, which are very time-saving for the generation of training dataset, Cloud-GAN performs the worst on SSIM value among all methods. This may be because it not only translates a cloudy image into a cloud-free image, but also preserves some texture features from the cloudy image that are used to restore the translated cloud-free image back to a cloudy image—and by this operation, the retained texture features include cloudy texture features, or the training of the generators, ending in a poor location of the optimization landscape. The instability of training adversarial networks also makes it difficult to generate realistic cloud-free images. AHF can filter some cloud components, but the low frequency components in background may be filtered out as well. MSRCF estimates the incident light by Gaussian filtering, which can help to reduce the influence of light but will not help much to remove thin clouds. This is the reason why AHF performs better than MSRCF in most cases on thin cloud removal. Because clouds have the least effects in SWIR bands, and VRE bands are less influenced by clouds than VIS bands, more background information is contained in these bands. By taking VRE/SWIR bands into consideration, CR-MSS can restore more information in Vis/NIR bands. This is another reason why CR-MSS can always obtain better performance on SSIM than compared methods.

Table 6 shows the computing time for all methods. The deep learning base methods are much more efficient than traditional methods. MSRCF takes the longest time, because the Gaussian filtering is very time-consuming, especially in such large images as Sentinel-2. RSC-Net-10 was the fastest method, only outputting 32 feature maps in each convolutional and deconvolutional layer, which reduces a lot of calculations. Although CR-MSS

had a longer computing time than compared deep learning-based methods, the difference is marginal, and CR-MSS obtained the best performance among all methods in most cases.

Table 6. Computing time of different cloud removal methods (in minutes and seconds on a Sentinel-2 image with size 10980×10980 pixels. Best results are marked in bold.

Method	CR-MSS	RSC-Net-10	U-Net-10	Cloud-GAN	AHF	MSRCP
Time	1 min 29 s	0 min 47 s	1 min 19 s	1 min 8 s	5 min 16 s	37 min 24 s

3.2. Influence of the Temporal Shift between Images

As mentioned in Section 2.2, the temporal shift between paired cloud and cloud-free images is 10 days in all experimental data. Table 7 shows the averaged PSNR and SSIM values of all methods on testing samples of different land covers. The highest values are marked in bold, and the symbol ‘/’ is used to represent the band that is not included in the training and testing. We can see that CR-MSS performs the best for SSIM values on urban areas, vegetation areas, and bare land on bands 2/3/4/8, except on band 2 for vegetation, which is slightly better for RSC-Net-10. It can be seen that on vegetation areas, CR-MSS has five PSNR and three SSIM values that are lower than those of RSC-Net-10 or U-Net-10, while only 2 PSNR and 1 SSIM values are lower on Urban areas and four PSNR and one SSIM values are lower on bare land. This is because during the same temporal shift of 10 days between training and testing images, the spectral information in Vis/NIR/VRE/SWIR bands over vegetation areas changes faster than over the two other land cover areas. With VRE/SWIR bands as output, the reference images change more in CR-MSS than baseline methods that only process Vis/NIR bands. Therefore, the temporal shift on vegetation areas influence the performance of CR-MSS more than baseline methods. However, it is worth noticing that CR-MSS obtains competitive SSIM values on vegetation areas, because SSIM contains the properties of the object structure that will not change much in vegetation areas within a short temporal shift. It can also be seen that CR-MSS obtains the best performance on SSIM values on urban and bare land areas on nine out of 10 bands, because the influences of temporal shift on the properties of the object structure are much less than on vegetation areas in the same temporal shift.

Table 7. Average PSNR and SSIM values over different land covers and all eight real testing image pairs from Table 2 (3782 testing samples). Best results are marked in bold.

Landcover	Index	Method	B2	B3	B4	B8	B5	B6	B7	B8A	B11	B12
Urban (1st, 2nd, 5th)	PSNR	CR-MSS	17.23	16.93	19.08	21.02	20.74	19.02	20.84	20.90	21.20	21.65
		RSC-Net-10	15.89	16.31	18.31	20.43	17.35	19.96	20.55	20.92	20.84	20.98
		U-Net-10	16.28	16.71	18.56	20.26	18.49	20.32	20.60	20.54	20.33	20.36
		Cloud-GAN	13.58	13.64	14.37	16.60	/	/	/	/	/	/
		AHF	13.06	13.78	15.68	18.01	/	/	/	/	/	/
		MRSCP	13.00	13.83	14.30	14.55	/	/	/	/	/	/
	SSIM	CR-MSS	0.67	0.69	0.76	0.84	0.80	0.81	0.83	0.84	0.86	0.86
		RSC-Net-10	0.67	0.65	0.75	0.80	0.78	0.82	0.83	0.84	0.85	0.85
		U-Net-10	0.67	0.64	0.74	0.78	0.79	0.81	0.82	0.83	0.84	0.84
		Cloud-GAN	0.45	0.48	0.52	0.50	/	/	/	/	/	/

Vegetation (6th, 7th, 8th)	PSNR	AHF	0.50	0.56	0.66	0.74	/	/	/	/	/	/
		MRSCP	0.46	0.53	0.63	0.64	/	/	/	/	/	/
		CR-MSS	21.40	19.30	21.50	21.96	22.08	19.59	21.55	22.42	22.77	23.64
		RSC-Net-10	19.08	19.10	21.09	22.78	21.78	23.12	23.25	23.84	22.74	22.40
		U-Net-10	18.83	19.88	20.70	21.92	19.83	21.86	22.27	22.78	21.60	21.97
		Cloud-GAN	16.77	15.78	17.03	16.43	/	/	/	/	/	/
	SSIM	AHF	16.83	15.48	18.09	21.75	/	/	/	/	/	/
		MRSCP	17.00	15.17	17.84	11.95	/	/	/	/	/	/
		CR-MSS	0.68	0.70	0.76	0.86	0.83	0.81	0.85	0.87	0.87	0.86
		RSC-Net-10	0.70	0.64	0.76	0.83	0.81	0.85	0.86	0.87	0.87	0.86
		U-Net-10	0.69	0.67	0.74	0.82	0.80	0.85	0.86	0.87	0.84	0.83
		Cloud-GAN	0.45	0.46	0.52	0.50	/	/	/	/	/	/
	PSNR	AHF	0.49	0.54	0.67	0.82	/	/	/	/	/	/
		MRSCP	0.49	0.51	0.63	0.47	/	/	/	/	/	/
		CR-MSS	18.50	18.45	19.10	20.07	20.08	19.32	20.01	20.35	21.10	20.49
		RSC-Net-10	18.43	17.85	19.13	17.55	19.56	18.62	18.19	17.78	20.00	20.52
		U-Net-10	18.65	18.09	19.03	19.48	19.54	20.15	19.86	19.76	19.51	19.90
		Cloud-GAN	14.37	14.94	15.32	14.71	/	/	/	/	/	/
	SSIM	AHF	12.73	13.60	14.28	15.45	/	/	/	/	/	/
		MRSCP	14.75	15.42	16.06	14.31	/	/	/	/	/	/
		CR-MSS	0.71	0.73	0.75	0.75	0.74	0.76	0.75	0.76	0.80	0.78
		RSC-Net-10	0.70	0.68	0.72	0.71	0.75	0.73	0.73	0.74	0.79	0.78
		U-Net-10	0.71	0.68	0.71	0.72	0.75	0.74	0.74	0.75	0.78	0.77
		Cloud-GAN	0.52	0.54	0.56	0.52	/	/	/	/	/	/
Bare land (3rd, 4th)	PSNR	AHF	0.48	0.53	0.60	0.67	/	/	/	/	/	/
		MRSCP	0.50	0.54	0.60	0.59	/	/	/	/	/	/
		CR-MSS	0.50	0.54	0.60	0.59	/	/	/	/	/	/

3.3. Influence of VRE/SWIR Bands

To analyze the influence of the VRE/SWIR bands, we conducted two other experiments with or without VRE/SWIR bands as inputs, and without VRE/SWIR bands as outputs. The experiment with Vis/NIR/VRE/SWIR bands as inputs and Vis/NIR bands as outputs is called CR-MSS-10-4. The experiment with Vis/NIR bands as inputs and outputs is called CR-MSS-4-4. CR-MSS-10-4 has Vis/NIR-In and VRE/SWIR-In in the input branches, while CR-MSS-4-4 only has Vis/NIR-In branch. Both CR-MSS-10-4 and CR-MSS-4-4 only have Vis/NIR-Out branch.

Figures 9–12 show cloud removal results of CR-MSS, CR-MSS-10-4 and CR-MSS-4-4 on four samples from the fifth, sixth, seventh, and eighth real testing pairs. The first and second rows show the cloud contaminated image and the corresponding cloud-free image. The following rows show the results of CR-MSS, CR-MSS-10-4, and CR-MSS-4-4,

respectively. It can be seen that results on Vis/NIR bands are quite good visually. However, as shown in Figure 10, the sample (from the sixth testing image pair) contains unevenly distributed clouds. Thus, the cloud effect in this sample is not completely removed by CR-MSS-10-4 and CR-MSS-4-4, while CR-MSS removes clouds in this sample on all experimental bands.

Figure 12 shows a sample (from the eighth testing image pair) affected by cloud shadows. It can be seen that the effect of cloud shadow is removed by CR-MSS, CR-MSS-10-4, and CR-MSS-4-4. The results on other four samples from the first, second, third, and fourth testing pairs are presented in Figure 13. The first two columns are the cloud contaminated image and corresponding cloud-free image. The last three columns are the results of CR-MSS, CR-MSS-10-4, and CR-MSS-4-4. It can be seen that most of the clouds over different land cover types can be removed by all CR-MSS-based methods. It can be seen that CR-MSS-based methods try to restore the original background information in the cloudy image, rather than directly transferring the cloudy image into the reference image 10 days apart from the input cloud image.

Table 8 shows corresponding PSNR and SSIM values of Figures 9–12. The average PSNR and SSIM values of real samples in Figure 13 and all 3782 real testing samples are listed in Tables 9 and 10, respectively. From Tables 8–10, we can see that the performances of CR-MSS are better than CR-MSS-10-4 and CR-MSS-4-4 most of the time. However, the PSNR and SSIM values of all CR-MSS-based methods are similar. Although CR-MSS-10-4 has the same input bands as CR-MSS, CR-MSS obtains better results on most bands. This is because the output of CR-MSS contains VRE/SWIR bands, which are not included in the output of CR-MSS-10-4, and the restoration of VRE/SWIR bands can add more supervision information on the training of CR-MSS than the restoration of Vis/NIR bands. CR-MSS-4-4 with only Vis/NIR bands as inputs and outputs performs the worst among all CR-MSS-based methods. This is because a lot of background information is damaged by clouds on visible bands. Although NIR band is less affected by clouds than Vis bands, the information it contains is limited and less than bands 8A/11/12 combined. Thus, CR-MSS and CR-MSS-10-4 obtain slightly better results than CR-MSS-4-4.

Table 8. Corresponding PSNR and SSIM values of four real samples in Figure 9–12. Best results are marked in bold.

Image	Index	Method	B2	B3	B4	B8	B5	B6	B7	B8A	B11	B12
Figure 9	PSNR	CR-MSS	17.40	17.07	17.95	20.74	21.21	17.45	19.73	21.98	19.30	18.00
		CR-MSS-10-4	15.12	13.91	16.06	19.57	/	/	/	/	/	/
		CR-MSS-4-4	14.11	13.72	14.80	20.16	/	/	/	/	/	/
	SSIM	CR-MSS	0.67	0.69	0.75	0.88	0.81	0.83	0.86	0.89	0.87	0.84
		CR-MSS-10-4	0.64	0.64	0.74	0.82	/	/	/	/	/	/
		CR-MSS-4-4	0.62	0.63	0.71	0.82	/	/	/	/	/	/
Figure 10	PSNR	CR-MSS	21.28	19.24	21.65	22.87	22.62	19.41	22.15	23.12	26.39	25.03
		CR-MSS-10-4	19.83	18.71	20.18	22.46	/	/	/	/	/	/
		CR-MSS-4-4	19.58	18.38	20.11	20.91	/	/	/	/	/	/
	SSIM	CR-MSS	0.70	0.69	0.81	0.88	0.88	0.79	0.87	0.88	0.92	0.92
		CR-MSS-10-4	0.69	0.69	0.81	0.88	/	/	/	/	/	/
		CR-MSS-4-4	0.69	0.69	0.79	0.87	/	/	/	/	/	/
Figure 11	PSNR	CR-MSS	23.56	21.75	24.84	24.24	23.54	22.89	23.15	24.72	23.08	20.41
		CR-MSS-10-4	26.20	23.51	24.91	23.08	/	/	/	/	/	/
		CR-MSS-4-4	27.71	25.19	22.63	23.15	/	/	/	/	/	/
	SSIM	CR-MSS	0.73	0.72	0.76	0.91	0.83	0.85	0.90	0.91	0.91	0.88
		CR-MSS-10-4	0.75	0.73	0.78	0.81	/	/	/	/	/	/
		CR-MSS-4-4	0.77	0.73	0.75	0.82	/	/	/	/	/	/
Figure 12	PSNR	CR-MSS	19.83	21.41	21.91	22.59	22.09	22.56	22.44	22.60	18.39	21.72
		CR-MSS-10-4	21.68	21.45	20.65	22.07	/	/	/	/	/	/

	CR-MSS-4-4	18.52	21.31	21.63	21.11	/	/	/	/	/	/
	CR-MSS	0.65	0.67	0.78	0.78	0.75	0.73	0.77	0.78	0.79	0.84
SSIM	CR-MSS-10-4	0.70	0.68	0.77	0.77	/	/	/	/	/	/
	CR-MSS-4-4	0.65	0.70	0.78	0.78	/	/	/	/	/	/

Table 9. Average PSNR and SSIM values of four real samples in Figure 13. Best results are marked in bold.

Index	Method	B2	B3	B4	B8	B5	B6	B7	B8A	B11	B12
PSNR	CR-MSS	19.85	19.83	22.18	23.50	23.29	22.22	23.93	23.56	23.36	24.77
	CR-MSS-10-4	18.87	18.58	20.42	22.87	/	/	/	/	/	/
	CR-MSS-4-4	18.89	18.93	20.23	23.23	/	/	/	/	/	/
SSIM	CR-MSS	0.73	0.74	0.77	0.85	0.80	0.84	0.85	0.86	0.88	0.86
	CR-MSS-10-4	0.73	0.75	0.77	0.78	/	/	/	/	/	/
	CR-MSS-4-4	0.70	0.71	0.73	0.78	/	/	/	/	/	/

Table 10. Average PSNR and SSIM values over all eight real testing image pairs from Table 2 (3782 testing samples). Best results are marked in bold.

Index	Method	B2	B3	B4	B8	B5	B6	B7	B8A	B11	B12
PSNR	CR-MSS	19.53	18.49	20.29	21.09	19/30	20.70	21.03	21.26	21.62	21.98
	CR-MSS-10-4	19.50	18.47	20.34	20.86	/	/	/	/	/	/
	CR-MSS-4-4	19.43	18.30	20.03	20.56	/	/	/	/	/	/
SSIM	CR-MSS	0.69	0.71	0.77	0.81	0.80	0.82	0.83	0.84	0.85	0.84
	CR-MSS-10-4	0.69	0.71	0.77	0.80	/	/	/	/	/	/
	CR-MSS-4-4	0.69	0.71	0.76	0.79	/	/	/	/	/	/

As the results of different methods and different number of input bands show, CR-MSS can achieve the best performance in these experiments on most bands. SSIM values of CR-MSS on VRE/NIR/SWIR bands are more acceptable than those on visible bands. This means that the cloud effect is less in bands VRE/NIR/SWIR than visible bands, which is the reason why CR-MSS is designed to take multi-spectral images as input. With more input and output bands, CR-MSS can make full use of the spectral information to restore cloud contaminated images.

Comparing results from Tables 4 and 10, it can be seen how well CR-MSS based methods perform on average over the eight testing image pairs compared to baseline methods when using the same input and output bands. For every Vis/NIR band taken separately, CR-MSS-4-4 has a PSNR slightly lower but close to each best result among the baseline methods. As it comes to SSIM values, CR-MSS-4-4 outperforms every baseline method on every Vis/NIR band. This demonstrates that, even when restricted to visible and NIR bands, the proposed CR-MSS method performs similarly well or better than baseline methods, depending on the measure. One of the main advantages of CR-MSS method remains its capability to handle more than visible and NIR bands seamlessly, as well as the ability to further improve results on visible and NIR bands when adding other bands as inputs.

As can be seen in Figures 4–13 and Tables 3, 4, 7, and 8, all methods perform best on vegetation areas under thin clouds, but worst on urban areas under thin clouds. This is because urban area has more textures than vegetation area, and the details of complex textures in images are very difficult to restore in many image processing tasks.

3.4. Spectral Preservation on Simulated Data

As explained in Section 3.1, we applied all methods to simulated cloud images and then computed NRMSE values between cloud removal results and real cloud-free images. The average NRMSE values in each band over all eight simulated cloud testing samples are shown in Table 11. The best values are marked in bold. The lower the NRMSE value,

the more spectral information is preserved. It can be seen that the lowest NRMSE values for bands 3/4/5/6/11/12 fall in CR-MSS based methods, while U-Net-10 performs better for band 2, and RSC-Net performs better for bands 7/8/8A. Cloud-GAN obtains the worst performance on spectral preservation; this may be because Cloud-GAN is a cycle network designed for image style transformation, which preserves more texture information than spectral information.

Table 11. Average NRMSE values over all eight simulated testing image pairs (3782 simulated testing samples). Best results are marked in bold.

Index	Method	B2	B3	B4	B8	B5	B6	B7	B8A	B11	B12
NRMSE	CR-MSS	0.5697	0.2830	0.2382	0.0190	0.0767	0.0298	0.0249	0.0161	0.0123	0.0221
	CR-MSS-10-4	0.5873	0.2840	0.2182	0.0194	/	/	/	/	/	/
	CR-MSS-4-4	0.4762	0.2335	0.2408	0.0208	/	/	/	/	/	/
	RSC-Net-10	0.6244	0.3216	0.3114	0.0177	0.1008	0.0325	0.0239	0.0101	0.0163	0.0337
	U-Net-10	0.4279	0.2621	0.2210	0.0249	0.0938	0.0352	0.0316	0.0220	0.0289	0.0594
	Cloud-GAN	1.3156	0.5831	1.1778	0.0785	/	/	/	/	/	/
	AHF	0.7123	0.2666	0.3874	0.0952	/	/	/	/	/	/
	MSRCP	1.1203	0.4694	0.8262	0.0615	/	/	/	/	/	/

4. Discussion

From the experimental results on the real image dataset, we can see that most of the time, CR-MSS performs the best qualitatively and quantitatively. This is because CR-MSS takes VRE/SWIR bands into consideration when removing cloud in Vis/NIR bands, which is not available in other deep learning-based cloud removal methods in remote sensing images, e.g., Cloud-GAN [64] and CGAN [65] only process RGB bands. Some methods such as Qin et al. [61] and RSC-Net-10 [62] take SWIR bands in Landsat-8 images as input when removing cloud, but it can only handle the images at the same spatial resolution. Although VRE/SWIR bands have lower spatial resolution than Vis/NIR bands in Sentinel-2A images, they are less influenced by cloud than Vis/NIR bands. Similar to super resolution reconstruction, CR-MSS can make full use of the less influenced low spatial resolution VRE/SWIR bands to reconstruct the missing information in much influenced high spatial resolution Vis/NIR bands. Therefore, CR-MSS can also be considered as a combination of cloud removal and super resolution reconstruction.

Experimental results on the simulated dataset show CR-MSS based methods can preserve more spectral information than baseline methods on most bands in simulated cloud images. We understand this simulated cloud dataset is not perfect, and these results should be taken only for their comparative value, not as an indication of how the proposed method would perform on a real dataset. This was just a complementary experiment to quantitatively assess spectral preservation in a way the real cloud dataset could not.

In Meraner et al. [49], Sentinel-2 low spatial resolution bands were first rescaled to 10 m resolution and then processed with Vis/NIR bands and Sentinel-1 SAR images. A hand designed rescaling algorithm was adopted in Meraner et al. [49] to rescale all data to the same resolution. Hand designed rescaling algorithms such as Nearest neighbor, Bilinear, and Cubic convolution have at least one inevitable drawback: the parameters in the algorithms are not optimal for specific tasks. This drawback can be avoided by training the convolution neural network to learn optimal parameters for specific tasks such as cloud removal automatically. Therefore, convolutional layers were adopted to handle multi-spectral bands at different spatial resolution in CR-MSS, and up-sampling low spatial resolution bands to high resolution is not necessary anymore. In addition, up-sampling low spatial resolution bands to high resolution will not produce more useful spatial information when real high resolution bands are already available and introduce more spectral information.

Since single image based cloud removal methods can hardly remove the influence of clouds completely, many multi-temporal based methods have been proposed to solve this problem. Multi-temporal cloud removal methods usually include three steps: first, select a time series images in the same region as the experimental data; then the cloud regions in each image are detected; finally, the cloud-free regions are mosaicked to get a completely cloud-free image. If the spectral information of each cloud-free regions in the mosaicked image are not consistent with each other, radiometric harmonization across the temporal stack will be conducted to solve this problem. Although the components in the mosaicked cloud-free image are completely cloud-free, there are still some limitations to the multi-temporal based methods. (1) Accurate cloud masks are necessary for each image in the multi-temporal stack; those masks are difficult to obtain, especially for thin clouds, and any labelling error compounds across the temporal stack. (2) Time series images used in the methods are acquired at different times, which means the ground information in the mosaicked cloud-free image is not consistent in time. (3) The longest time period of the time series images depends on the revisit time of the satellite and weather conditions: the longer the time period, the less meaningful the result is. (4) Multi-temporal cloud removal methods usually require more input data and processing than single-image methods. They are naturally justified as part of a whole multi-temporal chain including, e.g., multi-temporal classification or time series analysis. However, if the end application can be achieved with single image (e.g., classification) or dual images (bi-temporal change detection), often single-image cloud removal methods are still preferred in practice for their ease of use.

Although there are many complex and effective deep learning-based networks in the image processing field, U-Net architecture is the most widely adopted in image inpainting, image de-hazing, and image super resolution due to its simplicity and effectiveness. In view of this, CR-MSS adopted U-Net as the main architecture, with some modifications, so that it can be used for handling multispectral satellite images at different spatial resolutions. As a result, CR-MSS remains a relatively simple deep learning model that is easy to understand, implement, and deploy in operational conditions. CR-MSS was designed for thin cloud removal; however, it can also be easily applied to other remote sensing image processing tasks such as cloud detection and semantic segmentation, due to the simplicity and effectivity inherited from U-Net.

5. Conclusions

In this paper, we proposed a thin cloud removal method, CR-MSS, which used the VRE/SWIR spectral information to better restore the original background information contaminated by thin clouds in Vis/NIR bands in Sentinel-2A images. CR-MSS takes Vis/NIR and VRE/SWIR bands of Sentinel-2A images as inputs, grouped into two images at 10 m resolution and 20 m resolution, and fed into separated input/output branches without rescaling low spatial bands to high resolution by hand designed rescaling algorithm. Compared with baseline methods on real and simulated testing datasets, CR-MSS can achieve better overall performance for texture and spectral information preservation both qualitatively and quantitatively. Ablation experiments were also conducted to analyze the influence of VRE/SWIR bands on the performance of thin cloud removal in Sentinel-2A images. The experiment results show that even VRE/SWIR bands have lower spatial resolution than Vis/NIR bands; their spectral information can help restore the cloud contaminated pixels in Vis/NIR bands. Additionally, adding VRE/SWIR bands into input and output, CR-MSS performs better than only adding VRE/SWIR bands as input. The quantitative results derived from the simulation experiment showed CR-MSS based methods can preserve more spectral information than all baseline methods in bands 2/3/4/8.

In future work, we will improve and apply CR-MSS thin cloud removal to other optical satellites that acquire multispectral images with bands at different spatial resolutions (such as CBERS-04 and ZY-1 02D). We will also modify CR-MSS so that it can be applied to other remote sensing image processing tasks such as cloud detection and semantic

segmentation. The combination of generative adversarial networks and CR-MSS will also be taken into consideration to remove clouds in a semi-supervised way.

Author Contributions: conceptualization: J.L.; methodology: J.L.; visualization: Z.L.; funding acquisition: Z.H.; validation: J.L.; formal analysis: J.L., Z.W., Z.H., and M.M.; supervision: Z.W., M.M., and Z.H.; investigation: Y.W. and Z.L.; writing—original draft preparation: J.L. and Z.W.; writing—review and editing: J.L., M.M., and Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was jointly supported by the National Key R&D Program of China (No. 2017YFC0506200), by the National Natural Science Foundation of China (NSFC) (No. 41871227 and No.41501369), by the Natural Science Foundation of Guangdong (No.2020A1515010678), by the Basic Research Program of Shenzhen (No. JCYJ20190808122405692), and by the Academy of Finland through the Finnish Flagship Programme FCAI: Finnish Center for Artificial Intelligence (Grant No. 320183).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: [<https://github.com/Neoooolee/WHUS2-CR>].

Acknowledgments: The authors are grateful for the Sentinel-2 data services from the Copernicus Open Access Hub. They also thank J.Z. (Jiaqi Zhang) and G.S. (Shenguan Ting) for making the dataset. The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Foley, J.A.; DeFries, R.; Asner, G.P.; Barford, C.; Bonan, G.; Carpenter, S.R.; Chapin, F.S.; Coe, M.T.; Daily, G.C.; Gibbs, H.K.; et al. Global Consequences of Land Use. *Science* **2005**, *309*, 570–574, doi:10.1126/science.1111772.
2. Voltersen, M.; Berger, C.; Hese, S.; Schmullius, C. Object-based land cover mapping and comprehensive feature calculation for an automated derivation of urban structure types at block level. *Remote Sens. Environ.* **2014**, *154*, 192–201.
3. Rogan, J.; Franklin, J.; Roberts, D.A. A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery. *Remote Sens. Environ.* **2002**, *80*, 143–156, doi:10.1016/s0034-4257(01)00296-6.
4. Fisher, A.; Flood, N.; Danaher, T. Comparing Landsat water index methods for automated water classification in eastern Australia. *Remote Sens. Environ.* **2016**, *175*, 167–182.
5. Mueller, N.; Lewis, A.C.; Roberts, D.A.; Ring, S.; Melrose, R.; Sixsmith, J.; Lymburner, L.; McIntyre, A.; Tan, P.; Curnow, S.; et al. Water observations from space: Mapping surface water from 25 years of Landsat imagery across Australia. *Remote Sens. Environ.* **2016**, *174*, 341–352, doi:10.1016/j.rse.2015.11.003.
6. Zhang, Y.; Rossow, W.B.; Lacis, A.A.; Oinas, V.; Mishchenko, M.I. Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. *J. Geophys. Res. Space Phys.* **2004**, *109*, doi:10.1029/2003jd004457.
7. Liou, K.N.; Davies, R. Radiation and Cloud Processes in the Atmosphere. *Phys. Today* **1993**, *46*, 66–67, doi:10.1063/1.2809044.
8. Parmes, E.; Rauste, Y.; Molinier, M.; Andersson, K.; Seitsonen, L. Automatic Cloud and Shadow Detection in Optical Satellite Imagery Without Using Thermal Bands—Application to Suomi NPP VIIRS Images over Fennoscandia. *Remote Sens.* **2017**, *9*, 806.
9. Roy, D.; Ju, J.; Lewis, P.; Schaaf, C.; Gao, F.; Hansen, M.; Lindquist, E. Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. *Remote Sens. Environ.* **2008**, *112*, 3112–3130, doi:10.1016/j.rse.2008.03.009.
10. Lv, H.; Wang, Y.; Shen, Y. An empirical and radiative transfer model based algorithm to remove thin clouds in visible bands. *Remote Sens. Environ.* **2016**, *179*, 183–195, doi:10.1016/j.rse.2016.03.034.
11. Zhu, X.; Gao, F.; Liu, D.; Chen, J. A Modified Neighborhood Similar Pixel Interpolator Approach for Removing Thick Clouds in Landsat Images. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 521–525, doi:10.1109/lgrs.2011.2173290.
12. Cheng, Q.; Shen, H.; Zhang, L.; Yuan, Q.; Zeng, C. Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal MRF model. *ISPRS J. Photogram. Remote Sens.* **2014**, *92*, 54–68, doi:10.1016/j.isprsjprs.2014.02.015.
13. Menaka, E.; Kumar, S.S.; Bharathi, M. Cloud removal using efficient cloud detection and removal algorithm for high-resolution satellite imagery. *Int. J. Comput. Appl. Technol.* **2015**, *51*, 54, doi:10.1504/ijcat.2015.068400.
14. Tseng, D.-C.; Tseng, H.-T.; Chien, C.-L. Automatic cloud removal from multi-temporal SPOT images. *Appl. Math. Comput.* **2008**, *205*, 584–600, doi:10.1016/j.amc.2008.05.050.

15. Zhang, Y.; Wen, F.; Gao, Z.; Ling, X. A Coarse-to-Fine Framework for Cloud Removal in Remote Sensing Image Sequence. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5963–5974, doi:10.1109/tgrs.2019.2903594.
16. Cerra, D.; Bieniarz, J.; Beyer, F.; Tian, J.; Müller, R.; Jarmer, T.; Reinartz, P. Cloud Removal in Image Time Series Through Sparse Reconstruction from Random Measurements. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3615–3628, doi:10.1109/jstars.2016.2550084.
17. Lin, C.H.; Tsai, P.H.; Lai, K.H.; Chen, J.Y. Cloud removal from multitemporal satellite images using information cloning. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 232–241.
18. Chen, B.; Huang, B.; Chen, L.; Xu, B. Spatially and temporally weighted regression: A novel method to produce continuous cloud-free Landsat imagery. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 27–37.
19. Li, X.H.; Wang, L.; Cheng, Q.; Wu, P.; Gan, W.; Fang, L. Cloud removal in remote sensing images using nonnegative matrix factorization and error correction. *ISPRS J. Photogramm. Remote Sens.* **2019**, *148*, 103–113.
20. Zhang, Y.; Guindon, B.; Cihlar, J. An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images. *Remote Sens. Environ.* **2002**, *82*, 173–187, doi:10.1016/s0034-4257(02)00034-2.
21. Chen, S.; Chen, X.; Chen, J.; Jia, P.; Cao, X.; Liu, C. An Iterative Haze Optimized Transformation for Automatic Cloud/Haze Detection of Landsat Imagery. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 2682–2694, doi:10.1109/tgrs.2015.2504369.
22. Xu, M.; Pickering, M.; Plaza, A.J.; Jia, X. Thin Cloud Removal Based on Signal Transmission Principles and Spectral Mixture Analysis. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 1659–1669, doi:10.1109/tgrs.2015.2486780.
23. Liu, Z.; Hunt, B. A new approach to removing cloud cover from satellite imagery. *Comput. Vis. Graph. Image Process.* **1984**, *25*, 252–256, doi:10.1016/0734-189x(84)90107-5.
24. Shen, H.; Li, H.; Qian, Y.; Zhang, L.; Yuan, Q. An effective thin cloud removal procedure for visible remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2014**, *96*, 224–235, doi:10.1016/j.isprsjprs.2014.06.011.
25. Liu, J.; Wang, X.; Chen, M.; Liu, S.; Zhou, X.; Shao, Z.; Liu, P. Thin cloud removal from single satellite images. *Opt. Express* **2014**, *22*, 618–632, doi:10.1364/oe.22.000618.
26. Land, E.H. Recent advances in retinex theory. *Vis. Res.* **1986**, *26*, 7–21.
27. Jobson, D.; Rahman, Z.; Woodell, G. Properties and performance of a center/surround retinex. *IEEE Trans. Image Process.* **1997**, *6*, 451–462, doi:10.1109/83.557356.
28. Rahman, Z.-U.; Jobson, D.J.; Woodell, G.A. Multi-Scale Retinex for Color Image Enhancement. In Proceedings of the 3rd IEEE International Conference on Image Processing, Lausanne, Switzerland, 19 September 1996.
29. Jobson, D.; Rahman, Z.; Woodell, G. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Process.* **1997**, *6*, 965–976, doi:10.1109/83.597272.
30. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *1*, 97–115.
31. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *39*, 640–651.
32. Goodfellow, I.J.; Abadie, J.P.; Mirza, M.; Xu, B. Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst.* **2014**, *2*, 2672–2680.
33. Mao, X.; Shen, C.; Yang, Y. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 2802–2810.
34. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
35. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI, Munich, Germany, 5–9 October 2015; pp. 234–241.
36. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36, doi:10.1109/mgrs.2017.2762307.
37. Zhong, L.; Hu, L.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* **2019**, *221*, 430–443, doi:10.1016/j.rse.2018.11.032.
38. Marcos, D.; Volpi, M.; Kellenberger, B.; Tuia, D. Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 96–107, doi:10.1016/j.isprsjprs.2018.01.021.
39. Zhang, C.; Pan, X.; Li, H.; Gardiner, A.; Sargent, I.; Hare, J.; Atkinson, P.M. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS J. Photogramm. Remote Sens.* **2018**, *140*, 133–144, doi:10.1016/j.isprsjprs.2017.07.014.
40. Paoletti, M.E.; Haut, J.M.; Plaza, J.; Plaza, A. Deep learning classifiers for hyperspectral imaging: A review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 279–317, doi:10.1016/j.isprsjprs.2019.09.006.
41. Zhao, W.; Du, S. Spectral-Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4544–4554, doi:10.1109/tgrs.2016.2543748.
42. Pan, B.; Shi, Z.W.; Xu, X. MugNet: Deep learning for hyperspectral image classification using limited samples. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 108–119.
43. Othman, E.; Bazi, Y.; Alajlan, N.; Alhichri, H.; Melgani, F. Using convolutional features and a sparse autoencoder for land-use scene classification. *Int. J. Remote Sens.* **2016**, *37*, 2149–2167, doi:10.1080/01431161.2016.1171928.

44. Dai, X.; Wu, X.; Wang, B.; Zhang, L. Semisupervised Scene Classification for Remote Sensing Images: A Method Based on Convolutional Neural Networks and Ensemble Learning. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 869–873, doi:10.1109/lgrs.2018.2886534.
45. Deng, Z.; Sun, H.; Zhou, S.; Zhao, J.; Lei, L.; Zou, H. Multi-scale object detection in remote sensing imagery with convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 3–22, doi:10.1016/j.isprsjprs.2018.04.003.
46. Schilling, H.; Bulatov, D.; Middelman, W. Object-based detection of vehicles using combined optical and elevation data. *ISPRS J. Photogramm. Remote Sens.* **2018**, *136*, 85–105.
47. Bermudez, J.D.; Happ, P.N.; Feitosa, R.Q.; Oliveira, D.A.B. Synthesis of Multispectral Optical Images From SAR/Optical Multitemporal Data Using Conditional Generative Adversarial Networks. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1220–1224, doi:10.1109/lgrs.2019.2894734.
48. Li, W.B.; Li, Y.; Chan, J.C. Thick Cloud removal with optical and SAR imagery via convolutional mapping deconvolutional network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *99*, 1–15.
49. Meraner, A.; Eel, P.; Zhu, X.X.; Schmit, M. Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 333–346.
50. Gao, J.; Yuan, Q.; Li, J.; Zhang, H.; Su, X. Cloud Removal with Fusion of High Resolution Optical and SAR Images Using Generative Adversarial Networks. *Remote Sens.* **2020**, *12*, 191, doi:10.3390/rs12010191.
51. Wang, X.; Xu, G.; Wang, Y.; Lin, D.; Li, P.; Lin, X. Thin and Thick Cloud Removal on Remote Sensing Image by Conditional Generative Adversarial Network. *IEEE Int. Geosci. Remote Sens. Symp.* **2019**, 921–924, doi:10.1109/IGARSS.2019.8897958.
52. Chai, D.; Newsam, S.; Zhang, H.; Qiu, Y.; Huang, J. Cloud and cloud shadow detection in Landsat imagery based on deep convolutional neural networks. *Remote Sens. Environ.* **2019**, *225*, 307–316, doi:10.1016/j.rse.2019.03.007.
53. Jeppesen, J.H.; Jacobsen, R.H.; Inceoglu, F.; Toftegaard, T.S. A cloud detection algorithm for satellite imagery based on deep learning. *Remote Sens. Environ.* **2019**, *229*, 247–259, doi:10.1016/j.rse.2019.03.039.
54. Li, Z.; Shen, H.; Cheng, Q.; Liu, Y.; You, S.; He, Z. Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 197–212, doi:10.1016/j.isprsjprs.2019.02.017.
55. Wu, Z.; Li, J.; Wang, Y.; Hu, Z.; Molinier, M. Self-Attentive Generative Adversarial Network for Cloud Detection in High Resolution Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1792–1796, doi:10.1109/lgrs.2019.2955071.
56. Dare, P.; Dowman, I. An improved model for automatic feature-based registration of SAR and SPOT images. *ISPRS J. Photogramm. Remote Sens.* **2001**, *56*, 13–28, doi:10.1016/s0924-2716(01)00031-4.
57. Cai, B.; Xu, X.; Jia, K.; Qing, C.; Tao, D. DehazeNet: An End-to-End System for Single Image Haze Removal. *IEEE Trans. Image Process.* **2016**, *25*, 5187–5198, doi:10.1109/tip.2016.2598681.
58. Zhang, H.; Patel, V.M. Densely Connected Pyramid Dehazing Network. *IEEE Conf. Comput. Vis. Pattern Recognit.* **2018**, 3194–3203, doi:10.1109/CVPR.2018.00337.
59. Ren, W.; Ma, L.; Zhang, J.; Pan, J.; Cao, X.; Liu, W.; Yang, M.-H. Gated Fusion Network for Single Image Dehazing. *IEEE Conf. Comput. Vis. Pattern Recognit.* **2018**, 3253–3261, doi:10.1109/CVPR.2018.00343.
60. Chen, D.; He, M.; Fan, Q.; Liao, J.; Zhang, L.; Hou, D.; Yuan, L.; Hua, G. Gated Context Aggregation Network for Image Dehazing and Deraining. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, 7–11 January 2019; pp. 1375–1383.
61. Qin, M.; Xie, F.; Li, W.; Shi, Z.; Zhang, H. Dehazing for Multispectral Remote Sensing Images Based on a Convolutional Neural Network with the Residual Architecture. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1645–1655, doi:10.1109/jstars.2018.2812726.
62. Li, W.B.; Li, Y.; Chen, D.; Chan, J.C. Thin cloud removal with residual symmetrical concatenation network. *ISPRS J. Photogramm. Rem. Sens.* **2019**, *153*, 137–150.
63. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
64. Singh, P.; Komodakis, N. Cloud-Gan: Cloud Removal for Sentinel-2 Imagery Using a Cyclic Consistent Generative Adversarial Networks. *IEEE Int. Geosci. Remote Sens. Symp.* **2018**, 1772–1775, doi:10.1109/IGARSS.2018.8519033.
65. Sun, L.; Zhang, Y.; Chang, X.; Wang, Y.; Xu, J. Cloud-Aware Generative Network: Removing Cloud from Optical Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 691–695, doi:10.1109/lgrs.2019.2928840.
66. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. *Adv. Neural Inform. Process. Syst.* **2017**, *29*, 5998–6008.
67. Li, J.; Wu, Z.; Hu, Z.; Zhang, J.; Li, M.; Mo, L.; Molinier, M. Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 373–389, doi:10.1016/j.isprsjprs.2020.06.021.
68. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324.
69. Defourny, P.; Bontemps, S.; Lamarche, C.; Brockmann, C.; Boettcher, M.; Wevers, J.; Kirches, G.; Santoro, M. Land Cover CCI Product User Guide—Version 2.0. Esa. 2017. Available online: maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf (accessed on 10 April 2017).
70. Wang, Y.-K.; Huang, W.-B. A CUDA-enabled parallel algorithm for accelerating retinex. *J. Real Time Image Process.* **2012**, *9*, 407–425, doi:10.1007/s11554-012-0301-6.

71. Main-Knorn, M.; Pflug, B.; Louis, J.; Debaecker, V.; Müller-Wilm, U.; Gascon, F. Sen2Cor for Sentinel-2. In Proceedings of the Image and Signal Processing for Remote Sensing XXIII, Warsaw, Poland, 11–13 September 2017; Volume 10427.
72. Shen, H.; Wu, J.; Cheng, Q.; Aihemaiti, M.; Zhang, C.; Li, Z. A Spatiotemporal Fusion Based Cloud Removal Method for Remote Sensing Images with Land Cover Changes. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 862–874, doi:10.1109/jstars.2019.2898348.
73. Zheng, J.; Liu, X.-Y.; Wang, X. Single Image Cloud Removal Using U-Net and Generative Adversarial Networks. *IEEE Trans. Geosci. Remote Sens.* **2020**, 1–15, doi:10.1109/TGRS.2020.3027819.
74. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 5–8 May 2015.