

Thin Cloud Removal for Multispectral Remote Sensing Images Using Convolutional Neural Networks Combined With an Imaging Model

Yue Zi, Fengying Xie^{ID}, Member, IEEE, Ning Zhang, Zhiguo Jiang^{ID}, Member, IEEE, Wentao Zhu, and Haopeng Zhang^{ID}, Member, IEEE

Abstract—Multispectral remote sensing images are often degraded by clouds, resulting in the reduced efficiency and accuracy of image interpretation. Thin cloud removal is one of the most important and significant tasks for optical multispectral images. In this article, we propose a novel thin cloud removal method for multispectral images, which is a combination of traditional methods and deep learning methods. First, we adopt U-Net to estimate the reference thin cloud thickness map of the cloudy image. Then, a convolutional neural network named Slope-Net is designed to estimate the thickness coefficient of each band relative to the reference thin cloud thickness map to obtain the thin cloud thickness maps of different bands. Finally, the recovered clear image can be obtained by subtracting the thin cloud thickness maps from the cloudy image according to the traditional thin cloud imaging model. To train U-Net and Slope-Net, a wavelength-dependent thin cloud simulation method is presented to generate a labeled dataset composed of synthetic cloudy images, corresponding clear images, reference thin cloud thickness maps, and thickness coefficients. Qualitative and quantitative comparison experiments are conducted on both synthetic cloudy images and real cloudy images from the Landsat 8 Operational Land Imager. The results indicate that the proposed method can effectively remove thin clouds in multispectral images with various land cover types and maintain good color fidelity.

Index Terms—Convolutional neural network (CNN), multispectral remote sensing images, thin cloud removal, thin cloud thickness map.

I. INTRODUCTION

WITH the development of satellite technologies, remote sensing images are playing a critical role in modern earth observations. During the transmission of electromagnetic waves

Manuscript received November 23, 2020; revised January 25, 2021 and March 14, 2021; accepted March 19, 2021. Date of publication March 23, 2021; date of current version April 15, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFC1510905, in part by the Beijing Natural Science Foundation under Grant 4192032, and in part by the National Natural Science Foundation of China under Grant 61871011 and Grant 61771031. (Corresponding author: Fengying Xie.)

Yue Zi, Fengying Xie, Zhiguo Jiang, and Haopeng Zhang are with the Beijing Advanced Innovation Center for Biomedical Engineering and the Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China (e-mail: ziyyue91@buaa.edu.cn; xfy_73@buaa.edu.cn; jiangzg@buaa.edu.cn; zhanghaopeng@buaa.edu.cn).

Ning Zhang and Wentao Zhu are with the Shanghai Aerospace Electronic Technology Institute, Shanghai 201109, China (e-mail: dzs.zhangning@gmail.com; 337938145@qq.com).

Digital Object Identifier 10.1109/JSTARS.2021.3068166

before the waves are finally received by satellite sensors, the signal is easily degraded by the absorption and scattering of suspended atmospheric particles such as low-altitude haze and fog and medium- to high-altitude clouds [1]. These atmospheric effects reduce the quality of remote sensing images and limit their utilization. Hence, an effective haze and thin cloud removal method for single multispectral remote sensing images is of great importance.

Many methods have been developed to remove haze and thin clouds from remote sensing images, and they can briefly be classified into two main categories [2]: radiative-transmission-model-based methods and image-processing-based methods. Radiative-transmission-model-based methods such as 6S [3], LOWTRAN [4], MODTRAN [5], and ATCOR [6] often yield accurate results; however, these methods are complicated and require detailed parameters of meteorologic and geographic conditions when the data are recorded [2], [7]. Image-processing-based methods treat haze and thin cloud removal as computer vision problems, fully exploit the information in the acquired image, and require no additional parameters, which arouses broader research interests. In this article, we also address this problem with image processing techniques. Since haze and thin clouds exert similar visual influences in the computer vision field, we will generically address this problem as thin cloud removal in this article.

Traditional image-processing-based thin cloud removal methods rely on simplified models or priors. A physical cloud distortion model was first developed in 1977 [8], where Mitchell *et al.* removed the low-frequency cloud noise using homomorphic filtering. Liu *et al.* [9] improved the homomorphic filtering method to improve contrast and retain more details of satellite images. Shen *et al.* [10] adopted a local adaptive filtering algorithm in the frequency domain to remove thin clouds in Landsat ETM+ and GaoFen-1 images. These methods only consider the frequency features and ignore the spatial features of cloudy images, so they behave weakly in more complex scenes. Chavez [11], [12] used a dark-object subtraction technique to remove even atmospheric effects in multispectral images. However, the selection of ideal homogenous dark objects may be difficult, and this method is suitable only for dealing with even thin clouds. This technique was later improved by using a local search of dark objects to calculate a thin cloud thickness map for each band [13],

[14]. To solve the problem that the thin cloud thickness map is easily affected by bright ground objects, Liu *et al.* [15] proposed a ground radiance suppressed thin cloud thickness map to calculate a more precise thin cloud distribution. Xu *et al.* [16] and Zhou and Wang [17] removed thin clouds by calculating the linear regression coefficient between a visible or infrared band and the cirrus band. Markchom and Lipikorn [18] used the dark channel prior (DCP) [19] to estimate the scattering light in HSI color space and performed a subtraction in only the intensity channel to recover clear satellite images. Li *et al.* [20] proposed a two-stage thin cloud removal method based on homomorphic filtering and sphere model improved DCP. Xu *et al.* [21] found that cloudy images have a higher signal-to-noise ratio and developed a thin cloud removal method based on noise-adjusted principal components transformation (CR-NAPCT). These traditional model- or prior-based methods rely heavily on handcrafted features, and their restoration results have low accuracy and robustness for remote sensing images with various ground cover conditions and complex textures.

In recent years, with the strong nonlinear fitting ability of convolutional neural networks (CNNs), CNN-based methods have been widely used to tackle the atmospheric correction problem. Cai *et al.* [22] and Ren *et al.* [23] designed CNNs with different structures to learn the transmission map of a hazy image to recover clear natural scene images. Goncalves *et al.* [24] trained a network that directly learns the mapping from hazy images to clear images based on the powerful Inception-Res-V2 framework. Qin *et al.* [25] designed parallel deep CNNs with multiscale feature extracting blocks to recover clear imagery from Landsat 8 Operational Land Imager (OLI). Enomoto *et al.* [26] and Zhang *et al.* [27] revised the hand-engineered objective function of CNNs and used a learned similarity metric of generative adversarial networks to remove thin clouds in WorldView-2 and Landsat 8 OLI images. Li *et al.* [28] presented an end-to-end deep residual symmetrical concatenation network (RSC-Net) to estimate the clear output image with the cloudy input image. RSC-Net utilizes symmetrical convolutional-deconvolutional concatenations to capture the detailed features in restoration. These methods take full advantage of the automatic learning ability of the CNN to mine the high-level features of cloudy images so that they can deal with cloudy images under complex ground cover conditions. However, these methods generally use synthetic images to train end-to-end networks, and their simulation methods discard some useful prior knowledge, making it difficult to fully simulate real cloudy images. Therefore, these CNN-based methods usually achieve a good result in the simulation image but a common result in the real remote sensing image.

In this article, we propose a new thin cloud removal method based on an additive imaging model, which is a combination of traditional methods and deep learning methods. We first utilize CNNs, which have strong feature extraction ability and nonlinear regression ability, to estimate the reference thin cloud thickness map and thickness coefficients. Then, the traditional thin cloud imaging model is used to remove thin clouds. Our proposed method considers the different thin cloud thickness maps of each band to obtain better thin cloud removal results.

In summary, the proposed method has the following three main contributions.

- 1) A novel thin cloud removal framework based on an additive imaging model is proposed for multispectral remote sensing images, and the proposed framework combines the advantages of traditional methods with deep learning methods.
- 2) U-Net is used to estimate the reference thin cloud thickness map, which can obtain accurate thin cloud information, including both low- to medium-altitude thin clouds and high-altitude cirrus. In addition, a new CNN architecture named Slope-Net is designed to estimate the thickness coefficients, and then, the different thin cloud thickness maps of each band can be obtained.
- 3) A new wavelength-dependent thin cloud simulation method is proposed to generate adequate multispectral cloudy images for training U-Net and Slope-Net.

The remainder of this article is organized as follows. Section II introduces the related work of thin cloud removal. Section III elaborates the proposed method. Section IV describes the thin cloud simulation method for generating the dataset. Experimental results are shown in Section V. Finally, Section VI concludes this article.

II. RELATED WORK

Based on the previous research in [11] and [14], the digital number (DN) value received by the satellite sensor can be described as an additive model

$$x^s = x^g + x^c \quad (1)$$

where x^s is the DN value acquired by the sensor, x^g is the DN value of the ground cover without the influence of clouds, and x^c is the atmospheric effect, which depends on the thickness of the thin cloud. x^c is defined as a thin cloud thickness map, and the clear image can be recovered by subtracting the thin cloud thickness map from the cloudy image.

For multispectral remote sensing images, the thickness of thin clouds usually decreases with increasing wavelength, and the thin cloud thickness map of each band is linearly related [13]. Assuming that a reference thin cloud thickness map x_0^c is given, the proportional coefficient between the thin cloud thickness map of the i th band x_i^c and x_0^c is α_i . α_i is called the thickness coefficient of the i th band, and x_i^c can be expressed as

$$x_i^c = \alpha_i \times x_0^c. \quad (2)$$

Combining (1) and (2), the imaging model for the i th band can be represented as

$$x_i^s = x_i^g + \alpha_i \times x_0^c. \quad (3a)$$

Then, the clear image of the i th band x_i^g can be obtained from

$$x_i^g = x_i^s - \alpha_i \times x_0^c. \quad (3b)$$

Equation (3b) shows that if the reference thin cloud thickness map x_0^c and the thickness coefficient of the i th band α_i are estimated, we can obtain a clear image of the i th band x_i^g .

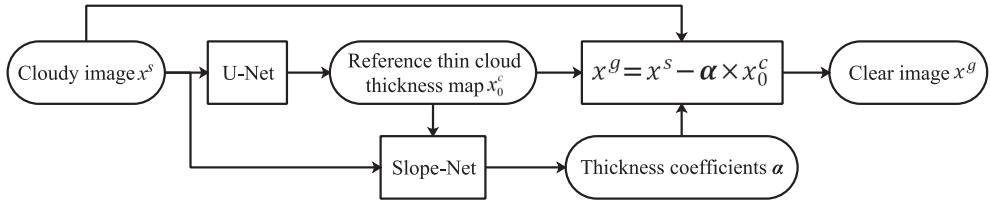


Fig. 1. Flowchart of the proposed thin cloud removal method.

TABLE I
LANDSAT 8 OLI SPECTRAL BANDS

Spectral Band	Wavelength
Band 1 - Coastal	0.433 – 0.453 μm
Band 2 - Blue	0.450 – 0.515 μm
Band 3 - Green	0.525 – 0.600 μm
Band 4 - Red	0.630 – 0.680 μm
Band 5 - Near Infrared	0.845 – 0.885 μm
Band 6 - Short Wavelength Infrared 1	1.560 – 1.660 μm
Band 7 - Short Wavelength Infrared 2	2.100 – 2.300 μm
Band 8 - Panchromatic	0.500 – 0.680 μm
Band 9 - Cirrus	1.360 – 1.390 μm

For the Landsat 8 OLI image (band information is shown in Table I), the cirrus band contains only cirrus information and does not contain ground cover information. Xu *et al.* [16] directly used the cirrus band as the reference thin cloud thickness map, and then, the local homogeneous land cover block, where the DN value of the clear image is approximately a constant, was selected to calculate the thickness coefficients via linear regression between the reference thin cloud thickness map and cloudy image, thereby recovering the clear image from the cloudy image. However, the cirrus band does not contain low- to medium-altitude thin cloud information, which makes this method unable to remove low- to medium-altitude thin clouds from cloudy images. In addition, some variations may exist in the local homogeneous block. Hence, the calculated thickness coefficient lacks robustness, which limits the adaptive ability of this method.

In this article, we present a new thin cloud removal method for Landsat 8 OLI images. The proposed method utilizes the strong feature extraction and nonlinear regression capabilities of CNNs to estimate the reference thin cloud thickness map and the thickness coefficient of each band to remove thin clouds from the cloudy image. Considering that the shortwave infrared bands are only marginally influenced by thin clouds, our thin cloud removal method is performed in the coastal, visible, and near-infrared bands (Bands 1–5).

III. METHODOLOGY

Fig. 1 shows the flowchart of the thin cloud removal method proposed in this article. First, we use U-Net to estimate the reference thin cloud thickness map from the cloudy image. Then, Slope-Net is employed to estimate the thickness coefficients to obtain the thin cloud thickness map of each band. Finally, we can obtain a clear image by subtracting the thin cloud thickness maps from the cloudy image.

A. Reference Thin Cloud Thickness Map Estimation

In this article, U-Net [29] is used to estimate the reference thin cloud thickness map, which is a popular CNN for image transformation. U-Net can obtain accurate thin cloud information, including both low- to medium-altitude thin clouds and high-altitude cirrus, from cloudy images without additional cirrus band information. The architecture of U-Net is shown in Fig. 2. Bands 1–5 of the Landsat 8 OLI image are input into U-Net, and the corresponding reference thin cloud thickness map is output. U-Net is a symmetrical encoder–decoder network with skip connections, where the encoder (left side) is used to capture multiscale semantic features, and the decoder (right side) is used to enable precise localization. The encoder contains five scales, and each scale is composed of two 3×3 convolutional layers and a 2×2 max pooling layer with stride 2 for downsampling. Each convolutional layer is followed by a batch normalization (BN) layer and a rectified linear unit (ReLU) layer. We double the number of feature channels after each max pooling layer. At each scale of the decoder, the feature map is upsampled by a 2×2 upconvolution layer with stride 2, and the number of feature channels is halved. Then, we concatenate two feature maps of the same scale from the encoder and the decoder. Two 3×3 convolution operations are performed, each followed by BN and ReLU activation functions. In the last layer, we use a 1×1 convolution operation to map each 64-D feature vector to the output, and then, the output is normalized to the range of (0, 1) by a sigmoid function.

B. Thickness Coefficient Estimation

According to (3a), the thickness coefficient of the i th band α_i can be calculated via linear regression between the reference thin cloud thickness map x_0^c and the cloudy image of the i th band x_i^s based on the homogenous land cover pixels, whose DN value on the clear image of the i th band x_i^g is approximately constant. However, it is difficult to directly select homogeneous land cover pixels from cloud-contaminated images. In this article, several cloudy images with different land cover backgrounds are selected to research the relationship between x_0^c and x_i^s . Taking Band 2 as an example, Fig. 3 shows the scatter plots of x_0^c versus x_2^s and the associated R^2 , where (a) is the true color cloudy image composite of Bands 2–4, (b) is x_0^c estimated by U-Net, and (c) is the scatter plot of x_0^c versus x_2^s . The first column is a cloudy image with homogeneous land cover. The scatter plot shows that x_0^c and x_2^s show a significant linear correlation, and their R^2 value is 0.9788. In this case, the thickness coefficient α_2 is equal to the slope of the fitting line

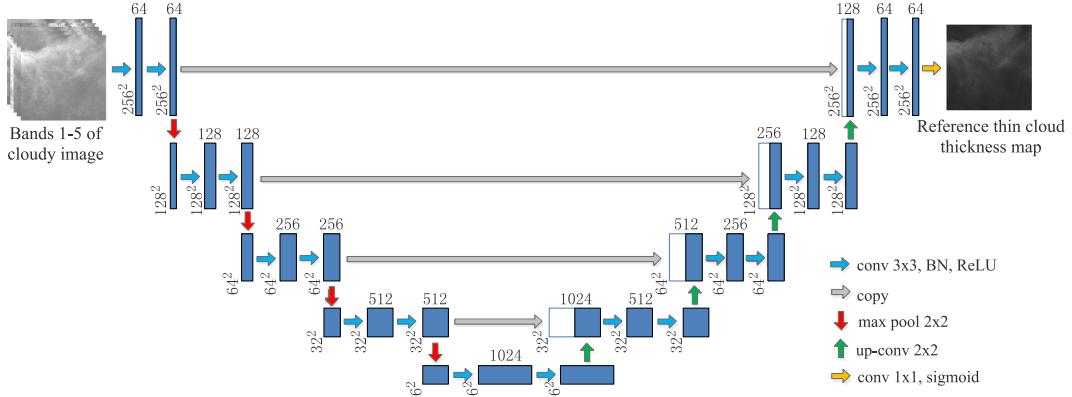
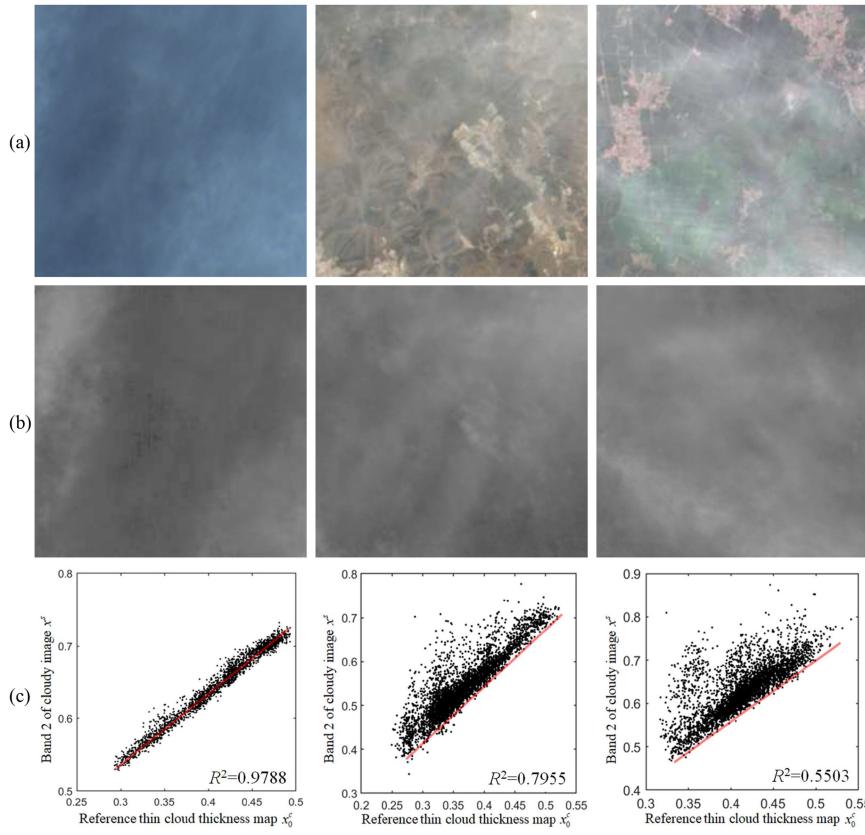


Fig. 2. Architecture of U-Net.

Fig. 3. Relationship between x_0^c and x_0^s . (a) True color cloudy image x_0^s . (b) Reference thin cloud thickness map x_0^c . (c) Scatter plot of x_0^c versus x_2^c .

in the scatter plot. The second and third columns are two cloudy images with many different land covers. We observe that x_0^c and x_2^c are not linearly related, and their R^2 values are 0.7955 and 0.5503, respectively. However, there is a virtual boundary line at the lower right of the scatter plot, which is composed of the dark pixels in the clear image x_2^g . x_0^c is not affected by land cover, while x_2^c will increase with the increase in x_2^g . Therefore, these brighter land cover pixels are situated in the upper portion of the virtual boundary line. These dark pixels surrounding the virtual boundary line can be regarded as discrete homogeneous

land cover pixels, and the thickness coefficient α_2 is equal to the slope of this virtual boundary line.

Based on the aforementioned observation and understanding, the thickness coefficient of the i th band α_i can be estimated by the slope of the virtual boundary line in the scatter plot of x_0^c versus x_i^s . In this article, a CNN called Slope-Net is designed to estimate the thickness coefficient of each band. The architecture of Slope-Net is shown in Fig. 4. Bands 1–5 of the cloudy image and the reference thin cloud thickness map are input into Slope-Net, and the thickness coefficients of five bands

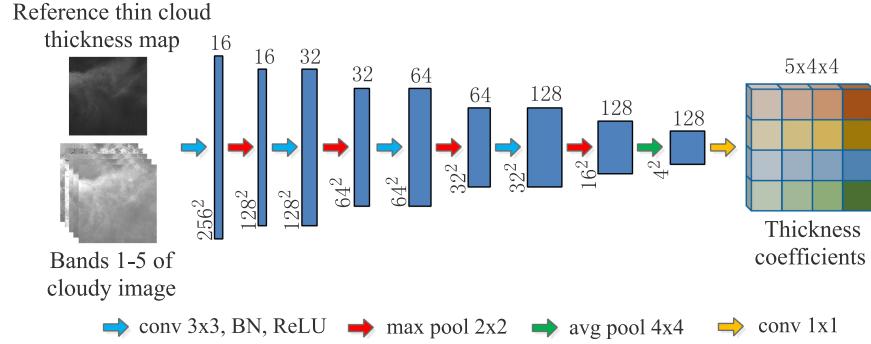


Fig. 4. Architecture of the designed Slope-Net.

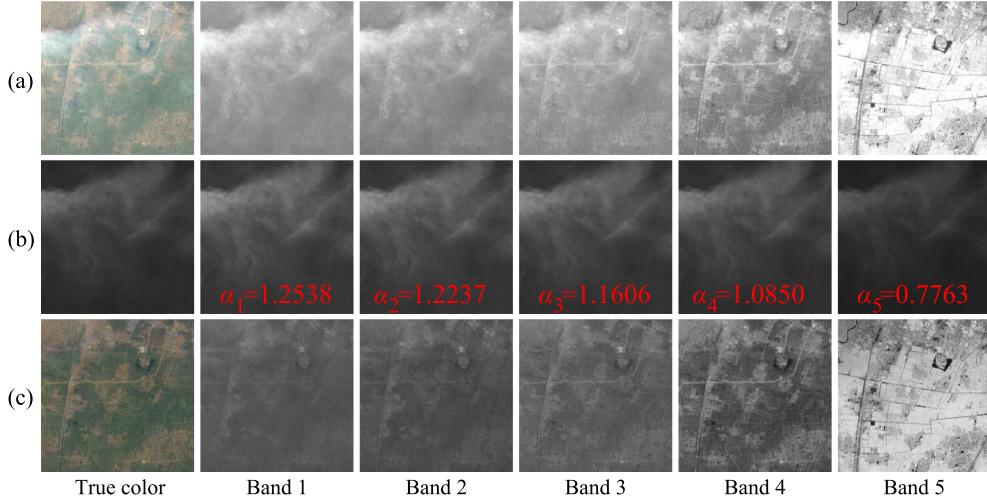


Fig. 5. Example of the thin cloud removal. (a) Cloudy image. (b) Thin cloud thickness map. (c) Thin cloud removal result.

are output. After four repeated operations of a 3×3 convolution with stride 1, followed by a BN, a ReLU and a 2×2 max pooling with stride 2, and a subsequent 4×4 average pooling operation with stride 4 are employed, we obtain a feature cube with size $128 \times 4 \times 4$, in which the receptive field size of each pixel is 94×94 . According to (3a), given a cloudy image and a reference thin cloud thickness map, its subblocks of arbitrary location and size have the same thickness coefficients. Therefore, we directly map each 128-D feature vector to the 5-D output by a 1×1 convolution operation at the last layer, and then, we obtain 16 sets of thickness coefficients simultaneously. Finally, we take the average of the 16 sets of thickness coefficients as our final result, thereby reducing the estimation error.

C. Thin Cloud Removal

After the reference thin cloud thickness map and the thickness coefficients are estimated by U-Net and Slope-Net, respectively, we can recover the multispectral clear image by subtracting the thin cloud thickness map of each band from the multispectral cloudy image according to (3b).

Fig. 5 shows an example of thin cloud removal for a multispectral image, where (a) is the real cloudy image, (b) is the thin cloud thickness map. The first column is the reference thin cloud thickness map estimated by U-Net. The thin cloud thickness map of each band is obtained by multiplying the reference thin cloud thickness map by the corresponding thickness coefficient estimated by Slope-Net. The thickness coefficient of each band is shown at the bottom of the thin cloud thickness map; (c) is the thin cloud removal result, which is obtained by subtracting the thin cloud thickness map (b) from the cloudy image (a). The thin clouds in each band are observed to be effectively removed.

IV. THIN CLOUD SIMULATION

We use a simulation method to generate abundant labeled images to train U-Net and Slope-Net. According to (3a), given a reference thin cloud thickness map and a set of thickness coefficients, we can synthesize a corresponding multispectral cloudy image from a multispectral clear image. DeepDive [24], McGAN [26], and SCR-GAN [27] used Perlin noise to generate the reference thin cloud thickness map randomly, and they assumed that the thickness of thin clouds was wavelength

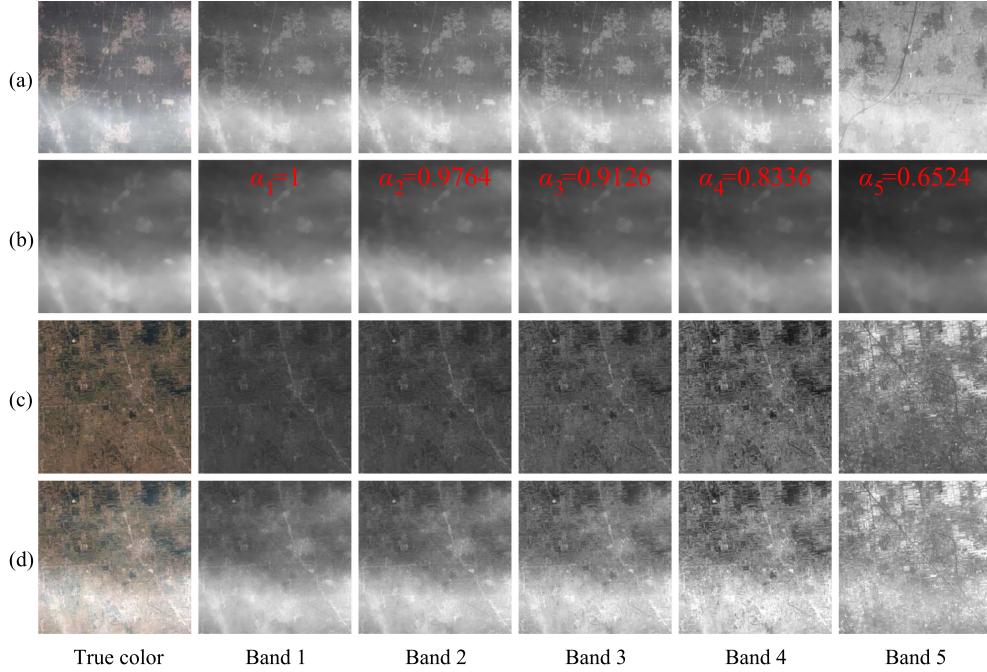


Fig. 6. Example of the thin cloud simulation. (a) Real cloudy image. (b) Thin cloud thickness map. (c) Real clear image. (d) Synthetic cloudy image.

independent, that is, the thickness coefficients of all bands were the same. However, according to [13] and [25], the thickness of thin clouds is wavelength dependent and decreases with increasing wavelength. To generate the synthetic cloudy image closer to real conditions, DH-CNN [25] extracted the reference thin cloud thickness map from the real cloudy image and then proposed a wavelength-dependent thin cloud simulation method according to the Rayleigh scattering law, but it was based on the multiplicative thin cloud imaging model. In this article, we propose a wavelength-dependent thin cloud simulation method based on the additive model.

We extract the reference thin cloud thickness map and thickness coefficient of each band from the real cloudy image. First, the thin cloud thickness map of each band is acquired by using a local window to search the dark pixels in the cloudy image [13]

$$x_i^c(p) = \min \{x_i^s(q) | q \in \Omega(p)\} \quad (4)$$

where $x_i^c(p)$ is the DN value of the thin cloud thickness map of the i th band at pixel p , and $x_i^c(p)$ is equal to the minimum value in a 3×3 local window $\Omega(p)$ centered on p in the cloudy image of the i th band x_i^s . Then, we select the thin cloud thickness map of Band 1 as the reference thin cloud thickness map, and the thickness coefficient of Band 1 is equal to 1. The thickness coefficients of other bands are calculated via linear regression between the thin cloud thickness maps of other bands and the reference thin cloud thickness map. Finally, given a real clear image and taking it with the reference thin cloud thickness map and the thickness coefficients into (3a), a corresponding synthetic cloudy image is obtained.

Fig. 6 shows a synthetic multispectral cloudy image, where (a) is a real cloudy image, from which the reference thin cloud thickness map and thickness coefficients are extracted, and (b)

is the thin cloud thickness map. The first column is the reference thin cloud thickness map, and the thin cloud thickness map of each band is obtained by multiplying the reference thin cloud thickness map by the corresponding thickness coefficient shown at the top of the thin cloud thickness map; (c) is a real clear image for simulation, and (d) is the synthetic cloudy image, which is obtained by adding the thin cloud thickness map (b) to the real clear image (c).

V. EXPERIMENTS

We evaluate the performance of our proposed method both qualitatively and quantitatively. The details of the experimental settings are introduced in Section V-A, and we discuss the effectiveness of the proposed modules in Section V-B. Finally, the experimental results of different methods are presented in Section V-C.

A. Experimental Settings

1) Datasets: Landsat 8 OLI images¹ were used to compose our training and test datasets. We downloaded 27 clear scenes that cover different landscapes such as urban areas, bare soil, vegetation, and water bodies and 17 scenes with cloud amounts ranging from 10% to 50% were used to generate synthetic thin clouds. These scenes were first cropped into 256×256 patches, and then, we collected 5000 clear patches and 5000 cloudy patches from clear scenes and cloudy scenes, respectively. Following the description of the thin cloud simulation method in Section IV, a reference thin cloud thickness map and a set of thickness coefficients can be extracted from each cloudy image.

¹Online. [Available]: <http://earthexplorer.usgs.gov/>

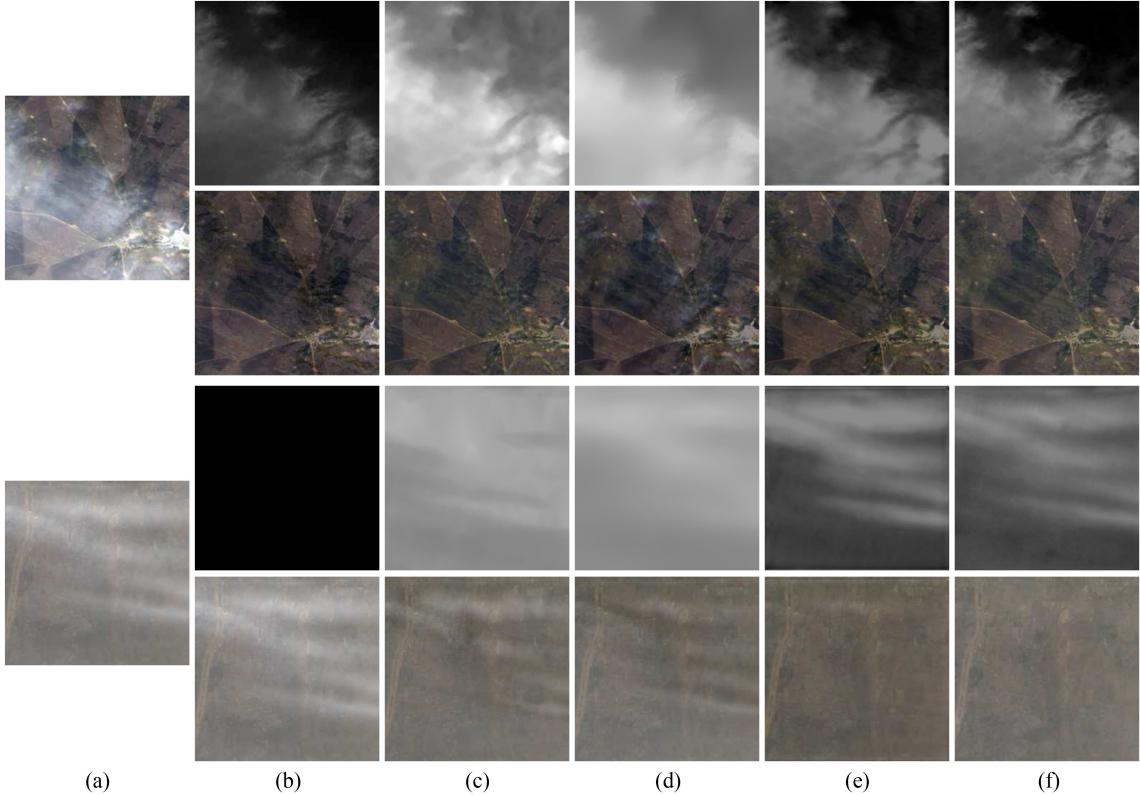


Fig. 7. Reference thin cloud thickness maps and thin cloud removal results of different methods. (a) Real cloudy image. (b)–(f) Reference thin cloud thickness maps and corresponding thin cloud removal results of Xu’s method [16], Makarau’s method [14], Liu’s method [15], Ren’s method [23], and our U-Net, respectively.

Then, a corresponding simulated cloudy image can be obtained from a clear image according to (3a). A total of 5000 sets of data were generated by simulation, of which 4500 sets were used as training data, and the remaining 500 sets were used as test data. In addition, we also established a real cloudy image test set, which contains 350 Landsat 8 OLI cloudy images with a size of 256×256 .

2) *Implementation Details*: The proposed thin cloud removal framework was implemented in PyTorch and trained on an Intel i7-6700 K 4.0-GHz CPU and an NVIDIA GTX 1080 GPU. Both U-Net and Slope-Net used the Euclidean distance between the output and the ground truth as the loss function, and they were trained by the Adam [30] optimizer with the parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 1 \times 10^{-8}$. We used a batch size of 1, and the epoch was set to 200. The initial learning rate was 0.0002, multiplied by 0.1 after every 50 epochs. Furthermore, the simulated cloudy image and reference thin cloud thickness map were used to train U-Net, and the simulated cloudy image, reference thin cloud thickness map, and thickness coefficients were used to train Slope-Net.

B. Effectiveness of the Proposed Modules

1) *U-Net*: In our proposed method, U-Net is used to estimate the reference thin cloud thickness map from the cloudy image. U-Net can obtain both low- to medium-altitude thin clouds and

high-altitude cirrus from cloudy images without additional cirrus band information. To demonstrate the superiority of U-Net, we compared it with Xu’s method [16], Makarau’s method [14], Liu’s method [15], and Ren’s method [23]. Xu’s method takes the cirrus band directly as the reference thin cloud thickness map. Makarau’s method uses the dark pixel searching method combined with the cirrus band to estimate the reference thin cloud thickness map, and Liu’s method uses the dark pixel searching and ground radiance suppressing method to estimate the reference thin cloud thickness map. In Ren’s method, a multiscale CNN is designed to obtain the reference thin cloud thickness map. Fig. 7 exhibits the results of different methods on two real cloudy images, where (a) is the real cloudy image and (b)–(f) are the reference thin cloud thickness maps and corresponding thin cloud removal results of Xu’s method, Makarau’s method, Liu’s method, Ren’s method, and our U-Net, respectively. Their thickness coefficients are estimated by Xu’s method according to their respective reference thin cloud thickness maps. In addition, there are high-altitude cirrus in the first image and low- to medium-altitude thin clouds in the second image. It can be seen that, when there is cirrus in the image, Xu’s method and Makarau’s method can obtain the accurate reference thin cloud thickness map as the cirrus band contains the cloud information, and then, the cirrus can be completely removed. However, when there is low- to medium-altitude thin cloud in the image, Xu’s reference thin cloud thickness map is invalid, and Makarau’s reference thin

TABLE II
AEEs (%) OF THICKNESS COEFFICIENTS FOR DIFFERENT METHODS ON THE SIMULATION TEST SET

Methods	AEE_1	AEE_2	AEE_3	AEE_4	AEE_5
Xu's [16]	9.76	13.31	22.03	28.95	52.62
Makarau's [14]	11.07	12.96	17.43	22.17	27.88
Liu's [15]	30.10	31.66	36.48	39.28	62.21
Slope-Net*	3.09	3.61	4.99	6.03	9.76
Slope-Net	2.22	2.77	3.74	4.37	8.45

cloud thickness map is not completely accurate because there is no cloud information in the cirrus band. Therefore, Xu's method cannot remove low- to medium-altitude thin clouds, and the restored result of Makarau's method has thin cloud residues. The reference thin cloud thickness map obtained by Liu's method is oversmoothed, resulting in that the cirrus and low- to medium-altitude thin clouds cannot be completely removed. Two CNN-based methods, Ren's method and our U-Net, can obtain an accurate reference thin cloud thickness map from the cloudy image without an additional cirrus band, regardless of whether there is high-altitude cirrus or low- to medium-altitude thin cloud in the image, to recover the clear image.

2) *Slope-Net*: In this article, Slope-Net is designed to estimate the thickness coefficient of each band. To verify the effectiveness and performance of our Slope-Net, we compared it with Xu's method [16], Makarau's method [14], Liu's method [15], and a variant of Slope-Net, referred to as Slope-Net, which outputs a set of thickness coefficients by replacing the 4×4 average pooling layer with a global average pooling layer (see Fig. 4). The comparative experiment is performed on a simulation test set with 500 images, in which the ground truth of the reference thin cloud thickness map is used as input. We used the average estimation error (AEE) to quantitatively evaluate the performance of the methods, which is defined as

$$AEE_i = \frac{1}{N} \sum_{j=1}^N \frac{|\hat{\alpha}_{ij} - \alpha_{ij}|}{\alpha_{ij}} \times 100\% \quad (5)$$

where AEE_i is the average estimation error of the i th band, N is the number of test images, $\hat{\alpha}_{ij}$ is the estimated value of the thickness coefficient of the i th band of the j th test image, and α_{ij} is the ground truth. Table II shows the AEEs of the thickness coefficients for different methods in each band. We can see that the AEEs of Xu's method, Makarau's method, and Liu's method are much greater than the AEEs of the two CNN-based methods. Xu's method must be combined with manual selection of homogeneous regions to gain better results, Makarau's method and Liu's method are unable to accurately identify cloud pixels and clear pixels when selecting cloud pixels to calculate the thickness coefficients via linear regression. In Bands 1–5, the results of Slope-Net are better than the results of Slope-Net by 0.87%, 0.84%, 1.25%, 1.66%, and 1.31%, respectively, indicating that Slope-Net can reduce the estimation error by outputting 16 sets of thickness coefficients.

TABLE III
METRIC STATISTICS OF DIFFERENT METHODS ON THE SYNTHETIC TEST SET

Methods	MSE	PSNR	SSIM
DeepDive [24]	0.0116	21.6713	0.8110
DH-CNN [25]	0.0051	25.0456	0.8957
RSC-Net [28]	0.0066	23.6840	0.9166
McGAN [26]	0.0006	33.4843	0.9715
SCR-GAN [27]	0.0008	32.8516	0.9657
Proposed	0.0006	34.2562	0.9832

C. Comparison With State-of-the-Art Methods

1) *Comparison on Synthetic Images*: We compared the proposed method with five state-of-the-art thin cloud removal methods, including DeepDive [24], DH-CNN [25], RSC-Net [28], McGAN [26], and SCR-GAN [27], which are all based on deep learning. DeepDive, DH-CNN, and RSC-Net belong to the CNN architecture, and McGAN and SCR-GAN belong to the generative adversarial network architecture.² We reproduced RSC-Net and SCR-GAN with default values. All five compared methods and our proposed method were supervised. To ensure an equal assessment, these compared methods change the input to five bands and use the same clear images and reference thin cloud thickness maps as our method to generate the simulation training and test datasets. Considering that DeepDive, DH-CNN, McGAN, and SCR-GAN are based on the multiplicative model, we used the simulation method in their original paper to generate the simulation images. The remaining RSC-Net uses our thin cloud simulation method to synthesize images.

Fig. 8 shows five instances of the thin cloud removal results on synthetic cloudy images, where (a) is the synthetic cloudy image, (b) is the ground truth, and (c)–(h) are the results of five compared methods and our method. We can observe that all the results of DeepDive have obvious color distortion. DH-CNN results show that there is some excessive thin cloud removal in the third column and some cloud residues in the fourth column. The results of RSC-Net in the last two columns have slight cloud residues. Obviously, the remaining three methods can effectively remove thin clouds in the five images, and the recovered results have good color fidelity.

We also use the mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) metrics to quantitatively evaluate the performance of the methods. A small MSE, large PSNR, and large SSIM indicate good thin cloud removal performance. The metric statistical results of the six methods on 500 synthetic test images are shown in Table III. It can be seen that the latter three methods are better than the former three. The former three methods do not use any dimension reduction operations in their network architectures, whereas the latter three methods utilize downsampling and upsampling operations to acquire a larger receptive field and better features. Among all the comparison methods, our method achieved the best value in all three metrics.

²Online. [Available]: The codes of DeepDive, DH-CNN, and McGAN were downloaded from https://github.com/puf3zin/deep_dehazing, <http://xfy.buaa.edu.cn/>, and <https://github.com/enomotokenji/mcgan-cvprw2017-pytorch>, respectively.

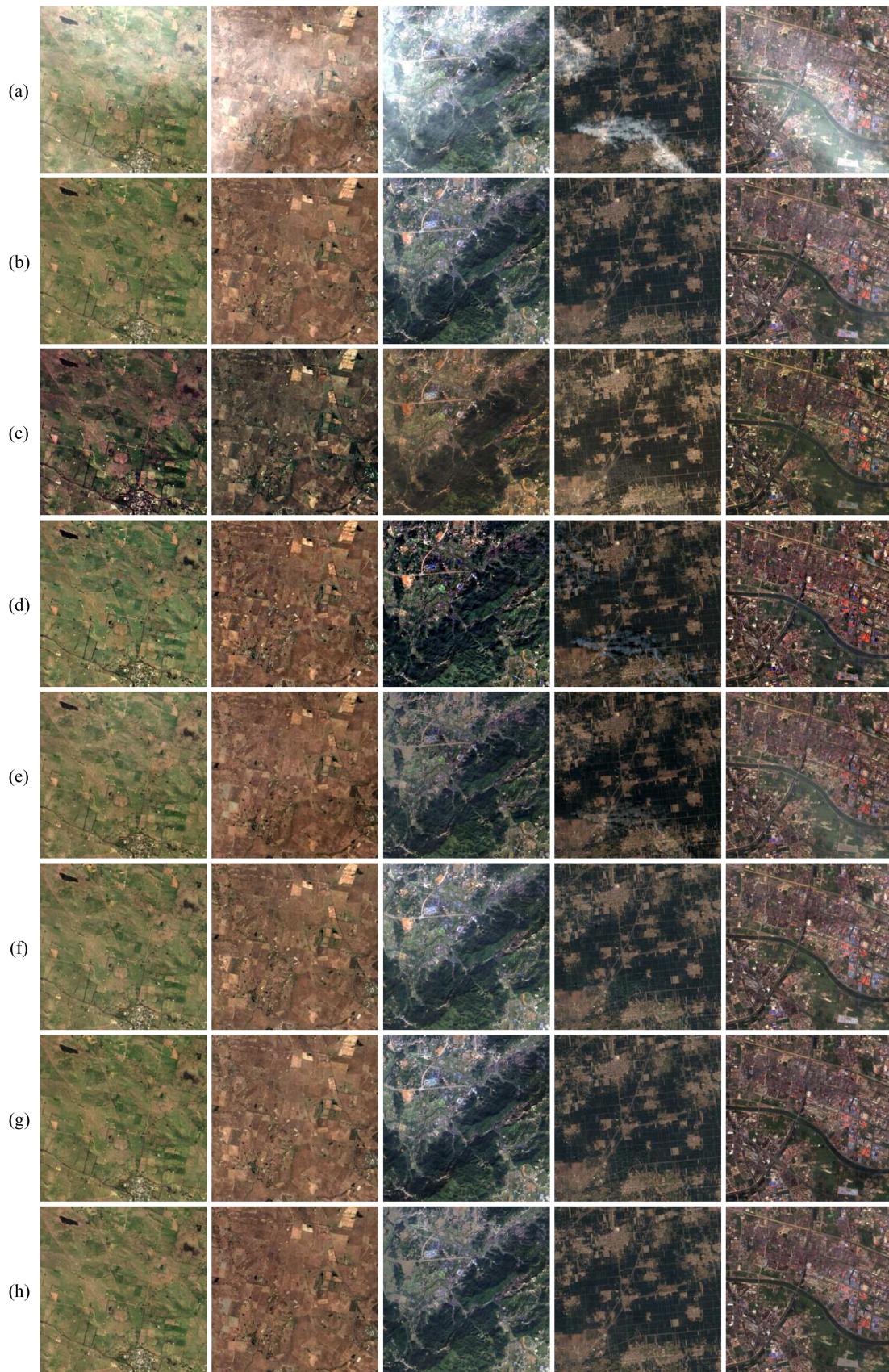


Fig. 8. Thin cloud removal results on synthetic cloudy images. (a) Synthetic cloudy images. (b) Ground truth. (c)–(h) Thin cloud removal results of DeepDive [24], DH-CNN [25], RSC-Net [28], McGAN [26], SCR-GAN [27], and our proposed method, respectively.

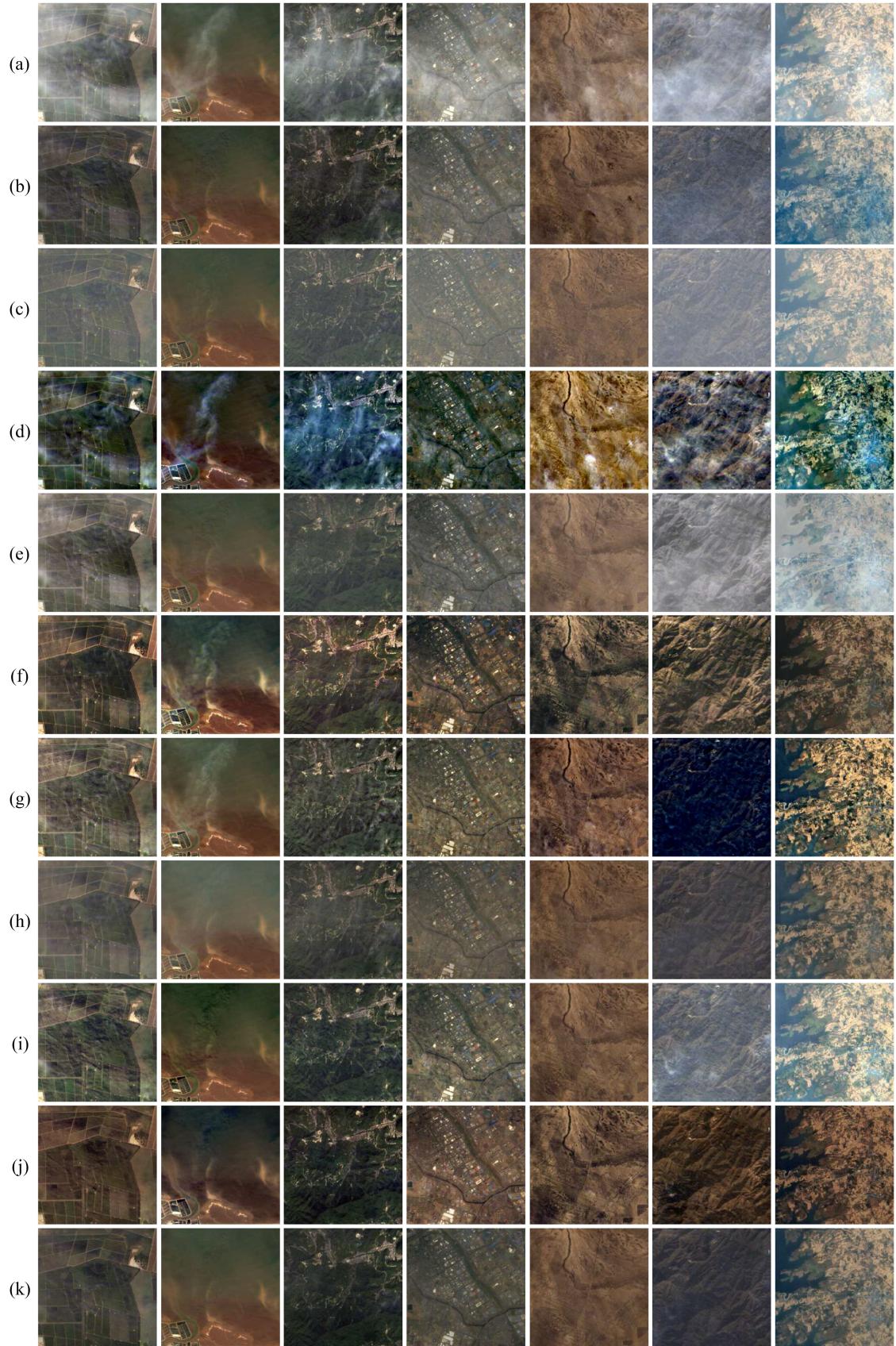


Fig. 9. Thin cloud removal results on real cloudy images. (a) Real cloudy images. (b)–(k) Thin cloud removal results of Xu's method [16], Makarau's method [14], Li's method [20], CR-NAPCT [21], DeepDive [24], DH-CNN [25], RSC-Net [28], McGAN [26], SCR-GAN [27], and our proposed method, respectively.

2) *Comparison on Real Images:* Fig. 9 shows several thin cloud removal examples for real cloudy images with different land cover types, including farmland, sea, forest, urban, desert, mountain, and coast. In addition to the five deep-learning-based comparison methods mentioned in the previous section, we also added four traditional unsupervised thin cloud removal methods for comparison, including Xu's method [16], Makarau's method [14], Li's method [20], and CR-NAPCT [21]. In addition, the code of Li's method was downloaded from,³ and we reproduced the remaining three methods with default values. It can be observed that the results of deep-learning-based methods are, on the whole, better than the results of traditional unsupervised methods. Xu's method and Makarau's method cannot completely remove the clouds in the images, and Li's results have serious cloud residues and color distortion. The results of CR-NAPCT are better than the results of the previous three methods, but there are still thin cloud residues in the first image and color distortion in the last image. The results of DeepDive, DH-CNN, and RSC-Net have cloud residues in the first three columns. Moreover, DeepDive and DH-CNN results have poor color fidelity in the last three columns. The McGAN results indicate that thin clouds are not totally removed in the first two columns and the last two columns. SCR-GAN results have obvious color deviation. Different from the compared methods, our proposed method achieved good results in terms of both thin cloud removal performance and color fidelity in all seven images.

Since atmospheric radiation and land cover change over time, an ideal cloud-free ground truth of the cloudy image is difficult to collect. We used two nonreference metrics, the dark channel feature (DC) [31] and change in hue (ΔH) [32], to quantitatively evaluate the performance of the methods, which are defined as

$$DC = \frac{1}{N_{R_c}} \sum_{p \in R_c} \min_{i \in \{r, g, b\}} \left(\min_{q \in \Omega(p)} (x_i^g(q)) \right) \quad (6)$$

$$\Delta H = \frac{1}{N_{R_{nc}}} \sum_{p \in R_{nc}} |h^g(p) - h^s(p)| \quad (7)$$

where x_i^g is one of the bands of the restored image, R_c represents the cloudy region in the original cloudy image, N_{R_c} is the number of pixels of R_c , and $\Omega(p)$ is a local patch centered at pixel p . The patch size is 5×5 . h^g and h^s are the hue channels in the HSI color space of the restored clear image and the original cloudy image, respectively. R_{nc} represents the noncloud region, and $N_{R_{nc}}$ is the number of pixels of R_{nc} . In addition, the cloudy region and noncloud region can be obtained by binarizing the reference thin cloud thickness map.

DC represents the thin cloud residual degree in the cloudy region. ΔH represents the change in hue in the noncloud region between the original image and restored image. A small DC and ΔH indicate good thin cloud removal performance.

Table IV shows the statistical results of different methods on 350 real cloudy images. We can clearly see that Xu's method, Makarau's method, CR-NAPCT, DeepDive, McGAN,

TABLE IV
METRIC STATISTICS OF DIFFERENT METHODS ON THE REAL TEST SET

Methods	DC	ΔH
Xu's [16]	0.2497	0.0320
Makarau's [14]	0.3276	0.0328
Li's [20]	0.1292	0.1575
CR-NAPCT [21]	0.3231	0.0453
DeepDive [24]	0.1786	0.1537
DH-CNN [25]	0.1434	0.1292
RSC-Net [28]	0.1970	0.0468
McGAN [26]	0.1827	0.0352
SCR-GAN [27]	0.1535	0.1478
Proposed	0.1351	0.0337

and RSC-Net have large DC values, indicating that they cannot completely remove thin clouds, and there are thin cloud residues in the restored images. Li's method, DeepDive, DH-CNN, and SCR-GAN obtain large ΔH values, which means that their thin cloud removal results have poor color fidelity. Among the ten methods, our method achieved the second-best DC value and the third-best ΔH value, and both of our DC and ΔH values are only marginally lower than the best values. Overall, our method can effectively remove thin clouds along with good color fidelity.

VI. CONCLUSION

Clouds often exist in multispectral remote sensing images, and these clouds reduce the image quality and result in incorrect interpretation. In this article, a novel thin cloud removal method is proposed for multispectral remote sensing images. The proposed method is based on an additive thin cloud imaging model, and it combines CNNs and a traditional imaging model. Considering the strong feature extraction ability and nonlinear regression ability of the CNN, we utilize U-Net to estimate the reference thin cloud thickness map, and then, Slope-Net is designed to estimate the thickness coefficient of each band. Thus, thin cloud thickness maps of different bands are obtained. Finally, based on the traditional thin cloud imaging model, we remove thin clouds by subtracting the thin cloud thickness maps from the cloudy image. Our proposed method achieves good thin cloud removal results by acquiring accurate thin cloud information, including both low- to medium-altitude thin clouds and high-altitude cirrus, and considering the different thin cloud thickness maps of each band. In addition, the collection of cloudy images and their corresponding ideal cloud-free images is difficult, and we present a wavelength-dependent thin cloud simulation method to produce sufficient labeled data for training U-Net and Slope-Net. Our simulation method extracts the reference thin cloud thickness maps and thickness coefficients from real cloudy images, thereby generating synthetic cloudy images that are very close to the real conditions. Qualitative and quantitative experiments are performed on synthetic images and real images from Landsat 8 OLI. The results demonstrate that compared with other state-of-the-art thin cloud removal methods, our method achieves better results in both thin cloud removal performance and color fidelity for multispectral images with various land cover types.

Although our method obtains satisfying results, there are still some limitations. On the one hand, according to the thin cloud

³Online. [Available]: <http://www.escience.cn/people/lijiayuan/index.html>

imaging model, the proposed method uses two CNNs to estimate the reference thin cloud thickness map and thickness coefficient, respectively. Compared with the end-to-end thin cloud removal methods, our proposed method will lead to accumulative estimation error. On the other hand, the proposed method uses the simulated data to train the U-Net and Slope-Net, because the real reference thin cloud thickness map and thickness coefficient of the real cloudy image are unavailable. However, the simulated cloudy images are not completely consistent with the real cloudy images, so the thin cloud removal results of real cloudy images are worse than the thin cloud removal results of simulated cloudy images. In the future, we aim to design an end-to-end network to remove thin clouds in remote sensing images, and train it with real unpaired images.

REFERENCES

- [1] Y. Du, B. Guindon, and J. Cihlar, "Haze detection and removal in high resolution satellite image with wavelet analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 1, pp. 210–217, Jan. 2002.
- [2] D. G. Hadjimitsis, C. R. I. Clayton, and V. S. Hope, "An assessment of the effectiveness of atmospheric correction algorithms through the remote sensing of some reservoirs," *Int. J. Remote Sens.*, vol. 25, no. 18, pp. 3651–3674, 2004.
- [3] E. F. Vermote, D. Tanre, J. L. Deuze, M. Herman, and J. J. Morcrette, "Second simulation of the satellite signal in the solar spectrum, 6S: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 675–686, May 1997.
- [4] R. G. Isaacs, W.-C. Wang, R. D. Worsham, and S. Goldenberg, "Multiple scattering LOWTRAN and FASCODE models," *Appl. Opt.*, vol. 26, no. 7, pp. 1272–1281, 1987.
- [5] A. Berk, L. S. Bernstein, and D. C. Robertson, "MODTRAN: A moderate resolution model for LOWTRAN 7," *Geophys. Lab.*, Bedford, MA, USA, Tech. Rep. GL-TR-89-0122, 1989.
- [6] R. Richter, "A spatially adaptive fast atmospheric correction algorithm," *Int. J. Remote Sens.*, vol. 17, no. 6, pp. 1201–1214, 1996.
- [7] F. Pacifici, N. Longbotham, and W. J. Emery, "The importance of physical quantities for the analysis of multitemporal and multiangular optical very high spatial resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6241–6256, Oct. 2014.
- [8] O. R. Mitchell, E. J. Delp, and P. L. Chen, "Filtering to remove cloud cover in satellite imagery," *IEEE Trans. Geosci. Electron.*, vol. 15, no. 3, pp. 137–141, Jul. 1977.
- [9] J. Liu *et al.*, "Thin cloud removal from single satellite images," *Opt. Express*, vol. 22, no. 1, pp. 618–632, Jan. 2014.
- [10] H. Shen, H. Li, Y. Qian, L. Zhang, and Q. Yuan, "An effective thin cloud removal procedure for visible remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 96, pp. 224–235, Oct. 2014.
- [11] P. S. Chavez, "An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data," *Remote Sens. Environ.*, vol. 24, no. 3, pp. 459–479, Apr. 1988.
- [12] P. S. Chavez, "Image-based atmospheric corrections-revisited and improved," *Photogramm. Eng. Remote Sens.*, vol. 62, no. 9, pp. 1025–1035, Sep. 1996.
- [13] A. Makarau, R. Richter, R. Muller, and P. Reinartz, "Haze detection and removal in remotely sensed multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5895–5905, Sep. 2014.
- [14] A. Makarau, R. Richter, D. Schlapfer, and P. Reinartz, "Combined haze and cirrus removal for multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 379–383, Mar. 2016.
- [15] Q. Liu, X. Gao, L. He, and W. Lu, "Haze removal for a single visible remote sensing image," *Signal Process.*, vol. 137, pp. 33–43, Aug. 2017.
- [16] M. Xu, X. Jia, and M. Pickering, "Automatic cloud removal for Landsat 8 OLI images using cirrus band," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 2511–2514.
- [17] B. Zhou and Y. Wang, "A thin-cloud removal approach combining the cirrus band and RTM-based algorithm for Landsat-8 OLI data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 1434–1437.
- [18] T. Markchom and R. Lipikorn, "Thin cloud removal using local minimization and logarithm image transformation in HSI color space," in *Proc. 4th Int. Conf. Frontiers Signal Process.*, Sep. 2018, pp. 100–104.
- [19] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
- [20] J. Li, Q. Hu, and M. Ai, "Haze and thin cloud removal via sphere model improved dark channel prior," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 472–476, Mar. 2019.
- [21] M. Xu, X. Jia, M. Pickering, and S. Jia, "Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform," *ISPRS J. Photogramm. Remote Sens.*, vol. 149, pp. 215–225, Mar. 2019.
- [22] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.
- [23] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M. H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 154–169.
- [24] L. T. Goncalves, J. D. O. Gaya, P. Drews, and S. S. D. C. Botelho, "DeepDive: An end-to-end dehazing method using deep learning," in *Proc. 30th Conf. Graph., Patterns Images*, 2017, pp. 436–441.
- [25] M. Qin, F. Xie, W. Li, Z. Shi, and H. Zhang, "Dehazing for multispectral remote sensing images based on a convolutional neural network with the residual architecture," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1645–1655, May 2018.
- [26] K. Enomoto *et al.*, "Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2017, pp. 1533–1541.
- [27] R. Zhang, F. Xie, and J. Chen, "Single image thin cloud removal for remote sensing images based on conditional generative adversarial nets," *Proc. SPIE*, vol. 10806, Aug. 2018, Art. no. 1080650.
- [28] W. Li, Y. Li, D. Chen, and J. C.-W. Chan, "Thin cloud removal with residual symmetrical concatenation network," *ISPRS J. Photogramm. Remote Sens.*, vol. 153, pp. 137–150, 2019.
- [29] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, Nov. 2015, pp. 234–241.
- [30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–15.
- [31] M. Qin, F. Xie, and Z. Jiang, "No reference assessment of image visibility for dehazing," in *Proc. Int. Conf. Image Graph.*, vol. 10666, Dec. 2017, pp. 664–674.
- [32] F. Xie, J. Chen, X. Pan, and Z. Jiang, "Adaptive haze removal for single remote sensing image," *IEEE Access*, vol. 6, pp. 67982–67991, 2018.



Yue Zi received the B.S. degree in detection, guidance and control technology from Beihang University, Beijing, China, in 2013, where he is currently working toward the Ph.D. degree with the Image Processing Center, School of Astronautics.

His research interests include remote sensing image processing, target detection, semantic segmentation, and deep learning.



Fengying Xie (Member, IEEE) received the Ph.D. degree in pattern recognition and intelligent system from Beihang University, Beijing, China, in 2009.

From 2010 to 2011, she was a Visiting Scholar with the Laboratory for Image and Video Engineering, The University of Texas at Austin, Austin, TX, USA. She is currently a Professor with the Image Processing Center, School of Astronautics, Beihang University. Her research interests include biomedical image processing, remote sensing image understanding and applications, image quality assessment, and object recognition.



Ning Zhang received the Ph.D. degree in optical engineering from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China, in 2012.

He is currently a Senior Engineer with the Shanghai Aerospace Electronic Technology Institute, Shanghai, China. His current research interests include remote sensing image processing and electronic system design.



Wentao Zhu received the M.S. degree in information and communication engineering from the Shanghai Academy of Spaceflight Technology, Shanghai, China, in 2020.

He is currently an Engineer with the Shanghai Aerospace Electronic Technology Institute, Shanghai. His current research interests include information processing, target detection, field-programmable gate array design, and image compression.



Zhiguo Jiang (Member, IEEE) received the Ph.D. degree in pattern recognition and intelligent system from Beihang University, Beijing, China, in 2005.

He is currently a Professor with the School of Astronautics, Beihang University. His research interests include remote sensing image analysis, target detection, tracking and recognition, and medical image processing.

Dr. Jiang is a standing member of the Executive Council of the China Society of Image and Graphics and also a member of the Executive Council of the Chinese Society of Astronautics. He is an Editor for the *Chinese Journal of Stereology and Image Analysis*.



Haopeng Zhang (Member, IEEE) received the Ph.D. degree in pattern recognition and intelligent system from Beihang University, Beijing, China, in 2014.

He is currently an Assistant Professor with the Image Processing Center, School of Astronautics, Beihang University. His main research interests include remote sensing image processing, multiview object recognition, 3-D object recognition and pose estimation, and other related areas in pattern recognition, computer vision, and machine learning.