IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 27, 2019, accepted January 14, 2020, date of publication January 17, 2020, date of current version January 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2967590

Cloud and Cloud Shadow Detection Based on Multiscale 3D-CNN for High Resolution Multispectral Imagery

YANG CHEN[1], LULIANG TANG[1], ZIHAN KAN[1], AAMIR LATIF[2], XIUCHENG YANG[3], MUHAMMAD BILAL[4], AND QINGQUAN LI[1,5]

[1]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[2]Institute of Geographic Sciences and Natural Resources Research, University of Chinese Academy of Sciences, Beijing 10010, China
[3]ICube laboratory, University of Strasbourg, 67000 Strasbourg, France
[4]School of Marine Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China
[5]Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060, China

Corresponding author: Luliang Tang (tll@whu.edu.cn)

This work was supported in part by grants from the National Key Research and Development Plan of China under Grant 2017YFB0503604 and Grant 2016YFE0200400, and in part by the National Natural Science Foundation of China under Grant 41971405, Grant 41671442, and Grant 41571430.

ABSTRACT Cloud and cloud shadow detection is one of the most important tasks for optical remote sensing image preprocessing. It is not an easy task due to the variety and complexity of underlying surfaces, such as the low-albedo objects (water and mountain shadow) and the high-albedo objects (snow and ice). In this study, an end-to-end multiscale 3D-CNN method is proposed for cloud and cloud shadow detection in high resolution multispectral imagery. Specifically, a multiscale learning module is designed to extract cloud and cloud shadow contextual information of different levels. In order to make full use of band information, four band-combination images are inputted into the multiscale 3D-CNN. A joint spectral-spatial information of 3D-convolution layer is developed to fully explore the joint spatial-spectral correlations feature in the input data. Overall, in the experiments undertaken in this paper, the proposed method achieved a mean overall accuracy of 97.27% for cloud detection, with a mean precision of 96.02% and a mean recall of 95.86%. For cloud shadow detection, the proposed method achieved a mean precision of 95.92% and a mean recall of 92.86%. Experimental results on two validation datasets (GF-1 WFV validation data and ZY-3 validation data) show that the proposed multiscale-3D-CNN method achieved good performance with limited spectral ranges.

INDEX TERMS Cloud detection, cloud shadow, convolution neural networks, multiscale 3D-CNN.

I. INTRODUCTION

Optical high-resolution remote sensing images (such as SPOT/Gaofen-1) are widely used for environment monitoring, geographical mapping, and change detection [1]. Clouds and cloud shadows obscure the spectral information of optical remote sensing sensors [2], and thus the presence of clouds and cloud shadows significantly influences the availability of optical high-resolution images, such as image fusion and change detection [3]. Therefore, the accurate identification of clouds and cloud shadows is one of the most important techniques for optical remote sensing applications.

Recently, a variety of cloud\shadow detection methods for remote sensing imagery have been proposed [4]. These methods can broadly be categorized into threshold based methods and image classification methods [5]. Previous threshold-based methods often utilize either predefined thresholds or adaptive thresholds to mask clouds in designed images [6]. Huang *et al.* [7] proposed an automated masking algorithm for cloud and cloud shadow detection using adaptive thresholds defined in Landsat images. Zhu and Woodcock [8] exploited Function of Mask (Fmask) algorithm to predict possible cloud locations through the scene based threshold, and then detect the cloud shadows by object geometry matching. Li *et al.* [9] proposed an automatic multi-feature combined (MFC) method to acquire the cloud

The associate editor coordinating the review of this manuscript and approving it for publication was Ali Kashif Bashir.

mask by threshold segmentation based on the multi-features (geometric, spectral, and textural features) and guided filtering, and calculate the cloud shadows utilizing the cloud and shadow matching and follow-up correction process. The above threshold-based method can successfully provide an accurate cloud and cloud shadow mask, but they are highly dependent on the sensors and sensitive to changes in atmospheric conditions and scene properties.

Image classification based on feature extraction and machine learning is also an effective method for cloud and cloud shadow detection. Li *et al.* [10] trained a Support Vector Machine (SVM) classifier to detect clouds from reflectance and texture information. Hollstein *et al.* [11] developed an overview of several ready-to-use machine learning to detect cloud and cloud shadow in Sentinel-2 images. Hughes and Hayes [12] uses a neural network approach to determine cloud, cloud shadow, and clear sky classification memberships of each pixel in Landsat OLI images. Generally, image classification methods yield more accurate cloud/shadow detection results than the threshold methods.

In optical remote sensing images, previous cloud shadows detection methods are often accomplished after cloud detection by geometrical matching [13]. It is difficult to detect cloud shadows, because their spectral signatures overlap with other low-albedo objects [14]. In the optical high-resolution images, it is not easy to separate clouds from some bright ground objects (such as snow, white buildings) when only using the spectral features [15]. Therefore, the accurate detection of cloud and cloud shadow is quite challenging for optical high-resolution images due to the limited spectral ranges (including blue, green, red, and near infrared bands) and the complexity of underlying surfaces.

In recently years, convolutional neural networks (CNN) have achieved great success in image classification [16], segmentation [17], and recognition tasks [18]. Deep learning-based semantic segmentation models can extract features automatically from input images in recent studies on cloud detection [19], such as Deeplab [20] and pyramid scene parsing network (PSPnet) [21]. The performance of segmentation networks cannot perform good in different objects, especially for small objects owing to the pooling effect. Shi *et al.* [22] used a single-branch CNN to extract cloud regions from superpixels. Xie *et al.* [23] developed a two-branch CNN to extract cloud regions from superpixels. Chen *et al.* [24] exploited the multiple-CNN model to detect cloud regions. Wang *et al.* [25] presented a new CNN to detect cloud and snow on an object level. Goff *et al.* [26] developed a fully connected CNN to extract cloud regions. Wieland *et al.* [27] exploited the modified U-Net 2D-CNN for cloud and cloud shadow segmentation. Chai *et al.* [28] uses an adaption of SegNet to detect cloud and cloud shadow in Landsat imagery.

However, the performance of some CNN methods are too dependent on superpixels segmentation accuracy. In addition, the spatial features and spectral features are extracted separately due to 2D-CNN is used in some cloud detection methods. The 2D-CNN based methods cannot fully extract the joint spatial-spectral correlations feature [29], which can be critical for cloud and cloud shadow detection. However, some scholars developed 3D-CNN model to extract deep spectral-spatial features directly from input data [30], [31]. But, these models cannot perform well in multi-scale spectral–spatial features. In order to produce good detection results, more information such as multi-scale contextual information should be taken into consideration.

In this paper, an end-to-end multiscale-3D-CNN architecture is proposed for cloud and cloud shadow detection in high resolution multispectral imagery. The proposed method enjoys the benefit from end-to-end deep learning and the performance of the proposed method is not dependent on superpixels segmentation accuracy. Specifically, the multiscale learning module is designed to extract cloud and cloud shadow contextual information of different levels. In order to fully explore the joint spatial-spectral correlations feature, a spectral-spatial information of 3D-convolution layer is developed for high resolution multispectral imagery. Experimental results on two validation datasets (GF-1 WFV validation data and ZY-3 validation data) show that the proposed multiscale 3D-CNN model achieves good performance and it does not require additional superpixels segmentation processing.

This paper introduces the cloud and cloud shadow detection method based on multiscale 3D-CNN. The major work of this paper is as follows:

(1) An end-to-end 3D-CNN method is proposed for cloud and cloud shadow detection in high resolution multispectral imagery.

(2) In order to fully explore the joint spatial-spectral correlations feature, a spectral-spatial information of 3D-convolution layer is developed.

(3) A multiscale learning module is designed to extract cloud and cloud shadow contextual information of different levels.

The rest of this paper is organized as follow. Section II describes the proposed cloud and cloud shadow detection method in detail. Section III presents the cloud and cloud shadow experimental results. Some discussions are offered in Section IV. Conclusion is summarized in Section V.

II. METHODOLOGY

In this section, the proposed framework of cloud and cloud shadow detection based on multi-scale 3D-CNN is introduced, and their main contributions are highlighted. First, a spectral-spatial information of 3D-convolution layer is developed to fully extract the joint spatial-spectral correlations feature. Second, the multi-scale 3D-CNN structure is discussed, which is used to extract cloud and cloud shadow contextual information of different levels. The overall structure of the proposed framework for cloud and cloud shadow detection from high resolution multispectral imagery is shown in Fig.1.

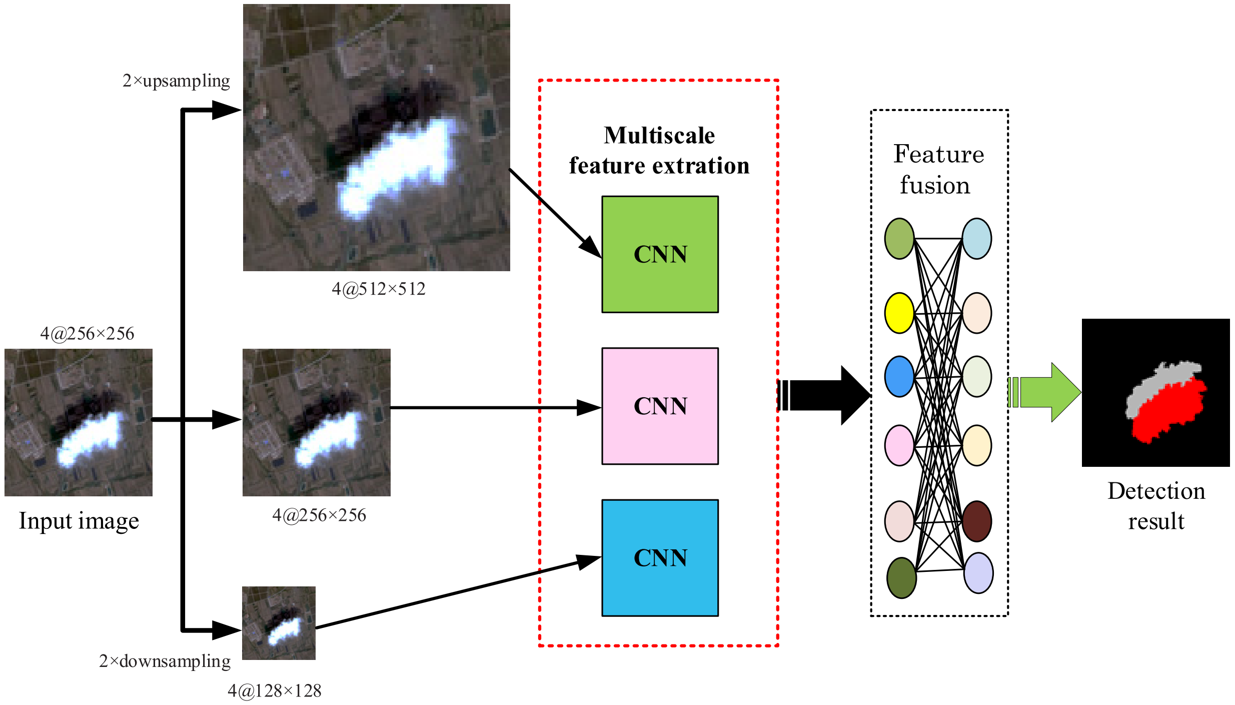


FIGURE 1. Framework of the proposed method for cloud and cloud shadow detection.

A. 3D-CONVOLUTION LAYER

The convolution layer produces new feature maps from previous feature maps and acts as multiple learnable filters in the input image [17]. In traditional CNN, 1-D convolution is used at the convolutional layers to extract spectral features from the input data [32]. 2-D convolution is used at the convolutional layers to extract spatial features from the input data [33]. Fig.2 (a) illustrates the main process of a 2-D convolution operation.

Recently, 3D-convolution layer is achieved to extracts features from both spatial and temporal dimensions by convolving a 3D kernel in the video processing task [34], [35]. The highly nonlinear semantic relationship between multispectral (MS) image bands indicates that higher level expression is essential for cloud and cloud shadow detection. However, it is difficult to directly extract spatial-spectral information from MS images in both the spatial and the spectral dimensions, when the 2D-CNN is used. The 3D-CNN structure seeks to fully extract the joint spatial-spectral correlations feature from the MS images. In 3D-CNN structure, in order to fully explore the joint spatial-spectral correlations feature, a joint spectral-spatial information of 3D-convolution layer is developed. Fig.2 (b) illustrates the main process of a joint spectral-spatial information 3D convolution operation. Formally, the joint spectral-spatial information of 3D-convolutional process is defined as follows.

$$G_{ij}^{xyz} = f(\sum_{k}\sum_{m=0}^{M_i-1}\sum_{n=0}^{N_i-1}\sum_{n=0}^{4} w_{ij}{}^{mnb} G_{(i-1)k}^{(x+m)(y+n)(z+b)} + b_{ij}) \tag{1}$$

where, b is the depth of the 3D kernel along spectral dimension, $G_{(i-1)k}^{(x+m)(y+n)(z+b)}$ indicates the (i-1)th layer of kth feature map, M_i is the length of the 3D kernel, N_i is the width of the 3D kernel, b_{ij} is the bias for this feature map, $f(\bullet)$ is the activation function.

In order to keep the spatial extent of the activations after convolutions, in this paper, a zero-padding method is used as it does not change activations and compensate for the number of lost pixels at the borders of the feature maps. In the CNN networks, the activation function is one of the significant factors, which brings nonlinearity into the CNN networks. Generally, between the convolutional layer and the other convolutional layer by activation function [36]. In the CNN networks, the Rectified Linear Unit (ReLU) function is conventionally used as an activation function because the neurons with rectified functions performing well to overcome saturation [37]. The ReLU is defined as follows [38]:

$$\mathrm{ReLU}(x) = \max(0, x) \tag{2}$$

B. STRUCTURE OF MULTISCALE 3D-CNN

Generally speaking, the cloud shadow is around the cloud, and the snow goes along the extending direction of a mountain [39]. This contextual information is very important for cloud and cloud shadow detection. It is key factors that how to effectively mine multiscale contextual relation between cloud and cloud shadow [28]. Inspired by Fully CNN (FCNN) [40] as backbone to extract semantic features of input images of different sizes, a multiscale 3D-CNN structure is designed to extract cloud and cloud shadow contextual information of different levels.

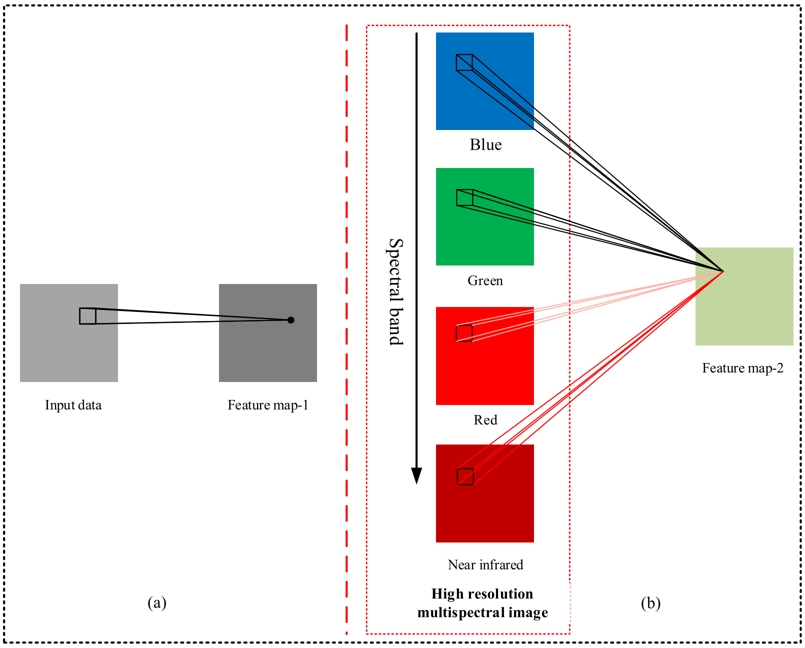


FIGURE 2. illustrates of the main differences between the 2D and the 3D convolutions. (a) 2D convolution operation. (b) 3D convolution operation.

As shown in Fig. 3, the proposed CNN structure is divided into three parts. The proposed CNN structural contains multi-scale sampling module, multi-scale learning module, and feature fusion module. In this study, each feature map is defined as a 3-D array with the size of height × width × depth, where width and height indicates the spatial size, and depth denotes the spectral size. Conv# indicates a 3D-convolutional layer with the size height × width × 4.

In the multi-scale sampling module (Input images preprocessing module), in order to get the multiscale input images, we first upsample the original images by a factor of 2 via bilinear interpolation to get the input images of size 512 × 512 × 4. Then, we downsample the original images as the input images of size 128 × 128 × 4. Finally, the different size patches (512 × 512 × 4, 256 × 256 × 4, and 128 × 128 × 4) are inputted into the multiscale learning module.

The multi-scale learning module consists of different mode: Global contextual mode, Local spatial mode, and Local contextual mode. The global contextual mode is designed to perceive global contextual information. The local spatial mode is designed to extract low-level spatial information. The local contextual mode is designed to perceive local contextual information.

Generally speaking, the smaller size convolutional kernels are used to exploit local features, the larger size convolutional kernels are used to exploit global features [41]. Therefore, in the multi-scale learning module, in order to extract both low-level spatial information and multi-scale semantic information, we set different sizes of the 3D-convolution kernel. As we know, global contextual information has strong contexts but weak local spatial information, whereas local spatial information has strong locations but weak context. Adding multiscale learning module to the CNN can improve the accuracy of cloud and cloud shadow detection. Global average pooling (GAP) has proven to be a good model as the global contextual features [42]. The GAP can reduce the plenty of parameters, sequentially GAP is used to heavily reduce the number of learning parameters in the very deep CNN structural [43], [47]. Therefore, in the multi-scale learning module, the final convolutional layer is followed by global average pooling.

In the feature fusion module, features of different levels are concatenated as the final output feature vector. The output is a 768-dimensional feature vector. The 768-dimensional feature vector is reshaped into a single 3-channel feature map, which belongs to cloud shadow, cloud, and background.

Based on FCNN, we design a multiscale-3D-CNN structure for cloud and cloud shadow detection.

The main modifications of multiscale-3D-CNN structure are highlighted as follows: (1) By replacing the fully connected layer with global average pooling, the number of learning parameters is heavily reduced, and the multilevel semantic informations is fused. (2) We remove the 2D-convolutional layer, in order to fully explore the joint spatial-spectral correlations feature, we add the joint spectral-spatial information of 3D-convolution layer. (3) We add the multi-scale learning module.

C. CLASSIFICATION AND LOSS MODULE

In this paper, we use the ReLU not only as an activation function [44], but also as the classification function at the last layer of our CNN. The predicted class for ReLU (PC-ReLU) classifier is used to classify features extracted in the

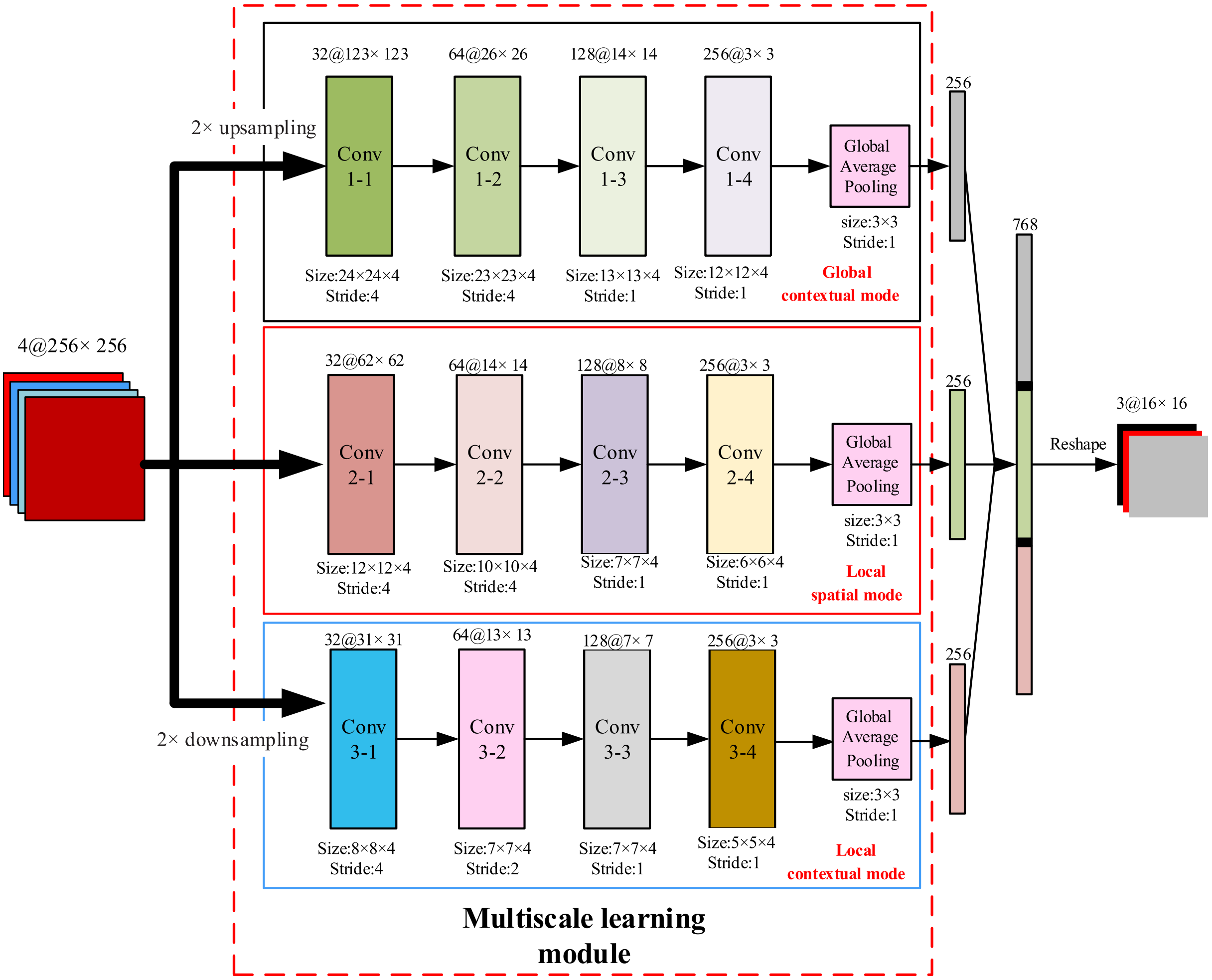


FIGURE 3. The proposed CNN structural for cloud and cloud shadow detection.

multiscale 3D-CNN. Suppose that x is the activation at the penultimate layer of CNN architecture, and y is the weight parameters. Therefore, in this paper, the PC- ReLU classifier can be expressed as:

$$\psi = \arg\max \max(0, \sum_{j}^{M-1} x_j y_j) \qquad (3)$$
$$j \in 1, 2, 3 \cdots M$$

To avoid overfitting problem, in this paper, the mean squared error as the loss function is calculated as follows [45]:

$$L(X_1, Y_i) = \frac{1}{2M} \sum_{i=1}^{M} \|X_1 - Y_i\|_2^2 \qquad (4)$$

where X_1 is an input image. Y_i is the training images.M is the number of samples.

D. ACCURACY EVALUATION

To evaluate the performance of the multi-scale 3D-CNN for cloud and cloud shadow, the overall accuracy (OA), the completeness (recall), and the correctness (precision) are used. The truth cloud and cloud shadows were manually drawn at the ENVI software platform. The completeness represents the ratio of correctly classified cloud and cloud shadow pixels among all true target pixels. The correctness represents the ratio of the correctly classified cloud and cloud shadow pixels and all predicted cloud pixels.

$$\text{completeness} = \frac{TP}{TP + FN} \qquad (5)$$

$$\text{correctness} = \frac{TP}{TP + FP} \qquad (6)$$

where *TP* represents the number of cloud and cloud shadow pixels that have been correctly classified, *TN* represents the number of background pixels that were correctly rejected *FN* represents the number of cloud and cloud shadow pixels classified as background pixels, *FP* represents the number of background pixels classified as cloud and cloud shadow pixels.

We used the overall accuracy, which is the percentage of correctly classified cloud pixels. The OA is defined as [46]:

$$OA = \frac{TN + TP}{T} \times 100\% \qquad (7)$$

where T is the total number of pixels in the test image.

III. EXPERIMENT AND ANALYSIS

A. DATA SET AND TRAINING

In order to evaluate the proposed multiscale-3D-CNN model and examine at its performance under the condition of limited spectral range, we use the two validation datasets (GF-1 WFV validation data and ZY-3 validation data). In order to make full use of information, the four band-combination images are inputted into the proposed multiscale-3D-CNN. The truth cloud and cloud shadows (reference images) are obtained by manually marking on the ENVI software platform. The FLAASH (Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes) method is used for GF-1 WFV and ZY-3 images. The radiometric calibration coefficient of GF-1 WFV and ZY-3 images FLAASH atmospheric correction can be downloaded from http://www.cresda.com/CN/Downloads /dbcs/index.shtml. Since the input size of the proposed multiscale-3D-CNN is 256 × 256, the GF-1 WFV and ZY-3 images of training samples were clipped into the size of 256 × 256.

1) GF-1 WFV VALIDATION DATA

In this paper, we use the public accessible GF-1 WFV validation data set released by Li *et al* [9]. The data set consists of 108 full scenes with the spatial resolution about 16 m. There are four bands (blue, green, red, and near infrared bands) in GF-1multispectral images. The data set is divided into three parts: training, validation, and test, the 48 images are used for training, 40 for validation, and 20 for test. The data set contain typical underlying surface, including mountain areas, urban areas, and ice/snow, etc.

2) ZY-3 WFV VALIDATION DATA

There are three visible bands and one near-infrared band in ZY-3 multispectral images. The data set contains 158 scenes from the satellite ZY-3 (http://clouds.sasmac.cn/query) with the spatial resolution about 5.8m. The 68 images are used for training, 60 for validation, and 30 for test. The data set contains different land-cover types, including mountain areas, water, forest, grassland, ice, snow, and so on.

3) NETWORK TRAINING

The designed multiscale-3D-CNN is implemented by using Python 3.5 on a personal computer with an E3-1505M v6 @3GHz, 32 GB DDR4 memory, and Nvidia Quadro M2200. The experiments were implemented using the software library Tensorflow. We trained the multiscale-3D-CNN for 100 epochs using stochastic gradient descent (SGD) with a minibatch size of 128 patches. The weights in each layer were initialized from a zero-mean Gaussian distribution with a standard deviation of 0.01. The learning rate started from 0.001 and was divided by 10 when the error plateaus, respectively. The weight decay and momentum were set to 0.001 and 0.1.

B. PERFORMANCE COMPARISON OF DEFFERENT CNN

In this section, the multiscale-3D-CNN is designed to detect cloud and cloud shadow, which can mine the semantic features of cloud and cloud shadow at three scales. We evaluate the proposed multiscale-3D-CNN with four different CNN, including 516+3D-CNN (Fig.3 black box), 256+3D-CNN (Fig.3 red box), 128+3D-CNN (Fig.3 blue box), and FCN-8 (https://github.com/shelhamer/fcn. berkeleyvision.org). FCN-8 is a well-known FCN, where the output scoremap is 1/8 × 1/8 size of the input image [40].

To visually compare the cloud and cloud shadow detection results, Fig. 4 shows some cloud and cloud shadow detection results of example images generated by different CNN structure on ZY-3 WFV validation data. Fig. 4 (g) shows ground truth (red region are cloud and gray region represents cloud shadow).

As shown in Fig. 4, we can see that our method can achieve good detection results. As shown in Fig. 4(b-g), the cloud and cloud shadow detection results of our proposed multiscale-3D-CNN framework are the most similar to the ground truth, while the single branch CNN framework have discernible detection errors in cloud shadow regions. In addition, all of the above CNN methods are much more accurate in detecting clouds than cloud shadows because the spectral features of cloud shadow are more similar to the background (water and mountain shadow).

In order to assess the detection performance of different CNN structure, we calculated the overall accuracy, the completeness (recall), and the correctness (precision) at twenty ZY-3 WFV images. Table 1 summarizes the average scores of cloud and cloud shadow mapping with the 516+3D-CNN, the 256+3D-CNN, the 128+3D-CNN, the FCN-8 [38], the proposed multiscale-3D-CNN respectively. From Table 1, it indicates that the proposed multiscale 3D-CNN structure has good metric values for both cloud and cloud shadow. The proposed multiscale 3D-CNN algorithm performs well accurately in cloud detection. The average overall accuracy of cloud detection is as high as 98.34%, and the average precision accuracy and the average recall accuracy are 97.03% and 96.01%, respectively. For cloud shadow detection, the average overall accuracy of cloud detection is up to 98.61%, and the average precision and the average recall are 96.83% and 95.17%, respectively.

C. COMPARISON WITH OTHER METHODS

In this paper, in order to verify the performance of the proposed method, we compared the proposed method to the MFC method [9] (http://sendimage.whu.edu.cn/en/mfc/), Deeplab method [20] (https://github.com/ rishizek/ /tensorflow-deeplab-v3), and PSPnet method [21] (https://hszhao.github.io/projects/pspnet/).

We propose multiscale 3D-CNN to learn the multilevel semantic information of various clouds and cloud shadows. The MFC method is a good performance traditional threshold cloud detection method. The MFC method is

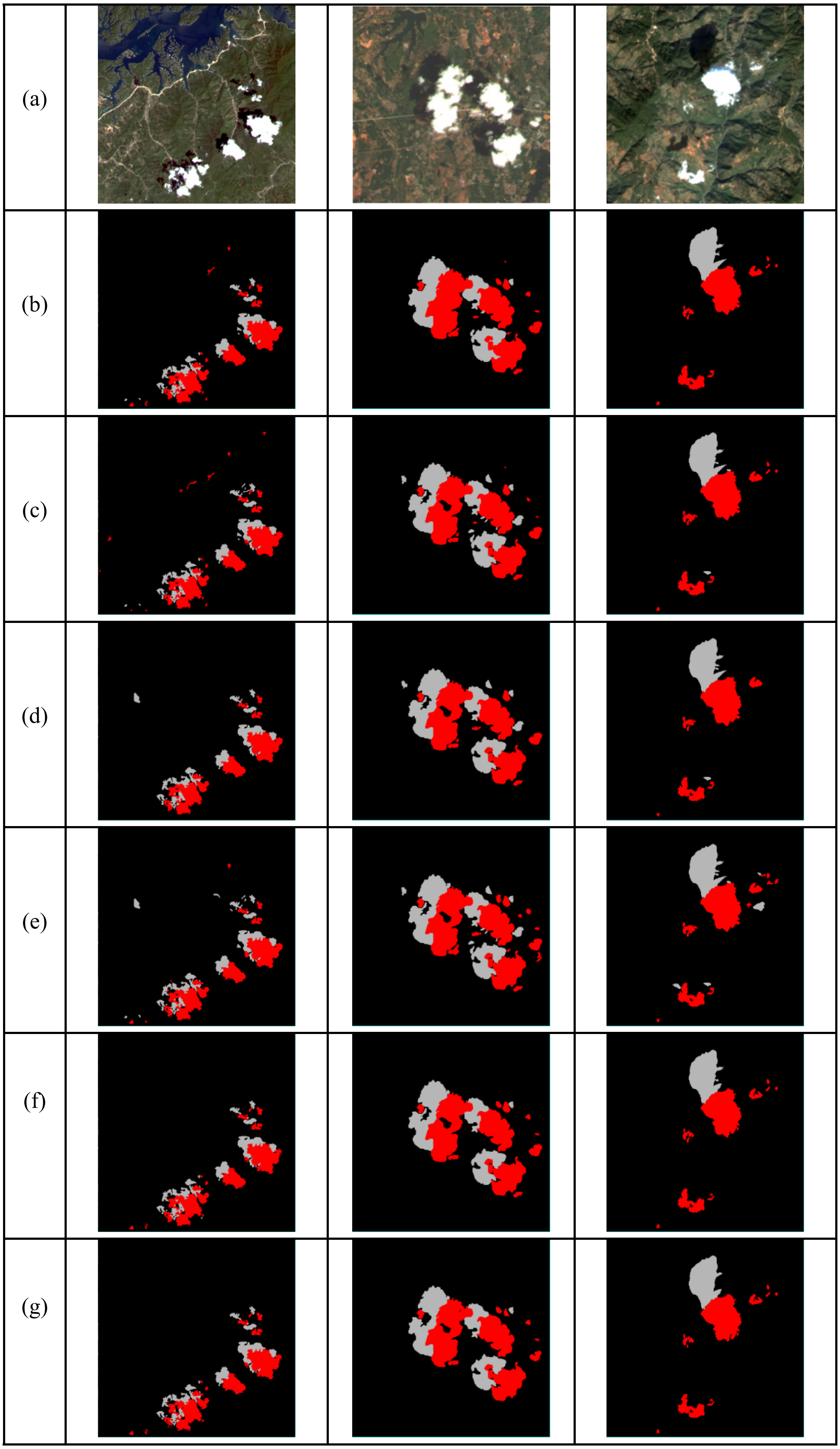


FIGURE 4. Visual comparison of different CNN. (a) Original image. (b) 516+3D-CNN. (c) 256+3D-CNN. (d) 128+3D-CNN. (e) FCN-8. (f) Our multiscale-3D-CNN. (g) Ground truth.

TABLE 1. Detection performance of different CNNs for cloud and cloud shadow.

	Overall accuracy		Precision		Recall	
	Cloud	Cloud shadow	Cloud	Cloud shadow	Cloud	Cloud shadow
516+3D-CNN	0.9231	0.9184	0.9137	0.9017	0.8907	0.8861
256+3D-CNN	0.9321	0.9309	0.9297	0.9241	0.9014	0.9041
128+3D-CNN	0.9417	0.9317	0.9384	0.9253	0.9182	0.9097
FCN-8	0.9367	0.9261	0.9301	0.9159	0.9107	0.9008
Our 3D-CNN	**0.9834**	**0.9861**	**0.9703**	**0.9683**	**0.9601**	**0.9517**

TABLE 2. Detection performance of different methods for cloud and cloud shadow.

	Overall accuracy		Precision		Recall	
	Cloud	Cloud shadow	Cloud	Cloud shadow	Cloud	Cloud shadow
MFC method	0.9598	0.9781	0.9421	0.8514	0.8929	0.8134
Deeplab method	0.9679	0.9702	0.9213	0.8409	0.9246	0.8072
PSPNet method	0.9783	0.9776	0.9401	0.8641	0.9318	0.8293
Our 3D-CNN	**0.9882**	**0.9892**	**0.9716**	**0.9402**	**0.9512**	**0.9234**

based on spectral information similarity to detect each pixel. Both the Deeplab and the PSPNet are good performed deep semantic segmentation networks for natural images.

Fig. 5 shows some of cloud and cloud shadow detection results of example images generated by different methods on GF-1 WFV validation data. In Fig.5, three GF-1 multispectral images are selected, including an image with a large piece of thick clouds and cloud shadow (see Fig.5 (a)), an image with snow/ice only (see Fig.5 (b)), and an image with small piece of thin cloud and cloud shadow (see Fig.5 (c)). Fig.5 (a) is the simplest situation that thin cloud and cloud shadow are large. Both the MFC method and the deep networks can achieve good detection results. In deep networks, the proposed multiscale 3D-CNN achieves a better performance for cloud shadow detection. As for the snow/ice case only, the MFC method cannot distinguish the cloud from the snow. The proposed multiscale-3D-CNN can accurately separate the cloud and snow in GF-1 multispectral images. In addition, we can find that MFC method only detects the thick clouds. As shown in Fig.5 (c), the MFC, the Deeplab, and the PSPNet also missed lots of thin clouds. The proposed multiscale-3D-CNN is more capable of detecting cloud and cloud shadow regions of different types, which is because the proposed method can extract cloud and cloud shadow contextual information of different levels.

To evaluate the effectiveness of the proposed multiscale 3D-CNN model in detecting cloud and cloud shadow regions, we calculated the overall accuracy, the completeness (recall), and the correctness (precision) at twenty GF-1 WFV images. A better cloud and cloud shadow detection algorithm has high values of overall accuracy, precision, and recall. Table 2 summarizes the average scores of cloud and cloud shadow mapping with the MFC method, the Deeplab method, the PSPNet method, and the proposed multiscale-3D-CNN, respectively. From Table 2, it can be seen that our multiscale-3D-CNN algorithm has the highest score in the overall accuracy, precision, and recall. The MFC method has high accuracy in cloud detection, while the accuracy of cloud shadow is poor. The above results show that our multiscale-3D-CNN model delivers more accurate detection results in cloud and cloud shadow detection.

IV. DISCUSSIONS

In the cloud and cloud shadow detection, the ice, snow, water, and mountain shadow are the main sources of noise that decrease the accuracy. In order to further assess the reliability of the proposed method, the performance of the proposed method is discussed when suppressing noise.

A. THE INTERFERENCE OF ICE AND SNOW

The spectrum for ice and snow is similar to that for cloud detection. The shortwave infrared band (SWIR, e.g., Sentinel-2A 1.55–1.75 μm band) is widely applied to distinguish the snow/ice from the clouds. However, it is a challenge to remove these noises from clouds in optical high

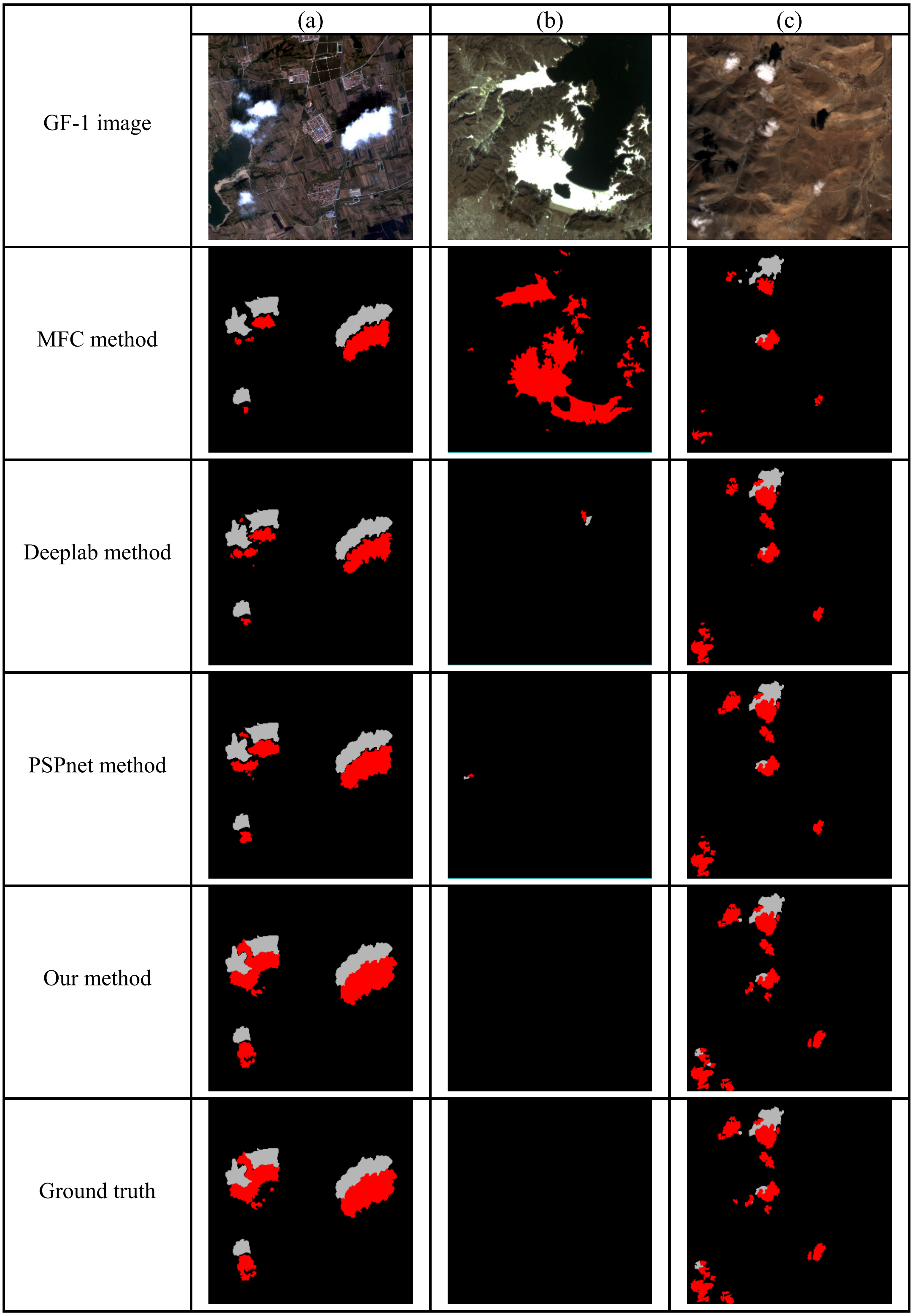


FIGURE 5. Visual comparison of different methods in GF-1 WFV validation data. (a)-(c) are three GF-1 images. (a) Large piece of thick cloud and shadow. (b) snow/ice only cases. (c) Small piece of thin cloud and cloud shadow.

resolution MS images because the high resolution MS images lack SWIR. To testify the reliability of the proposed method when suppressing ice\snow noise, we have compared the proposed method with the ANN method [12], the SVM method [10], and the MFC method [9] on noise ice\snow GF-1 WFV image. Both of the ANN method and the SVM method are good performance traditional based-classification method.

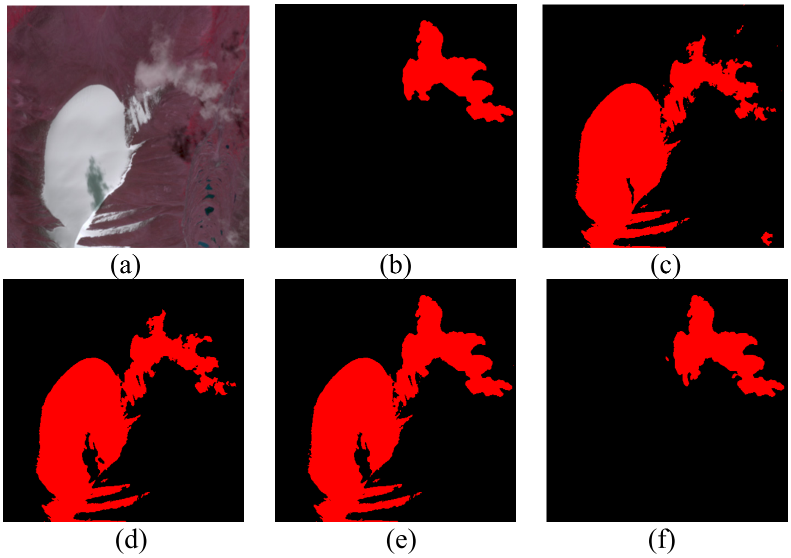


FIGURE 6. The performance comparison when suppressing ice\snow noises. (a)Original image. (b) Ground truth. (c)The ANN method. (d) The SVM method. (e) The MFC method. (f) The proposed method.

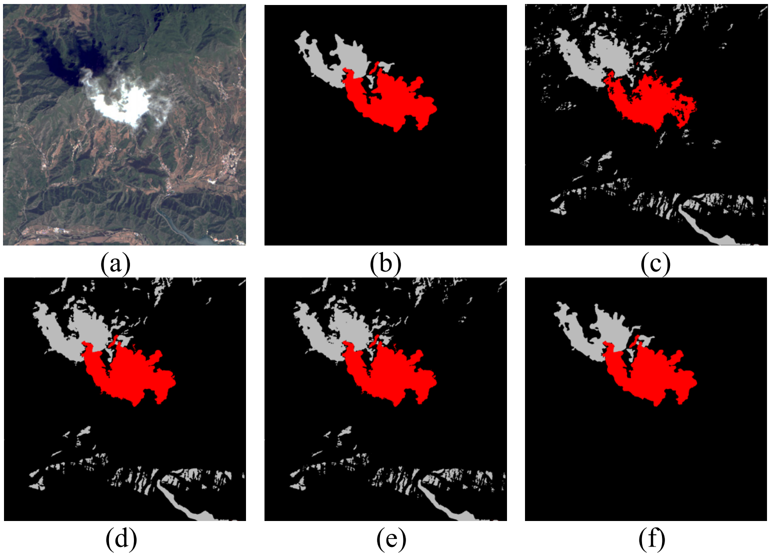


FIGURE 7. The performance comparison when suppressing water and mountain shadow noises. (a)Original image. (b) Ground truth. (c)The ANN method. (d) The SVM method. (e) The MFC method. (f) The proposed method.

Fig. 6 shows the cloud detection results of the different methods. It can be clearly seen that, in cloud-ice\snow co-existing cases, the proposed method can reliably extract cloud regions (see Fig.6(f)). Nevertheless, the ice and snow are misclassified as cloud with high probability in the ANN method, the SVM method, and the MFC method (see Fig. 6(c-e)). The results demonstrate that the traditional cloud detection methods have some limitations in ice\snow covered areas.

B. THE INTERFERENCE OF WATER AND MOUNTAIN SHADOW

The water and mountain shadow strongly influence cloud shadow detection because the spectral characteristics of water and mountain shadow are to be similar to the cloud shadow (see Fig.7(a)). To assess the performance of the proposed method when suppressing water and mountain shadow noise, the ZY-3 image with the noise water and mountain shadow is selected. In addition, we have compared the proposed method with the ANN method [12], the SVM method [10], and the MFC method [9].

A performance comparison of the water and mountain shadow noise scene is shown in Fig.7. For cloud shadow, traditional based-classification methods are impossible to distinguish between water bodies and mountain shadows when only using limited spectra information (see Fig.7(c-d)). However, the proposed method is able to capture discriminative semantic information of cloud shadows and solve the overestimation phenomenon (water and mountain shadow are misclassified as cloud shadow).

V. CONCLUSION

In this paper, the multi-scale 3D-CNN is proposed for cloud and cloud shadow detection using high resolution multispectral images. The proposed multiscale 3D-CNN has greatly improved the accuracy of cloud shadow detection while achieving good accuracy of cloud detection. In addition, even for the low-albedo objects which are easily confused with cloud shadows, such as water, and mountain shadow, the proposed multiscale 3D-CNN can distinguish them from cloud and cloud shadow.

On the one hand, in order to fully explore the joint spatial-spectral correlations feature, a spectral-spatial information of 3D-convolution layer is developed for high resolution multispectral imagery. On the other hand, in order to extract cloud and cloud shadow contextual information of different levels, a multi-scale learning module is designed. In order to make full use of band information, the four band-combination images are inputted into the multiscale 3D-CNN. In addition, the feature fusion module helps to handle features of different levels. Experimental results on two validation datasets (GF-1 WFV validation data and ZY-3 validation data) show that the proposed multiscale 3D-CNN model achieves a high accuracy with limited spectral information and it does not require additional superpixels segmentation processing.

In the future work, we will explore the possibility of using our multiscale-3D-CNN model to detect cloud and cloud shadow from different types of sensor images (such as Sentinel-2, Landsat, and Quickbird).

REFERENCES

[1] S. Le Hégarat-Mascle and C. André, "Use of Markov random fields for automatic cloud/shadow detection on high resolution optical images," *ISPRS J. Photogram. Remote Sens.*, vol. 64, no. 4, pp. 351–366, Jul. 2009.

[2] R. R. Irish, J. L. Barker, S. N. Goward, and T. Arvidson, "Characterization of the landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 10, pp. 1179–1188, Oct. 2006.

[3] X. Zhu, J. Chen, F. Gao, X. Chen, and J. G. Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sens. Environ.*, vol. 114, no. 11, pp. 2610–2623, Nov. 2010.

[4] A. Goodman and A. Henderson-Sellers, "Cloud detection and analysis: A review of recent progress," *Atmos. Res.*, vol. 21, nos. 3–4, pp. 203–228, May 1988.

[5] Y. Han, B. Kim, Y. Kim, and W. H. Lee, "Automatic cloud detection for high spatial resolution multi-temporal images," *Remote Sens. Lett.*, vol. 5, no. 7, pp. 601–608, Jul. 2014.

[6] E. Brocard, M. Schneebeli, and C. Matzler, "Detection of cirrus clouds using infrared radiometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 595–602, Feb. 2011.

[7] C. Huang, N. Thomas, S. N. Goward, J. G. Masek, Z. Zhu, J. R. G. Townshend, and J. E. Vogelmann, "Automated masking of cloud and cloud shadow for forest change analysis using landsat images," *Int. J. Remote Sens.*, vol. 31, no. 20, pp. 5449–5464, Oct. 2010.

[8] Z. Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in landsat imagery," *Remote Sens. Environ.*, vol. 118, pp. 83–94, Mar. 2012.

[9] Z. Li, H. Shen, H. Li, G. Xia, P. Gamba, and L. Zhang, "Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery," *Remote Sens. Environ.*, vol. 191, pp. 342–358, Mar. 2017.

[10] P. Li, L. Dong, H. Xiao, and M. Xu, "A cloud image detection method based on SVM vector machine," *Neurocomputing*, vol. 169, pp. 34–42, Dec. 2015.

[11] A. Hollstein, K. Segl, L. Guanter, M. Brell, and M. Enesco, "Ready-to-use methods for the detection of clouds, cirrus, snow, shadow, water and clear sky pixels in sentinel-2 MSI images," *Remote Sens.*, vol. 8, no. 8, p. 666, Aug. 2016.

[12] M. Hughes and D. Hayes, "Automated detection of cloud and cloud shadow in single-date landsat imagery using neural networks and spatial post-processing," *Remote Sens.*, vol. 6, no. 6, pp. 4907–4926, May 2014.

[13] X. Zhu and E. H. Helmer, "An automatic method for screening clouds and cloud shadows in optical satellite image time series in cloudy regions," *Remote Sens. Environ.*, vol. 214, pp. 135–153, Sep. 2018.

[14] J. D. Braaten, W. B. Cohen, and Z. Yang, "Automated cloud and cloud shadow identification in landsat MSS imagery for temperate ecosystems," *Remote Sens. Environ.*, vol. 169, pp. 128–138, Nov. 2015.

[15] A. Fisher, "Cloud and cloud-shadow detection in SPOT5 HRG imagery with automated morphological feature extraction," *Remote Sens.*, vol. 6, no. 1, pp. 776–800, Jan. 2014.

[16] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018.

[17] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 3431–3440.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.

[19] Z. Yan, M. Yan, H. Sun, K. Fu, J. Hong, J. Sun, Y. Zhang, and X. Sun, "Cloud and cloud shadow detection using multilevel feature fused segmentation network," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 10, pp. 1600–1604, Oct. 2018.

[20] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[21] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.

[22] M. Shi, F. Xie, Y. Zi, and J. Yin, "Cloud detection of remote sensing images by deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 701–704.

[23] F. Xie, M. Shi, Z. Shi, J. Yin, and D. Zhao, "Multilevel cloud detection in remote sensing images based on deep learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3631–3640, Aug. 2017.

[24] Y. Chen, R. Fan, M. Bilal, X. Yang, J. Wang, and W. Li, "Multilevel cloud detection for high-resolution remote sensing imagery using multiple convolutional neural networks," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 5, p. 181, May 2018.

[25] L. Wang, Y. Chen, L. Tang, R. Fan, and Y. Yao, "Object-based convolutional neural networks for cloud and snow detection in high-resolution multispectral imagers," *Water*, vol. 10, no. 11, p. 1666, Nov. 2018.

[26] M. Le Goff, J.-Y. Tourneret, H. Wendt, M. Ortner, and M. Spigai, "Deep learning for cloud detection," in *Proc. 8th Int. Conf. Pattern Recognit. Syst. (ICPRS)*, 2017, pp. 1–6.

[27] M. Wieland, Y. Li, and S. Martinis, "Multi-sensor cloud and cloud shadow segmentation with a convolutional neural network," *Remote Sens. Environ.*, vol. 230, Sep. 2019, Art. no. 111203.

[28] D. Chai, S. Newsam, H. K. Zhang, Y. Qiu, and J. Huang, "Cloud and cloud shadow detection in landsat imagery based on deep convolutional neural networks," *Remote Sens. Environ.*, vol. 225, pp. 307–316, May 2019.

[29] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.

[30] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, p. 67, Jan. 2017.

[31] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

[32] G. Storey, A. Bouridane, and R. Jiang, "Integrated deep model for face detection and landmark localization from 'in the wild' images," *IEEE Access*, vol. 6, pp. 74442–74452, 2018.

[33] Y. Chen, R. Fan, X. Yang, J. Wang, and A. Latif, "Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning," *Water*, vol. 10, no. 5, p. 585, May 2018.

[34] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[35] J. Booth, A. Roussos, S. Zafeiriou, A. Ponniahy, and D. Dunaway, "A 3D morphable model learnt from 10,000 faces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5543–5552,

[36] F. Hu, G.-S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14680–14707, Nov. 2015.

[37] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.

[38] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1797–1801, Oct. 2014.

[39] Y. Zhan, J. Wang, J. Shi, G. Cheng, L. Yao, and W. Sun, "Distinguishing cloud and snow in satellite images via deep convolutional network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1785–1789, Oct. 2017.

[40] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015.

[41] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large Kernel matters–improve semantic segmentation by global convolutional network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4353–4361.

[42] M. Lin, Q. Chen, and S. Yan, "Network in Network," 2013, *arXiv:1312.4400*. [Online]. Available: http://arxiv.org/abs/1312.4400

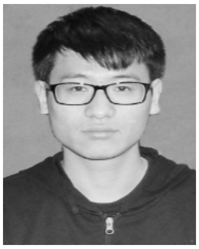
[43] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Proc. 13th ECCV*, 2014, pp. 346–361.

[44] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2013, p. 3,

[45] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolution network," in *Proc. ICML Deep Learn.*, 2015, pp. 1–5.

[46] R. G. Congalton, R. G. Oderwald, and R. A. Mead, "Assessing landsat classification accuracy using discrete multivariate analysis statistical techniques," *Photogramm. Eng. Remote Sens.*, vol. 49, no. 12, pp. 1671–1678, 1983.

[47] Y. Chen, L. Tang, X. Yang, R. Fan, M. Bilal, and Q. Li, "Thick clouds removal from multitemporal ZY-3 satellite images using deep learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, early access, 2019, doi: 10.1109/jstars.2019.2954130.



YANG CHEN received the M.Eng. degree from Liaoning Technical University, Fuxin, China, in 2019. He is currently pursuing the Ph.D. degree with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He was a jointly educates student with the China Academy of Surveying and Mapping, Beijing, China, from 2017 to 2019. His research interests include deep learning and intelligent remote sensing information processing.

LULIANG TANG received the Ph.D. degree from Wuhan University, Wuhan, China, in 2007. He is currently a Professor with Wuhan University. His research interests include space–time GIS, deep learning, GIS for transportation, and change detection.

ZIHAN KAN received the B.Sc. degree from Wuhan University, Wuhan, China, 2014. She is currently pursuing the Ph.D. degree with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. Her research addresses the issue of the interaction between human behaviors and the environment in the context of urban road networks.

AAMIR LATIF received the B.S. degree from the Lasbela University of Agriculture, Water and Marine Sciences, Pakistan, in 2011, and the M.S. degree from the Coastal and Ocean Management Institute, Xiamen University, in 2015. During his M.S., he worked on the Land use and land cover changes, and its ecological consequences on the coastal area of Karachi, Pakistan. He is currently pursuing the Ph.D. degree in ecology with the Institute of Geographic Sciences and Natural Resources Research, University of Chinese Academy of Sciences. His research interests focus on the vegetation dynamics and climate change on the Tibetan Plateau using remote sensing data.

XIUCHENG YANG received the B.Sc. degree in GIS from Sun Yat-sen University, China, in 2011, the M.S. degree in remote sensing and GIS, in 2015, from Peking University, China, and the Ph.D. degree in geoinformation from the University of Strasbourg, France, in 2018. He has authored about 30 articles in remote sensing journals and conferences. His research interests include 3D reconstruction of an urban scene from aerial imagery and remote sensing image analysis.

MUHAMMAD BILAL received the B.Sc. degree (Hons.) in space science (remote sensing/GIS, atmospheric science) from the University of the Punjab, Lahore, Pakistan, in 2008, the M.S. degree in meteorology (specialization in remote sensing and GIS) from COMSATS University Islamabad, Pakistan, in 2010, and the Ph.D. degree in photogrammetry and remote sensing from The Hong Kong Polytechnic University (PolyU), Hong Kong, in 2014. From 2014 to 2017, he worked at the PolyU as a Postdoctoral Fellow with the Nanjing University of Information Science and Technology (NUIST), Nanjing, China, where he joined a Professor, in October 2017. In June 2018, he was awarded the special title Distinguished Professor by the Jiangsu Province, China. He has authored widely on these topics in top *Aerosol Remote Sensing* journals. His research interests include the applications of satellite remote sensing for air quality monitoring, aerosol retrieval algorithms, and atmospheric correction.

QINGQUAN LI received the Ph.D. degree in geographic information system (GIS) and photogrammetry from the Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1998. He is currently a Professor with Shenzhen University, Guangdong, China, and also with Wuhan University, Wuhan. His research areas include dynamic data modeling in GIS, surveying engineering, and intelligent transportation systems.