Cloud Detection in Remote Sensing Images Based on Multiscale Features-Convolutional Neural Network

Zhenfeng Shao, Yin Pan, Chunyuan Diao, and Jiajun Cai

***Abstract*— Cloud detection in remote sensing images is a challenging but significant task. Due to the variety and complexity of underlying surfaces, most of the current cloud detection methods have difficulty in detecting thin cloud regions. In fact, it is quite meaningful to distinguish thin clouds from thick clouds, especially in cloud removal and target detection tasks. Therefore, we propose a method based on multiscale features-convolutional neural network (MF-CNN) to detect thin cloud, thick cloud, and noncloud pixels of remote sensing images simultaneously. Landsat 8 satellite imagery with various levels of cloud coverage is used to demonstrate the effectiveness of our proposed MF-CNN model. We first stack visible, near-infrared, short-wave, cirrus, and thermal infrared bands of Landsat 8 imagery to obtain the combined spectral information. The MF-CNN model is then used to learn the multiscale global features of input images. The high-level semantic information obtained in the process of feature learning is integrated with low-level spatial information to classify the imagery into thick, thin and noncloud regions. The performance of our proposed model is compared to that of various commonly used cloud detection methods in both qualitative and quantitative aspects. Compared to other cloud detection methods, the experimental results show that our proposed method has a better performance not only in thick and thin clouds but also in the entire cloud regions.**

***Index Terms*— Cloud detection, convolutional neural network (CNN), deep learning, multiscale features (MF), remote sensing images.**

I. Introduction

With the rapid development of remote sensing technology, remote sensing images have been widely used in the fields of earth observation [1], resource survey, natural disaster prediction, environmental pollution monitoring, etc. However, because of the significant influence of atmospheric density and cloud layer change on remote sensing processes, most of the remotely sensed images encounter different levels of cloud contamination. Global cloud data from the international satellite cloud climatology project-flux data (ISCCP-FD) show that more than 66% of the earth's surface area is often covered with cloud [2]. The attenuation and even loss of some image information caused by cloud not only reduces the quality and utilization of remote sensing data dramatically but also causes the difficulty of the analysis and application of remote sensing images [3], [4]. In order to improve the usability of remote sensing images, it is indispensably essential to conduct cloud detection before any task-specific remote sensing analysis.

In recent years, a large number of cloud detection methods have been proposed. These methods can be roughly divided into two categories: threshold-based methods and classification-based methods. Threshold-based methods are to determine proper thresholds of spectral reflectance or brightness temperature via specific channels for different sensors to identify cloud regions. Over the years, several cloud detection algorithms have been developed, such as ISCCP cloud mask algorithm, AVHRR Processing scheme Over Clouds, Land, and Ocean cloud mask algorithm, Clouds from the Advanced Very High Resolution Radiometer cloud mask algorithm, CO_2 slicing and Moderate Resolution Imaging Spectroradiometer (MODIS) cloud mask algorithm [5]–[9]. These threshold methods are widely adopted for cloud detection because of their high accuracy and reliable robustness. However, for complex land surface regions of various cloud cover types, it is difficult to identify proper thresholds to detect cloud accurately.

For moderate-spatial-resolution and low-spectral-resolution sensors like Landsat, many automated cloud detection algorithms have been developed based on a single Landsat image. Iris [10] and Iris *et al.* [11] had proposed the Automated Cloud Cover Assessment system to estimate the percentage of clouds for each Landsat scene. Oreopoulos *et al.* [12] assessed the performance on Landsat-7 images of a modified version of a cloud-masking algorithm originally developed for clear-sky compositing of MODIS images at northern midlatitudes. Huang *et al.* [13] proposed an automated masking algorithm for cloud and cloud shadow detection using clear forest pixels as a reference to define cloud boundaries in a spectral-temperature space and predicting the shadows according to cloud height and sun illumination geometry. However, due to the lack of ability to distinguish warm clouds or snow/ice in high latitude areas, Zhu and Woodcock [14] proposed Function of Mask (Fmask) algorithm to acquire the cloud mask

Manuscript received July 15, 2018; revised November 19, 2018 and December 13, 2018; accepted December 19, 2018. Date of publication January 24, 2019; date of current version May 28, 2019. This work was supported in part by National Key R&D Program of China under Grant 2018YFB0505400 and in part by the Natural Science Foundation of China under Grant 61671332, Grant 41771452, Grant 41771454, and Grant 41890820. *(Corresponding author: Yin Pan.)*

Z. Shao, Y. Pan, and J. Cai are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: shaozhenfeng@whu.edu.cn; yinpanwhu@163.com; cai_jiajun@foxmail.com).

C. Diao is with the Department of Geography and GIScience, University of Illinois at Urbana–Champaign, Urbana, IL 61801 USA (e-mail: chunyuan@illinois.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2018.2889677

through the scene-based threshold and probability mask, and calculate the cloud shadows by object matching. Nevertheless, the thresholds calculated based on lots of absolutely clear-sky pixels of single Landsat imagery in the Fmask algorithm are not constant, so limited by spectral domains, Fmask algorithm may not work correctly for complex surfaces, such as urban, snow, and mountain [15], [16]. Zhang *et al.* [17] utilized haze optimized transform (HOT) response level as an alternative measure of cloud optical depth and estimated the variation of incident visible radiation reduction in the corresponding shadow patch by the spatial distribution of the HOT response in a given cloud patch. Although this method realized spatial matching between cloud and cloud shadow objects automatically, the HOT threshold selected manually would affect the detection results directly.

A series of significant successes in the field of image classification [18] has been achieved due to the breakthrough in pattern recognition, machine learning, and computer vision [19]. There is no need to determine large-scale thresholds or optimal eigenvalues for classification-based methods, which have been increasingly applied to detect cloud-cover regions [20]. Movia *et al.* [21] combined some effective indexes with unsupervised classification methods to realize the shadow detection and removal in RGB VHR images for land-use analysis. Surya and Simon [22] performed color transformation and generated a ratio image using the spectral image rationing technique, and then applied the fuzzy C-means clustering method to detect clouds. Tian *et al.* [23] proposed probabilistic neural network classifiers to track cloud temporal changes in a sequence of images by utilizing the temporal contextual information, and Maximum likelihood criterion was adopted in both training and updating schemes. Vivone *et al.* [24] introduced a novel penalty term within the classical maximum *a posteriori* probability–Markov random field approach to reducing the high misclassification rate for pixels close to cloud edges. Through leveraging geometric, texture or color information of the tagged images, classification-based cloud detection methods have been continuously advanced in recent years to improve the training process for more accurate cloud detection [25]–[30].

In comparison with the threshold methods, the classification-based cloud detection methods have distinct advantages in automating the detection process and improving the detection performance. In [31], the spectral, texture, and structural features of each pixel had been extracted artificially, and then, the integrating feature information was used for realizing cloud and cloud shadow detection in fuzzy autoencoder neural network. However, this method must select the favorable features manually and require a large number of feature calculations to achieve good detection accuracy. In order to extract features automatically, convolutional neural network (CNN) [32]–[38] has recently been applied in target detection, which regards pixels or segmented super-pixels as research objects. In the above methods, the features of cloud, cloud shadow, or snow are automatically extracted by these different models, and the model parameters are optimized through training samples to realize the detection task. In addition, through the process of sample training, the complex nonlinear relationship between the label value and the input image can be appropriately constructed. More detailed information on existing cloud detection methods based on deep learning is shown in Table I.

TABLE I
SUMMARY OF DIFFERENT CLOUD DETECTION METHODS BASED ON DEEP LEARNING

Deep learning method	Based on pixels or objects	Detect contents
Fuzzy AutoEncode [31]	pixels	cloud, cloud shadow
Optimizing CNN [32]	pixels	cloud
CNN (based on CIFAR-10) [33]	objects	cloud
CNN (based PSPNet) [34]	pixels	cloud, cloud shadow
Deep Convolutional Network [35]	pixels	cloud, snow
Residual CNN [36]	pixels	haze removal
CNN [37]	objects	cloud
Two-Branch CNN [38]	objects	thin cloud, thick cloud

However, the task of distinguishing thin clouds from thick clouds, which is crucial for thin cloud removal and other image analysis, has not been paid enough attention by the majority of existing cloud detection methods. The thin clouds are translucent, and their spectral information always mixes with the underlying surface, which brings challenges to cloud detection. Xie *et al.* [38] regarded the super-pixels as research objects, and then used the center of super-pixels in the image block as a basic unit to conduct the training and classification based on a two-branch CNN model. Although this method can be used to detect thin cloud regions, the features extracted in this model only contain the information of local neighborhood within small image patch, which ignores the multiscale contextures in the classification process. In addition, the detection efficiency of thin and thick cloud is too dependent on image segmentation accuracy.

In response to the deficiencies of the cloud detection methods discussed above, in this paper, we stack the visible, near-infrared (NIR), shortwave infrared (SWIR), thermal infrared (TIR) bands of the Landsat 8 sensor to create the combination images of 10 bands. Then, we take each pixel of combination images as the basic research unit and construct the multiscale features-convolution neural network (MF-CNN). In our model, the multiscale global features of the thick and thin clouds will be extracted automatically, which contains the global context information of various aggregation levels. In addition, the high-level semantic information in different scales produced by feature learning will integrate with corresponding low-level spatial information in the process of classification, and the contextures at multiscale will assist the detection for thick and thin clouds.

The remainder of this paper is organized as follows. In Sections II and III, data source and proposed methodology for cloud detection are described. Section IV presents the cloud detection experiment and corresponding results. Further

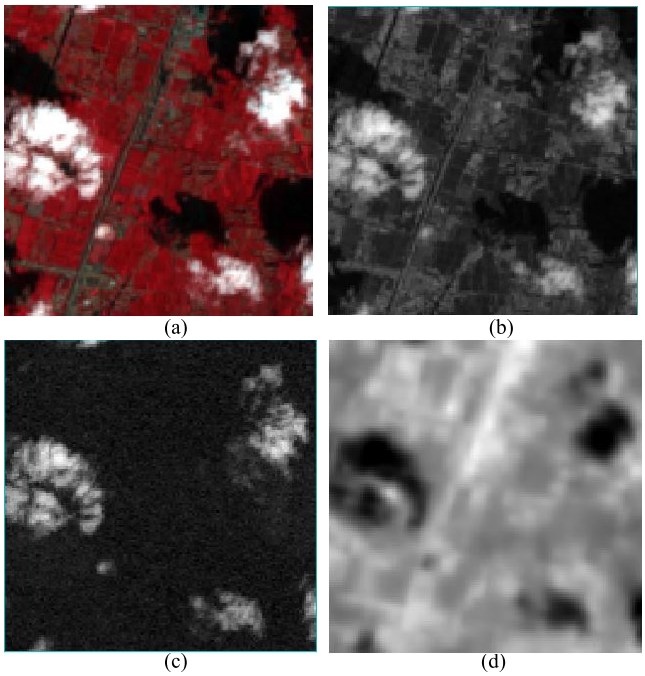


Fig. 1. Image display with different bands. (a) NIR-R-G combination image. (b) SWIR image. (c) Cirrus image. (d)TIRS image.

discussions are arranged in Section V. Finally, Section VI gives a summary of our work.

II. Materials

Landsat 8 satellite remote sensing imagery is widely used in the image analysis because of its multiple bands, high resolution, wide coverage, and short revisits. Nevertheless, the frequent cover of the clouds greatly limits the usage of captured images. Therefore, cloud detection is an essential part of the image processing for Landsat 8 images. The Landsat 8 satellite images can be freely downloaded from the USGS website.

Considering the thin cirrus clouds are difficult to detect just by visible and thermal infrared bands, Gao and Kaufman [39] and Zhu *et al.* [40] proposed that the strong water vapor band (1.36–1.38 μm) is very effective in separating the thin cirrus from the ground surface. Therefore, in order to improve the separation of thick and thin clouds, the cirrus band is indispensable. In addition, as the TIR and SWIR bands contain sufficient spectral and temperature-related information of thick and thin clouds, we stack the visible, NIR, SWIR, cirrus, and TIR bands to obtain the combination images with 10 bands. The spatial resolution of visible, NIR, SWIR bands is 30 m, while the thermal infrared bands have a spatial resolution of 100 m, which will be resampled to 30 m. The correction level of the experimental images is Level 1A. Table II shows the information of the selected bands.

As shown in Fig. 1(a), the thick and thin cloud pixels in the NIR and visible bands have distinct spectral characteristics, which can be easily separated from the background. As for the SWIR band, the similar spectral characteristics are displayed in Fig. 1(b). Besides, Fig. 1(c) shows that the cirrus band is apparently valid for the detection of cirrus clouds. Due to the lower temperature of the cloud regions, the thick and thin clouds correspond to the dark regions in the TIR band image [Fig. 1(d)]. In order to leverage various spectral and temperature information contained in Landsat 8 images, the above-described bands will be used in our cloud detection.

TABLE II
Information of Combination Image Bands

Band number	Band width (μm)	Band name	Spatial resolution(m)
Band1	0.43-0.45	Coastal	30
Band2	0.45-0.51	Blue	30
Band3	0.53-0.59	Green	30
Band4	0.64-0.67	Red	30
Band5	0.85-0.88	NIR	30
Band6	1.57-1.65	SWIR 1	30
Band7	2.11-2.29	SWIR 2	30
Band9	1.36-1.38	Cirrus	30
Band10	10.6-11.19	TIRS 1	30(resampled)
Band11	11.5-12.51	TIRS 2	30(resampled)

In our work, a total of 107 Landsat 8 satellite images are downloaded from the USGS website, which is mainly selected according to different cloud coverages and the underlying surface. The acquisition dates of these images are distributed between March 2014 and September 2018. In fact, due to the wide coverage of the Landsat 8 satellite, the entire image of Landsat 8 satellite may not be fully covered by clouds. Most of the images will have a wide range of clear pixels such as vegetation, bare soil or water. However, for the follow-up experiments of thick and thin cloud detection, this kind of wide-ranging continuous noncloud pixels will make the training data unbalanced completely, thus failing to achieve better detection. Therefore, for each scene, we select the cloud containing regions or representative underlying surfaces carefully, and the size of which varies from 300 by 400 pixels to 900 by 1000 pixels.

According to the needs of the subsequent experiment, the 107 regions from different Landsat 8 satellite images need to be divided into a test set and a training set. The training data are derived from 63 images, and the test set is from the remaining 44 images. In the process of dividing data sets, both of them take into account various cloud coverage information and underlying surface environment. Most of their images contain both cloud and noncloud regions. Cloud regions include small, medium, and large size clouds; the underlying surface environment includes urban buildings, vegetation, agricultural, water, and snow.

Limited by the performance of hardware devices, the 63 band-combination images of the training set in our study are divided into image blocks with the size of 128×128. A total number of 1236 Landsat 8 image blocks constitute the training data set. Similarly, the segmentation of 44 combination images in the test set produces 806 image blocks with the size of 128×128. In order to prove the effectiveness of the proposed method, the two data sets include thin and thick clouds with different cloud contents, as well as clear pixels that are easily confused with clouds, such as snow-covered surfaces

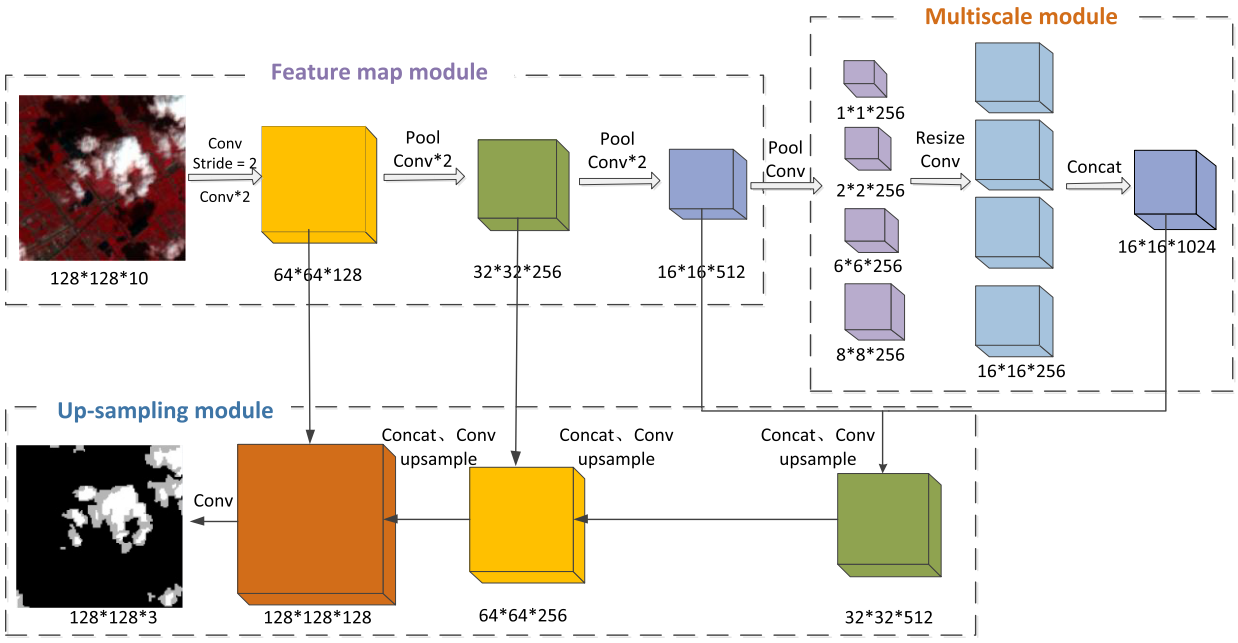


Fig. 2. Model structure of MF-CNN.

TABLE III
DETAILED INFORMATION OF IMAGES

Image parameters	Landsat 8
Product level	Level 1
Number of bands	10
Spatial resolution	30 m
Block size(pixel)	128*128
Number of blocks in training set	1236 (from 63 Landsat images)
Number of blocks in test set	806 (from 44 Landsat images)

and bright buildings. The summarized information of these images is provided in Table III.

III. METHODOLOGY

Deep learning has recently achieved a huge breakthrough in artificial intelligence and computer vision areas. As a deep learning method, CNN has improved the performance dramatically for a wide range of computer vision tasks such as image classification, saliency detection, object recognition, and semantic segmentation [41], [42]. Fully CNN (FCNN), proposed by Shelhamer *et al.* [43], can conduct intensive prediction without fully connected layers. This structure enables the segmentation map to generate images with any size and also improves the processing speed compared with the traditional CNN. Based on the principle of FCNN, MF-CNN is designed to extract the multiscale global features to characterize thick and thin clouds in the combination images of Landsat 8 satellite. In addition, the integration of low-level spatial information and high-level semantic information provides more detailed information of clouds so as to realize the cloud detection in pixel level more accurately.

A. Structure of MF-CNN

As shown in Fig. 2, the overall model structure is divided into three parts, which contains feature map module, multi-scale module, and up-sampling module. Each feature map in the model is a 3-D array with the size of height × width × depth, where height and width are spatial dimensions, and depth is the feature or channel dimension. In this paper, *Conv*# denotes a convolutional layer, *Maxpool*# denotes a max pooling layer, and *Avepool*# is an average pooling layer. A nonlinear rectified linear unit function is denoted by *Relu*#, a batch normalization layer is by Bn#, and a dropout layer is by *Dropout*#.

In the feature map module, the first layer is the combined image block with the size of 128 128 10, which is regarded as the input data. The *Conv**2 means that there are two convolutional layers, and the structure of this module can be described concisely as

Input(128 × 128 × 10)—Conv1(64 × 64 × 64)—Relu1—Conv2(64 × 64 × 96)—Bn2—Relu2—Conv3(64 × 64 × 128)—Bn3—Relu3—Maxpool—Conv4(32 × 32 × 192)—Bn4—Relu4—Conv5(32 × 32 × 256)—Bn5—Relu5—Maxpool—Conv6(16 × 16 × 256)—Bn6—Relu6—Conv7(16 × 16 × 512)—Bn7—Relu7.

In the module mentioned above, the size of filter is 3 × 3 in all convolutional layers, the stride of which is two pixels in *Conv1* and one pixel in other convolutional layers. The max-pooling layers perform max pooling over 2 × 2 spatial neighborhoods with a stride of two pixels on the output of the convolution layers. Therefore, through this module, the height and width of the output layer are one-eighth of the combined image block.

In the multiscale module, the output features of *Relu*7 (16 × 16 × 512) is regarded as the input layer. To obtain the multiscale global features that are characteristic of thick and thin clouds in combination images, we set different sizes of average pooling filter. Each average pooling layer is followed

by convolutional layers and bilinear interpolation (denoted by *Bl*#) so that four parallel feature maps with the same height and width of *Relu*7 can be obtained. We can describe this process as

$$\left.\begin{array}{l}\text{Relu—Avepool(16)—Conv\_m1}(1\times 1\times 256)\text{—Bl}\\ \text{—Conv\_m1}_2(16\times 16\times 256)\\ \text{Relu—Avepool(8)—Conv\_m2}(2\times 2\times 256)\text{—Bl}\\ \text{—Conv\_m2}_2(16\times 16\times 256)\\ \text{Relu—Avepool(4)—Conv\_m4}(4\times 4\times 256)\text{—Bl}\\ \text{—Conv\_m4}_2(16\times 16\times 256)\\ \text{Relu—Avepool(2)—Conv\_m8}(8\times 8\times 256)\text{—Bl}\\ \text{—Conv\_m8}_2(16\times 16\times 256)\end{array}\right\}$$
$$\text{concat}(16\times 16\times 1024).$$

The sizes of parallel average pooling filters are 16×16, 8×8, 4×4, 2×2 respectively, and the stride of corresponding filter is the same as its height. For each branch, the stride of filter in convolutional layers is one pixel, and the filter size of the first convolutional layer is 1×1, and the other is 3×3. Finally, by concatenating (denoted by Concat#) the four output convolutional layers, the global feature maps containing multiscale information can be obtained.

After the learning process of feature map module and multiscale module, the low-level spatial information of combination images has been transformed into high-level semantic information. In order to achieve better detection accuracy of thick and thin clouds than the model with only high-level information, the high-level semantic and low-level spatial information are applied simultaneously in the up-sampling module. The Concat# layer in multiscale module is connected with the feature layer Relu7($16\times 16\times 512$). Through the bilinear interpolation, the up-sampling feature map (denoted by Upsampling#) will become two times larger, and the process can be described as

$$\begin{array}{l}\text{Concat}(16\times 16\times 1024)+\text{Relu7}(16\times 16\times 512)\text{—Conv\_up1}\\ (16\times 16\times 512)\text{—Relu—Upsampling1}(32\times 32\times 512)\\ +\text{Relu5}(32\times 32\times 256)\text{—Conv\_up2}(32\times 32\times 256)\\ \text{—Bn—Relu—Upsampling2}(64\times 64\times 256)+\text{Relu3}(64\\ \times 64\times 128)\text{—Conv\_up3}(64\times 64\times 128)\text{—Bn—Relu}\\ \text{—Upsampling3}(128\times 128\times 128)\text{—Dropout}\\ \text{—Conv\_out}(128\times 128\times 3).\end{array}$$

In this module, the symbol "+" represents the connection between the layers, and the size of filter in the convolutional layers is 3×3. The final up-sampled result is followed by a dropout layer to avoid over-fitting, and through a convolutional layer with 1×1 filter, the classification result of thick and thin cloud in pixel level will be obtained.

B. Training of MF-CNN

In the training stage, the combination images with spectral information of nine bands in the training set are regarded as input data to train MF-CNN model. First, through the parameters setting of initial weight in the whole model, we can extract the feature map of input image with the size of one-eighth of the input size in the feature map module. Combined with multiscale module, the extracted features are subsampled to different global scales. Through the interpolation and concatenation, the multiscale global information is obtained. Finally, the combination of high-level semantic and low-level spatial information in the up-sampling module accomplishes the initial classification of each pixel. To minimize the difference between the MF-CNN-based classification result and ground truth the Adam (adaptive moment estimation) optimizer [44] is used to adjust the model parameters dynamically.

Multiple parameters in the MF-CNN need to be initialized in the training stage. The filter weights of each convolutional layer are initialized by drawing randomly from Gaussian distribution with mean of zero and standard derivation of 0.01, and the biases in convolutional layers are initialized with the constant of 0.1. We use Adam optimizer with initial learning rate as 0.001. The exponential decay rates for the first and second moment estimates are set to 0.9 and 0.999, respectively. In order to avoid over-fitting, the dropout rate is set to 0.5, which means that half of the features are reduced randomly during the training stage. The batch size is chosen as 12, and the training of model is up to 50×200 iterations.

Through the trained parameters of the MF-CNN model, the multiscale global features of any combination image with the size of $128\times 128\times 10$ can be extracted. Combined the extracted high-level semantic information with the low-level spatial information, the input image is classified into thick cloud, thin cloud, and noncloud in pixel level.

C. Accuracy Assessment

This paper evaluates the detection performance of thick and thin clouds from different aspects, both qualitatively and quantitatively. Qualitative assessment is to visualize the cloud classification masks generated by the MF-CNN model, and then compare the performance of different methods from visual interpretation. As for the quantitative assessment, the effectiveness of different methods is described in three aspects: thin cloud, thick cloud, and the entire cloud (thin cloud and thick cloud) detection accuracy. First of all, the precision and recall values of thick and thin clouds are calculated, respectively, and then the indexes of right rate (RR), error rate (ER), false alarm rate (FAR), and the ratio of RR to ER (RER) are applied to evaluating the detection accuracy of all cloud regions (including thick cloud and thin cloud)

$$\text{precision} = \text{CP/DP} \tag{1}$$
$$\text{recall} = \text{CP/ GN} \tag{2}$$

where CP is the number of pixels correctly detected as cloud, DP is the total number of pixels detected as cloud, and GN is the number of cloud pixels in ground truth. For each cloud detection method, the precision and recall values of thick and thin cloud will be calculated separately. The higher precision and recall values indicate better performance of cloud detection methods. There may be some tradeoffs between precision and recall values in evaluating our results. In this paper, we use the F_Score evaluation index to integrate these two measures

$$\text{F\_Score} = 2\times \text{precision}\times \text{recall}/(\text{precision}+\text{recall}) \tag{3}$$

where F_Score combines the evaluation results of precision and recall, when F_Score is higher, it shows that the detection method is more effective.

In addition, the algorithm performance for the entire cloud detection also needs to be evaluated. The definition of RR is the same as the recall. The ER and the FAR are defined as follows:

$$\mathrm{ER} = (\mathrm{CN} + \mathrm{NC})/\mathrm{TN} \tag{4}$$

$$\mathrm{FAR} = \mathrm{NC}/\mathrm{GN} \tag{5}$$

where CN is the number of cloud pixels detected as noncloud pixels, NC is the number of noncloud pixels detected as cloud pixels, and TN is the total number of pixels in the input image. The higher values of RR and lower values of ER or FAR indicate a better cloud detection method. However, the above situation is too ideal. Most of the cases are high RR accompanied by high FAR, or low ER accompanied by low RR. Thus, it may not be comprehensive to evaluate the effectiveness of the model using only one of them. Here we use RER, which is defined as the RER, to integrate the measures of RR and ER in model evaluation

$$\mathrm{RER} = \mathrm{RR}/\mathrm{ER}. \tag{6}$$

IV. Experiments and Results

The training procedures are conducted by using python on the PC with Intel Xeon CPU E3-1240 at 3.70 GHz and Quadro K620 VGA compatible controller, and the experiment is implemented under the TensorFlow framework.

A. Experiment Settings

As described in Section II, this paper uses a total of 2042 blocks of 10-band combined images with 128 pixels by 128 pixels for the experiment, containing various underlying surface information. Among them, 1236 image blocks in the training set are obtained from the 63 Landsat 8 satellite images, and 806 image blocks in the test set are obtained from the other 44 scene images.

We define the translucent cloud as thin cloud, and the cloud that completely covers the surface information is defined as thick cloud. The ground truth of thick, thin, and noncloud pixels is marked manually. In order to evaluate the effectiveness of the MF-CNN model in detecting thick cloud and thin cloud of different shapes and sizes, various levels of cloud coverage images are contained in the training set and test set. At the same time, in addition to vegetation and water, clear pixels (noncloud pixels) in the two data sets include underlying surfaces that are easily confused with clouds, such as bright buildings and snow-covered ground.

To illustrate that the proposed method can avoid the bright ground surface (clear pixels) being misclassified into cloud pixels, the proportion of cloudless image blocks in the test data set is slightly higher, while the distribution of image blocks with different cloud coverage ratios are roughly similar in the two data sets. The details of the image blocks distribution are summarized in Table IV.

In the training set, there are 1 816 966 thick cloud pixels, 1 885 608 thin cloud pixels and 16 548 050 clear pixels. The test set contains 903 698 thick cloud pixels, 1 057 859 thin cloud pixels, and 11 243 947 clear pixels. It can be seen that in each data set, the proportion of thin cloud and thick cloud pixels is not much different. After analyzing the number distribution of image blocks with different cloud coverages in the two data sets, we calculate the proportion distribution of the image blocks with different coverages of thin clouds and thick clouds, respectively (shown in Fig. 3).

TABLE IV
IMAGE DISTRIBUTION OF CLOUD COVERAGE IN TRAINING AND TEST DATA SETS

Cloud coverage	Number in training set	Ratio in training set	Number in test set	Ratio in test set
rate = 0%	294	23.8%	311	38.6%
0% < rate < 10%	300	24.2%	177	22.0%
10% < rate ≤ 20%	164	13.3%	80	9.9%
20% < rate ≤ 30%	155	12.6%	65	8.0%
30% < rate ≤ 40%	113	9.1%	60	7.5%
40% < rate ≤ 100%	210	17.0%	113	14.0%

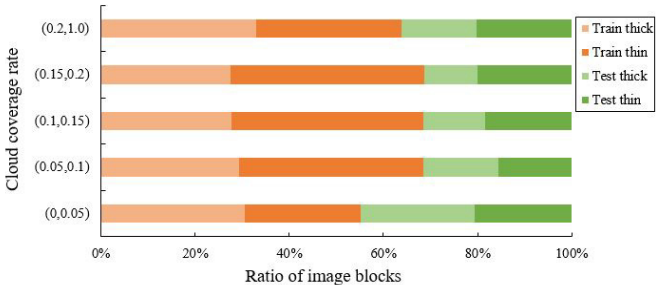


Fig. 3. Ratio of image blocks with different thick and thin cloud coverages.

In this paper, we propose the MF-CNN model to learn the multiscale global features of various thick and thin clouds by integrating high-level semantic information with low-level spatial information. In the training stage, the parameters in the MF-CNN model are optimized through the Adam optimizer. The trained model is utilized to classify the combination images into thin, thick, or no-cloud regions in the test stage. To evaluate the effectiveness of our proposed MF-CNN model, we replace the multiscale module with the conventional convolutional layer and abandon low-level spatial information with different scales. The above structure based on MF-CNN is regarded as self-contrast model, which will not contain any multiscale feature information. Besides, the FCNN model mentioned in [35] is also applied to comparison experiment, which obtains image features through multiple convolutional and pooling layers, and then different feature layers are up-sampled to the same size of original input image directly, and the final concatenated features are applied to distinguish cloud and snow by softmax classifier in the end. In addition to the methods discussed above, SVM [45] and RF [46] are also compared with our proposed method.

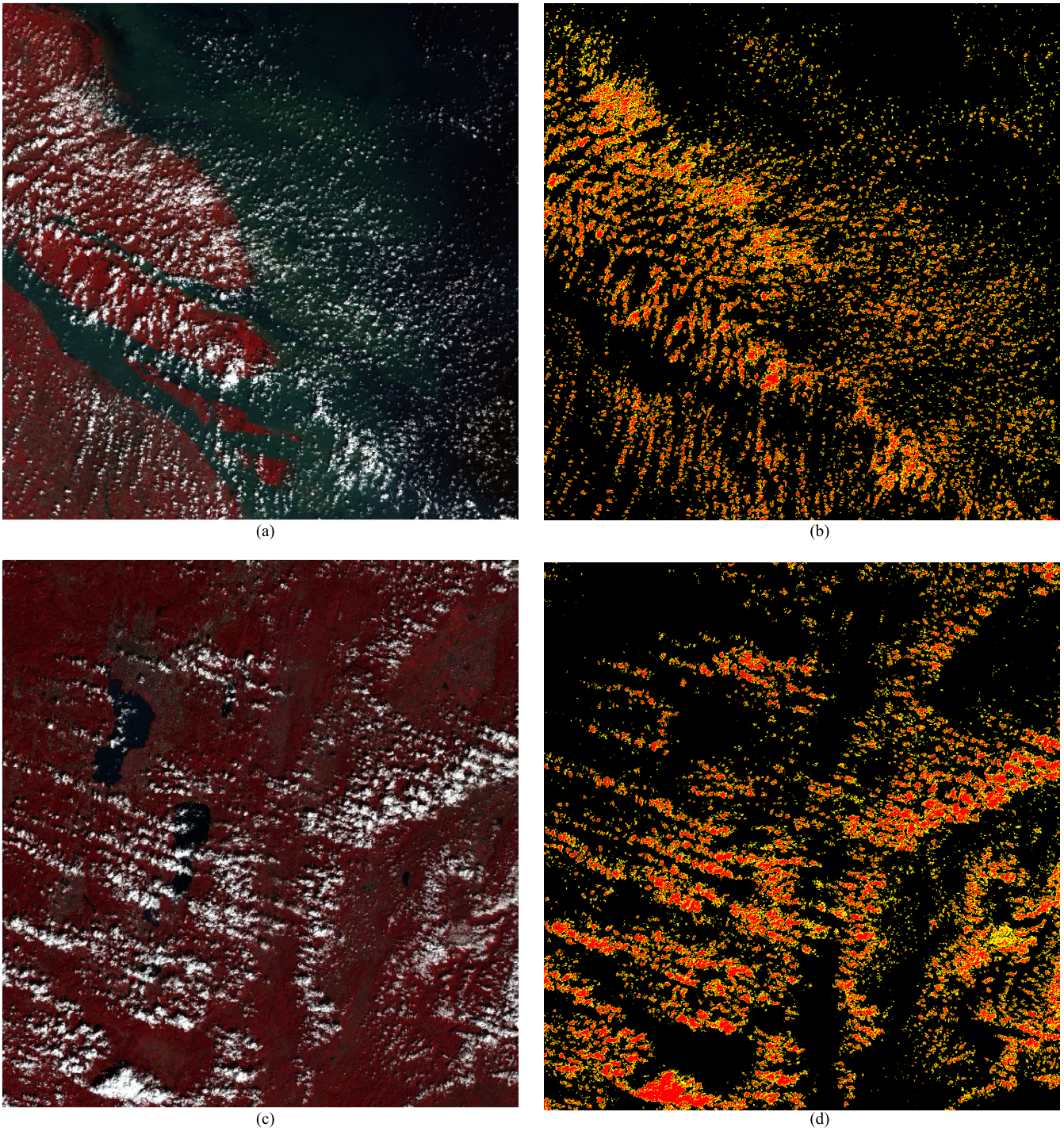


Fig. 4. Spliced visual cloud detection results of Landsat 8 image with our method. (a) Original combination image with NIR, red, green, band (p118_r38 and 20141003). (b) Spliced visual cloud detection results of Landsat 8 image (p118_r38 and 20141003) with our method. (c) Original combination image with NIR, red, green band (p129_r43 and 20161021). (d). Spliced visual cloud detection results of Landsat 8 image (p129_r43 and 20161021) with our method.

B. Detection Performance of Thick and Thin Clouds

According to the training and testing principle of MF-CNN model in Section III, each image block with the size of 128 128 will be divided into three categories (thick clouds, thin clouds, and nonclouds). Through the splicing of classified image blocks, the detection results of whole Landsat 8 remote sensing images can be obtained, and Fig. 4 shows that the detection results of thin and thick cloud in land areas and coastal areas, surface, red areas indicate thick cloud pixels, and yellow areas indicate thin cloud pixels. It can be seen from Fig. 4 that there is no obvious stitching line in the splicing classification results. Whether it is in land areas or coastal areas, the overall detection effect of thin and thick cloud is good, even the small isolated thin and thick clouds can be detected accurately.

To comprehensively present the detection results of thin and thick cloud with different methods in detail, Figs. 5–9 (detection area is 9.6×9.6 km^2) show visual results of cloud detection with different methods in the situation of

TABLE V
DETECTION PERFORMANCE OF DIFFERENT METHODS FOR THICK AND THIN CLOUDS. COLUMNS PRECISION_C, RECALL_C, AND F_SCORE_C DESCRIBE THICK CLOUDS, WHILE PRECISION_T, RECALL_T, AND F_SCORE_T DESCRIBE THIN CLOUDS

Methods	precision_c	recall_c	F_Score_c	precision_t	recall_t	F_Score_t
SVM	0.8648	0.7984	0.8303	0.7409	0.5879	0.6556
RF	0.8628	0.8194	0.8406	0.7396	0.5923	0.6578
FCNN	0.8701	0.7620	0.8125	0.7011	0.5943	0.6433
Self-Contrast	0.8666	0.6461	0.7403	0.5899	0.4784	0.5283
Our Method	0.9074	0.8946	0.8920	0.7813	0.7693	0.7753

various underlying surface (water, urban buildings, snow) easily confused with the cloud and different cloud amounts. Figs. 5–9 present the original combination images (NIR—red-green band), the ground truth value, and detection results of FCNN method in [35], self-contrast method, RF method, SVM method, our proposed method, and Fmask method in turn. The Fmask algorithm does not distinguish between thin and thick clouds, so entire clouds (thin and thick clouds) are represented, respectively. The black areas in the figure indicate the ground in white. From Figs. 5–9, we can see that the detection results in thick and thin clouds by our proposed method are the most similar to the ground truth values, while other methods have more detection errors. In general, the detection results are better when the underlying surface is vegetation. For other underlying surfaces that are easily confused with clouds, there are more cases of commission and omission. The detailed analysis will be further discussed in Section V. In addition to the visual results, we detect the thick and thin clouds in the 806 test images and quantitatively evaluate the detection performance through three evaluation indexes, namely, the precision values, recall values, and F_Score. The statistical results of detection performance by different methods are given in Table V.

The precision_c and precision_t indicators represent the average precision values for thick and thin clouds, respectively, and the recall_c and recall_t indicators are the average recall values for thick and thin clouds. However, the evaluation results of the above two indicators are inconsistent in some cases, so the F_Score_c and F_Score_t considering the above two indicators are calculated for thick and thin clouds. If the F_Score is higher, the performance of the corresponding detection method is better.

In Table V, compared with the other four methods, the proposed method has significant improvement in precision, recall, and F_Scores values in thick clouds and thin clouds, which demonstrates the MF-CNN model proposed in this paper has obvious superiority in cloud detection. In general, the detection result of thick clouds for each method is better than that of thin clouds. The main reason is that thick clouds have distinct spectral characteristics, which can be distinguished from the background. However, the spectra of semitransparent thin clouds are more difficult to be distinguished from the other objects, as spectra of these thin clouds are generally mixed with those of various underlying surface objects. In addition, in Table V, the recall values of thick clouds and thin clouds in the four comparison methods are lower than their precision values, which means that these methods have omissions in the detection of thick or thin cloud pixels even if they have a certain degree of detection accuracy. Among these comparison methods, the detection result from the self-contrast method is the worst, whether in thin cloud detection or thick cloud detection. That is mainly because, in the self-contrast method, rich low-level spatial information and multiscale global features are lost in the process of feature extraction, and only the high-level semantic information is included in the final classification, which makes the detection result lack detailed information and is only suitable for rough detection.

C. Detection Performance of Entire Cloud Regions

After evaluating the detection efficiency of thick and thin clouds, the comprehensive detection performance of all cloud pixels still needs to be evaluated. In addition to the four comparison methods mentioned above, the classic Fmask cloud detection method is added for comparison, and Figs. 5(h)–9(h) indicate the detection result of Fmask. It can be seen from the visual detection results, the method does not confuse the clouds with bright buildings (Fig. 7), water (Fig. 8), and snow (Fig. 9), but there is significant over-detection of the clouded areas. In order to further quantify the detection effect of the five comparison methods and the proposed method in the entire cloud areas (thin cloud and thick cloud regions), several accuracy assessment measures, namely, the RR, ER, FAR, and RER, are used in this paper. We calculate the average of RR, ER, FAR, and RER for all the test images and the results are shown in Table VI.

As can be seen from Table VI, the RR of the Fmask is very high, indicating that the method is quite effective in cloud detection, but the high RR value is accompanied by the high ER value (about 1 to 3 times of other methods) and incredibly high FAR value (about 10 to 30 times of other methods). This means that the method has obvious over-detection in cloud detection. Through comparing the cloud detection effect of the other four selected models with our proposed method, the RR value of the self-contrast model is the lowest, accompanied by higher ER value, which results in the worst performance in RER index (the lowest RER value) when considering both the RR and ER. On the contrary, the RER index of our method considering both RR and ER is markedly higher than that of all other methods. Although FAR of our model is slightly higher, our detection method guarantees the lowest ER value while ensuring the higher RR value. It demonstrates that the proposed method not only has advantages in the detection of thin clouds and thick clouds but also has superiority for the detection of entire cloud regions.

V. DISCUSSION

A. Qualitative Analysis of the Proposed Method in Cloud Detection

As shown in Figs. 5–9, in general, the thick and thin cloud detection results of our proposed method are the most similar to the ground truth, while the results of the SVM and RF

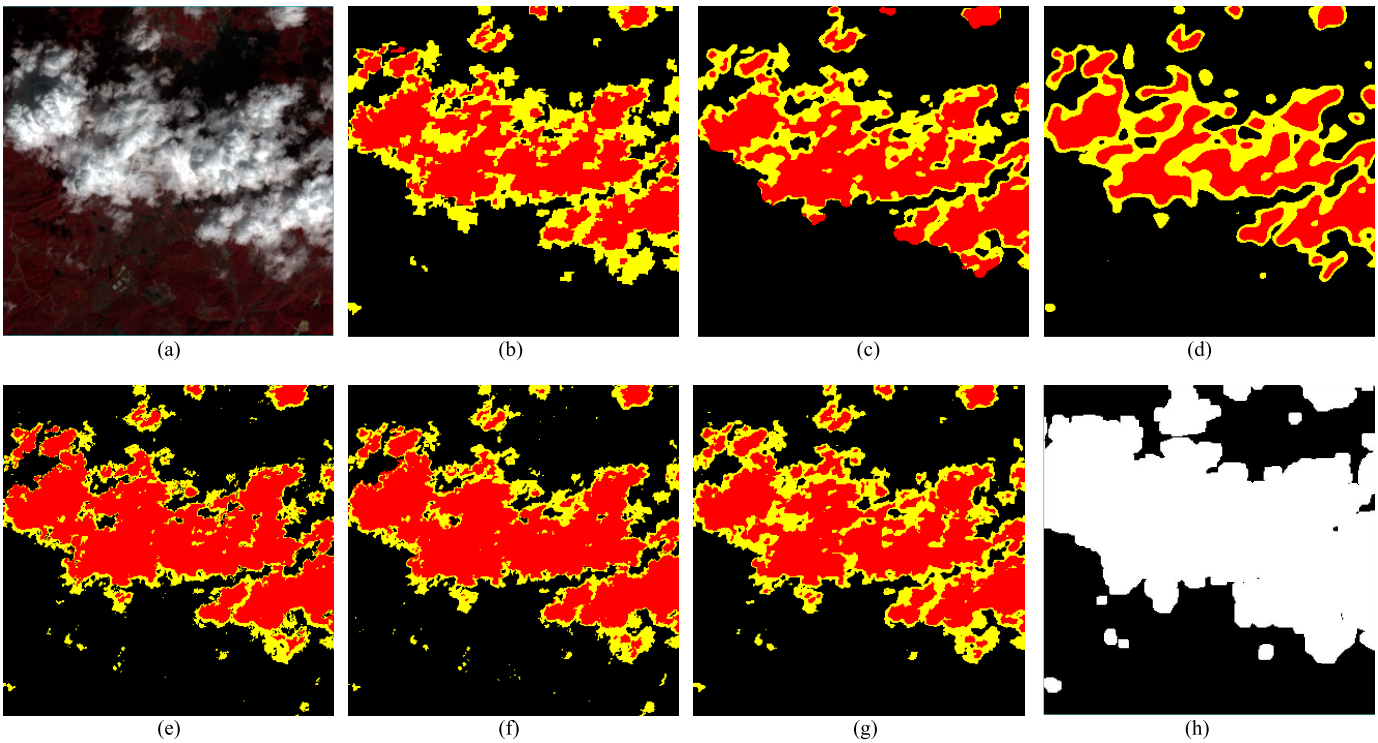


Fig. 5. Visual comparisons of different cloud detection methods in the partial scene of Landsat 8 (p129_r43 and 20161021). (a) Original combination image with NIR, red, green band. (b) Ground truth image. (c) Cloud detection result of FCNN method. (d) Cloud detection result of our method without multiscale features (self-contrast method) (e). Cloud detection result of RF method. (f) Cloud detection result of SVM method. (g) Cloud detection result of our proposed method. (h) Cloud detection result of Fmask method.

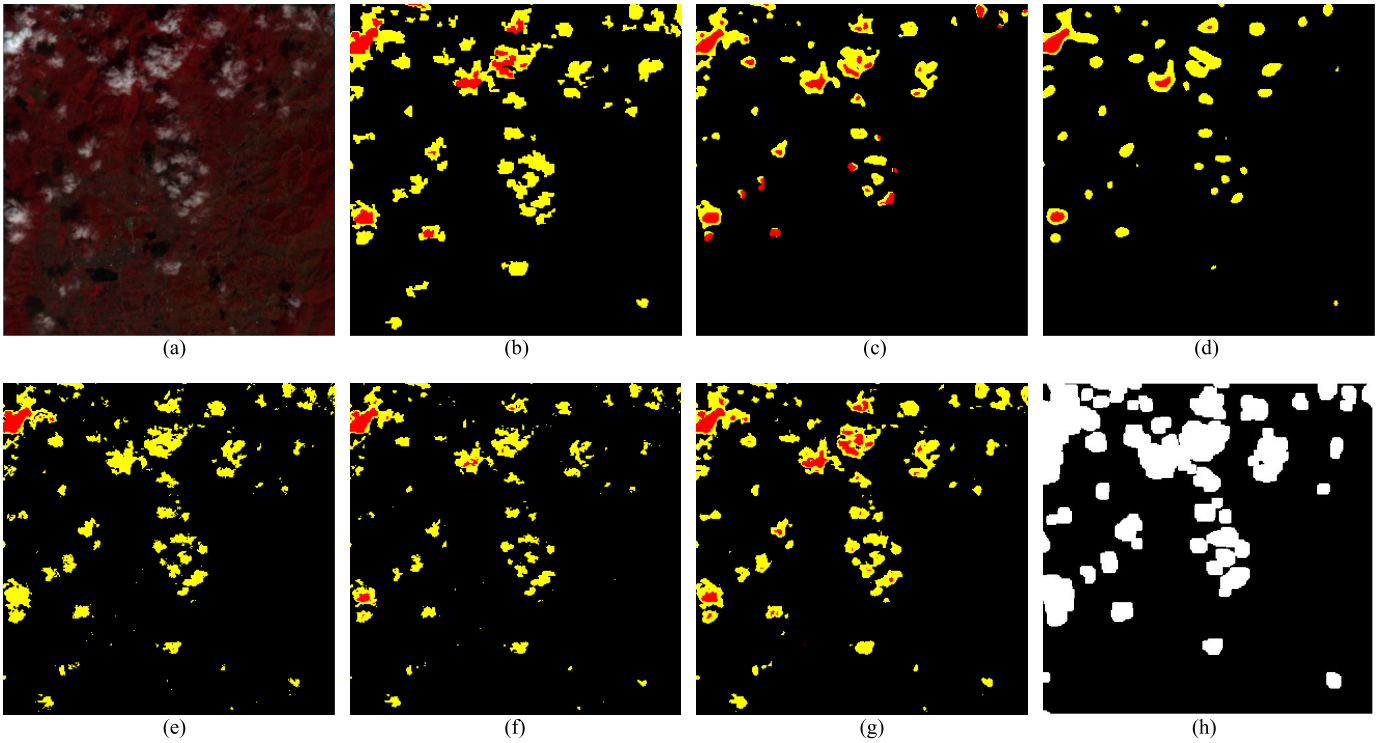


Fig. 6. Visual comparisons of different cloud detection methods in the partial scene of Landsat 8 (p119_r40 and 20171002). (a) Original combination image with NIR, red, green band. (b) Ground truth image. (c) Cloud detection result of FCNN method. (d) Cloud detection result of our method without multiscale features (self-contrast method). (e) Cloud detection result of RF method. (f) Cloud detection result of SVM method. (g) Cloud detection result of our proposed method. (h) Cloud detection result of Fmask method.

methods are discrete and have discernible detection errors in thin-cloud regions. As for the method in [35] and self-contrast structure based on MF-CNN model, their detection results are smooth and can hardly capture the subtle information in the shape of thick and thin clouds. All of the above cloud detection methods are much more accurate in detecting

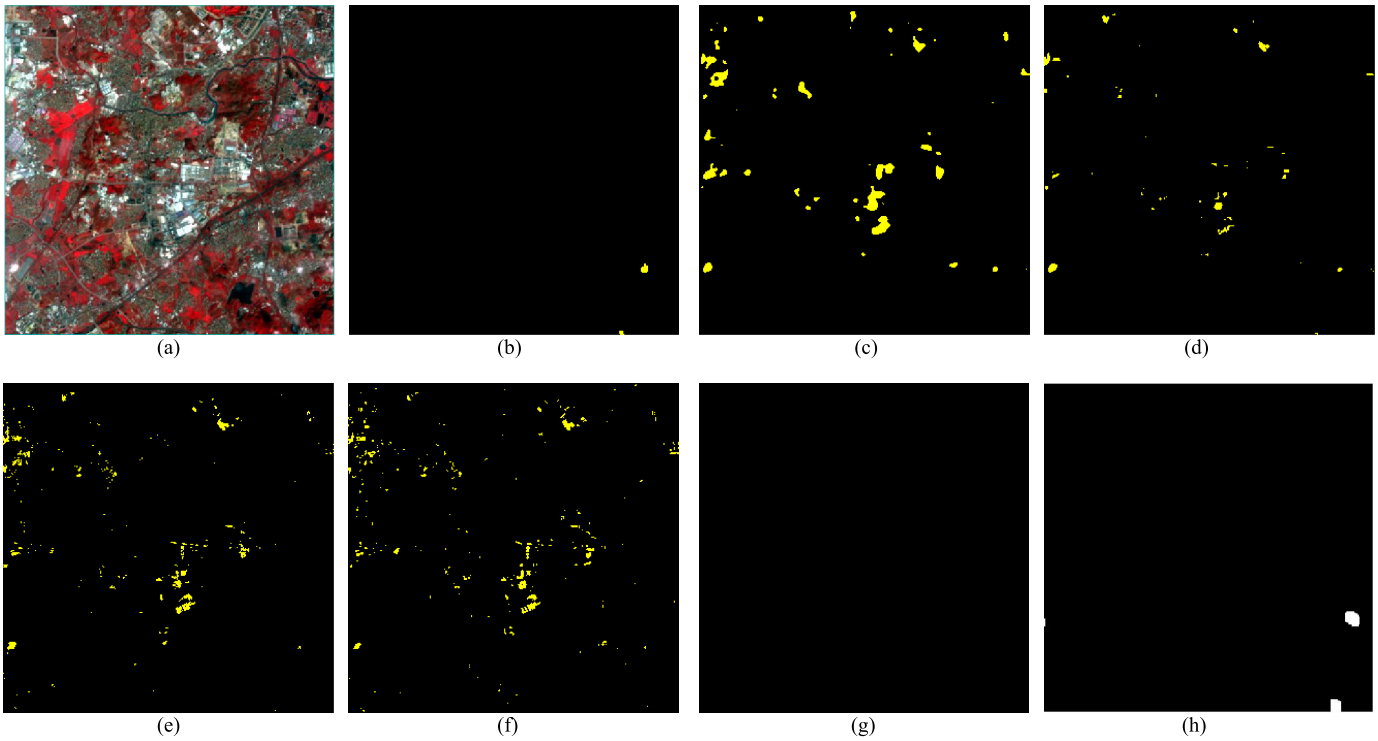


Fig. 7. Visual comparisons of different cloud detection methods in the partial scene of Landsat 8 (p119_r43 and 20180122). (a) Original combination image with NIR, red, green band. (b) Ground truth image. (c) Cloud detection result of FCNN method. (d) Cloud detection result of our method without multiscale features (self-contrast method). (e) Cloud detection result of RF method. (f) Cloud detection result of SVM method. (g) Cloud detection result of our proposed method. (h) Cloud detection result of Fmask method.

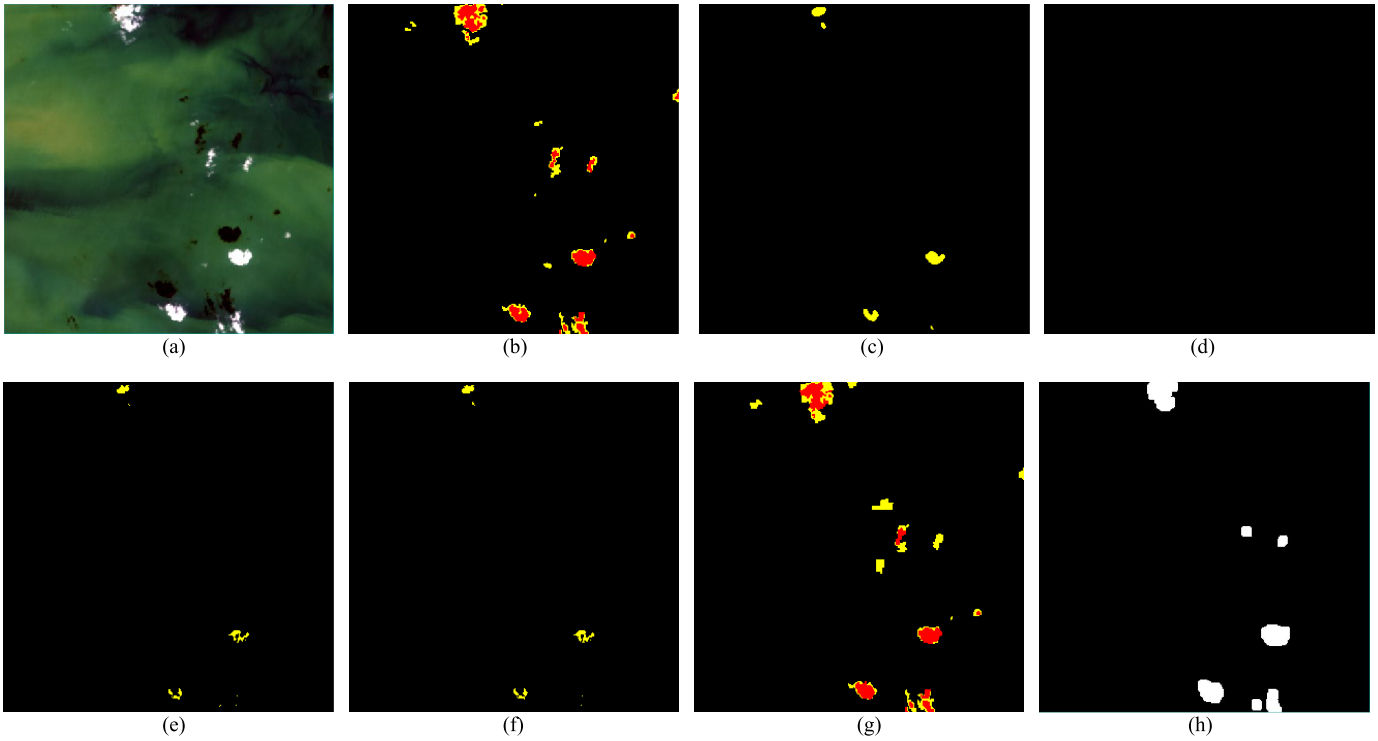


Fig. 8. Visual comparisons of different cloud detection methods in the partial scene of Landsat 8 (p118_r40 and 20141003). (a) Original combination image with NIR, red, green band. (b) Ground truth image. (c) Cloud detection result of FCNN method. (d) Cloud detection result of our method without multiscale features (self-contrast method). (e) Cloud detection result of RF method. (f) Cloud detection result of SVM method. (g) Cloud detection result of our proposed method. (h) Cloud detection result of Fmask method.

thick clouds than thin clouds because the spectral features of thick clouds are more distinctive and differentiable from the background. While the thin clouds are semitransparent and are usually mixed with different underlying surfaces, it is more difficult to identify those regions covered by thin clouds.

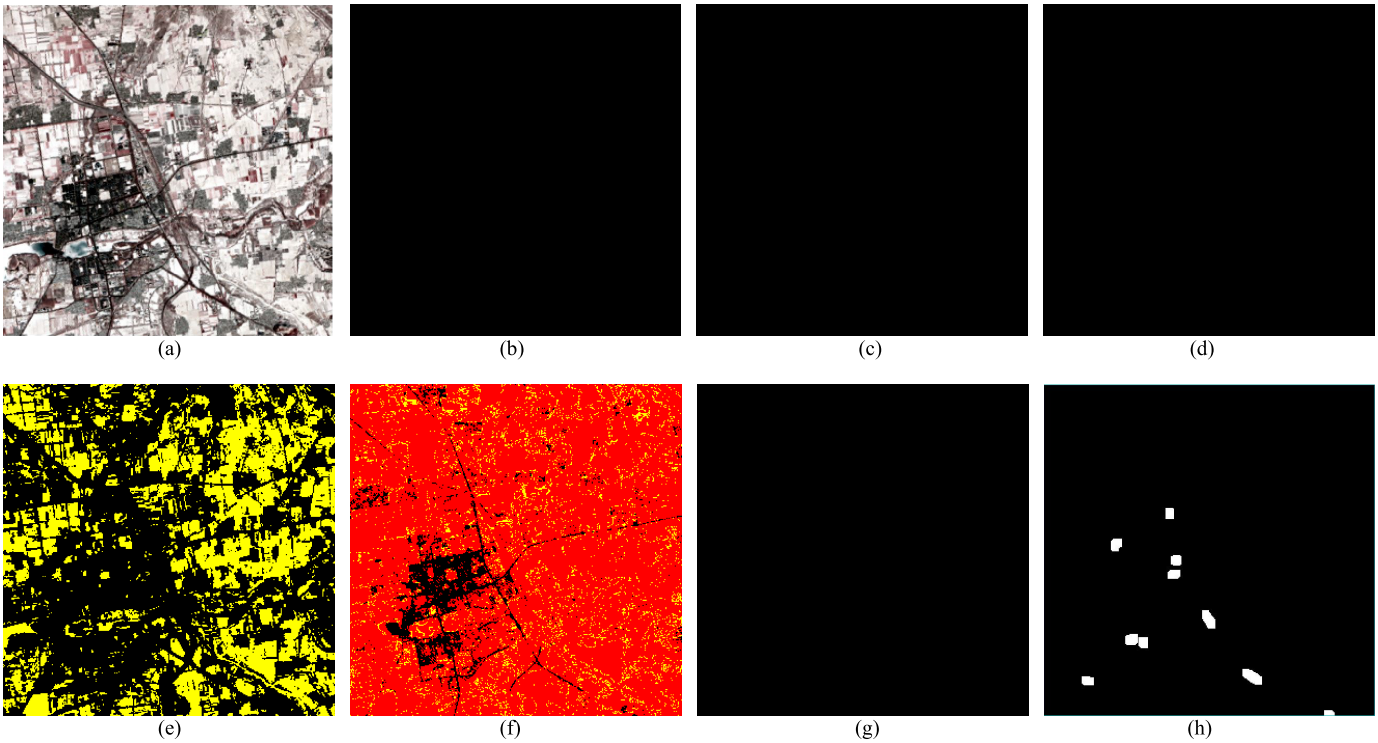


Fig. 9. Visual comparisons of different cloud detection methods in the partial scene of Landsat 8 (p123_r32 and 20160214). (a) Original combination image with NIR, red, green band. (b) Ground truth image. (c) Cloud detection result of FCNN method. (d) Cloud detection result of our method without multiscale features (self-contrast method). (e) Cloud detection result of RF method. (f) Cloud detection result of SVM method. (g) Cloud detection result of our proposed method. (h) Cloud detection result of Fmask method.

TABLE VI
DETECTION PERFORMANCE OF DIFFERENT METHODS FOR ENTIRE CLOUDS

Methods	RR	ER	FAR	RER
SVM	0.8340	0.0549	0.0517	15.19
RF	0.8440	0.0601	0.0275	14.05
FCNN	0.8236	0.0637	0.0371	12.93
Self-Contrast	0.7361	0.0887	0.0290	8.30
Fmask	0.9923	0.1049	0.6341	9.46
Our Method	0.9340	0.0385	0.0693	24.22

As for the Fmask detection method, it can detect the entire cloud areas (thin clouds and thick clouds). Since the method takes into account the relative positional relationship between the clouds and the cloud shadows, the method does not cause confusion between the bright underlying surface and the clouds, which can be reflected primarily in Figs. 7–9. However, as can be seen from Figs. 5 and 6, this method has obvious over-detection problems in the cloud regions. In fact, the commission errors are mainly from the cloud boundary because Fmask would make a buffer of three pixels for each cloudy pixel (Zhu and Woodcock, 2012).

Through visually comparing the results between model detection results and ground truth values, the methods of SVM and RF can distinguish thick cloud pixels accurately in combination images to a certain extent, but the thick cloud detection results usually contain many thin cloud pixels (as shown in Fig. 5). As for the thin cloud regions, both of these two methods are difficult to identify them accurately. Some complex or bright background pixels are misclassified to thin clouds (as shown in Figs. 7 and 9 and partially discrete thin cloud pixels similar to noise in Figs. 5 and 6), and some thin cloud pixels are missed in the process of detecting (as shown in Figs. 5–8). The cloud detection results of these two methods are quite similar. The SVM and RF algorithms are based on spectral feature similarity to classify each pixel. They identify the cloud regions only through the spectral features of each pixel, without taking into account the structural neighborhood information. In addition, these relatively simple model structures have limited feature learning for complex scenes, such as spectral information mixture of thin clouds and the ground surface, or the bright underlying scene that is easily confused with clouds.

As for the detection method of FCNN in [35] and self-contrast method, only the general distribution of thin clouds and thick clouds can be detected. The edges of cloud regions are not distinguished accurately. In Fig. 5, the FCNN detection method works well on the thick clouds, but there are apparent misclassification and omission for the thin cloud pixels in Figs. 6–8. In addition, the detection results of isolated or marginal thin clouds are poor. That is because feature layers constructed at different scales are sampled up to the size of input image simultaneously. Multiscale feature concatenation layers have no progressive learning process via convolutional layers, which omits much structural information in the image and renders over-smooth edge of clouds. Besides,

the lack of multiscale global features in the learning strategy leads to the loss of detailed information on thin and thick clouds, while the detection results of self-contrast model are worse, which can only detect the approximate range of thin and thick clouds, as shown in Figs. 5 and 6. Sometimes, for smaller clouds, it even has significant omissions (as shown in Figs. 6 and 8). In addition, for bright buildings that are easily confused with clouds, this method most likely divide them into thin clouds, as shown in Fig. 7. The reason for the above phenomenon is that the feature information of the thick clouds or semitransparent thin clouds mixed with the underlying surface has certain similarities with the bright buildings. Though the self-contrast model has the process of progressive learning via convolutional layers, it does not extract multiscale global features of thick and thin clouds through the multiscale module, nor does it integrate low-level spatial information, which will result in the loss of spatial information and reducing the ability of feature learning in the model. The dearth of spatially structured information and the limitations of feature learning capabilities reduce its detection accuracy, and therefore, the contours of the detected thick or thin cloud regions are different from the ground truth values. Although the above two cloud detection methods based on CNN have errors in the process of thin and thick cloud detection, they are easier to learn the effective features of clouds and snow from the two SWIR and cirrus bands because of their better feature learning ability than RF and SVM algorithms, which is conducive to distinguishing snow and clouds.

In summary, our proposed model is more similar to the ground truth, which is attributed to the multiscale global features learning strategy that learns the context information of cloud regions at multiscale. In addition, the low-level spatial and high-level semantic information integrated into the progressive up-sampling learning process supplement the information. The proposed model is more capable of detecting cloud regions of different types and can achieve more accurate detection results.

B. Quantitative Analysis of the Proposed Method's Effectiveness in Cloud Detection

In the quantitative comparison in Section IV, we only compare the average detection accuracy of different methods for all the test images in pixel level. To further illustrate the effect of different detection methods in thin and thick cloud detection, we calculate the F_Score of thick and thin cloud for each test image containing cloud and compare the detection result of each selected method with that of our proposed method in the level of image blocks. The distribution of detection accuracy for test images is shown in Fig. 10.

As shown in Fig. 10, the horizontal and vertical axes represent the F_Score value of thin clouds and thick clouds of test images, respectively. The blue points are the detection results of our proposed method, and the points of other four colors are the results of comparison methods. Compared with other methods, the F_Score values of test images detected by our proposed method are concentrated in the upper right corner of Fig. 10. It indicates that our proposed model can detect thin and thick clouds more accurately than other methods. The overall distributions of F-Score value, such as SVM, RF, FCNN model, and self-contrast method, are more fragmented, indicating these methods are less stable in cloud detection. Besides, the general distribution trend of F-Scores for these methods is obviously shifted to the left, which means the detection performance of thin clouds is poorer. As for the thick cloud detection results, considering the compactness and distribution position in the coordinate system, the thick cloud detection results of the other four comparison methods are slightly worse than that of the method proposed in this paper. This conclusion is also consistent with the analysis in Section IV.

Due to the inability to distinguish thin clouds from thick clouds in the Fmask algorithm, its detection result is not evaluated in the above discussion. To evaluate the effectiveness of our proposed model in detecting entire cloud regions, we calculate the RR, ER, FAR, and the RER values of each test image block containing cloud.

According to the distribution of these values, we divide the RR, ER, FAR, and RER values into different intervals, respectively, and then count the number of test images within a certain range for each cloud detection method. The horizontal axis represents different numerical intervals, the vertical axis represents the number of test image blocks in the corresponding interval, and different color bars correspond to the six different detection methods. The distributions of the four evaluation indexes for different methods are shown in Figs. 11–14.

In Fig. 11, the RR values of our proposed model and Fmask algorithm are mainly distributed in the intervals of 0.95 to 1.00 and 0.90 to 0.95, while the values of the RF, SVM, FCNN, and self-contrast methods are mainly distributed in the other four lower accuracy intervals. The self-contrast method performs the worst in terms of right detection rate, and the most of values in the entire cloud detection is less than 0.75. The above results also mean that the method proposed in this paper and the Fmask algorithm is more advantageous in terms of the right detection rate in entire cloud (thick and thin cloud) detection.

Although the Fmask algorithm has a high accuracy rate, this method has obvious disadvantages in ER and FAR. As shown in Figs. 12 and 13, the ER values of the Fmask algorithm are mainly distributed in the intervals of 0.06 to 0.08 and greater than 0.08. The FAR values of total test blocks are mainly distributed in the intervals of 0.15 to 0.2 and greater than 0.2. This means that the Fmask algorithm has obvious over-detection problems in the process of cloud detection, which will detect noncloud pixels as clouds. The ER values of the self-contrast method are distributed evenly in the first four intervals with smaller ER values, mainly in the interval of larger than 0.08, while the other three methods (FCNN, RF, SVM) are generally uniformly distributed in the five different intervals. The method proposed in this paper performs well in ER and concentrates on the first three intervals with lower ER. As for the FAR in cloud detection of test block images containing clouds, the FAR values of FCNN, self-contrast, RF and SVM methods are mainly distributed in the first three intervals with smaller values, and our proposed

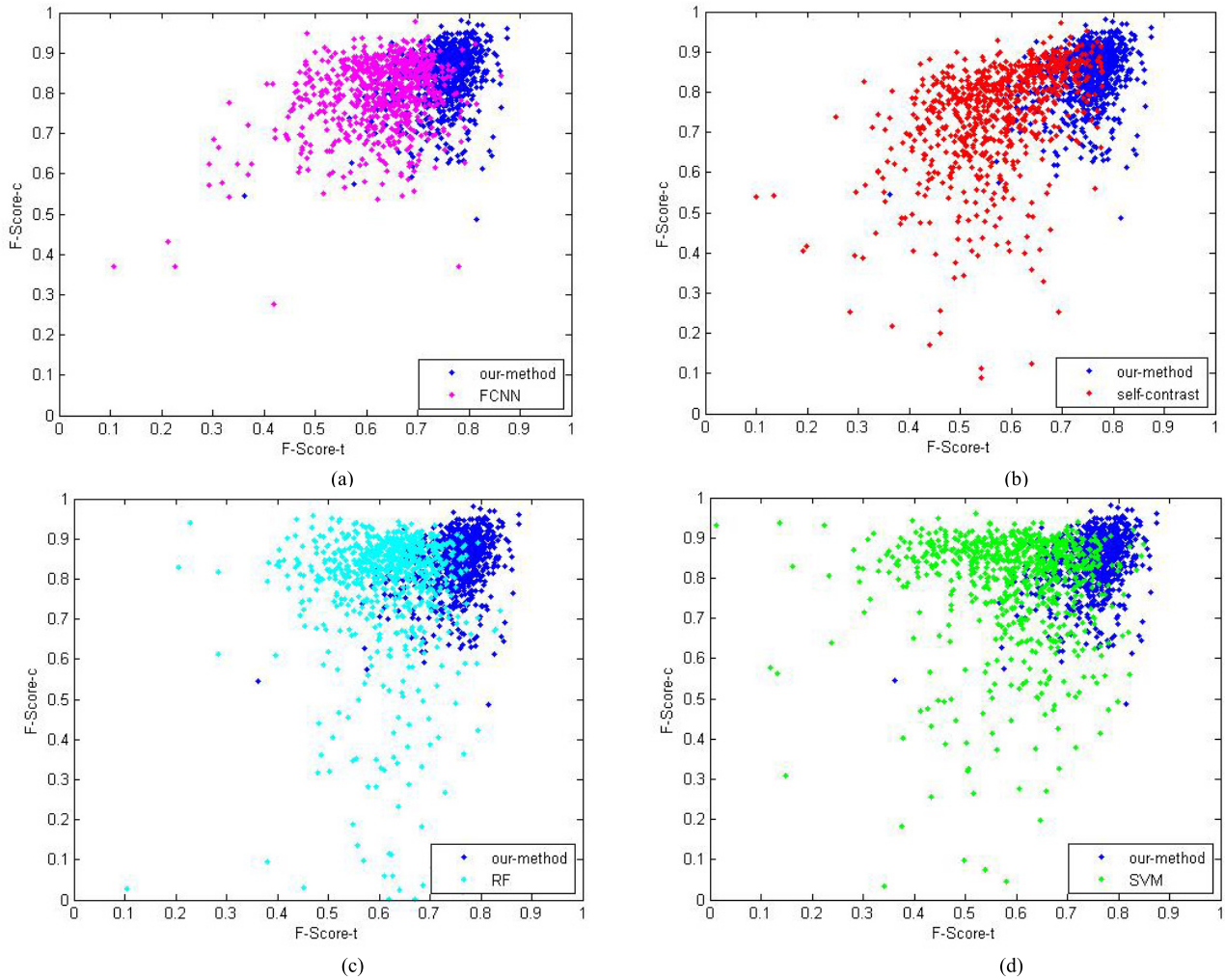


Fig. 10. Distribution of cloud detection accuracy in test data set. (a) Comparison of our method with FCNN model. (b) Comparison of our method with self-contrast model. (c) Comparison of our method with RF. (d) Comparison of our method with SVM.

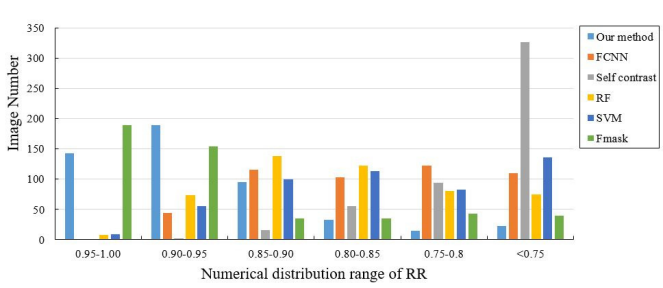


Fig. 11. RR distribution of the test data set.

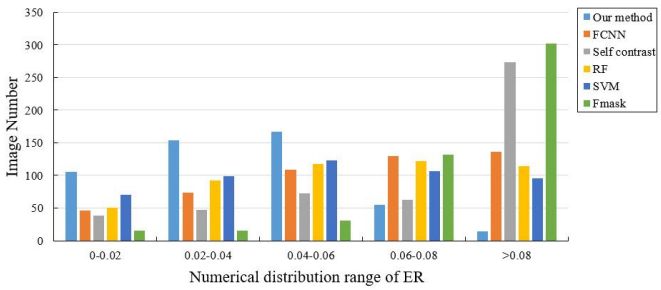


Fig. 12. ER distribution of the test data set with cloud.

method is mainly distributed in the third intervals. Yet the performance of Fmask method is worst in this evaluation index obviously, which is consistent with the conclusion in Table VI.

In order to comprehensively evaluate the RR and ER of different methods, we use the ratio of RR and ER to calculate the RER, and apply this indicator to measure the effectiveness of different methods in the entire cloud detection. In Fig. 14, we can see that the RER values of FCNN, RF, and SVM are concentrated in the first three intervals, and the self-contrast and Fmask methods are mainly distributed in the smallest interval. In contrast, the RER values of our proposed method have almost no image block in the smallest interval and mainly distributed in the largest intervals.

Overall, according to the analysis in Figs. 11–14, the proposed method has a better performance in thin and thick cloud detection. In addition, the classic Fmask algorithm is added as a comparison method in the process of entire cloud detection. The analysis of four evaluation indexes shows that although the Fmask algorithm performs well on the RR index, it is accompanied by a high ER value because of over-detection

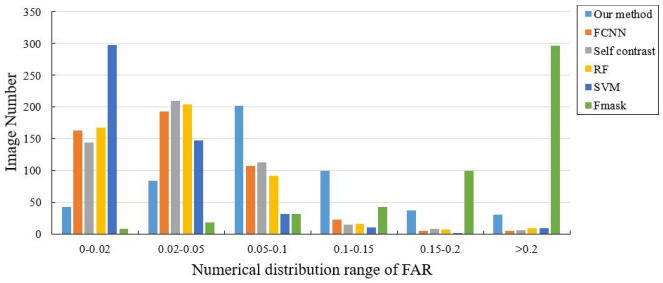


Fig. 13. FAR distribution of the test data set with cloud.

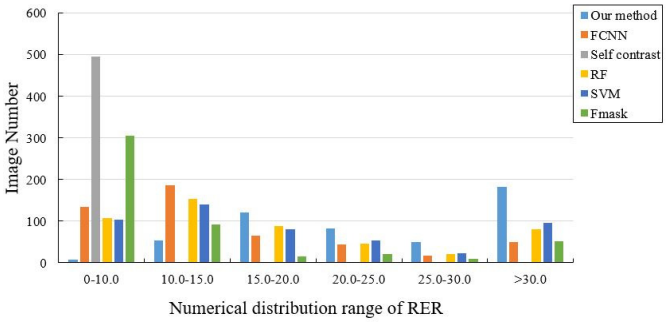


Fig. 14. RER distribution of the test data set with cloud.

in the cloud. The method proposed in this paper has apparent advantages in ER and RER indexes while ensuring the RR value.

VI. Conclusion

For Landsat 8 satellite images, the MF-CNN model proposed in this paper has greatly improved the accuracy of thin cloud detection in pixel-level while achieving high accuracy of thick cloud detection. As for the entire cloud detection task, the proposed method ensures high RR while controlling the ER and avoids the over-detection problem. In addition, even in the bright underlying surface easily confused with clouds, such as bright buildings or snow, the method proposed in this paper will not cause false detection.

On the one hand, multiscale global features extracted from the MF-CNN model can characterize thin or thick clouds from different scales, which is beneficial to obtain more abundant features for the subsequent classification task. On the other hand, the integration of low-level spatial and high-level semantic information in the gradual up-sampling learning process supplements the traditional spectral-based features at multiscales for the detection targets, which facilitates the identification of complex clouds with various types and shapes.

In order to evaluate the effectiveness of the MF-CNN model, the cloud detection results of traditional machine learning, deep learning, and classic Fmask methods are used for experimental comparison. To further evaluate the performance of the proposed method, the F_Score of thick and thin clouds, and RR, ER, FAR, RER of the entire cloud are calculated to describe the performance of the detection methods comprehensively. Through both qualitative and quantitative analysis, we found that the detection performance of our method on thin, thick, or entire cloud detection is superior to that of other methods.

Despite the high accuracy achieved by this paper, identifying thin-cloud regions are a challenging task. In the future, we consider integrating the spectral information with the contour information of the thin and thick clouds, and employ deeper convolutional network model to obtain more abundant cloud features to improve the performance of cloud detection further.

References

[1] K. Anderson, B. Ryan, W. Sonntag, A. Kavvada, and L. Friedl, "Earth observation in service of the 2030 agenda for sustainable development," *Geo-Spatial Inf. Sci.*, vol. 20, no. 2, pp. 77–96, 2017.

[2] Y. Zhang, W. B. Rossow, A. A. Lacis, V. Oinas, and M. I. Mishchenko, "Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data," *J. Geophys. Res., Atmos.*, vol. 109, p. D19105, Oct. 2004.

[3] H. Lv, Y. Wang, and Y. Shen, "An empirical and radiative transfer model based algorithm to remove thin clouds in visible bands," *Remote Sens. Environ.*, vol. 179, pp. 183–195, Jun. 2016.

[4] Q. Li, W. Lu, and J. Yang, "A hybrid thresholding algorithm for cloud detection on ground-based color images," *J. Atmos. Ocean. Technol.*, vol. 28, pp. 1286–1296, Oct. 2011.

[5] W. B. Rossow and L. C. Garder, "Cloud detection using satellite measurements of infrared and visible radiances for ISCCP," *J. Climate*, vol. 12, pp. 2341–2369, Dec. 1993.

[6] K. T. Kriebel, G. Gesell, M. Kästner, and H. Mannstein, "The cloud analysis tool APOLLO: Improvements and validations," *Int. J. Remote Sens.*, vol. 24, no. 12, pp. 2389–2408, 2003.

[7] C. L. Liu and B. F. Wu, "Application of cloud detection algorithm for the AVHRR data," *J. Remote Sens.*, vol. 8, pp. 677–687, 2004.

[8] B.-C. Gao, P. Yang, and R.-R. Li, "Detection of high clouds in polar regions during the daytime using the MODIS 1.375-μm channel," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 474–481, Feb. 2003.

[9] S. A. Ackerman, K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley, "Discriminating clear sky from clouds with MODIS," *J. Geophys. Res., Atmos.*, vol. 103, no. D24, pp. 32141–32157, 1998.

[10] R. R. Irish, "Landsat 7 automatic cloud cover assessment," *Proc. SPIE, Algorithms Multispectral, Hyperspectral, Ultraspectral Imag. VI*, vol. 4049, pp. 348–355, Aug. 2000, doi: 10.1117/12.410358.

[11] 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, 2006.

[12] L. Oreopoulos, M. J. Wilson, and T. Várnai, "Implementation on Landsat data of a simple cloud-mask algorithm developed for MODIS land bands," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 597–601, Jul. 2011.

[13] C. Huang *et al.*, "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.

[14] 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.

[15] Y. Oishi, H. Ishida, and R. Nakamura, "A new Landsat 8 cloud discrimination algorithm using thresholding tests," *Int. J. Remote Sens.*, pp. 1–21, Aug. 2018, doi: 10.1080/01431161.2018.1506183.

[16] S. Qiu, B. He, Z. Zhu, Z. Liao, and X. Quan, "Improving Fmask cloud and cloud shadow detection in mountainous area for Landsats 4–8 images," *Remote Sens. Environ.*, vol. 199, pp. 107–119, Sep. 2017.

[17] Y. Zhang, B. Guindon, and X. Li, "A robust approach for object-based detection and radiometric characterization of cloud shadow using haze optimized transformation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5540–5547, Sep. 2014.

[18] L. Tracewski, L. Bastin, and C. C. Fonte, "Repurposing a deep learning network to filter and classify volunteered photographs for land cover and land use characterization," *Geo-Spatial Inf. Sci.*, vol. 20, no. 3, pp. 252–268, 2017.

[19] S. Wang, Y. Li, and D. Wang, "Data field for mining big data," *Geo-Spatial Inf. Sci.*, vol. 19, no. 2, pp. 106–118, 2016.

[20] T. Bai, D. Li, K. Sun, Y. Chen, and W. Li, "Cloud detection for high-resolution satellite imagery using machine learning and multi-feature fusion," *Remote Sens.*, vol. 8, no. 9, p. 715, 2016.

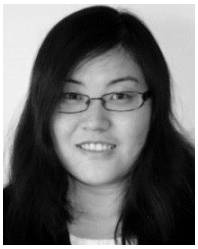
[21] A. Movia, A. Beinat, and F. Crosilla, "Shadow detection and removal in RGB VHR images for land use unsupervised classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 119, pp. 485–495, Sep. 2016.
[22] S. R. Surya and P. Simon, "Automatic cloud detection using spectral rationing and fuzzy clustering," in *Proc. 2nd Int. Conf. Adv. Comput., Netw. Secur.*, Dec. 2013, pp. 90–95.
[23] M. R. Azimi-Sadjadi, W. Gao, T. H. V. Haar, and D. Reinke, "Temporal updating scheme for probabilistic neural network with application to satellite cloud classification—Further results," *IEEE Trans. Neural Netw.*, vol. 12, no. 5, pp. 1196–1203, Sep. 2001.
[24] G. Vivone, P. Addesso, R. Conte, M. Longo, and R. Restaino, "A class of cloud detection algorithms based on a MAP-MRF approach in space and time," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5100–5115, Aug. 2014.
[25] R. Zhang, D. Sun, S. Li, and Y. Yu, "A stepwise cloud shadow detection approach combining geometry determination and SVM classification for MODIS data," *Int. J. Remote Sens.*, vol. 34, no. 1, pp. 211–226, 2013.
[26] P. Li, L. Dong, H. Xiao, and M. Xu, "A cloud image detection method based on SVM vector machine," *Neurocomputing*, vol. 169, no. 2, pp. 34–42, Dec. 2015.
[27] K. Tan, Y. Zhang, and X. Tong, "Cloud extraction from Chinese high resolution satellite imagery by probabilistic latent semantic analysis and object-based machine learning," *Remote Sens.*, vol. 8, no. 11, p. 963, 2016.
[28] X. Hu, Y. Wang, and J. Shan, "Automatic recognition of cloud images by using visual saliency features," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 8, pp. 1760–1764, Aug. 2015.
[29] C. Ma, F. Chen, J. Liu, and J. Duan, "A new method of cloud detection based on cascaded AdaBoost," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 18, no. 1, p. 012026, 2014.
[30] Q. Zhang and C. Xiao, "Cloud detection of RGB color aerial photographs by progressive refinement scheme," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7264–7275, Nov. 2014.
[31] Z. Shao, J. Deng, L. Wang, Y. Fan, N. S. Sumari, and Q. Cheng, "Fuzzy AutoEncode based cloud detection for remote sensing imagery," *Remote Sens.*, vol. 9, no. 4, p. 311, 2017.
[32] T. Johnston, S. R. Young, D. Hughes, R. M. Patton, and D. White, "Optimizing convolutional neural networks for cloud detection," in *Proc. Mach. Learn. HPC Environ.*, 2017, Art. no. 4.
[33] M. L. Goff, J. Y. Tourneret, H. Wendt, M. Ortner, and M. Spigai, "Deep learning for cloud detection," in *Proc. Int. Conf. Pattern Recognit. Syst.*, 2018, p. 10.
[34] Z. Yan *et al.*, "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.
[35] 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.
[36] H. Jiang and N. Lu, "Multi-scale residual convolutional neural network for haze removal of remote sensing images," *Remote Sens.*, vol. 10, no. 6, p. 945, 2018.
[37] 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.*, Jul. 2016, pp. 701–704.
[38] 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.
[39] B.-C. Gao and Y. J. Kaufman, "Selection of a 1.3758-μm channel for remote sensing of cirrus clouds and stratospheric aerosols from EOS/MODIS," *Proc. SPIE, Passive Infr. Remote Sens. Clouds Atmos.*, vol. 1934, pp. 109–116, Sep. 1993, doi: 10.1117/12.154909.
[40] Z. Zhu, S. Wang, and C. E. Woodcock, "Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images," *Remote Sens. Environ.*, vol. 159, pp. 269–277, Mar. 2015.
[41] 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. 2017.
[42] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1265–1274.
[43] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
[44] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, Dec. 2014, pp. 1–15. [Online]. Available: https://arxiv.org/abs/1412.6980
[45] C. Latry, C. Panem, and P. Dejean, "Cloud detection with SVM technique," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2007, pp. 448–451.
[46] B. C. Ko, J.-Y. Kwak, and J.-Y. Nam, "Wildfire smoke detection using temporospatial features and random forest classifiers," *Opt. Eng.*, vol. 51, no. 1, p. 7208, 2012.

Zhenfeng Shao received the Ph.D. degree in aerial photogrammetry from Wuhan University, Wuhan, China, in 2004.

He is currently a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His research interests include remote sensing.

Yin Pan received the B.Eng. degree in surveying and mapping from Wuhan University, Wuhan, China, in 2017, where she is currently pursuing the M.Eng. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing.

Her research interests include deep learning and image processing.



Chunyuan Diao received the Ph.D. degree in geography from the State University of New York at Buffalo, Buffalo, NY, USA.

She is currently an Assistant Professor with the Department of Geography and GIScience, University of Illinois at Urbana–Champaign, Urbana, Il, USA. Her research interests include the confluence of remote sensing, GIScience, and biogeography.

Jiajun Cai received the B.Eng. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2016, where he is currently pursuing the M.Eng. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing.

His research interests include deep learning and image processing.