

Multi-spectral cloud detection based on a multi-dimensional and multi-grained dense cascade forest

Ming Shao^a and Yao Zou^{b,*}

^aFudan University, School of Management, Shanghai, China

^bDonghua University, Glorious Sun School of Business and Management, Shanghai, China

Abstract. Cloud detection in satellite images is a vital step for cloud/land recognition, cloud/snow discrimination, and cloud shadow removal. Accurate cloud detection plays an important role in land resource management, environmental pollution monitoring, and land target recognition. Deep learning (DL) algorithms have shown great progress in cloud detection. However, as the complexity of the DL-based model increases, cloud detection efficiency decreases. DL-based cloud detection models are unable to successfully balance the performance-efficiency tradeoff. In our study, a multi-dimensional and multi-grained dense cascade forest (MDForest) is proposed for multi-spectral cloud detection. MDForest is a deep forest structure that automatically extracts low-level and high-level features from satellite cloud images end-to-end; a multi-dimensional and multi-grained scanning mechanism is introduced to capture the spectral information of multi-spectral satellite images while enhancing the representation learning ability of cascade forest. The experimental results on the HJ-1A/1B dataset show that MDForest improves the performance of cloud detection and possesses a good inference efficiency compared with DL-based cloud detection methods, which makes the proposed MDForest satisfy the application where good performance and high efficiency are both required. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JRS.15.028507](https://doi.org/10.1117/1.JRS.15.028507)]

Keywords: cloud detection; multi-spectral; multi-dimensional; cascade forest.

Paper 200897 received Dec. 31, 2020; accepted for publication Jun. 15, 2021; published online Jun. 28, 2021.

1 Introduction

Accurate cloud detection in satellite images is a vital step for land object recognition, land resource management, and environmental pollution detection.^{1,2} The difficulties of satellite image transmission and the high similarity that exists between cloud regions and ground objects (such as ice, snow, and fog) make cloud detection a challenge,^{3,4} which drives researchers to develop a fast and accurate cloud detection method. With the development of remote sensing technology, the effective integration and processing of multiple spectral satellite images provide a way to achieve accurate cloud detection by extracting rich information from multi-spectral satellite images.⁵

In recent years, the robustness of cloud detection methods has been greatly improved. Based on the optimization pattern of these methods, three groups have emerged: (1) threshold-based methods,^{6,7} (2) machine learning (ML)-based handcraft feature designing approaches,^{8,9} and (3) deep learning (DL)-based algorithms.^{2,10} Because of their reliable performance and simplicity, function of mask^{11,12} and automated cloud cover assessment^{13,14} algorithms are two representative approaches that have been widely employed for cloud detection. Aiming at addressing the poor generalization of the threshold-based method, Yuan and Hu¹⁵ combined support vector machine (SVM) and haze optimization transform technique to distinguish the fog, cloud-free, and thick cloud areas from satellite imagery, by calculating the correlation between the spectral responses. Many ML-based methods that combine handcraft feature engineering for cloud detection have been broadly explored, these works include SVM-based multi-feature fusion,^{16–18}

*Address all correspondence to Yao Zou, zouyaodhu@mail.dhu.edu.cn

cloud detection implemented by random forest using spectral reflection characteristics,¹⁹ and fast cloud detection realized by decision tree-based models.²⁰ ML-based cloud detection models are a group of algorithms that highly rely on the domain knowledge of meteorology experts. Compared with ML-based cloud detection models, DL-based techniques extract low-level to high-level features end-to-end and have shown great promise to distinguish cloudless areas and cloudy regions.^{21,22} Based on the texture, spectral, and structural information, Shao et al.¹⁰ combined a neural network and fuzzy theory to improve the performance of cloud detection. Subsequently, Shao et al.²³ proposed a multi-scale convolutional neural network (CNN) to automatically extract spatial and spectral information from multi-spectral satellite cloud images, which further enhance the prediction result of cloud detection. With the benefit from the end-to-end low-level and high-level feature extraction mechanism, many DL-based cloud detection models have been established to reduce the detection error. These models include fully CNN,^{24,25} deep CNN,²⁶ cascade CNN,²⁷ and residual network.²²

DL-based algorithms have made great progress in the improvement of cloud detection. However, there are some defects of DL-based methods that limit their practical application in a cloud detection task. (1) It is widely known that the training of DL-based models requires a great amount of labeled training data while the sample labeling process in cloud detection task cost high.²⁸ (2) NN is a pre-defined framework that may lead to under-parameterization or over-parameterization, making the training of NN a complex hyperparameter fine-tuning process.²⁹ (3) DL-based cloud detection models are widely criticized because of their black-box attribution, which relates to an implicit learning process.^{30,31}

Based on the above considerations, an improved multi-dimensional dense deep forest is proposed for multi-spectral cloud detection. The proposed algorithm has the following characteristics: (1) it is a deep framework that ensemble by the decision trees, (2) end-to-end learning, (3) multi-dimensional multi-scale scanning mechanism, and (4) dense connection. The above features make the proposed methods have the following advantages. (1) The deep structure based on tree algorithm effectively extracts the features of cloud samples from satellite cloud images, improving the efficiency of training and reasoning. (2) End-to-end learning avoids complex manual feature design, allowing it to extract low-level and high-level features from the original satellite images. (3) The multi-dimensional and multi-scale scanning mechanism effectively captures the spatial and spectral information of cloud images. (4) The dense connection improves the feature reuse abilities of deep forest, enabling the information extracted from each layer to be exploited more efficiently.

2 Methodology

2.1 Proposed Method

Given a dataset $S^{(0)} = \{x_i^{(0)}, y_i\}_{i=1}^N$ with N cloud detection samples, where $x_i^{(0)}$ represents the original input sample and y_i is the label corresponding to the i 'th sample. In the cloud detection task, the input x is the cloud detection image samples that contain four spectral information; $y_i \in \{1, 2, 3\}$, where $y_i \in \{1\}$ represents the cloud-free area, $y_i = 2$ is the area of thin cloud, and $y_i = 3$ indicates the thick cloud area. In this study, a multi-dimensional and multi-grained dense deep forest (MDForest) is proposed for multi-spectral cloud detection.

Figure 1 shows the structure of MDForest. MDForest consists of two parts: multi-dimensional and multi-grained scanning and dense cascade forest.^{32,33} Multi-dimensional and multi-grained scanning realizes the re-representation process of features. The multi-scale features are captured by a multi-grained scanning mechanism to enhance the representation learning ability of the cascade forest. Also, it possesses the ability to process multiple spectrums, which can effectively distinguish the similarity between the spectrums in satellite images. Cascade forest is a deep forest that simulates the representation learning of neural networks and achieves good prediction accuracy and efficiency through hierarchical image information processing. To realize multi-spectral cloud detection, we first collect the satellite images with four spectrums from Chinese satellites HuanJing-1A and HuanJing-1B (HJ-1A/1B)^{34,35} and effectively integrate them into multi-spectral satellite images. Then, the multi-dimensional and multi-grained

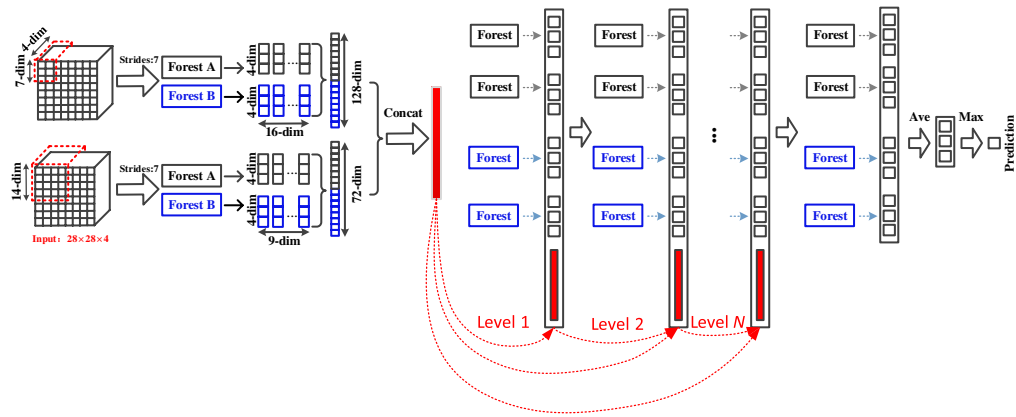


Fig. 1 The structure of MDForest.

scanning mechanism is applied to extract spatial information and spectral information from cloud samples. Next, a cascaded forest is established to realize the representation learning based on the re-represented features. HDForest is constructed in a cascading way, making HDForest an ensemble framework to process information layer-by-layer. In addition, the designing of a densely connected structure, which borrowed from DenseNet,³⁶ is introduced in this study to avoid overfitting by maximizing the utilization of spatial information and spectral information.

2.2 Dense Cascade Forest

Figure 2 shows the structure of the dense cascade forest. According to the hierarchical processing mechanism of the neural network, a deep forest structure can be described as $x^l = F^{(l)}(x^{(l-1)}) = [F^{(l1)}(x^{(l-1)}), F^{(l2)}(x^{(l-1)}), \dots, F^{(lm)}(x^{(l-1)})]$, where $l \in \{1, 2, \dots, L\}$ represents the probabilistic output of $l - 1$ 'th layer. $F^{(l)}$ represents the cascaded forest of level l and $F^{(li)}$ the i 'th random forest in the cascade forest of the level l . As can be seen from Fig. 2, a dense

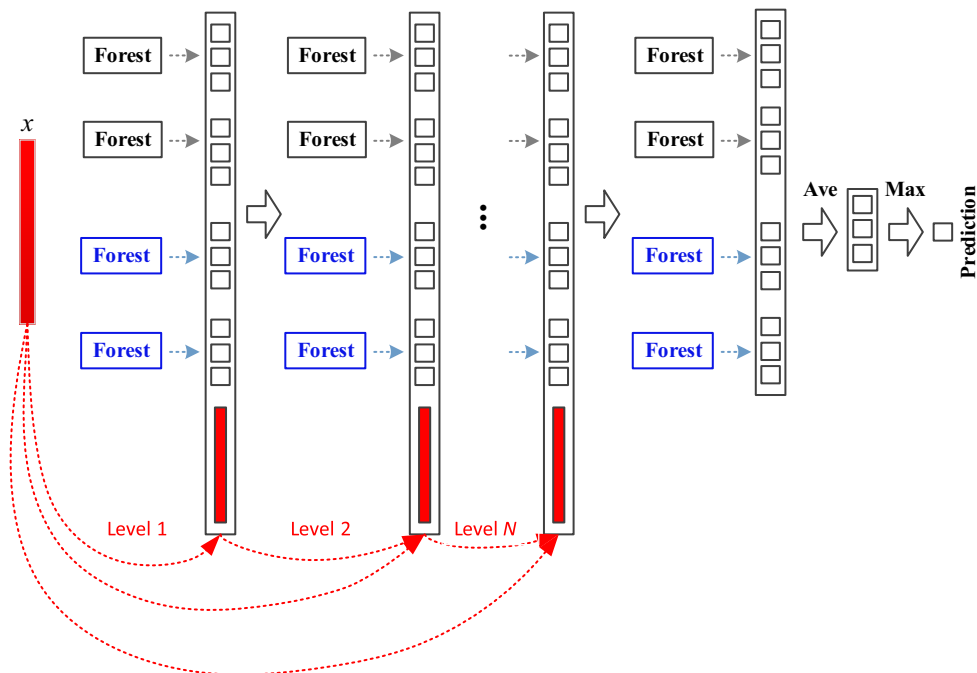


Fig. 2 The structure of dense cascade forest.

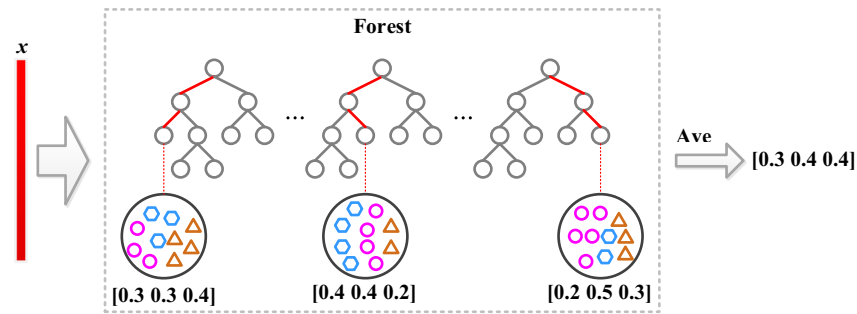


Fig. 3 Illustration of a random forest.

cascade forest simulates the layer-by-layer information processing mechanism of the neural network, the output of $l - 1$ 'th layer is considered as the input of l 'th layer. However, such a structure is prone to overfitting when the number of layers increases, which hinders the diversity of individual learners. Each layer of the cascade forest is only related to the previous layer, which leads to the homogenization of based learners. One way to solve this problem is to design a dense connection structure to maximize the feature reuse rate. In this study, to improve the utilization of features, the dense connection is borrowed to build a dense deep forest, thus avoiding the overfitting problem. Based on above analysis, the layer-by-layer information processing mechanism can be re-expressed as

$$\begin{cases} x^{(l)} = [x^{(0)}, F^{(l)}(x^{(l-1)})] = [x^{(0)}, F^{(l-1)}(x^{(l-2)}), F^{(l-2)}(x^{(l-3)}), \dots, F^{(1)}(x^{(0)})], & l = \{2, \dots, L\} \\ x^{(l)} = F^{(l)}(x^{(0)}) = [F^{(l1)}(x^{(0)}), F^{(l1)}(x^{(0)}), \dots, F^{(lm)}(x^{(0)})], & l = 1 \end{cases} \quad (1)$$

The final prediction results are averaged by the output of the last layer. Figure 3 shows the probabilistic prediction generation process of a random forest. As shown in Fig. 3, the construction of l 'th level cascaded forest is not only related to the output of level $l - 1$, but also all the outputs of the layers before the level $l - 1$. Each level of the cascade forest is the concatenation results of the probabilistic prediction that are produced by four random forests while the prediction of random forest can be calculated according to Fig. 3.

As shown in Fig. 3, the probabilistic predictions of random forest can be concretely calculated by the following rules: (1) observe the class distribution on each leaf node of each decision tree in the random forest; (2) calculate the proportion of samples of different classes at all the leaf nodes; (3) take the probabilistic result at leaf node as the prediction result of decision tree for the given instance; (4) average all the probabilistic results of all the decision trees as the prediction of the random forest. Based on the prediction process of a random forest and the structure of the dense cascade forest, we concatenate all the class vectors that are generated by four random forests as the output of each level of the dense cascade forest. To further alleviate the overfitting problem, K -fold cross-validation³⁷⁻³⁹ is incorporated in each layer of the dense cascade forest to get the robust probabilistic prediction result. As can be seen from Fig. 3, given an instance x , cloud detection can be modeled as an identification process of cloud-free, thin cloud, and thick cloud regions. As a result, a decision tree generates a three-dimensional class vector. The output of a forest is regarded as the average of the outputs of all the decision trees in a random forest.

In this work, four random forests are used as base learners to establish a cascade layer of HDForest, two of which are general random forests and two of which are completely random forests. Each random forest consists of 800 decision trees. Each tree in a completely random forest randomly selects a feature as the parent node for splitting. The training of the decision tree finished when there are less than 10 samples at each leaf node. Different from the completely random forest, a general random forest randomly selects \sqrt{d} features as candidates for node splitting, where d is the number of features. In the growth of each tree, we utilize the Gini index as the criteria for node splitting.

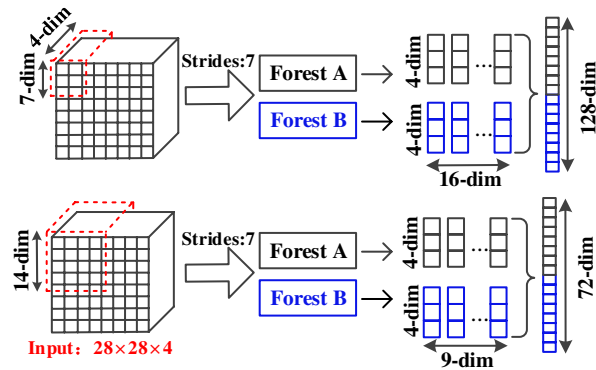


Fig. 4 The illustration of the multi-dimensional and multi-grained scanning mechanism.

2.3 Multi-Dimensional and Multi-Grained Scanning

The multi-dimensional and multi-grained scanning mechanism is inspired by the convolution operation of the CNN, which refers to a sliding window similar to CNN. In this study, the input size of each satellite cloud image sample is $28 \times 28 \times 4$, suppose we set the dimension of the sampling window as 7 and the sliding step size as 7, $[\frac{28-7}{7} + 1]^2 = 16$ subsamples can be obtained after performing a single sliding window on an original cloud sample with size $28 \times 28 \times 4$. Likewise, in the process of multi-dimensional and multi-grained scanning, if we perform two grained windows with sizes 7 and 14 with a sliding stride of 7 for feature representation, we can get 25 subsamples that are slid from an original cloud sample, including 16 subsamples with the dimension of $7 \times 7 \times 4$ and 9 subsamples with the dimension of $14 \times 14 \times 4$. Subsequently, each subsample is used for training two completely random forests and two general random forests, and each random forest predicts one subsample into a C -dimensional class vector (in this study, cloud sample images are classified into three categories: cloudless, thin cloud, and thick cloud regions, therefore, $C = 3$). All the outputs of the four random forests can be concatenated as the representation of original samples. In conclusion, an original cloud sample with size $28 \times 28 \times 4$, which performed multi-dimensional and multi-grained scanning with windows of size $7 \times 7 \times 4$ and size of $14 \times 14 \times 4$, can be transformed into a probabilistic space with the dimension of $4 \times 25 \times 3 = 300$, where 4 is the number of random forests, 25 represents the number of samples that are slid from an original cloud sample, and 3 is the dimension of each predictive class probabilistic vector (Fig. 4).

3 Results

The experimental data are collected from HJ-1A and HJ-1B. HJ-1A/1B are Chinese environmental and disaster monitoring satellites. The HJ-1A satellite is equipped with a CCD camera and a hyperspectral imager while HJ-1B has a CCD camera and an infrared camera. The design principles of the two CCD cameras on the HJ-1A and HJ-1B satellites are the same. They are placed symmetrically at the sub-satellite points, bisecting the field of view and observing in parallel. HJ-1A and HJ-1B jointly complete the push-broom imaging with a swath of 700 km, a resolution of 30 m, and 4 spectrum channels. In this study, cloud detection is modeled as the recognition of cloudless, thin cloud, and thick cloud. To well balance the feasibility of patch-wise cloud detection and the quality of the cloud detection dataset, we extract satellite cloud samples with size 28×28 and 4 spectrum channels from satellite imagery. Table 1 shows band information for the four spectral satellite pictures. In this study, 28,800 cloud detection samples are collected for the experiment using HJ-1A/1B satellite images; 80% of the samples are picked randomly for training, with 20% of the remaining samples for testing.

A brief comparison of single-spectral cloud detection is explored and numerical techniques are adopted to validate the effectiveness of MDForest. In the study, several standard cloud sensing ML/DL algorithms, which include SVM, decision tree, random forest, neural network, CNN, ResNet-34,⁴⁰ and MDForest, are initially chosen for single-spectral cloud identification. In the

Table 1 HJ-1A/1B CCD camera channel parameters.

Channel	Waveband			
	1	2	3	4
Wavelength (μm)	0.43 to 0.52	0.52 to 0.60	0.63 to 0.69	0.76 to 0.90

implementation of SVM, the radial basis function kernel⁴¹ is selected for the SVM. The decision tree is trained under the determination of the criteria of the Gini index, and the number of decision trees in the random forest is determined by hyperparameters optimization of grid search. The structure of the neural network used in our experiment is a structure with four hidden layers, each hidden layer is composed of n neurons, where n is grid searched from the value set $\{64, 128, 256\}$, each hidden layer is activated with ReLU function.⁴² In the output layer, the SoftMax activation function is performed to get the probabilistic prediction. Learning rate, which is an important parameter to the final predictive performance, is determined by grid searching from the space of $\{10^{-4}, 10^{-3}, 10^{-2}\}$.

In the implementation of CNN-based cloud detection, we adopt a CNN structure that is similar to LeNet.⁴³ CNN is composed of convolution layers and pooling layers. In this study, each pool size of each pooling layer is set to 2. A convolution layer is first performed on single/multi-spectral cloud detection images to extract the low-level features. We consider the parameter of filters in each convolution layer as a hyperparameter that needs to be finetuned, and the searching space of the number of filters in each convolution layer is set $\{64, 128, 256\}$. In the first convolutional layer, kernel size is set to 5 to enlarge the receptive field. Similar to the activation pattern of NN, each convolution layer is activated by the ReLU function, and the learning rate of CNN is searched from the value set $\{10^{-4}, 10^{-3}, 10^{-2}\}$.

Considering the running environment and the scale of the multi-spectral cloud detection, ResNet-34^{40,41} is selected for further comparison. The training strategy of ResNet-34 inherited the training pattern from NN and CNN. In this study, all NN-based cloud detection methods are optimized by Adam optimizer.

In the design of MDForest, three grained sliding windows with a sliding ratio of 0.125, 0.25, and 0.5 are first performed on original satellite cloud images. Next, one cascade layer accomplishes the realization of transforming raw cloud samples into a probabilistic feature space. Dense cascade layers are designed to learn the information from the transformed features, each layer of dense cascade forest consists of four random forests with T , where T is the number of decision trees searched from the value set of $\{100, 200, 300, 400, 500, 600\}$. Since MDForest is an ensemble algorithm that is integrated by random forests, random forest is the ensemble of decision trees. Therefore, MDForest can be regarded as an “ensemble in ensemble” algorithm. The high tree-based ensemble pattern makes MDForest a robust algorithm that has fewer hyperparameters to be finetuned. Therefore, in this study, we only focus on the fine-tuning of the parameter of the number of the decision tree in each random forest; the other parameters are consistent with the default implementation of random forest, which is realized in the scikit-learn⁴⁴ package.

Table 2 shows the performance comparison of various cloud detection algorithms on single-spectral satellite images, all the experiments are run on CPU i7-10700K with a memory size of 48 GB, and the comparison results are the averaged results of the predictions on the four wavebands.

As can be seen from Table 2, SVM performs worst for single-spectral satellite cloud detection and MDForest outperforms other cloud detection methods. Compared with SVM, the decision tree algorithm improves the performance of cloud detection while achieving quick training speed and high inference efficiency, which indicates that tree-based approaches are better choices for single-spectral cloud detection than SVM. Compared with the prediction result of the single classifier such as decision tree and SVM, the random forest classifier gets a higher accuracy for single-spectral cloud detection, which demonstrates the effectiveness of ensemble-based approaches for cloud detection. In comparison, random forest outperforms CNN and neural network, which indicates that random forest can be a good solution for single-spectral cloud

Table 2 Performance comparison of various cloud detection methods based on single-spectral satellite images. All the results are averaged on the four spectrum channels.

Algorithm	Testing accuracy (%)	Training time (s)	Testing time (s)
SVM	50.325	18.232	0.421
Decision tree	75.729	9.170	0.005
Random forest	84.132	21.013	0.207
Neural network	79.231	603.199	0.107
CNN	83.221	5932.431	0.995
MDForest	85.243	386.019	0.878

Table 3 Performance comparison of various cloud detection methods based on multi-spectral satellite images.

Algorithm	Testing accuracy (%)	Training time (s)	Testing time (s)
Random forest	89.438	54.497	0.427
Neural network	87.104	902.312	0.531
CNN	88.979	5863.212	0.920
ResNet-34	91.521	51023.231	1.766
MDForest	92.313	488.353	1.322

detection. In the comparison of training time and testing time, as can be seen from Table 3, tree-based single-spectral cloud detection methods are more efficient than neural network-based single-spectral cloud detection.

To further testify the performance of the various cloud detection methods based on single-spectral satellite images, we present the prediction results on the full single-spectral satellite images based on various cloud detection methods in Fig. 6. Figure 6(a) is the real satellite cloud image; Fig. 6(b) represents the prediction result of SVM; Fig. 6(c) indicates the prediction result of random forest, which is formed by 1000 decision trees; Fig. 6(d) shows the prediction result of the neural network, which has three hidden layers; Fig. 6(e) presents the predicted image of the CNN; Fig. 6(f) provides the prediction result of MDForest with single-dimensional and multi-grained scanning mechanism.

As can be seen from Figs. 5(a) and 5(b), SVM tends to misclassify the land region into the thin cloud area, the cloud areas that are predicted by SVM greatly differ from the real cloud distribution in Fig. 5(a). In contrast, random forest and neural network-based methods reduce the misclassification error of the cloudless area. In the comparison of the CNN and neural network, CNN shows the superiority in the prediction of the cloudless area, which mainly credits the spatial information extraction of cloud images. Finally, as can be seen from the comparison of neural network-based cloud detection methods and forest-based methods, forest-based cloud detection methods show better predictive ability, the misclassification regions of forest-based areas are smaller than that of neural network-based cloud detection methods. In conclusion, the cloud detection algorithms have a large proportion of false detection regions in the prediction of a single spectral satellite image, which motivated our exploration of multi-spectral cloud detection.

To further improve the performance of cloud detection, we integrate spectral information for multi-spectral cloud detection. Since random forest and neural network-based cloud detection show good performance in cloud detection, we focus on the performance comparison of neural network-based multi-spectral cloud detection. Figure 6 shows the performance comparison of

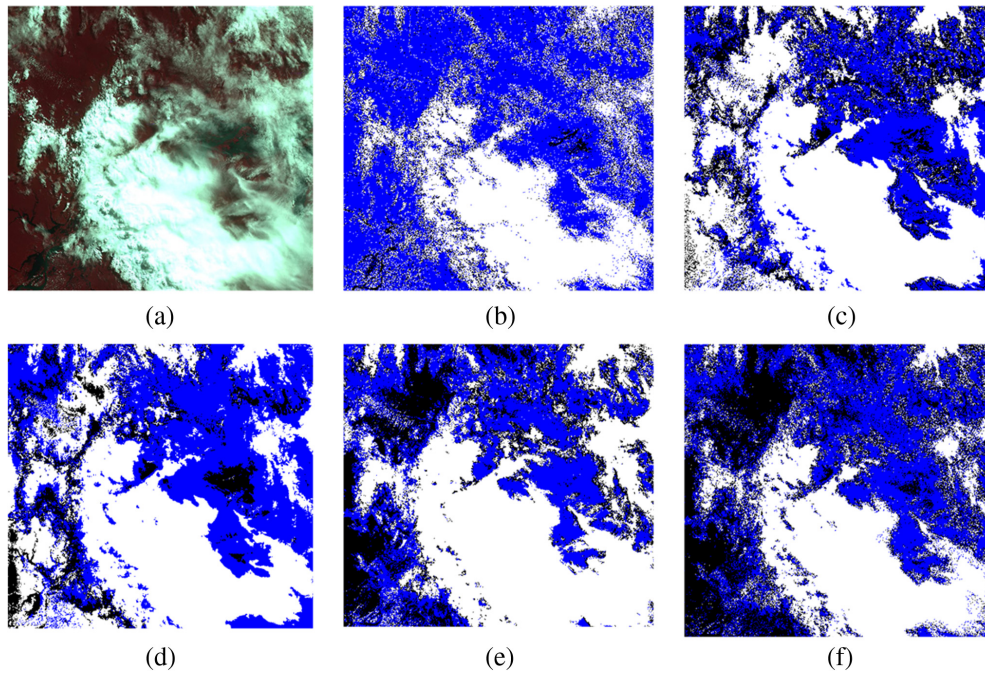


Fig. 5 Prediction visualization of a satellite image from HJ-1A/1B: in the illustration of the prediction images, black color is the prediction of the cloudless area, the blue color represents the regions of thin cloud, and the white color indicated the predicted area is covered with thick cloud. (a) The original cloud image, (b) the prediction of SVM on the single-spectral satellite image, (c) the prediction random forest, (d) the prediction of the neural network, (e) the prediction of CNN on fourth waveband, and (f) the prediction of MDForest on fourth waveband.

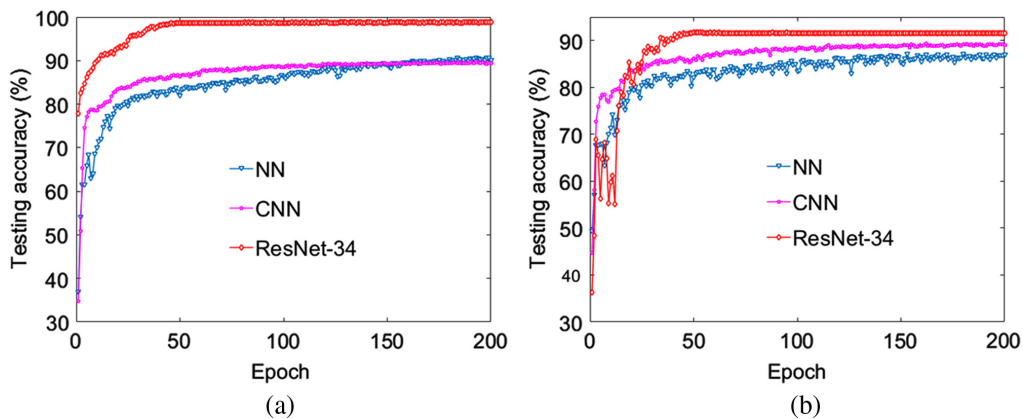


Fig. 6 Performance comparison of neural network-based cloud detection methods using multi-spectral satellite images. (a) The comparison results of the training accuracy curves of the neural network, CNN, and ResNet-34. (b) The comparison results of testing accuracy curves of the neural network, CNN, and ResNet-34.

neural network-based multi-spectral cloud detection methods. As can be seen from Fig. 6, ResNet outperforms the CNN and neural network, which indicates that a deeper network structure has a stronger ability to extract features from low-level to a higher level. In addition, as shown in Fig. 6, the training accuracy of the neural network is slightly higher than that of CNN while its testing accuracy is slightly worse than that of the neural network, which implies that CNN is more suitable for multi-spectral cloud detection due to its superior spatial and spectral feature extraction ability. Based on the prominent performance of CNN-based multi-spectral

cloud detection, we further compared the performance of neural network-based cloud detection methods to the forest-based cloud detection methods to verify the effectiveness of MDForest.

Table 3 shows the performance comparison of the neural network-based and forest-based cloud detection methods. As shown in Table 3, MDForest achieves comparable accuracy to ResNet while MDForest costs less on training/testing single-spectral cloud images than ResNet-34. As can be seen from Table 3, though ResNet improves the performance of cloud detection based on multi-spectral satellite images, the high complexity of ResNet limited its applicability for practical cloud detection, which is not suitable for fast cloud detection. By comparison, MDForest satisfied the need where the high efficiency and good accuracy of cloud detection are both required. Combined with Tables 2 and 3, it can be seen that random forest is a robust cloud detection method that has good detection efficiency, but from the perspective of the realization of accurate cloud detection, MDForest is the best choice.

Since NN-based cloud detection methods based on multi-spectral satellite images are good at extracting spatial and spectral information, we further verify the generalization ability of neural network-based cloud detection methods and MDForest. Figure 7 shows the comparison of the prediction images of neural network-based methods and MDForest. Figure 7(a) shows the prediction of the neural network based on a multi-spectral satellite image, and the real cloud image is shown in Fig. 5(a). Figure 7(b) presents the prediction of CNN based on the multi-spectral satellite image, Fig. 7(c) provides the prediction result of ResNet based on the multi-spectral satellite image, and Fig. 7(d) shows the prediction result of MDForest based on the multi-spectral satellite image.

As can be seen from Fig. 7, which refers to the original satellite images in Fig. 5(a), the prediction of the neural network is the poorest in the above comparison methods since the neural network cannot effectively utilize spectral information. In addition, the learning mechanism of the neural network is concluded as a fitting process, which cannot well extract the spatial information of multi-spectral cloud samples, making the improvement insufficient compared with other spatial and spectral information learning approaches. As shown in Fig. 7(a), the cloud detection results of the neural network have a high false detection on the area of thin cloud area, the predicted thick cloud region is larger than the other three cloud detection methods.

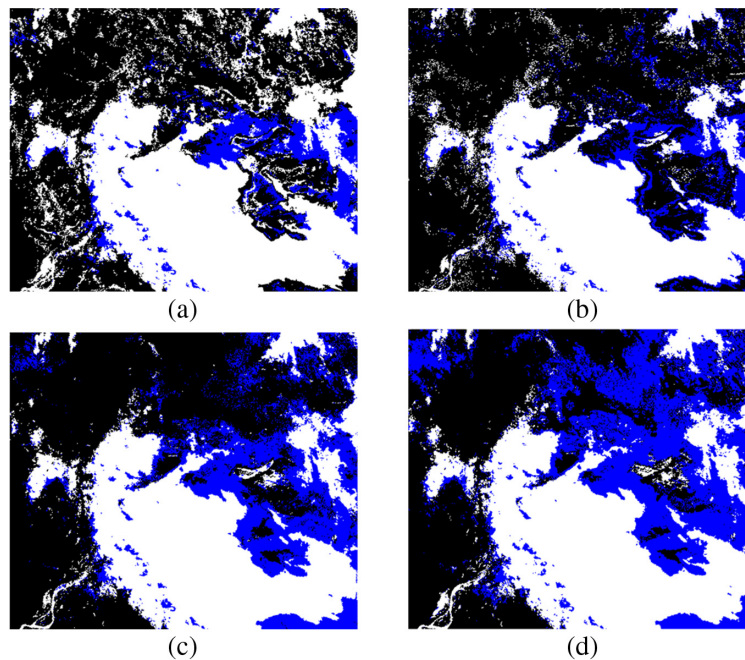


Fig. 7 Prediction results of advanced neural network-based cloud detection methods and MDForest using multi-spectral satellite image #1. (a) The prediction of the NN on the multi-spectral satellite image; (b) the prediction of CNN on the multi-spectral satellite image; (c) the result that is predicted by ResNet based on the multi-spectral satellite image; (d) the multi-spectral cloud detection result of MDForest.

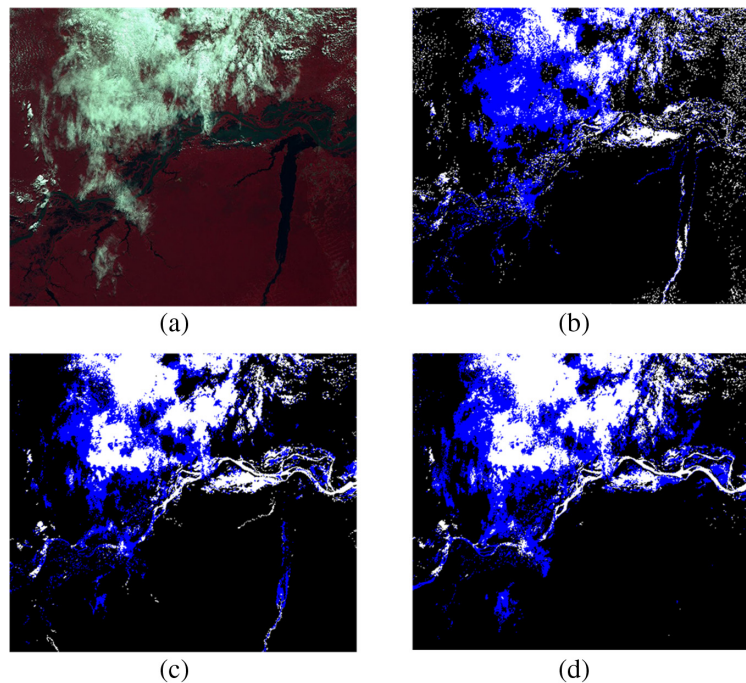


Fig. 8 Prediction results of advanced CNN-based cloud detection methods and MDForest for multi-spectral satellite image #2. (a) The real satellite image #2; (b) the prediction result of CNN for multi-spectral satellite image #2; (c) the predicted image of ResNet-34 for multi-spectral satellite image #2; (d) the prediction result of MDForest for multi-spectral satellite image #2.

Compared with the real images in Fig. 5(a), the misclassification error of thin cloud to thick cloud is high, making the neural network not a suitable method for multi-spectral cloud detection. In comparison, the prediction results of CNN, ResNet, and MDForest are close to the real image, and the misclassification areas of thin cloud have been greatly reduced. As can be seen from the comparison of Figs. 7(b) and 7(c), the prediction of ResNet is comparable to the prediction result of CNN since ResNet is a deeper framework that inherits the learning pattern and feature extraction mechanism from CNN. As illustrated in Fig. 7, MDForest achieves the best prediction result. Although the misclassification areas that are predicted by MDForest on the complex ground cover (such as the river) are large when compared to CNN's predicted image, MDForest's accuracy in predicting thin cloud patches is significantly improved.

To further validate the effectiveness of MDForest for multi-spectral cloud detection, two groups on two random satellite multi-spectral cloud images are compared. Figures 8 and 9 show the prediction results of CNN, ResNet-34, and MDForest; Fig. 8(a) shows the real satellite image for the multi-spectral satellite image #2; Fig. 8(b) shows the prediction result of CNN for the multi-spectral satellite image #2; Fig. 8(c) shows the prediction of ResNet-34; Fig. 8(d) presents the prediction results of MDForest the multi-spectral satellite image #2. Figure 9(a) shows the real satellite image #3; Fig. 9(b) shows the prediction result of CNN for the multi-spectral satellite image #3; Fig. 9(c) shows the predicted image of ResNet-34 for multi-spectral satellite image #3; Fig. 9(d) shows the prediction result of MDForest for multi-spectral satellite image #3.

As can be seen from Fig. 8, CNN and ResNet misclassified more thick cloud areas into thin cloud regions than MDForest. Based on the comparison of prediction images of the neural network, ResNet, and MDForest using multi-spectral satellite imagery, the prediction result of MDForest has a higher similarity with the real image on the distribution of cloud-free areas, thin cloud regions, and the land covered by thick cloud. In the prediction of the cloud-free area, MDForest shows significant improvement compared with the prediction results of CNN and ResNet-34. As can be seen from Fig. 9, in the area that is covered with rivers, MDForest gets a relatively accurate prediction result while CNN and ResNet-34 classified more areas of rivers into thin cloud areas or thick cloud regions.

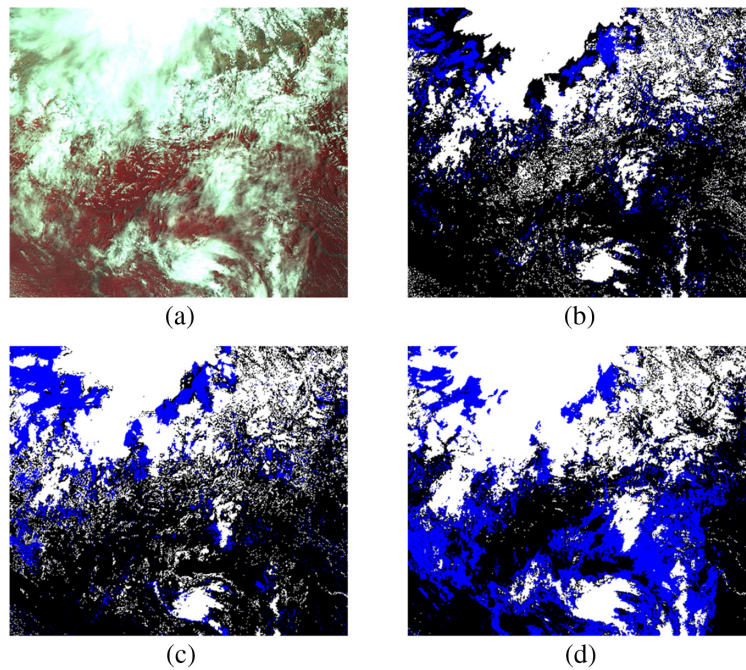


Fig. 9 Prediction results of advanced neural network-based cloud detection methods and MDForest for multi-spectral satellite image #3. (a) The real satellite image #3; (b) the prediction result of CNN for the multi-spectral satellite image #3; (c) predicted image of ResNet-34 for multi-spectral satellite image #3; (d) the prediction result of MDForest for multi-spectral satellite image #3.

MDForest achieves good cloud detection performance using multi-spectral satellite imagery due to the robustness of tree-based deep structure and the multi-dimensional and multi-grained scanning mechanism. Compared with neural network-based cloud detection methods, the parameters fine-tuning process of MDForest is much simpler. Consequently, we further study the influence of different parameter settings on the performance of cloud detection. Figure 10 shows the performance of MDForest under a different number of trees in a random forest and

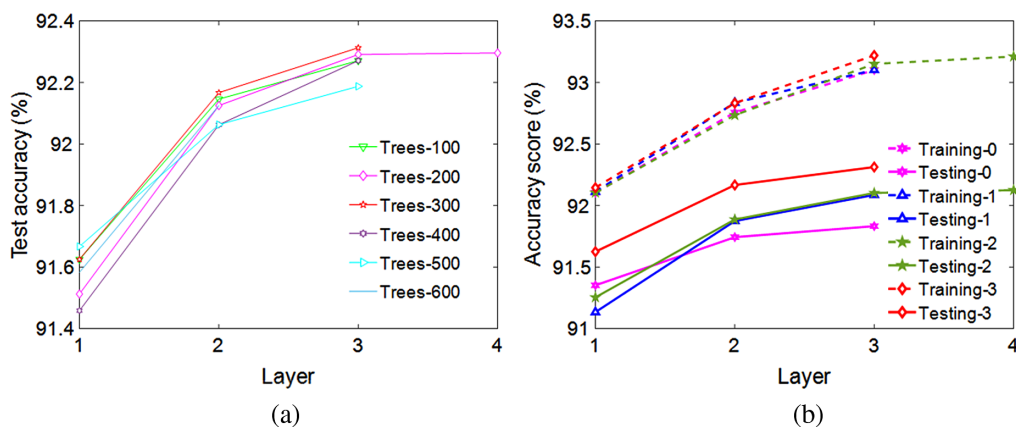


Fig. 10 The performance of MDForest under different parameter settings. (a) The influence of the number of trees in each forest on the performance of cloud detection using multi-spectral satellite images. (b) The performance comparison with different scale sliding windows, training-0 is the training curve of MDForest without multi-dimensional and multi-grained scanning mechanism; testing-0 is the testing curve of MDForest without multi-dimensional and multi-grained scanning mechanism; training-1 is the training curve of MDForest with 1-grained multi-dimensional scanning; testing-1 is the testing curve of MDForest with 1-grained multi-dimensional scanning, and so on.

different grained scanning. Figure 10(a) shows the testing accuracy comparison under different numbers of trees in a random forest, Fig. 10(b) shows the comparison of training accuracy and testing accuracy comparison under different grained scanning. As can be seen from Fig. 10(a), MDForest is an adaptive deep forest that deepens its structure according to the complexity of input data. When the number of decision trees in a random forest is set to 200, MDForest deepens its structure to layer 4 while the three layered MDForest seems to be a model complex enough to deal with the input data. Such character makes MDForest a superior data-driven cloud detection method to NN-based cloud detection methods whose optimal structure is determined by manually fine-tuning. Figure 10(a) demonstrates that MDForest gets the optimal testing accuracy in the setting of each forest is ensembled by 300 decision trees; when the number of decision trees is determined as lower than 300, more decision trees for random forest in HDForest imply the better performance HDForest can get; when the number of decision trees in a random forest exceeds the threshold value of 300, overfitting occurs. Moreover, Fig. 10(b) indicates that more grained sliding windows are beneficial to the performance improvement of cloud detection using multi-spectral satellite images, which further verifies the effectiveness of MDForest.

4 Conclusions

Though CNN-based methods realized high-performance multi-spectral cloud detection by extracting and integrating spatial and spectral information, the performance improvement on cloud detection was based on the increase of model complexity, which hindered the progress for fast cloud detection. In this study, we proposed a multi-dimensional and multi-grained dense deep forest for cloud detection using multi-spectral satellite imagery. The proposed method was a deep forest structure that simulated the layer-by-layer processing mechanism of NN. In addition, the multi-layered structure gave the proposed method the representation learning ability, allowed it automatically to extract features of satellite cloud images end-to-end. Moreover, the multi-dimensional and multi-grained scanning mechanism possessed the power to deal with multi-spectral satellite cloud detection, which further improved the performance of cloud detection. Finally, a densely connected structure was borrowed in the proposed method to avoid the overfitting problem. Experimental results on HJ-1A/1B demonstrated that the proposed MDForest improved the performance of multi-spectral cloud detection while getting a better cloud detection efficiency, which can be regarded as an alternative to CNN-based methods for multi-spectral cloud detection.

MDForest improved the efficiency and performance of cloud detection based on multi-spectral satellite cloud images. However, there are still some problems that need to be addressed in future work, which include: (1) the recognition of ground area is not ideal enough, in the future work, some noisy samples such as lakes, rivers, and fogs will be collected to enrich the diversity of cloud detection dataset; (2) MDForest improves the efficiency of cloud detection, implementation on some mobile devices can achieve the goal of fast cloud detection. Therefore, in our future work, embedding MDForest into some hardware would be a good solution for practical cloud detection.

Acknowledgments

This research received no external funding. Conceptualization, Ming Shao; methodology, Yao Zou; validation, Ming Shao; formal analysis, Yao Zou; writing—original draft preparation, Ming Shao; writing—review and editing, Ming Shao; supervision, Yao Zou; project administration, Yao Zou. All authors have read and agreed to the published version of the manuscript. The authors declare no conflicts of interest.

References

1. F. Xie et al., "Multilevel cloud detection in remote sensing images based on deep learning," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **10**(8), 3631–3640 (2017).

2. M. Shi et al., "Cloud detection of remote sensing images by deep learning," in *Int. Geosci. and Remote Sens. Symp.*, Institute of Electrical and Electronics Engineers Inc., pp. 701–704 (2016).
3. M. Segal-Rozenhaimer et al., "Cloud detection algorithm for multi-modal satellite imagery using convolutional neural-networks (CNN)," *Remote Sens. Environ.* **237**, 111446 (2020).
4. H. Shang et al., "Diurnal cycle and seasonal variation of cloud cover over the Tibetan Plateau as determined from Himawari-8 new-generation geostationary satellite data," *Sci. Rep.* **8**(1), 1–8 (2018).
5. X. Y. Zhuge, X. Zou, and Y. Wang, "A fast cloud detection algorithm applicable to monitoring and nowcasting of daytime cloud systems," *IEEE Trans. Geosci. Remote Sens.* **55**(11), 6111–6119 (2017).
6. H. Fu et al., "Cloud detection for FY meteorology satellite based on ensemble thresholds and random forests approach," *Remote Sens.* **11**(1), 44 (2018).
7. R. Gupta, S. J. Nanda, and U. P. Shukla, "Cloud detection in satellite images using multi-objective social spider optimization," *Appl. Soft Comput. J.* **79**, 203–226 (2019).
8. C. Deng et al., "Cloud Detection in satellite images based on natural scene statistics and Gabor features," *IEEE Geosci. Remote Sens. Lett.* **16**(4), 608–612 (2019).
9. J. Wei et al., "Dynamic threshold cloud detection algorithms for MODIS and Landsat 8 data," in *Int. Geosci. and Remote Sens. Symp.*, Institute of Electrical and Electronics Engineers Inc., pp. 566–569 (2016).
10. Z. Shao et al., "Fuzzy AutoEncode based cloud detection for remote sensing imagery," *Remote Sens.* **9**(4), 311 (2017).
11. 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.* **159**, 269–277 (2015).
12. S. Foga et al., "Cloud detection algorithm comparison and validation for operational Landsat data products," *Remote Sens. Environ.* **194**, 379–390 (2017).
13. P. Li et al., "A cloud image detection method based on SVM vector machine," *Neurocomputing* **169**, 34–42 (2015).
14. C. Huang et al., "An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks," *Remote Sens. Environ.* **114**(1), 183–198 (2010).
15. Y. Yuan and X. Hu, "Bag-of-words and object-based classification for cloud extraction from satellite imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **8**(8), 4197–4205 (2015).
16. T. Bai et al., "Cloud detection for high-resolution satellite imagery using machine learning and multi-feature fusion," *Remote Sens.* **8**(9), 715 (2016).
17. Y. Sui, B. He, and T. Fu, "Energy-based cloud detection in multispectral images based on the SVM technique," *Int. J. Remote Sens.* **40**(14), 5530–5543 (2019).
18. 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.* **8**(11), 963 (2016).
19. J. Wei et al., "Cloud detection for Landsat imagery by combining the random forest and superpixels extracted via energy-driven sampling segmentation approaches," *Remote Sens. Environ.* **248**, 112005 (2020).
20. L. Gomez-Chova et al., "Cloud detection machine learning algorithms for PROBA-V," in *Int. Geosci. and Remote Sens. Symp.*, Institute of Electrical and Electronics Engineers Inc., pp. 2251–2254 (2017).
21. P. Tannor and L. Rokach, "AugBoost: gradient boosting enhanced with step-wise feature augmentation," in *Int. Joint Conf. Artif. Intell.*, pp. 3555–3561 (2019).
22. M. Xia et al., "Cloud/snow recognition for multispectral satellite imagery based on a multi-dimensional deep residual network," *Int. J. Remote Sens.* **40**(1), 156–170 (2019).
23. Z. Shao et al., "Cloud detection in remote sensing images based on multiscale features-convolutional neural network," *IEEE Trans. Geosci. Remote Sens.* **57**(6), 4062–4076 (2019).
24. A. Francis, P. Sidiropoulos, and J.-P. Muller, "CloudFCN: accurate and robust cloud detection for satellite imagery with deep learning," *Remote Sens.* **11**(19), 2312 (2019).

25. S. Mohajerani and P. Saeeedi, "Cloud-Net: an end-to-end cloud detection algorithm for Landsat 8 imagery," in *Int. Geosci. and Remote Sens. Symp.*, pp. 1029–1032 (2019).
26. J. Yang et al., "CDnet: CNN-based cloud detection for remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.* **57**(8), 6195–6211 (2019).
27. S. Ji et al., "Simultaneous cloud detection and removal from bitemporal remote sensing images using cascade convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.* **59**(1), 732–748 (2021).
28. L. Deng and D. Yu, "Deep learning: methods and applications," *Found. Trends Signal Process.* **7**(3–4), pp. 197–387 (2013).
29. L. V. Utkin and M. A. Ryabinin, "A Siamese deep forest," *Knowl.-Based Syst.* **139**, 13–22 (2018).
30. C. Rudin, "Algorithms for interpretable machine learning," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery and Data Mining*, Association for Computing Machinery (ACM), New York, pp. 1519–1519 (2014).
31. M. Du, N. Liu, and X. Hu, "Techniques for interpretable machine learning," *Commun. ACM* **63**(1), 68–77 (2019).
32. Z.-H. Zhou and J. Feng, "Deep forest," *Natl. Sci. Rev.* **6**(1), 74–86 (2017).
33. M. Pang et al., "Improving deep forest by confidence screening," in *Proc.- IEEE Int. Conf. Data Mining*, Institute of Electrical and Electronics Engineers Inc., pp. 1194–1199 (2018).
34. Z. Yang et al., "An improved scheme for rice phenology estimation based on time-series multispectral HJ-1A/B and polarimetric RADARSAT-2 data," *Remote Sens. Environ.* **195**, 184–201 (2017).
35. M. Singha, B. Wu, and M. Zhang, "Object-based paddy rice mapping using HJ-1A/B data and temporal features extracted from time series MODIS NDVI data," *Sensors* **17**(12), 10 (2016).
36. G. Huang et al., "Densely connected convolutional networks," in *Proc. 30th IEEE Conf. Comput. Vision and Pattern Recognit., CVPR*, pp. 2261–2269 (2017).
37. T. T. Wong and N. Y. Yang, "Dependency analysis of accuracy estimates in k-fold cross validation," *IEEE Trans. Knowl. Data Eng.* **29**(11), 2417–2427 (2017).
38. Y. Liu et al., "Fast cross-validation for kernel-based algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(5), 1083–1096 (2020).
39. J. Bi et al., "Impacts of snow and cloud covers on satellite-derived PM2.5 levels," *Remote Sens. Environ.* **221**, 665–674 (2019).
40. K. He et al., "Deep residual learning for image recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognit.*, pp. 770–778 (2016).
41. G. Venkatesh, E. Nurvitadhi, and D. Marr, "Accelerating deep convolutional networks using low-precision and sparsity," in *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. – Proc.*, Institute of Electrical and Electronics Engineers Inc., pp. 2861–2865 (2017).
42. M. Khalid et al., "Empirical evaluation of activation functions in deep convolution neural network for facial expression recognition," in *43rd Int. Conf. Telecommun. Sign. Process., TSP 2020*, pp. 204–207, Institute of Electrical and Electronics Engineers Inc. (2020).
43. A. El Sawy, H. El-Bakry, and M. Loey, "CNN for handwritten arabic digits recognition based on LeNet-5," *Adv. Intell. Syst. Comput.* **533**, 565–575 (2017).
44. F. Pedregosa et al., "Scikit-learn: machine learning in Python," *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

Ming Shao received his PhD in management science and engineering from Shanghai Donghua University, China, in 2010. He is currently engaged in postdoctoral research at Fudan University. He has expertise in the domains of cloud computing, machine learning (ML), big data analysis, and so on.

Yao Zou is currently pursuing his PhD in the field of ML, big data analysis at Shanghai Donghua University, China. His expertise mainly includes remote sensing applications for the environment and natural sources.