IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 5, 2021, accepted September 17, 2021, date of publication September 20, 2021,
date of current version September 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3114185

Enhanced Multi-Dimensional and Multi-Grained Cascade Forest for Cloud/Snow Recognition Using Multispectral Satellite Remote Sensing Imagery

MENG XIA, (Graduate Student Member, IEEE), ZHIJIE WANG, FANG HAN, AND YANTING KANG

College of Information Science and Technology, Donghua University, Shanghai 201620, China

Corresponding author: Zhijie Wang (wangzj@dhu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 11972115 and Grant 11572084.

ABSTRACT Cloud/snow recognition is one application of satellite remote sensing imagery in natural disaster monitoring. Deep learning technology has contributed to the improvement of the performance of cloud/snow recognition. However, deep learning-based methods cannot well balance the performance and efficiency of cloud/snow recognition. In this paper, an augmented multi-dimensional and multi-grained Cascade Forest is proposed for cloud/snow recognition. The multi-dimensional deep forest structure with the representation learning ability allows it to capture the spatial and spectral information of cloud/snow satellite imagery accordingly equipped with good recognition efficiency. Besides, a simple augmentation Random Erasing method is introduced for enhancing the robustness of cloud/snow recognition. The experimental results on the HJ-1A/1B dataset show that the proposed method improves the performance of cloud/snow recognition by extracting spectral information from multi-spectral satellite imagery. In addition, based on the tree-based structure, the proposed method well balances the performance and efficiency of cloud/snow recognition, which can be considered as an alternative to the Neural Network for cloud/snow recognition.

INDEX TERMS Cloud/snow recognition, multi-dimensional and multi-grained, random erasing, representation learning.

I. INTRODUCTION

The plateau areas are covered with snow all year round, and snow disasters happen irregularly. The lack of snow disaster monitoring will seriously affect the development of agriculture in the plateau areas and will cause irretrievable and disastrous effects on the production and life of people and social order. Therefore, it is particularly important to perform a good job of forecasting heavy snow in plateau areas. The development of remote sensing technology provides an effective observation means for snow disaster monitoring. However, there are still some factors affecting the monitoring of snow disasters, among which cloud interference is the most important factor causing difficulties in the monitoring of snow disasters. The imagery captured by remote sensing satellites has the characteristics of a wide range and high spatial resolution. Using image techniques can identify the distribution of snow and clouds. For satellite remote sensing imagery containing clouds and snow, the pixels of cloud and snow in the satellite remote sensing images are very close, the local textures are similar, and the spectral features of the panchromatic band of clouds and snow are similar. Therefore, accurate recognition of clouds and snow have become one difficulty in the monitoring of snow disasters. The main existing challenges in cloud/snow recognition are follows. 1) Cloud hinder the propagation of light signals, resulting in low measurement accuracy of remote sensing data. 2) The local texture of cloud and snow has a high degree of similarity, and there is also a certain similarity between the spectra, so it is difficult to extract the spectral features and texture features of cloud and snow in satellite imagery. 3) Other factors on the ground (mountains, shadows, etc.) have a serious impact on cloud/snow recognition; 4) The distribution of snow and cloud is mainly in plateau areas, but there are few samples of cloud maps in plateau areas.

The research methods of cloud/snow recognition mainly start with the spectral information [1] and texture information [2], [3] of cloud and snow. Bai *et al.* [4] compared and

The associate editor coordinating the review of this manuscript and approving it for publication was Jon Atli Benediktsson.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

analyzed the spectrum, texture, and other features between the cloud and the background in remote sensing imagery, then they proposed a cloud recognition method using machine learning and multi-feature fusion, which combines spectral information with texture features. The results are far more accurate than spectral-based or texture-based methods. Traditional machine learning algorithms based on feature engineering [5], [6] are easy to understand, and the model design is simple, but they have the following shortcomings. 1) Traditional machine learning models are time-consuming and labor-intensive to extract the spectral and texture features of satellite imagery. 2) Traditional machine learning methods have few parameters and cannot extract complex semantic features, resulting in poor model generalization performance. 3) Machine learning algorithms based on feature engineering usually require complex feature extraction and analysis, which makes it impossible to effectively capture the spatial features of the image.

Thanks to the application of large-scale graphics computing equipment and technology, Neural Networks have developed rapidly. The Deep Neural Network (DNN) can effectively capture feature information by transmitting the labeled data to the multi-layer Neural Network for training [7], [8]. It is an end-to-end network model that can alleviate the arduous and complicated process of cloud/snow recognition. The process of feature extraction and feature analysis can extract the relevant information of the original image to the greatest extent, so the accuracy has been greatly improved. Hughes and Hayes [9] used DNN for cloud/snow recognition, and its research results have been able to train a Deep Neural Network to distinguish between cloud and snow. Pilipović and Risojević [10] proposed a cloud/snow recognition algorithm that had used an improved pre-trained convolutional Neural Network for end-to-end training. The application of the pre-training model greatly improved the accuracy and speed of cloud/snow recognition in remote sensing imagery. Xie *et al.* [11] conducted a multi-level cloud recognition method based on deep learning, which uses convolutional layers to extract multi-scale features, then exploited improved simple linear iterative clustering (SLIC) to accurately obtain cloud boundary information. This method effectively distinguished the cloud and other backgrounds, and improve the efficiency of cloud/snow recognition. Liu *et al.* [12] combined the super-pixel level detection method with the convolutional Neural Network and made full use of the texture information of the imagery, which greatly improved the accuracy of recognition. However, this method took a lot of time to annotate remote sensing imagery. Although the accuracy was improved, the recognition speed was reduced. Le Goff's [13] and Zhu and Helmer [14] researches utilized the texture and spatial features of satellite remote sensing imagery to analyze clouds and snow, made full use of the advantages of Neural Networks in feature extraction, and combined the texture features of clouds and snow with spectral features, which had seen an increase both in recognition accuracy and recognition speed. Xia *et al.* [15] proposed to use the multi-dimensional deep residual network (M-ResNet) to extract the image features and spectral information of satellite images, which solved the problems of gradient diffusion and model degradation, thereby improving the recognition accuracy. The cloud/snow recognition algorithm based on deep learning involves a Deep Neural Network. As the number of layers increases, it will cause the problem of gradient disappearance or gradient explosion. At the same time, the Deep Neural Network model has the shortcomings of a large number of parameters, slow training and testing speed, and poor generalization performance.

Although the existing cloud/snow recognition algorithms based on deep learning [16] can make good use of the spectral and texture features of remote sensing images [17], [22], almost all models need to have a pre-set structure. Some models cannot accurately distinguish between clouds and snow with similar spectra, which leads to large deviations in the recognition results compared with the original remote sensing imagery. Meanwhile, methods based on deep learning require an enormous amount of labeled image data to train the network and require high computing equipment [18], [19]. This requires a lot of computing time and resources, making many high-level models impossible to apply to real life.

Ensemble learning [23] is a type of machine learning algorithm, which improves the recognition accuracy of the learner by integrating multiple weak learners, and can avoid the problems of gradient disappearance [20] and model degradation [21]. Brown de Colstoun *et al.* [24] developed an enhanced decision tree classifier for land cover recognition on multi-temporal remote sensing imagery and achieved an overall accuracy of 82% on ground truth data. Cheng and Lin *et al.* [25] used supervised learning technology with multi-resolution function to integrate Random Forest and Support Vector Machine to the remote sensing imagery. Xia *et al.* [26] applied multi-scale Cascade Forest to remote sensing satellite imagery recognition. The results showed that the ensemble learning algorithm had great advantages in cloud/snow recognition. Ensemble learning can make full use of the characteristics of the imagery and improve the generalization ability and predictive ability of the model. However, the current application of ensemble algorithms in cloud/snow recognition still has the following shortcomings. 1) Most ensemble learning algorithms do not fully consider the correlation between space, texture, and spectrum of remote sensing images, and feature extraction is difficult. 2) Affected by fewer meteorological observation stations in plateau areas, fewer satellite imagery datasets can be used. However, the existing cloud/snow recognition algorithms do not consider the advantage of image enhancement methods to improve the reliability and diversity of the data. 3) Most of the existing ensemble learning methods are based on the decision tree, which requires higher device memory.

Based on the problems of traditional algorithms, deep learning algorithms, and existing ensemble algorithms in cloud/snow recognition, this paper develops a multi-grained sampling Cascade Forest (RES-gcForest) algorithm based

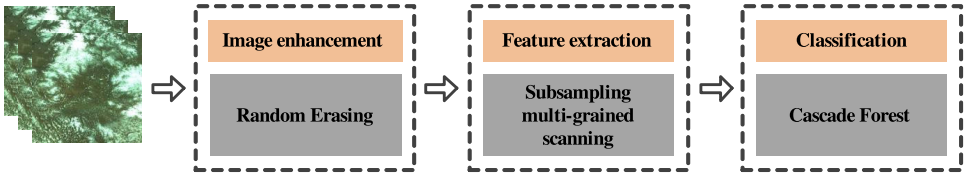


FIGURE 1. The structure diagram of RES-gcForest.

on the Random Erasing image enhancement method, which can achieve high accuracy and fast cloud/snow recognition. The RES-gcForest model exploits Random Erasing to perform image enhancement operations on image data, thereby improving the diversity and reliability of the data, reducing the risk of over-fitting due to unreliable or fewer data. The most important thing is to make the model robust to remote sensing imagery. Therefore, RES-gcForest can better learn small datasets. At the same time, representation learning converts the original data into a data form that can be recognized by the machine, avoiding the trouble of manually extracting features, so that the computer learns how to extract features while learning to use features. Scanning with multiple windows and different granularities in RES-gcForest can extract more features than CNN, and better extract the spatial information and spectral information of the image. Therefore, it is possible to extract as much pixel information and texture information from the original image as possible while maintaining the internal spatial correlation information of the image. Sub-sampling the instances after multi-grained scanning can reduce the computational complexity of the model. RES-gcForest does not need to determine the structure in advance, because each layer adopts supervised learning, so the number of layers of RES-gcForest is adaptive. Compared with the time consumed by deep learning caused by parameter adjustment, RES-gcForest reduces the number of parameters in the same receiving field and improves the learning ability. Because the RF in RES-gcForest can be processed in parallel, the training speed of the RES-gcForest model is further improved to a certain extent. The main contributions of the algorithm proposed in this paper are as follows:

1) Exploit Random Erasing to perform image enhancement operations on the image dataset to improve the robustness of the model and reduce the risk of over-fitting.
2) Multi-grained and multi-window scanning can improve the diversity of extracted features. Taking advantage of two different Random Forests further increases the diversity of features.
3) The sub-sampling of transformation instances in the model reduces the computational parameters and the computational complexity of the model. The parallel processing of Random Forest in the model can further reduce the memory requirement and time cost, which can improve the training speed of the model.
4) Each layer of RES-gcForest adopts supervised learning, so the number of layers of RES-gcForest is adaptive, and there is no need to manually set the number of cascading layers. The entire training process of RES-gcForest only has forward propagation, which is very different from the deep learning of gradient update through error backpropagation. Forward propagation avoids fluctuations caused by various factors when errors are propagated back.

The rest of this article is organized as follows. The second part describes the RES-gcForest model. The third section introduces the dataset and experimental settings. The fourth section analyzes the experimental results. Finally, the fifth section gives a conclusion. In the experimental part, the effects of various algorithms in single-spectrum and multi-spectral remote sensing imagery cloud/snow recognition were compared. The comparison methods are gcForest, Cascade Forest, Support Vector Machine, Random Forest, Neural Network, and Convolutional Neural Network. The experimental results show that the accuracy of the RES-gcForest method on single-spectrum or multi-spectral images is the best among all methods, and it has good prediction results and performance.

II. METHODOLOGY

Fig. 1 shows the RES-gcForest model. This paper applies the idea of image enhancement and random sampling to the process of satellite image processing and feature extraction, using multiple sliding windows of different granularities to extract features from satellite images, then randomly sample the extracted features, and finally input to the cascade forest to get the predicted result. The RES-gcForest model divides the training process into three stages. 1) Use Random Erasing to perform image enhancement operations on satellite cloud and snow images. 2) Input the processed images into three sliding windows with different granularities to extract features, Then the extracted features are randomly sampled. Then input the randomly sampled features into the Random Forest to obtain the class probability vector. 3) The class probability vectors obtained from the three windows are spliced, and the new feature vector obtained by splicing is used as the input of the Cascade Forest. The image enhancement method used by RES-gcForest can not only reduce the risk of over-fitting, make the model robust to satellite cloud images, but also enrich the dataset of satellite cloud images in plateau areas that are already scarce. Introducing the idea of sub-sampling can reduce the redundancy of features and speed up the training of the model.

The RES-gcForest is a tree-based structure that simulates the representation learning pattern of NN-based cloud/snow recognition. Multi-grained scanning is borrowed from the mechanism of the convolutional layer of a convolutional neural network [27] while cascade forest is a hierarchical framework that realizes end-to-end representation learning for cloud/snow recognition, making the variant of the deep forest a spatial-aware model for cloud/snow recognition.

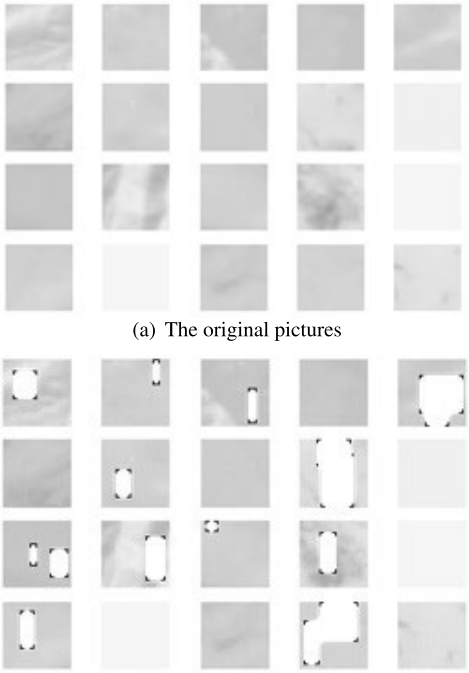
RES-gcForest is a variant of the deep forest. In RES-gcForest, Random Erasing is performed to augment the cloud/snow dataset due to the insufficient training cloud/snow training samples and the high cost of the cloud/snow labeling process. Cascade forest is a hierarchical framework that realizes layer-by-layer representation learning by ensemble parallel forest-based layers in a cascading fashion. Multi-grained scanning simulates the feature extraction pattern and extracts features with multi-grained sliding windows, which borrows the conception of multiple instance learning, re-representing the original cloud/snow satellite images into a probabilistic space. More scale sliding windows imply more adequate feature extraction.

A. RANDOM ERASING

Because there are few cloud and snow samples in plateau areas, and most of the cloud/snow recognition algorithms do not perform image enhancement operations on the original satellite images, the model is not robust to the image, and the recognition accuracy is low. In order to improve the generalization ability of the model and enhance the robustness of the model to image data, this paper introduces a data enhancement method, namely Random Erasing [28]. It can be easily implemented in most existing machine learning models.

Random Erasing is a lightweight method that does not require any additional parameter learning or memory consumption. It also can be integrated with any image processing method that uses machine learning models, and does not need to change the learning strategy of machine learning; Random Erasing can expand small-scale image data, and it is also a supplement to the regularization method. The robustness of the model will increase when the Random Erasing is applied.

Fig. 2 shows the result of the Random Erasing of part of the data used in this article. The picture is a single-channel picture, the upper side represents the original picture, and the down side is the picture data after Random Erasing. It can be clearly seen that picture after Random Erasing has an erased rectangular area on the basis of the original picture.



(a) The original pictures

(b) Pictures after Random Erasing

FIGURE 2. Pictures before and after random erasing.

B. SUB-SAMPLING MULTI-GRAINED SCANNING

Deep Neural Networks (DNN) are powerful in processing feature relationships [29]. For example, Convolutional Neural Networks (CNN) [30] are more effective in processing images with important relationships between pixels. Recurrent Neural Networks (RNN) [31] are more effective in processing sequence data with important sequential relationships. Inspired by this idea, a model multi-grained scanning similar to a window slider was proposed by Zhou and Feng [32]. In multi-grained scanning, the data that the sliding window slides each time is taken as a sample. The sliding operation can learn the feature vector about the input from the original input. Then input into two Random Forests respectively, and use the class probability vectors output by the two Random Forests as the features of the new sample.

Multi-grained scanning can get many instances, but there is also the problem of excessive redundancy of instances. The resulting problem is that the feature vector after multi-grained scanning is a dense feature vector, which reduces the efficiency of subsequent Cascade Forest training and inference, and increases memory consumption and running time. Based on the above problems, we introduce the operation of Pang *et al.* [33] to randomly sample the features of multi-grained scanning to improve training efficiency.

The number of pictures in the training set is 32, 000, the picture size is 28×28, and there are four channels. For the specific introduction of the dataset, please refer to section . Sub-sampling multi-grained scanning uses three different granularity windows to scan the input images in sequence. Fig. 3 is a schematic diagram of sub-sampling multi-grained scanning.

As shown in Fig. 3, when the sampling rate is 0.2, sub-sampling can not only reduce the number of conversion instances but also reduce the number of conversion features from 484 to 96 orders of magnitude, greatly reducing the amount of calculation and calculation cost. While sampling, the spatial information and pixel information of the image will also be saved. New features are generated in each window, and features from the three windows are cascaded to identify fewer and more difficult samples.

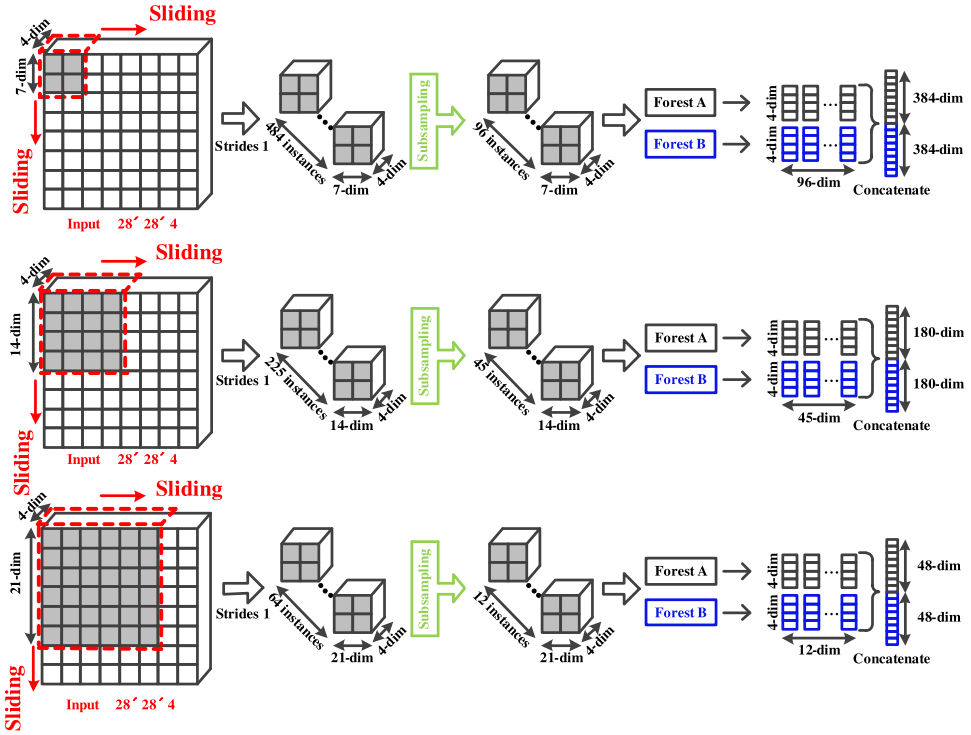


FIGURE 3. Schematic diagram of feature extraction using sub-sampling multi-grained scanning. Each image uses three different sizes with a dimension of 4, a granularity of 7, a step size of 1, and a dimension of 4, a granularity of 14, a step size of 1, and a dimension of 4, a granularity of 21, and a step size of 1. The granularity window performs feature extraction on the image, and the sampling rate is 0.2.

The special operation of sub-sampling multi-grained scanning is that the data that each window slides to is applied as a sample, and the input feature vectors are learned from the input image after image enhancement, which are the instances corresponding to Figure 3. Then do sub-sampling on the scanned instance. Finally, it passes through two Random Forests, a complete Random Forest (Forest A) and a normal Random Forest (Forest B). The class probability vector output by the Random Forest is used as the feature of the new instances.

The following uses mathematical formulas to describe sub-sampling multi-grained scanning.

The size of the picture input to the multi-granularity scan is 28×28, assuming that the size of the multi-grained scanning window is as follows: $k_1 \times k_1 \times 4$, $k_2 \times k_2 \times 4$, $k_3 \times k_3 \times 4$, where k_1, k_2, k_3 represent the particle size and are all less than 28. After a window with a granularity of k_1 passes through a sliding step of S_{stride_1}, the number of instances N obtained is:

$$N = (\frac{28 - k_1}{S_{stride\_1}} + 1)^2.$$

After sub-sampling with a sampling rate of S_{ratio_1}, the number of instances $N_{subsampling}$ becomes:

$$N_{subsampling} = S_{ratio\_1} \times (\frac{28 - k_1}{S_{stride\_1}} + 1)^2.$$

Input $N_{subsampling}$ instances into two Random Forests (an ordinary Random Forest and a complete Random Forest) respectively, and then $N_{subsampling}$ class probability vectors are obtained in turn. Combine the class probability vectors obtained from two Random Forests to obtain a new feature vector $\mathbf{X}_1$ of D_{dim} dimension, where:

$$D_{dim} = 8 \times S_{ratio\_1} \times (\frac{28 - k_1}{S_{stride\_1}} + 1)^2.$$

Combine $\mathbf{X}_1$ with the feature vectors $\mathbf{X}_2$ and $\mathbf{X}_3$ obtained from the other two sliding windows, where the sampling rate corresponding to $\mathbf{X}_2$ is S_{ratio_2}, the step size is S_{stride_2}, and the sampling rate corresponding to $\mathbf{X}_3$ is S_{ratio_3}, The step size is S_{stride_3}. Get the input feature $\mathbf{X}$ of the Cascade Forest, the dimension of $\mathbf{X}$ is:

$$\begin{aligned} D_{dim\_\mathbf{X}} = 8 \times [&S_{ratio\_1} \times \left(\frac{28 - k_1}{S_{stride\_1}} + 1\right)^2 \\ &+ S_{ratio\_2} \times \left(\frac{28 - k_2}{S_{stride\_2}} + 1\right)^2 \\ &+ S_{ratio\_3} \times \left(\frac{28 - k_3}{S_{stride\_3}} + 1\right)^2] \end{aligned}$$

C. CASCADE FOREST

Fig. 4 is a structural diagram of the cascading forest. Each layer of the Cascade Forest is an integration of Random Forests composed of decision trees. In the cascading process,

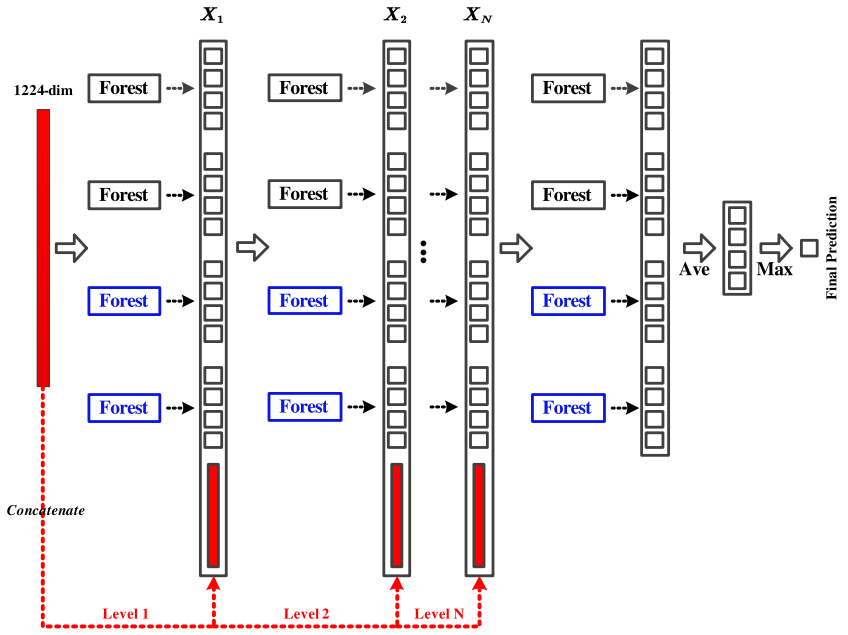


FIGURE 4. The structure of Cascade Forest. Each layer of the cascade contains two ordinary Random Forests (black) and two completely Random Forests (blue).

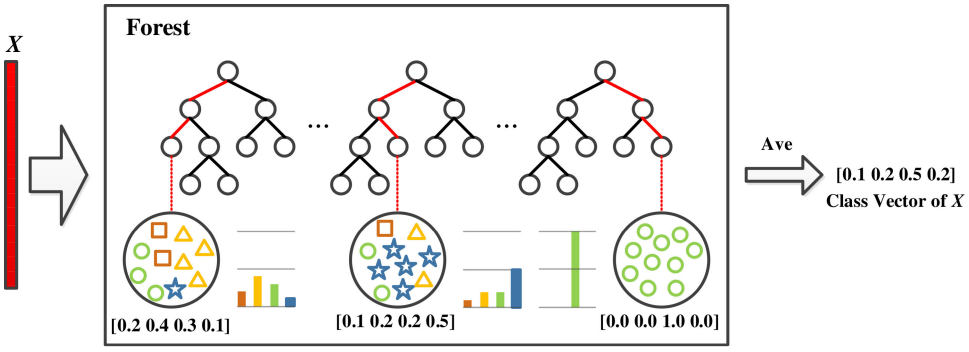


FIGURE 5. Illustration of class vector generation. Different tags in leaf nodes indicate different classes.

different types of Random Forests are used to increase the diversity of features, because the diversity of features is crucial to the construction of the overall structure. For simplicity, two completely Random Forests and two ordinary Random Forests [34] are used. Both the complete Random Forest and the ordinary Random Forest are composed of 1000 decision trees. The difference is that each tree in the ordinary Random Forest randomly selects $\sqrt{d}$ candidate features (d is the number of input features, that is, $D_{dim\_\mathbf{X}}$), and then uses the Gini coefficient [35] to filter the split nodes. Each tree in a completely Random Forest randomly selects a feature as the split node of the split tree and then grows until each leaf node. The main difference between the two types of forests is the candidate feature space. The complete Random Forest selects features randomly to split in the complete feature space, while the ordinary Random Forest selects split nodes through the Gini coefficient in a random feature subspace.

There is a big difference between Cascade Forest and the Deep Neural Network. The feature extraction process of the Deep Neural Network is guided and updated by high-level error backpropagation. In the case of the Deep Neural Networks, the gradients passed down from high-levels are susceptible to fluctuations due to various factors. The entire training process of RES-gcForest is only forward propagation.

Because the training of each Random Forest between the single layers of the Cascade Forest is independent of each other, and the training of the decision trees within a single Random Forest is also independent of each other, RES-gcForest can be calculated in parallel. Given an input feature vector, each Random Forest trains the input feature vector. The class probability vectors of different categories are obtained at the leaf nodes of each decision tree. Then it takes the average of all decision trees in the same random forest to generate an estimate of the class distribution, as shown in Fig. 5. Red highlights the path the instance traverses to the leaf node. The estimated class distribution forms a class probability vector, which is then connected with

the original feature vector to be input to the next cascade layer.

Fig. 5 is a schematic diagram of simplified Random Forest probability vector generation for four recognition problems. $\mathbf{X}$ represents the input vector. The input vector will find a path in each tree to find its subspace. The subspaces found in different decision trees may be different, so it can count different categories to get the proportion of each category, and then average the proportions of all trees to get the probability of each category in the entire forest distributed. Taking the $D_{dim\_\mathbf{X}}$ dimension $\mathbf{X}$ obtained by multi-grained scanning as the input of the Cascade Forest, after passing through the four Random Forests of the first layer of the Cascade Forest, four four-dimensional class probability vectors will be obtained, which are $\mathbf{X}_{l_1}^1, \mathbf{X}_{l_1}^2, \mathbf{X}_{l_1}^3, \mathbf{X}_{l_1}^4$, concatenate the above four class probability vectors and $\mathbf{X}$ to obtain a new feature vector $\mathbf{X}_1$. Input $\mathbf{X}_1$ to the second layer of the Cascade Forest, and you will get four class probability vectors. Follow the steps of the first layer to continue until the N-th layer, and get a new feature vector $\mathbf{X}_N$. Input $\mathbf{X}_N$ into four Random Forests to generate four class probability vectors, and then average the four class probability vectors to obtain a class probability vector, and take the largest class as the final prediction output.

In order to reduce the risk of over-fitting, the class probability vector generated by each Random Forest is generated by k-fold cross-validation [36]. Each feature will be used as training data $k-1$ times to generate $k-1$ class probability vectors, and then all class probability vectors will be averaged to generate the final class probability vector as the enhancement feature of the next layer. After extending to the new cascade layer, the performance of the entire cascade can be evaluated on the validation set. If there is no obvious performance improvement, the training process will be terminated. Therefore, the number of cascade layers is automatically determined. This is the most critical and important thing in the Cascade Forest. Each layer of the Cascade Forest is simple supervised learning, because the way it constructs the Cascade Forest allows the Cascade Forest to automatically determine the number of layers. Contrary to most deep Neural Networks with fixed model complexity, Cascade Forest judges the termination time of training by evaluating each layer in the Cascade Forest, and adaptively determines its model complexity. This makes it applicable to training data of different scales, not limited to large-scale training data.

The RES-gcForest model firstly enhances the satellite imagery of the plateau area. Since there are fewer weather stations in the plateau area and the quality of the collected samples is low, it is necessary to enhance the image of the collected satellite images. After that, a multi-window and multi-grained scanning with random sampling is performed. The main purpose of the scan is to perform feature extraction. The extraction technique is similar to the CNN [37]. The use of multi-window and multi-grained can enrich the diversity of features, so that the spectral features and texture features of the original image can be converted into examples for recognition as much as possible. The instances obtained through multi-grained scanning are randomly sampled, and then the class vectors generated by the two Random Forests are regarded as new sample features. Finally, the above sample features are fused and input into the Cascade Forest. The Cascade Forest is inspired by deep Neural Networks to process the features after multi-granularity scanning layer by layer.

III. EXPERIMENTAL SETTINGS

A. DATASET AND EXPERIMENTAL EQUIPMENT

The image samples used in the article are all from the environment and natural disaster forecasting satellite HJ-1A/1B. The HJ-1A/1B satellite data with four channels are used for analysis. Tab. 1 shows the four HJ-1A/1B The wavelength of the visible channel.

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

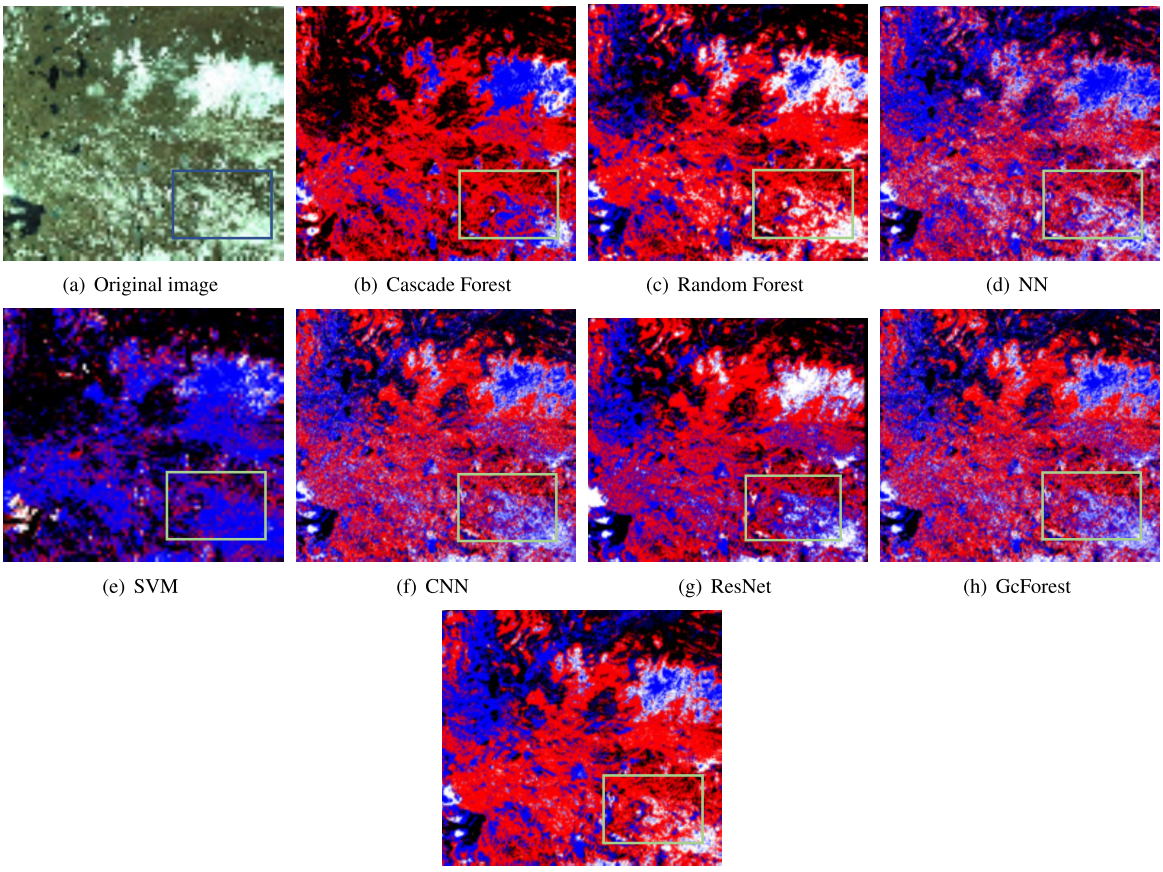
	Wave band			
Channel	1	2	3	4
Wavelength (mm)	0.43-0.52	0.52-0.60	0.63-0.69	0.76-0.90

In this study, we realized patch-wise cloud/snow recognition by collecting and extracting cloud and snow satellite samples from high-resolution satellite images whose space resolution is 30 meters and size is about 150000×150000. We first extract the cloud/snow patches from full-scale satellite images with the size of $128 \times 128 \times 4$, and $64 \times 64 \times 4$, where 4 is the number of spectrums. However, when we performed the prediction on the full-scale satellite images, the prediction results are coarser than that of $28 \times 28 \times 4$ due to the prediction results in Fig. 7 and Fig. 8 are visualized by patch-wise predicting on the full-scale large satellite images. Considering the feasibility of patch-wise cloud/snow recognition, we chose an appropriate size of 28×28 to establish a cloud/snow dataset to realize accurate patch-wise cloud/snow recognition. Fig. 6, Fig. 7, and Fig. 9 are the prediction results on the full-scale satellite images, where we first perform the training to get a robust cloud/snow recognition model and then performing patch-wise predictions on the full-scale satellite images whose scale range from 150000×150000.

$40,000$ images of 28×28 size are taken, and the samples in the dataset included: cloud/snow-free data, cloud-only data, snow-only data, and cloud/snow mixed data. Each category of images contains $10,000$ samples. 70% of the total number of the dataset are taken as a training set and 10% of the total number of the dataset are applied as a validation set and the remaining 20% as a test set. In this experiment, the following test environments were used for all experiments: CPU (i7-9700K), RAM (32GB), and GPU (GTX2070).

B. EXPERIMENTAL DESIGN

RES-gcForest is used to discern the full range of multi-spectral satellite remote sensing images in Tibet. Experiment conducts Cascade Forest, Random Forest (RF),

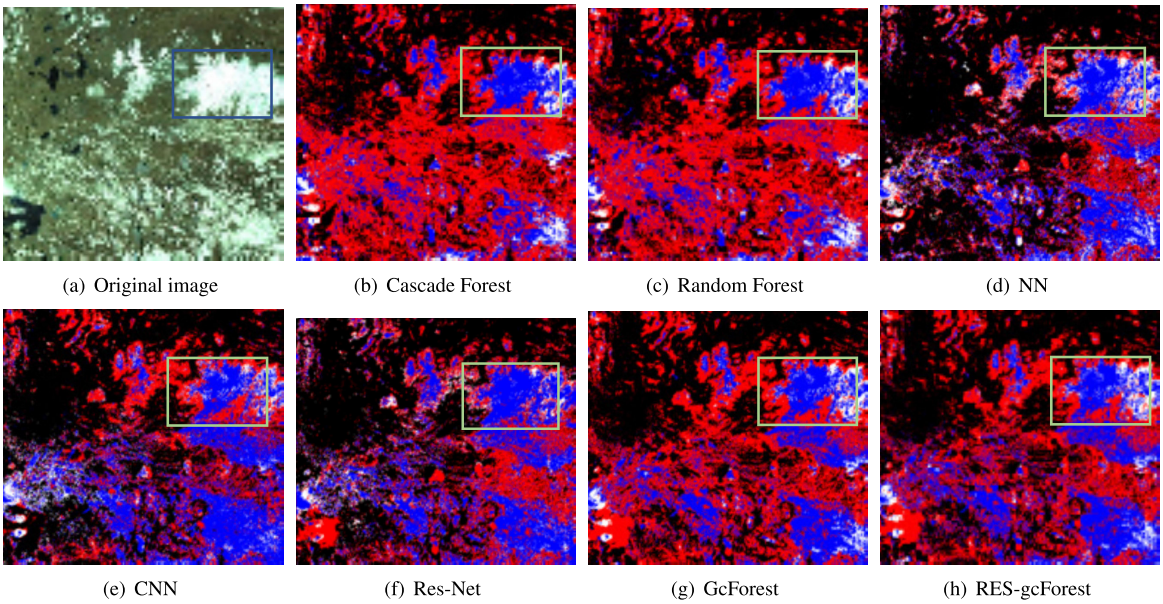


(a) Original image (b) Cascade Forest (c) Random Forest (d) NN

(e) SVM (f) CNN (g) ResNet (h) GcForest

(i) RES-gcForest

FIGURE 6. Single-Spectrum cloud/snow recognition results using different algorithms. In the predicted images, black is a cloud/snow-free area, blue represents a snow-only area, red indicates a cloud-only area, and white denotes a cloud/snow mixed area.



(a) Original image (b) Cascade Forest (c) Random Forest (d) NN

(e) CNN (f) Res-Net (g) GcForest (h) RES-gcForest

FIGURE 7. Multispectral cloud/snow recognition results using different algorithms: black is a cloud/snow-free area, blue represents a snow-only area, red indicates a cloud-only area, and white denotes a cloud/snow mixed area.

Neural Network (NN), and Convolutional Neural Network (CNN), ResNet, and gcForest as comparison algorithms. The experiment involves three groups of comparisons. 1) Comparison of prediction effects of RES-gcForest with

other algorithms on single-spectral images. 2) Comparison of prediction effects of RES-gcForest with other algorithms on multi-spectral images. 3) Compare the parameters in the RES-gcForest model and analyze the influence of the parameter changes on the experimental results.

C. COMPARISON MODEL

In order to evaluate the performance of the RES-gcForest, compared with other existing technologies, including Cascade Forest, Random Forest (RF) [33], Neural Network (NN), Convolutional Neural Network (CNN), gcForest, and ResNet.

All algorithms are trained and tested on GPU and CPU. Through parameter optimization, we found that when the window granularity is (14, 7, 3), the sampling rate is 0.64, the number of decision trees in the scan part of the Random Forest is 60, and the number of Random Forest decision trees in the Cascade Forest is 1000. It can get results with higher accuracy than other algorithms, so the next comparison test uses this as a benchmark.

Both CNN and NN are designed with a six-layer structure, 150 epochs, batch_size is 100, and the final accuracy rate is averaged. The number of two Random Forest decision trees in Cascade Forest is 1000, the number of decision trees in RF is 1000, the decision tree of gcForest scanning part of Random Forest is 60, the cascade part is 1000, and the window size is (14, 7, 3). The sampling rate of RES-gcForest is set to 0.64, the decision tree for the scanning part of the Random Forest is 60, the decision tree for the cascade part is 1000, and the window size is (14, 7, 3).

We chose a sub-sampling rate of 0.2 to accelerate the training and predicting process of RES-gcForest, and compare it with other algorithms. In the original framework of deep forest, multi-grained scanning is designed to enhance the representation learning of cascade forest, transforming the original satellite images into a probabilistic space. As can be seen from the structure of multi-grained scanning, we can get multiple sub-instances by performing multi-scale sliding windows on original satellite images and transform the original satellite image into a class probability vector. More sliding windows are beneficial for the feature extraction but lead to the redundant feature representation, which would slow down the training of cascade forest. Therefore, we perform a sub-sampling to accelerate the training of RES-gcForest. The ratio of 0.2 is concluded from the experimental trials that both considering the training efficiency and the overall performance of cloud/snow recognition. Such that, RES-gcForest can be a good choice where small-scale accurate cloud/snow recognition is desirable.

IV. EXPERIMENTAL RESULTS

A. COMPARISON OF SINGLE SPECTRUM EXPERIMENT

It can be obtained from the experiment that all algorithms have the highest test accuracy for Channel 2, so the experimental results of all algorithms on Channel 2 are compared, and the results are shown in Tab. 2.

TABLE 2. Single-Spectrum cloud/snow recognition performance comparison based on different algorithms.

Algorithm	Test accuracy (%)	Training time (s)	Testing time (s)
Cascade Forest	77.48	2165.96	89.49
RF	74.49	**108.79**	**3.157**
NN	73.39	1763.23	11.01
SVM	63.12	1512.93	501.12
CNN	76.50	2041.23	27.58
ResNet	80.12	3422.34	40.32
GcForest	80.65	5637.49	328.363
RES-gcForest	**81.47**	1252.80	25.93

It can be seen from Tab. 2 that the traditional SVM algorithm has the lowest recognition accuracy for single-spectrum satellite image cloud/snow recognition, and the average test accuracy of the test set is only 63.12%. Followed by the Neural Network algorithm, the accuracy rate is 73.39%. The accuracy of Random Forest is better than NN, and the test accuracy of RES-gcForest is the highest among all algorithms. The results also show that the training speed of RES-gcForest is faster than CNN, gcForest, and Cascade Forest. The accuracy of RES-gcForest is 6.50% higher than the average accuracy of CNN, 1.02% higher than gcForest, and 5.15% higher than Cascade Forest. Meanwhile, RES-gcForest is superior to ResNet in performance. In terms of test time, Random Forest has the fastest test time, but the accuracy rate is very low. GcForest has the slowest test time and the longest time, but the accuracy rate is higher. RES-gcForest can guarantee the highest accuracy without much test time consumption. Fig. 6 shows single-spectrum satellite cloud/snow images of different recognition algorithms. In the predicted image, black is a cloud/snow-free area, blue is a snow-only area, red is a cloud-only area, and white is a cloud/snow mixed area.

It can be seen from Fig. 6 that using different algorithms to train single-spectrum satellite image data, NN has the worst overall performance among the above algorithms. The visual result of NN is that most cloud/snow-free areas are classified as snow-only areas, while the cloud/snow mixed area at the bottom of the satellite image is detected as snowfall areas. Compared with the results of NN, the ability to detect cloud/snow-free areas in the Random Forest recognition image has been improved. The cloud/snow mixed area at the bottom of the satellite image can be correctly detected, but the incorrectly classified areas are not significantly reduced. Fig. 6(b) and Fig. 6(e) Compared with NN and Random Forest algorithms, the use of Cascade Forest and CNN has improved the detection effect of only cloud areas, but CNN has a poor detection effect on cloud/snow mixed areas, and the same problem also exists in gcForest. In Fig. 6(g), due to the feature redundancy in ResNet, the original image information is seriously lost. Therefore, compared with CNN, the recognition effect is not greatly improved. In the lower right corner of the figure, the effect of RES-gcForest on cloud and snow recognition in this area is more significant than that of other algorithms in this area. Among all the algorithms,

the result of RES-gcForest is the best. Compared with the first four algorithms, the misrecognition area is reduced, but compared with the original image, the recognition image is still not ideal, such as the snow-only area in the lower right corner of the satellite image. It is mistakenly divided into a cloud-snow mixed area, and the cloud-free and snow-free areas in the upper left corner of the image are predicted to be a cloud and snow area. Except for the Cascade Forest algorithm, the rest of the algorithms will predict the cloud-free and snow-free area in the upper left corner as a cloud and snow area. The above algorithm does not have high accuracy for cloud/snow recognition of a single spectral satellite image, and the recognition effect on the entire HJ-1A / 1B cloud and snow image has not reached the expected result. all algorithms are inaccurate in the single-channel prediction effect, because all the algorithms obtain limited texture and spectral information in the single-channel image data, and cannot fully extract the useful information of the channel, resulting in that the accuracy rate of the algorithm is low and the false recognition rate is high.

B. COMPARISON OF MULTISPECTRAL EXPERIMENT

Because of the low accuracy of RES-gcForest and the comparison algorithm in single-spectral cloud and snow recognition, we try to train multi-spectral satellite images and predict 8000 28 × 28 multi-spectral satellite images. The parameter settings for the comparison of the multi-spectrum experiment are consistent with the single spectrum, and the experimental results are shown in Tab. 3. Obviously, the accuracy of the test on the multi-spectrum test set is higher than that on the single spectrum.

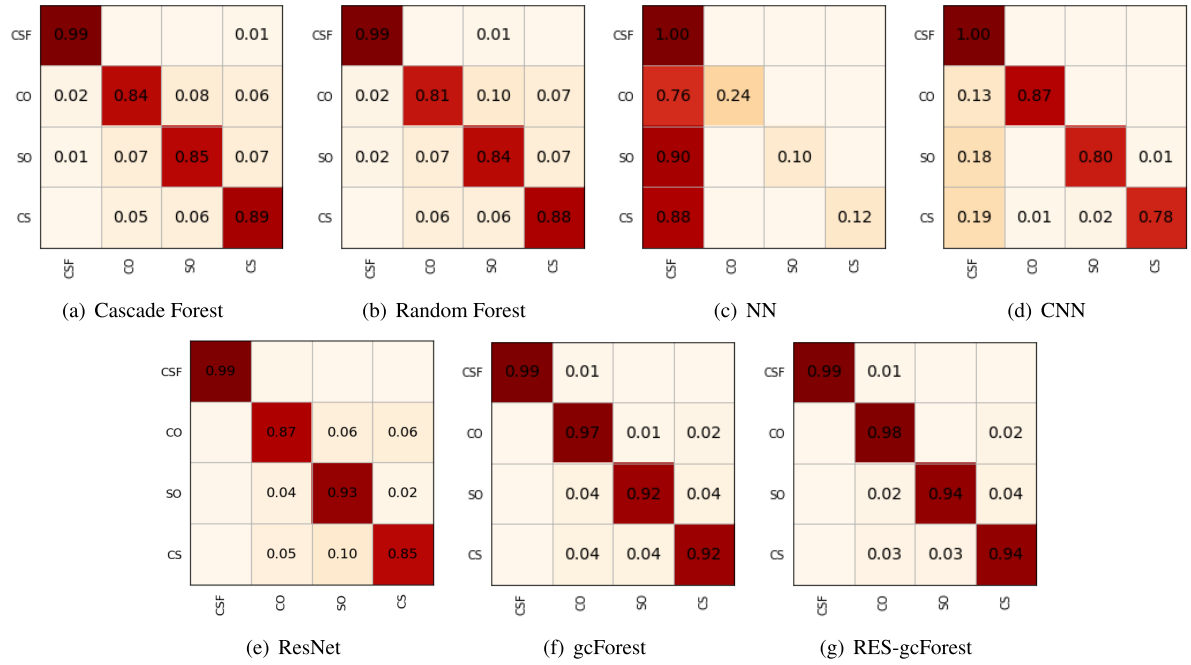
TABLE 3. Comparison of multi-spectral cloud/snow recognition performance based on different algorithms.

Algorithm	Test accuracy (%)	Training time (s)	Testing time (s)
Cascade Forest	88.92	4233.81	78.07
RF	88.09	**672.19**	**7.03**
NN	88.79	2785.92	38.43
CNN	94.02	3454.84	40.31
ResNet	94.54	5573.79	53.52
gcForest	94.87	4878.13	96.31
RES-gcForest	**95.28**	2990.14	60.32

The results in Tab. 3 show that RES-gcForest has the best performance on the multi-spectral cloud/snow recognition, and the accuracy is the highest. In terms of test accuracy, RES-gcForest is 0.43% higher than gcForest, but because of the introduction of image enhancement and sub-sampling, RES-gcForest can use less time and cost to obtain better performance than gcForest, and at the same time memory consumption is higher than gcForest is reduced, the training speed is increased by 38.70%, and the test speed is increased by 37.37%. ResNet not only take a long time to train, but also have lower accuracy than RES-gcForest. Among all algorithms, Random Forest has the worst test effect, with an accuracy rate of only 88.09% on the test set, but due to its own structural advantages, its test speed is the best. The accuracy of Cascade Forest is only 0.94% higher than Random Forest, but the training time is more than twice the training time of Random Forest, and the test time is more than ten times that of Random Forest. The accuracy of NN and CNN is better than Cascade Forest, but lower than gcForest and RES-gcForest. Because the multi-spectral data increases the connection between the original images, the performance of ResNet can be brought into full play. In Fig. 7, the rectangular area in the upper right corner represents the area where the cloud and snow are mixed. According to visual perception, the visual effects of ResNet, gcForest and RES-gcForest are significantly better than other algorithms. The training and testing speed of RES-gcForest is faster than NN and CNN. Figure 7 shows the recognition results of multi-spectral satellite cloud/snow images based on different recognition algorithms.

It can be seen from Fig. 7 that using different algorithms to train multispectral satellite image data, the visual effect of NN in the above algorithm is the worst. In NN, many satellite images are predicted as snow-falling areas only in cloud areas, and other areas are predicted as cloud-free and snow-free areas. Although the recognition accuracy of NN is very high, the visual effect is poor. Although the recognition accuracy of Random Forest and Cascade Forest is not high, they can distinguish rough areas more accurately. Compared with NN, the visual effect is improved. However, these two algorithms are in the snow area in the lower right corner of the satellite image. The prediction effect is poor. CNN has mis-recognized the snowfall area in the lower right corner, predicting a part of the cloud-only area as a snowfall area. ResNet can distinguish the cloud-snow mixed area better, but their recognition in the upper right corner area is not as good as that of RES-gcForest and gcForest. Compared with gcForest, RES-gcForest has more obvious visual effects in the recognition of the snow area at the bottom of the satellite image. But the recognition of gcForest in the snow area in the lower right corner is not as effective as RES-gcForest. Compared with the other seven algorithms, RES-gcForest visually matches the recognition area of the original image.

The recognition results of each model are plotted as a confusion matrix, as shown in Fig. 8. In Fig. 8, the horizontal represents the predicted label and the vertical represents the true label. CSF represents the cloud/snow-free area, CO represents the cloud-only area, SO represents the snow-only area, and CS represents the cloud-snow mixed area. Fig. 8(a) shows that Cascade Forest has a better recognition effect on cloud/snow-free, snow-only, cloud-only, and cloud/snow mixed areas. This is because Cascade Forest can effectively extract the texture and spectral features of clouds and snow so that they can be recognized more accurately. However, it can be seen from the confusion matrix of Cascade Forest that its recognition in the cloud-only, snow-only, and cloud-snow mixed areas is still somewhat inadequate, indicating that the texture and spectral characteristics of clouds and snow have not been fully utilized. For the same reason, it can be seen that the Random Forest in Fig. 8(b) also has shortcomings



(a) Cascade Forest (b) Random Forest (c) NN (d) CNN

(e) ResNet (f) gcForest (g) RES-gcForest

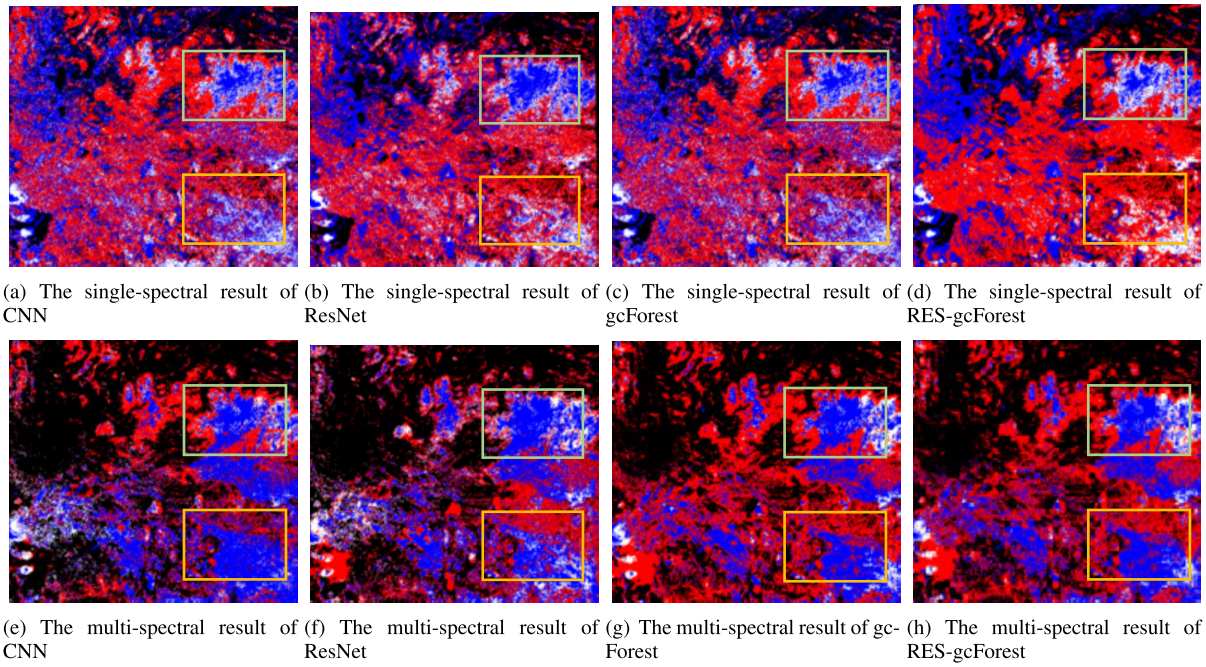
FIGURE 8. Confusion matrix diagram of different cloud/snow recognition algorithms: CSF represents the cloud/snow-free area, CO denotes the cloud-only area, SO indicates the snow-only area, and CS is the cloud-snow mixed area.

in the cloud-only, snow-only, and cloud-snow mixed areas. Fig. 8(c) and Fig. 8(d) show that NN and CNN have poor prediction results for cloud-only, snow-only, and cloud-snow mixed regions. The neural network algorithm has poor generalization ability in cloud and snow classification. This is because the characteristics of the multi-spectral cloud and snow image have great fluctuations for the neural network, and the texture characteristics of cloud and snow cannot be distinguished well. In the process of back propagation of the neural network, it is affected by various factors, resulting in an increase in misclassification. However, it can be seen from the figure that NN and CNN are very accurate in identifying cloud/snow-free areas. In Fig. 8(e), ResNet for recognition of snow-only areas has been further improved compared to CNN. GcForest has high accuracy in cloud/snow recognition. However, it can be seen from the confusion matrix that the RES-gcForest method proposed in this article has the highest detection accuracy. The structure of multi-grained and multi-scale feature extraction in proposed model can not only effectively extract texture features, but also other effective semantic features. In addition, the application of image enhancement improves the robustness of the model to image data. Therefore, the RES-gcForest method not only has excellent recognition accuracy for clouds and snow but also has excellent recognition accuracy for other categories.

Fig. 9 shows the comparison results of single-spectral image recognition and multi-spectral image recognition using CNN, ResNet, gcForest, and RES-gcForest, respectively. It can be seen from the comparison chart of single spectrum and multi-spectrum in Fig. 9 that the recognition effect of the three methods on multi-spectral images is much better than that of single-spectrum recognition images, and there are fewer misrecognized areas. As can be seen in Fig. 9(a) and Fig. 9(e), CNN can not only improve the accuracy of recognition on multispectral but also improve the visual recognition effect. The multi-spectral CNN can identify cloud-only area, cloud-free and snow-free area, and snow-only area, reducing the number of mis-recognized snow-only regions and snow-only regions. In the multispectral experiment, comparing the area framed in the figure, it is obvious that the recognition effect of gcForest and RES-gcForest is better than that of ResNet. However, compared with gcForest and RES-gcForest, the visual effect of CNN is still poor, and there are still more areas of misrecognition. Due to the shallow network framework, CNN cannot extract advanced features from multi-spectral images. Therefore, the generalization ability of multi-spectral data recognition is not ideal, resulting in many misrecognition areas in visual effects. Obviously, gcForest and RES-gcForest can effectively use the spectral feature information between multi-spectral data and have better multi-spectral data generalization capabilities than single-spectral data.

C. COMPARISON OF RES-GCFOREST MODEL PARAMETERS

A small number of decision trees cannot effectively learn feature information, but too many decision trees in Random Forests can easily lead to over-fitting [36], resulting in a decline in the generalization ability of the model. As the number of decision trees increases, the training time and testing time of the model also increase at the same time.



(a) The single-spectral result of CNN (b) The single-spectral result of ResNet (c) The single-spectral result of gcForest (d) The single-spectral result of RES-gcForest

(e) The multi-spectral result of CNN (f) The multi-spectral result of ResNet (g) The multi-spectral result of gcForest (h) The multi-spectral result of RES-gcForest

FIGURE 9. Comparison of CNN, ResNet, gcForest, CNN and RES-gcForest algorithms based on single-spectral image recognition and multi-spectral image recognition: black denotes a cloud/snow-free area, blue represents a snow-only area, red indicates a cloud-only area, and white is a cloud/snow mixed area.

TABLE 4. The influence of the number of different decision trees in the Cascade Forest on the performance of the RES-gcForest model. The other parameters in the model are set as follows. $W = (14, 7, 3), S = 0.81, D = 7, T_S = 60$.

T_c	Training		Testing	
	Accuracy (%)	Time (s)	Accuracy (%)	Time (s)
1000	96.75	3579.73	95.21 68.82	68.82
900	96.74	3019.02	95.13	62.71
800	96.43	2620.34	95.16	57.19
700	96.30	2203.08	95.10	51.67
600	96.11	1516.31	94.91	50.73
500	96.03	1578.41	94.95	59.77
400	95.94	1243.37	94.69	47.93
300	96.32	949.94	94.80	42.28
200	96.12	689.19	94.85	27.18
100	96.27	463.01	94.86	15.78

The size of the sampling rate also affects the results. In order to reflect the influence of the parameters in the RES-gcForest model on its results, such as the number of decision trees in the Random Forest, the size of the sub-sampling sampling rate and other factors on the results, each influencing factor was analyzed quantitatively and qualitatively. The influence of the number of decision trees and the sampling rate on the recognition results is tested by the control variable method. W represents the size of the sliding window, S represents the size of the sub-sampling sampling rate, D represents the sliding step size, T_s represents the number of decision trees in the multi-granularity scan, and T_c represents the number of decision trees in the Cascade Forest.

Tab. 4 shows the influence of the number of different decision trees in the Cascade Forest on the performance of the RES-gcForest model.

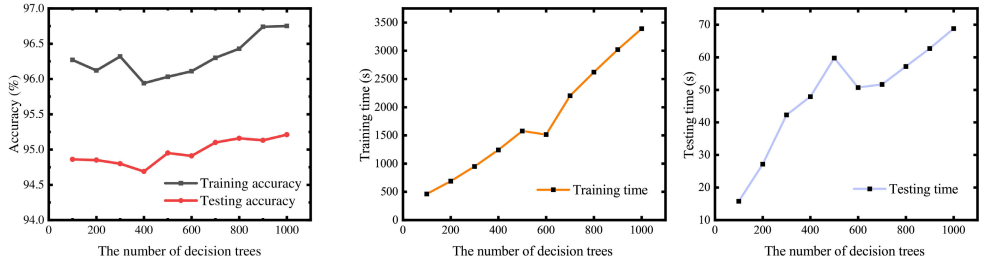
It can be seen from Tab. 4 that as the number of decision trees in the Cascade Forest increases, all evaluation indicators are on an upward trend, and there will occasionally be some fluctuations, but the overall level is on the rise. In order to more intuitively see the influence of the number of decision trees in the cascading forest on the model, the results are drawn as a line graph, as shown in Fig. 10.

From the curve in Fig. 10, as the number of decision trees in the Cascade Forest continues to increase, all indicators are in an upward phase. But for the training accuracy, the balance will be reached at a certain critical point. If the number of decision trees continues to increase, the accuracy will not increase significantly. The number of decision trees in the Cascade Forest is not the better, the more decision trees, the more time it takes to train and test, but the accuracy is not improved, so it is necessary to find a balance between time consumption and accuracy point.

Tab. 5 shows the effect of the number of different decision trees in sub-sampling multi-grained scanning on the performance of the RES-gcForest model.

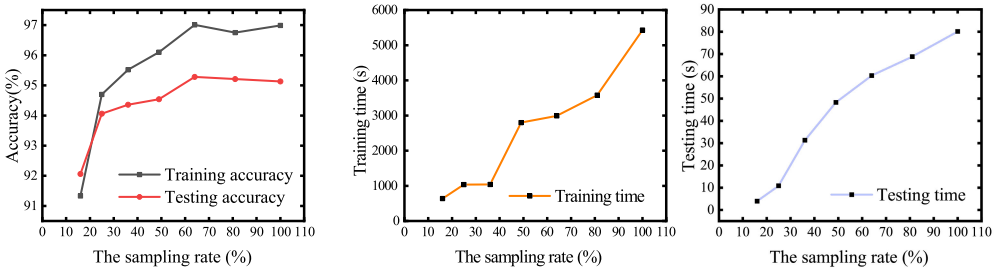
Tab. 5 is a quantitative analysis of the number of decision trees in Random Forests in sub-sampling multi-grained scanning. Tab. 5 shows that in sub-sampling multi-grained scanning, the change in the number of decision trees in the Random Forest has little effect on the final test accuracy, but the training time will increase with the increase in the number of decision trees. The test time also increased slightly.

Tab. 6 shows the effect of different sampling rates in sub-sampling multi-grained scanning on the performance of the RES-gcForest model.



(a) Training and testing accuracy curve (b) Training time curve (c) Testing time curve

FIGURE 10. The accuracy curves and time curves of the number of decision trees in the Cascade Forest on the model.



(a) Training and testing accuracy curve (b) Training time curve (c) Testing time curve

FIGURE 11. The effect of sampling rate on the model in sub-sampling multi-grained scanning.

TABLE 5. The effect of the number of different decision trees in sub-sampling multi-grained scanning on the performance of the RES-gcForest model. The other parameters in the model are set as follows. $W = (14, 7, 3)$, $S = 0.81$, $D = 7$, $T_c = 500$.

T_s	Training		Testing	
	Accuracy (%)	Time (s)	Accuracy (%)	Time (s)
10	95.61	956.45	94.39	42.27
20	95.74	1051.16	94.16	38.74
30	95.79	1135.47	94.36	47.80
40	95.39	1186.62	94.46	53.12
50	95.99	1230.38	94.95	52.84
60	96.03	1278.41	94.95	59.77
70	95.96	1331.23	94.65	57.15
80	96.00	1379.36	94.52	56.27
90	96.02	1439.32	94.56	61.72
100	96.11	1505.89	94.61	68.6

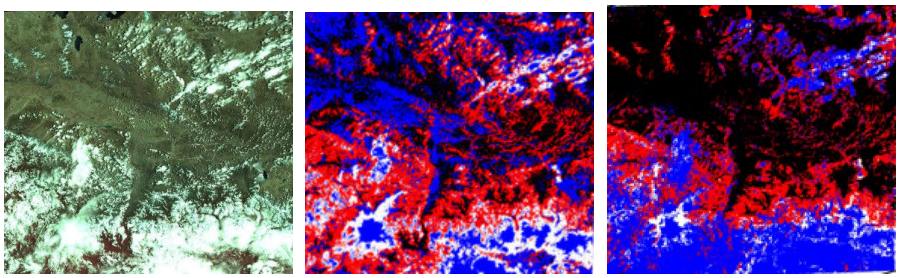
TABLE 6. The effect of sub-sampling ratio on the performance of model. The other parameters in the model are set as follows. $W = (14, 7, 3)$, $D = 7$, $T_s = 60$, $T_c = 1000$.

Network parameter S	Training accuracy (%)	Training time (s)	Testing accuracy (%)	Testing time (s)
0.16	91.33	637.42	92.06	3.90
0.25	94.70	1035.75	94.06	10.87
0.36	95.52	1037.75	94.36	31.28
0.49	96.10	2798.79	94.54	48.28
0.64	97.01	2990.14	95.28	60.32
0.81	96.75	3579.73	95.21	68.82
1.00	96.99	5430.87	95.13	80.10

It is obvious from Tab. 6 that as the sampling rate increases, training time and testing time also increase, because the increase in sampling rate leads to more features, which slows down the training speed and testing speed of the model. Meanwhile, the accuracy of the model is also increasing, but there is a zero point. When this critical point is reached, even if the sampling rate increases again, the accuracy of the model will decrease. Because of the excessive amount of data, the model has the risk of over-fitting. So for the RES-gcForest model, a reasonable sampling rate can get the best results. Fig. 11 visualizes the data in Tab. 6.

In Fig. 11(a), as the sampling rate increases, both the training accuracy and the test accuracy improve, but they will tend to flatten and reach a peak. Since the sampling rate continues to increase, the accuracy rate will tend to decrease. Fig. 11(b) and Fig. 11(c) can show that the training time and testing time both increase with the increase of the sampling rate.

Summarizing the results of the impact of the above parameters on the RES-gcForest model, it can be found that the parameters that have a greater impact on the model are the number of decision trees in the Cascade Forest and the sampling rate in sub-sampling multi-grained scanning. As the number of decision trees in the Cascade Forest increases, the training time and testing time of the model will increase accordingly. When a certain number of decision trees is reached, the performance will reach the best, which is called the equilibrium point. After this balance point, no matter how to increase the number of decision trees and sampling rate of the model, the overall performance of the model is poor. Because as the number of decision trees increases, the training time and testing time will increase significantly, resulting in a lower overall performance of the model. The above



(a) Original image (b) Prediction result of single-spectral image (c) Prediction result of multi-spectral image

FIGURE 12. Single-Spectral and multi-spectral image recognition based on RES-gcForest: black is a cloud/snow-free area, blue indicates a snow-only area, red denotes a cloud-only area, and white represents a cloud/snow mixed area.

results can also be concluded that the number of decision trees in sub-sampling multi-grained scanning has little effect on the model.

In order to verify the stability and generalization effect of RES-gcForest on multispectral data training and recognition, as well as the robustness and generalization of the model to image data, 50 HJ-1A/1B satellite images were carried out using the RES-gcForest model. Recognition. All results prove the effectiveness and superiority of the RES-gcForest method. Fig. 12 shows the comparison between the prediction result of another satellite image of RES-gcForest and the original image.

V. CONCLUSION

In this paper, in view of the low utilization of features of satellite remote sensing images in traditional algorithms, and a large amount of model parameters in deep learning algorithms, and the low degree of distinction between clouds and snow, this paper proposes an improved deep forest cloud based on Random Erasing Snow recognition model, this model has a small number of parameters, high feature utilization, can effectively distinguish between clouds and snow, and the model training time is greatly reduced compared with deep learning algorithms. The RES-gcForest model employs Random Erasing, an image enhancement method, to ensure the robustness of the model to images, and takes advantage of multi-window scanning to extract image features to ensure the integrity of image spatial features and extract more features. Then through this operation of sub-sampling to reduce the number of model parameters and improve the learning ability of the model.

The experiments are compared with traditional machine learning methods and the latest deep learning methods. Experimental results show that RES-gcForest is superior to current machine learning methods and deep learning methods in accuracy, and has certain advantages in speed. The traditional method and the deep neural network algorithm do not achieve the expected effect in the single spectral dataset training, while the RES-GCForest in the single spectral dataset can reach 81.47% of the test accuracy, it can be seen from the comparison figure that the visual recognition effect of RES-GCForest is better than other algorithms. Compared with single spectral images, RES-gcForest has a great improvement in cloud/snow recognition ability in multi-spectral images, and can effectively identify cloud and snow.

Although RES-gcForest has advantages in cloud/snow recognition, it still has disadvantages in some aspects. Therefore, the future research is as follows: 1) RES-gcForest model cannot connect every single spectral channel of cloud and snow image, and future research needs to explore the correlation between different spectra; 2) The representation learning ability of the RES-gcForest model is always limited. When processing a large amount of cloud image data, its advantages are not as obvious as those of neural networks. In future studies, we can try to combine the advantages of RES-gcForest feature extraction and the robustness of the model with the strong representation learning ability of deep learning. A more accurate and fast model will be proposed, which will be applied on large-scale high-resolution datasets.

REFERENCES

[1] B. Tian, M. K. Shaikh, M. R. Azimi-Sadjadi, T. H. Haar, and D. Reinke, "A study of cloud classification with neural networks using spectral and textural features," *IEEE Trans. Neural Netw.*, vol. 10, no. 1, pp. 138–151, Jan. 1999.

[2] C.-H. Li, B.-C. Kuo, C.-T. Lin, and C.-S. Huang, "A spatial–contextual support vector machine for remotely sensed image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 784–799, Mar. 2012.

[3] E. Lasota, W. Rohm, C.-Y. Liu, and P. Hordyniec, "Cloud detection from radio occultation measurements in tropical cyclones," *Atmosphere*, vol. 9, no. 11, p. 418, Oct. 2018.

[4] 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, Aug. 2016.

[5] H. Wang, Y. He, and H. Guan, "Application of support vector machines in cloud detection using EOS/MODIS," *Proc. SPIE*, vol. 7088, Aug. 2008, Art. no. 70880M.

[6] S. Mahajan and B. Fataniya, "Cloud detection methodologies: Variants and development—A review," *Complex Intell. Syst.*, vol. 6, pp. 251–261, Dec. 2020.

[7] M. Xia, Y. Li, Y. Zhang, L. Weng, and J. Liu, "Cloud/snow recognition of satellite cloud images based on multiscale fusion attention network," *J. Appl. Remote Sens.*, vol. 14, no. 3, Feb. 2020, Art. no. 032609.

[8] X. Wu, Z. Shi, and Z. Zou, "A geographic information-driven method and a new large scale dataset for remote sensing cloud/snow detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 174, pp. 87–104, Apr. 2021.

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

[10] R. Pilipović and V. Risojević, "Evaluation of convnets for large-scale scene classification from high-resolution remote sensing images," in *Proc. IEEE 17th Int. Conf. Smart Technol. (EUROCON)*, Jul. 2017, pp. 932–937.
[11] 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.
[12] H. Liu, D. Zeng, and Q. Tian, "Super-pixel cloud detection using hierarchical fusion CNN," in *Proc. IEEE 4th Int. Conf. Multimedia Big Data (BigMM)*, Sep. 2018, pp. 1–6.
[13] M. Le Goff, J.-Y. Tourneret, H. Wendt, M. Ortner, and M. Spigai, "Deep learning for cloud detection," in *Proc. 8th Int. Conf. Pattern Recognit. Syst. (ICPRS)*, Jul. 2017, pp. 1–6.
[14] X. Zhu and E. H. Helmer, "An automatic method for screening clouds and cloud shadows in optical satellite image time series in cloudy regions," *Remote Sens. Environ.*, vol. 214, pp. 135–153, Sep. 2018.
[15] M. Xia, W. Liu, B. Shi, L. Weng, and J. Liu, "Cloud/snow recognition for multispectral satellite imagery based on a multidimensional deep residual network," *Int. J. Remote Sens.*, vol. 40, no. 1, pp. 156–170, Jan. 2019.
[16] Z. Li, H. Shen, Q. Cheng, Y. Liu, S. You, and Z. He, "Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors," *ISPRS J. Photogramm. Remote Sens.*, vol. 150, pp. 197–212, Apr. 2019.
[17] J. Guo, J. Yang, H. Yue, H. Tan, C. Hou, and K. Li, "CDnetV2: CNN-based cloud detection for remote sensing imagery with cloud-snow coexistence," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 700–713, Jan. 2021.
[18] 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.
[19] S. Mohajerani, T. A. Krammer, and P. Saeedi, "Cloud detection algorithm for remote sensing images using fully convolutional neural networks," Oct. 2018, *arXiv:1810.05782*. Accessed: Aug. 13, 2021. [Online]. Available: http://arxiv.org/abs/1810.05782
[20] T. N. Phan, V. Kuch, and L. W. Lehnert, "Land cover classification using Google earth engine and random forest classifier—The role of image composition," *Remote Sens.*, vol. 12, no. 15, p. 2411, Jul. 2020.
[21] C. Wang, S. Platnick, K. Meyer, Z. Zhang, and Y. Zhou, "A machine-learning-based cloud detection and thermodynamic-phase classification algorithm using passive spectral observations," *Atmos. Meas. Techn.*, vol. 13, no. 5, pp. 2257–2277, May 2020.
[22] M. Shao and Y. Zou, "Multi-spectral cloud detection based on a multi-dimensional and multi-grained dense cascade forest," *J. Appl. Remote Sens.*, vol. 15, no. 2, Jun. 2021, Art. no. 028507.
[23] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Frontiers Comput. Sci.*, vol. 14, no. 2, pp. 241–258, Apr. 2020.
[24] E. C. B. de Colstoun, M. H. Story, C. Thompson, K. Commisso, T. G. Smith, and J. R. Irons, "National park vegetation mapping using multitemporal Landsat 7 data and a decision tree classifier," *Remote Sens. Environ.*, vol. 85, no. 3, pp. 316–327, May 2003.
[25] H.-Y. Cheng and C.-L. Lin, "Cloud detection in all-sky images via multi-scale neighborhood features and multiple supervised learning techniques," *Atmos. Meas. Techn.*, vol. 10, no. 1, pp. 199–208, Jan. 2017.
[26] M. Xia, N. Tian, Y. Zhang, Y. Xu, and X. Zhang, "Dilated multi-scale cascade forest for satellite image classification," *Int. J. Remote Sens.*, vol. 41, no. 20, pp. 7779–7800, Oct. 2020.
[27] M. Xia, W. Liu, K. Wang, W. Song, C. Chen, and Y. Li, "Non-intrusive load disaggregation based on composite deep long short-term memory network," *Expert Syst. Appl.*, vol. 160, Dec. 2020, Art. no. 113669.
[28] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, Apr. 2020, Art. no. 7.
[29] G. Katz, "The marabou framework for verification and analysis of deep neural networks," in *Computer Aided Verification*. Cham, Switzerland: Springer, 2019, pp. 443–452.
[30] L. Wang, Y. Chen, L. Tang, R. Fan, and Y. Yao, "Object-based convolutional neural networks for cloud and snow detection in high-resolution multispectral imagers," *Water*, vol. 10, no. 11, p. 1666, Nov. 2018.
[31] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.
[32] Z.-H. Zhou and J. Feng, "Deep forest," Jul. 2017, *arXiv:1702.08835*. Accessed: Aug. 13, 2021. [Online]. Available: http://arxiv.org/abs/1702.08835
[33] M. Pang, K.-M. Ting, P. Zhao, and Z.-H. Zhou, "Improving deep forest by confidence screening," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2018, pp. 1194–1199.
[34] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
[35] S. S. Sundhari, "A knowledge discovery using decision tree by Gini coefficient," in *Proc. Int. Conf. Bus., Eng. Ind. Appl.*, Jun. 2011, pp. 232–235.
[36] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artif. Intell.*, San Francisco, CA, USA, vol. 2, Aug. 1995, pp. 1–7.
[37] M. Xia, X. Zhang, W. Liu, L. Weng, and Y. Xu, "Multi-stage feature constraints learning for age estimation," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2417–2428, 2020.
[38] Y. Gu, B. Wylie, S. Boyte, J. Picotte, D. Howard, K. Smith, and K. Nelson, "An optimal sample data usage strategy to minimize overfitting and underfitting effects in regression tree models based on remotely-sensed data," *Remote Sens.*, vol. 8, no. 11, p. 943, Nov. 2016.

MENG XIA (Graduate Student Member, IEEE) received the bachelor's degree from the School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing, China, in 2019. He is currently is pursuing the Ph.D. degree with the College of Information Science and Technology, Donghua University, Shanghai, China, in 2021. His research interests include ensemble learning, machine learning, and deep learning.

ZHIJIE WANG received the B.E., M.E., and Ph.D. degrees from the College of Information Science and Technology, Donghua University, Shanghai, China, in 1991, 1994, and 1997, respectively, and the postdoctoral degree from The University of Tokyo, Tokyo, Japan, in 2002. He is currently a Professor with the College of Information Science and Technology, Donghua University. His main research interests include computational neuroscience, machine learning, and deep learning.

FANG HAN received the B.S. and M.S. degrees from Beijing Jiaotong University, Beijing, China, in 2003 and 2006, respectively, and the Ph.D. degree from Beihang University, Beijing, in 2009. She visited the University of Aberdeen, Aberdeen, U.K., for one year as a visiting Ph.D. student, in 2008, and New York University, New York, USA, as a Visiting Scholar for one year, in 2016. She is currently a Professor with the College of Information Science and Technology, Donghua University, Shanghai, China. Her main research interests include computational neuroscience and deep learning.

YANTING KANG received the bachelor's degree from the College of Physics and Electronic Engineering, Hainan Normal University, Haikou, China, in 2019. He is currently is pursuing the degree with the College of Information Science and Technology, Donghua University, Shanghai, China, in 2019. His research interests include deep learning, GAN networks, and image restoration.