Very High Resolution Remote Sensing Imagery Classification Using a Fusion of Random Forest and Deep Learning Technique—Subtropical Area for Example

Luofan Dong, Huaqiang Du, Fangjie Mao, Ning Han, Xuejian Li, Guomo Zhou, Di'en Zhu, Junlong Zheng, Meng Zhang, Luqi Xing, and Tengyan Liu

***Abstract*—Recently, convolutional neural networks (CNNs) showed excellent performance in many tasks, such as computer vision and remote sensing semantic segmentation. Especially, the ability to learn high-representation features of CNN draws much attention. And random forest (RF) algorithm, on the other hand, is widely applied for variables selection, classification, and regression. Based on the previous fusion models that fused CNN with the other models, such as conditional random fields (CRFs), support vector machine (SVM), and RF, this article tested a method based on the fusion of an RF classifier and the CNN for a very high resolution remote sensing (VHRRS) based forests mapping. The study area is located in the south of China and the main purpose was to precisely distinguish Lei bamboo forests from the other subtropical forests. The main novelties of this article are as follows. First, a test was conducted to confirm if a fusion of CNN and RF make an improvement in the VHRRS information extraction. Second, based on RF, variables with high importance were selected. Then, a test was again conducted to confirm if the learning from the selected variables will further give better results.**

***Index Terms*—Classification, convolutional neural networks (CNNs), random forest (RF), subtropical forest, very high resolution remote sensing (VHRRS).**

I. Introduction

Remote sensing (RS) techniques, which now cover large scientific fields, have established themselves as popular and effective methods for monitoring the environmental changes and land use cover dynamics [1]–[4]. Very high resolution remote sensing (VHRRS) images, which contain more valuable features, saliently the spatial information, draw much attention these years [3]. And moreover, with the development of aerospace technology and sensor technology, the VHRRS images are getting easy and inexpensive to obtain that, in return, raises a problem that enormous data remain underutilized. Thus, driven by the explosion of the remotely sensed datasets, the establishment of accurate and effective methods for remotely sensed imagery information extraction is a prerequisite for applications and deep-in investigations of RS technology [5].

An elementary pixel in a certain image is the fusion of multiple objects on the ground [6]. As a result, the spectrum of a certain cell is not only determined by the main landcover but also by the proportions and characters of all objects. Even though VHRRS images have a better spatial resolution, within-class spectral variation is still outstanding and sometimes even worse [7], especially for vegetation [8]. Due to the extreme complexity of the vegetation composition and diversity of vegetation growth in the forest, it is universal that the same forest types exhibit different spectral characteristics. Better methods for information extraction and land cover discrimination are significantly needed to take advantage of spatial information of the VHRRS.

The booming of machine learning (ML) algorithms in decades proved to all that learning features from datasets are more efficient and practical than defining the features. These ML methods, for instance, support vector machine (SVM), classification and regression tree (CART), k-nearest neighbor (KNN), and RF are widely employed these years in the remote sensing fields [9]–[12]. Mostly, these nonparametric algorithms work well to construct the relationships between inputs and outputs, for instance, extracting and retrieving of forest quantities information, such as biomass and soil moisture, and most frequently classification [13]–[15]. Due to this character called "learning by itself," various data have been taken into practice, such as microwave, light detection and ranging, satellite remote

Manuscript received May 19, 2019; revised July 29, 2019, September 13, 2019, and October 25, 2019; accepted November 9, 2019. Date of publication December 16, 2019; date of current version February 12, 2020. This work was supported in part by the National Natural Science Foundation under Grants 31670644 and U1809208, in part by the State Key Laboratory of Subtropical Silviculture Foundation under Grant zy20180201, in part by the Joint Research Fund of Department of Forestry of Zhejiang Province and Chinese Academy of Forestry under Grant 2017SY04, and in part by the Zhejiang Provincial Collaborative Innovation Center for Bamboo Resources and High-efficiency Utilization under Grant S2017011. *(Corresponding\penalty -\@M author: Huaqiang Du.)*

The authors are with the State Key Laboratory of Subtropical Silviculture, Lin'an 311300, China, with the Key Laboratory of Carbon Cycling in Forest Ecosystems and Carbon Sequestration of Zhejiang Province, Lin'an 311300, China, and also with the School of Environmental and Resources Science, Zhejiang A&F University, Lin'an 311300, China (e-mail: dongluofan@gmail.com; dhqrs@126.com; mfangjie@gmail.com; hangis2002@163.com; xuejianli201609@163.com; zhougm@zafu.edu.cn; 1096178563@qq.com; 1459165815@qq.com; 781792079@126.com; x522874591@163.com; lty706695603@163.com).

Digital Object Identifier 10.1109/JSTARS.2019.2953234

sensing, and the fusion of multisource and multiresolution data [16]–[18].

However, there are two problems that always occur. For one thing, most of these ML algorithms are based on the given variables and lack the ability to extract information. For another, the performance of ML algorithms is extremely case specific. Additional methods for feature extraction and selection, such as object-based image analysis (OBIA), textural variables, and vegetation indices, are always needed. For example, Qian *et al.* [19] applied the OBIA on the VHRRS to obtain object-level features and reduced the dimension. Then, the four ML algorithms including SVM, CART, normal Bayes (NB), and KNN were compared, and it turned out that the best ML model depends on a specific case. Besides, the view that one-fits-all algorithm does not exist because of the influence of available samples, spatial resolution, and type of sensors could be found from other works and reviews as in [6], [20], and [21].

Among these ML algorithms, RF is prevalently employed for the selection of variables, especially when hundreds of spectral and spatial features are obtained [22], [23]. Other attributes, such as tolerability to high dimensionality, efficiency, and being insensitive to over-fitting, make it one of the most popular predictors [24]. Both pixel-wise and object-based land-cover mapping using RF are well investigated [25]–[28]. Features, such as spectrum and texture, that differ from each other while involving a high order of self-correlation need to be reduced for efficiency. And additionally, it is helpful to reveal the relations between these predictors and targets, such as spectral profiles and forests' quality [26], [28]–[30]. However, there is no doubt that the determination of the initial variables needs expertise experiments.

Automagical extraction of high-representation variables, known as representation learning [31], could be achieved by deep learning (DL). According to recent neuroscience, deep feature representation could be learned hierarchically from simple concepts [18]. For examples, recurrent neural network, generative adversarial networks, deep reinforcement learning, and convolutional neural networks (CNNs), well-known branches of DL, have made excellent achievements in the tasks that were labeled "formidable," such as complex pattern recognition, professional chess competition, natural language processing, and unmanned vehicle [32]–[38].

The CNNs, especially, draw much attention. Applying CNNs for remote sensing tasks, including object detection, classification, and scene recognition, has been brisk for years [39]. The aforementioned challenges, including a lack of enough annotation samples, intricacy of CNNs, complexity, and quantity of the RS data [40], are to some extent extenuated by using CNNs. Amid works concerning CNNs, the most frequently used data are hyperspectral and VHRRS images [8], [41]–[43], which are more complicated in terms of dimension and spectral variation. For CNNs, with powerful learning ability and data augmentation technology (i.e., rotation, random noise, and crop) to avoid overfitting [32], [44], many works have shown that applying CNNs on remote sensing is robust. For example, Zhan *et al.* [45] applied CNNs to handle weakly distinguished samples, such as cloud and snow. Using the high-resolution free data including Landsat 8 and Sentinel-2, Kussul *et al.* [46] employed one -dimensional (1-D) and 2-D CNNs on a large area (28 000 km^2) for discrimination of 11 complex land covers. On the other hand, open recognition competitions, such as large-scale visual recognition competition spurred the bloom of CNNs. Structures, such as *V*gg, ResNet, fully convolutional network (FCN), are borrowed for remote sensing applications [47]–[50]. Furthermore, in order to resolve specific remote sensing related difficulties, some methods are proposed to exclusively dedicate to RS [51]. For instance, Zhang *et al.* [51] obtained subimage by applying OBIA and vector analysis. Compared with the popular region proposal in computer vision fields, OBIA could handle various shapes and sizes of the real land covers. The multibranch parallel network models based on GoogLeNet [34] and skip-layer architectures based on ResNet [35] are investigated [52], [53].

Additionally, the fusion of the multiimage or multimodel based on CNNs was well exploited these years. Compared with the complex network architectures, fusion models and data are simple and informative. For example, Scarpa *et al.* [54] employed a three-layer CNN to fuse Sentinel-1 (synthetic aperture radar sensor) image and Sentinel-2 (multiresolution optical sensor) image. Radar, which is weather insensitive, serves as a compensation for a normalized vegetation index (NDVI) obtained from optical images. Zhang *et al.* [18], based on the traditional multiple layer perception and CNNs, proposed a hybrid model MLP-CNN. Audebert *et al.* [55] fused two state-of-the-art CNN models, the ResNet and SegNet, and took the digital surface model into consideration. Fu *et al.* [56] proposed an improved FCN combined with the conditional random fields (CRFs) as postprocessing, furthermore, making an improvement on conventional models [57]–[59].

A combination of CNN and RF, in which the former is regarded as an extractor and the latter is regarded as a classifier, comes out with spontaneity. Several works have been proposed recently [60]–[63]. For instance, the DCNR model proposed in [60] takes cubic samples, containing spectral-spatial information, as inputs of CNN and RF as a classifier. But for VHRRS data, cubic samples, which gather several neighboring pixels, are far from the informative enough [60]. Thus, this article concentrates on information extraction of VHRRS. And further, except for learning high-level representation from the imagery itself, there is the feasibility that learning from both spectral and other low-level affiliate information derived from imagery will achieve better performance, costly but worthwhile.

In this article, we take subtropical forests as the study area for two reasons. First and most important, subtropical forests are more complex compared to the uniform northern forests in terms of forests composition, crown structure, and diversity of tree types. Even an image with a spatial resolution of 1.2 m cannot separate a single tree from the subtropical forests. Thus, despite the heterogeneity of spectrum, high representation features explored from the spectrum and spatial context are valuable for the subtropical forests and other similar tasks. Second, monitoring of subtropical forests is of high value because subtropical forests cover a large area and play a crucial role in the terrestrial ecosystem functions [6], [64], [65].

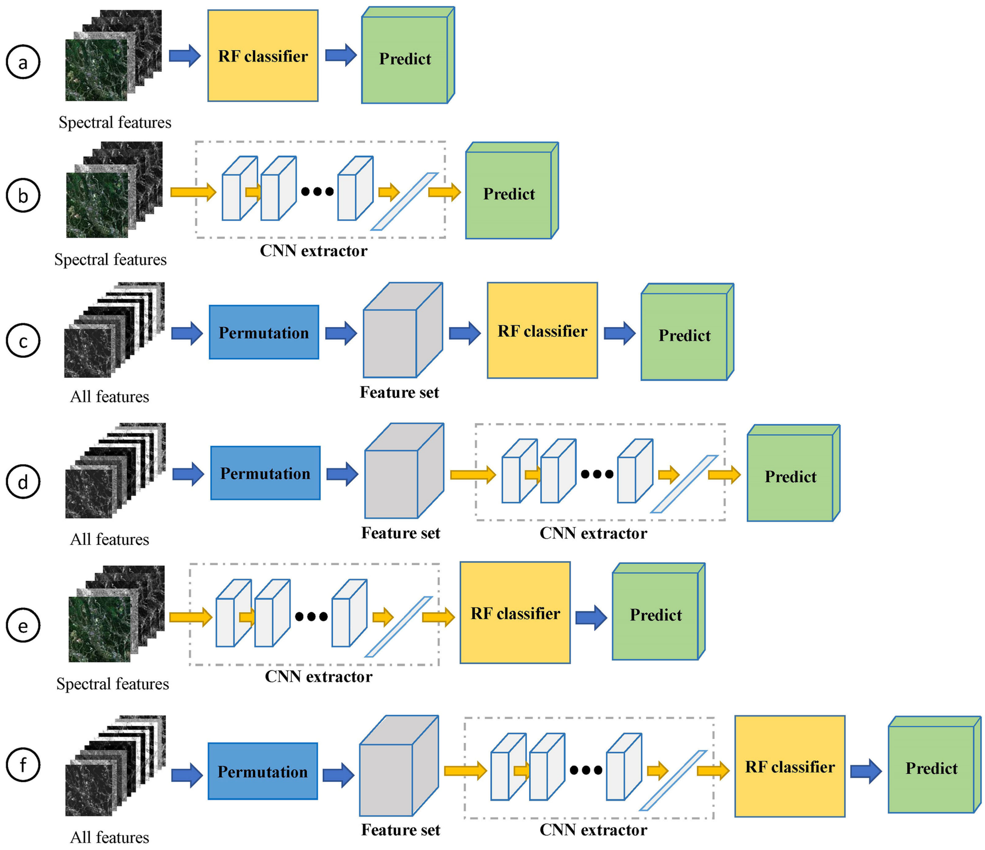


Fig. 1. Overview of our approaches. (a) Conventional pixel-based and spectrum-based RF classification. (b) Conventional patch-based CNNs classification. (c) RF classification based on features selected by permutation. (d) Patch-based CNNs classification based on features selected by permutation. (e) Fusion of RF and CNN based on spectral information. (f) Fusion of RF and CNN based on features selected by permutation.

The innovations of this article mainly include the following.

1) We tested if the fusion of CNN and RF make an improvement of VHRRS information extraction.
2) Based on RF, variables with high importance were selected. We tested if learning from the selected variables will further better the results.

In this article, the experiments are takenoff based on worldview2 VHRRS images. First, we tested common CNNs, with structure and parameters optimization, and RF, with parameters' optimization as well on the corrected spectral [see Fig. 1(a) and (b)]. Then, the traditional meaningful variables, such as vegetation indices and textural features based on gray-level cooccurrence matrix (GLCM), were obtained and, together with spectral features were signed "All features." Permutation and several RF-based functions were employed to score all the features and reduce the dimension. The performance of both the models on "feature set" was tested [see Fig. 1(c) and (d)]. Finally, in the fusion model convolution-RF (ConvRF), RF was employed on the high-level features extracted by CNNs [see Fig. 1(e) and (f)].

II. Materials and Methods

A. Study Area and Data

The research area is located in Taihuyuan (see Fig. 2), south of China, which has a monsoon-type climate, warm and humid, with sufficient sunshine, abundant rainfall, and four distinct seasons. According to the field investigation, the forest of Taihuyuan is dominated by coniferous forest, broad-leaved forest, and bamboos. Therefore, the classification system is set to include ten land cover types: coniferous, building, farmland, bamboos, broadleaf, road, tea garden, water, bare land, and cutting site.

For this study, the classification was based on worldview-2 remotely sensed imagery with four spectral bands (red, blue, green, and infrared) and a spatial resolution of 1.2 m. The used imagery was obtained in June 2016. Radiation calibration and atmosphere correction were applied. Since the present computer hardware condition still poses a limitation on data concerning possible processing volume, it is necessary to divide images into proper sizes. In this article, we divided the image into patches of 224 × 224 according to the setting of some famous works of the ImageNet competition [32]–[35], [44].

B. Features Setting

Variables including vegetation indices and textural features are taken into consideration, as listed in Table I. Vegetation indices include the NDVI, normalized difference water index (NDWI), difference vegetation index (DVI), and ratio vegetation index (RVI).

The GLCM is employed as the representative of the statistical texture features [66], [67]. Since being proposed in 1979,

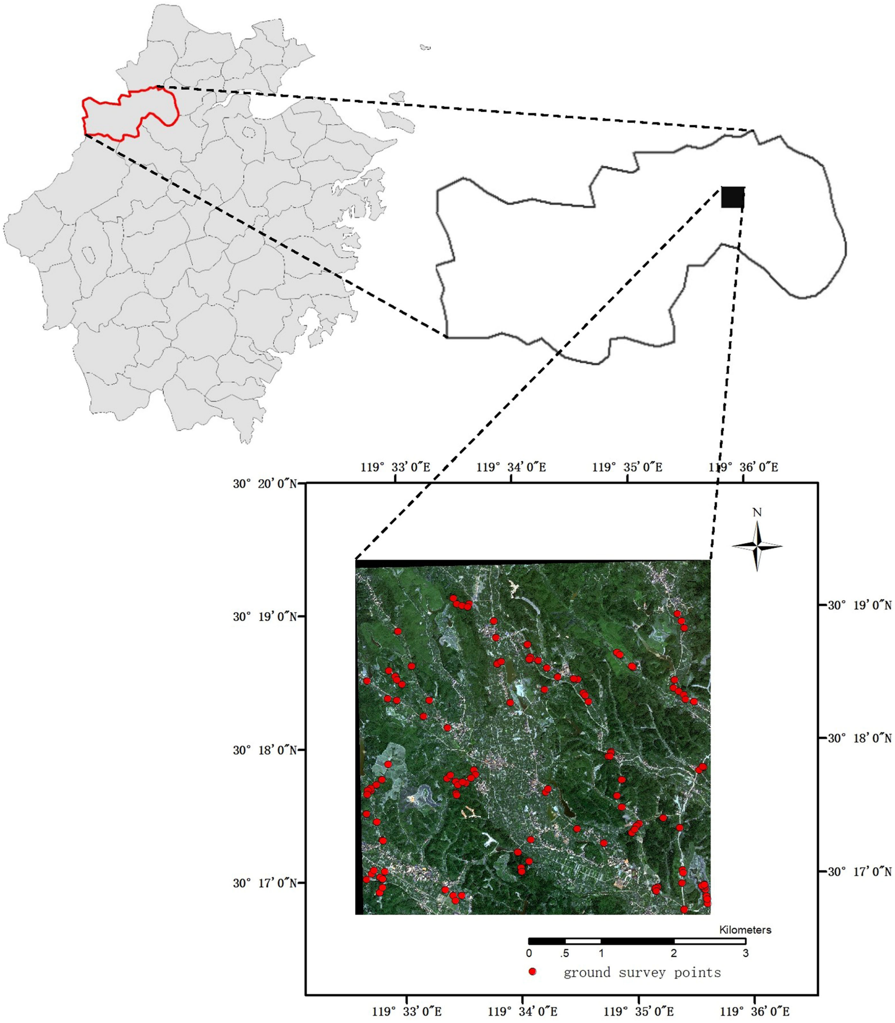


Fig. 2. Location of the study area and the field survey points.

GLCM, based on probability statistics, has been widely employed in classification tasks as well as the other applications. The GLCM is defined as a region with *N* grayscale values that correspond to an $N \times N$ matrix. The value of each pixel (i, j) in the matrix represents the probability that the gray value is *j* as the distance from grayscale value *i*.

C. Random Forest

The RF is a branch of bagging algorithm [68], and it exhibits superior performance in cases with noise and weak discrimination data and is insensitive to the initialization of parameters. An RF selects several samples randomly (with bootstrap) to build a decision tree (without pruning) every iteration and constructs the whole RF by building numerous decision trees. Then, all trees "vote" for the most popular class—the output.

The RF-based feature selection (permutation) has been widely used in many domains and is robust for variables involved with high-dimension and high-order correlation [30]. For example, for a variable x_i, the permutation approach replaces all x_i with a random value and classifies the permutation as noise, thereby breaking the original association between x_i and the result *Y*. Meanwhile, by using the Gini coefficient, the RF determines which variables to be used to minimize the purity decrease of a predictor. In order to reduce the dimension of the data, the random forest cross-validation (rfcv) for feature selection, a function from the add-in package RF, is widely applied for the determination of the variable number. This function sequentially reduces the number of predictors (ranked by variable importance) via a nested cross-validation procedure.

The RF model is generally less sensitive to parameter settings than some other predictors. The optimization of an RF is mostly based on two parameters: *mtry* and *ntrees*, where *mtry* is the number of variables used in splitting a node and *ntrees* is the number of trees in an RF. The optimal value of *mtry* is determined by traversing all possible values. According to the research of Breiman [22], the generalization error of the RF converges as the number of trees increases, a characteristic absent in most other classifiers. In other words, the model performs better and better as the value of *ntrees* increases [29]. Therefore, the optimization of *ntrees* is to balance the classification accuracy with computational effectiveness.

TABLE I
ALL THE INVOLVED FEATURES ARE LISTED, INCLUDING SPECTRAL BANDS OF IMAGERY, VEGETATION INDICES, AND TEXTURE FEATURES BASED ON GLCM

	Feature	Formula
Spectral	Red band (R)	/
	Green band (G)	/
	Blue band (B)	/
	Near Infrared band (NIR)	/
Vegetation Indices	Normalized Difference Vegetation Index (NDVI)	$NDVI = (NIR - R)/(NIR + R)$
	Normalized Difference Water Index (NDWI)	$NDWI = (G - NIR)/(G + NIR)$
	Difference Vegetation Index (DVI)	$DVI = NIR - R$
	Ratio Vegetation Index (RVI)	$RVI = NIR/R$
Texture based on GLCM	Mean (MEA)	$MEA = u_i = \sum_i^N \sum_j^N iP_{i,j} \quad u_j = \sum_i^N \sum_j^N jP_{i,j}$
	Variance (VAR)	$VAR = \sigma_i = \sum_i^N \sum_j^N (i - u_i)P_{i,j} \quad \sigma_j = \sum_i^N \sum_j^N (j - u_j)P_{i,j}$
	Homogeneity (HOM)	$HOM = \sum_i^N \sum_j^N P_{i,j}{}^2$
	Contrast (CON)	$CON = \sum_i^N \sum_j^N P_{i,j}(i - j)$
	Dissimilarity (DIS)	$DIS = \sum_i^N \sum_j^N P_{i,j}\lvert i - j\rvert$
	Entropy (ENT)	$ENT = \sum_i^N \sum_j^N 1\text{n}\, P_{i,j}$
	Augular Second Moment (ASM)	$ASM = \sum_i^N \sum_j^N \frac{P_{i,j}}{1 + (i - j)^2}$
	Correlation (COR)	$COR = \sum_i^N \sum_j^N \frac{(i - u_i)(j - u_j)P_{i,j}}{\sigma_i\sigma_j}$

D. Convolutional Neural Network

A CNN is a multilayer neural network, stacked by convolutional layers and pooling layers alternatively, which learns from an enormous amount of data with convolutional filters and then employs fully connected layers. The processes of training CNN mainly consist of forward-propagation and backpropagation.

1) Forward Propagation: With multiple nonlinear layers stacked back and forth, the forward propagation procedure constructs multiple highly abstract representations and computes the output. Many strategies have been developed for forward propagation, including the convolutional layer, pooling layer, fully connected layer (FC), skip layer, etc., [31], [69]. The brief explanations of these approaches are as follows.

1) Convolutional layer: As the main composite of CNN, the convolutional layer is designed to learn high-level representation features. A significant contribution of the convolutional layer is the sharing of parameters (i.e., the convolution filter) inside a layer, which reduces the number of trainable parameters and accelerates the training processes. Activate functions, which greatly improves the expression ability of the CNNs, are applied after CNNs for further nonlinearization, including Sigmoid, Tanh, ReLU, etc.
2) Pooling layer: Pooling layer, the same as subsampling, is applied for extracting the most significant feature. Another advantage of applying pooling is translation invariance. There are several pooling methods widely used including max-pooling and average-pooling.
3) FC layer: FC is a multilayer neural network applied as the last few layers of the CNN for computing the probability that a certain pixel belongs to a class, i.e., a learnable classifier.
4) Skip layer: The outputs of every layer are connected to the final layer and directly involves classification [35], [69]. This structure works similar to a voting system in which all the outputs vote for the most popular result. Since the gradient disappearance problem plagues many models, skip layer connection is useful and enables deeper network. A lot of investigations show that the fusion of former and latter feature maps of CNN could improve the results [24].

2) Backward Propagation and Optimization: The backward propagation procedure computes the gradient of the loss with respect to parameters, by which all the trainable parameters are updated to minimize the loss [31]. Loss, which is obtained after forward-propagation and computed by a loss function, leads the direction of the optimizer. It is crucial to construct an appropriate loss function for optimization. In our study, cross-entropy loss function, the commonly used function in classification research, is applied.

E. Fusion of CNN and RF

The motivation is to apply CNNs to extract high-level representation features. In the traditional CNN, an FC layer is applied as the final decision basis to compute the possibility that a pixel belongs to each class. However, there are some problems with applying the FC layer as a classifier, such as easily overfitting especially with inadequate samples, not robust enough, and computationally intensive. Many works try to replace the FC layer in their model with other structures, such as CNN as decision basis in the FCN [56], [59], [70]. An RF is apparently more complex and persistent to overfitting. Thus, we replace the FC layer in CNNs with an RF classifier to achieve better outcomes.

In the fusion procedure, a remote sensing image and labels are needed. First, CNN is trained in the first place by end-to-end using the image and labels. In this process, CNNs are trained until relatively low loss and high accuracy are achieved. Then, feature maps are obtained by processing images with the trained CNNs. The annotation maps are used again to obtain samples from feature maps for the RF. An RF classifier is trained through the optimization process we mentioned above with feature maps and labels. The land-cover map is completed and output by RF.

In this article, two CNNs are trained with spectral features and features selected by the RF permutation algorithm from 200 features, as explained in Table I, respectively. As it is well known that active region (every pixel in a feature map contains information from neighbors) gets larger when CNN gets deeper, feature maps are obtained layer by layer in order to test the performance of every CNN extractor. The results are named according to the CNN that is used and the layers that the feature map obtained, for example, Spectral-Conv1, Spectral-Conv2, …, Spectral-Conv8, Selected-Conv1, Selected-Conv2, …, Selected-Conv8. Approximately 4000 pixels are selected randomly from the labels and are separated into training data and test data with a ratio of 0.75 to 0.25 for the construction of the RF classifier. The classification results of all the RFs are obtained to make the further comparison.

III. Result

In this section, the results of all models are reported. The samples of our study are selected based on a field survey (see Fig. 2) and the visual interpretation, and are divided into training samples and validation samples randomly with a ratio of 0.75 to 0.25. There are many methods for the validation of classification performance. In this article, user accuracy rate, producer accuracy rate, overall accuracy rate, and Kappa coefficient (Kappa) based on confusion matrix are employed for the validation of the results [71].

A. RF and CNN With Spectral Data

In this section, the conventional RF and CNN are tested on spectral data. Here the RF model with spectral data is denoted as Spectral-RF and CNN with spectral is denoted as Spectral-DL.

1) Construction of the RF: This part details how the Spectral-RFs are constructed on spectral data. About 400 samples for each class were randomly selected from labeled samples and were randomly divided for training and testing with a ratio of 0.75 to 0.25 [22]. The optimal values of *mtry* and *ntrees* are determined by traversing all possible values and picking up the ones with the lowest out-of-bag (OOB) errors. First, *ntrees* is set to a relatively big value. Due to the convergence of RF when the number of the *ntrees* grows, the traverse of *mtry* will not be impacted by *ntrees*. With the optimal *mtry*, the suitable value of *ntrees* is obtained. The performance of the Spectral-RF is obtained based on the test samples. Finally, the constructed Spectral-RF is applied to the whole image to get the classification map.

2) Construction of the CNN: Parameters setting of the Spectral-DL are listed in the following. The convolutional filters are initialized with a window size of 3 × 3 [32]–[34], [72]. The learning rate is set to 0.0003; the weight decay parameter is set to 0.0005.

Training samples are of high importance in training CNNs. According to our field survey, only relatively pure pixels are labeled as samples. Then, the labeled map is divided into small patches as well as the raw image. Some subimages are eliminated if in which labeled pixels are less than 5%. Then, we get about 200 subimages each with a size of n@224 × 224, where *n* indicates the number of layers. A total of 25% of the subimages are chosen for testing, and the rest for the training. Here, we get about 150 images as training samples and about 50 images for testing. Since large amounts of samples are always needed in the training of a CNN model, rotations, and mirror flips were employed to times the training data, while testing samples remained unchanged.

3) Results of RF and CNN: The results of RF and CNN are reported in forms of tables (see Table II), confusion matrices (see Fig. 3), and classification maps (see Fig. 4).

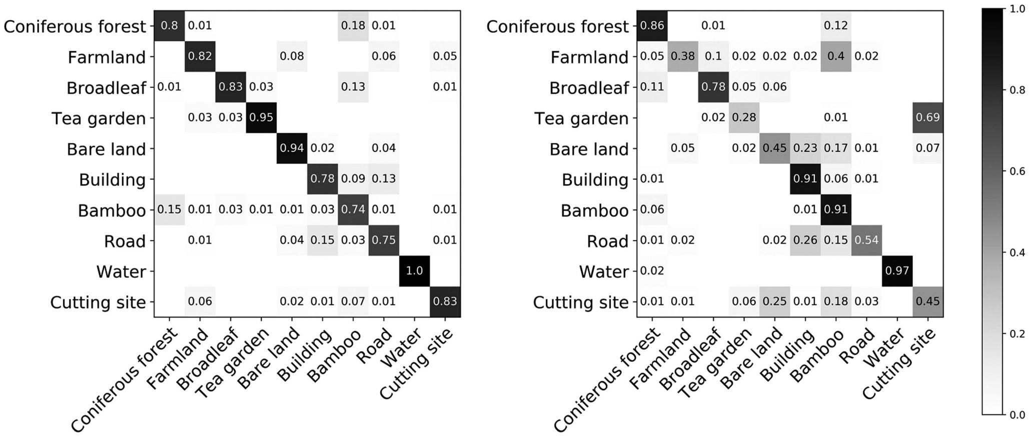
As elaborated in Table II, OA and Kappa coefficient of Spectral-RF are 0.830 and 0.809, respectively, while OA and Kappa coefficient of the Spectral-DL are 0.784 and 0.689, respectively. In both models, body of water is well recognized, which indicates water is distinctive by using spectral data exclusively. Although both trained with spectral data, Spectral-RF outperforms Spectral-DL, especially in terms of broadleaf forests, tea gardens, and cutting sites.

As represented in confusion matrices (see Fig. 3), which clearly show the misclassification, Spectral-RF keeps relatively uniform accuracy rates among all the classes. The confusion between several vegetation exists in Spectral-RF but not as

TABLE II
CLASSIFICATION RESULTS OF CONVENTIONAL RF AND DL

Feature	N	OA	Kappa		1	2	3	4	5	6	7	8	9	10
Spectral-RF	4	0.830	0.809	UA	0.804	0.818	0.826	0.946	0.938	0.780	0.736	0.750	1.000	0.831
				PA	0.837	0.867	0.922	0.933	0.854	0.762	0.675	0.711	1.000	0.902
Spectral-DL	4	0.784	0.689	UA	0.862	0.375	0.784	0.277	0.453	0.910	0.912	0.540	0.971	0.454
				PA	0.746	0.843	0.123	0.391	0.555	0.658	0.851	0.788	0.979	0.627

In this experiment, we take spectrum as inputs (R, G, B, NIR). *N* indicates the number of variables. OA indicates the overall accuracy rate. Kappa indicates the kappa coefficient. The number from 1 to 10 represents land types; 1: Coniferous forest; 2: Farmland; 3: Broadleaf; 4: Tea garden; 5: Bare land; 6: Building; 7: Bamboo; 8: Road; 9: Water; 10: Cutting site.



Spectral-RF

(a)

Spectral-DL

(b)

Fig. 3. (a) Confusion matrix of Spectral-RF. (b) Confusion matrix of Spectral-DL.

salient as does Spectral-DL, that only manages to well extract between several main vegetation types, such as coniferous forest and bamboos. A total of 40% of the farmland pixels are misclassified by the Spectral-DL as bamboo forests. And for the tea gardens, even worse, 69% of which are misclassified as cutting sites.

As shown in Fig. 4, the "salt-and-pepper effect" is obvious in Spectral-RF, whereas "edge effect" is outstanding in Spectral-DL. In scene 1 of Fig. 4, the Spectral-DL fails at the boundary while keeps integrality inside a patch. In scene 2 of Fig. 4, cutting sites are not well classified mainly because some economic nurseries are clear-cut and others are not. In scene 3 of Fig. 4, the "salt-and-pepper effect" of Spectral-RF could be well caught, while Spectral-DL makes better.

B. CNN and RF With Selected Variables

In this section, we first conduct variable selection. Then, the RF and CNN are trained with the selected features. Here the RF model with all feature sets was denoted as AllFeature-RF, the RF model with the selected feature set is denoted as Selected-RF, and CNN with selected feature set is denoted as Selected-DL.

1) Variable Selection: Eight kinds of GLCM textures are selected, including mean, variance, homogeneity, contrast, dissimilarity, entropy, angular second moment, and correlation, all of which are computed in four directions (i.e., 0°, 45°, 90°, and 135°) and four spectral bands (e.g., red, green, blue, and NIR) [66], [67], [73]. To reduce the number of variables, the mean values of the four directions are calculated. The window sizes are set from 3 to 13 with a step of 2–3, 5, 7, 9, 11, 13. The stacked GLCM variables and vegetation indices (i.e., NDVI, NDWI, DVI, and RVI) have 200 layers, which are too "heavy" for most methods—although RF could, theoretically, handle such kind of a complex data, variable selection is practically needed for economic purpose.

The variable selection is based on the RF. First, 4000 samples with 200 variables are randomly selected. Second, as we explained in the previous section, the best values of the *mtry* and *ntrees* are obtained. With the determined values, rfcv is applied with tenfold cross validation and the results are shown in Fig. 5 to determine the performance of the RF as the number of variables getting less. As shown in Fig. 5, 20 variables are stable enough to approach a relatively low OOB error.

Permutation is employed to rank all the 200 variables in terms of variable importance. The decrease accuracy of every land cover type and mean decrease accuracy (MDA) over all classes are shown in Fig. 6, based on which 20 variables are selected for further investigations. For example, in most classes, features related to the NIR band and G band get high importance scores. The decrease accuracy of every land cover type is listed to avoid some extreme exceptions. For example, if a given variable reached a substantial importance score for a certain land cover

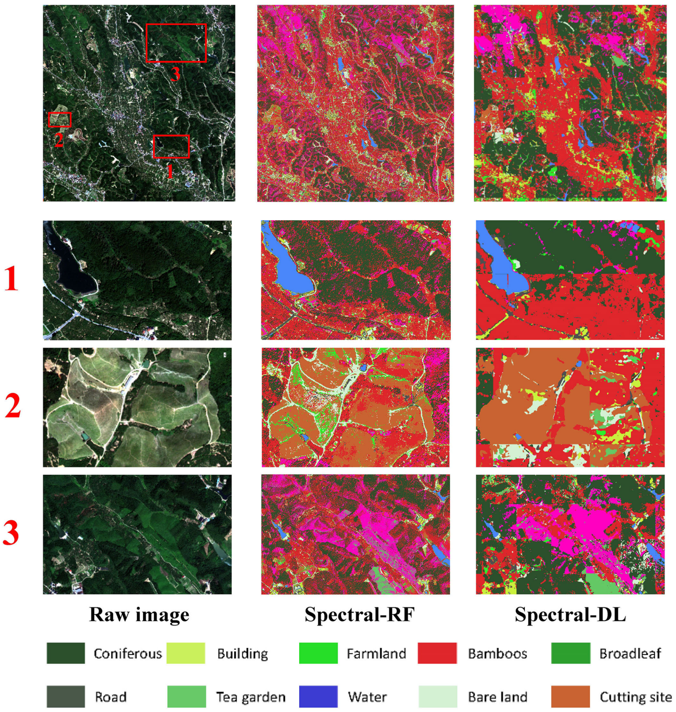


Fig. 4. Left column shows the ground truth. Mid column shows the results of Spectral-RF. The right column shows the results of Spectral-DL.

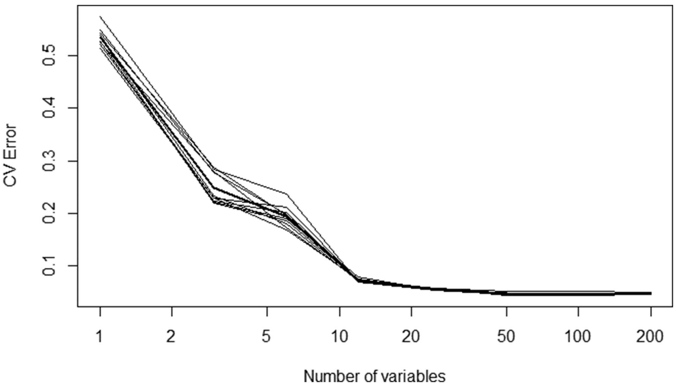


Fig. 5. Effect of the number of variables on OOB error obtained by rfcv for feature selection with all features.

type, we would have to reconsider the use of MDA. In our study, after the analysis (see Fig. 6), we picked 20 variables based on the MDA (see Table II).

2) Results of RF and CNN: With the selected variables listed in Table III, we reconstruct RF and CNN. In order to make a comparison, the results of AllFeature-RF are reported in Table IV, together with the results of Selected-RF and Selected-DL. Furthermore, confusion matrices and classification maps of Selected-RF and Selected-DL are reported in Figs. 7 and 8, respectively.

As reported in Table IV, OA and Kappa coefficient of AllFeature-RF are 0.958 and 0.953, respectively; OA and Kappa coefficient of the Selected-DL are 0.957 and 0.952, respectively; OA and Kappa coefficient of the Selected-DL are 0.942 and 0.922, respectively. Both Selected-RF and Selected-DL have great improvement compared to spectral-based ones. For example, AllFeature-RF is 0.128 and 0.144 higher than that of the Spectral-RF in OA and Kappa, respectively; the Selected-RF is 0.127 and 0.143 higher than that of the Spectral-RF in OA and Kappa, respectively; the Selected-DL is 0.158 and 0.233 higher than that of the Spectral-RF in OA and Kappa, respectively.

As shown in Fig. 7, both AllFeature-RF and Selected-RF show that most of the land types are correctly classified. Errors clearly exist inside vegetation and nonvegetation. For instance, bamboos are classified as coniferous forests, and buildings are classified as bare land and road. In the case of Selected-DL, problems became outstanding that tea garden, broadleaf forest, and farmland, where most of the errors lie in, mixed with each other with high error classification rates.

The spectral-RF indicates the extent to which land-use types could be recognized by the spectrum. Cutting sites, for example, are well discriminated for exhibiting high value in the R band (see Fig. 6). By contrast, most of the light of the Red band is absorbed by vegetation for photosynthesis. Body of water has an NDWI value of approaching 0; bare lands have NDWI values about –0.4 to –0.2; and vegetation have lower NDWI values.

C. *Fusion of RF and CNN*

With trained CNNs, the fully connected layers of CNNs with RF classifier and feature maps of all layers of the CNN are obtained. Qualities of all the feature maps are tested. Here the integrated models ConvRF trained with spectral data are denoted as Spectral-ConvRFn ($n = 1, 2, \ldots, 8$), where n indicates which layer it is in the CNN, and ConvRF trained with spectral data are denoted as Selected-ConvRFn ($n = 1, 2, \ldots, 8$). The confusion matrices of the Spectral-ConvRFn and Selected-ConvRFn are reported in Figs. 9 and 10. The OA and Kappa coefficient are shown in Fig. 11. And for convenient comparison, only Spectral-ConvRF7, Spectral-ConvRF8, Spectral-ConvRF7, and Spectral-ConvRF8 are reported especially, with classification maps in Fig. 12 and accuracy rates in Table V.

As shown in Table V, the OA and Kappa coefficient of the Spectral-ConvRF7 are 0.953 and 0.948, respectively; OA and Kappa coefficient of Spectral-ConvRF8 are 0.962 and 0.958, respectively; OA and Kappa coefficient of Selected-ConvRF7 are 0.987 and 0.986, respectively; OA and Kappa coefficient of Spectral-ConvRF8 are 0.991 and 0.990, respectively. The Selected-ConvRF7 and Selected-ConvRF8 are the best methods (reach the highest OA and Kappa coefficient) in our study. Selected-ConvRF8 outperforms Selected-ConvRF7 slightly as well as Spectral-ConvRF8 outperforms Spectral-ConvRF7 slightly.

Together with Figs. 9–11, it is obvious that the accuracy rates, as well as stability, increase gradually as the layers go deeper. However, misclassifications exist among all the classes. As an economic crop, Lei bamboo covers most of the land in our research area. Intense artificial managements are employed to cultivate Lei bamboo. For example, Fig. 12 shows a village and its surroundings, where the traditional crops mix with Lei

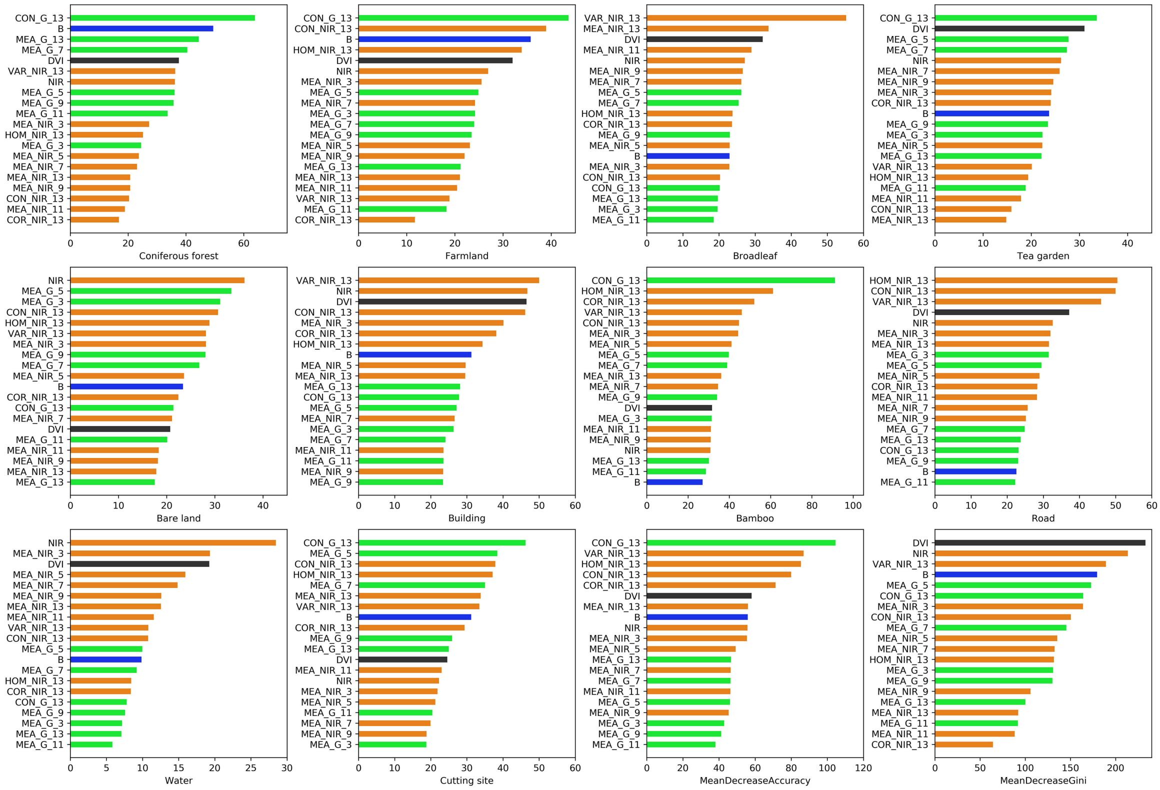


Fig. 6. Variable importance scores of all features. Variable naming is explained as (feature name)-(based spectral band)-(window size), for example, MEA_G_13 indicates the MEA based on the green band with a window size of 13 × 13. Bar colors indicate the spectral bands; green: green band, red: red band, blue: blue band, orange: NIR band, and black: vegetation index variables.

TABLE III
FEATURES SELECTED FOR FURTHER RESEARCH

Spectral	Vegetation index	Texture		
		MEA of NIR	MEA of G	Other Texture
B	DVI	MEA_NIR_3	MEA_G_3	CON_G_13
NIR		MEA_NIR_5	MEA_G_5	VAR_NIR_13
		MEA_NIR_7	MEA_G_7	HOM_NIR_13
		MEA_NIR_9	MEA_G_9	COR_NIR_13
		MEA_NIR_11	MEA_G_11	CON_NIR_13
		MEA_NIR_13	MEA_G_13	

Variable naming is explained as (feature name)-(based spectral band)-(window size), for example, MEA_G_13 indicates the MEA based on the green band with a window size of 13 × 13.

TABLE IV
CLASSIFICATION ACCURACY COMPARISON OF ALLFEATURE-RF, SELECTED-RF, AND SELECTED-DL

Feature	N	OA	Kappa		1	2	3	4	5	6	7	8	9	10
AllFeature-RF	200	0.958	0.953	UA	0.988	0.986	0.958	0.952	0.935	0.962	0.913	0.961	1.000	0.966
				PA	0.931	0.859	0.968	0.988	0.966	0.950	0.984	0.937	1.000	0.988
Selected-RF	20	0.957	0.952	UA	0.970	0.918	0.952	0.927	0.988	0.965	0.941	0.955	1.000	0.966
				PA	0.946	0.882	0.962	0.950	0.934	0.953	0.960	0.988	1.000	1.000
Selected-DL	20	0.942	0.922	UA	0.992	0.866	0.609	0.420	0.850	0.967	0.996	0.858	0.999	0.996
				PA	0.982	0.988	0.201	0.981	0.986	0.936	0.926	0.859	0.998	0.997

N indicates the number of variables. OA indicates the overall accuracy rate. Kappa indicates the kappa coefficient. The number from 1 to 10 represents land types; 1: Coniferous forest; 2: Farmland; 3: Broadleaf; 4: Tea garden; 5: Bare land; 6: Building; 7: Bamboo; 8: Road; 9: Water; 10: Cutting site.

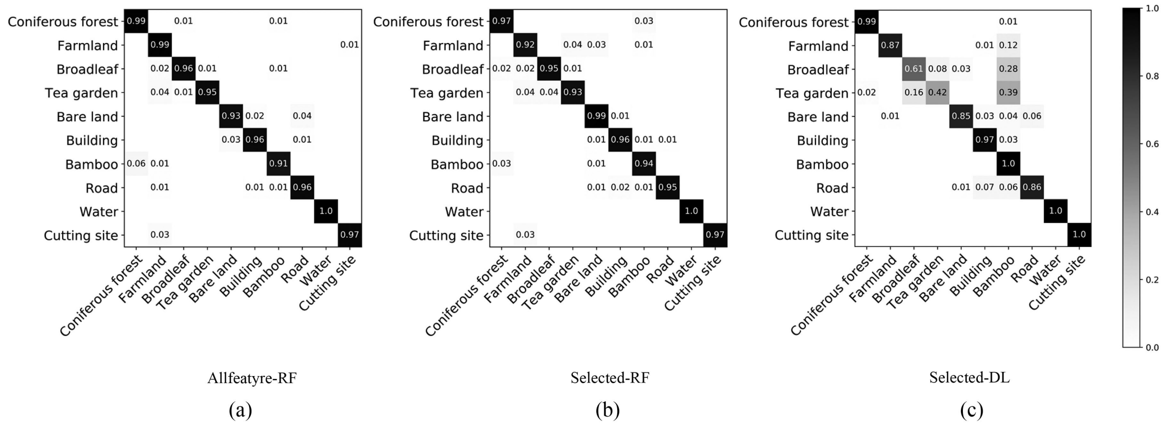


Fig. 7. (a) Confusion matrix of AllFeature-RF. (b) Confusion matrix of Selected-RF. (c) Confusion matrix of Selected-DL.

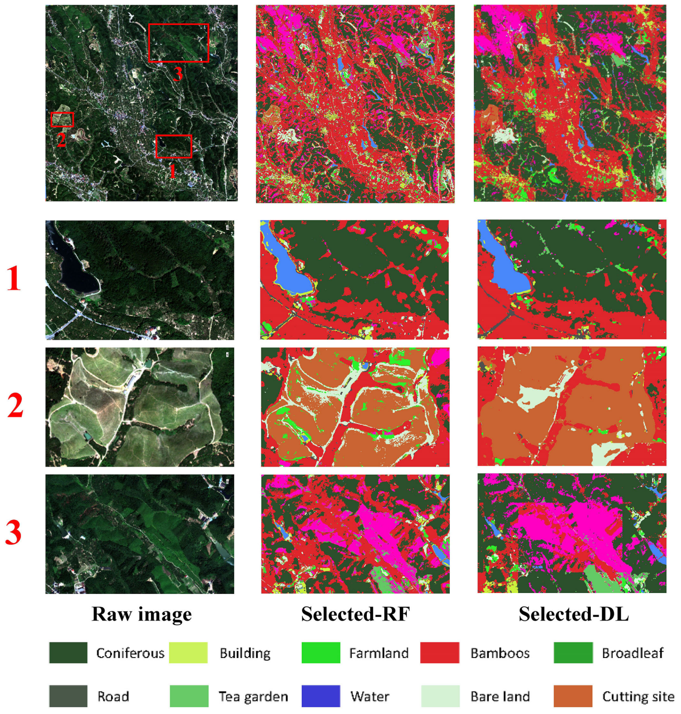


Fig. 8. Result of Selected-RF and Selected-DL where the left column shows the ground truth, the mid column shows the result of Selected-RF, and the right column shows the result of Selected-DL.

bamboo. Some roads inside the farmland are classified as bare land, but it is rational for that since these roads vary in materials such as stones, soils, and even sand. Obvious bounds still exist in ConvRF7, which is modified in ConvRF8.

IV. Discussion

The VHRRS land use classification tasks face many difficulties, for instance, obstruction of an interclass variation, underutilization of information, limitation of computation, and needs of expertise experiments. Although RF and CNN, as state-of-the-art methods, show considerable performance and address some of these problems, there could be a great improvement, as shown in our study, by means of adding additional information and combination of the RF and CNN. Some details are discussed in the following sections.

A. CNN Structures

For these years, numerous CNN structures have been proposed. Although we are not here in this part to explore the best CNN structure, we are testing a considerable better structure as a counterpart of RF. Hence, CNNs with different numbers of layers are constructed, which are denoted as layer_n. Some training processes have been shown in Fig. 13, where all the models are trained from scratch. And the details of the structures are reported in Table VI. It is obvious that the cross-entropy loss (CL) of layer_5 is slightly higher than others and FCN8s got down fast. And with the structures going deeper, the CLs converge. However, it is well known that the accuracy rate does not change along with the loss curve. Therefore, the last ten epochs' parameters of these structures, as well as layer_8 with different training ratios, are collected and tested to produce a violin plot (see Fig. 14). One is unlikely to reach that the deeper a CNN is, the better the results are because layer_16 and layer_20 do not converge well. Comprehensively taking effectiveness and economies into consideration, we took layer_8 as a representation of the CNN model.

B. Analysis of RF and CNN

In both cases (Spectral- and Selected-), RFs better CNNs (see Tables II and IV). Although it is hard to conclude that the RF is better than CNN for the reason that CNN is sensitive to parameters setting and structure designs, we could conclude that RF is more robust than CNN and less sensitive to initialization.

In the case of Spectral-RF (see Fig. 3), it is obvious that some Lei bamboo forests are misclassified as other land types (except for water). There are several reasons. First, in order to conveniently regulate the Lei bamboo, the most important economic forest in the study area, Lei bamboo is always grown along with village and road. And, as a typical zone covered by the subtropical forest, the study area is of a complex forest distribution where broad-leaved trees are mixed with coniferous.

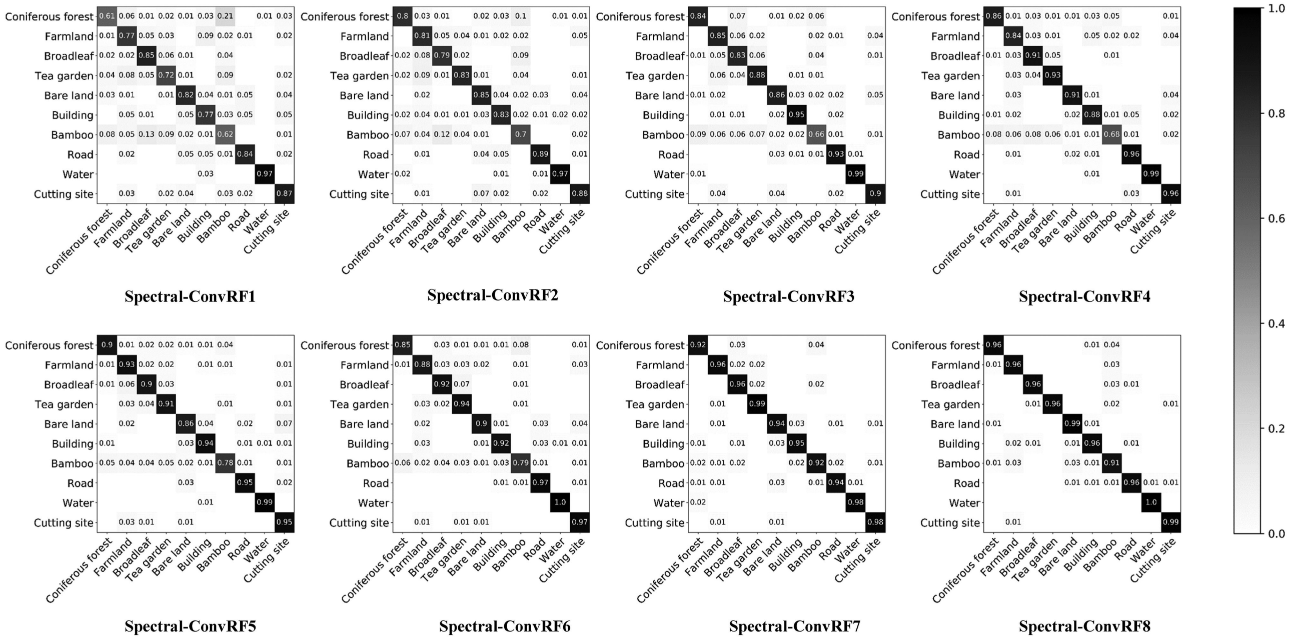


Fig. 9. Confusion matrices of Spectral-ConvRF. In the expression of Spectral-ConvRFn, n indicates the layer of CNN.

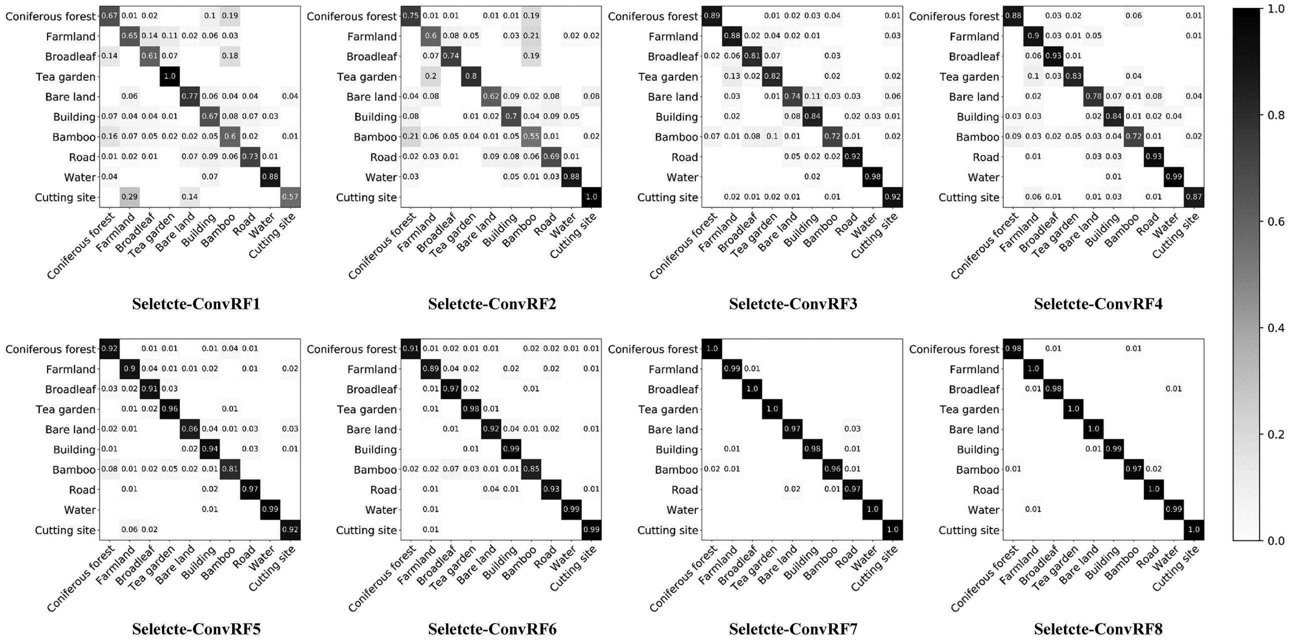
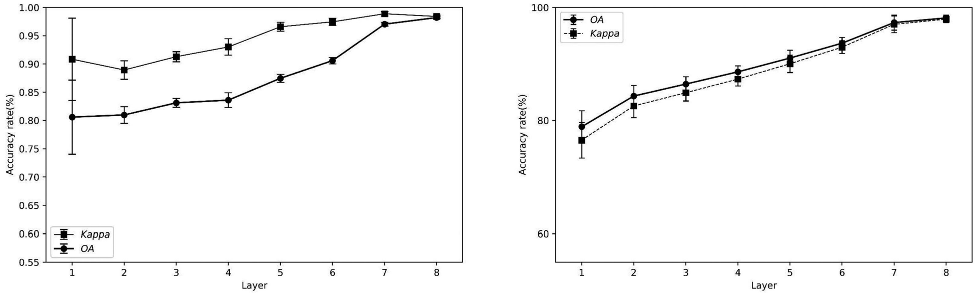


Fig. 10. Confusion matrices of Selected-ConvRF. In the expression of Spectral-ConvRFn, n indicates the layer of CNN.

However, the invasion of the Lei bamboo into other forest types leads to a further mixture between bamboos and other vegetation.

The uses of elaborated chosen features, NIR and green bands as well as their relevant texture features and vegetation indices (see Fig. 6), make improvements both on RF and CNN. The GLCM, especially, made great improvement in the results. According to previous studies, radiational distortion that makes significant impact on the spectrums does not alter the GLCM textures very much [74]. Therefore, comparing to use four bands, the incorporation of the GLCM that derived from the four bands better the results over nearly all the land-cover types. However, some classes of CNN are badly discriminated. This is due to imbalanced training samples, for example, there are limited areas covered with tea gardens and broadleaf forests. But for RF, stratified random sampling is applied, this flaw is moderated a lot. And the edge effect of the CNN cannot be addressed here for the rigid separation of image. Therefore, further research works about the irregular and overlap separation of images are needed (see Figs. 4 and 8).

All models in our study performed well on water discrimination. Although some of the body of the water exhibits green color (with a perspective of RGB), NIR band and the vegetation



(a)

(b)

Fig. 11. (a) OA and Kappa coefficient of Spectral-ConvRFn, where n indicates the layer of CNN. (b) OA and Kappa coefficient of Selected-ConvRFn, where n indicates the layer of CNN.

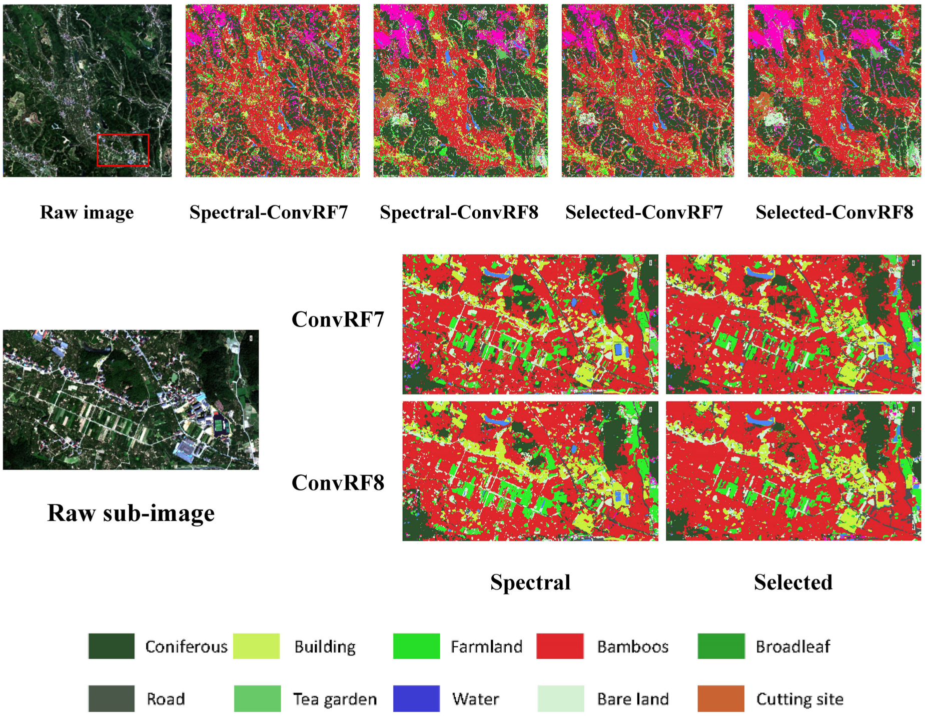


Fig. 12. Classification maps of Spectral-ConvRF7, Spectral-ConvRF8, Selected-ConvRF7, and Selected-ConvRF8.

TABLE V
CLASSIFICATION ACCURACY COMPARISON OF ALLFEATURE-RF, SELECTED-RF, AND SELECTED-DL

Feature	OA	Kappa		1	2	3	4	5	6	7	8	9	10
Spectral-ConvRF7	0.953	0.948	UA	0.924	0.958	0.959	0.990	0.942	0.949	0.917	0.941	0.984	0.977
			PA	0.948	0.958	0.928	0.944	0.927	0.949	0.933	0.977	0.992	0.977
Spectral-ConvRF8	0.962	0.958	UA	0.956	0.959	0.960	0.955	0.985	0.959	0.915	0.956	1.000	0.989
			PA	0.964	0.943	0.984	1.000	0.957	0.959	0.896	0.982	0.991	0.978
Selected-ConvRF7	0.987	0.986	UA	1.000	0.990	1.000	1.000	0.970	0.981	0.962	0.970	1.000	1.000
			PA	0.979	0.980	0.991	1.000	0.980	1.000	0.990	0.951	1.000	1.000
Selected-ConvRF8	0.991	0.990	UA	0.980	1.000	0.980	1.000	1.000	0.989	0.971	1.000	0.990	1.000
			PA	0.990	0.982	0.990	1.000	0.991	1.000	0.990	0.981	0.990	1.000

OA indicates the overall accuracy rate. Kappa indicates the kappa coefficient. The number from 1 to 10 represents land types; 1: Coniferous forest; 2: Farmland; 3: Broadleaf; 4: Tea garden; 5: Bare land; 6: Building; 7: Bamboo; 8: Road; 9: Water; 10: Cutting site.

indices, most of which are computed involved with the NIR band, make water distinctive (see Fig. 6)—body of water absorbs the NIR spectrum while vegetations reflect. However, this could be a disadvantage that some VHRRS and unmanned aerial vehicle images do not contain bands except for RGB [7], [8].

C. Analysis of the Fusion of CNN and RF

The ConvRF takes advantage of the two machine learning methods and frees humans from feature selection. Although the lack of deep-in understanding of the decision process of DL has labeled it “questionable,” the ability of the CNN models to

TABLE VI
DETAILS OF THE CNN LAYERS INFORMATION

	size (MB)	8@3×3	16@3×3	32@3×3	64@3×3	128@3×3	256@3×3	512@3×3	4096@3×3	Skip layer
Layer-5	0.39	1	1	1	1	1	\	\	\	√
Layer-6	1.52	1	1	1	1	1	1	\	\	√
Layer-7	6.05	1	1	1	1	1	1	1	\	√
Layer-8	15	1	1	1	1	1	1	2	\	√
Layer-9	15.2	1	1	1	2	1	1	2	\	√
Layer-10	15.7	1	1	1	2	2	1	2	\	√
Layer-16	39.1	1	1	1	3	3	3	4	\	√
Layer-20	39.3	1	2	3	4	3	3	4	\	√
Layer-8-norm	15	1	1	1	1	1	1	2	\	×
FCN8s	**675**				2	2	2	3	2	√

Processes of how the image goes through are provided. For example, 8@3 × 3 indicates that sizes of the convolutional filter are 3 × 3 and the number of obtained feature maps is 8.

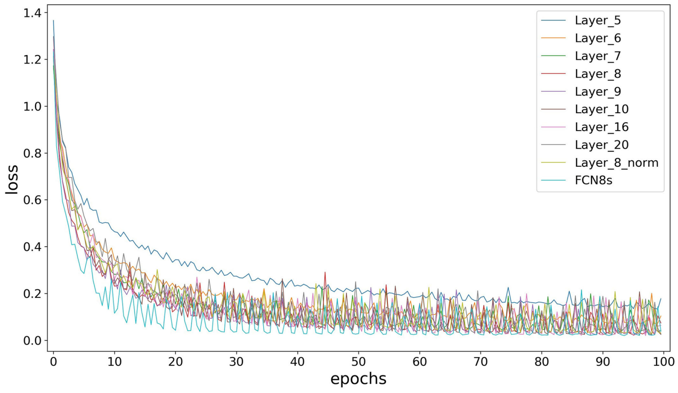


Fig. 13. Cross-Entropy loss of different structures. Every model was trained for 100 epochs and then the losses were recorded.

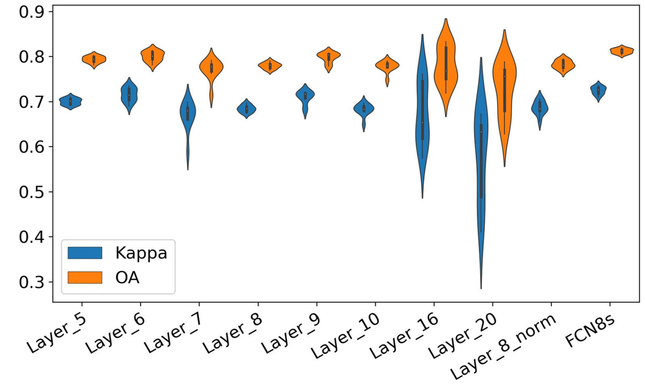


Fig. 14. Every model was trained for 100 epochs and then the OA and Kappa distributions of the last ten epochs with different CNN structures were recorded.

learn high-level representation from the large amounts of data has been widely accepted. The high-level representation, kinds of abstract features and cannot be understood by conventional language, is undeniably informative [24], [34]. Especially for VHRRS, CNN learns the complex spatial pattern surrounding the target pixel [51]. In contrast, the construction of the RF is under "visible" and rational control.

Edge effect, the most outstanding problem in CNN, results from the lack of information on the edge of a subimage. However, another reason is the limitation of hardware that we cannot take the whole image as input. An overlap during segmentation will be a possible resolution. As shown in Fig. 12, ConvRF7 still exhibits the edge-effect, whereas gets smoother in ConvRF8. A possible reason is that a 3 × 3 average pool layer is applied after ConvRF.

V. CONCLUSION

In this article, we employed ConvRF for the VHRRS classification task. The CNN works as a high-level representation extractor to utilize spectral-spatial features. The RF is constructed and takes the place of FC layer in the CNN as a classifier. This combination of the RF and CNN demonstrates better results and involves less artificial efforts. As shown in our study, textures and vegetation indices improve OA from 0.830 to 0.957. High representation extracted by the CNN improves the OA to 0.962. And further, if the CNN learning from a delicately designed dataset (band data together with other low-level information like vegetation indices and textures) in the first place, greater improvement (with an OA of 0.991), although costly, could be obtained.

The fusion of CNN and RF could take advantage of the texture information contained in high spatial resolution imagery and robustness of the RF as a classifier. The disadvantage of our work is that the fusion model is computation intensive. Better hardware, such as memory and graphics processing unit would be helpful. For RS tasks, such as classification of land use in the urban area, shapes and edges are crucial. The ConvRF could be unsatisfactory.

More attention could be focused on the size of the patches in this fusion model. And given that the DL technology is growing rapidly, more state-of-the-art CNN models could be taken into consideration. The fusion of multilevel feature representation has a great potential improvement on the existing model.

References

[1] P. Griffiths *et al.*, "Forest disturbances, forest recovery, and changes in forest types across the Carpathian ecoregion from 1985 to 2010 based on Landsat image composites," *Remote Sens. Environ.*, vol. 151, no. 8, pp. 72–88, 2014.

[2] Z. Zhu and C. E. Woodcock, "Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change," *Remote Sens. Environ.*, vol. 152, pp. 217–234, 2014.

[3] N. Han, K. Wang, L. Yu, and X. Zhang, "Integration of texture and landscape features into object-based classification for delineating Torreya using IKONOS imagery," *Int. J. Remote Sens.*, vol. 33, no. 7, pp. 2003–2033, 2012.

[4] J. Dong *et al.*, "Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine," *Remote Sens. Environ.*, vol. 185, pp. 142–154, 2016.

[5] L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive SVM for semisupervised classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3363–3373, Nov. 2006.

[6] I. Ali, F. Greifeneder, J. Stamenkovic, M. Neumann, and C. Notarnicola, "Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data," *Remote Sens.*, vol. 7, no. 12, pp. 16398–16421, 2015.

[7] B. Yang, Z. Dong, Y. Liu, F. Liang, and Y. Wang, "Computing multiple aggregation levels and contextual features for road facilities recognition using mobile laser scanning data," *ISPRS J. Photogramm. Remote Sens.*, vol. 126, pp. 180–194, 2017.

[8] J. Sherrah, "Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery," 2016, *arXiv:1606.02585*.

[9] T. Panboonyuen, K. Jitkajornwanich, S. Lawawirojwong, P. Srestasathiern, and P. Vateekul, "Road segmentation of remotely-sensed images using deep convolutional neural networks with landscape metrics and conditional random fields," *Remote Sens.*, vol. 9, no. 7, 2017, Art. no. 680.

[10] T. Qu, Q. Zhang, and S. Sun, "Vehicle detection from high-resolution aerial images using spatial pyramid pooling-based deep convolutional neural networks," *Multimedia Tools Appl.*, vol. 76, pp. 21651–21663, 2017.

[11] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ, USA: Prentice-Hall, 1994, pp. 71–80.

[12] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[13] S. S. Joibary, "Forest attributes estimation using aerial laser scanner and TM data," *Forest Syst.*, vol. 22, no. 3, pp. 484–496, 2013.

[14] M. Jia, L. Tong, Y. Chen, Y. Wang, and Y. Zhang, "Rice biomass retrieval from multitemporal ground-based scatterometer data and RADARSAT-2 images using neural networks," *J. Appl. Remote Sens.*, vol. 7, no. 1, 2013, Art. no. 073509.

[15] R. Mahabir, A. Croitoru, A. Crooks, P. Agouris, and A. Stefanidis, "A critical review of high and very high-resolution remote sensing approaches for detecting and mapping slums: Trends, challenges and emerging opportunities," *Urban Sci.*, vol. 2, no. 1, 2018, Art. no. 8.

[16] F. Löw, C. Conrad, and U. Michel, "Decision fusion and non-parametric classifiers for land use mapping using multi-temporal RapidEye data," *ISPRS J. Photogramm. Remote Sens.*, vol. 108, pp. 191–204, 2015.

[17] H. Ghassemian, "A review of remote sensing image fusion methods," *Inf. Fusion*, vol. 32, pp. 75–89, 2016.

[18] C. Zhang *et al.*, "A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 140, pp. 133–144, 2018.

[19] Y. Qian, W. Zhou, J. Yan, W. Li, and L. Han, "Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery," *Remote Sens.*, vol. 7, no. 1, pp. 153–168, 2015.

[20] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *Int. J. Remote Sens.*, vol. 39, no. 9, pp. 2784–2817, 2018.

[21] P. Thanh Noi and M. Kappas, "Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery," *Sensors*, vol. 18, no. 1, 2018, Art. no. 18.

[22] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[23] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, 2016.

[24] Y. Tao, M. Xu, Z. Lu, and Y. Zhong, "DenseNet-based depth-width double reinforced deep learning neural network for high-resolution remote sensing image per-pixel classification," *Remote Sens.*, vol. 10, no. 5, 2018, Art. no. 779.

[25] K. J. Archer and R. V. Kimes, "Empirical characterization of random forest variable importance measures," *Comput. Statist. Data Anal.*, vol. 52, no. 4, pp. 2249–2260, 2008.

[26] S. Boonprong, C. Cao, W. Chen, and S. Bao, "Random forest variable importance spectral indices scheme for burnt forest recovery monitoring—Multilevel RF-VIMP," *Remote Sens.*, vol. 10, no. 6, 2018, Art. no. 807.

[27] B. Melville, A. Lucieer, and J. Aryal, "Object-based random forest classification of Landsat ETM+ and WorldView-2 satellite imagery for mapping lowland native grassland communities in Tasmania, Australia," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 66, pp. 46–55, 2018.

[28] C. Pelletier, S. Valero, J. Inglada, N. Champion, and G. Dedieu, "Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas," *Remote Sens. Environ.*, vol. 187, pp. 156–168, 2016.

[29] M. Kühnlein, T. Appelhans, B. Thies, and T. Nauss, "Improving the accuracy of rainfall rates from optical satellite sensors with machine learning—A random forests-based approach applied to MSG SEVIRI," *Remote Sens. Environ.*, vol. 141, no. 141, pp. 129–143, 2014.

[30] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, "Bias in random forest variable importance measures: Illustrations, sources and a solution," *BMC Bioinf.*, vol. 8, no. 1, 2007, Art. no. 25.

[31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, pp. 84–90, 2017.

[33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[34] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[36] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[37] D. Silver *et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[38] O. Firat, K. Cho, and Y. Bengio, "Multi-way, multilingual neural machine translation with a shared attention mechanism," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol.*, 2016, pp. 866–875.

[39] X. Huang, X. Han, S. Ma, T. Lin, and J. Gong, "Monitoring ecosystem service change in the City of Shenzhen by the use of high-resolution remotely sensed imagery and deep learning," *Land Degradation Develop.*, vol. 30, no. 12, pp. 1490–1501, 2019.

[40] L. Ying, H. Zhang, X. Xue, Y. Jiang, and S. Qiang, "Deep learning for remote sensing image classification: A survey," *Wiley Interdisciplinary Rev. Data Mining Knowl. Discovery*, no. 8, 2018, Art. no. e1264.

[41] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, 2019.

[42] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[43] P. Li, P. Ren, X. Zhang, Q. Wang, X. Zhu, and L. Wang, "Region-wise deep feature representation for remote sensing images," *Remote Sens.*, vol. 10, no. 6, 2018, Art. no. 871.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[45] 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.

[46] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, May 2017.

[47] L. Jiao, M. Liang, H. Chen, S. Yang, and X. Cao, "Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5585–5599, Oct. 2017.

[48] H. Li, T. Chao, Z. Wu, C. Jie, and D. Min, "RSI-CB: A large scale remote sensing image classification benchmark via crowdsource data," 2017, *arXiv:1705.10450*.

[49] X. Sun, S. Shen, X. Lin, and Z. Hu, "Semantic labeling of high resolution aerial images using an ensemble of fully convolutional networks," *J. Appl. Remote Sens.*, vol. 11, no. 4, 2017, Art. no. 042617.

[50] M. Papadomanolaki, M. Vakalopoulou, S. Zagoruyko, and K. Karantzalos, "Benchmarking deep learning frameworks for the classification of very high resolution satellite multispectral data," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. III-7, pp. 83–88, 2016.

[51] C. Zhang *et al.*, "An object-based convolutional neural network (OCNN) for urban land use classification," *Remote Sens. Environ.*, vol. 216, pp. 57–70, 2018.

[52] H. Zhu, L. Jiao, W. Ma, F. Liu, and W. Zhao, "A novel neural network for remote sensing image matching," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2853–2865, Sep. 2019.

[53] Y. Tan, S. Xiong, and Y. Li, "Automatic extraction of built-Up areas from panchromatic and multispectral remote sensing images using double-stream deep convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 3988–4004, Nov. 2018.

[54] G. Scarpa, M. Gargiulo, A. Mazza, and R. Gaetano, "A CNN-based fusion method for feature extraction from sentinel data," *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 236.

[55] N. Audebert, B. Le Saux, and S. Lefèvre, "Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks," *ISPRS J. Photogramm. Remote Sens.*, vol. 140, pp. 20–32, 2018.

[56] G. Fu, C. Liu, R. Zhou, T. Sun, and Q. Zhang, "Classification for high resolution remote sensing imagery using a fully convolutional network," *Remote Sens.*, vol. 9, no. 5, 2017, Art. no. 498.

[57] Y. Cai *et al.*, "A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach," *Remote Sens. Environ.*, vol. 210, pp. 35–47, 2018.

[58] C. Zhang and Z. Xie, "Combining object-based texture measures with a neural network for vegetation mapping in the Everglades from hyperspectral imagery," *Remote Sens. Environ.*, vol. 124, no. 124, pp. 310–320, 2012.

[59] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[60] L. Tao, J. Leng, L. Kong, G. Song, B. Gang, and W. Kai, "DCNR: Deep cube CNN with random forest for hyperspectral image classification," *Multimedia Tools Appl.*, vol. 3, pp. 1–23, 2018.

[61] G. Xu, M. Liu, Z. Jiang, D. Söffker, and W. Shen, "Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning," *Sensors*, vol. 19, no. 5, 2019, Art. no. 1088.

[62] Q. Qiu, J. Lezama, A. Bronstein, and G. Sapiro, "ForestHash: Semantic hashing with shallow random forests and tiny convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 442–459.

[63] X. Chen, Z. Cao, Y. Xiao, and Z. Fang, "Hand pose estimation in depth image using CNN and random forest," *Proc. SPIE*, vol. 10609, 2018, Art. no.106091K.

[64] X. Xu and E. Hirata, "Decomposition patterns of leaf litter of seven common canopy species in a subtropical forest: N and P dynamics," *Plant Soil*, vol. 273, no. 1/2, pp. 279–289, 2005.

[65] D. Haboudane, J. R. Miller, E. Pattey, P. J. Zarco-Tejada, and I. B. Strachan, "Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture," *Remote Sens. Environ.*, vol. 90, no. 3, pp. 337–352, 2004.

[66] R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, May 1979.

[67] M. H. Bharati, J. J. Liu, and J. F. Macgregor, "Image texture analysis: Methods and comparisons," *Chemometrics Intell. Lab. Syst.*, vol. 72, no. 1, pp. 57–71, 2004.

[68] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.

[69] G. Huang, Z. Liu, K. Q. Weinberger, and V. D. M. Laurens, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.

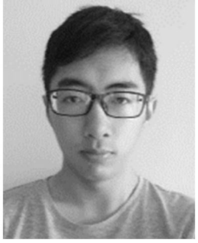
[70] H. R. Roth *et al.*, "An application of cascaded 3D fully convolutional networks for medical image segmentation," *Comput. Med. Imag. Graph.*, vol. 66, pp. 90–99, 2018.

[71] S. V. Beijma, A. Comber, and A. Lamb, "Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data," *Remote Sens. Environ.*, vol. 149, pp. 118–129, 2014.

[72] J. Dai *et al.*, "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 764–773.

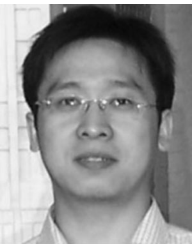
[73] D. A. Clausi, "Texture segmentation of SAR Sea Ice Imagery," Ph.D. dissertation, Dept. Syst. Design Eng., Univ. Waterloo, Waterloo, Ontario, Canada, 1996.

[74] D. Boyd and F. Danson, "Satellite remote sensing of forest resources: Three decades of research development," *Prog. Phys. Geography*, vol. 29, no. 1, pp. 1–26, 2005.



Luofan Dong received the B.S. degree in human geography and urban-rural planning from Zhejiang A&F University, Hangzhou, China, in 2017. He is currently working toward the M.S. degree with Zhejiang A&F University, Zhejiang, China.

His research interests include digital image processing, machine learning, and forest resource monitoring using remotely sensed data.

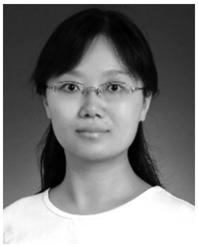


Huaqiang Du received the B.S. degree in forestry and the M.S. degree in forest management from Northeast Forestry University, Heilongjiang, China, in 1999 and 2002, respectively, and the Ph.D. degree from Beijing Forestry University, Beijing, China, in 2005.

He is currently a Professor with the School of Environment and Resources Science, Zhejiang A&F University, Zhejiang, China. His research interests include digital image processing, forest resource monitoring using multisource remotely sensed data, and forest carbon estimation with remote sensing techniques.

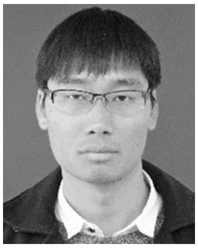
Fangjie Mao received the B.S. degree in computer science and technology from the PLA Naval University of Engineering, Wuhan, China, in 2010, and the M.S. degree in forest management and the Ph.D. degree in bamboo resources and efficient utilization from Zhejiang A&F University, Hangzhou, China, in 2013 and 2016, respectively.

He is currently a Lecturer with the School of Environment and Resources Science, Zhejiang A&F University, Zhejiang. His research interests include forest carbon estimation and monitoring with the combination of remote sensing techniques and ecosystem models and ecosystem model development.



Ning Han received the B.S. degree in geographic information system from Zhengzhou University, Zhengzhou, China, in 2006, and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2011.

She is currently an Associate Professor with Zhejiang A&F University, Hangzhou, China. Her research interests include the research of satellite-based forest resources monitoring and carbon cycling.



Xuejian Li received the B.S. degree in geographical science from Qiqihar University, Qiqihar, China, in 2014, and the M.S. degree in forest management in 2017 from Zhejiang A&F University, Zhejiang, China, where he is currently working toward the Ph.D. degree in bamboo resources and efficient utilization.

His research focuses on the temporal and spatial evolution of bamboo forest phenology and its response to climate change.

Guomo Zhou received the B.S. degree in forestry from Zhejiang A&F University, Zhejiang, China, in 1982, the M.S. degree in forest management from Beijing Forestry University, Beijing, China, in 1987, and the Ph.D. degree in soil science from Zhejiang University, Zhejiang, China, in 2006.

He is currently a Professor with the School of Environment and Resources Science and the President of Zhejiang A&F University. His research interests include forest carbon monitoring, climate change, carbon management, and sustainable forest management.

Di'en Zhu received the B.S. degree in forestry from Southwest Forestry University, Kunming, China, in 2015, and the M.S. degree in forest management from Zhejiang A&F University, Zhejiang, China, in 2018. He is currently working toward the Ph.D. degree with Beijing Forestry University, Beijing, China.

His research interests include forest resources monitoring and extracting using multiple remote sensing data.

Junlong Zheng received the B.S. degree from Zhejiang A&F University, Zhejiang, China, in 2017. He is currently working toward the M.S. degree in forest management. His research interets include temporal and spatial distribution of carbon sources and sinks in subtropical forests and their response to climate change.

Meng Zhang received the B.S. degree from Xinyang Normal University, Henan, China, in 2017. She is currently working toward the M.S. degree with Zhejiang A&F University, Zhejiang, China.

Her research interests include forest resources monitoring and extracting using multiple remote sensing datasets.

Luqi Xing received the B.S. degree in geographic information system and the M.S. degree in forestry management both from Zhejiang A&F University, Zhejiang, China, in 2015 and 2019, respectively.

Her research interests include assimilation of multiresolution LAI in bamboo forest and its application in carbon cycle simulation.

Tengyan Liu received the B.S. degree in geographical science from Xingtai University, Xingtai, China, in 2016, and the M.S. degree in forest management from Zhejiang A&F University, Zhejiang, China, in 2019.

Her research interest focuses on temporal-spatial evolution of forest carbon stock in the subtropical forests.