A Novel Multi-Training Method for Time-Series Urban Green Cover Recognition From Multitemporal Remote Sensing Images

Wenye Wang, Shenghua Wan, Pengfeng Xiao, *Senior Member, IEEE*, and Xueliang Zhang, *Member, IEEE*

***Abstract*—Urban green space plays a crucial role in the construction of ecological city and livable environment. While multitemporal remote sensing images provide strong support for urban green cover monitoring, they often suffer from *data shifting*, where the data distribution varies from phase to phase. Designing a general multitemporal framework to extract urban green cover is challenging, mainly due to possible time-consuming data labeling and inconsistent prediction. To address that, we propose multi-training, a novel method for land cover classification on multitemporal remote sensing images. Multi-Training is a two-stage method to independently train classifier on each phase in the training stage and then to combine the information from all the classifiers in the communication stage. As a semi-supervised learning method, multi-training adopts a new rule to obtain the confidence of unlabeled samples' prediction, which reduces the dependence on labeled data and increases the result's consistency between phases. The experimental results show that multi-training outperforms self-training, co-training, tri-training, and super-training on both accuracy and consistency on multitemporal remote sensing image datasets. Furthermore, we have analyzed the necessary parameters in our method and conclude that the number and the combination of phases will dominate the prediction results.**

***Index Terms*—Multitemporal remote sensing images, multi-training, semi-supervised learning, urban green cover, vegetation recognition.**

I. Introduction

VEGETATION, as an important part of the urban ecological landscape, plays a key role in air filtration, microclimate regulation, noise reduction, and water quality amelioration [1]. Therefore, inventorying the temporal and spatial information of urban green cover is beneficial for decision making about natural landscape management and planning [2]. With the rapid development of remote sensing techniques, satellite data have been one of the most important data sources to extract urban green cover, which brings new impetus to the monitoring of urban vegetation. The multitemporal remote sensing images provide more information about the land cover in time series than single temporal image. Thus, it supplies strong support for land classification and landscape change detection [3]. However, the classification of remote sensing images under multitemporal conditions is mainly faced with these problems.

1) It is time-consuming and labor-tedious to prepare a large number of labeled samples for training models with supervised learning methods.
2) The labeled samples in one phase cannot be directly applied to another phase because the spectral information of vegetation is always influenced by phenology, weather, and atmospheric conditions [4].
3) The classification results in each phase always differ from each other which cannot guarantee the consistency in time series.

To solve these three problems of land cover classification based on multitemporal remote sensing images, the existed techniques can be divided into three categories: deep learning, transfer learning, and semi-supervised learning (SSL).

Some researchers considered applying deep learning methods to land cover the classification of multitemporal images for its ability of automatically exploring the best representation of data's features. For example, recurrent neural networks (RNNs) were constructed to capture the relationship on time-series images [5], [6]. Long short-term memory was used to deal with long-time-series images to solve the problems of gradient disappearance and gradient explosion [7], [8]. One-dimensional convolutional neural networks (1DCNN) [9] and three-dimensional convolutional neural networks (3DCNN) [10] were designed to classify crops from multitemporal remote sensing images. In 1DCNN, the crop was classified with the characteristics of time dimension. In 3DCNN, the time direction was the third dimension besides the image length and width. In addition, several hybrid models were proposed. For example, Rußwurm and Körner [11] added two-dimensional convolutional layers as spatial feature extractors and connected recurrent cells in a bidirectional way to reduce temporal biases toward later images. Interdonato et al. [12] combined CNN and RNN into one end-to-end architecture to learn diverse spectral–spatial–temporal feature representation. Although deep learning methods have

Manuscript received 28 May 2022; revised 6 August 2022 and 19 October 2022; accepted 31 October 2022. Date of publication 4 November 2022; date of current version 10 November 2022. This work was supported by the National Natural Science Foundation of China under Grant 41871235 and Grant 42071297. *(Corresponding author: Pengfeng Xiao.)*

The authors are with the Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Key Laboratory for Land Satellite Remote Sensing Applications of Ministry of Natural Resources, School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China (e-mail: wwy13594137516@gmail.com; wanshenghua0and1@gmail.com; xiaopf@nju.edu.cn; zxl@nju.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3218919

made great performance, it takes huge time cost to collect a great number of labeled samples for training.

Transfer learning is another mainstream method to deal with these problems. The purpose of this method is to apply the classifier trained in source domain to target domain and make full use of the knowledge learned in source domain [13]. In recent years, transfer learning methods have been developed for multitemporal remote sensing image classification and land cover classification [14], [15], [16]. However, these methods can work well only if the source domain can be fully trained, i.e., the training data in source domain are adequate. Another problem is that if too many training phases are involved, it can be difficult for the model to perform well on all phases.

SSL is a kind of approach to utilize the relationship and structure among the unlabeled samples to improve the model's performance on datasets [17]. In most cases, the number of unlabeled samples is large enough that contains sufficient information and SSL is proposed to utilize the unlabeled samples to improve the model's performance [18]. SSL methods can be categorized into four classes based on the assumptions as follows:

1) generative methods [19];
2) low-density separation methods, such as transductive support vector machines [20];
3) graph-based methods [21];
4) divergence-based methods, such as cooperative training (co-training) method [22].

These SSL methods have achieved remarkable results in machine learning tasks, and many researchers have applied these methods to classify land cover and land use on remote sensing images [23], [24], [25], [26]. In addition, several researchers modified the SSL methods to fit the classification problem on multitemporal remote sensing images.

Co-Training was originally a semi-supervised method designed for two independent sufficient views that trained two classifiers separately and chose unlabeled samples for each other [22]. Further studies showed that the assumption of two conditionally independent views was not necessary [27]. The key for the co-training approaches to succeed is that there exists a large difference between the classifiers [28]. By using output smearing [30] and fine-tuning networks to create diversity between three convolutional neural networks, trinet [29] was able to classify multiple images and it used multiple dropouts to ensure the accuracy of prediction. However, the computational cost of this method is huge for training three deep neural networks and using multiple dropouts. Co-Training was extended to identify and extract snow cover under two phases of remote sensing images, which worked out greatly [31]. It independently trained two classifiers on unchanged area of each phase and updates the unlabeled samples whose prediction was greater than the threshold into training set. Before training, co-training adopted the fixed rule, Chi-Square distance [32], to compute the data difference between phases to distinguish the changed and unchanged areas, but it could not handle multitemporal data directly because the data distribution shifts seriously.

We propose a novel multi-training method for land cover classification of multitemporal remote sensing images based on a small number of training samples. In the previous works, when the number of labeled samples is limited, the updated samples with pseudolabels given by the model always contain much noise, which can decrease the model's performance in turn [33], [34]. Our motivation is to fully use the multitemporal information to eliminate the noise and update pseudolabels with high confidence. Because of the influence of phenology in multitemporal remote sensing images, vegetation always changes greatly during a long period, which brings great difficulty for green cover recognition. We focus on the accuracy and the consistency of vegetation recognition between different phases when evaluating the performance of the methods. We also divide other land covers into subclasses to decrease the misclassification between vegetation and nonvegetation.

The main contributions of the study are as follows.

1) We propose a method for utilizing information from multitemporal remote sensing images and classifying land covers based on a very small training set, even though each class has only one labeled sample.
2) We explore a new mode in this method to solve the classification problem of multitemporal remote sensing data and convert the data's difference to the power of model's classification ability.
3) We adopt a dynamic way to view different phases' data from a global aspect and make progress both on the accuracy and consistency compared with the previous works.

The rest of this article is organized as follows. The proposed method is illustrated in Section II. The study area and the data are introduced in Section III. In Section IV, the experimental setup and related parameters are described. The results and comparative analysis are represented in Section V. Section VI presents the discussion. Finally, Section VII concludes this article.

II. Methodology

In this section, we first introduce how we design the algorithm for multitemporal problems. Then, we demonstrate the most important part of the proposed method that how to choose and update the unlabeled samples into the training set. Additionally, we explain why we select the random forest (RF) as the base classifier. At the end of this section, we present the proposed method with pseudocodes and explain some training details.

A. Multitemporal Extensions

We propose the multi-training method for the land cover classification problem in multiple phases. The key problem of classifying land cover on multitemporal images is how to make effective use of abundant information to improve the performance of the model. We are motivated by the fact that the joint information from all phases can maximally capture the relationship of land cover and enhance the consistency during the period, and we aim to improve the final classification result by combining the predictions on each phase. To do so, we randomly select training samples to initialize the base classifier on each

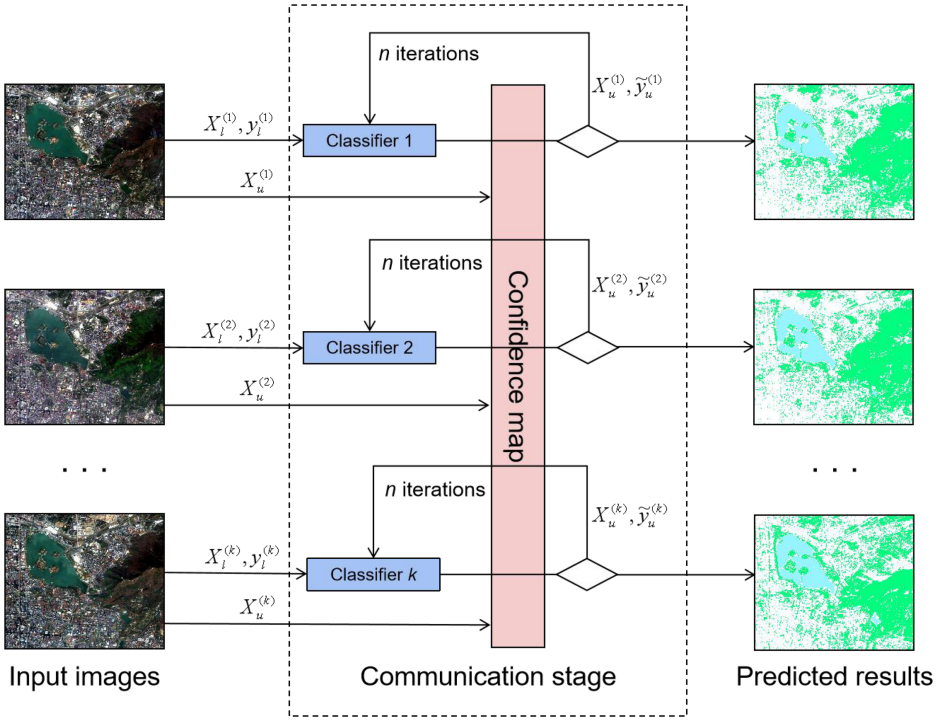


Fig. 1. Flowchart of the proposed multi-training method. The dashed line box shows the communication stage, which is divided into evaluating and updating steps.

phase and update the unlabeled samples whose confidence value is beyond the threshold. Then, given the updated training set, we iteratively optimize the classifiers until the training process terminates. The framework of multi-training is intuitively shown in Fig. 1.

We follow the assumption made by Zhu et al. [31] that changed areas commonly account for a small part of the image during a period. Based on this assumption, the initial classifiers can be trained independently on their own training data without worrying about the temporal inconsistency of the selected samples. Then, after each round of training, all the classifiers communicate with each other in two steps: evaluating and updating. During the evaluating step, all the classifiers' predictions are combined to compute the joint confidence value of the pixels at the same position under multiple phases. At the updating step, the samples with high joint confidence values will be added to the training set of each classifier. By sharing information between each phase, classifiers learn from each other in the communication stage, increasing their robustness and consistency. After finite iterations, the accuracy and consistency of the predictions given by multiple classifiers can be gradually improved. The distribution's difference in multitemporal remote sensing data can guarantee the heterogeneity of these classifiers, which can be regarded as a random data augmentation.

Multi-Training trains classifiers for each phase. The prediction process is to predict the test samples by the classifier trained on corresponding phase. In the accuracy evaluation, the prediction accuracy under each phase can be obtained. Therefore, the overall evaluation for multiple phases is to calculate the mean value of the accuracy of each phase.

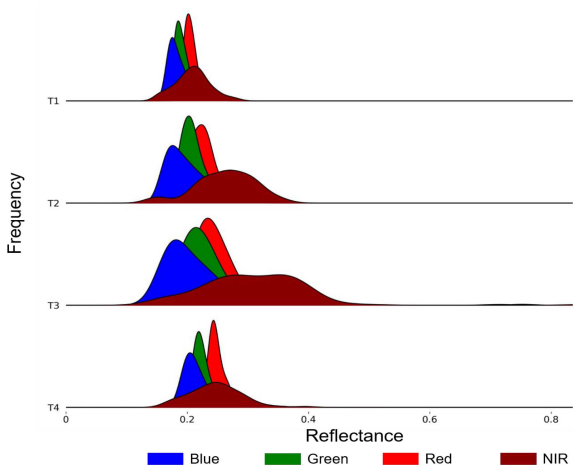


Fig. 2. Data shifting of four-temporal images (T1, T2, T3, and T4) acquired within one year. The data distribution varies a lot in different phases, which brings difficulty to multitemporal classification.

B. Selection of Unlabeled Samples for Training

Data shifting, in which the data distribution varies from phase to phase, is an inevitable problem when processing with multitemporal data, as shown in Fig. 2. Many reasons cause this problem, including the following.

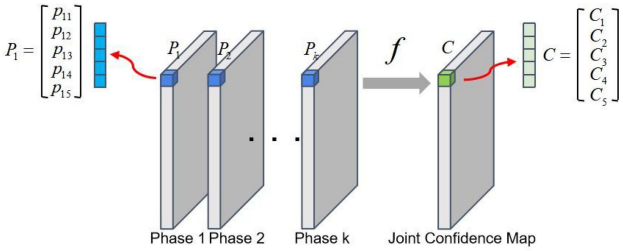


Fig. 3. Computing process of the joint confidence map. After the predicted probability results given by each base classifier in each phase, we combine all the results and calculate the joint confidence map by (1).

1) The influence of solar radiation and atmospheric condition at different phases [35] and changes in the position and height of the satellite lead to the difference in data.
2) Phenological changes in vegetation result in the changes in the spectral spectrum under multiple phases [36].
3) The land cover will be changed by human activities during a long period, which causes the data's feature space in the same area to seriously shift.

For the reason of data shifting, directly applying the labeled data from one phase to another will lead to low classification accuracy and inconsistent results.

To make full use of the multiple data information and enlarge the training set in each phase jointly, we need to determine which sample's class is unchanged so that the rich information of samples on the multiple phases can be utilized for training to improve classification accuracy. In the previous study [31], the changed and unchanged areas are divided into two classes by the unsupervised Kittler–Illingworth thresholding algorithm before running the co-training algorithm. Then, the updating process will take place in the unchanged areas. However, setting a fixed measurement over multiple phases will lead to a dilemma. It is difficult to judge whether a pixel's class changes or not based on the fixed criteria because the distribution of data in different phases has a large deviation.

There is another problem that, when the number of labeled samples is limited, the updated labeled samples in SSL always contain much noise that may mislead the classification of the model. Therefore, to overcome this difficulty, the model should update the samples with high confidence to filter noise.

In the proposed method, we replace explicitly measuring the distance between different phases' data in the previous method [31] by implicitly calculating the joint confidence value based on the classifiers' predicted probability on their own phase to find the unchanging area, as shown in Fig. 3. The joint confidence value on multiphases can not only help model to determine sample's class and change information but also eliminate the noise of updating the pseudolabeled samples. The specific calculation steps of joint confidence are demonstrated below. First, we train a classifier independently on each phase. After training, each classifier gives a probability prediction map of the whole image on their own phase. On each pixel of the probability map, there is a probability vector whose size is equal to the number of land cover classes and each value represents how likely the pixel belongs to this class. Then, we combine all the probability results

Algorithm 1: Multi-Training.

Input: The dataset of multiple phases, $D = \{D_1, \ldots, D_k\}$
The set of labeled samples for each phase, $L = \{L_1, \ldots, L_k\}$
The number of initial labeled samples for each class, $N_l^{(0)}$
The number of updating unlabeled samples for each epoch, N_u
The maximum iterations, N_{iter}
Output: The set of trained models, $f = \{f_1, f_2, \ldots, f_k\}$
Algorithm:
1. Initialize models, $f = \{f_1, f_2, \ldots, f_k\}$
2. **While** epoch $< N_{\text{iter}}$ **do**
3. Train models on each own labeled sample
4. Make predictions on the whole data, generate the probability map P_i for each phase
5. Compute the joint confidence map with the function
6. Calculate the updating threshold of each class by the confidence map
th $=$ mean$(\{p | p \in P_i, i \text{ denotes the } i\text{th class}\})$
7. Randomly select the unlabeled samples whose confidence is higher than th for each class with setting number N_u
8. **End while**

to obtain the final confidence value during the whole episode

$$C(f_1, f_2, \ldots, f_k) = \frac{\left(\prod_{i=1}^{k} P(y_i|x_i, \theta_i)\right)^{\frac{2}{k}}}{Z \cdot \frac{1}{k} \sum_{i=1}^{k} P(y_i|x_i, \theta_i)} \tag{1}$$

where $f_1, f_2, \ldots, f_k$ denote the models; $P(y_i|x_i, \theta_i)$ denotes the probability map on the ith phase; and $C(f_1, f_2, \ldots, f_k)$ denotes the joint confidence map. The idea of (1) comes from the harmonic mean. Dividing the numerator and denominator of (1) by the numerator will result in a structure similar to that of the harmonic mean. The implicit operation of counting the reciprocal makes (1) more sensitive to small values, which ensures that the results of large joint confidence come from inputs that are all large values. The $\frac{2}{k}$ power of the numerator is to balance the dimension and Z is to restrict the summation of the output confidences to 1.

We combine the class prediction information of samples in multiple phases into a confidence map over the whole period by (1). When computing the joint confidence, there may be three conditions.

1) If the probabilities given by all classifiers are high, the generated joint confidence value will also be high, which implies that the pixel's class has not changed and it belongs to this class in this period with high probability.
2) If some classifiers' probabilities are high but others' probabilities are low, it means that the class of the sample probably has changed.
3) If most classifiers' probabilities are low, the joint confidence value must be low, suggesting that this sample does not belong to the class.

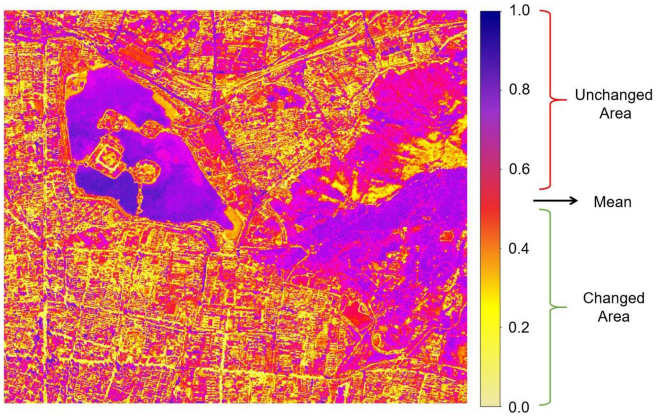


Fig. 4. Changing map generated by the proposed method to show where the land cover possibly changed.

Compared with the voting methods [37], this mapping function provides a potential way to judge the pixels' class and change condition, and the final results will depend on the global joint confidence value. The changing map, where each value in this map maintains the maximum joint confidence given by (1), is intuitively represented in Fig. 4.

After each round of classifiers making predictions, the joint confidence map can be obtained by using the above-mentioned method, and then the training set can be updated. We calculate the average confidence value of one class at the updating step and set it as the threshold. Then, we randomly select the fixed number of pixels whose confidence value is higher than the threshold as a new training sample. To ensure the randomness, a fixed number of samples are randomly selected to add to the training set. The above process is executed repeatedly until the final accuracy converges.

C. Base Classifier

RF is a widely used model in machine learning community that ensembles many decision tree classifiers to predict the final result [38]. RF adopts random selection on data and features to construct many independent decision trees and combines their results by majority voting. RF can fully mine data information and achieve a great result in classification even with a limited number of samples. Therefore, we choose RF as the base classifier to justify whether the proposed algorithm can still improve the performance during the period under the condition that the base classifiers have made full use of the information on each phase.

It is worth noting that other classifiers that can give the predicted class probability of samples (e.g., 1DCNN [9], naive Bayesian model [39], and support vector machine [40]) are also able to be extended as the base classifier. In the proposed method, the RF classifier is adopted as a base classifier, but it is not the only feasible choice. In fact, under the framework of the proposed method, we can choose the most suitable classifier according to the specific problem settings. Furthermore, we performed a small set of experiments replacing the classifier from RF to 1DCNN to verify this statement.

D. Theoretical Analysis of Multi-Training

The main motivation of the proposed multi-training method is similar to that of ensemble learning, whereas it is slightly different from the general idea of ensemble learning. Ensemble learning combines all the classifiers' results on the data and then improves the final result by voting. We used the idea of ensemble learning to provide higher confidence for the updating of unlabeled samples based on the SSL framework. The data's similarity between different phases makes the base classifier under each phase have more reference data to make predictions, and the data's divergence provides the model with more disagreements. In the process of determining the updating samples, the discriminant equation of the algorithm adopts the idea of harmonic mean's definition so that the final joint confidence considers the prediction results of all classifiers. However, we only consider the confidence value as an updating standard, rather than directly assigning labels to the unlabeled data. Until all the samples have completed the joint confidence computing, the training set will be updated. Since this updating process considers the image information of the whole period, the prediction of each pixel is more robust than ensemble learning. Therefore, the spatial and temporal consistency of the classification results is ensured by the proposed method.

E. Process of Multi-Training

The multi-training process is presented in Algorithm 1 with pseudocodes. There are some details listed in the following text that need to be considered, which will affect the final results.

1) To test the accuracy and the stability of the proposed method, we need to randomly select the examples many times. The training set and the test set should be split before repeated sampling to ensure that the training data can only sample from the precut training set. The model cannot have access to the test set in advance.
2) To compare the performance between different methods, we need to fix the random seeds to guarantee that different methods share the same training set.
3) When updating the unlabeled samples into the training set, the samples whose confidence is higher than the threshold should be randomly selected. Randomized updating makes the training process more generalized.

III. Study Area and Data

A. Study Area

The study area is in Nanjing City, the capital of Jiangsu Province, China. Nanjing is situated in the lower reaches of the Yangtze River Plain. Located in the subtropical monsoon climate zone, it is greatly different in vegetation between seasons, which provides powerful support for multitemporal training. As the national ecological garden city, Nanjing is the most forested city in Jiangsu Province, which is an ideal study area for recognition of urban green cover. Thus, we chose Nanjing as the study area and carried out the study in the center of Nanjing that includes the vegetation in the parks, on the mountains, and along the streets.

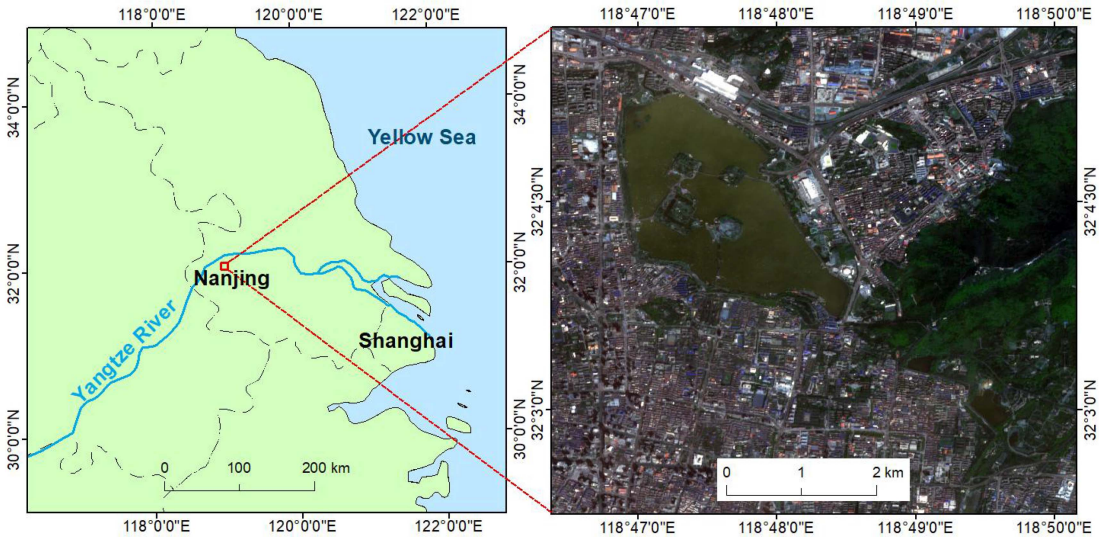


Fig. 5. Location of the study area (left) and the Sentinel-2 MSI image (right) with true color composite acquired on August 15, 2016.

TABLE I
ACQUIRED DATE OF THE EIGHT SENTINEL-2 MSI IMAGES

No.	Date	No.	Date
1	February 11, 2017	5	March 13, 2018
2	April 2, 2017	6	June 11, 2018
3	October 9, 2017	7	September 9, 2018
4	December 18, 2017	8	January 22, 2019

B. Data

This study used Sentinel-2 multispectral instrument (MSI) images (obtained from google earth engine) of Xuanwu Lake area in Nanjing, as shown in Fig. 5. Sentinel-2 is a wide swath multispectral imaging mission supporting land monitoring studies, including the monitoring of vegetation, soil, and urban area, as well as the observation of inland waterways and coastal areas. The image contains 772 × 653 pixels, and each has four spectral bands with a spatial resolution of 10 m. The four spectral bands are blue (0.46–0.52 μm), green (0.54–0.58 μm), red (0.50–0.80 μm), and near infrared (0.78–0.90 μm). We selected eight images from February 11, 2017 to January 22, 2019, with roughly every three months an image, as shown in Table I. The images selected from different months were to verify the proposed method's ability to overcome the identification challenge caused by the difference of vegetation phenology, and the images chosen from different years were to test its performance for annual vegetation information updating.

To augment the existing data and classify green cover efficiently, we adopted some commonly used radiometric indices based on the availability. The selected nine indices are shown in Table II. We sampled from the data and calculated the correlation degrees between the features, as shown in Fig. 6. For two indices with large correlation coefficients, keep one of them to remove redundancy. The five retained indices (NDVI, MSAVI, MTVI, SAVI, and VARI) and the original four spectral bands were combined to construct the features of the images.

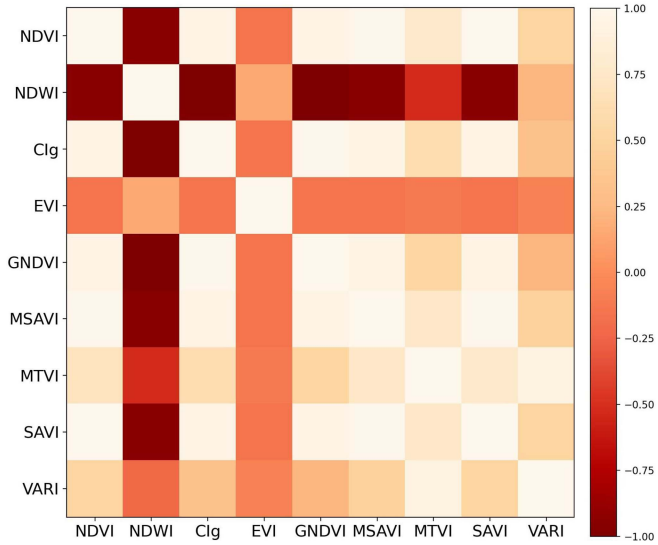


Fig. 6. Hot map of correlation degree between the features.

The classification scheme of the study area includes six classes: tree, low vegetation, water, road, building, and shadow. For more refined vegetation extraction, we divided the vegetation into finer divisions, referring to the ISPRS dataset Vaihingen. A relatively rough classification scheme was used for other feature categories in which shadow represents a part of the earth's surface that is blocked from sunlight by buildings. It is noted that the classes other than vegetation were merged in the final results.

In each phase, we manually sampled the whole images, and 1000 labeled pixels were obtained for each category. The labeled samples were different in different phases. We further divided the labeled samples into training set and test set according to the 1:4 ratio, as shown in Table III. More precisely, the training set here should be named training pool for the reason that to test the efficiency of the proposed method, we only sampled a very

TABLE II
LIST OF RADIOMETRIC INDICES SELECTED AS FEATURES

Radiometric index	Equation
Normalized Difference Vegetation Index [41]	$NDVI = \frac{NIR - R}{NIR + R}$
Normalized Difference Water Index [42]	$NDWI = \frac{G - NIR}{G + NIR}$
Green Chlorophyll Index [43]	$CIg = \frac{NIR}{G} - 1$
Enhanced Vegetation Index [44]	$EVI = \frac{2.5(NIR - R)}{NIR + 6R - 7.5B + 1}$
Green Normalized Difference Vegetation Index [45]	$GNDVI = \frac{NIR - G}{NIR + G}$
Modified Triangular Vegetation Index [46]	$MTVI = \frac{1.5(1.2(NIR - G) - 2.5(R - G))}{\sqrt{(2NIR + 1)^2 - (6NIR - 5\sqrt{R}) - 0.5}}$
Soil Adjusted Vegetation Index [47]	$SAVI = \frac{NIR - R}{NIR + R + L}(L + 1)$
Modified Soil Adjusted Vegetation Index [48]	$MSAVI = 0.5(2(NIR + 1) - 2\sqrt{(2NIR + 1)^2 - 8(NIR - R)})$
Visible Atmospherically Resistant Index [49]	$VARI = \frac{G - R}{G + R - B}$

TABLE III
NUMBER OF SAMPLES ON EACH CLASS IN TRAINING AND TEST SETS

Class	Training set	Test set
Vegetation	200	800
Shadow	200	800
Water	200	800
Road	200	800
Building	200	800
Total	1000	4000

TABLE IV
PARAMETERS USED IN THE EXPERIMENTS

Parameter	Definition
N_t	Number of phases combined
N_l	Number of labeled samples for each class initially
N_u	Number of updating unlabeled samples for each class in each epoch
N_{iter}	Number of iterations of multi-training

TABLE V
PARAMETER SETTING OF EACH EXPERIMENT

Experiment	N_l	N_u	N_t
General performance comparison	1	5	4
Updating of unlabeled samples	1	5	4
Selection of parameter N_l	1:2:19	5	4
Selection of parameter N_u	1:2:19	5:5:50	4

small number of samples from it independently with a fixed size to form the training set during the training process in the repeated experiments. The unlabeled sample set used in the experiments is the whole pixels from images, excluding the labeled part.

IV. EXPERIMENTAL DESIGN

A. Validation Metric

In our experiments, F1-score was selected as the validation metric because it requires both high-precision and recall values and is defined as follows:

$$precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2}$$

$$recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} \tag{4}$$

where TP (true positive) is the number of pixels which are classified correctly; FP (false positive) is the number in which negatives are misidentified as positives; and FN (false negative) is the number that positives are misidentified as negatives.

B. Experimental Setup

We designed five experiments to evaluate the performance of the proposed method as well as the updating process of the unlabeled data and the effect of labeled and unlabeled samples' selection. The average F1-score and the standard deviations (SDs) under several independently repeated experiments are used to represent the accuracy of the classification task and the stability of the proposed method under different phases, respectively. Some necessary parameters and their explanations are listed in Tables IV and V.

1) Comparison of Different Methods: The first experiment was designed to compare the performances of different methods. We tested the proposed method and three other representative methods, i.e., self-training [50], co-training [31], tri-training [53], and super-training. Self-training method is an SSL method applied on each phase independently [50]. It aims to update the high-confidence samples into the training set through each training round. In the experiment, the classifiers trained from self-training predicted the probability of unlabeled samples, and these probabilities generated the mean value as the threshold of sample selection. Randomly selected a fixed number of samples greater than the threshold to join the sample set. In the experiment of co-training, the threshold was obtained in the same way as self-training. If the probability of an unlabeled pixel was greater than its respective threshold in both phases, the pixel was

added to the alternative set. Then, a fixed number of samples were randomly selected in the alternative set to join the sample set. Tri-Training is the semi-supervised paradigm used by trinet. In the experiment of tri-training, three models were used. If two models predict a sample consistently then the sample is given to the third model for training and the prediction by the two models is used as a pseudolabel. Super-Training is a supervised learning method with five times as many labeled samples as the other semi-supervised methods. For each iteration, super-training selects samples from the labeled sample set, unlike the semi-supervised methods that select unlabeled samples. To be fair, these methods also used RF as the classifier, and the parameter settings were the same as those of multi-training. These methods were compared with the proposed methods from different training mechanisms and different labeled samples.

Self-training was trained in a single phase, and its prediction was also predicted in a single phase. The overall evaluation of self-training is to take the average value of accuracy in each phase. Co-Training was conducted in two phases. In the case of multiple phases, the co-training accuracy of a certain phase was the average of the accuracies of the models trained by the combination of this phase and all other phases. For example, there are four phases T1, T2, T3, and T4. To calculate the accuracy under T1 phase, it is the mean value of the accuracy of the three models trained by T1&T2, T1&T3, and T1&T4. The overall evaluation of co-training is to take the average value of accuracy in each phase. Tri-Training was conducted in three phases. Similar to co-training, the accuracy of a certain phase was the average of the accuracies of the models trained by the combination of this phase and other two phases.

2) Effect of N_l and N_u: In the second experiment, we compared the effects of different numbers of labeled and unlabeled samples on the model's accuracy to find the optimal parameters of the proposed method. Since the optimal N_l is related to N_u [32], we simultaneously changed N_l and N_u for each class before training to analyze the parameters' sensitivity. In this experiment, N_l ranges from 1 to 19 with step 2, and N_u varies from 5 to 50 with step 5. We recorded the average F1-score and SDs on each parameter setting. We also compared the co-training algorithm's performance on the same parameter settings.

3) Consistency of Multitemporal Classification Results: Classification of multitemporal remote sensing images always faces the difficulty that the results are inconsistent on spatial and temporal. Therefore, the classification consistency is a crucial target to test the performance of multitemporal land cover classification algorithms. In this experiment, we modified the original prediction difference of the classifiers (PDC) in the previous work [31] to obtain a new consistency measurement on a multitemporal setting as follows:

$$PDC = 1 - \frac{1}{N}\sum_{j=1}^{N} I\left(\varphi(\{f_i\left(x_j^i\right)\}_{i=1}^{N_t}) > \frac{N_t}{2}\right) \quad (5)$$

where $f_i(x_j^i)$ denotes the predicted label of each classifier on each pixel; $\varphi(\cdot)$ returns the maximum number of inputs whose values are the same; and $I(\cdot)$ denotes the indicator function that returns 1 if the input is true else 0. We compared the PDC of multi-training and co-training on the parameter setting in which N_l ranged from 1 to 9 with a transition at 2, N_u ranged from 5 to 50 with a transition at 5, and N_t was fixed as 4.

4) 1DCNN as the Classifier: To further verify the applicability of the proposed semi-supervised method, the classifier was replaced from RF to 1DCNN [9] for an experiment. The experimental conditions were the same as in Experiment 1 except that the classifier was replaced. Still, the accuracies of the proposed method, self-training, co-training, tri-training, and super-training were compared. In this experiment, due to the small size of the samples, we designed a lightweight 1DCNN, consisting of three convolutional layers, three Relu layers, and one fully connected layer. To increase the dimensionality of the input data, we used five indices in addition to the four spectral bands, just as in the previous experiments. Two scenes of ISPRS Vaihingen images were used for pretraining. To enable the 1DCNN to converge during the initial training, five samples per class were used instead of one sample per class as in the previous experiments.

5) Effect of Phases: To explore the best applicable conditions of the proposed method, we designed an experiment to evaluate the influence of N_t and phases' combinations on the improvements of classification accuracy and stability by the proposed method. The number of phases should be considered as a key factor influencing the experimental results for the reason that the proposed method is designed for multitemporal remote sensing images. N_t was changed from 3 to 8 in this experiment to explore the changing trend of F1-score and SDs. And we selected every four phases from all the images as the inputs of the multi-training algorithm to test the accuracy via different phases' combination.

V. Results

A. Comparison of Different Methods' Classification Results

To visualize the results of vegetation recognition and analyze them qualitatively, we depicted the classification results under four phases and highlighted the area of vegetation recognition in Fig. 7.

In the aspect of visualization, it is clear that the vegetation extracted by the proposed method is more delicate and accurate. Meanwhile, there is a clear distinction between the spatial location of the tree and the low vegetation. Although the training process is based on the pixels, the final results are still very consistent in the area with a wide coverage of vegetation. In contrast, the vegetation area extracted by the self-training method on a single phase is relatively broken. In the urban blocks, it is apparent that most of the vegetation distributed along the road were recognized by the multi-training algorithm. Thus, we can conclude that the proposed method has the ability to extract the vegetation with great consistency during long period. The results in Fig. 7 can also reflect the influence of seasonal changes on vegetation. For example, in mountain areas, the recognized vegetation is sparser in winter than in summer. From the results, we can see that the proposed method can even obtain similar classification results as supervised training algorithm, which is trained on a large-scale labeled training set.

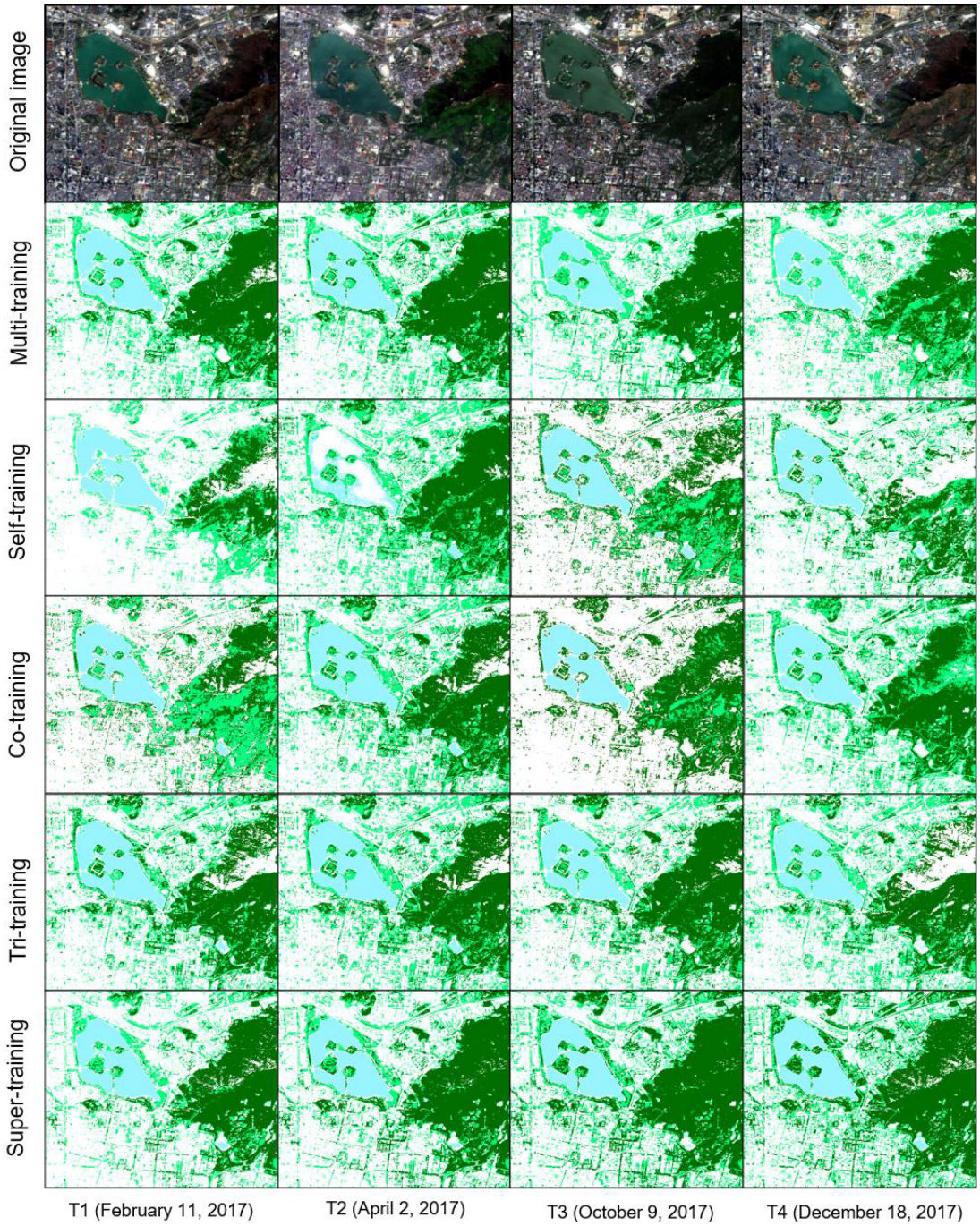


Fig. 7. Original images and classification results in the case $N_l = 1$ and $N_u = 5$. The images on the first row are the original Sentinal-2 images of the study area under four phases, and the next five rows below are multi-training, self-training, co-training, tri-training, and super-training's classification results. Dark green, light green, and blue represent tree, low vegetation, and water.

To further illustrate the efficiency of the multi-training algorithm, we implemented a quantitative comparison on the performance of the proposed method and other aforementioned classification methods testing on the multitemporal dataset. The average F1-score of the four phases for the six classes and the average value of all classes are shown in Table VI. In the contrast experiments, N_l was set as 1 and N_u was set as 5. The random seed was fixed to ensure that different methods shared the same training set. Table VI indicates that the multi-training algorithm can significantly improve the recognition accuracy of tree, low vegetation, and shadow from 0.762 to 0.832, from 0.652 to 0.728, and from 0.698 to 0.801, respectively. A possible explanation is that the distribution of vegetation and shadow is quite different in multiple phases due to phenological influence and seasonal alternation. Therefore, the classifiers can capture more additional information about these classes than other land covers from different views, which increases the confidence of determining the true class of the unlabeled samples. With

TABLE VI
AVERAGE F1-SCORE IN THE CASE $N_l = 1$, $N_u = 5$, AND RF AS CLASSIFIER

Class	Multitraining			Self-training			Cotraining			Tritraining			Supertraining		
	Initial	Final	Improved	Initial	Final	Improved	Initial	Final	Improved	Initial	Final	Improved	Initial	Final	Improved
Tree	0.762	0.832	**9.19%**	0.752	0.762	1.33%	0.751	0.767	2.13%	0.756	0.779	3.04%	0.926	0.965	4.21%
Low Vegetation	0.652	0.728	**11.66%**	0.647	0.654	1.08%	0.648	0.668	3.09%	0.653	0.672	2.91%	0.897	0.912	1.67%
Shadow	0.698	0.801	**14.76%**	0.690	0.722	4.64%	0.679	0.724	6.63%	0.680	0.731	7.50%	0.912	0.968	6.14%
Water	0.801	0.846	5.62%	0.787	0.81	2.92%	0.782	0.839	**7.29%**	0.789	0.822	4.18%	0.931	0.974	4.62%
Road	0.651	0.703	7.99%	0.625	0.612	–2.08%	0.622	0.634	1.93%	0.630	0.643	2.06%	0.812	0.906	**11.58%**
Building	0.462	0.512	10.82%	0.419	0.451	7.64%	0.425	0.472	11.06%	0.422	0.457	8.29	0.726	0.884	**21.76%**
Average	0.671	0.737	**9.84%**	0.653	0.668	2.30%	0.651	0.684	5.07%	0.655	0.684	4.42%	0.862	0.935	8.47%

The bold highlights the biggest improved accuracy among five methods per row.

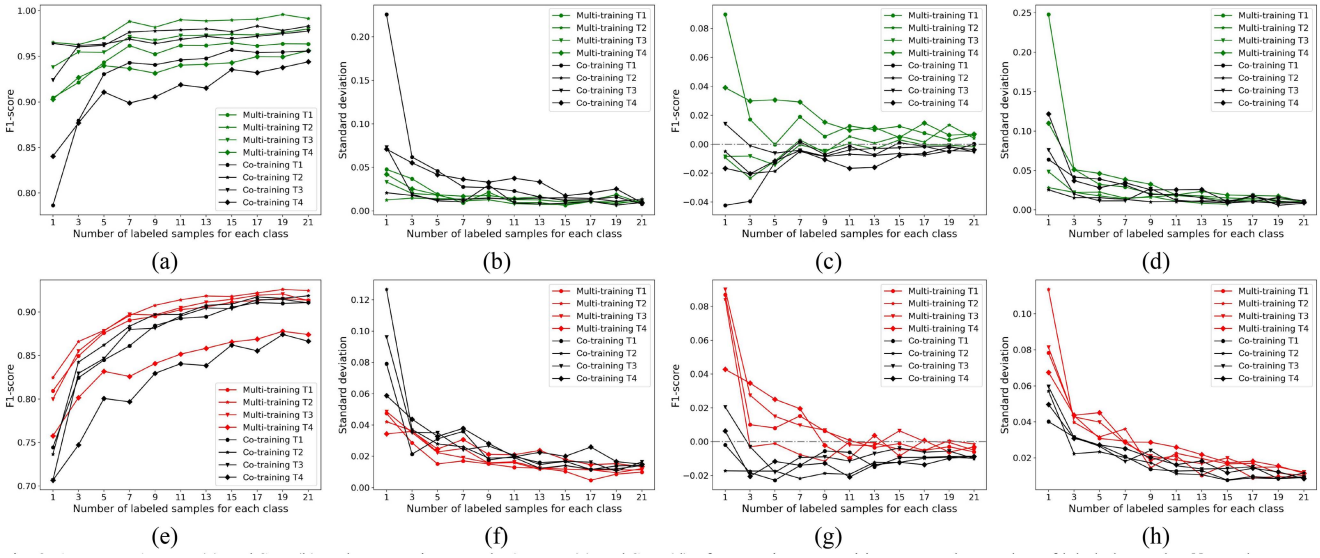


Fig. 8. (a) Average F1-score, (b) SDs, (c) average improved F1-score, and (d) SDs of vegetation recognition versus the number of labeled samples N_l. (e) Average F1-score, (f) SDs, (g) average improved F1-score and (h) SDs of total recognition accuracy of all classes versus the number of labeled samples N_l.

abundant information, the model has the ability to add the samples with more accurate pseudolabels into the training set during updating. The water's classification accuracy improves most by co-training method, from 0.782 to 0.839, an increase of 7.29% units. This is because the co-training algorithm was mainly aimed at the coupdating between the two phases; thus, it has good improvement on the water area that changes little over the whole period. Generally, the identification of roads and buildings is a quite difficult task, especially when the training samples are not enough. Therefore, with a large training set, the supervised learning method greatly improves the classification accuracy in these two classes. From the results in Table VI, all the land cover's prediction accuracy of multi-training algorithm has improved obviously and the average F1-score improvement outperformed the other compared methods. This shows that the multi-training method has significant applicability, and thus, it can be used in the recognition of land cover when labeled samples are limited.

B. *Influence of N_l and N_u*

We analyzed different N_l's effects on the average F1-score and SDs of the final results and improved the accuracy on the vegetation and all classes in Fig. 8. From Fig. 8(a) and (e), by changing N_l in the initial training set, we can find the changing trend of the final classification accuracy. It is easy to understand that both the proposed method and co-training can improve the final classification accuracy as N_l increases. Especially when N_l is increased from 1 to 5 in each class, the final classification accuracy of the model is improved significantly. In all cases, the average F1-scores of the proposed method are higher than those of the co-training method, which shows that the proposed method can improve the accuracy of land cover classification in multitemporal data. In Fig. 8(b) and (f), as N_l increases, the SDs of the classification accuracy decrease gradually and the SDs of the proposed method are lower than that of the co-training. This is because more labeled samples can help the model to classify better and obtain more stable results. Through the communication step between all classifiers, the proposed method can learn the difference on different phases and stabilize quickly. In Fig. 8(c), (d), (g), and (h), it can be seen that the average F1-scores and SDs of the improvement between the initial and the final accuracy decrease as the N_l decreases. It may be that if the training set contains enough labeled samples, the model has already learned very well on this initial dataset, which causes the model's ability to be restricted and cannot learn more useful

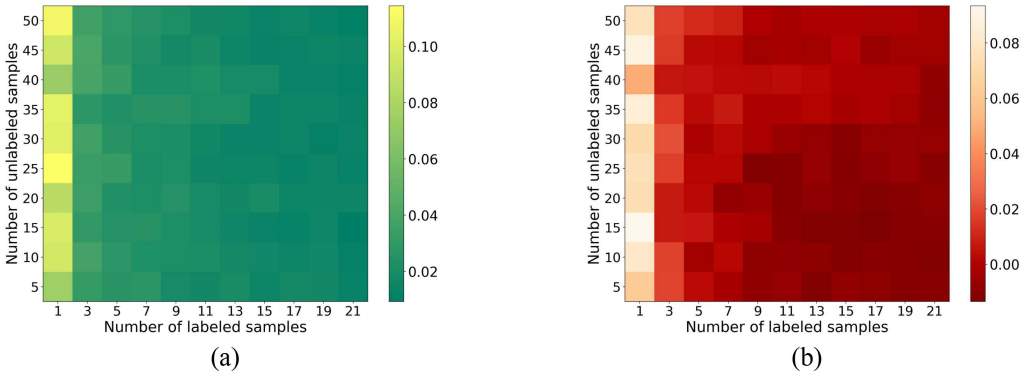


Fig. 9. Impact of different N_l and N_u values on the results. (a) Represents the improved F1-scores based on different N_l and N_u. (b) Represents the improved corresponding SDs. The horizontal axis represents N_l and the vertical axis represents N_u.

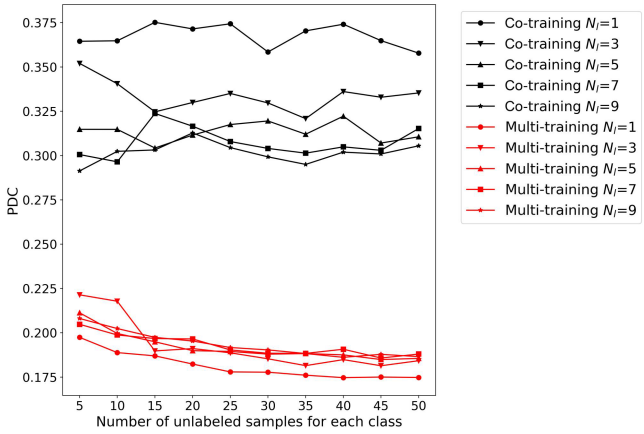


Fig. 10. Average prediction difference of the classifiers (PDC) of co-training and multi-training under four phases.

information in the following updating process. The large gap between $N_l = 1$ and $N_l = 5$ in these four curves strongly supports this explanation. In general, the proposed method shows significant improvements in accuracy and stability over the co-training.

The improvements of the average F1-scores and SDs based on different N_l and N_u are shown in Fig. 9. We can clearly see that, for a small N_l, fewer unlabeled samples result in a higher F1-score and lower SDs. However, when N_l is large enough, the result is reversed, and more unlabeled samples are needed to train a model with both high accuracy and stability. The results in this experiment corroborate the findings of previous work [31]. Therefore, a small N_u is suggested when the training set only includes very few labeled samples, while a large N_u is suitable if sufficient labeled samples are in the dataset.

C. Consistency of Multitemporal Classification Results

To further illustrate the consistency of the proposed method, we compared the PDC of multi-training and co-training under four phases. In each pair of this comparison experiment, we fixed the number of labeled samples and justified the unlabeled samples' number. In Fig. 10, it is clear that the PDC of co-training is obviously higher than that of multi-training and does not show any tendency with increasing N_u. This is because, in multi-training, classifiers can communicate with each other to select high-confidence unlabeled training samples into a new training set, which provides more opportunity to maintain consistency between different phases. The curves of multi-training in Fig. 10 reflect a clear pattern via the number of unlabeled samples that the PDC values decrease significantly with the increase of N_u. The possible reason is that with the increase of N_u, there are more high-confidence samples in the training set, which will further enhance the prediction consistency of all the classifiers in the next iteration. It results in the low PDC values, given large number of unlabeled samples for each class during training, which provides an instructional parameter setting method for future applications.

D. Classification Accuracy With 1DCNN as Classifier

The initial accuracy of each method is improved in the case of using 1DCNN as a classifier based on five labeled samples, as shown in Table VII. Multi-Training has a significant increase in the final accuracy after SSL, and the improvement is significantly higher than that of the other semi-supervised methods, which further proves the effectiveness and applicability of multi-training. In addition, the improved accuracy of the building is large, up to 18.03%, showing that the feature extraction ability of 1DCNN for complex features can help multi-training perform better.

VI. Discussion

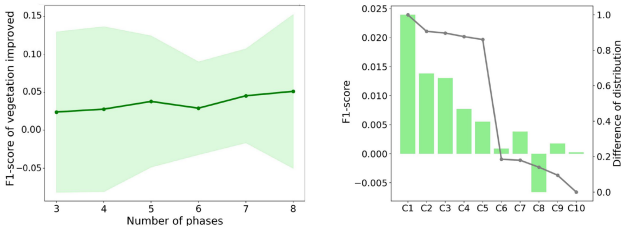
A. Effect of Phases

In Fig. 11(a), with increasing of N_t, the improvement of the average F1-scores of vegetation increases obviously, from 0.02 under $N_t = 3$ to 0.05 under $N_t = 8$. However, whether this trend can be maintained when N_t is larger needs further research. The increase of N_t not only make the model more accurate but also make it more robust, especially from 3 to 7, which is beneficial to determine the unchanged area and selection of unlabeled samples when updating the training set. However, if N_t is too large, i.e., the images are acquired in a long period, the SDs of the model's accuracy will be large although F1-score is

TABLE VII
AVERAGE F1-SCORE IN THE CASE $N_l = 5$, $N_u = 5$, AND 1DCNN AS CLASSIFIER

Class	Multitraining			Self-training			Cotraining			Tritraining			Supertraining		
	Initial	Final	Improved	Initial	Final	Improved	Initial	Final	Improved	Initial	Final	Improved	Initial	Final	Improved
Tree	0.851	0.912	**7.16%**	0.831	0.842	1.32%	0.833	0.865	3.84%	0.839	0.873	4.05%	0.931	0.972	4.40%
Low Vegetation	0.822	0.901	**9.61%**	0.816	0.827	1.34%	0.817	0.847	3.67%	0.808	0.835	3.34%	0.902	0.946	4.87%
Shadow	0.878	0.923	**5.12%**	0.865	0.844	−2.42%	0.867	0.889	2.53%	0.875	0.892	1.94%	0.945	0.973	2.96%
Water	0.901	0.963	**6.88%**	0.883	0.878	−0.56%	0.889	0.921	3.60%	0.896	0.926	3.34%	0.951	0.976	2.62%
Road	0.811	0.876	8.01%	0.801	0.787	−1.74%	0.805	0.826	2.61%	0.797	0.815	2.25%	0.845	0.924	**9.34%**
Building	0.732	0.864	**18.03%**	0.726	0.705	−2.89%	0.723	0.765	5.81%	0.730	0.757	3.69%	0.827	0.907	9.67%
Average	0.832	0.906	**8.89%**	0.820	0.813	−0.79%	0.822	0.852	3.62%	0.824	0.849	3.09%	0.900	0.949	5.49%

The bold highlights the biggest improved accuracy among five methods per row.



(a) (b)

Fig. 11. (a) Effect of N_t on the improvements of average F1-score of vegetation. The solid line represents the average F1-score and the light-shaded region represents the SD. (b) Improved F1-score and difference of data distribution versus the combination of different phases.

still increasing, as shown in the area where $N_t \geq 7$ in Fig. 11(a). It seems possible that the joint confidence value is not credible when N_t is too large, which causes the updating process to no longer be stable.

To explore the effect of phases' combinations, we calculate the difference of data distribution for all phases. Because the proposed method is for multitemporal training, only computing the difference between two phases cannot truly represent the difference of the phases in the combination. Therefore, we use the following equation to calculate the difference in the whole combination:

$$Diff\,(T_1,\,T_2,\ldots,\,T_k) = \frac{1}{N}\left(\sum_{f=1}^{N}\left(\frac{1}{f}\sum_{i=1}^{N} std\left(x_1^{(f)},\,x_2^{(f)},\,\ldots,\,x_k^{(f)}\right)\right)\right) \tag{6}$$

where T_i denotes the *i*th phase and $x_i^{(f)}$ denotes the *f*th band feature of data in the *i*th phase.

The result calculated by the equation above can represent the difference among multiple phases and satisfies the setting of the problem. We chose the five-phase combinations with the largest distribution difference and the five-phase combinations with the smallest distribution difference as the inputs of the proposed method. Our goal is to test under which conditions above the proposed method can obtain better performance on the land cover classification of multitemporal remote sensing images.

As shown in Fig. 11(b), there is a high correlation between the difference in data distribution and the improvement of the vegetation recognition accuracy. The large difference in the phase combination can help the model to obtain a great improvement of accuracy. The difference between phases makes the divergence of classifiers expand so that the co-training can be carried out better [27]. And in the communication stage, larger differences can make the model learn more information. In the process of iterative training, the classifiers will become more and more similar, resulting in the failure of co-training and even the decline of accuracy [52]. If the difference between phases is small, this problem will be more serious. It can be seen in Fig. 11(b) that the improved accuracy of the model trained on phase combination with small difference is small even negative.

B. Limitations and Future Improvements

In this study, we propose a novel semi-supervised method for time-series urban green cover recognition by utilizing the differences between images of each phase, which avoids the tedious sample labelling and alleviates the problems caused by data shifting. However, in our experiments, we used RF as the classifier and performed a feature selection and deredundancy process, which slightly increased the workload. In other application scenarios, deep learning models that can automatically extract features can be used as classifiers, which can make the process simpler.

Additionally, even though a safe method is used to assign the high-confidence samples pseudolabels, there is no guarantee that all pseudolabels are correct. Incorrect pseudolabels may aggravate the error during iteration leading to a decrease in classification accuracy. An appropriate noise-tolerant learning method may be a solution for this.

VII. CONCLUSION

In this study, a multi-training method for multitemporal remote sensing images is designed to recognize urban green cover. The study demonstrates that the proposed method has higher accuracy and consistency in vegetation recognition than the previous methods. In the case of few samples, the proposed method performs better and has more advantages. Furthermore, the difference between the observed spectral distributions of multitemporal data is exploited in a mutual learning way, which

provides a new strategy to make use of multitemporal images. For better model accuracy, N_u and N_l should match in size. In addition, the large differences between multitemporal data improve the model's classification performance. Although this study mainly focuses on the recognition of urban vegetation, the proposed method can be easily migrated and applied to land cover classification and updating of land cover maps in time series.

REFERENCES

[1] P. Bolund and S. Hunhammar, "Ecosystem services in urban areas," *Ecol. Econ.*, vol. 29, no. 2, pp. 293–301, May 1999, doi: 10.1016/S0921-8009(99)00013-0.

[2] D. Wen, X. Huang, H. Liu, W. Liao, and L. Zhang, "Semantic classification of urban trees using very high resolution satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1413–1424, Apr. 2017, doi: 10.1109/JSTARS.2016.2645798.

[3] A. M. Dewan and Y. Yamaguchi, "Using remote sensing and GIS to detect and monitor land use and land cover change in Dhaka metropolitan of Bangladesh during 1960–2005," *Environ. Monit. Assessment*, vol. 150, no. 1, 2009, Art. no. 237.

[4] T. Schmidt, C. Schuster, B. Kleinschmit, and M. Förster, "Evaluating an intra-annual time series for grassland classification—How many acquisitions and what seasonal origin are optimal?," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3428–3439, Aug. 2014.

[5] A. Sharma, X. Liu, and X. Yang, "Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks," *Neural Netw.*, vol. 105, pp. 346–355, 2018.

[6] D. H. T. Minh et al., "Deep recurrent neural networks for winter vegetation quality mapping via multitemporal SAR Sentinel-1," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 464–468, Mar. 2018, doi: 10.1109/LGRS.2018.2794581.

[7] M. RuBwurm and M. Korner, "Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 1496–1504, doi: 10.1109/CVPRW.2017.193.

[8] X. Jia et al., "Incremental dual-memory LSTM in land cover prediction," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining.*, 2017, pp. 867–876.

[9] L. Zhong, L. Hu, and H. Zhou, "Deep learning based multi-temporal crop classification," *Remote Sens. Environ.*, vol. 221, pp. 430–443, Feb. 2019, doi: 10.1016/j.rse.2018.11.032.

[10] S. Ji, C. Zhang, A. Xu, Y. Shi, and Y. Duan, "3D convolutional neural networks for crop classification with multi-temporal remote sensing images," *Remote Sens.*, vol. 10, no. 1, 2018, Art. no. 75.

[11] M. Rußwurm and M. Körner, "Multi-temporal land cover classification with sequential recurrent encoders," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 4, Mar. 2018, Art. no. 129, doi: 10.3390/ijgi7040129.

[12] R. Interdonato, D. Ienco, R. Gaetano, and K. Ose, "DuPLO: A DUal view point deep learning architecture for time series classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 149, pp. 91–104, Mar. 2019, doi: 10.1016/j.isprsjprs.2019.01.011.

[13] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, 2016, Art. no. 9.

[14] Y. Guo, X. Jia, and D. Paull, "A domain-transfer support vector machine for multi-temporal remote sensing imagery classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 2215–2218.

[15] H. Lyu and H. Lu, "A deep information based transfer learning method to detect annual urban dynamics of Beijing and Newyork from 1984–2016," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 1958–1961.

[16] Z. Wang, H. Zhang, W. He, and L. Zhang, "Phenology alignment network: A novel framework for cross-regional time series crop classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2021, pp. 2934–2943, doi: 10.1109/CVPRW53098.2021.00329.

[17] B. Banerjee and K. M. Buddhiraju, "A novel semi-supervised land cover classification technique of remotely sensed images," *J. Indian Soc. Remote Sens.*, vol. 43, no. 4, pp. 719–728, Dec. 2015, doi: 10.1007/s12524-014-0370-z.

[18] X. Zhu, *Semi-Supervised Learning Literature Survey*. Madison, WI, USA: Univ. Wisconsin-Madison, 2008.

[19] B. M. Shahshahani and D. A. Landgrebe, "The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 5, pp. 1087–1095, Sep. 1994, doi: 10.1109/36.312897.

[20] T. Joachims, "Transductive inference for text classification using support vector machines," in *Proc. 16th Int. Conf. Mach. Learn.*, 1999, vol. 99, pp. 200–209.

[21] M. I. Jordan, *Learning in Graphical Models*. Cambridge, MA, USA: MIT Press, 1999.

[22] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. 11th Annu. Conf. Comput. Learn. Theory*, Madison, WI, USA, 1998, pp. 92–100, doi: 10.1145/279943.279962.

[23] E. L. Pencue-Fierro, Y. T. Solano-Correa, J. C. Corrales-Muñoz, and A. Figueroa-Casas, "A semi-supervised hybrid approach for multitemporal multi-region multisensor landsat data classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5424–5435, Dec. 2016.

[24] D. Hong, N. Yokoya, G.-S. Xia, J. Chanussot, and X. X. Zhu, "X-ModalNet: A semi-supervised deep cross-modal network for classification of remote sensing data," *ISPRS J. Photogramm. Remote Sens.*, vol. 167, pp. 12–23, Sep. 2020, doi: 10.1016/j.isprsjprs.2020.06.014.

[25] W. Han, R. Feng, L. Wang, and Y. Cheng, "A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 23–43, Nov. 2018, doi: 10.1016/j.isprsjprs.2017.11.004.

[26] G. Camps-Valls, T. V. B. Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007, doi: 10.1109/TGRS.2007.895416.

[27] W. Wang and Z.-H. Zhou, "Analyzing co-training style algorithms," in *Proc. Eur. Conf. Mach. Learn.*, 2007, pp. 454–465, doi: 10.1007/978-3-540-74958-5_42.

[28] W. Wang and Z.-H. Zhou, "A new analysis of co-training," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 1135–1142.

[29] D.-D. Chen, W. Wang, W. Gao, and Z.-H. Zhou, "Tri-net for semi-supervised deep learning," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Stockholm, Sweden, 2018, pp. 2014–2020, doi: 10.24963/ijcai.2018/278.

[30] L. Breiman, "Randomizing outputs to increase prediction accuracy," *Mach. Learn.*, vol. 40, no. 3, pp. 229–242, 2000.

[31] L. Zhu, P. Xiao, X. Feng, X. Zhang, Y. Huang, and C. Li, "A co-training, mutual learning approach towards mapping snow cover from multi-temporal high-spatial resolution satellite imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 122, pp. 179–191, Dec. 2016, doi: 10.1016/j.isprsjprs.2016.11.003.

[32] A. D'Addabbo, G. Satalino, G. Pasquariello, and P. Blonda, "Three different unsupervised methods for change detection: An application," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2004, vol. 3, pp. 1980–1983.

[33] D. T. Nguyen, C. K. Mummadi, T. P. N. Ngo, T. H. P. Nguyen, L. Beggel, and T. Brox, "SELF: Learning to filter noisy labels with self-ensembling," in *Proc. Int. Conf. Learn. Representations*, 2020, pp. 1–15.

[34] E. Arazo, D. Ortego, P. Albert, N. E. O'Connor, and K. McGuinness, "Pseudo-labeling and confirmation bias in deep semi-supervised learning," in *Proc. Int. Joint Conf. Neural Netw.*, Glasgow, U.K., 2020, pp. 1–8, doi: 10.1109/IJCNN48605.2020.9207304.

[35] S. M. Abdolrassoul and B. J. Turner, "A comparison of four common atmospheric correction methods," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 4, pp. 361–368, Apr. 2007.

[36] R. Bindschadler et al., "The landsat image mosaic of Antarctica," *Remote Sens. Environ.*, vol. 112, no. 12, pp. 4214–4226, Dec. 2008, doi: 10.1016/j.rse.2008.07.006.

[37] P. Bartlett, Y. Freund, W. S. Lee, and R. E. Schapire, "Boosting the margin: A new explanation for the effectiveness of voting methods," *Ann. Statist.*, vol. 26, no. 5, pp. 1651–1686, Oct. 1998, doi: 10.1214/aos/1024691352.

[38] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[39] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[40] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2011, Art. no. 27.

[41] T. N. Carlson and D. A. Ripley, "On the relation between NDVI, fractional vegetation cover, and leaf area index," *Remote Sens. Environ.*, vol. 62, no. 3, pp. 241–252, 1997.

[42] B.-C. Gao, "Normalized difference water index for remote sensing of vegetation liquid water from space," *Proc. SPIE*, vol. 2480, 1995, pp. 225–236.
[43] A. A. Gitelson, Y. Gritz, and M. N. Merzlyak, "Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves," *J. Plant Physiol.*, vol. 160, no. 3, pp. 271–282, 2003.
[44] Z. Jiang, A. R. Huete, K. Didan, and T. Miura, "Development of a two-band enhanced vegetation index without a blue band," *Remote Sens. Environ.*, vol. 112, no. 10, pp. 3833–3845, 2008.
[45] A. A. Gitelson, Y. J. Kaufman, and M. N. Merzlyak, "Use of a green channel in remote sensing of global vegetation from EOS-MODIS," *Remote Sens. Environ.*, vol. 58, no. 3, pp. 289–298, 1996.
[46] P. J. Zarco-Tejada, J. R. Miller, T. L. Noland, G. H. Mohammed, and P. H. Sampson, "Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1491–1507, Jul. 2001.
[47] A. R. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sens. Environ.*, vol. 25, no. 3, pp. 295–309, Aug. 1988, doi: 10.1016/0034-4257(88)90106-X.
[48] J. Qi, A. Chehbouni, A. R. Huete, Y. H. Kerr, and S. Sorooshian, "A modified soil adjusted vegetation index," *Remote Sens. Environ.*, vol. 48, no. 2, pp. 119–126, May 1994, doi: 10.1016/0034-4257(94)90134-1.
[49] A. Gitelson, R. Stark, U. Grits, D. C. Rundquist, Y. Kaufman, and D. Derry, "Vegetation and soil lines in visible spectral space: A concept and technique for remote estimation of vegetation fraction," *Int. J. Remote Sens.*, vol. 23, no. 13, pp. 2537–2562, 2002.
[50] C. Rosenberg, M. Hebert, and H. Schneiderman, "Semi-supervised self-training of object detection models," in *Proc. 7th IEEE Workshop Appl. Comput. Vis.*, 2005, vol. 1, pp. 29–36, doi: 10.1109/ACVMOT.2005.107.
[51] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proc. IEEE*, vol. 90, no. 7, pp. 1151–1163, Jul. 2002.
[52] S. Qiao, W. Shen, Z. Zhang, B. Wang, and A. Yuille, "Deep co-training for semi-supervised image recognition," in *Proc. 15th Eur. Conf. Comput. Vis.*, 2018, pp. 142–159.
[53] Z.-H. Zhou and M. Li, "Tri-training: Exploiting unlabeled data using three classifiers," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 11, pp. 1529–1541, Nov. 2005, doi: 10.1109/TKDE.2005.186.

Wenye Wang received the B.S. degree in physical geography and environmental resource in 2021 from Nanjing University, Nanjing, China, where he is currently working toward the M.S. degree in remote sensing of resources and environment.

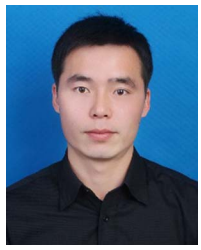
His research interests include semantic segmentation and deep learning for remote sensing.

Shenghua Wan received the B.S. degree in geographic information science in 2021 from Nanjing University, Nanjing, China, where he is currently working toward the Ph.D. degree in computer science and technology.

His research interests include machine learning and reinforcement learning.

Pengfeng Xiao (Senior Member, IEEE) received the B.M. degree in land resource management from Hunan Normal University, Changsha, China, in 2002, and the Ph.D. degree in cartography and geographical information system from Nanjing University, Nanjing, China, in 2007.

From 2007 to 2009, he was a Lecturer with the School of Geography and Ocean Science, Nanjing University, where he was an Associate Professor, from 2010 to 2018. Since 2019, he has been a Professor with Nanjing University. He was a Visiting Scholar with the Department of Geography, University of Giessen, Frankfurt, Germany, from 2011 to 2012, and the Department of Environmental Science, Policy, and Management, University of California at Berkeley, Berkeley, CA, USA, from 2014 to 2015. He has authored 4 books and more than 60 articles. His current research interests include high-resolution remote sensing image analysis, remote sensing of snow cover, and land use and land cover change.



Xueliang Zhang (Member, IEEE) received the B.S. degree in geographical information system and the Ph.D. degree in remote sensing of resources and environment from Nanjing University, Nanjing, China, in 2010 and 2015, respectively.

From 2014 to 2015, he was a Visiting Student with Informatics Institute, University of Missouri, Columbia, MO, USA. From 2016 to 2018, he was an Associate Researcher with the Department of Geographic Information Science, Nanjing University. He is currently an Associate Professor with the Department of Geographic Information Science, Nanjing University. His research interests include high-resolution remote sensing image analysis, semantic segmentation, and deep learning for remote sensing.