MULTI-TRAINING BASED RECOGNITION OF URBAN GREEN COVER ON MULTI-TEMPORAL HIGH RESOLUTION REMOTE SENSING IMAGES

Shenghua Wan[1], Pengfeng Xiao[1]*, Wenye Wang[1]

[1] Department of Geographic Information Science, Nanjing University, Nanjing, Jiangsu 210023, China b

Commission VI, WG VI/4

KEY WORDS: Vegetation extraction, Multi-temporal, Multi-training

ABSTRACT:

Urban green space plays a crucial role in the construction of ecological city and livable environment. The monitoring of urban greening information is of great significance to urban management. Multi-temporal high-resolution remote sensing images are important data sources for urban green cover information updating, while effectively identifying vegetation information from these images is one of the key part which often faces many challenges. Firstly, to achieve good results by supervised learning, large scale training samples are usually needed. However, the labeling of training samples on remote sensing images is time-consuming and laborious. Secondly, because of the data distribution shifting from time to time, the training samples under a single time-phase can not be directly applied to other phases, which results in a great waste of labeled samples. It is necessary to construct the framework of information extraction of vegetation and other kinds of ground objects in multi-temporal images in order to deal with such problems. Based on semi-supervised learning and high resolution remote sensing images under multiple time phases, an algorithm for ground object classification is proposed. This method simultaneously solve the dataset shift and ill-posed problem of multi-temporal remote sensing image classification, and effectively improves the accuracy of small sample training on each image while ensuring the consistency of classification results. It brings a new mode to the information extraction of multi-temporal remote sensing images.

1. INTRODUCTION

Vegetation, as an important part of urban ecological landscape, plays a key role in air purification, water conservation, noise reduction and environmental purification. It is of great practical significance to accurately monitor the information of urban green space in time. Remote sensing image is an important resource to extract urban vegetation cover information. The commonly used data sources are Landsat, MODIS, and now high-resolution remote sensing images bring new impetus to the monitoring of urban vegetation. The property of high spatial resolution provide more accurate vegetation extraction results and sufficient data in time phase. The abundant data of multi-temporal remote sensing images provide strong support for the classification of land cover, which can distinguish land cover objects more effectively than the single-phase classification. However, the classification of remote sensing images under multi-temporal conditions is mainly faced with these problems: (1) Using supervised learning method to extract vegetation information can obtain better results, while it also needs a large number of labeled samples .(2) Because the spectral information of vegetation are influenced by seasonal change, the labeled samples in one phase can not be applied in another phase. Multi-temporal remote sensing images bring more opportunities and challenges to urban vegetation extraction.

To solve the problem of land cover classification based on multi-temporal remote sensing images, some researchers directly apply machine learning methods to stacked multi-temporal data which do not consider the correlation between different time phases and only classify land cover objects on single time phase in isolation. With the development of deep learning techniques, some research combined with recurrent neural network (RNN) in deep learning to mine remote sensing image information on time series for the reason that RNN has the advantage to process time series data, but at meantime it brings huge time cost and needs the great number of labeled samples. In view of this problem, some researchers propose a framework of semi-supervised learning which can be based on small samples. Cooperative training method is applied to identify and extract snow cover in temporal remote sensing images. Some used tri-training method to execute LULC classification. Some have used semi-supervised fuzzy C means clustering method to dynamically extract vegetation information in time series remote sensing images.

Domain adaptation is a representative method of transfer learning whose data contains source domain and target domain. 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. This kind of method is usually based on sample migration, feature-based migration or model-based migration. This method has been widely used in remote sensing field such as remote sensing image classification, land use/land cover classification result map updating. It takes advantage of the feature that the object on similar temporal will not change significantly. But the direct problem is that to be learned well enough, a large number of source domain labeled samples are also needed to train classifiers.

When the labeled training samples are insufficient, the classifier can not get an ideal result through supervised learning. In most time, the number of unlabeled samples is sufficient which will implement data information. Semi-supervised learning is proposed to solve how to utilize the unlabeled data. The main idea is to train an initial classifier by using labeled training samples, and then add new training samples into the training

* Corresponding author

set according to certain rules in the training process, and iteratively make the classifier update. A semi-supervised learning field can be divided into four methods, generative methods (Shahashahani and Landgrebe, 1994), low-density separation algorithms, such as TSVM (Joachims, 1999; Vapnik, 1998), graph-based methods (Jordan, 1998), divergence based methods, such as cooperative training (Blum and Mitchell, 1998). These semi-supervised learning methods have achieved remarkable results in machine learning related tasks. Many scholars use cooperative training methods to classify remote sensing images between phases, but due to the complexity of the remote sensing image data's distribution, such as homomorphism, and the migration of data distribution, these methods can not achieve ideal results directly.

This work proposes a method for land use types classification based on a small number of samples on multi-temporal remote sensing images. Based on the idea of semi-supervised learning, the remote sensing image data of multi phases in the same year is selected. After pre-processing correction and registration, the corresponding four classifiers are used. Under the training condition of a very small number of samples, the prediction probability map of each time is generated iteratively, and the invariant value of the same sample is calculated by combining the probability output value of the four classifiers. Note that this value can essentially refer to the change of a single pixel, and point out the category of the pixel at meantime. According to the setting threshold, we randomly select a certain number of samples from each class added to the training set, and the classifiers train repeatedly until the iteration was completed.

In summary, there are four main contributions in our paper. Firstly, we propose a method that can utilize information from multi-temporal remote sensing images and classify the land use types well especially the vegetation. Then, we solve the problem that different phase's data distribution shifts greatly and we convert it to the power of model's classification. Finally, our method can view different phase's data globally, and make progress both on the accuracy and consistency compared with previous related approaches.

The rest of this article will be organized in six sections: the section 2 is the introduction of methods, the study area and data are introduced in section 3, the experiments are represented in section 4, the presentation and comparative analysis of the results are in section 5, and the discussion and summary are described in section 6, 7.

2. METHODOLOGY

2.1 Base Classifier

Random forest is a widely used method in machine learning which can be regard as ensemble of decision tree classifier. It adopts random selection on data and features, constructs many independent decision trees, and combines the results of all decision trees. Random forest can improve the prediction accuracy without significant increase in computation. The following is the process of random forest :(1) Assuming that there are N samples, N samples are randomly selected to train a decision tree as a sample at the node of the decision tree root. (2) When each sample has M attributes, the nodes of the decision tree are divided when cracking, m attributes are randomly selected from the M (m is much less than M), and then an optimal attribute is selected by the method of decision tree selection. (3) The decision tree formation process splits repeatedly according to (2) until it can not. (4) Build a large number of decision trees according to (1)(2)(3) to form random forests. Following the above procedure, a random forest model with excellent performance can be trained. In our algorithm design, random forest classifier is not the only feasible base classifier. In fact, in our algorithm, the base classifier needs to satisfy that can give the probability of which category each sample belongs to. It is convenient for the following step to select samples with certain probability. So the key feature of the base classifier is that it can predict the probability distribution of the sample set, and our method can be extended by using different base learners without too many constraints.

2.2 Multi-temporal extensions of Base Classifier

2.3 Selection of unlabeled sample for training

When processing multi-temporal remote sensing image data, it is inevitable to encounter data offset problem. The reasons for the inconsistent distribution of data come from various sources: (1) due to the influence of atmospheric radiation at different times, the change of position and height angle of the satellite and the imaging conditions of the satellite are quite different, which leads to the difference of remote sensing images even in the same region; (2) the influence of seasonal variation is not considered at the single phase, the objects in the images will also change in the spectral spectrum under multiple phases; and (3) for the reason of the changes of the objects themselves, the data distribution in the feature space in same area may be seriously shift, even the land cover type will change. Because of the problem of data distribution offset, if the samples in one image be applied directly in other phases, it will often lead to the problem of low classification accuracy and inconsistent classification results.

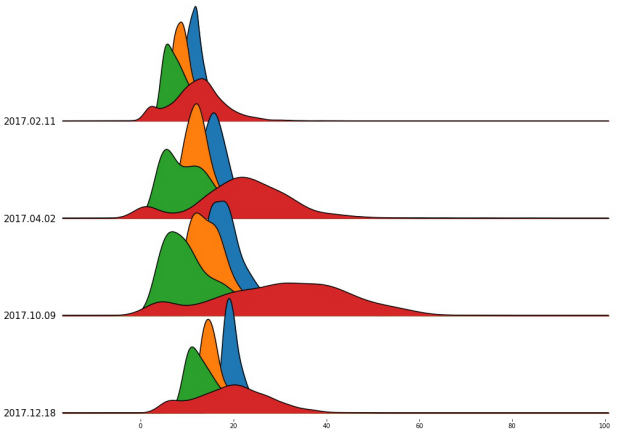

Figure 1. Data distribution on different phases

In order to make full use of the time-phase data information, the most important step is to determine the unchanged region and the changed region of the ground object, so that the information redundancy on the multi-temporal area can be utilized for multi-temporal training to improve classification accuracy. Most of the algorithms of multi-temporal cooperative training need to set the measurement criteria before training, divide the changed area and the unchanged area, and then train

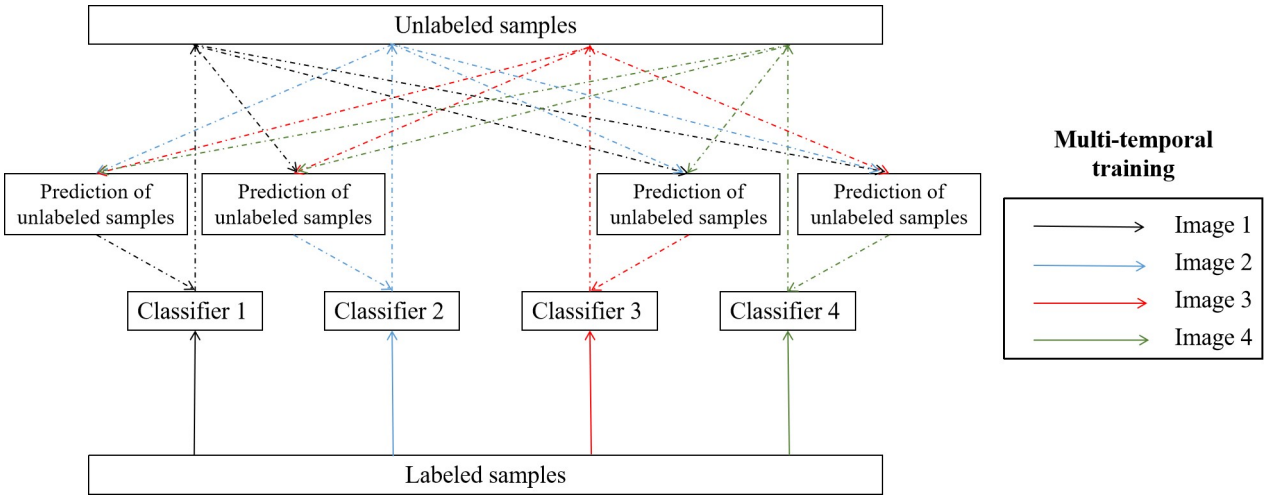

Figure 2. Illustration of how to train multiple classifiers on multiple phases.

on the unchanged area. Setting a unified measurement standard between multiple time phases is equivalent to falling into the error of data deviation illustrated before. Because the data distribution on each image has shifted, the division seems very naive in the case of information intervention in numerous data domains. For example, we have two groups of people trying to find out the tallest of each group. It is known that the first group is generally shorter and the second group is taller and we have already find the tallest one of the first group. We can not directly say that the one in the second group whose height is equal to that of the people is tallest of the second group for the reason that the two groups' height distributions are greatly different. So we must to judge the unchanged area based on its own data domain. In our multi-temporal training method, we abandon the previous method of explicitly measuring the unchanging region according to the distance between data domains, and use implicit prediction probability based on each classifier its own time phase. The joint confidence on multi-phases is calculated to determine pixel's category and unchanged information. The specific calculation steps of joint confidence are demonstrated here. Firstly, on each phase we train a classifier independently. After training, each classifier gives a probability prediction map P_k of pixels on each phase. On each pixel of probability map, there is an array of probability whose size is the number of land cover types and the value on array represents how likely the pixel belongs to this kind of land cover type. Then we combine all the probability maps during the whole episode following the function.

$$C(M_1, M_2, \ldots, M_k) = \frac{(\prod_{i=1}^{k} P_i)^{\frac{2}{k}}}{\frac{1}{k}\sum_{i=1}^{k} P_i} \tag{1}$$

Here, P_i denotes probability map on each single phase, M_1, M_2, ..., M_k denotes the models and $C(M_1, M_2, \ldots, M_k)$ denotes the joint confidence map.

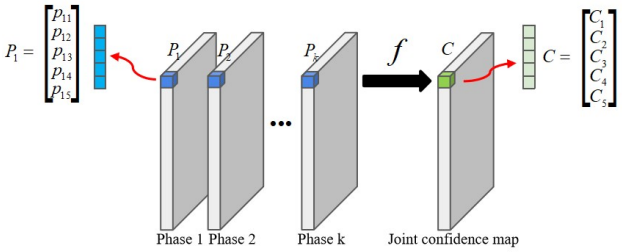

Figure 3. Illustration of combining the multiple phases predicted probability map to generate joint confidence map.

In this function, we combine and reduce the category prediction information of multiple phases into a confidence map over the period of time. The goal of the function is mapping pixel's prediction probability on multi-phases to a confidence value of the whole episode. The confidence computation method is followed by the form of general harmonic mean. It has two advantages: (1) when deciding the class on one pixel during this time, the probability given by multiple classifiers needed to be considered together. If the probabilities given by all classifiers on one pixel are greatly high, then the confidence generated will also be very high implies that it is highly likely that the pixel's type has not changed and its true type is this in this period of time. If some classifiers' probabilities are high but some are low, then the final confidence given will be determined by the gap between the low and high probability values. If the probability given by these classifiers which support the pixel's type unchanged are significantly high, then the confidence can be high possibly. And the low probability can be regarded as some special reason resulting it like seasonal change or illumination change. However, if most classifier's probabilities are very low, the final confidence must be very low suggests that the type of pixel has changed; (2) compared with the voting methods, this mapping function provides a potential way to let this classifiers judge the pixels' type and change state. On this step, joint confidence map only provides a preliminary result of land cover type by the confidence value. The crucial part to update the samples into the training set depends on the subsequent process. Only the samples whose confidence is greater than the threshold can be add into the training set.

Here, we visualize the changing map generated by the ap-

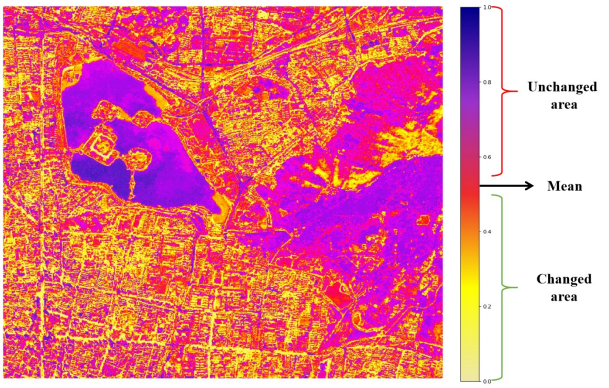

Figure 4. The visualization of changed and unchanged area on the joint confidence map

proach to show where the type has likely changed and where the type has likely unchanged. After each round of classifiers making predictions, we can obtain a class confidence map for this period of time by using the above method, and then update the training set based on the framework of semi-supervised learning. We calculate the average confidence of one class at each round of updates, and set it as a threshold. The pixels whose confidence value of its own class of pixels is higher than this threshold will be picked, and a fixed number of samples are randomly selected to add into the training set. We repeat this process for the setting update epochs.

2.4 Multi-training

We design the following algorithm for this problem and follow the process for training.

Algorithm 1 Framework of Multi-training

Input: The dataset of multiple phases, $D = \{D^1, \ldots, D^k\}$;
The set of labelled samples for each phase, $D_l = \{D_l^1, \ldots, D_l^k\}$;
The initial labelled samples number for each class, $N_l^{(0)}$;
The updating unlabelled samples number for each epoch, N_u;
The updating threshold, λ;
Output: The set of trained models, $M = \{M_1, \ldots, M_k\}$;

1: Initialize models, $M = \{M_1, \ldots, M_k\}$;
2: **for** epoch $<$ endEpochs **do**
3: Train models on each own labelled samples;
4: Make predictions on the whole data, generate the probability map M_i for each phase;
5: Compute the joint confidence map with the function 1;
6: Calculate the updating threshold of each class by the confidence map, $th = \lambda \times mean(\{p|p \in P_i, i\ denotes\ class\})$;
7: Randomly select the unlabelled samples whose confidence is higher than th for each class with setting number N_u;
8: **end for**
9: **return** M;

When we perform the training according to the above algorithm, there are some details need to pay attention to, which will affect the final effect.

3. STUDY AREA AND DATA

3.1 Study area

The study area, Nanjing City,is stood in the lower reaches of the Yangtze River Plain. Located in the subtropical monsoon climate zone, it is great different in vegetation between seasons, which provides powerful support for MULTI-TEMPORAL training. Selected as The National Ecological Garden City, Nanjing City continues to be the first with the forest coverage rate in Jiangsu Province. The main tree species in Nanjing City are Oriental plane, Ginkgo, Camphor and Zelkova schneideriana. So we chose Nanjing as our study area and our imagery focused on Xuanwu Lake area in the center of Nanjing.

3.2 Data

This study used SENTINEL-2-MSI images of Xuanwu Lake area in Nanjing, as shown in 5. The images contains 772653 pixels, and they has four bands with a spatial resolution of 8.98 m. Four bands are Blue(0.46 0.52μm), Green(0.54 0.58μm), Red(0.50 0.80μm), NIR(0.78 0.90μm). Sentinel-2 is a wide-swath, high-resolution, multi-spectral imaging mission supporting Copernicus Land Monitoring studies, including the monitoring of vegetation, soil water cover, as well as observation of inland waterways and coastal areas.

We selected multiple images from August 15,2016 to June 30,2020, every three months. All images are preprocessed to eliminate cloud interference. The three-month sampling interval ensures the difference of the vegetation, while the images of the same month between years ensure the similarity of the vegetation. Through these operations, we obtained varied multi-temporal imagery in Nanjing.

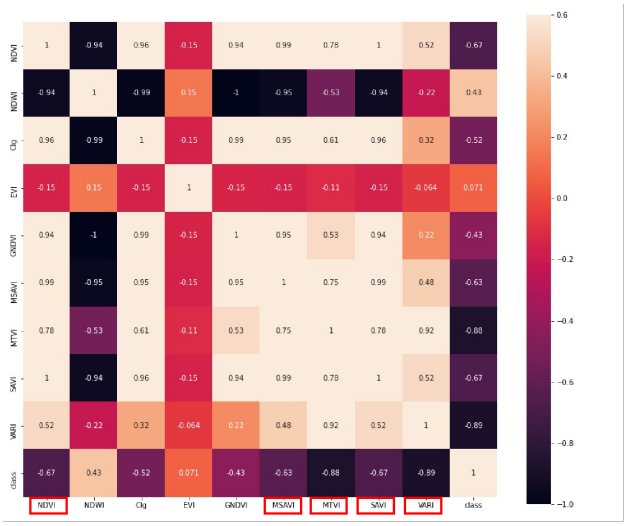

Figure 6. The correlation of features and label.

In order to augment our existing data, we try to construct some meaningful indices on the basis of previous researches. We sample from the data and calculate the correlation degree between the feature and the ground category. In the following figure, we show the correlation degree of these indicators. The first five index () and the original four bands are combined to form 9 dimensional features, which are the final features of the data set.

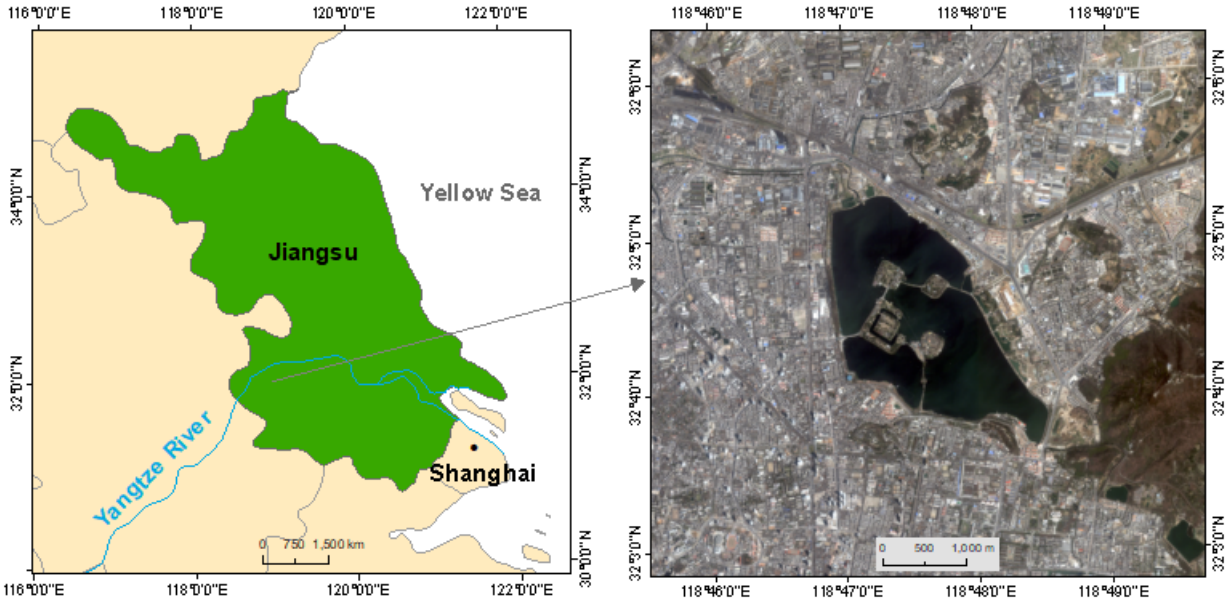

Figure 5. Illustration of the study area.

On each phase, we randomly marked the same number of samples on image and the number of each class labeled are 1,000. We further divide the labeled data set into training set and testing set according to 1:4 ratio. More precisely, the training set here should be named training pool for the reason that subsequent training samples are selected from it with a fixed size independently and repeatedly to form the final training set during the experiment.

4. EXPERIMENTAL DESIGN

4.1 Validation metric

In our experiments, F1-score is selected as the validation metric which is defined in for its advantages that it requires both recall and precision high.

$$F1 - score = \frac{2}{\frac{1}{P} + \frac{1}{R}} \tag{2}$$

4.2 Experimental setup

Two experiments were designed to evaluate the performance of multi-training as well as the effect of unlabeled samples' selection and multiple temporal combination. F1-score is used to represent the advantages and disadvantages of the model, and the average value and variance under multiple independent repeated experiments to reflect the accuracy and stability of the model under different time phases and parameter combinations. In our experiments, some necessary parameters and their explanations are listed in the table. (a)In order to explore the influence of the combination of time phase on the algorithm and decide the best phase combination method, we adjust the number of time phase to test the improved accuracy of the algorithm for small sample training sets on multi-temporal remote sensing images. In our experiment, the number of time phases is 3, 4, 5, 6, 7, 8, and images started from 2017. The average improved accuracy under multiple time phases is set as the final measurement. Because other methods are not directly related to the number of time phases, only our method in this paper is discussed at this part. (b)In the second experiment, we start to compare average accuracy and standard deviation between different methods on this task. The second experiment contains three sub-experiments which are designed to evaluate the number of labeled and unlabeled samples and the confidence threshold's effect to the model's accuracy. To achieve this goal, we change the number of labeled samples for each class before training and the number of unlabeled samples when updating the training set and adjust the threshold of confidence when selecting the unlabeled samples. In out experiment, Nl ranges from 1 to 19 with step 2, Nu varies from 5 to 50 with step 5 and changes from 0.8 to 1.2 with step 0.05. Here we represent the experiments' name and parameters setting in each experiment.

5. RESULTS

5.1 Mapping results of Multi-training

In order to represent the results of vegetation extraction intuitively and analyze qualitatively, we compared the classification results under four phases and highlighted the area of vegetation extraction. In order to facilitate positioning, we also visualize the Xuanwu Lake in the result maps. The images on the left of the figure below is the study area of Sentinal image under four phases, the middle is the vegetation result map extracted by our algorithm, and the right column is the vegetation extraction result obtained by semi-supervised learning method of the single phase. From the comparison of results given by our algorithm and single-phase, we can see that the vegetation extracted by our algorithm is more delicate and accurate. Although based on the classification of pixels, the results of our extraction are still very consistent in the area with wide coverage of vegetation. In contrast, the vegetation area extracted by single-phase semi-supervised learning method is relatively broken. In the identification of vegetation cover in urban blocks, we can clearly find that most of the vegetation is distributed along the road in our result maps. The vegetation extracted by our algorithm has bette consistency during this peroid and can reflect the influence of phase change on vegetation, such

as the decrease and increase of mountain vegetation caused by seasonal change. Our multi-training algorithm has a significant advantage in mutli-temporal vegetation extraction.

5.2 Performance of Multi-training

5.3 Influence of unlabeled sample selection

5.4 Influence of different number of phases

Because our algorithm designed is for multi-temporal remote sensing images' vegetation extraction, the number of time phases will affect the performance of the algorithm. The number of time phases is modified from 3 to 8 in our experiment, and the adjacent phases are separated from each other for two to three months to explore the trend of F1-score and standard deviation. From the results, we can see that with the increase of time phase number, the average improved F1-score of vegetation extraction increases obviously, from 2 percentage points under 3 time phase to 5 percentage points under 8 time phase. This result can be explained by the equation.1. The criterion of confidence in our algorithm is provided by the prediction results of classifier on multiple time phases. We relax the equation to its up-bound which is the predicted probabilities' average value.

$$\frac{(\prod_{i=1}^{k} P_i)^{\frac{2}{k}}}{\frac{1}{k}\sum_{i=1}^{k} P_i} \leq \frac{(\frac{1}{k}\sum_{i=1}^{k} P_i)^2}{\frac{1}{k}\sum_{i=1}^{k} P_i} = \frac{1}{k}\sum_{i=1}^{k} P_i \tag{3}$$

When phases' number increases, the variance decreases and the average stabilizes. The increase of the number can make the result more robust, which is beneficial to the judgment of the unchanged area and the selection of unlabeled samples when updating the training set. The increasing number of phases makes the algorithm more stable, so that the multi-training coincides with its original intention. The further improvement with the increase of the number of phases is the characteristic that the general algorithm considers the single phase in isolation does not have.

6. DISCUSSIONS

7. CONCLUSIONS

ACKNOWLEDGEMENTS

Acknowledgements of support for the project/paper/author are welcome.

REFERENCES

APPENDIX

Any additional supporting data may be appended, provided the paper does not exceed the limits given above.

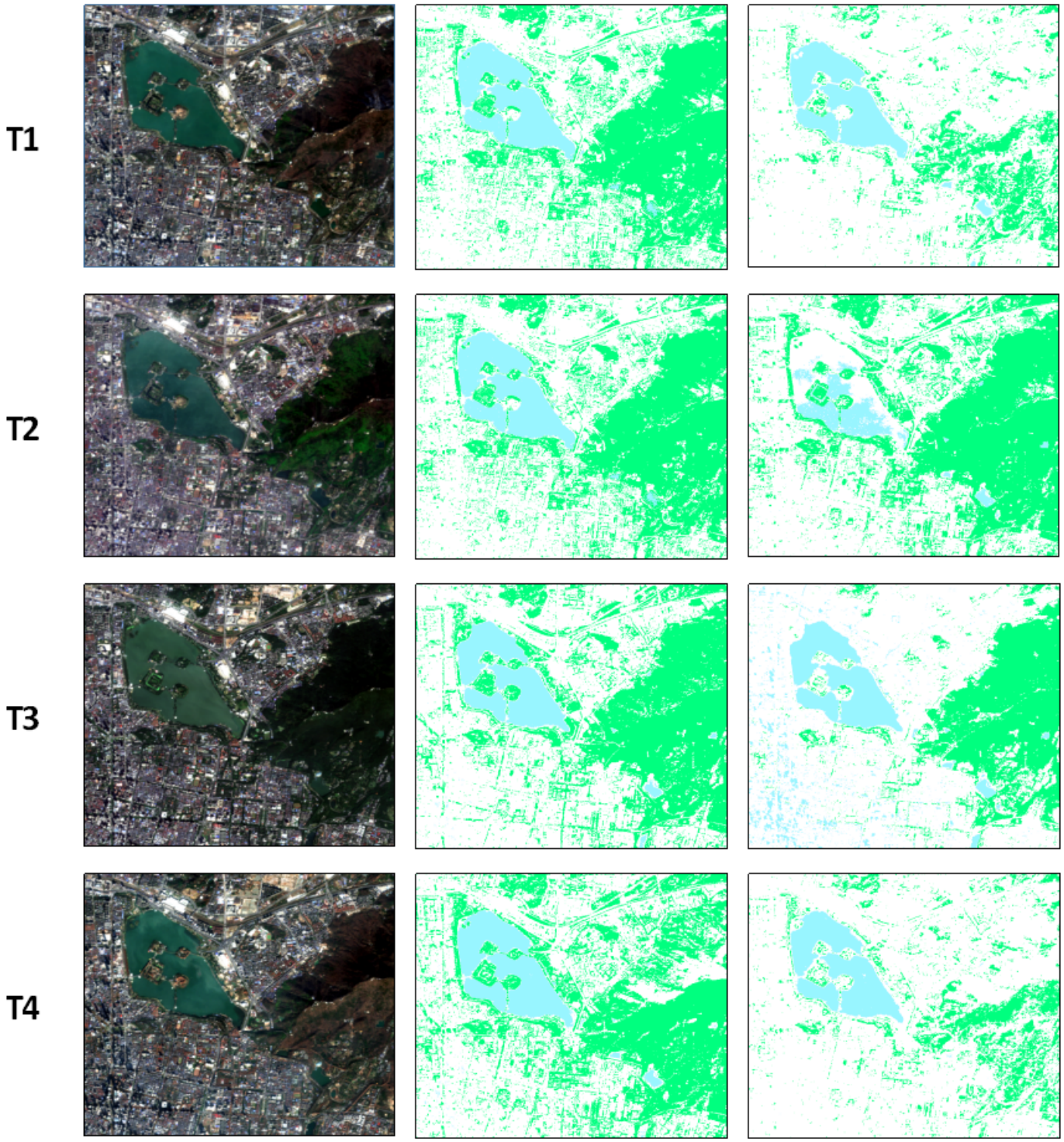

Figure 7. Four images on the focus area with true colour composite (left) and their results of vegetation cover extraction by our algorithm (middle) and single-phase training (right).