

# Remote Sensing Image Transfer Classification Based on Weighted Extreme Learning Machine

Yang Zhou, Jie Lian, *Member, IEEE*, and Min Han, *Senior Member, IEEE*

**Abstract**—It is expensive in time or resources to obtain adequate labeled data for a new remote sensing image to be categorized. The cost of manual interpretation can be reduced if labeled samples collected from previous temporal images can be reused to classify a new image over the same investigated area. However, it is reasonable to consider that the distributions of the target data and the historical data are usually not identical. Therefore, the efficient strategy of transferring the beneficial information from historical images to the target image hits a bottleneck. In order to reuse sufficient historical samples to classify a given image with scarce labeled samples, this letter presents a novel transfer learning algorithm for remote sensing image classification based on extreme learning machine with weighted least square. This algorithm adds a transferring item to an objective function and adjusts historical and target training data with different weight strategies. Experiments on two sets of remote sensing images show that the presented algorithm reduces the requirement for target training samples and improves classification accuracy, timeliness, and integrity.

**Index Terms**—Extreme learning machine (ELM), image classification, remote sensing, transfer learning, weighted least square.

## I. INTRODUCTION

REMOTE sensing monitoring is a dynamic process which requires observations of multitemporal remote sensing images. However, it is time-consuming and difficult to collect adequate training samples to a new remote sensing image classification task, which may seriously restrict the rapid thematic mapping [1]. Obviously, the small-sample-size problem can lead to unsatisfactory accuracy as the training samples of the target image are scarce [2]. In fact, the samples collected from historical images in the past can provide rich prior knowledge for the new classification task. However, influenced by some possible reasons (e.g., alterations in atmospheric radiation, dissimilar illumination conditions, phenological changes, different land cover, etc.), the spectral values usually change in different image acquisition dates. Hence, historical images and target images might obey different probability statistical distributions, although obtained in the same area [3], which do not meet

the basic assumption of statistical learning theory. The existing algorithms lack effective strategies to reuse historical samples, resulting in the waste of historical resources. Therefore, how to transfer the prior knowledge in historical data to the target model is the key issue to solve [4].

Transfer learning is an important attempt to solve the aforementioned problems, but to our knowledge, few related works have been done. Bruzzone *et al.* [5] designed an advanced algorithm for updating land use map by applying domain-adaptation SVM. Matasci *et al.* [6] employed TrAdaBoost algorithm to settle the transfer problem for hyperspectral image classification. Liu *et al.* [7] proposed the TrCbrBoost algorithm which successfully makes use of source domain data to realize land use mapping of target data set when true labeled data are unavailable. An SVM ensemble approach is presented by Huang *et al.* [8], where a series of spectral-spatial features was considered. They further carried out the multitemporal transfer learning method to detect urban areas from high-resolution imagery [9]. Nonetheless, the research of transfer learning about remote sensing image classification is still in its infancy, and further study on how to rapidly and directly transfer useful historical information to the target model is in demand [10].

Under the circumstance of scarce available labeled target data and sufficient labeled historical data, this letter proposes a transfer learning algorithm for remote sensing image classification based on extreme learning machine (ELM) [11] with weighted least square method, which is called the weighted ELM for transfer classification (WELMTC) algorithm. It measures domain similarity in the objective function in the form of regularization. With different weight adjustment strategies for historical samples and target samples, this algorithm transfers prior knowledge in the historical image to the target model.

## II. WELMTC REMOTE SENSING IMAGE CLASSIFICATION

ELM is an effective method for remote sensing image classification, providing good generalization performance at extremely fast learning [11]. However, same as other machine learning algorithms, the ELM is under the assumption that the training data sets are drawn from the same distribution. Therefore, information from an old image cannot be directly used to remedy the insufficient labeled data for a new image. Different from the traditional neural network learning algorithms which require setting a lot of training parameters, ELM need not adjust the input weights and hidden layer biases and can obtain a unique optimal solution if the activation function is infinitely differentiable. Only the output weights need to be learned, so once ELM is trained, the information about the image data is

Manuscript received January 28, 2016; accepted March 9, 2016. Date of publication August 1, 2016; date of current version September 16, 2016. This work was supported by the project of the National Natural Science Foundation of China under Grant 61374154 and the Special Fund for Basic Research on Scientific Instruments of the National Natural Science Foundation of China under Grant 51327004.

The authors are with the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China (e-mail: minhan@dlut.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LGRS.2016.2568263

recorded by output weights  $\hat{\beta}_1$ . To transfer information from the initial model to the target model, a transfer item [12], [13]  $\|\beta_2 - \hat{\beta}_1\|$  is introduced into the cost function, and an iterative process is conducted to adjust the output weights of ELM until all of the labeled data from the target image can be recognized correctly.

#### A. WELMTC

Let  $S_1$  and  $S_2$  indicate two remote sensing images acquired over the same study area at different times, where  $S_1$  is the historical image and  $S_2$  is the target image needed to be classified.  $L_1 = \{(\mathbf{x}_i^1, \mathbf{t}_i^1) | \mathbf{x}_i^1 \in S_1\}_{i=1}^{N_1}$  and  $L_2 = \{(\mathbf{x}_i^2, \mathbf{t}_i^2) | \mathbf{x}_i^2 \in S_2\}_{i=1}^{N_2}$  are data sets from different images composed of  $N_1$  and  $N_2$  labeled samples, where  $\mathbf{x}_i = [x_{i1}, \dots, x_{in}]^T \in R^n$  is the feature vector and  $\mathbf{t}_i = [t_{i1}, \dots, t_{im}]^T \in R^m$  is the label vector. The  $L_1$  historical data set is the supplementary of the  $L_2$  target data set and provides additional information for the classification of the target image.

For data set  $L_1$ , the ELM with activation function  $g(\bullet)$  and  $\tilde{N}$  hidden nodes are mathematically modeled as

$$\sum_i^{\tilde{N}} \beta_i^1 g(\mathbf{w}_i^1 \cdot \mathbf{x}_j^1 + b_i^1) = t_j^1, j = 1, \dots, N_1 \quad (1)$$

where  $\mathbf{w}_i = [w_{i1}, \dots, w_{in}]^T$  is the input weight vector,  $\beta_i = [\beta_{i1}, \dots, \beta_{im}]^T$  is the output weight vector, and  $b_i$  is the bias of the  $i$ th hidden node. The aforementioned equation is able to be rewritten briefly as  $\mathbf{H}_1 \beta_1 = \mathbf{T}_1$ , where the initial ELM hidden layer output matrix  $\mathbf{H}_1$  can be calculated as

$$\mathbf{H}_1 = \begin{bmatrix} g(\mathbf{w}_1^1 \cdot \mathbf{x}_1^1 + b_1^1) & \cdots & g(\mathbf{w}_{\tilde{N}}^1 \cdot \mathbf{x}_1^1 + b_{\tilde{N}}^1) \\ \vdots & \ddots & \vdots \\ g(\mathbf{w}_1^1 \cdot \mathbf{x}_{N_1}^1 + b_1^1) & \cdots & g(\mathbf{w}_{\tilde{N}}^1 \cdot \mathbf{x}_{N_1}^1 + b_{\tilde{N}}^1) \end{bmatrix}_{N_1 \times \tilde{N}} \quad (2)$$

$$\beta_1 = \begin{bmatrix} \beta_1^1 \\ \vdots \\ \beta_{\tilde{N}}^1 \end{bmatrix}_{\tilde{N} \times m} \quad \text{and} \quad \mathbf{T}_1 = \begin{bmatrix} \mathbf{t}_1^1 \\ \vdots \\ \mathbf{t}_{N_1}^1 \end{bmatrix}_{N_1 \times m} \quad (3)$$

The initial ELM is obtained by calculating the output weights as follows:  $\hat{\beta}_1 = \mathbf{H}_1^\dagger \mathbf{T}_1$  using historical data set  $L_1$ , such that  $\hat{\beta}_1 = \arg \min_{\beta_1} \|\mathbf{T}_1 - \mathbf{H}_1 \beta_1\|$ , where  $\mathbf{H}_1^\dagger$  is the Moore–Penrose generalized inverse of  $\mathbf{H}_1$ .

The initial ELM output weights are calculated directly by the least square estimation (LSE) algorithm. Each training sample is given the same weight. However, in the applications of multi-temporal remote sensing image classification, the contributions of different samples are different for the target model, so it is necessary to assign different weights for different samples. The use of the weighted least squares estimation (WLSE) [14] algorithm realizes the weighted processing of ELM. The optimization objective function of the initial ELM based on WLSE is defined as

$$J_1 = \|\mathbf{T}_1 - \mathbf{H}_1 \beta_1\| \Sigma_1 \quad (4)$$

where  $\Sigma_1 = \text{diag}(\sigma_1^1, \dots, \sigma_n^1) \in R^{n \times n}$ ,  $\sigma_1^1, \sigma_2^1, \dots, \sigma_n^1 > 0$  is a positive definite diagonal weight matrix, and  $\sigma_i^1$  is the weight

of  $i$ th sample  $(\mathbf{x}_i^1, \mathbf{t}_i^1)$  in data set  $L_1$ . The partial derivative of the objective function is

$$\begin{aligned} \frac{\partial J_1}{\partial \beta_1} &= \frac{\partial [(\mathbf{T}_1 - \mathbf{H}_1 \beta_1)^T \Sigma_1 (\mathbf{T}_1 - \mathbf{H}_1 \beta_1)]}{\partial \beta_1} \\ &= -\mathbf{H}_1^T \Sigma_1 \mathbf{T}_1 - \mathbf{H}_1^T \Sigma_1^T \mathbf{T}_1 + \mathbf{H}_1^T \Sigma_1 \mathbf{H}_1 \beta_1 + \mathbf{H}_1^T \Sigma_1^T \mathbf{H}_1 \beta_1 \\ &= -\mathbf{H}_1^T (\Sigma_1 + \Sigma_1^T) (\mathbf{T}_1 - \mathbf{H}_1 \beta_1). \end{aligned} \quad (5)$$

Then, set (5) as zero

$$\frac{\partial J_1}{\partial \beta_1} = -\mathbf{H}_1^T (\Sigma_1 + \Sigma_1^T) (\mathbf{T}_1 - \mathbf{H}_1 \beta_1) = 0. \quad (6)$$

Therefore, the output weights of historical ELM can be calculated as

$$\hat{\beta}_1 = (\mathbf{H}_1^T \Sigma_1 \mathbf{H}_1)^{-1} \mathbf{H}_1^T \Sigma_1 \mathbf{T}_1. \quad (7)$$

Data set  $L_2$  from the target image is used to transfer the initial ELM to adapt to the new image  $S_2$ . The objective function to minimize is as follows:

$$J_2 = (1 - \lambda) \|\mathbf{T}_2 - \mathbf{H}_2 \beta_2\| \Sigma_2 + \lambda \|\beta_2 - \hat{\beta}_1\| \quad (8)$$

where  $\Sigma_2$  is the weight matrix of the target samples in  $L_2$ .  $\mathbf{H}_2 \in R^{N_2 \times \tilde{N}}$  is the output matrix of the hidden layer calculated by  $L_2 = \{(\mathbf{x}_i^2, \mathbf{t}_i^2) | \mathbf{x}_i^2 \in S_2\}_{i=1}^{N_2}$ , and the input weights and biases are the same as the initial ELM:  $\mathbf{w}_i^2 = \mathbf{w}_i^1$ ,  $b_i^2 = b_i^1$ ,  $i = 1, \dots, \tilde{N}$ .  $\|\beta_2 - \hat{\beta}_1\|$  is the item to determine the direction of transfer learning, and parameter  $\lambda$  determines the intensity of transfer learning which is chosen by repeated experiments. The output weights are transferred by

$$\hat{\beta}_2 = \arg \min_{\beta_2} J_2. \quad (9)$$

The partial derivative of objective function  $J_2$  is

$$\frac{\partial J_2}{\partial \beta_2} = (1 - \lambda) (-2\mathbf{H}_2^T \Sigma_2 \mathbf{T}_2 + 2\mathbf{H}_2^T \Sigma_2 \mathbf{H}_2 \beta_2) + 2\lambda (\beta_2 - \hat{\beta}_1). \quad (10)$$

Then, set (10) to zero to calculate the output weights after transfer learning

$$\hat{\beta}_2 = [(1 - \lambda) \mathbf{H}_2^T \Sigma_2 \mathbf{H}_2 + \lambda \mathbf{I}_{\tilde{N}}]^{-1} [(1 - \lambda) \mathbf{H}_2^T \Sigma_2 \mathbf{T}_2 + \lambda \hat{\beta}_1]. \quad (11)$$

Plug the output weight calculated by historical data set  $L_1 = \{(\mathbf{x}_i^1, \mathbf{t}_i^1) | \mathbf{x}_i^1 \in S_1\}_{i=1}^{N_1}$  [as shown in (7)] into (11) to obtain the final result of output weights of WELMTC

$$\begin{aligned} \hat{\beta}_2 &= [(1 - \lambda) \mathbf{H}_2^T \Sigma_2 \mathbf{H}_2 + \lambda \mathbf{I}_{\tilde{N}}]^{-1} \\ &\times [(1 - \lambda) \mathbf{H}_2^T \Sigma_2 \mathbf{T}_2 + \lambda (\mathbf{H}_1^T \Sigma_1 \mathbf{H}_1)^{-1} \mathbf{H}_1^T \Sigma_1 \mathbf{T}_1]. \end{aligned} \quad (12)$$

In the special case that  $\Sigma_1 = \Sigma_2 = \mathbf{I}$ , ELM learns both historical and target samples without distinguishing the different contributions of different data. The training result is  $\hat{\mathbf{T}} = \mathbf{H}_2 \hat{\beta}_2$ . If  $\|\hat{\mathbf{T}} - \mathbf{T}_2\| \neq 0$ , then adjust the weight of samples, and go back to (11) to continue the transfer learning until all of the samples in  $L_2$  can be recognized correctly.

### B. Adjustment Strategy of the Sample Weight and Algorithm Process

In the process of transfer learning, not all available samples can improve the performance of classification. Therefore, an adjustment strategy of weighting samples is introduced to control the effect of samples in different domains. The initial  $\Sigma_1$  and  $\Sigma_2$  are given artificially. If a sample from the target image is misclassified, we consider that it has not been fully learned. In order to emphasize this target sample in  $L_2$ , we need to increase the weight of it according to the AdaBoost algorithm [15]. On the contrary, if the auxiliary sample from the historical images obtains error result, it means that it is inconsistent with the distribution of target image, so the historical sample is not suitable to establish the target classification model. According to Hedge ( $\beta$ ) strategy [16], we lower the weight of the misclassified sample in  $L_1$  to rapidly reduce the influence of this sample in the training process.

Concretely, during the training process, if the sample cannot learn correctly, its weight will be adjusted by

$$\sigma = \begin{cases} \frac{\sigma_i^1}{(1+\sqrt{2\ln N_2})}, & \sigma_i \text{ is the weight for } L_1 \\ \frac{\sigma_j^2(1-\gamma)}{\gamma}, & \sigma_j \text{ is the weight for } L_2 \end{cases} \quad (13)$$

where  $\gamma$  is the error rate with  $\hat{\beta}_2$  on target data set  $L_2 = \{(\mathbf{x}_i^2, \mathbf{t}_i^2) | \mathbf{x}_i^2 \in S_2\}_{i=1}^{N_2}$ , which can be calculated as

$$\gamma = \sum_{j=1}^{N_2} \frac{\sigma_j^2 h_j}{\sum_{j=1}^{N_2} \sigma_j^2}, \quad h_j = \begin{cases} 1, & \text{if } \max(\hat{\mathbf{t}}_j) \neq \max(\mathbf{t}_j) \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

After several iterations, the different distribution training samples that can better fit the same distribution samples will hold larger weights, whereas the different distribution training samples that are not similar to the same distribution ones will hold lower weights. Therefore, the historical samples with large weights will help in assisting the target samples to establish a better transfer learning classifier.

The algorithm of WELMTC is given as follows.

---

**Algorithm:** Weighted Extreme Learning Machine for Transfer Classification (WELMTC)

---

**Input:** Two training data sets  $L_1 = \{(\mathbf{x}_i^1, \mathbf{t}_i^1) | \mathbf{x}_i^1 \in S_1\}_{i=1}^{N_1}$  and  $L_2 = \{(\mathbf{x}_i^2, \mathbf{t}_i^2) | \mathbf{x}_i^2 \in S_2\}_{i=1}^{N_2}$

**Initialization:**

1. Initialize the sample weights
 
$$\begin{cases} \sigma_i^1 = 1/N_1, & i = 1, \dots, N_1 \\ \sigma_j^2 = 1/N_2, & j = 1, \dots, N_2 \end{cases}$$
2. Set ELM input weights  $\mathbf{w}$  and node bias  $b$  randomly.
3. Set  $\rho = 1/(1 + \sqrt{2\ln N_2})$

**While**  $\|\mathbf{H}_2\hat{\beta}_2 - \mathbf{T}_2\| > \varepsilon$

1. Calculate the output weights  $\hat{\beta}_2$
2. Calculate the error rate  $\gamma$  with  $\hat{\beta}_2$  on target data set  $L_2 = \{(\mathbf{x}_i^2, \mathbf{t}_i^2) | \mathbf{x}_i^2 \in S_2\}_{i=1}^{N_2}$

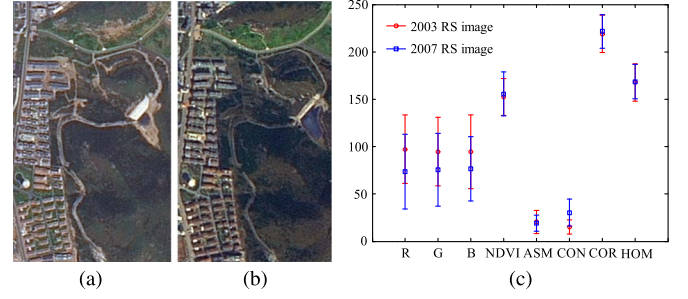


Fig. 1. (a) Remote sensing image of 2003. (b) Remote sensing image of 2007. (c) Range of input features for two images; markers represent the mean values, and bars represent the standard deviation.

3. Update  $\hat{\rho} = \gamma/(1 - \gamma)$

4. Update sample weights as follows: 
$$\begin{cases} \sigma_i^1 = \sigma_i^1 \rho^{h_i} \\ \sigma_j^2 = \sigma_j^2 \hat{\rho}^{-h_j} \end{cases},$$

$$h_i = \begin{cases} 1, & \text{if } \max(\hat{\mathbf{t}}_i) \neq \max(\mathbf{t}_i) \\ 0, & \text{otherwise} \end{cases},$$

$$h_j = \begin{cases} 1, & \text{if } \max(\hat{\mathbf{t}}_j) \neq \max(\mathbf{t}_j) \\ 0, & \text{otherwise} \end{cases}$$

**End While**

**Output:** The final classification result

$$\hat{\mathbf{T}} = \mathbf{H}_2\hat{\beta}_2$$

### III. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, two sets of remote sensing images are applied to verify the validity of the proposed WELMTC algorithm.

#### A. SPOT-5 Remote Sensing Images of Dalian, China

The image collected on May 5, 2003 [Fig. 1(a)], is regarded as the historical image, and the image collected on April 13, 2007 [Fig. 1(b)], is the target image, of which both spatial resolutions are 2.5 m with the size of  $393 \times 236$ . Six classes of land use include water bodies, dense forest, sparse forest, bare lands, buildings, and grasslands. Regardless of the spectral features, we introduce the normalized differential vegetation index and four texture features including angular second moment, contrast, correlation, and homogeneity [11]. The different distributions between two images can be seen in Fig. 1(c), especially in the sixth input variable, where even the mean value is outside the other's standard deviation.

In order to ensure that the samples can better reflect the characteristics of the image, an adaptive method is applied to obtain training data according to the variation intensity of different land use categories [16]. In flat or background region with fewer details, we lower the sampling density, whereas in regions with abundant details, we increase the sampling density. In order to simulate the circumstance that target samples are scarce and the historical samples are sufficient, we build target data set  $L_2$  using a small amount of sample and historical data set  $L_1$  with plenty of samples. Table I shows two data sets acquired by visual interpretation and field trips.

TABLE I  
DESCRIPTION OF DATA SET:  $L_1$  HISTORICAL  
DATA SET AND  $L_2$  TARGET DATA SET

Dataset	Water bodies	Dense forest	Sparse forest	Bare lands	Building	Grass lands
$L_1(2003)$	49	245	245	147	147	245
$L_2(2007)$	27	45	45	45	27	45

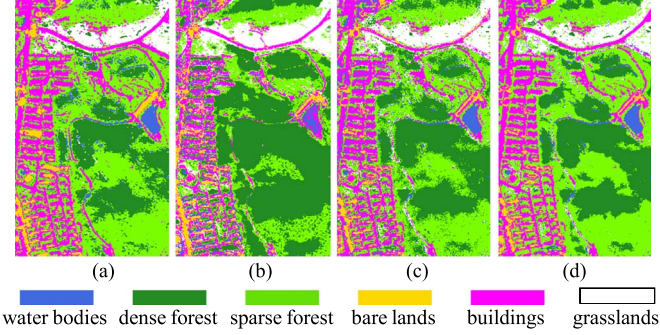


Fig. 2. Comparison of classification results for the target image in 2007. (a) Classification result of the initial ELM on  $L_2$ . (b) Classification result of OS-ELM on  $L_1$  and  $L_2$ . (c) Classification result of WELMTC ( $\Sigma_1 = \Sigma_2 = \mathbf{I}$ ). (d) Classification result of WELMTC.

TABLE II  
ACCURACY OF CLASSIFICATION

Methods	ELM	OS-ELM	WELMTC ( $\Sigma=\mathbf{I}$ )	WELMTC
Water	1.0000	0.7449	1.0000	1.0000
Dense forest	0.9082	1.0000	1.0000	1.0000
Sparse forest	0.9653	0.7204	0.9245	0.9918
Bare land	0.7449	0.9082	0.8367	0.8061
Building	0.6667	0.5714	0.7619	0.7619
Grass	0.9837	0.926	0.9673	0.9755
OA	0.9107	0.8393	0.9388	0.9592
Kappa	0.8833	0.7882	0.9195	0.9463

Fig. 2 shows the classification results of the target image in Fig. 1(b) based on different methods. Then, on this basis, quantitative evaluation of the classification accuracy is conducted. Test data are obtained from the target image by  $7 \times 7$  random group sampling strategy. We choose 1568 test data randomly and give the real labels to each sample. Table II shows the classification accuracy including the overall accuracy and kappa coefficient of different types of land use based on repeated experiments for 50 times. The activation function applied in this experiment is a simple sigmoidal function  $g(x) = 1/(1 + \exp(-x))$ , and the number of hidden nodes is 160, considering both the training accuracy and time.

By comparing the classification results in Fig. 2 and classification accuracies in Table II, it can be noticed that the proposed WELMTC has the best classification result. It not only improves the classification accuracy but also enhances the integrity of the classification result.

As comparison method, Fig. 2(a) is obtained by initial ELM. It shows that data set  $L_2$  from the target image is not sufficient to obtain satisfactory result, which leads to a serious landscape fragmentation. In Fig. 2(b), although online sequential ELM (OS-ELM) [17] implements the successive learning of two data sets, it lacks transfer learning strategy, which makes it impossible to improve the result by adding historical data set. Suffered from spectral drift, the accuracy is instead reduced

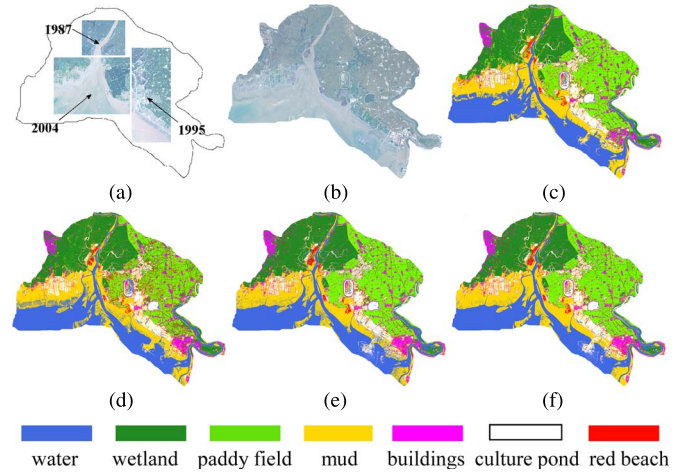


Fig. 3. Classification results for the target image in 2009. (a) Historical data distribution. (b) Target image in 2009. (c) Classification result of ELM. (d) Classification result of historical data set added target data set. (e) Classification result of WELMTC ( $\Sigma_1 = \Sigma_2 = \mathbf{I}$ ). (f) Classification result of WELMTC.

when historical samples are added stiffly to target samples comparing with only using target samples.

The proposed WELMTC is in the domain-adaptation framework. It can reuse historical data set effectively and obtain better classification results [Fig. 2(c) and (d)]. To analyze the effect of the sample weight adjustment strategy, a comparison result in Fig. 2(c) is conducted by setting the weight matrix as the unit matrix ( $\Sigma_1 = \Sigma_2 = \mathbf{I}$ ) for the transfer learning. In Fig. 2(d), it further improves the learning effect because the sample weight strategy controls the effect of samples from different domains, which can transfer the prior knowledge selectively.

## B. TM Remote Sensing Images of Panjin, China

In order to further prove the applicability of the proposed method to larger remote sensing images with more complex scenes, a series of TM multitemporal images with the size of  $1776 \times 2177 \times 7$  taken in Panjin, a world famous China Wetland Capital, is used in this section. As shown in Fig. 3(a), historical data are extended so that 735, 1764, and 1225 samples are acquired from the 1987, 1995, and 2004 images, respectively. Here, 980 samples collected from the image in 2009 are regarded as the target data, where land use is divided into water, wetland, paddy field, mud, buildings, culture pond, and red beach (*Suaeda heteroptera*). The average accuracies of 50 times repeated experiments on 3675 test data are shown in Table III.

It can be seen from Table III and Fig. 3 that the proposed method further improves the classification accuracy for large target remote sensing images. The range extension of historical data set helps in building complete priori information, thus further increasing the classification accuracy.

For the large remote sensing image classification task, the time consumption of the algorithm may directly influence the timeliness of remote sensing monitoring. Therefore, in order to further illustrate the efficiency of the proposed algorithm, it is compared with some other available algorithms based on different amounts of data derived from TM remote sensing images of Panjin. The time requirement results are shown in Table IV,



TABLE III

ACCURACY OF CLASSIFICATION USING DIFFERENT DATA SETS: MIXED DATA SET REFERS TO TARGET DATA SET ADDED HISTORICAL DATA SET

	target dataset	historical dataset			mixed dataset			WELMTC( $\Sigma=I$ )			WELMTC		
data sources	2009	1995	+1987	+2004	1995	+1987	+2004	1995	+1987	+2004	1995	+1987	+2004
OA(%)	86.48	70.37	72.76	73.50	79.89	80.11	82.97	87.84	88.19	90.69	89.77	92.70	95.70
Kappa	0.8254	0.6122	0.6529	0.6551	0.8297	0.7427	0.7771	0.8421	0.8477	0.8789	0.8687	0.9144	0.9444

TABLE IV

TIME COMPARISON OF THE ALGORITHMS: THE NODE NUMBERS OF ELM AND BAYESIAN ARTMAP (BAM) ARE BOTH 100, AND THE NUMBER OF TARGET DATA ARE 500

sample size	$10^3$		$10^4$		$10^5$	
	train(s)	test(s)	train(s)	test(s)	train(s)	test(s)
BAM	0.1263	0.0463	1.0287	0.3179	9.9836	3.1693
SVM	0.0124	0.0058	0.3818	0.1869	21.0745	10.6440
ELM	0.0158	0.0032	0.6788	0.0477	1.9789	0.3550
WELMTC	0.0172	0.0038	0.6809	0.0341	2.0598	0.3704

where the experimental conditions are the following: Intel(R) Core(TM) i7 4720HQ, Quad-Core, 64-b operating system, 8-GB memory, and 2.60 GHz.

As shown in Table IV, even when samples increase to large-scale, the proposed method can reach a fast speed whereas SVM and BAM trains and tests for a long time at the same level of accuracy about 95%. That is because the ELM algorithm has good generalization performance with extremely fast learning.

#### IV. CONCLUSION

This letter has proposed a domain-adaptation algorithm WELMTC for remote sensing image classification for the circumstance of scarce labeled samples in the target image and abundant labeled samples in the historical image over the same area. It solves the problematic strategy of transferring useful information from the historical image to the target one. The proposed method is applied to SPOT-5 and TM remote sensing images. Then, the time requirement of the proposed algorithm is compared with some other available algorithms. From the contrast of classification results, the conclusion can be drawn that WELMTC reduces the requirement for target training samples and effectively improves the classification accuracy, timeliness, and integrity.

#### REFERENCES

- [1] Y. S. Zhang *et al.*, "Semi-supervised manifold learning based multigraph fusion for high-resolution remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 2, pp. 464–468, Feb. 2014.
- [2] L. La *et al.*, "Transfer learning with reasonable boosting strategy," *Neural Comput. Appl.*, vol. 24, no. 3, pp. 807–816, Mar. 2014.
- [3] J. M. Leiva-Murillo, L. Gómez-Chova, and G. Camps-Valls, "Multitask remote sensing data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 151–161, Jan. 2013.
- [4] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. A. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.
- [5] L. Bruzzone and M. Marconcini, "Toward the automatic updating of land-cover maps by a domain-adaptation SVM classifier and a circular validation strategy," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1108–1122, Apr. 2009.
- [6] G. Matasci, M. Volpi, M. Kanevski, L. Bruzzone, and D. Tuia, "Semi-supervised transfer component analysis for domain adaptation in remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3550–3564, Jul. 2015.
- [7] Y. Liu and X. Li, "Domain adaptation for land use classification: A spatio-temporal knowledge reusing method," *ISPRS J. Photogramm. Remote Sens.*, vol. 98, pp. 133–144, Sep. 2014.
- [8] X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 257–272, Jan. 2013.
- [9] X. Huang, H. Liu, and L. Zhang, "Spatiotemporal detection and analysis of urban villages in Mega City regions of China using high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3639–3657, Jul. 2015.
- [10] S. J. Pan and Q. A. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [11] N. B. Chang, M. Han, W. Yao, L. C. Chen, and S. Xu, "Change detection of land use and land cover in an urban region with SPOT-5 images and partial Lanczos extreme learning machine," *J. Appl. Remote Sens.*, vol. 4, no. 1, Nov. 2010, Art. no. 043551.
- [12] M. Long, J. Wang, G. Ding, S. J. Pan, and P. S. Yu, "Adaptation regularization: A general framework for transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1076–1089, May 2014.
- [13] Z. Deng, K.-S. Choi, Y. Jiang, and S. Wang, "Generalized hidden-mapping ridge regression, knowledge-leveraged inductive transfer learning for neural networks, fuzzy systems and kernel methods," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2585–2599, Dec. 2014.
- [14] Y. Y. Liu, Z. F. Li, T. L. Yang, and Z. Bao, "An adaptively weighted least square estimation method of channel mismatches in phase for multichannel SAR systems in azimuth," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 439–443, Feb. 2014.
- [15] A. Kadkhodaie-Ilkhchi, S. T. Monteiro, F. Ramos, and P. Hatherly, "Rock recognition from MWD data: A comparative study of boosting, neural networks, and fuzzy logic," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 680–684, Oct. 2010.
- [16] W. Y. Dai, Q. Yang, G. R. Xue, and Y. Yu, "Boosting for transfer learning," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 193–200.
- [17] X. Wang and M. Han, "Online sequential extreme learning machine with kernels for nonstationary time series prediction," *Neurocomputing*, vol. 145, pp. 90–97, Dec. 2014.