Combining Multiple Classifiers for Domain Adaptation of Remote Sensing Image Classification

Hongkang Wei, Li Ma ⓘ , *Member, IEEE*, Yong Liu, and Qian Du ⓘ , *Fellow, IEEE*

Abstract—This article investigates the effectiveness of multi-classifier fusion technique on domain adaptation for remote sensing image classification. Since it is impossible to find a domain adaptation method that is optimal for different datasets, and it is also difficult to select the best base classifier for domain-invariant features, multiple domain adaptation fusion (MDAF) method and the multiple base classifier fusion (MBCF) method are proposed to achieve a more stable and superior classification performance. The most crucial step of the weighted fusion approach is to assign weights for classifiers. It is known that different classifiers have varied performances on different subsets of data, and therefore a samplewise adaptive weight is more desirable than a fixed one. For each sample, a desired weight should be able to characterize the reliability of a classifier, so that the advantages of different classifiers can be exploited. We propose a neighborhood consistency based adaptive weighting method, which assigns a large weight to a classifier on a sample if the prediction of the sample is consistent to the predictions of its local neighbors. Experiments with three remote sensing images demonstrate the efficiency of the proposed weighting strategy in the proposed MDAF and MBCF methods.

Index Terms—Classification, domain adaptation, multiclassifier fusion, neighborhood consistency, remote sensing.

I. INTRODUCTION

REMOTE sensing images are important and increasingly available for earth observation [1]. However, classification of remote sensing images faces the problem of limited labeled data. Semisupervised learning [2] and dimensionality reduction [3], [4] are popular methods to decrease the labeling cost. Alternatively, "borrowing" labeled information from some temporally or spatially separate image is an attractive strategy for classifying image with few labels or even without labels. However, directly reusing the labeled data may not perform well on the target image due to the spectral drift between the two images [5], [6]. Domain adaptation has great

Hongkang Wei is with the School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan 430074, China (e-mail: hongkong9771@outlook.com).

Li Ma and Yong Liu are with the School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan 430074, China (e-mail: mali@cug.edu.cn; 99086669@qq.com).

Qian Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA (e-mail: du@ece.msstate.edu).

potential to address this problem since it attempts to classify the target image by transferring knowledge from the source image.

Domain adaptation algorithms are often categorized as classifier based methods and feature-based methods. Classifier based approaches aim to learn an adaptive classifier for target data by using knowledge from the source data [7]–[9]. Generally, support vector machine (SVM) or extreme learning machine (ELM) classifier can be utilized. An adaptive SVM or ELM classifier can be learned by combining the source and target labeled data or introducing a feature alignment constraint to the objective function. Feature-based approaches aim to learn a common feature space where data distributions become similar between the source and target domains. In the common feature space, the classifier trained by the source labeled data can perform well on the target data. The domain invariant features can be learned by minimizing the distribution differences between domains, where the data distributions are often described by the sample means [10]–[13], sample covariance matrix [14], [15], subspace eigenvectors [16], [17], or data manifold [18]–[21]. Lately, deep learning methods have been successfully applied for domain adaptation. The maximum mean discrepancy (MMD) and correlation alignment (CORAL) strategies can be introduced into the deep network to generate both invariant and discriminate features [22], [23]. Adversarial learning is also effective for domain adaptation by using discriminators to match different domains. Ganin *et al.* proposed a domain adversarial neural network for feature alignment [24]. Pei *et al.* utilized multiple domain discriminators to achieve a conditional distribution adaptation [25].

In this article, we focus on the unsupervised domain adaptation scenario, where the source images are assumed to have abundant labeled information while the target image does not have any labeled samples. Unsupervised domain adaptation methods have been successfully applied to remote sensing field. Matasci *et al.* employed MMD strategy in transfer component analysis method to learn a shared feature space [10]. Yang *et al.* utilized manifold alignment methods for image classification, where the two feature spaces are aligned by minimizing the distances between the corresponding data pairs [18]. Sun *et al.* applied subspace alignment for scene classification of remote sensing data [26]. In our previous work, we achieved unsupervised domain adaptation by aligning the class centroids and class covariance matrix [15]. We also employed ELM and deep learning-based domain adaptation approaches for remote sensing image classification [27]–[29].

Although domain adaptation has achieved great success in classification of remote sensing images, there still exist some problems. One problem is that it is difficult to find a domain adaptation method that is optimal for different datasets. Different approaches may perform variedly on different images. Another problem is about the feature-based domain adaptation methods, which focus on the strategy to obtain domain invariant features. After obtaining a common feature space, SVM or k nearest neighbors (kNN) classifier is often used for classification. Although the selection of base classifier does not affect the comparison of different feature alignment strategies, it has influence on the classification accuracy. Therefore, how to select an optimal base classifier is a problem. For these two problems, we propose a multiclassifier fusion method, which combines the classification results of different domain adaptation approaches or different base classifiers, rather than selecting the optimal one which may not exist. Combining the classification results of multiple classifiers may exploit the advantages of different methods, and obtain a more stable and superior classification performance than any single classifier.

Many studies have indicated that the fusion of multiple classifiers can improve the classification accuracy than a single classifier. For multiple classification results, the popular fusion rules contain maximum rule, minimum rule, summation rule, production rule, median rule, weighted fusion rule, and so on [30]–[33]. Among these fusion strategies, the weighted method is the most effective. It aims to assign weights for each classifier, and the final classification result equals the weighted summation of the results from multiple classifiers. Zhao *et al.* assigned the classifier weight to be the classification accuracy on the validation data [34], [35]. Huang *et al.* proposed a P-fusion method that calculates the classifier weight as the difference in the probabilistic prediction results [32]. Gao *et al.* exploited the similarity between classification results and clustering structure to obtain the weight of a classifier [33].

In this article, we investigate multiclassifier fusion method for the domain adaptation problem. Aiming to solve the two aforementioned problems, we propose two multiclassifier fusion based algorithms for domain adaptation: one focuses on multiple domain adaptation fusion (MDAF), and the other aims to multiple base classifier fusion (MBCF) on domain invariant features. MDAF is a domain adaptation method that combines predictions from multiple domain adaptation classifiers, while MBCF is a fused base classifier that applied to the domain invariant features obtained by a feature-based domain adaptation method. In the experiments, we mainly considered four domain adaptation classifiers in this article, including joint distribution adaptation (JDA), CORAL, subspace alignment (SA), and manifold alignment (MA). In MBCF, three base classifiers are considered, i.e., SVM, kNN, and linear discriminant analysis (LDA). By combining multiple classifiers, it is expected that more useful knowledge can be transferred and a superior classification performance can be achieved.

The most crucial issue of the weighted fusion method is to assign weight for each classifier in an unsupervised manner. Since different classifiers have different properties and domain adaptation ability, they may perform variedly on different subsets of a target image. One classifier may be effective on one class, but another classifier may perform superior on a different one. Therefore, a fixed weight of a classifier is not suitable and a sample-wise adaptive weight is more desirable. For each target sample, we first search its K nearest neighbors, and then employ the neighborhood consistency criterion on their prediction results to determine the weight of each classifier. Since nearest neighbors have similar spectral properties and tend to be from the same class, the prediction result of a target sample is regarded to be reliable if its pseudolabel is consistent to the pseudolabels of its neighbors. Therefore, the classifier weight on a sample is proportional to the consistency degree in its neighborhood.

The proposed multiclassifier fusion approach offers the following properties.

1) The neighborhood consistency criterion is able to determine the effectiveness of a classifier on each specific sample. Using the sample-wise adaptive weights, the advantages of different classifiers can be explored.
2) Based on the multiclassifier fusion strategy, MDAF and MBCF algorithms are proposed to obtain a more stable and superior classification performance. Moreover, there is no strict limitation for selecting the specific classifiers.
3) To the best of our knowledge, this is the first attempt to introduce the multiple classifier fusion strategy to domain adaptation for remote sensing image classification.

The rest of this article is organized as follows. Section II introduces the domain adaptation problem. Section III presents the multiclassifier fusion approach for domain adaptation. Several related multiclassifier fusion methods are described and compared in Section IV. Experimental results are discussed in Section V and the conclusion is drawn in Section VI.

II. DOMAIN ADAPTATION FOR REMOTE SENSING IMAGES

Traditional classifier requires that the training data and testing data should have the same data distribution. For classification of remote sensing images, this requirement may not be satisfied, since spectra may vary significantly between domains. We utilize two multitemporal hyperspectral images, which were captured by EO-1 Hyperion sensor in May and June 2001 in Botswana to show the spectral drift. Fig. 1 plots the mean spectral signatures of two classes in the two images, where spectral properties of the same class have obvious differences between domains.

Domain adaptation aims to reduce the domain shift and make use of the prior knowledge of the source domain to learn a classifier for the target domain. In this article, we investigate unsupervised domain adaptation approach, where the labeled information in target domain is not available. Let $\mathbf{X}^s \in \mathrm{R}^{D \times Ns}$ and $\mathbf{X}^t \in \mathrm{R}^{D \times Nt}$ denote source and target data, respectively, where D is the spectral dimensionality, N_s denotes the number of labeled data in source domain, and N_t means the number of unlabeled data in target domain. Source data are with class labels $\mathbf{Y}^s \in \mathrm{R}^{1 \times Ns}$ while no labeled information is available for target data. Both source and target domains contain the same C classes.
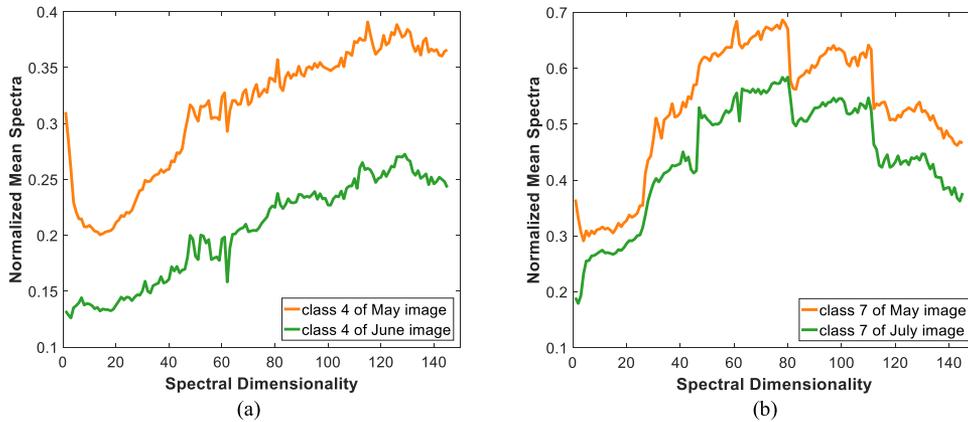
Fig. 1. Spectral signatures of the same class in different images. (a) Class 4 (Firescar) in May image and June image. (b) Class 7 (Savanna) in May image and July image.
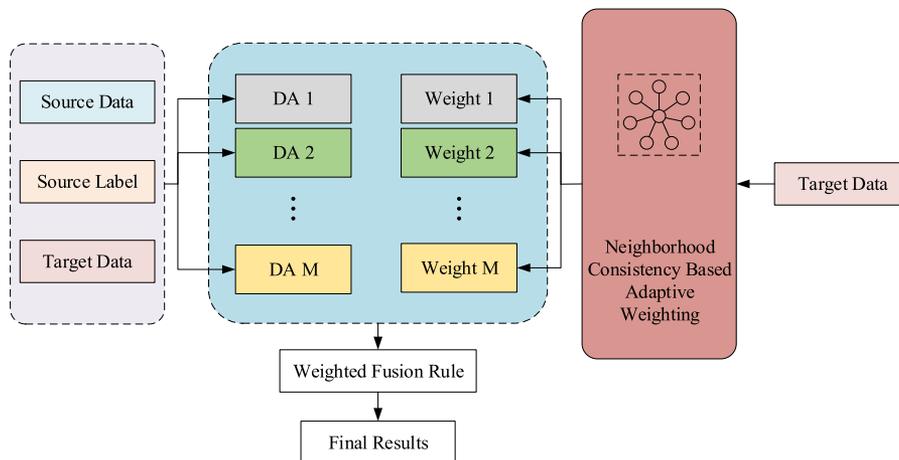
Fig. 2. Flowchart of the multiple domain adaptation fusion (MDAF) method.

We briefly describe the idea of four popular domain adaptation methods.

1) JDA approach aims to jointly adapt the marginal distribution and class-conditional distribution, which applied MMD strategy to achieve the adaptation purpose.

2) CORAL achieves domain adaptation by aligning the covariance matrix of source and target domains, which explores the second-order statistics to obtain the marginal distribution adaptation.

3) SA represents the source and target domains as subspaces spanned by eigenvectors, which aligns the source subspace eigenvectors with the target ones.

4) MA uses the corresponding data pairs as bridges to match manifolds of source and target domains, which not only aligns the two manifolds but also preserves the local geometry of each manifold.

III. PROPOSED MULTICLASSIFIER FUSION APPROACH FOR DOMAIN ADAPTATION

In this section, we first describe the two multiclassifier based methods for domain adaptation. One combines multiple domain adaptation classifiers, and the other combines multiple base classifiers on domain invariant features. Then, we provide the weighted fusion rule, where the final fused result is defined as the summation of the weighted probabilistic predictions of the classifiers. Finally, the neighborhood consistency based adaptive weighting method is presented.

A. Application of Multiclassifier Fusion to Domain Adaptation

MDAF and MBCF are two examples of using multiclassifier fusion for domain adaptation. The flowcharts of the proposed MDAF and MBCF approaches are illustrated in Figs. 2 and 3, respectively.

In Fig. 2, multiple domain adaptation classifiers are denoted as $DA_1, DA_2, \ldots, DA_M$, where M is the number of classifiers. Each domain adaptation classifier produces probabilistic prediction results of target data. Based on the neighborhood consistency weighting method, we obtain M weights. After obtaining the multiple prediction results and their associated weights, the final classification result can be achieved through the weighted fusion rule. It is worth noting that the output of each DA_m is the classification results, so the feature-based domain adaptation
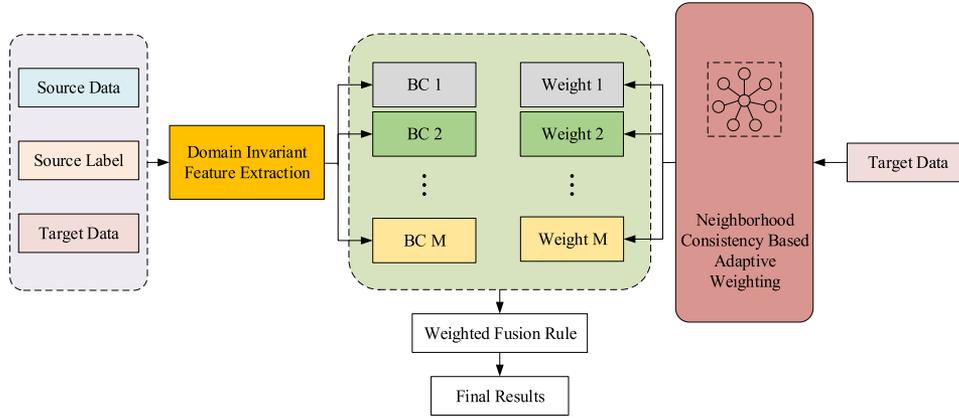
Fig. 3. Flowchart of the multiple base classifier fusion (MBCF) method.

method should be followed by a classifier like SVM to produce the classification results.

In Fig. 3, MBCF combines multiple base classifiers denoted as BC_1, BC_2, ..., BC_M. The base classifier is conducted on the domain invariant features obtained by a feature-based domain adaptation method. Similarly, the base classifiers yield probabilistic prediction results of target data, and the neighborhood consistency based adaptive weighting method is utilized to obtain the weights. It is noteworthy that invariant feature can be extracted by any feature-based domain adaptation method.

Both MDAF and MBCF utilize the multiclassifier fusion strategies, but they are applied to different scenarios. MDAF is a domain adaptation method that combines prediction results from multiple domain adaptation classifiers. On the contrary, MBCF itself is not a domain adaptation method, but is applied to classify the aligned features obtained by a feature-based domain adaptation method. In MDAF and MBCF, although there is no limitation to select the multiple domain adaptation classifiers or base classifiers, it is preferred to select classifiers with diversity, since they have different properties and can provide different and complementary knowledge.

B. Multiclassifier Fusion Rule

Multiclassifier fusion aims to combine multiple prediction results from different classifiers by assigning weight for each classifier. The final fusion result is a weighted sum of prediction results. The prediction result of a data sample can be a hard label or a soft label. The soft probabilistic prediction results contain more information than the hard label and are often used in the fusion methods.

The most crucial step is to design adaptive sample-wise weight. For a data point $\mathbf{x}$, the probabilistic prediction result from the mth classifier is denoted as $\mathbf{p}_m(\mathbf{x}) \in R^{C \times 1}$, and the weight of the mth classifier on this data point is denoted as $w_m(\mathbf{x})$. Then, the fused classification result $\mathbf{p}^f(\mathbf{x}) \in R^{C \times 1}$ is represented as

$$\mathbf{p}^f(\mathbf{x}) = \sum_{m=1}^{M} \hat{w}_m(\mathbf{x}) \times \mathbf{p}_m(\mathbf{x}) \tag{1}$$

where the normalized weight is calculated by

$$\hat{w}_m(\mathbf{x}) = \frac{w_m(\mathbf{x})}{\sum_{m=1}^{M} w_m(\mathbf{x})}. \tag{2}$$

Fig. 4 illustrates the necessity of using sample-wise adaptive weights rather than fixed ones. Taking the Hyperion BOT June-May data as an example, where June image and May image are source domain and target domain, respectively, we employed JDA and MA domain adaptation methods to classify the target data. Fig. 4(a) and (b) shows the classification results of JDA and MA domain adaptation algorithms, respectively. By plotting target data from class 3 with circles and class 5 with triangles, and by coloring the correct predictions or false predictions with green or red, the classification performance can be visualized. Although JDA has a much higher overall accuracy than MA on this data, JDA does not perform well on all the classes and MA has its advantage on some classes. From the results, JDA performs well on class 5 but yields more false predictions than MA on class 3, while MA obtains satisfactory classification on class 3 but misclassified a lot of samples in class 5. The results suggest that using fixed weight is inappropriate and a sample-wise adaptive weight can take full advantage of different classifiers.

C. Neighborhood Consistency-Based Weighting Method

We make use of target data to evaluate the domain adaptation performance. It is known that nearest neighbors have similar spectral properties and tend to be from the same class, and thus a "good" classifier should predict them as the same category. Therefore, the prediction results of target data can be utilized to evaluate the classifier. For each target data sample, if its predicted label of a classifier is consistent to the predicted labels of its nearest neighbors, the prediction result on this sample is regarded to be reliable and a large weight is assigned to the classifier. On the contrary, if most of the nearest neighbors have different pseudolabels, the prediction on this sample is considered unreliable and a small weight is set to the classifier.

For a target data point $\mathbf{x}$, its K nearest neighbors are tested. Predicted by the mth classifier, the pseudolabel of $\mathbf{x}$ is denoted as y^m, and the pseudolabels of its K nearest neighbors are $y_1{}^m$,
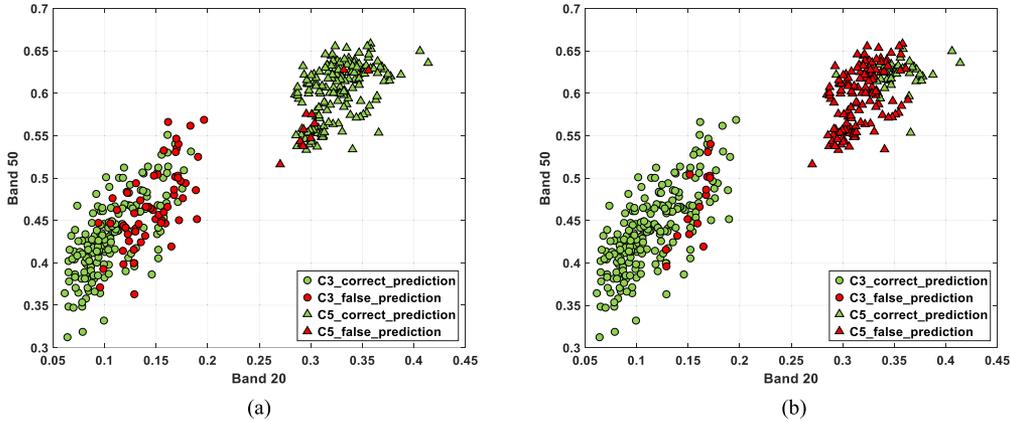
Fig. 4.   Different classifiers perform differently on different classes. (a) Prediction results of JDA for class 3 and class 5 of BOT June-May data. (b) Prediction results of MA for class 3 and class 5 of BOT June-May data.
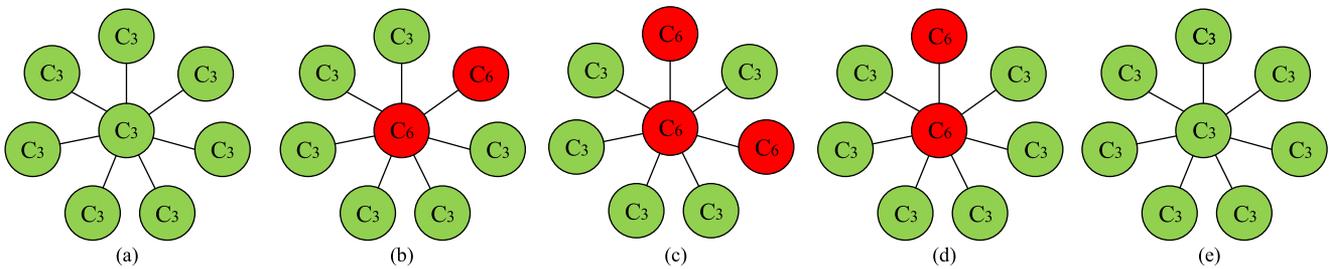
Fig. 5.   Neighborhood consistency-based weighting for multiple domain adaptation classifiers on a target sample from class 3. (a) Real labels of this neighborhood. (b) JDA misclassified the point as class 6 and its weight is 1/7. (c) CORAL yields false pseudolabel as class 6 and its weight is 2/7. (d) SA misclassified the point as class 6 and its weight is 1/7. (e) MA achieves correct prediction and its weight is 1.

$y_2^m, \ldots, y_K^m$. The neighborhood consistency based weight for the mth classifier is calculated as

$$w_m(\mathbf{x}) = \frac{\sum_{k=1}^{K} \delta(y_k^m = y^m)}{K} \quad (3)$$

where δ is the Kronecker delta. Two methods are used to select neighbors, one is based on the spectral similarity measured by Euclidean distance, and the other is based on the spatial distance. In this article, the spatial neighbors of a data point $\mathbf{x}$ are defined as pixels within an $\mathbf{x}$-centered squared spatial window. It is worth mentioning that a small window is preferred, since a large spatial window may include neighbors that are from different classes and cannot satisfy the neighborhood consistency constraint.

Fig. 5 illustrates the neighborhood consistency based weighting method. The four domain adaptation methods include JDA, CORAL, SA, and MA. For a target sample from class 3, its seven nearest neighbors are selected. Fig. 5(a) shows the real labels of this neighborhood, where all the neighbors belong to class 3 and are denoted as green nodes. Fig. 5(b) shows the prediction results of JDA on this neighborhood, where the green neighbors denote correct predictions as class 3 and red neighbor represents false prediction to another class. JDA misclassified this sample as class 6. Since most neighbors have different predictions and only one neighbor has the same prediction, according to formula (3), the weight for JDA is 1/7. Fig. 5(c) and (d) illustrates the classification results of CORAL and SA, respectively. Both CORAL and SA yield false pseudolabel for

this point, and the weights are 2/7 and 1/7 for them, respectively. Fig. 5(e) shows the results of MA, which achieves correct predictions for this point and all its neighbors. Therefore, the weight for MA is 1. From the results, it can be observed that based on the neighborhood consistency criterion, the classifier with false prediction on a sample has a small weight and the classifier with correct prediction is assigned with a large weight. The weighting method is able to effectively evaluate the performance of different classifiers.

The proposed multiclassifier fusion algorithm is described in the following. When the classifiers are domain adaptation approaches, the algorithm is called as MDAF, and when the classifiers are base classifiers for classifying domain-invariant features, it is referred to as MBCF.

IV. RELATED WORKS AND DISCUSSION

The following discussions describe some related multiclassifier fusion strategies.

A. Traditional Fusion Operations on the Probabilistic Prediction Results

The traditional multiclassifier fusion strategies [30], [31] generally conduct some operations on the probabilistic prediction results of multiple classifiers, including maximum rule, minimum rule, sum rule, product rule, and median rule. For a data point $\mathbf{x}$, let $p_{m,c}(\mathbf{x})$ characterize the probability of $\mathbf{x}$ belonging

---

Algorithm: Multiclassifier Fusion Approach.

Input: 1) Labeled source data { $\mathbf{X}^s$, $\mathbf{Y}^s$ } and unlabeled target data $\mathbf{X}^t$.
     2) M classifiers (DA_1, DA_2, …, DA_M in MDAF or BC_1, BC_2, …, BC_M in MBCF).

Procedure:
1. Obtain the pseudo-labels and probabilistic predictions of target data by each classifier.
2. Calculate the weight $w_m(\mathbf{x})$ on each target data $\mathbf{x} \in \mathbf{X}^t$ for the m-th classifier, $m = 1$, …, M, according to (3). Normalize the classifier weight by (2).
3. Combine the multiple classifiers and obtain the final probabilistic prediction of each target data by using (1).

Output: The predicted labels of target data.

---

to the cth class predicted by the mth classifier. The probabilistic prediction matrix $\mathbf{P}(\mathbf{x})$ of all the classifiers on $\mathbf{x}$ is denoted as

$$\mathbf{P}(\mathbf{x}) = \begin{bmatrix} p_{1,1}(\mathbf{x}) & \cdots & p_{1,C}(\mathbf{x}) \\ \vdots & p_{m,c}(\mathbf{x}) & \vdots \\ p_{M,1}(\mathbf{x}) & \cdots & p_{M,C}(\mathbf{x}) \end{bmatrix} \quad (4)$$

where each row denotes the probabilistic prediction vector of one classifier on sample $\mathbf{x}$, and each column represents the probabilities of $\mathbf{x}$ belonging to the cth class predicted by all the classifiers.

Suppose the final fused prediction result on sample $\mathbf{x}$ is denoted as $\mathbf{p}^f(\mathbf{x}) = [p_1^f(\mathbf{x}), p_2^f(\mathbf{x}), ..., p_C^f(\mathbf{x})]$, where $p_c^f(\mathbf{x})$ represents the fused probability of $\mathbf{x}$ belonging to the cth class.

The maximum (Max) rule defines $p_c^f(\mathbf{x})$ as

$$p_c^f(\mathbf{x}) = \max_{m=1,2,...,M} p_{m,c}(\mathbf{x}) \quad c = 1, 2, \ldots, C. \quad (5)$$

The minimum (Min) rule is defined as

$$p_c^f(\mathbf{x}) = \min_{m=1,2,...,M} p_{m,c}(\mathbf{x}) \quad c = 1, 2, \ldots, C. \quad (6)$$

The sum rule (Sum) summarizes the probabilities in each column and is given as

$$p_c^f(\mathbf{x}) = \sum_{m=1}^{M} p_{m,c}(\mathbf{x}) \quad c = 1, 2, \ldots, C. \quad (7)$$

The product rule (Product) regards the multiplication of the probabilities in each column as the fused probability of $\mathbf{x}$ belonging to each class

$$p_c^f(\mathbf{x}) = \prod_{m=1}^{M} p_{m,c}(\mathbf{x}) \quad c = 1, 2, \ldots, C. \quad (8)$$

The median rule (Median) selects the median value in each column and is defined as

$$p_c^f(\mathbf{x}) = \underset{m=1,2,...,M}{\text{median}} \, p_{m,c}(\mathbf{x}) \quad c = 1, 2, \ldots, C. \quad (9)$$

These fusion rules conduct different operations on the probabilistic results. Differently, we aim to fuse the classifiers by assigning weights to them.

B. Global Weight for Each Classifier

Many studies [34], [35] calculate global weights for classifiers. They generally divide the labeled data into training data and validation data. Training data are used to train a classifier and validation data are used to calculate the classification accuracy. The weight for the classifier is directly proportional to the overall accuracy on the validation data. For unsupervised domain adaptation, target domain does not include labeled samples. If we obtain the validation data from labeled source data, the classification accuracy on validation data cannot represent the performance on target data since they have different data distributions. Therefore, the accuracy based global weight cannot be applied to domain adaptation problem.

C. Locally Weighted Ensemble (LWE) Approach

LWE [33] is a weighted fusion method, which calculates the weight of a classifier via constructing two graphs $\mathbf{G}_M$ and $\mathbf{G}_T$ on the test samples. $\mathbf{G}_M$ is constructed on the prediction results of the classifier, where the samples are connected if they are classified into the same class. $\mathbf{G}_T$ is constructed on the clustering results obtained by k-means method, where the samples in the same cluster are connected. The classifier weight on a sample $\mathbf{x}$ is proportional to the similarity of its local graph between $\mathbf{G}_M$ and $\mathbf{G}_T$

$$w_m(\mathbf{x}) = \frac{\sum_{v_1 \in V_M} \sum_{v_2 \in V_T} 1(v_1 = v_2)}{|V_M| + |V_T|} \quad (10)$$

where $\mathbf{V}_M$ and $\mathbf{V}_T$ are the sets of connected neighbors of $\mathbf{x}$ in $\mathbf{G}_M$ and $\mathbf{G}_T$. The more common neighbors exist in the two local neighborhood, the higher weight we set to the classifier.

Note that our weighting strategy is similar to the LWE approach, since both of them utilize the prediction results in a local neighborhood. However, there are two differences. One is that LWE utilizes k-means method to obtain the local clusters while we employ K spectral neighbors or spatial neighbors from a spatial square window to construct the local neighborhood. The other difference is that the LWE is for text classification, spam filtering, and intrusion detection, while our approach is for remote sensing image classification. Moreover, we have demonstrated the efficiency of the neighborhood consistency constraint in some previous works [12], [36]. One utilizes this constraint for obtaining a superior translating vector to move the target samples to the source domain [12]. One exploits this constraint to select more reliable pseudolabels to construct the relationship matrix between two domains in manifold alignment framework [36]. We further explore the application of the constraint to multiclassifier fusion approach in this article.

D. P-Fusion Approach

Reference [32] proposed an SVM ensemble approach for combining spectral, structure, and semantic features of remote
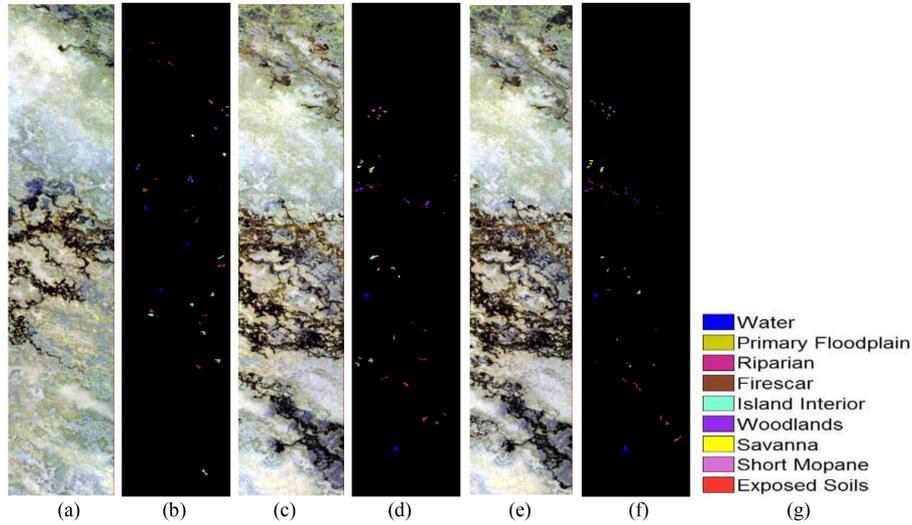
Fig. 6.   BOT data in May, June, and July. (a) BOT image in May. (b) Ground truth of image in May. (c) BOT image in June. (d) Ground truth of image in June. (e) BOT image in July. (f) Ground truth of image in July. (g) Class legend.

sensing data. A P-fusion approach is utilized to combine multiple SVM classifiers, where the weight is calculated by the SVM-based posterior probability. For the mth SVM classifier, its probabilistic prediction results $\hat{p}_{m,1}(\mathbf{x}), \hat{p}_{m,2}(\mathbf{x}), \ldots, \hat{p}_{m,C}(\mathbf{x})$ on a testing sample $\mathbf{x}$ are arranged in a descending order. Then, the weight of this classifier is defined as

$$w_m(\mathbf{x}) = \sum_{c=1}^{C-1} \left[\hat{p}_{m,c}(\mathbf{x}) - \hat{p}_{m,c+1}(\mathbf{x}) \right] \cdot \frac{1}{c}. \quad (11)$$

Both the P-fusion approach and our proposed fusion approach calculate weights for classifiers, but the strategies to obtain the weights are different. Moreover, the P-fusion approach was not applied to domain adaptation problem in [32].

V. EXPERIMENT RESULTS AND ANALYSIS

A. Data Description

Hyperspectral images from three sensors were exploited for experiments. One was obtained by Hyperion instrument of NASA EO-1 satellite, one by the NASA Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) instrument, and the third by the Digital Globe Worldview-2 satellite. Source and target data were chosen from different temporal images or spatially separate images.

The NASA EO-1 satellite collected 242-band data at a 30-m spatial resolution over a 7.7-km strip covering 357–2576-nm portion of the spectrum in 10-nm spectral resolution. Uncalibrated and noisy bands that cover water absorption features were removed, and the remaining 145 bands were used for the analysis. The experimental data with 1476×256 pixels were acquired over the Okavango Delta, Botswana (BOT) in May, June, and July 2001. All of them contain nine common classes. Source and target images can be obtained by choosing any two of the three images, and then we can get six data pairs for experiments. Three pseudocolor images and label information are shown in Fig. 6.

The NASA AVIRIS instrument acquired 224-band data at an 18-m spatial resolution from an altitude of approximately 20 km, which covers 400–2500-nm portion of the spectrum in 10-nm spectral resolution. After removing water absorption and low SNR bands, the remaining 176 bands were used for experiments. The experimental data with 512×614 were collected over the Kennedy Space Center (KSC), Florida, on March 23, 1996. The spatially separated images are denoted as KSC1 and KSC2. Both of them contain ten identified classes. Two pseudocolor images and label information are shown in Fig. 7.

The Digital Globe Worldview-2 satellite collected 8-band data including red, green, blue, near-infrared 1, coastal, yellow, red edge and near-infrared 2 at a 1.8-m spatial resolution, which covers 400–1040-nm portion of the spectrum. Two images acquired in July 2011 and July 2012 over the same area in Wuhan, China, were used for the domain adaptation. The images with ground reference are displayed in Fig. 8. The class names and the number of samples in each class of these images are listed in Table I.

We chose ten data pairs for experiments. For BOT data, six data pairs are denoted as May-June, June-May, May-July, July-May, June-July, and July-June. For KSC data, two data pairs were used and named as KSC1-KSC2 and KSC2-KSC1. For Worldview-2 data, two data pairs (2011-2012 and 2012-2011) were used for evaluating domain adaptation methods. We randomly chose 400 samples per class from each Worldview-2 image as source and target data. Twenty replications of the experiments were conducted, and the mean accuracy was calculated for evaluation. In each data pair, the first and the second datasets are denoted as source and target domain, respectively.

B. Compared Multiclassifier Fusion Strategies

The proposed neighborhood consistency weighting based fusion methods were compared with seven multiclassifier fusion strategies, including Max, Min, Sum, Product, Median [30], [31], LWE [33], and P-fusion method [32]. For our neighbor
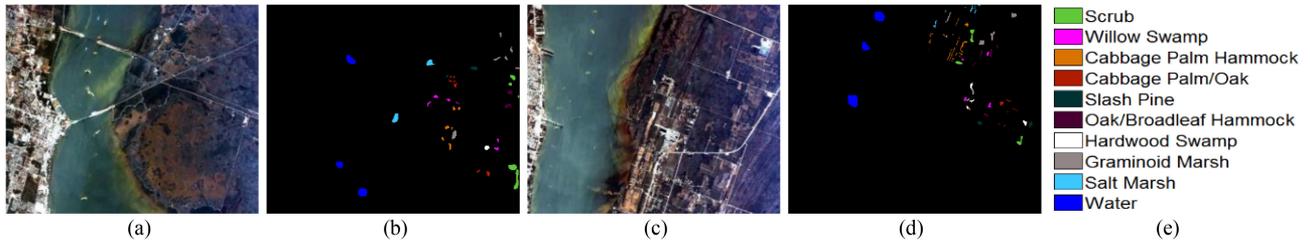
Fig. 7. KSC data. (a) KSC1. (b) Ground truth of KSC1. (c) KSC2. (d) Ground truth of KSC2. (e) Class legend.
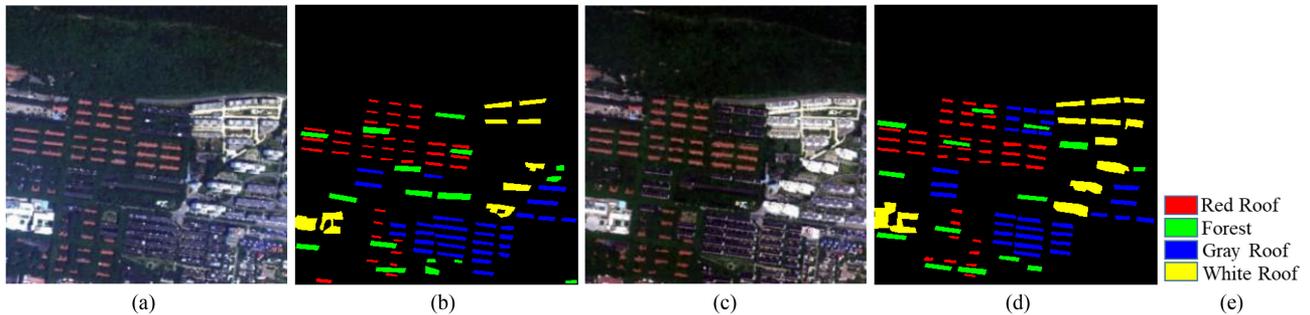
Fig. 8. Worldview-2 data in 2011 and 2012. (a) Worldview-2 image in 2011. (b) Ground truth of image in 2011. (c) Worldview-2 image in 2012. (d) Ground truth of image in 2012. (e) Class legend.

TABLE I
CLASS NAMES AND THE NUMBER OF SAMPLES OF BOT, KSC, AND WORLDVIEW-2 IMAGES

BOT					KSC				Worldview-2			
ID	Class Name	May	June	July	ID	Class Name	KSC1	KSC2	ID	Class Name	2011	2012
1	Water	158	195	185	1	Scrub	761	422	1	Red Roof	2511	2963
2	Primary Floodplain	228	192	96	2	Willow Swamp	243	180	2	Forest	3592	3144
3	Riparian	237	179	164	3	Cabbage Palm Hammock	256	431	3	Gray Roof	4425	4528
4	Firescar	178	196	186	4	Cabbage Palm/Oak	252	132	4	White Roof	3082	5301
5	Island Interior	183	197	131	5	Slash Pine	161	166				
6	Woodlands	199	218	169	6	Oak/Broadleaf Hammock	229	274				
7	Savanna	162	189	171	7	Hardwood Swamp	105	248				
8	Short Mopane	124	166	152	8	Graminoid Marsh	431	453				
9	Exposed Soils	111	156	96	9	Salt Marsh	419	156				
					10	Water	927	1392				

consistency based weighting method, there are two ways to select neighbors, and thus both spectral and spatial neighbors are utilized in the experiments. The MDAF with the two neighbor selection methods are denoted as MDAF_Spe and MDAF_Spa, respectively. The MBCF with the two neighbor selection methods are denoted as MBCF_Spe and MBCF_Spa, respectively.

There is no free parameter in Max rule, Min rule, Sum rule, Product rule, Median rule, or the P-fusion rule. LWE contains one parameter, which denotes the number of clusters. The value of c' was tested from 2 to 10 as recommended by the reference [33]. The proposed fusion approach contains one parameter, which is the number of neighbors in the neighborhood consistency weighting method. When the spectral neighbors were

used, different values of K were tested, which are from 5 to 49 with a step of 2. When the spatial neighbors are utilized, three different sizes of spatial windows were tested, including 3×3, 5×5, and 7×7.

C. Results of MDAF Approach

In MDAF, we mainly considered four popular feature-based domain adaptation methods, which are JDA [37], CORAL [14], SA [26], and MA [38]. We also tested the efficiency of MDAF on three deep adaptation networks, including deep adaptation network (DAN) [22], multiple adversarial domain adaptation network (MADA) [25], and correlation alignment for deep

TABLE II
OA (%) OF MDAF AND THE COMPARED FUSION STRATEGIES

Data sets		BOT							KSC			Worldview-2		
		May-June	June-May	May-July	July-May	June-July	July-June	average	KSC1-KSC2	KSC2-KSC1	average	2011-2012	2012-2011	average
DA Algorithms	JDA	89.69	81.39	90.30	80.63	90.81	91.82	87.44	64.58	68.21	66.40	95.53	93.79	94.66
	CORAL	87.09	80.76	87.78	76.77	94.15	92.42	86.49	66.61	60.78	63.69	88.10	94.38	91.24
	SA	85.55	81.20	85.33	76.33	92.96	90.52	85.32	64.53	60.07	62.30	95.08	93.41	94.25
	MA	77.96	71.46	77.19	66.84	93.04	88.21	79.11	62.92	**69.56**	66.24	95.62	93.10	94.36
Max		88.86	81.71	90.22	81.01	94.81	93.42	88.34	67.46	64.64	66.05	95.48	94.41	94.95
Min		89.16	80.89	90.07	79.68	95.11	93.19	88.02	67.49	65.04	66.27	95.09	94.32	94.71
Sum		89.87	83.42	89.78	81.14	94.52	92.89	88.60	67.23	67.18	67.21	96.22	94.43	95.33
Product		89.87	82.78	90.00	80.89	94.52	92.89	88.49	67.28	66.38	66.83	96.03	94.46	95.25
Median		89.69	84.24	89.70	80.38	94.07	92.59	88.45	66.74	67.34	67.04	96.21	94.35	95.28
LWE		89.63	84.24	89.04	81.77	94.15	92.30	88.52	68.68	64.93	66.81	96.42	94.29	95.36
P-fusion		89.51	83.04	90.00	81.46	94.44	92.89	88.56	**69.05**	64.83	66.94	96.12	94.42	95.27
MDAF_Spe		90.40	84.62	90.59	82.28	94.81	93.01	89.29	68.42	66.54	67.48	96.44	94.55	95.50
MDAF_Spa		**91.41**	**85.25**	**91.70**	**83.29**	**95.78**	**93.96**	**90.23**	67.96	67.73	**67.84**	**96.47**	**94.79**	**95.63**

domain adaptation (D_CORAL) [23]. It should be noted that the four feature-based methods aim to achieve domain invariant features. After obtaining the common features, SVM is applied for classification purpose, although other classifiers can also be utilized. The selected base classifier will not affect the fusion of different domain adaptation methods, since it is the diversity of multiple domain invariant features that results in the effectiveness of the fusion.

Five experiments were conducted. First, MDAF was employed on four feature-based domain adaptation methods. Second, three of the four feature-based methods were selected to further illustrate the fusion performance. Third, the classification accuracy on each class was illustrated to observe how the MDAF algorithm integrates the advantages of different domain adaptation classifiers. Fourth, the classification maps of Worldview-2 image were presented. Fifth, MDAF was applied to fuse the results of three deep adaptation networks.

1. MDAF on Four Feature-Based Domain Adaptation Methods: JDA has two parameters: λ is the regularization parameter and p is the dimensionality of the common feature space. CORAL is parameter-free. SA has one parameter p, which denotes the dimensionality of the common feature space. MA has two parameter: σ is the heat kernel parameter and p is the dimensionality of the common feature space. The two parameters of JDA were fixed to $\lambda = 0.001$ and $p = 50$. The parameter of SA was fixed to $p = 10$. The two parameters of MA were fixed to $\sigma = 0.01$ and $p = 50$. For Worldview-2 multispectral images, the dimensionality is fixed to be 8 for all the four methods. Note that optimal parameters are not selected for these methods, since the purpose here is to evaluate the fusion performance. The fusion can be demonstrated to be effective if the fused accuracy is higher than the accuracy of each individual classifier.

The overall accuracy (OA) of the four domain adaptation algorithms and the multiclassifier fusion results with nine fusion

strategies on the ten datasets are listed in Table II. Several observations can be drawn from comparing these methods.

1) Among the four domain adaptation methods, we cannot find an optimal one for all the datasets. Different approaches performance varied on different images. In general, JDA has the best performance while MA yields the lowest accuracies on most of the data pairs.

2) Most of the fusion methods outperform the single domain adaptation classifier, indicating that fusing multiple results can obtain a more stable and superior performance.

3) Compared with the other fusion methods, MDAF_Spe and MDAF_Spa yield superior performances on most of the datasets. It indicates that the proposed fusion strategy that utilizes the neighborhood consistency based weights can better evaluate the classifier performance and provide more effective weights.

4) For KSC data, the proposed MDAF_Spe and MDAF_Spa achieve the highest average overall accuracies. P-fusion method obtains the highest accuracy on KSC1-KSC2 data pairs but its accuracy on KSC2-KSC1 is low. MA domain adaptation method achieves the best performance on KSC2-KSC1.

5) MDAF_Spa obtains higher accuracies than MDAF_Spe on most of the datasets. The reason may be that spatial neighbors are more likely from the same class and false neighbors with similar spectra can be avoided.

2. MDAF on Three Feature-Based Domain Adaptation Methods: To further evaluate the performance of MDAF method, we conducted MDAF on three domain adaptation classifiers. Six BOT data pairs were utilized for illustration. Fig. 9 shows the results of the selected domain adaptation methods and the fused results of MDAF, where (a) shows the fused results of CORAL, SA, MA, (b) illustrates the fused results of JDA, SA, and MA, (c) shows the performance of fusing JDA, CORAL, and MA, and (d) plots the fused results of JDA, CORAL, and SA. The
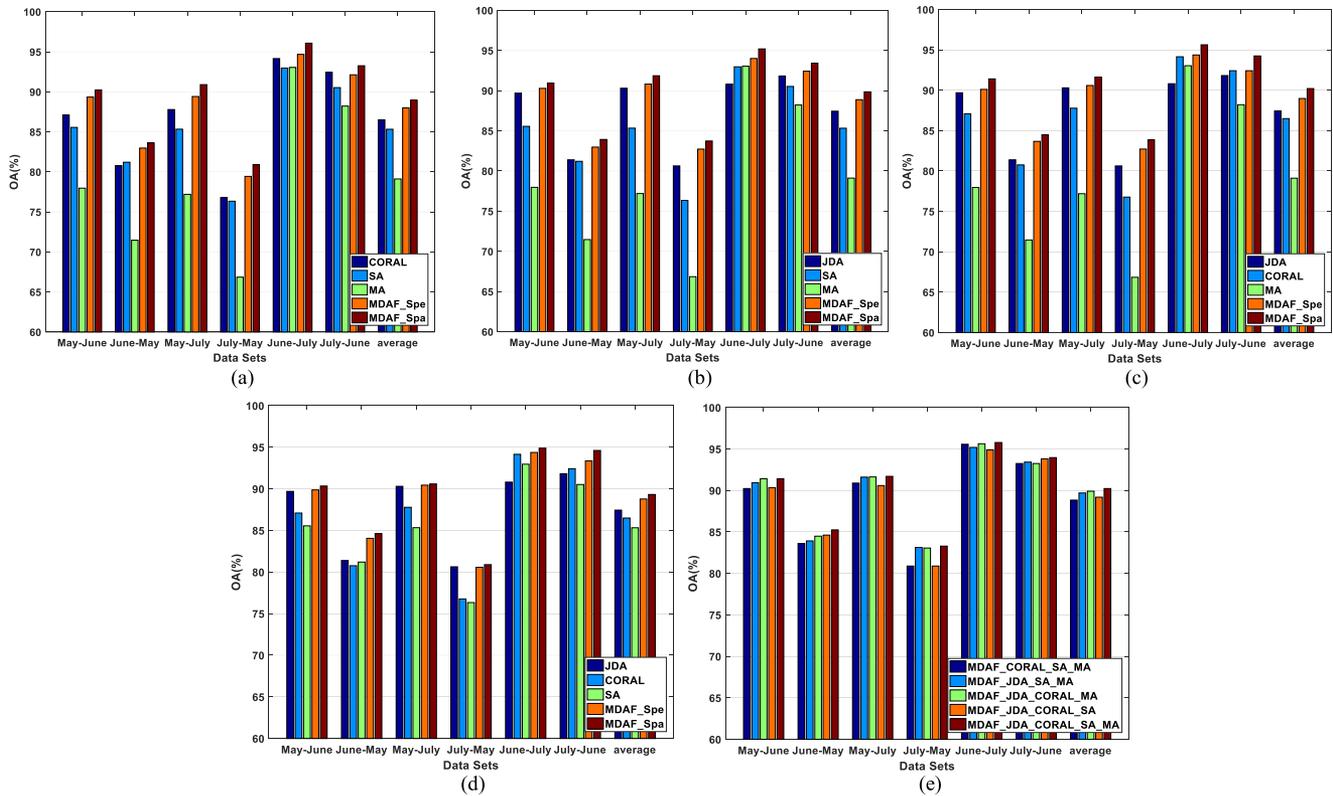
Fig. 9. MDAF on three feature-based domain adaptation methods and comparison with fusing four methods using BOT images. (a) MDAF on CORAL, SA, and MA. (b) MDAF on JDA, SA, and MA. (c) MDAF on JDA, CORAL, and MA. (d) MDAF on JDA, CORAL, and SA. (e) Comparison of MDAF on three or four feature-based domain adaptation methods.

effectiveness of MDAF can be demonstrated no matter what the three classifiers are. MDAF achieves higher accuracies than any single classifier on most of the datasets. Moreover, MDAF_Spe outperforms MDAF_Spe on all the datasets. We can also observe that for some datasets like June-May and July-May, the accuracies of MA are quite low (71.46% and 66.84%). However, the low accuracies of MA do not affect the performance of fusion. MDAF_Spa obtained 13%–17% improvements with respect to MA.

To observe the influence of the selected domain adaptation methods in MDAF, we picked the results of MDAF_Spa from Table II and Fig. 9(a)–(d), and plotted them in the same figure in Fig. 9(e). The fusion results on three classifiers or four classifiers are similar on most of the datasets. In addition, the MDAF on four classifiers obtains a slightly superior performance compared to the results of MDAF on three classifiers. They indicate that the number of domain adaptation methods do not significantly affect the fusion performances.

3. Classification Accuracy on Each Class: To better understand how the MDAF algorithm integrates the advantages of different domain adaptation algorithms, Table III shows the per-class classification accuracies using BOT June-May data pair. MDAF_Spa(3) denotes the MDAF_Spa on three feature-based domain adaptation algorithms (JDA, CORAL, and SA), while MDAF_Spa(4) denotes MDAF_Spa that fuses all the four feature-based methods. It can be observed that MA algorithm obtains the lowest OA. However, MA does not influence the

fusion performance but provides positive effect. By fusing MA with the other three classifier, MDAF_Spa(4) obtains a further improvement compared to MDAF_Spa(3). From observing the accuracies of each class, although the OA of MA is the lowest, MA obtains much higher accuracies on class 3 and class 7 than the other three domain adaptation methods. It achieves 15%–23% improvements on class 3 and 24%–47% improvements on class 7 with respect to JDA, CORAL, and SA. The good performance of MA on the two classes provides a positive effect to the fusion. We can also see that the accuracy of MA on class 5 is quite low (24.04%), but MDAF is not affected by this low accuracy and can still obtain 90.16% result on this class. This is because the other three classifiers obtain satisfactory classification on class 5, and our fusion method can well evaluate the effectiveness of these classifiers.

4. Classification Results of the Whole Image by the MDAF Approach: We chose the Worldview-2 "2011-2012" data pair to illustrate the classification performance on the whole images. It is worth noting that the BOT and KSC images were obtained from upland and wetland areas, which are not easy to visualize the differences in the whole classification maps. Thus, we only provided the classification maps of Worldview-2 images. The classification maps are shown in Fig. 10, where (a)–(d) denote the results of SVM classification on features generated by JDA, CORAL, SA, and MA, respectively, and (e) shows the result of the proposed MDAF_Spa method. Since there is no ground truth for the whole image, the reference image was obtained by

TABLE III
PER-CLASS CLASSIFICATION ACCURACY OF BOT JUNE-MAY DATA PAIR WITH THE MDAF ALGORITHM

ID	Class Name	JDA	CORAL	SA	MDAF_Spa(3)	MA	MDAF_Spa(4)
1	Water	98.73	99.37	97.47	99.37	100.00	100.00
2	Primary Floodplain	49.12	63.60	63.16	60.96	41.67	58.77
3	Riparian	75.11	70.04	78.48	81.01	93.25	86.92
4	Firescar	98.88	93.82	88.20	97.75	96.63	98.31
5	Island Interior	93.99	92.90	85.79	96.17	24.04	90.16
6	Woodlands	86.43	93.97	87.44	96.98	57.29	94.97
7	Savanna	64.20	41.36	58.02	54.94	88.27	63.58
8	Short Mopane	100.00	99.19	99.19	100.00	91.13	100.00
9	Exposed Soils	82.88	84.68	84.68	83.78	62.16	83.78
	OA%	81.39	80.76	81.20	84.62	71.46	**85.25**
	AA%	83.26	82.10	82.49	85.66	72.72	**86.28**
	kappa	0.7898	0.7829	0.7877	0.8263	0.6774	**0.8334**
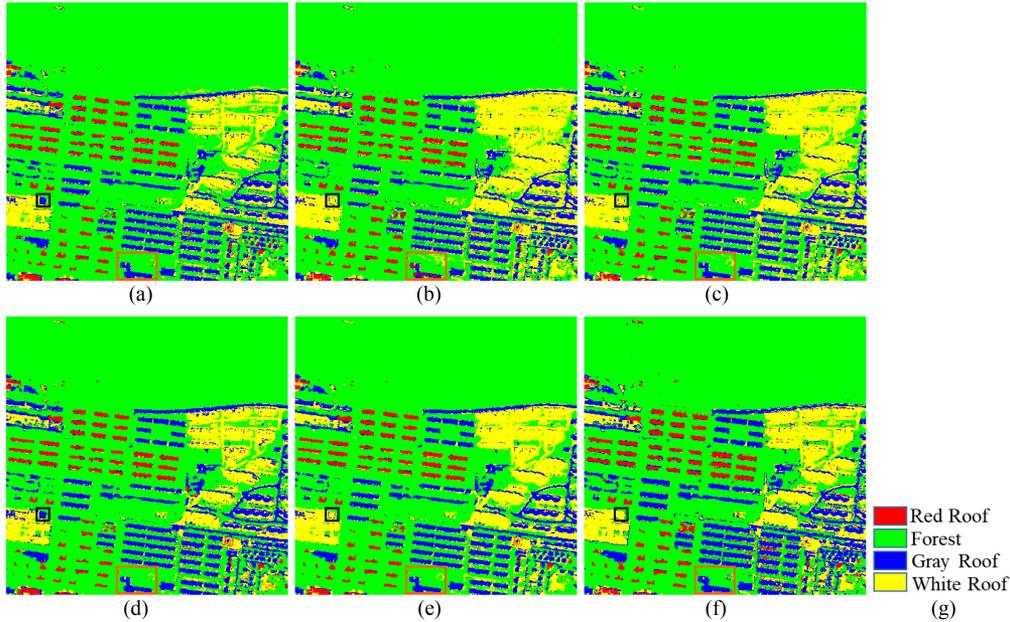
Fig. 10. Classification results of Worldview-2 "2011–2012" data pair. (a) JDA result. (b) CORAL result. (c) SA result. (d) MA result. (e) MDAF result. (f) Reference obtained by SVM using target labeled data as training data. (g) Class legend.

the SVM classifier trained on the target labeled data, as shown in Fig. 10(f). For a better comparison, we selected two local regions, which are denoted with black and red windows. From comparing the black windows in Fig. 10(a)–(d), JDA and MA misclassified many pixels of white roof (yellow) as gray roof (blue), while CORAL and SA obtained correct classification. The fused result of MDAF in Fig. 10(e) is correct, indicating its effective evaluations of the four classifiers. From observing the red windows, CORAL misclassified many pixels of gray roof (blue) as red roof (red), and some pixels of forest (green) were also misclassified as white roof (yellow). MDAF was not influenced by the errors of CORAL and obtained correct classification.

5. MDAF on Three Deep Domain Adaptation Networks: MDAF can be utilized to combine multiple domain adaptation methods, including both feature-based methods and classifier-based methods. We employed MDAF to fuse three deep learning based domain adaptation methods using BOT data in this experiments. Deep learning-based domain adaptation methods are most popular nowadays and can simultaneously obtain aligned features and adaptive classifier. They can be regarded as classifier-based methods. We selected three deep learning-based domain adaptation methods, including DAN that utilized MMD domain adaptation strategy, MADA that employed multiple domain discriminator to obtain class-wise feature alignment, and D_CORAL that introduced correlation alignment strategy

TABLE IV
OA (%) OF DIFFERENT FUSION METHODS ON THREE DEEP ADAPTATION NETWORKS

BOT	Deep Adaptation Networks			Max	Min	Sum	Product	Median	LWE	P-fusion	MDAF_Spe	MDAF_Spa
	DAN	MADA	D_CORAL									
May-June	88.21	90.46	89.16	90.34	90.34	89.99	90.05	89.57	90.17	89.81	90.88	**91.17**
June-May	84.87	81.65	85.00	87.03	87.47	86.90	87.15	85.76	87.15	86.65	87.66	**88.48**
May-July	87.19	90.22	87.93	90.07	90.22	89.93	90.15	88.00	90.22	89.48	90.37	**90.52**
July-May	74.94	75.76	76.20	76.71	76.46	76.71	76.77	76.14	76.46	76.27	**77.47**	77.09
June-July	91.85	94.30	92.59	94.44	94.44	94.15	94.30	92.37	94.15	94.00	94.07	**94.74**
July-June	88.80	89.04	87.86	89.40	89.16	88.98	88.98	88.45	89.16	88.45	89.16	**89.45**
average	85.98	86.90	86.46	88.00	88.02	87.78	87.90	86.72	87.88	87.44	88.27	**88.58**

TABLE V
OA (%) OF MBCF AND COMPARED FUSION STRATEGIES ON THE JDA PRODUCED FEATURES

Data sets			BOT							KSC			Worldview-2		
			May-June	June-May	May-July	July-May	June-July	July-June	average	KSC1-KSC2	KSC2-KSC1	average	2011-2012	2012-2011	average
JDA	Base Classifiers	LDA	86.49	83.86	86.07	80.70	92.59	92.00	86.95	62.61	57.98	60.30	93.71	93.33	93.52
		kNN	88.92	79.11	88.30	76.39	90.07	90.34	85.52	65.10	**63.45**	64.28	91.46	86.37	88.91
		SVM	89.34	82.28	90.37	74.37	91.70	92.00	86.68	64.69	62.55	63.62	95.13	92.75	93.94
	Max		88.57	83.99	89.56	80.13	92.37	92.54	87.86	64.06	60.73	62.40	94.84	90.77	92.81
	Min		88.98	84.37	88.74	78.80	92.52	92.59	87.67	63.80	61.21	62.51	94.85	90.91	92.88
	Sum		89.16	84.68	90.44	79.43	92.22	92.89	88.14	64.92	62.63	63.78	95.21	93.18	94.20
	Product		88.86	84.24	89.41	78.67	92.30	92.77	87.71	64.30	61.18	62.74	94.94	90.90	92.92
	Median		90.34	83.16	90.15	78.42	91.33	92.24	87.61	65.02	63.16	64.09	95.31	93.33	94.32
	LWE		90.05	86.20	90.74	80.38	91.19	92.89	88.57	65.36	62.39	63.88	95.19	93.47	94.33
	P-fusion		88.68	84.24	90.07	79.87	92.59	92.89	88.06	64.30	62.24	63.27	95.04	92.97	94.01
	MBCF_Spe		89.87	86.01	**92.60**	**82.03**	92.52	92.42	89.24	**66.06**	62.98	**64.52**	**95.61**	93.23	94.42
	MBCF_Spa		**91.17**	**86.33**	91.85	81.90	**92.96**	**93.25**	**89.58**	66.04	63.00	**64.52**	95.57	**93.53**	**94.55**

in a neural network. For the parameters, DAN has a parameter representing the weight of the multiple-kernel MMD loss, MADA contains a weight of the domain classification loss, and the D_CORAL includes a weight of the CORAL loss. These parameters were chosen as the recommended values in [22], [25], and [23], respectively. The classification results of the three algorithms and the nine fusion strategies on BOT datasets are listed in Table IV. Similar observations can also be drawn as the fusion of feature-based methods. Both MDAF_Spe and MDAF_Spa outperformed the three adaptation networks, and obtained superior performance to other fusion methods.

D. Result of MBCF Approach

In MBCF, three base classifiers are utilized, which are LDA [39], kNN, and SVM. LDA and kNN were implemented by using the built-in functions of MATLAB. The number of the neighbors in kNN is fixed to 7 in all the experiments, and it can also be set to other values which will not influence the MBCF performance. The SVM classifier was realized by the libsvm toolbox, and fivefold cross validation is used to obtain the optimal parameters.

MBCF is employed to obtain a superior fused base classifier for domain invariant features, which can be obtained by

any feature-based domain adaptation method. We utilized JDA, CORAL, SA, and MA to generate the domain invariant features, employed LDA, kNN, and SVM for classification, and applied MBCF to fuse the three classification results for each domain invariant features. In MBCF, the parameter $\lambda = 0.1$ in JDA and $\sigma = 0.2$ in MA. The dimensionality is fixed to be 50 for BOT and KSC hyperspectral data and 8 for Worldview-2 multispectral images.

The results are listed in Tables V, VI, VII, and VIII. Similar observations can also be drawn as the MDAF approach. The base classifiers perform variedly on different datasets, and it is impossible to select one that is optimal to all the data. All the nine fusion methods are able to obtain superior performances than a single base classifier on most of the datasets, suggesting the effectiveness of fusion. MBCF_Spe and MBCF_Spa obtain higher accuracies than the other seven multiclassifier fusion strategies on most of the datasets, indicating the advantage of the neighborhood consistency based weighting method. MBCF_Spa outperforms MBCF_Spe, suggesting the advantage of using spatial neighbors.

From comparing the results of MBCF on different domain invariant features, it can be observed that the performance of

TABLE VI
OA (%) OF MBCF AND COMPARED FUSION STRATEGIES ON THE CORAL PRODUCED FEATURES

Data sets			BOT							KSC			Worldview-2		
			May-June	June-May	May-July	July-May	June-July	July-June	average	KSC1-KSC2	KSC2-KSC1	average	2011-2012	2012-2011	average
CORAL	Base Classifiers	LDA	84.48	81.08	83.56	78.23	92.30	89.51	84.86	69.90	59.88	64.89	94.80	93.39	94.10
		kNN	84.66	77.97	85.19	75.76	90.52	89.57	83.94	65.39	61.73	63.56	95.69	93.53	94.61
		SVM	87.09	80.76	87.78	76.77	94.15	92.42	86.49	66.61	60.78	63.69	89.39	93.99	91.69
	Max		87.32	81.77	88.07	78.48	93.56	91.77	86.83	69.98	60.02	65.00	96.36	93.82	95.09
	Min		88.57	82.09	89.33	79.49	93.56	91.88	87.49	68.50	61.50	65.00	96.27	93.88	95.08
	Sum		88.51	82.72	88.96	78.99	94.22	92.18	87.60	69.23	61.00	65.12	96.32	94.02	95.17
	Product		88.63	82.22	89.04	79.56	93.93	91.59	87.49	68.29	61.47	64.88	96.32	93.96	95.14
	Median		88.21	82.34	88.37	78.73	93.93	92.18	87.29	68.09	61.71	64.90	96.16	94.08	95.12
	LWE		89.10	82.47	89.33	79.81	93.33	92.24	87.71	68.99	61.36	65.18	96.10	93.67	94.89
	P-fusion		88.21	82.85	88.37	79.11	94.00	92.00	87.42	69.69	60.65	65.17	96.40	93.96	95.18
	MBCF_Spe		89.69	83.92	**90.81**	80.70	94.30	92.06	88.58	69.20	61.58	65.39	96.47	93.97	95.22
	MBCF_Spa		**90.52**	**85.13**	90.67	**81.20**	**95.41**	**93.48**	**89.40**	**70.52**	**61.84**	**66.18**	**96.77**	**94.53**	**95.65**

TABLE VII
OA (%) OF MBCF AND COMPARED FUSION STRATEGIES ON THE SA PRODUCED FEATURES

Data sets			BOT							KSC			Worldview-2		
			May-June	June-May	May-July	July-May	June-July	July-June	average	KSC1-KSC2	KSC2-KSC1	average	2011-2012	2012-2011	average
SA	Base Classifiers	LDA	76.36	79.30	80.52	77.28	85.63	88.15	81.21	62.17	**63.27**	62.72	94.64	93.30	93.97
		kNN	86.26	77.41	86.89	74.43	90.89	87.97	83.97	**68.84**	58.38	63.61	95.73	93.18	94.45
		SVM	88.86	77.22	88.89	76.27	**94.30**	91.59	86.19	65.67	58.93	62.30	94.68	93.56	94.12
	Max		85.31	81.08	88.52	79.24	90.89	90.64	85.95	66.09	59.70	62.90	95.97	93.78	94.88
	Min		86.20	81.71	89.63	78.86	91.19	90.82	86.40	68.21	60.05	64.13	95.97	93.77	94.87
	Sum		88.45	81.39	89.70	80.06	92.37	90.76	87.12	67.18	60.25	63.72	96.10	93.85	94.98
	Product		86.73	81.71	89.63	79.05	91.41	90.88	86.57	68.11	60.11	64.11	96.03	93.80	94.92
	Median		89.63	80.32	89.04	78.35	92.89	90.34	86.76	67.44	59.88	63.66	96.02	93.80	94.91
	LWE		86.97	82.59	90.96	80.57	92.89	90.88	87.48	67.54	59.19	63.37	95.73	93.68	94.71
	P-fusion		87.32	81.33	89.56	79.81	92.30	91.23	86.92	66.68	61.02	63.85	96.05	93.87	94.96
	MBCF_Spe		89.45	82.53	91.19	81.90	93.33	91.47	88.31	68.34	60.33	64.34	96.24	93.88	95.06
	MBCF_Spa		**89.87**	**83.80**	**91.56**	**82.91**	93.70	**91.65**	**88.91**	68.60	60.60	**64.60**	**96.48**	**94.37**	**95.43**

MBCF depends on the features. For example, JDA obtains superior feature alignment performance to MA on BOT data, and thus MBCF in conjunction with JDA obtains higher accuracies than MBCF on features of MA. It is worth noting that the purpose of MBCF is to obtain a superior classification accuracy by fusing multiple base classifiers, rather than improving the domain invariant features. In addition, MDAF is regarded to be more stable, since MBCF relies on the feature generator.

E. Sensitivity Analysis of Parameter in the Neighborhood Consistency Weighting Strategy

The proposed MDAF or MBCF contains one parameter, which is the number of neighbors in the neighborhood consistency weighting method. We conducted sensitivity analysis of the number of spectral neighbor (K) and the size of the spatial neighborhood window, respectively.

For the spectral neighbors, different values of K were tested, which are from 5 to 49 with a step of 2. The MDAF_Spe performed on four feature-based domain adaptation methods and MBCF_Spe on JDA produced features were chosen for the analysis. The classification results with different neighbors K on BOT images were shown in Fig. 11(a), (b). The overall accuracies of MDAF_Spe and MBCF_Spe on all the data pairs are stable with the changed values of K, which indicated the proposed method is insensitive to this parameter. However, since the adaptive weight is sample-wised, the neighborhood consistency criterion may require different number of neighbors for different samples. If a target sample has overlapping spectra with another target class, a small neighborhood is desirable to obtain a homogeneous neighborhood. If a target sample has spectral drift

TABLE VIII
OA (%) OF MBCF AND COMPARED FUSION STRATEGIES ON THE MA PRODUCED FEATURES

Data sets			BOT							KSC			Worldview-2		
			May-June	June-May	May-July	July-May	June-July	July-June	average	KSC1-KSC2	KSC2-KSC1	average	2011-2012	2012-2011	average
MA	Base Classifiers	LDA	76.48	79.94	75.56	75.76	91.33	89.34	81.40	69.41	68.50	68.95	93.65	93.10	93.37
		kNN	73.46	77.91	71.85	72.91	90.59	90.40	79.52	68.24	70.48	69.36	95.47	93.42	94.44
		SVM	80.27	76.96	**78.89**	74.49	92.15	89.75	82.09	69.28	70.56	69.92	95.64	92.76	94.20
	Max		77.90	79.68	75.33	77.09	92.74	90.05	82.13	70.39	70.43	70.41	95.28	93.61	94.45
	Min		79.09	79.62	75.70	77.47	92.81	90.23	82.49	69.82	71.30	70.56	95.41	93.79	94.60
	Sum		79.80	79.75	76.44	77.28	92.30	90.82	82.73	70.39	70.67	70.53	95.55	93.80	94.68
	Product		79.21	79.75	75.78	77.34	92.74	90.64	82.58	70.06	71.33	70.70	95.47	93.88	94.68
	Median		80.04	79.05	77.19	76.58	92.37	90.40	82.60	69.51	70.45	69.98	95.68	93.76	94.72
	LWE		79.09	80.19	78.15	77.03	92.96	90.88	83.05	69.43	71.25	70.34	95.29	93.59	94.44
	P-fusion		79.03	79.81	76.37	77.22	92.59	90.28	82.55	70.34	70.67	70.51	95.42	93.74	94.58
	MBCF_Spe		80.75	**80.63**	76.44	77.47	93.33	90.82	83.24	70.89	71.51	71.20	95.83	93.81	94.82
	MBCF_Spa		**80.98**	80.57	76.00	**77.91**	**93.63**	**91.53**	**83.44**	**71.69**	**71.54**	**71.62**	**95.96**	**94.40**	**95.18**
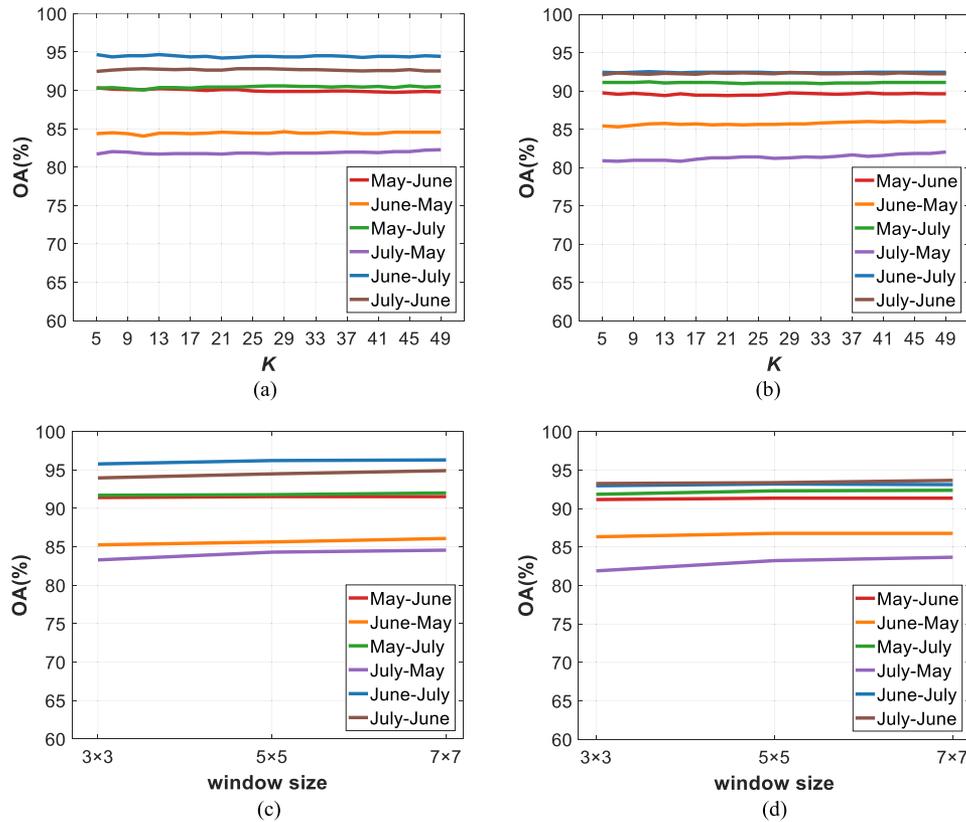
Fig. 11. Sensitivity analysis of the parameters K and window size in the neighborhood consistency weighting approach. (a) MDAF_Spe on BOT data. (b) MBCF_Spe on BOT data. (c) MDAF_Spa on BOT data. (d) MBCF_Spa on BOT data.

and become spectrally similar to another source class, it may be misclassified. In this case, a large neighborhood is preferred to avoid all its neighbors are misclassified. Although the overall accuracies on all the target samples indicate the insensitive of the parameter K, an adaptive number of neighbors should be more suitable to obtain a superior performance.

For the spatial neighbors, three different sizes of spatial windows were tested, including 3×3, 5×5, and 7×7. The MDAF_Spa performed on four domain adaptation methods and MBCF_Spa on JDA produced features were chosen for the analysis. The classification results with different window sizes on BOT images were shown in Fig. 11(c), (d). It can

be observed that the overall accuracies are slightly improved with the increasing window sizes. Since the target data used for quantitative evaluation are from the labeled patches in BOT images, the large window still has a good spatial homogeneity and contains more neighbors that can be used to evaluate the classification consistency. For the other data points in target image, the spatial window with a large size may not be suitable, since the homogeneity cannot be guaranteed especially for the pixels in the boundaries of different land cover types. Therefore, we suggest that a small spatial window such as 3×3 and 5×5 is used in the spatial neighborhood consistency based weighting method.

VI. CONCLUSION

We proposed MDAF and MBCF algorithms for multiclassifier fusion for domain adaptation. An adaptive samplewise weight is assigned to each classifier on each sample, which can exploit the advantages of classifiers on a per-sample basis. We also proposed a neighborhood consistency criterion on each target sample to calculate the weight of each classifier. In the experiments with Hyperion, AVIRIS, and Worldview-2 remote sensing images, the proposed fusion method not only outperformed the single classifier, but also offered better performance than other seven multiclassifier fusion methods. Results also demonstrated that effective fusion can integrate the advantages of different classifiers. A classifier with a low overall accuracy can still provide positive effect on the fusing results, since it may have outstanding performance on some classes. Moreover, the proposed method does not limit the selection of fused classifiers, although the methods that can provide diverse knowledge are preferred.

This article focuses on single source domain and single target domain, and the proposed method should also be suitable for fusing classification results from multiple source domains. In addition, using adaptive neighbors and combining other weighting strategies can further improve the fusion performance, which are our future work.

ACKNOWLEDGMENT

REFERENCES

[1] S. Mei, R. Jiang, X. Li, and Q. Du, "Spatial and spectral joint super-resolution using convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4590–4603, Jul. 2020.

[2] L. Ma, M. M. Crawford, X. Yang, and Y. Guo, "Local-manifold-learning-based graph construction for semisupervised hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2832–2844, May 2015.

[3] Y. Zhou, J. Peng, and C. L. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1082–1095, Feb. 2015.

[4] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 118–139, Jun. 2019.

[5] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 41–57, Jun. 2016.

[6] C. Persello and L. Bruzzone, "Kernel-based domain-invariant feature selection in hyperspectral images for transfer learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2615–2626, May 2016.

[7] J. Hoffman, E. Rodner, J. Donahue, T. Darrell, and K. Saenko, "Efficient learning of domain-invariant image representations," in *Proc. Int. Conf. Learn. Represent.*, AZ, Scottsdale, May 2013, pp. 1–9.

[8] L. Duan, D. Xu, and I. W. Tsang, "Learning with augmented features for heterogeneous domain adaptation," in *Proc. Int. Conf. Mach. Learn.*, Edinburgh, Scotland, Jun. 2012, pp. 711–718.

[9] X. Li, L. Zhang, B. Du, L. Zhang, and Q. Shi, "Iterative reweighting heterogeneous transfer learning framework for supervised remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 2022–2035, May 2017.

[10] G. Matasci, M. Volpi, M. Kanevski, L. Bruzzone, and D. Tuia, "Semisupervised 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.

[11] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.

[12] L. Zhu and L. Ma, "Class centroid alignment based domain adaptation for classification of remote sensing images," *Pattern Recognit. Lett.*, vol. 83, no. 2, pp. 124–132, Jan. 2016.

[13] J. Peng, W. Sun, L. Ma, and Q. Du, "Discriminative transfer joint matching for domain adaptation in hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 972–976, Jun. 2019.

[14] B. Sun, J. Feng, and K. Saenko, "Correlation alignment for unsupervised domain adaptation," 2016.

[15] L. Ma, M. M. Crawford, L. Zhu, and Y. Liu, "Centroid and covariance alignment-based domain adaptation for unsupervised classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2305–2323, Apr. 2019.

[16] B. Frenando, A. Habrad, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 2960–2967.

[17] B. Banerjee and S. Chaudhuri, "Hierarchical subspace learning based unsupervised domain adaptation for cross-domain classification of remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 11, pp. 5099–5109, Oct. 2017.

[18] H. L. Yang and M. M. Crawford, "Spectral and spatial proximity-based manifold alignment for multitemporal hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 51–64, Jan. 2016.

[19] D. Tuia, M. Volpi, M. Trolliet, and G. Camps-Valls, "Semisupervised manifold alignment of multimodal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7708–7720, Dec. 2014.

[20] D. Tuia and G. Camps-Valls, "Kernel manifold alignment for domain adaptation," *PLoS One*, vol. 11, no. 2, Feb. 2016, Art. no. e0148655.

[21] D. M. Gonzalez, G. Camps-Valls, and D. Tuia, "Weakly supervised alignment of multisensor images," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Milano, Italy, Jul. 2015, pp. 2588–2591.

[22] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. Int. Conf. Mach. Learn.*, Lille, France, Jul. 2015, pp. 97–105.

[23] B. Sun and K. Saenko, "Deep CORAL: Correlation alignment for deep domain adaptation," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, Netherlands, Oct. 2016, pp. 443–450.

[24] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 1–35, Jan. 2016.

[25] Z. Pei, Z. Cao, M. Long, and J. Wang, "Multi-adversarial domain adaptation," 2018, *arXiv: 1809.02176*.

[26] H. Sun, S. Liu, S. Zhou, and H. Zou, "Unsupervised cross-view semantic transfer for remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 13–17, Jan. 2016.

[27] L. Zhou and L. Ma, "Extreme learning machine-based heterogeneous domain adaptation for classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 11, pp. 1781–1785, Nov. 2019.

[28] Z. Liu, L. Ma, and Q. Du, "Class-wise distribution adaptation for unsupervised classification of hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, Jun. 2020, vol. 59, no. 1, pp. 508–521, Jan. 2021.

[29] M. Chen, L. Ma, W. Wang, and Q. Du, "Augmented associative learning-based domain adaptation for classification of hyperspectral remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6236–6248, 2020.

[30] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[31] A. Mi, L. Wang, and J. Qi, "A multiple classifier fusion algorithm using weighted decision templates," *Sci. Program.*, vol. 2016, pp. 1–10, Aug. 2016.

[32] X. Huang and L. P. 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.

[33] J. Gao, W. Fan, J. Jiang, and J. Han, "Knowledge transfer via multiple model local structure mapping," in *Proc. ACM Int. Conf. Knowl. Discovery Data Mining.*, Beijing, China, Aug. 2008, pp. 283–291.

[34] Z. Zhao and A. Cai, "Combining multiple SVM classifiers for adult image recognition," in *Proc. IEEE Conf. Net. Infr. Digital Content.*, Beijing, China, Sep. 2010, pp. 149–153.

[35] J. Zhao, Y. Chen, X. Zhuang, and Y. Xu, "Posterior probability based multi-classifier fusion in pedestrian detection," in *Proc. Int. Conf. Genetic Evol. Comput.*, Nanchang, China, Oct. 2014, pp. 323–329.

[36] C. Luo and L. Ma, "Neighbor consistency based unsupervised manifold alignment for classification of remote sensing image," in *Proc. 10th IAPR Workshop Pattern Recognit. Remote Sens.*, Beijing, China, Aug. 2018, pp. 1–8.

[37] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 2200–2207.

[38] B. Liu, G. Gao, and Y. Gu, "Unsupervised multitemporal domain adaptation with source labels learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1477–1481, Mar. 2019.

[39] H. Sifaou, A. Kammoun, and M. Alouini, "High-dimensional linear discriminant analysis classifier for spiked covariance model," *J. Mach. Learn. Res.*, vol. 21, no. 112, pp. 1–24, May 2020.

Hongkang Wei received the B.S. degree in communication engineering from the China University of Geosciences, Wuhan, China, in 2019, where he is currently working toward the M.S. degree in information and communication engineering at School of Mechanical Engineering and Electronic Information.

His research interests include pattern recognition, computer vision, and hyperspectral data analysis.

Li Ma (Member, IEEE) received the B.S. degree in biomedical engineering and M.S. degree in pattern recognition and intelligent system from Shandong University, Jinan, China, in 2004 and 2006, respectively, and the Ph.D. degree in pattern recognition and intelligent system from Huazhong University of Science and Technology, Wuhan, China, in 2011.

During 2008–2010, she was a Visiting Scholar with Purdue University, Indiana, USA. She also visited Mississippi State University, Mississippi, USA, for five months in 2018. She is currently an Associate Professor with the School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan. Her research interests include hyperspectral data analysis, pattern recognition, and remote sensing applications.
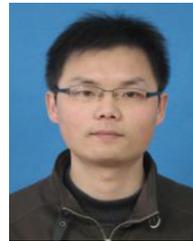
Yong Liu received the B.S. degree in computer science from Wuhan University of Technology, Wuhan, China, in 2000, the M.S. degree in internet technologies from University of Wollongong, Australia, in 2004, and the Ph.D. degree in geological engineering from China University of Geosciences, Wuhan, China, in 2011.

During 2013–2014, he was a Visiting Scholar with the University of Connecticut, USA. He is currently an Associate Professor with the School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan. His research interests include intelligence algorithm, information process, and their applications.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland at Baltimore, Baltimore, MD, USA, in 2000.

She is the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA, and also an Adjunct Professor with the College of Surveying and Geo-informatics, Tongji University, Shanghai, China. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is a Fellow of the SPIC International Society for Optics and Photonics. She was a recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society. She was the Chair of the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She served as the Co-Chair for the Data Fusion Technical Committee of the IEEE GRSS from 2009 to 2013. She was the General Chair for the fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing held at Shanghai, in 2012. She served as an Associate Editor for the IEEE JOURNAL OF TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS), the *Journal of Applied Remote Sensing*, and the IEEE Signal Processing Letters. She has served as the Editor-in-Chief of the IEEE JSTARS (2016–2020). She also organized several international workshops and journal special issues on remote sensing image processing and analysis.