

An Active Relearning Framework for Remote Sensing Image Classification

Qian Shi^{1b}, Xiaoping Liu, *Member, IEEE*, and Xin Huang^{1b}, *Senior Member, IEEE*

Abstract—Classification is an important technique for remote sensing data interpretation. In order to enhance the performance of a supervised classifier and ensure the lowest possible cost of the training samples used in the process, active learning (AL) can be used to optimize the training sample set. At the same time, integrating spatial information can help to enhance the separability between similar classes, which can in turn reduce the need for training samples in AL. To effectively integrate spatial information into the AL framework, this paper proposes a new active relearning (ARL) model for remote sensing image classification. In particular, our model is used to relearn the spatial features on the classification map, which contributes significantly to enhancing the performance of the classifier. We integrate the relearning model into the AL framework, with the aim to accelerate the convergence of AL and further reduce the labeling cost. Under the newly developed ARL framework, we propose two spatial–spectral uncertainty criteria to optimize the procedure for selecting new training samples. Furthermore, an adaptive multiwindow ARL model is also introduced in this paper. Our experiments with two hyperspectral images and two very high resolution images indicate that the ARL model exhibits faster convergence speed with fewer samples than traditional AL methods. Our results also suggest that the proposed spatial–spectral uncertainty criteria and the multiwindow version can further improve the performance when implementing ARL.

Index Terms—Active learning (AL), classification, high resolution, hyperspectral, relearning.

I. INTRODUCTION

IN RECENT years, a variety of supervised classification methods have been proposed to increase the classification accuracy when analyzing remotely sensed images [1]–[7]. However, classification performance is strongly dependent on the quality of the samples used to train the classifier. In the last few years, active learning (AL) methods have

been successfully applied to efficiently select training sample sets, by minimizing the number of training samples that are necessary to keep the discriminative capabilities as high as possible [8]–[11]. Starting from a small-size and suboptimal training set, AL repeatedly selects the most informative unlabeled samples, according to a certain criterion computed on a model outcome. The chosen candidates are then presented to the user, who determines their labels. These new queried samples are then added to the training set. This procedure runs iteratively until a stopping criterion is met, e.g., a predefined accuracy level on an independent test set, or a given training set size [12].

In this regard, how to define the uncertainty criterion is crucial for AL heuristics, and this can be roughly defined as the reciprocal of the confidence of a classifier on the label assignment on an unlabeled sample set [13]. Thus, various AL approaches differ in the definition of the uncertainty criterion for the model. The first group of AL approaches is based on multiclass-level uncertainty (MCLU), such as the support vector machine (SVM). The samples distributed close to the separating hyperplane are defined as the uncertain ones [10], [14]–[16], i.e., the ones with lowest classifier confidence on a given set of candidates. A second group of AL approaches are based on a classifier that can output the posterior probability. In [17], the breaking ties (BT) criterion [17], [18] was introduced to find the samples with the minimum difference between the two highest posterior probabilities. In [19], the samples that mostly change the existing belief in the class distributions were selected. A third group of AL approaches define the uncertain samples as the ones that show the maximum disagreement between the predictions of different classification models. For instance, in [20], the entropy of the predictions was used as the criterion to select new samples. In [21], a multiview-disagreement-based AL framework was used to define each subset of the feature space as a committee, and the samples with the maximum disagreement between different committees were then selected as the most informative samples.

As pointed out in [22] and [23], it is necessary to consider a batch of samples in order to reduce the computational burden and iterations. However, the pixels close to the border of an object are usually mixed pixels. Thus, using only an uncertainty criterion will inevitably result in sampling several mixed pixels lying close to the discriminative hyperplane and close to each other. To overcome such proximal sampling issue, a diversity constraint is enforced. The unlabeled samples

Manuscript received August 3, 2017; revised October 20, 2017; accepted January 2, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 41101336 and Grant 61601522, in part by the Program for New Century Excellent Talents in University of China under Grant NCET-11-396, in part by the Science Fund for Excellent Young Scholars under Grant 41322009, in part by the China Post-Doctoral Science Foundation under Grant 2015M580753, and in part by the Key National Natural Science Foundation of China under Grant 41531176. (Corresponding authors: Xiaoping Liu; Xin Huang.)

Q. Shi and X. Liu are with the School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China (e-mail: liuxp3@mail.sysu.edu.cn).

X. Huang is with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: huang_wu@163.com).

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/TGRS.2018.2800107

are clustered into many clusters, and only one sample in each cluster can be chosen to represent the cluster. There are different ways of selecting the most informative sample added in the query sample set. For example, in [22], the centroid of each cluster was chosen. In [23], the kernel k -means is used to extract the centroid of each cluster, which has been proved better results than k -means clustering way. In [20], the sample with the minimum support vector (SV) distance in each cluster was selected as the informative sample.

AL approaches have been widely used in remote sensing imagery classification, and have demonstrated great potential in improving the quality of the training sample set [24]. In particular, integrating spatial information into the AL framework has been shown to be able to further reduce the number of samples needed by the classifier and, at the same time, lessen the time and effort of manual image interpretation by experts.

A commonly used approach for exploring spatial information is to optimize the selection criterion for the classifier. With the assumption that the training sample set should represent the whole image (not only in the spectral domain but also in the spatial domain), the distribution of the query samples in the geographic space is also taken into consideration. In [25] and [26], the uncertain samples selected from the spectral domain were reselected in the spatial domain, and the samples with a larger spatial distance to the current sample set were identified as new query samples. This method for increasing the distances between the query points generally encourages the querying of fewer correlated and more representative samples.

A second group of approaches aim at achieving AL in a region-based manner. In [27] and [28], region-based AL heuristics were developed to guide the sample selection toward a limited number of compact spatial batches rather than scattered points. In this way, the uncertainty can be evaluated based on both the spectral and spatial domains. In [29], the pixels located within a homogenous object were assumed to be similar to each other, and thus this property can be considered as the diversity criterion to encourage a spatially uncorrelated sample set. In [30], with the assumption that the samples belonging to the same clusters share the same label, the homogenous patches were used as the reference to label the new samples located in the same homogeneous patch. In [31], the new samples were used to guide the process to generate a more optimized hierarchical clustering tree.

The aforementioned AL methods use spatial information to guide the selection of the new samples. The ultimate purpose of selecting these new samples is to enhance the performance of the classifier while minimizing the labeling cost. It is worth noting that spatial features can be directly used as a source of complementary information to further accurately describe the similarity between samples is an appropriate approach.

In such a process, the improvement in the classification accuracy is significant in the early iterations; however, when the quality of the training sample set has increased to a certain level, the learning rate of the AL process will start to decrease. The separating hyperplane begins to adjust with small change, and as a result, the classification accuracy increases gently. Thus, we called this stage as “fine-tuning

state.” In this state, despite continually adding a comparable number of samples, the improvement in the accuracy becomes insignificant. Unfortunately, the time-consuming fine-tuning state cannot be avoided due to the low starting accuracy, which is usually much less than the given accuracy. Therefore, in order to reduce the labeling cost needed for a given accuracy, it is crucial to improve the starting accuracy and shorten the period of the fine-tuning state. To achieve that purpose, a direct and powerful approach is to improve the predictive performance of the classifier. To this end, instead of designing complex classification algorithms, effective utilization of the spatial information to further accurately describe the similarity between samples is an appropriate approach.

In summary, an AL framework for remote sensing image classification should not only select the most informative samples according to the uncertainty from the spatial and spectral domains, but should also construct a more powerful classification model with sufficient consideration of the spatial information. There are many ways to improve the classification accuracy with the help of spatial information [3], [32]–[39]. The most common way is to extract spatial features on the image, for example, morphology texture features [3], [32], GLCM features [33], and so on. By adding spatial features, the feature space can be optimized, and the similarity between samples can be accurately described. Another kind way to utilize the spatial information is based on the assumption that spatial neighborhood samples prefer to share the same label, for example, Markov random field (MRF) [36], condition random field [37], and so on. And the third type is object-based methods, which segment the spatially adjacent pixels into spectrally similar objects and then conduct image analysis on the objects as the minimum unit of information [38], [39]. Meanwhile, the semisupervised learning methods can also be used to improve the classification map. And [40] proposed the integration of concepts of AL and SSL, and as a result, it jointly leverages the advantages of both approaches.

In this paper, we focus on extracting spatial features to improve the classification accuracy. In order to enhance the class separability and improve the classification accuracy for each iteration in the AL process, a recently developed iterative learning strategy, which is known as “relearning” [41], can be integrated into the AL framework. The relearning model aims to extract a new spatial feature that could describe the spatial arrangement of the land-cover classes based on the classification map, and, therefore moreover, as reported in [41], the classification accuracy can be dramatically improved after only a few iterations, with the aid of the primitive co-occurrence matrix (PCM) considered in the relearning process. We should note that the improvement in classification accuracy is achieved without adding additional samples into the classifier. Thus, the relearning model can improve such accuracy at each iteration in AL. By integrating the relearning model into the AL framework, we propose a novel application, called hereinafter “active relearning” (ARL). It should be stressed that ARL is a new concept and not a simple combination of two methods. AL improves the classification accuracy by adding new training samples in each iteration; on the other hand, the relearning model can enhance the separability of

different land-cover classes according to the classification map in each iteration in AL. We should also note that the ability to enhance the separability can be constantly increased because of the improvement of the classification map as the iterations are increased. Furthermore, the relearning process results in an improvement in the classification performance, by not only improving the starting accuracy of the fine-tuning state in AL but also speeding up the rate of the convergence process.

In order to improve the adaptability and stability of the newly introduced ARL framework to different scenes, we also propose an adaptive window size algorithm, the so-called adaptive ARL model, which can adaptively select an appropriate window size for each pixel to properly characterize the spatial arrangement. On the other hand, in terms of designing two kinds of uncertainty criteria from both the spatial and spectral domains, we make full use of the homogeneous regions' output by the relearning model. A lower confidence for a spatial region indicates probable serious misclassification between these similar samples, and thus sampling these uncertain regions can help to effectively distinguish these similar samples and make the discriminant hyperplane more accurate. Thus, we consider both the uncertainty of the spatial regions and the uncertainty of the unlabeled samples.

In summary, when only spectral information is used, the learning rate of AL will fall into a fine-tune state at a lower accuracy level, which will waste large number of samples in the subsequent iterations. Thus, spatial information is necessary for AL process. The relearning model is a novel spatial feature extraction way, and we first proposed the concept of "ARL," which embeds the relearning process into the AL framework in an iterative manner. Under this framework, we can improve the performance of AL in two ways: 1) we propose an adaptive window size to enhance the stability and adaptability of the relearning model and 2) we propose a new spatial-spectral criterion to select new samples under the ARL framework.

The remainder of this paper is organized as follows. Section II introduces the background knowledge. Section III describes the newly proposed ARL framework, which consists of two parts: 1) the spectral-spatial uncertainty strategy and 2) the adaptive multiwindow ARL model. Section IV presents the experimental results. Section V concludes this paper with some remarks.

II. BACKGROUND

This section introduces the background to the relearning model and the AL framework, based on which we propose a new ARL method.

A. Relearning Model

The relearning model aims at optimizing or relearning the labeling that is derived from an initial classification, according to some implicit rules related to the spatial arrangement of the image primitives [40].

Let us consider a remote sensing image $I = \{\mathbf{x}_i\}_{i=1}^N$, which is composed of N samples (arranged in n_{col} columns and n_{line} lines) with n_s spectral bands. The classifier trained by

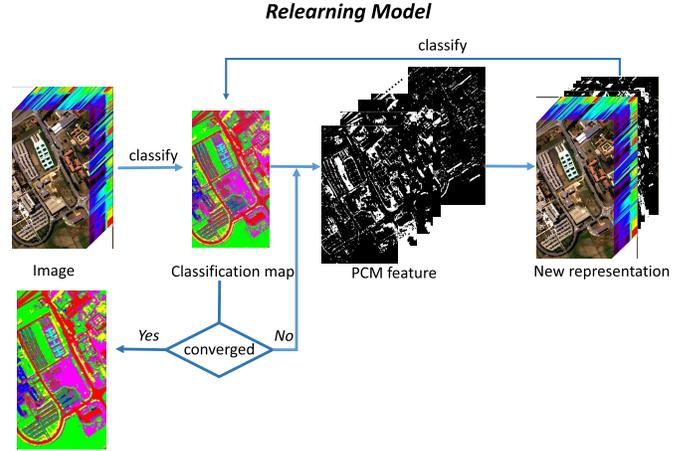


Fig. 1. Flowchart illustrating the relearning model.

the given training set L^{tm} is deployed to provide classification map I_C , which contains n_c information classes. Based on the classification I_C , PCM feature $I_P = \{\mathbf{p}_i\}_{i=1}^N$ can be generated. In each PCM feature \mathbf{p}_i , the element c_{ht} calculates the co-occurrence relationship between class h and class t in the $s \times s$ window centered by \mathbf{x}_i . This co-occurrence relationship can be represented by the times of co-occurrence between class h and class t . And $h, t \in \{1, 2, \dots, C\}$, and thus there are $n_c \times n_c$ elements in the \mathbf{p}_i .

As shown in Fig. 1, with the inclusion of the PCM feature I_P , I_{sp} represents the composed image when stacking I and I_P . Based on the new data representation, the classifier is retrained and the new classification map I_C^* is generated by assigning a predicted label to each sample in image I_{sp} . I_C is updated by I_C^* , and the process of relearning is repeated until the difference between I_C and I_C^* is minimized.

The basic principle of the PCM is illustrated in Fig. 2. Its calculation is described in the following steps.

Step 1: The classes defined in classified image I_C are viewed as the primitives of the image, providing an implicit representation of a scene.

Step 2: For each sample \mathbf{x}_i , the PCM feature can be calculated based on the classification map I_C . In the window centered at sample \mathbf{x}_i , $c_{ht, \text{dir}}$ represents the number of times that classes h and t occur in the given window, with $\text{dir} = (0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ)$, $h, t \in \{1, 2, \dots, n_c\}$.

Step 3: The element c_{ht} on all directions is summed as follows:

$$c_{ht} = \sum_{\text{dir}} c_{ht, \text{dir}}. \quad (1)$$

And $\mathbf{p}_i = \{c_{ht}\}_{h,t=1}^{n_c}$, which means that \mathbf{p}_i is the set of c_{ht} in the given $s \times s$ window centered at \mathbf{x}_i .

Step 4: The PCM feature can be represented by $I_P = \{\mathbf{p}_i\}_{i=1}^N$. Thus, the spectral feature I and PCM feature I_P are combined to generate new representation. And based on the training sample set L^{tm} , the postclassification map I_C^* can be generated.

It is important to note that the relearning model can suppress the pixelwise classification noise (salt-and-pepper) and smooth

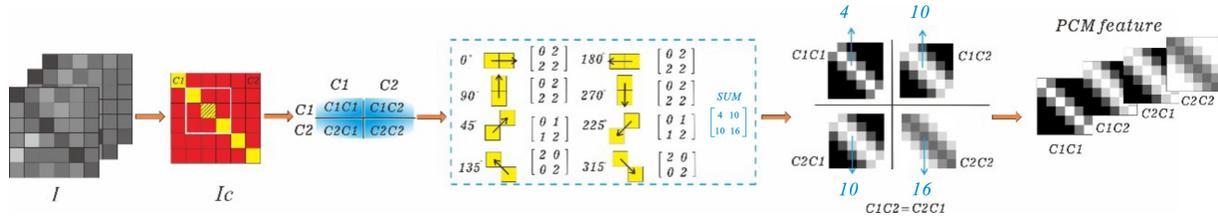


Fig. 2. Illustration of the PCM. PCM feature was extracted on the classification map I_c . There are two classes in I_c . The window size is set as 3. And we calculate the occurrence of class pair $[c_{11}, c_{12}, c_{21}, c_{22}]$ on eight different directions. And then, we sum the occurrence value on all the directions. For the sample x_{33} , $c_{11} = 4$, $c_{12} = 10$, $c_{21} = 10$, and $c_{22} = 16$. Thus, the PCM feature for x_{33} is $[c_{11}, c_{12}, c_{21}, c_{22}] = [4, 10, 10, 16]$.

Active Learning Framework

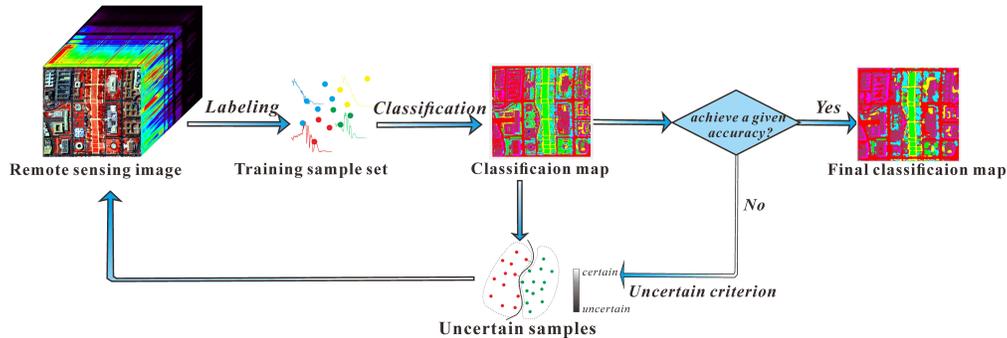


Fig. 3. AL procedure.

the raw classification. More significantly, the relearning model can also reduce misclassification (i.e., enhance class separability) by learning the implicit correlations or rules from the spatial arrangement of the labels and the class outputs within a neighborhood.

B. Active Learning

We first describe the overall AL framework. As shown in Fig. 3, AL is carried out in an iterative way, where the supervisor S interacts with the system by labeling the most uncertain samples from the pool U at each iteration. At the first stage, an initial training set L_0^{tm} with a few labeled samples is used for training the classifier G . After initialization, the uncertainty criterion F is used to select a set of uncertain samples. Then, based on the diversity criterion D , the most representative samples are selected from the uncertain samples. The supervisor S assigns the true class label y_i to these new selected samples. These new labeled samples T_{iter} (where iter refers to the iteration number) are then added to $L_{\text{iter}}^{\text{tm}}$ ($L_{\text{iter}}^{\text{tm}} = L_{\text{iter}-1}^{\text{tm}} + T_{\text{iter}}$), and the classifier G is retrained using the updated training set. L^{st} is the test sample set. This closed loop of querying and retraining continues until the classification accuracy of L^{st} is higher than the given accuracy.

The uncertainty criterion and diversity criterion used in this paper are further described as follows. The uncertainty is evaluated by the confidence of the classifier in correctly classifying the considered samples. In this paper, the SVM [42] and sparse multinomial logistic regression (SMLR) [43] are used as the classifiers to test the performance of the ARL framework. MCLU is adopted based on the SVM classifier used to analyze the uncertainty within the one-against-all (OAA) architecture. The MCLU technique selects the most uncertain samples

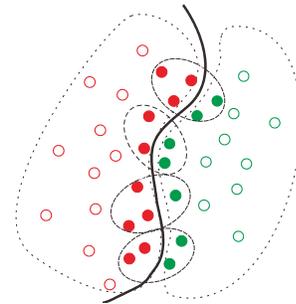


Fig. 4. Removing the redundancy in the uncertain sample set. The filled circles represent the uncertain samples selected by the uncertainty criterion, and they are divided into many clusters. The samples closest to the discrimination plane in each cluster are selected as the most representative ones.

according to their distance to the hyperplanes of the binary SVM classifier with the OAA strategy. Meanwhile, the BT, which focuses on analyzing the boundary regions between two classes, is used as the uncertainty criterion for SMLR, in order to obtain more diverse sample sets. The aforementioned uncertainty criteria are described in detail in the following.

1) *Multiclass-Level Uncertainty*: For the samples x_i , the distance value to the n hyperplanes can be represented by $\{f_1(x_i), \dots, f_n(x_i)\}$. The difference between the first- and second-largest distance values to the hyperplanes can then be represented as

$$d_i = f_{r_1 \max}(x_i) - f_{r_2 \max}(x_i). \quad (2)$$

If d_i is small, the decision for $r_{1 \max}$ is not reliable, and there is a possible conflict with class $r_{2 \max}$. Therefore, the uncertainty of x_i is high.

2) *Breaking Ties*: The BT criterion focuses on analyzing the boundary regions between two classes, with the goal of obtaining more diverse sample sets [23]. Suppose p_{ia}

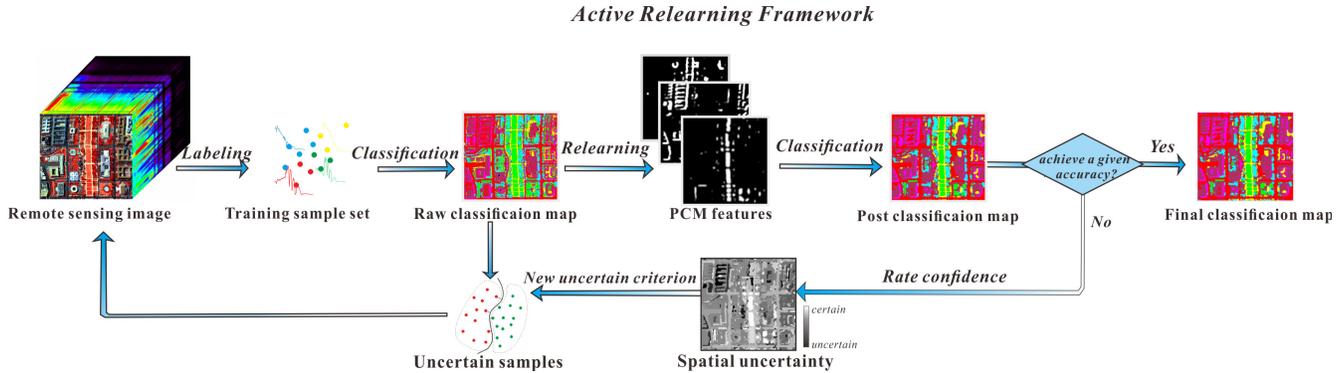


Fig. 5. Proposed ARL framework.

is the largest and p_{i_b} is the second-largest probability for a sample x_i , where i_a represents the index of the largest probability for sample x_i and i_b represents the index of the second-largest probability for sample x_i . The decision criterion is then

$$d_i = p_{i_a} - p_{i_b}. \quad (3)$$

Intuitively, if the value of d_i is small, the tie between p_{i_a} and p_{i_b} is strong, which indicates that the classification confidence is low.

We sort d_i by ascending order and select the n_{spectral} samples in the front part of the sorting as the new samples, which are represented by \mathbf{T}_{iter} , with iter representing the iterations.

Clustering-based diversity (CBD) is used as the diversity criterion in this paper. As shown in Fig. 4, k -means clustering is used to divide the uncertain samples \mathbf{T}_{iter} into a series of clusters \mathbf{C}_{iter} . In [23], k -means clustering in the kernel space is used to find the clusters. In the diversity criterion, it selects the batch of samples at each iteration according to the identification of the most uncertain sample of each cluster. And this way has been performed better than selecting the center of the cluster, which is called enhanced CBD (ECBD).

III. ACTIVE RELEARNING

This section introduces the newly proposed ARL model. As mentioned above, the relearning model can effectively enhance the performance of the classifier by describing the spatial arrangement; as a result, the accuracy of the classification map can be significantly increased. However, the relearning model changes only the data representation, and does not update the labeled sample set. At this point, the AL model can enlarge the training sample set in an interactive way with the classifier. Thus, we can utilize the relearning model and the AL model to jointly improve the classification accuracy.

As shown in Fig. 5, the proposed ARL framework can be briefly described as follows. In the beginning, it is classified with the initial training sample set L_0^{trn} , generating raw classification map \mathbf{I}_C . PCM feature \mathbf{I}_p can be generated by the relearning model based on classification map \mathbf{I}_C . The new representation of can be transferred as \mathbf{I}_{sp} by stacking the spectral bands \mathbf{I}_s and the PCM feature \mathbf{I}_p . Thus, the new classification map \mathbf{I}_C^* can be generated with the same training sample set. At this time, we test the accuracy of \mathbf{I}_C^* using test sample set L^{tst} . If the classification result has achieved a given

accuracy, the AL process can be stopped, and classification map \mathbf{I}_C^* is the final classification map. If the accuracy is less than the given accuracy, we select the most uncertain samples from unlabeled sample set U .

Algorithm 1 Proposed ARL Framework

Inputs:

- I : the image to be classified;
- L_0^{trn} : initial training sample set;
- L^{tst} : test sample set;
- U : unlabeled sample set;
- t : number of new samples added to L^{trn} at every iteration of the AL process;
- $size$: the window size to calculate the PCM feature;

Output:

Classification map \mathbf{I}_C^*

Repeat

1. Train the SVM classifier with the current training set L^{trn} , while estimating its free parameters by cross validation, and classify the image to obtain classification map \mathbf{I}_C ;
2. Calculate the PCM feature based on classification map \mathbf{I}_C and retrain the classifier to obtain the new classification map \mathbf{I}_C^* ;
3. Select the most informative sample set \mathbf{T}_{iter} according to the uncertainty criterion and the diversity criterion;
4. Label the selected sample set \mathbf{T}_{iter} ;
5. Add the labeled samples \mathbf{T}_{iter} to the training set L^{trn} and remove them from U

Until the accuracy of \mathbf{I}_C^* obtains the given accuracy on L^{tst}

Under the newly introduced ARL framework, we make full use of the homogeneous regions in \mathbf{I}_C^* to optimize the design of the uncertainty criterion. Based on both the region and sample uncertainty, we introduce new spectral-spatial uncertainty criteria to select the most informative samples, which are described in Section III-A. On the other hand, in order to improve the adaptability and stability of the ARL process with regard to different scenes, we propose an adaptive window size to measure the spatial arrangement, described in Section III-B.

A. Spectral–Spatial Uncertainty Criteria

Due to the smoothing effect of the relearning model, the class labels present in the new classification map \mathbf{I}_C^* can show a form of homogenous representation in the local regions. Therefore, we extract these homogeneous regions according to the distribution of the class labels. We use region grow algorithm to segment postclassification map to generate patches. To guarantee the representativeness of the spatial regions, we merge the regions whose number of pixels is smaller than 40 pixels. Euclidean distance is used to describe the similarity between the samples. We set the similarity threshold value as 20.

The spatial confidence of each region is not only related to the confidence of the pixels in each patch but also related to the homogeneity of the class labels appearing on the original classification map \mathbf{I}_C . The spatial confidence index dp_j of each region R_j can be calculated as

$$dp_j = \text{var}_j \times \sum_{t=1}^{m_j} d_{tj}/m_j \quad (4)$$

$$\text{var}_j = \max(\text{count}_{jc})/m_j \quad (1 < c < n_c) \quad (5)$$

where m_j is the number of samples in patch R_j , count_{jc} is the total number of the c th class labels in patch R_j , and n_c is the number of class labels. d_{tj} is the uncertainty of the t th sample in the j th patch on the raw classification map.

In this paper, we assume that uncertain samples with a lower predicted confidence d_{tj} , located in spatial regions with a lower spatial confidence dp_j , are more likely to be the informative samples for the remote sensing image classification. Two different criteria are proposed to integrate the spatial information in the process of training sample collection.

Algorithm 2 Spa1 Uncertainty Criterion

1. Compute the uncertainty of each sample x_i based on the spectral uncertainty criterion, and select the most uncertain sample set T_{spectral} ;
 2. Cluster the T_{spectral} sample set into n_k clusters $\mathbf{K}_{\text{iter}_i}$;
 3. Calculate the spatial uncertainty dp_j of the homogeneous region R_j which corresponds to the h -th sample x_h in $\mathbf{K}_{\text{iter}_i}$;
 4. Select the samples with the smallest dp_j in each $\mathbf{K}_{\text{iter}_i}$ as the uncertain sample set $\mathbf{T}_{\text{spe_spa}}$;
-

- 1) Similar to the traditional batch mode AL method [23], the first step is to select a batch of samples based on the spectral uncertainty. The second step is then to cluster these samples into a set of clusters, with each cluster represented by $\mathbf{K}_{\text{iter}_i}$. We assume that sample x_h in the i th cluster $\mathbf{K}_{\text{iter}_i}$ corresponds to a homogeneous patch R_j with uncertainty dp_j . The sample corresponds to the smallest dp_j in each cluster, which is selected to represent $\mathbf{K}_{\text{iter}_i}$. Since this is the first way to combine the spatial information, we refer to it as ‘‘Spa1.’’ Algorithm 2 summarizes the proposed selection strategy.
- 2) The spectral and spatial uncertainty can be considered simultaneously. Thus, we calculate the spectral

uncertainty and the spatial uncertainty for each sample in the unlabeled set U . Considering sample x_i in patch R_j , the new uncertainty can be represented as

$$dsp_i = d_i + dp_j \quad (6)$$

Thus, a batch of samples can be selected according to the value of dsp_i . We then cluster these samples into many clusters, and the most uncertain sample with the smallest dsp_i in each cluster is selected as the most representative sample. We call this selection strategy ‘‘Spa2.’’ Algorithm 3 summarizes the proposed selection strategy.

Algorithm 3 Spa2 Uncertainty Criterion

1. Compute the uncertainty of each sample x_i based on the proposed uncertainty and select the most uncertain sample set $T_{\text{spe_spa}}$;
 2. Cluster the $T_{\text{spe_spa}}$ sample set into n_k clusters $\mathbf{C}_{\text{iter}_s}$ ($1 \leq s \leq N_S$);
 3. Select the smallest dsp_i in each $\mathbf{C}_{\text{iter}_s}$ as the uncertain samples;
-

B. Adaptive Multiwindow Active Relearning Model

According to Section II, the relearning model can effectively enhance the classifier’s performance because the local spatial arrangement of the class labels is taken into consideration. Since, in most cases, nearby pixels have a similar spatial arrangement in local patches, the PCM features of nearby pixels will exhibit high similarity. Consequently, the relearning model often leads to a smoother classification result. Furthermore, describing the spatial arrangement in a local patch can also enhance the separability between adjacent ground classes.

However, when the window size is set larger, the similarity of the spatial arrangement between nearby pixels also increases. As a result, the PCM features between nearby pixels are more similar, which is not appropriate for distinguishing the edges of ground objects. In contrast, if the window is set to a smaller size (such as 3×3), then the similarity of the PCM features between nearby pixels would be reduced. As a result, a weaker smoothing effect would be applied, which is not effective for noise removal purposes. However, in this case, the edge information in the image can be better preserved. Therefore, from the above analysis, we can see that a fixed window size is not suitable for the whole image, and thus, a small window size should be applied on the edges between two ground classes, while a larger window size should be applied in the homogenous regions.

To address this problem, in this paper, we propose a direct way to search for the most suitable window size for each sample.

As shown in Fig. 6, pixel 1 is located in a homogeneous region and pixel 2 is located in the border area between two objects of interest. The homogeneity of each sample x_i can be measured by calculating and comparing the label consistency in different window sizes. The consistency is calculated by the frequency of the occurrence of the central pixel’s label in the local window.

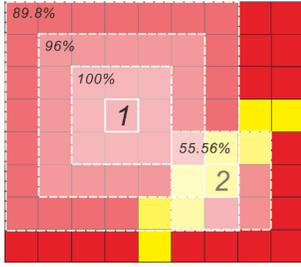


Fig. 6. Example of consistency in two situations.

For the classification map I_C , in order to extract the main structural information and suppress the influence of noise, the PCM features extracted from the larger window size ($size_0 = 13$) on all the pixels are first used to generate a smoother classification map I_R^* . We therefore treat the new classification map I_R^* as the reference to calculate the homogeneity of each sample.

According to the example in Fig. 6, we calculate the consistency of the spatial arrangement with an increasing window size. The label consistency can be calculated as

$$homo_i = \sum_{j=1}^{n_{all_i}} n_{same_i} / n_{all_i} \quad (7)$$

where $n_{all_i} = size_i \times size_i$ and n_{same_i} is the number of samples whose label is the same as x_i in the local window. We calculate $homo_i$ from the smallest window size ($size_i = 3$). In most cases, when the window size becomes larger, $homo_i$ becomes smaller due to the occurrence of other class labels. Thus, we apply a threshold H ($0 \leq H \leq 1$) to the consistency criterion: when $homo_i$ is less than H from the very start, the corresponding window size gives the optimal one for pixel x_i .

As seen in Fig. 6, for pixel 1, the neighborhood samples located in a 3×3 window show 100% consistency, those located in a 5×5 window show 96% consistency, and those located in a 7×7 window show 89.8% consistency. For pixel 2, the neighborhood samples located in a 3×3 window show only 55.56% consistency. If we set the homogeneity threshold as 0.8, the local window size for pixel 1 is 7×7 , whereas the window size for pixel 2 is 3×3 .

When a smaller threshold H is used, a larger $size_i$ is generated. Thus, we set the bound value as $size_i = size_0$ to avoid setting the value of H too small. Algorithm 4 summarizes the strategy adopted for determining the window size for each pixel.

IV. EXPERIMENTS

Our experimental section is divided into four parts. First of all, the remote sensing data sets used in the experiments are introduced in Section IV-A. The following parts correspond to the contributions of this paper. The main contribution is the proposed ARL framework, and thus, we compare the performance of the ARL model and the AL model in Section IV-B. Under the ARL framework, we propose new spectral-spatial uncertainty criteria and an adaptive multiwindow version, and thus, we assess their effectiveness in Sections IV-C and IV-D, respectively.

Algorithm 4 Proposed Multiwindow Relearning Model

Inputs:

I_C : the raw classification map;

H : the threshold of the homogeneity of neighborhood samples;

Outputs:

The $size_i$ of each sample x_i

The PCM feature of each sample x_i

Set the window $size_0 = 13$, calculate the PCM feature based on classification map I_C , and retrain the classifier to obtain the new classification map I_R^*

For each sample x_i

Initialization:

$size_i = 3$;

Repeat

- 1) On the reference classification map I_R^* , calculate the label homogeneity $homo_i$ of the neighborhood samples located in window $size_i \times size_i$;
- 2) If the homogeneity of the neighborhood samples is larger than the given threshold: $homo_i > H$, increase the window size: $size_i = size_i + 1$;

Until $homo_i \leq H$ or $size_i = size_0$

Calculate the PCM feature of x_i with the window $size_i \times size_i$.

A. Data Description

Experiments were conducted on a series of remote sensing data sets to validate the performance of the proposed method.

- 1) *Hyperspectral Data Sets*: The airborne visible infrared imaging spectrometer (AVIRIS) image collected over the Indian Pines test site is a widely used hyperspectral data set for classification applications. The image consists of 145×145 pixels, with 220 spectral bands [see Fig. 7(a)], and contains 12 crop types and a total of 10 171 labeled pixels for algorithm testing. Classification of this image is a very challenging problem because of the highly mixed class signatures and the unbalanced number of labeled pixels per class [44].
- 2) The second hyperspectral image data set was acquired by the airborne reflective optics system imaging spectrometer and covers an urban area of Pavia, northern Italy [see Fig. 7(b)]. The image scene, with a size of 610×340 pixels, is centered at the University of Pavia. After removing 12 noisy/water absorption bands, 103 spectral channels were utilized in the experiments. Nine ground-truth classes, with a total of 43 923 samples, were considered in the experiments. Several groups of classes in this data set have similar spectral properties, such as trees-grass and roofs-trails-roads, which reduces the spectral separability and increases the classification difficulty.
- 3) *HSR Data Sets*: High spatial resolution (HSR) data provide detailed ground information. In this paper, two HSR data sets were used for the validation and comparison of the various algorithms addressed in this paper. For the QuickBird Wuhan and WorldView-2 Hainan data sets,

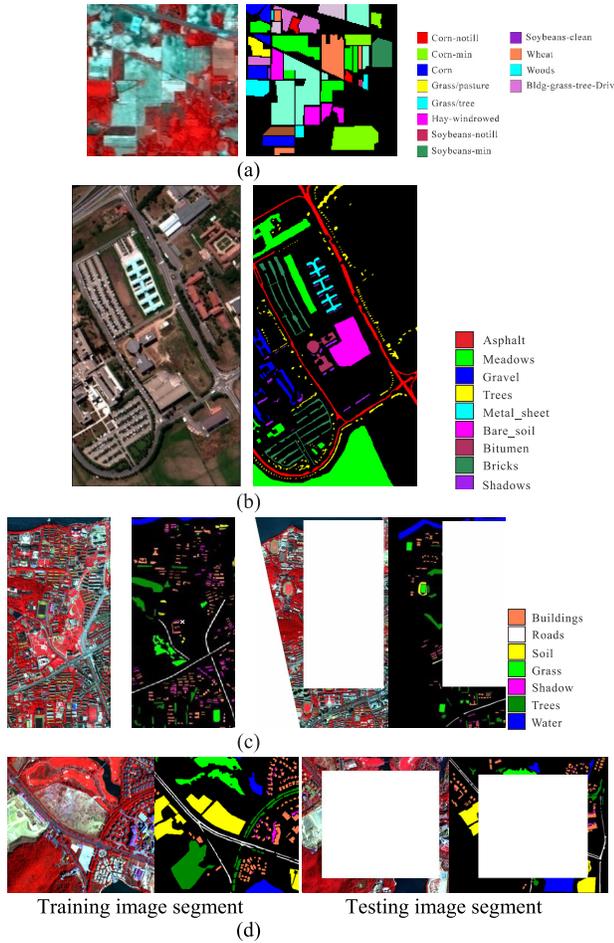


Fig. 7. Data sets and their reference samples. (a) AVIRIS Indian Pines image. (b) Pavia University image. (c) QuickBird Wuhan image. (d) WorldView-2 Hainan image.

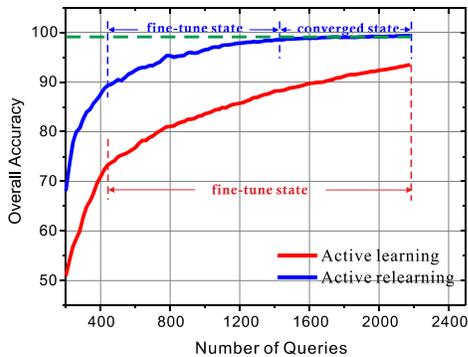


Fig. 8. Trends of the overall accuracy of AR and ARL with regard to the training iterations.

the challenge when interpreting these HSR images is to discriminate between the spectrally similar classes, e.g., grass-trees, water-shadow, and roads-buildings-soil.

We separate the data set into two disjoint sample sets: learning set and testing set. The training samples are selected from the learning set, while the test set was used to test the performance of different AL methods. We use two ways to separate the data set. For the hyperspectral data set, we separate

TABLE I
NUMBER OF LEARNING AND TEST SAMPLES USED IN EXPERIMENTS.
(a) AVIRIS INDIAN PINES IMAGE. (b) PAVIA UNIVERSITY IMAGE

(A) AVIRIS INDIAN PINES IMAGE		
Class	# Learning Samples	# Test Samples
Corn-notill	574	860
Corn-min	334	500
Corn	94	140
Grass/pasture	199	298
Grass/tree	299	448
Hay-windrowed	196	293
Soybeans-notill	387	581
Soybeans-min	987	1,481
Soybeans-clean	246	368
Wheat	85	127
Woods	518	776
Bldg-grass-tree-Drives	152	228
Total	4,068	6,103

(B) PAVIA UNIVERSITY IMAGE		
Class	# Learning Samples	# Test Samples
Asphalt	2,652	3,979
Meadows	7460	11,189
Gravel	840	1,259
Trees	1,226	1,838
Metal_sheets	538	807
Bare_soil	2,011	3,017
Bitumen	532	798
Bricks	1,473	2,209
Shadows	379	568
Total	17,110	25,666

(C) QUICKBIRD WUHAN IMAGE		
Class	# Learning samples	# Test samples
Roads	11643	6529
Grass	3639	3904
Buildings	10886	9237
Soil	5275	2696
Shadow	7377	1161
Trees	1013	6529
Water body	3217	3904
Total	43050	31644

(D) WORLDVIEW-2 HAINAN IMAGE		
Class	# Learning samples	# Test samples
Roads	7910	3668
Grass	3567	1789
Buildings	17348	4841
Soil	7309	108
Shadow	1228	199
Trees	12568	1518
Water body	4191	7018
Total	54121	19141

the data set in a random way. This randomly separating way can make train sample set and test sample set satisfy the same distribution. While for the HSR data set, we separate the data set in the way of selecting region of interest (ROI), which

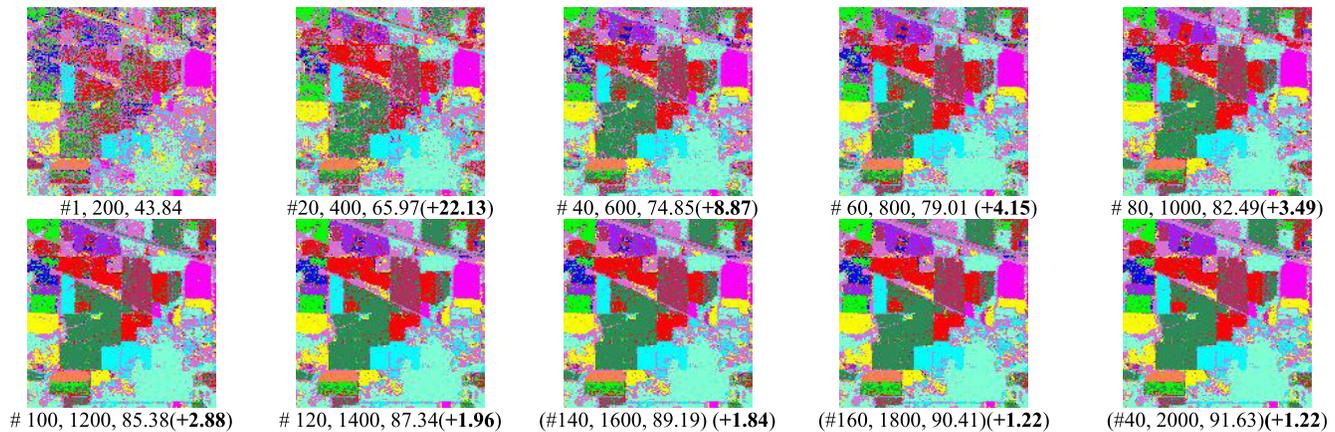


Fig. 9. Original classification maps I_C generated in different iterations on AVIRIS data set. The first number denotes the iterations, the second number is the number of samples used for the classification, the third number is the classification accuracy, and the fourth number in brackets is the improvement in accuracy achieved in the last 20 iterations.

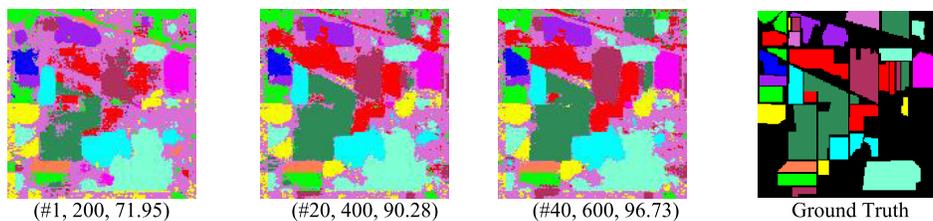


Fig. 10. Classification maps I_C^* generated in corresponding iterations on AVIRIS data set.

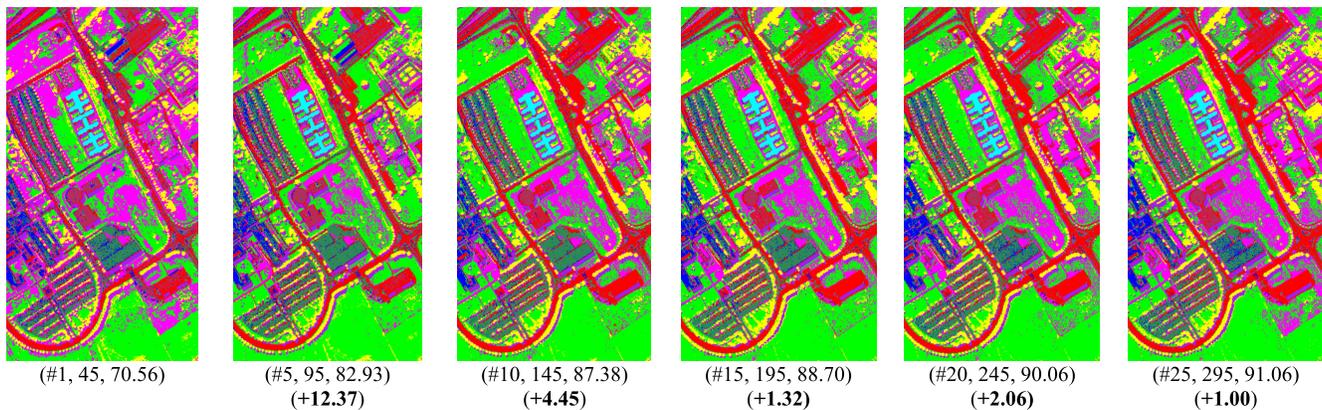


Fig. 11. Original classification maps I_C generated in different iterations on PAVIA data set. The first number denotes the iterations, the second number is the number of samples used for the classification, the third number is the classification accuracy, and the fourth number in brackets is the improvement in accuracy achieved in the last 20 iterations.

is shown in Fig. 7(c) and (d). And this selecting way is more closed to practical application. The numbers of training sample sets and testing sample sets are shown in Table I.

B. Comparison of the Performance of ARL and AL Models

The objective of this section is to show the superiority of the proposed ARL model over the AL model. To show the performance in a more intuitive way, we use the AVIRIS Indian Pines data set as an example. The classifier is the SMLR algorithm. The sparsity parameter was empirically fixed to 0.001 and the smoothness parameter to 2. The learning and test samples are shown in Table I(a).

We use 200 training samples as the initial samples. And we add 20 samples in each iteration for the AL and ARL. The window size of PCM feature is set as 7. The overall accuracy on the test sample set in each iteration can be shown in Fig. 8. Figs. 9 and 10 present the classification map for every 20 iterations.

As shown in Fig. 8, the accuracy of ARL is significantly higher than that of AL. One of the most important observations is that the starting accuracy of the fine-tuning state of ARL is about 90%, which is much closer to the full accuracy, whereas the starting point of AL is only 73.5%. Furthermore, we can see that within about 65 iterations, the fine-tuning state of ARL

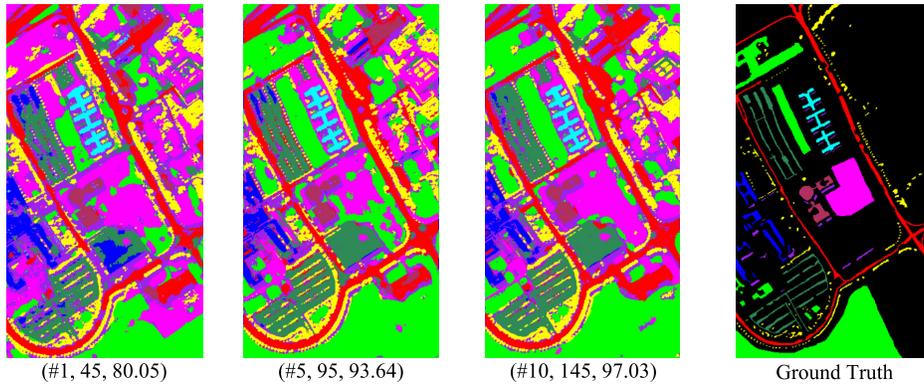


Fig. 12. Classification maps generated in corresponding iterations on PAVIA data set.

has finished and the learning process converges to the given accuracy. However, the fine-tuning state of AL still continues over 200 iterations.

We also show the evolution of classification map \mathbf{I}_C with the increase in iterations in Fig. 9. First, the accuracy increases in the first 20 iterations (22.13%), which is higher than the following iterations. Meanwhile, between 40 and 60 iterations, the rising tendency shows a steep fall (from 8.87% to 4.15%), after which the falling tendency starts to show a steady state. Comparing the classification maps with the ground truth, the misclassifications on the first two classification maps are widespread over the whole image, and some areas are even totally classified to wrong classes. However, in the third classification map (#40 iterations), which is in the fine-tuning state, the misclassified pixels are more like speckles distributed over the whole image. More importantly, from the classification maps shown in the following classifications, we can see that most of the new samples added in the process of AL are used to overcome the specklelike noise on the classification map \mathbf{I}_C . At this time, the direct benefits from the spectral bands are very limited. However, the specklelike noise can be removed with great precision by the relearning model, which can be clearly observed from classification map \mathbf{I}_C^* in Fig. 10. As shown in Fig. 10, the third classification map (#40 iterations) of the ARL process, with just the first 600 samples, can give a high accuracy (96.7%).

We also present the result of PAVIA_U data set in Fig. 11. First, the accuracy increase in the first 20 iterations (12.37%) is higher than the following iterations. Meanwhile, between 10 and 15 iterations, the rising tendency shows a steep fall (from 4.45% to 1.32%). About 20 iterations, the overall accuracy of classification map \mathbf{I}_C is 91.06%. However, we still can find many noises on the classification map \mathbf{I}_C . As shown in Fig.12, the overall accuracy of relearning model could get 97.03% by only 10 iterations. Thus, the relearning model can get higher accuracy with less samples.

C. Effectiveness of the ARL Model and the Proposed Spectral-Spatial Uncertainty Criteria

In this section, we focus on the ARL performance and the proposed uncertainty criteria.

TABLE II
LEARNING ALGORITHMS INVESTIGATED IN THIS PAPER

Algorithm	Classifier	Uncertainty criterion	Spatial information
MCLU-ECBD	SVM	MCLU	No
MCLU-ECBD-GLCM			Yes
MCLU-ECBD-MRF			
MCLU-ECBD-PCM			
MCLU-Spa1-PCM			
MCLU-Spa2-PCM			
BT-ECBD	SMLR	BT	No
BT-ECBD-GLCM			Yes
BT-ECBD-MRF			
BT-ECBD-PCM			
BT-Spa1-PCM			
BT-Spa2-PCM			

1) *Design of the Active Learning Experiments:* In the experiments, for all the data sets, all the available samples were split into two sets, corresponding to learning set L^{tm} and test set L^{st} . The specific numbers of learning and test samples are shown in Table II. The initial training samples were selected randomly from the learning set U for both data sets, starting with five samples per class; in each iteration, 40 most uncertain samples are selected in the uncertain stage, and 10 samples are selected after diversity criterion. Thus, 10 samples are added at each iteration.

The AL algorithm runs until the accuracy of the learning sample set achieved full accuracy. As the bound of the ARL, full accuracy was achieved by the classification map \mathbf{I}_C^* generated with all the available sample sets. As the low bound on the accuracy, passive learning (or random sampling) was evaluated by random sampling from the pool of candidates. The performances were assessed by the average overall accuracy and its standard deviation in 10 independent runs. In each run, the initial training samples were chosen randomly.

At each iteration, all the approaches sampled the same number of pixels. In the experiments, SMLR algorithm and SVM were used to evaluate the performance of compared

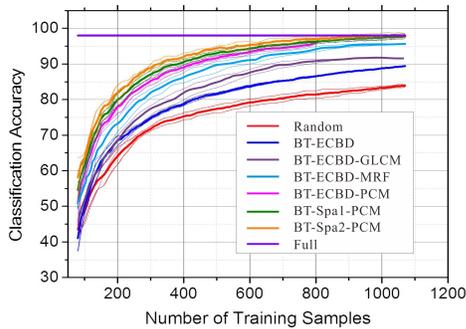


Fig. 13. AL curves achieved on the AVIRIS Indian Pines data set with different sample selection strategies.

methods. According to our experience, SMLR could present better performance on the hyperspectral image, and SVM does better on HSR image set. Thus, SMLR was used as the classifier for the hyperspectral image set, while SVM was used as the classifier for the HSR image set.

For the SVM classifier, a radial basis function (RBF) kernel was adopted, and the values for the regularization parameter C and the spread γ of the RBF kernel parameters were chosen by performing grid-search model selection. For the other classifier, SMLR, as noted in [17], the sparsity parameter was empirically fixed to 0.001 and the smoothness parameter to 2.

Batch mode selection was used in the experiments. As mentioned in Section II, BT and MCLU, which are two typical ways to select informative samples, were used as the most basic ways to select new samples. ECBD [23] was used as the typical method to remove the redundancy in the uncertain samples selected by the uncertainty criterion. Considering the structure of the batch mode selection, we refer to the traditional batch mode AL models as BT-ECBD and MCLU-ECBD. After we add the relearning models to the AL process, we refer to them as BT-ECBD-PCM and MCLU-ECBD-PCM. Furthermore, we used two selection methods to consider the spatial information, Spa1 and Spa2, and thus, we refer to the proposed methods as BT-Spa1-PCM and BT-Spa2-PCM, which are based on the BT selection method, and MCLU-Spa1-PCM and MCLU-Spa2-PCM, which are based on the MCLU selection method. PCM features were added in the feature set, which could contribute to improving the classification accuracy, as a result, reducing the iterations of AL model. In Section IV-C2, we also compare other ways to extract spatial features. For example, gray-level cooccurrence matrix (GLCM) is used to extract the texture information for image. Meanwhile, we also compare the MRFs, which is an effective postprocessing way, since they are able to smooth the classification. Thus, GLCM and MRFs are also added into AL model to compare with PCM feature. Table II shows the algorithms investigated in this paper.

2) *Experimental Results*: The experimental results obtained for all the data sets confirm the superiority of the ARL model. Compared with the AL model, the ARL model gives a faster rate of convergence; at the same time, the converged accuracy is higher than in the AL model.

Considering the AVIRIS Indian Pines image, the full accuracy for the SMLR classifier is equal to 98.2%. In Fig. 13,

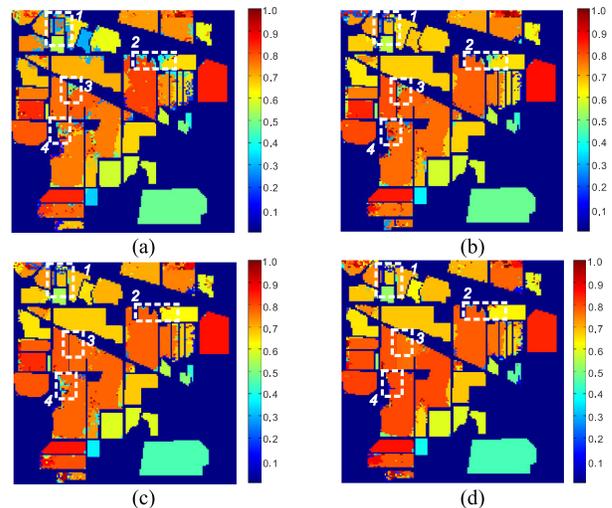


Fig. 14. Comparison of the postprobability maps of the second iteration according to the different strategies by adding 30 samples on the AVIRIS data set. (a) Initial classification map in the first iteration (OA: 68.49%). (b) BT-ECBD (OA: 72.31%). (c) BT-Spa1 (OA: 74.28%). (d) BT-Spa2 (OA: 75.67%).

the AL curves show the results as a function of the number of training samples, averaged over 10 runs of the algorithm, each with a different initial set. The shaded areas represent the standard deviation over the 10 considered runs. From the results shown in Fig. 13, it can be seen that the uncertainty criteria considering the spatial uncertainty show a better performance. BT-ECBD-GLCM and BT-ECBD-MRF could get higher accuracy compared with BT-ECBD; however, they are still lower than BT-ECBD-PCM. In particular, only about 900 samples (19% of the entire set) are needed by BT-Spa1-PCM, BT-Spa2-PCM, and BT-ECBD-PCM.

The performance of BT-Spa2-PCM is better than that of BT-Spa1-PCM. In BT-Spa1-PCM, spectral uncertainty is first used to select a batch of samples T_{spectral} from the unlabeled set U , and spatial uncertainty is used to reselect these uncertain samples from T_{spectral} . The spectral uncertainty therefore plays a dominant role in BT-Spa1-PCM. In BT-Spa2-PCM, however, the spatial uncertainty and spectral uncertainty are equally important. Thus, BT-Spa2-PCM can give a more informative sample set for the classifier.

To better understand the proposed strategies, a detailed comparison of the selection of new training samples with the different strategies is shown in Fig. 14, which depicts the spatial confidence dp_j for each pixel in the image. Fig. 14(b)–(d) shows the new spatial confidence dp_j after adding 30 new training samples by the different strategies (BT-ECBD, BT-Spa1, and BT-Spa2). We marked four regions with low spatial confidence in the first iteration and observed the improvement of these regions in the second iteration.

After adding 30 new training samples, the spatial confidence in rectangle 1 improves using all three strategies. However, when comparing the confidence improvement in rectangles 2 and 3, the superiority of BT-Spa1 and BT-Spa2 over BT-ECBD is obvious. Furthermore, in rectangle 4, only BT-Spa2 shows an obvious confidence improvement. Thus, BT-Spa2 shows

TABLE III
OA, KAPPA, AVERAGE CLASS ACCURACIES, AND STANDARD DEVIATIONS (Σ) FOR (a) 10 AND (b) 50 ITERATIONS
OF THE AL PROCESS ON THE AVIRIS INDIAN PINES DATA SET

Method	(a)			(b)		
	10 iterations			50 iterations		
	BT-ECBD-PCM	BT-Spa1-PCM	BT-Spa2-PCM	BT-ECBD-PCM	BT-Spa1-PCM	BT-Spa2-PCM
OA	78.76	81.19	82.83	90.09	92.23	92.50
σ_{OA}	3.37	3.07	4.15	3.18	3.51	3.95
Kappa	0.7649	0.7891	0.8041	0.8814	0.9086	0.9109
σ_{Kappa}	0.041	0.038	0.028	0.039	0.037	0.033
AA	81.51	83.55	84.95	90.32	92.09	92.37
σ_{AA}	4.13	3.95	3.45	3.89	2.53	2.89
Corn-notill	72.51	77.81	79.99	90.77	92.01	92.21
Corn-min	59.87	62.91	64.46	81.65	81.40	82.53
Corn	72.43	72.35	80.37	78.78	81.30	84.35
Grass/pasture	88.69	92.19	91.79	94.52	95.42	95.79
Grass/tree	95.91	94.17	94.32	97.72	97.23	99.55
Hay-windrowed	100.00	100.00	100.00	100.00	100.00	100.00
Soybeans-notill	73.31	71.41	71.70	81.68	84.21	85.56
Soybeans-min	72.77	74.81	80.38	90.60	92.32	90.82
Soybeans-clean	77.93	80.62	81.15	88.48	94.82	96.24
Wheat	100.00	100.00	100.00	100.00	100.00	100.00
Woods	92.26	92.98	93.59	93.98	95.66	95.30
Bldg-grass-tree-Drives	84.33	85.57	84.94	85.80	85.13	85.75

the best performance in selecting new samples, BT-Spa1 is second, and BT-ECBD last.

The obtained results are shown in greater detail in Table III. In particular, we consider the performances obtained after [see Table III(a)] 10 and [see Table III(b)] 50 iterations of the iterative process, which correspond to [see Table III(a)] 160 and [see Table III(b)] 560 samples used to train the classifier, respectively. From the results in Table II, we can see that BT-Spa1-PCM and BT-Spa2-PCM show higher classification accuracies than BT-ECBD-PCM. In the 10th iteration, the improvement in overall accuracy based on BT-Spa1-PCM is 2.3%, and for BT-Spa2-PCM, it is 4.3%. When the iterations increase to 50, the accuracy difference between the different selection strategies decreases, with improvements in overall accuracy for BT-Spa1-PCM of 1.5% and for BT-Spa2-PCM of 1.95%. From Table III, we can see that BT-Spa2-PCM shows the highest improvement. Compared with BT-ECBD-PCM, the superiority of BT-Spa1-PCM and BT-Spa2-PCM is obvious in the early iterations.

For the Pavia University data set, SMLR was used as the classifier, and the full accuracy for ARL is as high as 99.2%. The Pavia University data set contains rich spatial information, and Fig. 15 shows the learning curves of the Pavia University data set. Once again, the proposed BT-Spa1-PCM and BT-Spa2-PCM perform better than BT-ECBD-PCM. And the performance of BT-ECBD-PCM is better than BT-ECBD-GLCM and BT-ECBD-MRF. In particular, in the 19th iteration (about 273 samples), the improvement of BT-Spa1-PCM and BT-Spa2-PCM is more obvious, at about 2% and 3%, respectively.

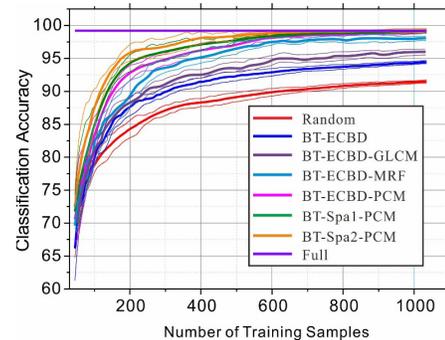


Fig. 15. Learning curves achieved on the Pavia University data set with the different selection strategies.

Table IV shows the overall accuracy, kappa, and averaged accuracy for each class, based on the different strategies after 10 iterations (145 samples) and 50 iterations (545 samples). From the numeric comparison, in the 10th iteration, the improvement in overall accuracy based on BT-Spa1-PCM is 2.29% and the improvement based on BT-Spa2-PCM is 3.32%. When the iterations increase to 50, the improvement in overall accuracy based on BT-Spa1-PCM is 1.4% and the improvement based on BT-Spa2-PCM is 1.92%. Overall, we can again see that BT-Spa2-PCM shows the best performance.

The learning curves for the two HSR data sets are shown in Fig. 16. For the QuickBird Wuhan data set, SVM was used as the classifier, and the full accuracy of ARL is 98.41%. As shown in Fig. 16(a), MCLU-Spa1-PCM needs about 400 samples to converge to full accuracy, and MCLU-Spa2-PCM

TABLE IV
OA, KAPPA, CLASS ACCURACIES, AND STANDARD DEVIATIONS (Σ) FOR (a) 10 AND (b) 50 ITERATIONS
OF THE AL PROCESS ON THE PAVIA UNIVERSITY DATA SET

Method	(a)			(b)		
	10 iterations			50 iterations		
	BT-ECBD-PCM	BT-Spa1-PCM	BT-Spa2-PCM	BT-ECBD-PCM	BT-Spa1-PCM	BT-Spa2-PCM
OA	93.23	94.39	95.76	96.34	97.34	97.39
σ_{OA}	3.94	3.23	3.02	3.31	3.52	3.62
Kappa	0.9123	0.9246	0.9413	0.9456	0.9573	0.9542
σ_{Kappa}	0.034	0.038	0.030	0.031	0.034	0.028
AA	93.72	95.61	96.21	96.09	97.57	97.94
σ_{AA}	3.72	3.17	3.26	3.28	3.25	3.13
Asphalt	94.34	95.99	97.15	96.05	97.72	97.81
Meadows	92.37	96.86	96.29	96.26	97.56	97.69
Gravel	83.81	85.93	86.77	93.12	96.01	96.45
Trees	90.97	92.40	94.14	93.78	97.27	98.37
Metal sheet	100.00	100.00	100.00	100.00	100.00	100.00
Bare soil	94.13	96.66	97.07	99.72	99.77	99.91
Bitumen	96.90	97.82	98.10	97.22	98.02	98.92
Bricks	94.34	95.99	97.15	96.05	97.72	97.81
Shadows	92.37	96.86	96.29	96.26	97.56	97.69

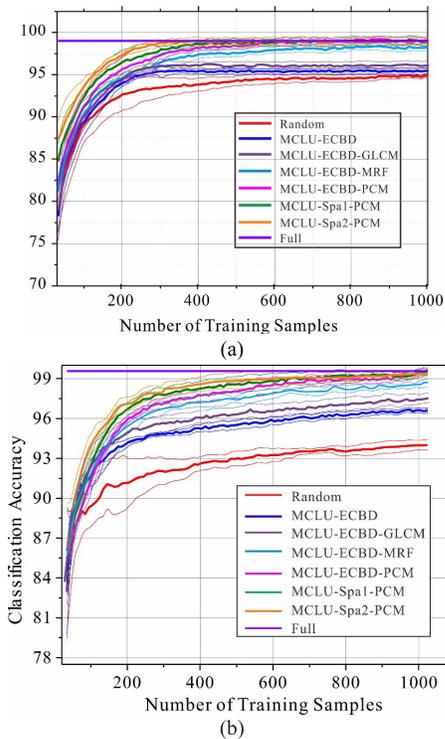


Fig. 16. Learning curves on the two HSR data sets with different selection strategies. (a) QuickBird Wuhan data set. (b) WorldView-2 Hainan data set.

needs only 380 samples, whereas MCLU-ECBD-PCM needs about 560 samples. As shown in Fig. 16(b), after adding 1000 samples, the full accuracy for ARL is 99.43%; however, the converged accuracy of MCLU-ECBD is about 97.21%, and the converged accuracies of MCLU-Spa1-PCM and MCLU-Spa2-PCM are about 99.72%. We can

see that the relearning process can effectively improve the converged accuracy. Meanwhile, the samples needed to obtain convergence are clearly fewer than in the original AL process.

The obtained results are shown in greater detail in Tables V and VI. In particular, we consider the performances obtained after [see Tables V(a) and VI(a)] 10 and [see Tables V(b) and VI(b)] 30 iterations of the iterative process, which correspond to [see Tables V(a) and VI(a)] 135 and [see Tables V(b) and VI(b)] 335 samples used to train the classifier, respectively. For the QuickBird Wuhan data set, in the 10th iteration, the improvement in overall accuracy for BT-Spa1-PCM is 1.10% and for BT-Spa2-PCM, it is 2.36%. In the 30th iteration, the improvement in overall accuracy based on BT-Spa1-PCM is 1.07% and the improvement based on BT-Spa2-PCM is 1.04%. BT-Spa2-PCM obtains the highest accuracy in all classes. The accuracy comparison shows that BT-Spa1-PCM and BT-Spa2-PCM can converge to a satisfactory accuracy faster, with fewer samples.

Considering the WorldView-2 Hainan data set, in the 10th iteration, compared with BT-ECBD-PCM, the accuracy improvement based on BT-Spa1-PCM is 0.33% and the improvement based on BT-Spa2-PCM is 0.96%. In the 30th iteration, the accuracy improvement based on BT-Spa1-PCM is 0.21% and the improvement based on BT-Spa2-PCM is 0.29%, which is a smaller improvement than the results obtained in the 10th iteration. We can see that the performance on the two HSR data sets again proves the effectiveness of the proposed ARL process and the corresponding spatial-spectral uncertainty criteria.

We also compare the processing time of different AL methods. There are two kinds of stop conditions for iterations. The first one is that the overall accuracy on the test sample set has achieved the given accuracy, while the second one is

TABLE V
OA, KAPPA, CLASS ACCURACIES, AND STANDARD DEVIATIONS (Σ) FOR (a) 10 AND (b) 30 ITERATIONS
OF THE AL PROCESS ON THE QUICKBIRD WUHAN DATA SET

Method	(a)			(b)		
	10 iterations			30 iterations		
	BT-ECBD-PCM	BT-Spa1-PCM	BT-Spa2-PCM	BT-ECBD-PCM	BT-Spa1-PCM	BT-Spa2-PCM
OA	94.94	96.04	97.30	97.88	98.95	98.92
σ_{OA}	2.821	3.156	3.304	3.329	3.415	3.051
Kappa	0.9276	0.9340	0.9431	0.9611	0.9664	0.9670
σ_{Kappa}	0.030	0.036	0.028	0.035	0.038	0.037
AA	95.26	97.14	97.87	97.64	99.04	99.58
σ_{AA}	3.89	3.47	3.09	3.42	3.27	2.89
Buildings	95.90	97.32	98.47	97.61	99.31	99.43
Roads	93.95	98.25	97.83	97.81	99.10	99.31
Soil	85.37	87.42	88.31	94.63	97.68	97.91
Grass	95.43	99.70	99.72	92.34	93.78	95.98
Shadow	100.00	100.00	100.00	100.00	100.00	100.00
Trees	95.84	98.08	98.60	100.00	100.00	100.00
Water	98.54	99.42	99.66	98.81	99.52	100.00

TABLE VI
OA, KAPPA, CLASS ACCURACIES, AND STANDARD DEVIATIONS (Σ) FOR (a) 10 AND (b) 30 ITERATIONS
OF THE AL PROCESS ON THE WORLDVIEW-2 DATA SET

Method	(a)			(b)		
	10 iterations			30 iterations		
	BT-ECBD-PCM	BT-Spa1-PCM	BT-Spa2-PCM	BT-ECBD-PCM	BT-Spa1-PCM	BT-Spa2-PCM
OA	96.46	96.93	97.42	98.08	98.29	98.37
σ_{OA}	3.52	3.25	3.14	3.42	3.21	3.15
Kappa	0.9454	0.9423	0.9524	0.9634	0.9712	0.9731
σ_{Kappa}	0.0321	0.034	0.031	0.035	0.032	0.035
AA	95.67	96.66	96.92	97.47	97.77	98.01
σ_{AA}	3.47	3.17	3.26	3.28	3.19	3.25
Buildings	94.29	95.80	97.09	96.01	97.59	97.87
Roads	92.19	96.79	96.19	96.28	97.54	97.69
Soil	83.91	86.08	86.69	93.11	96.10	96.36
Grass	93.95	94.32	94.21	90.74	92.37	94.40
Shadow	100.00	100.00	100.00	100.00	100.00	100.00
Trees	94.17	96.41	97.02	99.57	99.69	99.96
Water	96.96	98.01	98.25	97.14	98.08	98.88

TABLE VII
PROCESSING TIMES OF DIFFERENT AL STRATEGIES (SECONDS)
WHEN ACHIEVING A GIVEN ACCURACY

	AL	ARL	ARL+MultiSize	ARL+ spa 2
AVIRIS	2687.28 (#257)	57.67 (#38)	59.64 (#36)	54.58 (#34)
PAVIA	687.95 (#27)	197.67 (#14)	190.67 (#12)	182.64 (#11)
QB	--	489.76 (#47)	479.54(#44)	421.54(#42)
WV	--	506.8 (#35)	541.8(#32)	521.57(#32)

(The number in bracket is the iterations needed to satisfy given accuracy)

that the iterations have got the maximum. Thus, we evaluate the AL algorithms under these two kinds of the situations.

Table VII shows the processing time of AL curves under the condition that when the overall accuracy is getting the given

accuracy, the iteration could stop. And the given accuracy of AVIRIS and PAVIA_U data set is set as 96%. Meanwhile, the given accuracy of QB and WV data set is set as 97.5%. AL could not get the given accuracy on QB and WV data sets. Table VII shows the processing time of AL under fixed iterations.

From Tables VII and VIII, we can find that the processing time of ARL is larger than AL when iteration times are fixed. However, under the condition that the overall accuracy should achieve the given accuracy, the processing times of AL are larger than AL. Thus, we can find that though the processing time of ARL per iteration is larger than ARL, the total processing time is far less than the AL.

Comparing ARL and ARL + MultiSize, we could find that the processing times of these two ways are close, which

TABLE VIII
PROCESSING TIMES OF DIFFERENT AL STRATEGIES (SECONDS)
WHEN ACHIEVING FIXED ITERATIONS

	AL	ARL	ARL+MultiSize	ARL+ spa 2
AVIRIS	18.89(#60)	87.27(#60)	89.37(#60)	112.54(#60)
PAVIA	40.26(#25)	537.67(#25)	559.34(#25)	593.38(#25)
QB	17.54(#25)	212.76(#25)	233.57(#25)	274.23(#25)
WV	21.96(#25)	279.56 (#25)	295.21(#25)	310.45(#25)

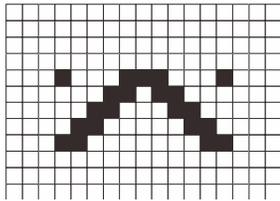


Fig. 17. Classification image with two classes (distinguished by black and white colors).

indicate that calculating the window size for each sample does not costing much time. Comparing ARL and ARL + Spa2, we can see that the processing times of two processing ways are close. Thus, the processing time of running new uncertain criterion is short. And we can find that ARL + spa2 needs less iterations. And corresponding processing time to get satisfied accuracy is least compared with other strategies in Table VII.

D. Performance of the Adaptive Multiwindow (Multiscale) ARL Model

ARL enhances the performance of the classifier by considering the spatial arrangement of the labels. However, the window size is the most influential parameter in the relearning model. In this regard, we further propose an adaptive multiwindow (multiscale) ARL model, which was validated by the experiment.

1) *Simulation Experiments*: Simulation experiments were conducted with a classification map with two classes, in order to allow us to intuitively analyze the results.

Fig. 17 shows a classification map consisting of a black class and a white class, where the black class shows rich structural information and the white class consists of a homogeneous patch, but with two outliers. This example demonstrates a typical challenge for $size_i$ to generate the PCM feature. On the one hand, the spatial structure information and class separability should be well preserved, and in this case, the PCM feature generated by a small $size_i$ would be more suitable. On the other hand, the noise and outliers should be removed to correct the wrong classification, so a larger $size_i$ would be more suitable in this circumstance to smooth the area. Thus, a simple and intuitive classification map was used to validate the superiority of the proposed adaptive relearning model.

We compared the PCM features generated by the multisize window and fixed size window. The PCM feature consists of three bands. In particular, band1 includes the spatial arrangement of a class2–class2 pair, band2 represents a class1–class2 pair, and band3 represents a class1–class1 pair. We represent the PCM feature visually in Fig. 18.

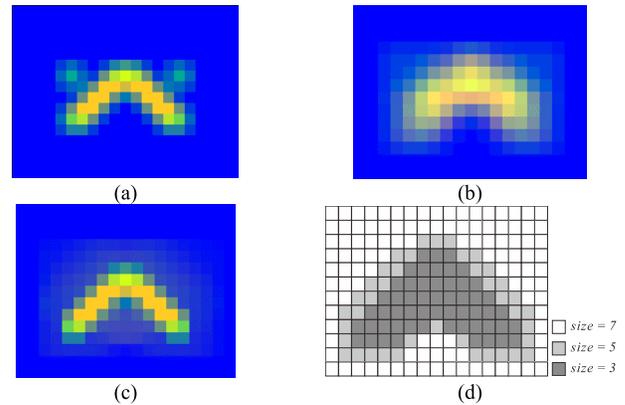


Fig. 18. PCM feature map generated by a fixed window size and the multisize window process (R: band1, G: band2, and B: band3). (a) $size_i = 3$. (b) $size_i = 5$. (c) Multisize window process. (d) $size_i$ for each pixel.

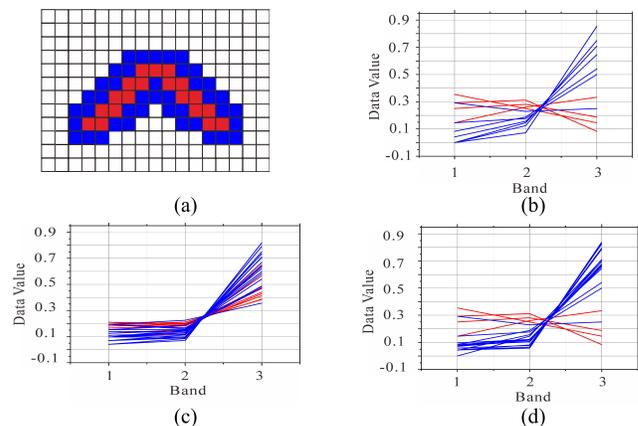


Fig. 19. Separability between the two classes when different window sizes are used to generate the PCM feature. (a) ROI. (b) $Size_i = 3$ (dis = 0.4643). (c) $Size_i = 5$ (dis = 0.1435). (d) Multiwindow (dis = 0.4601).

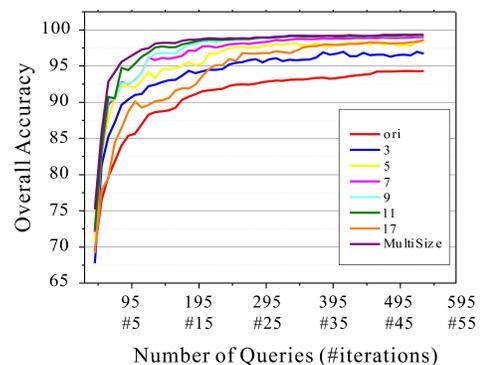


Fig. 20. Learning curves obtained using different window sizes for the Pavia University data set.

First, we present the denoising effects of the PCM feature. As shown in Fig. 18(a), when $size_i$ is set as 3, the two outliers are still apparent. As shown in Fig. 18(b), when $size_i$ is set as 5, the outliers can be removed; however, the border between the two classes is obscured. As shown in Fig. 18(c), the multisize window PCM feature can not only preserve the border between the two classes but also remove the outliers in the classification map. Fig. 18(c) represents the window

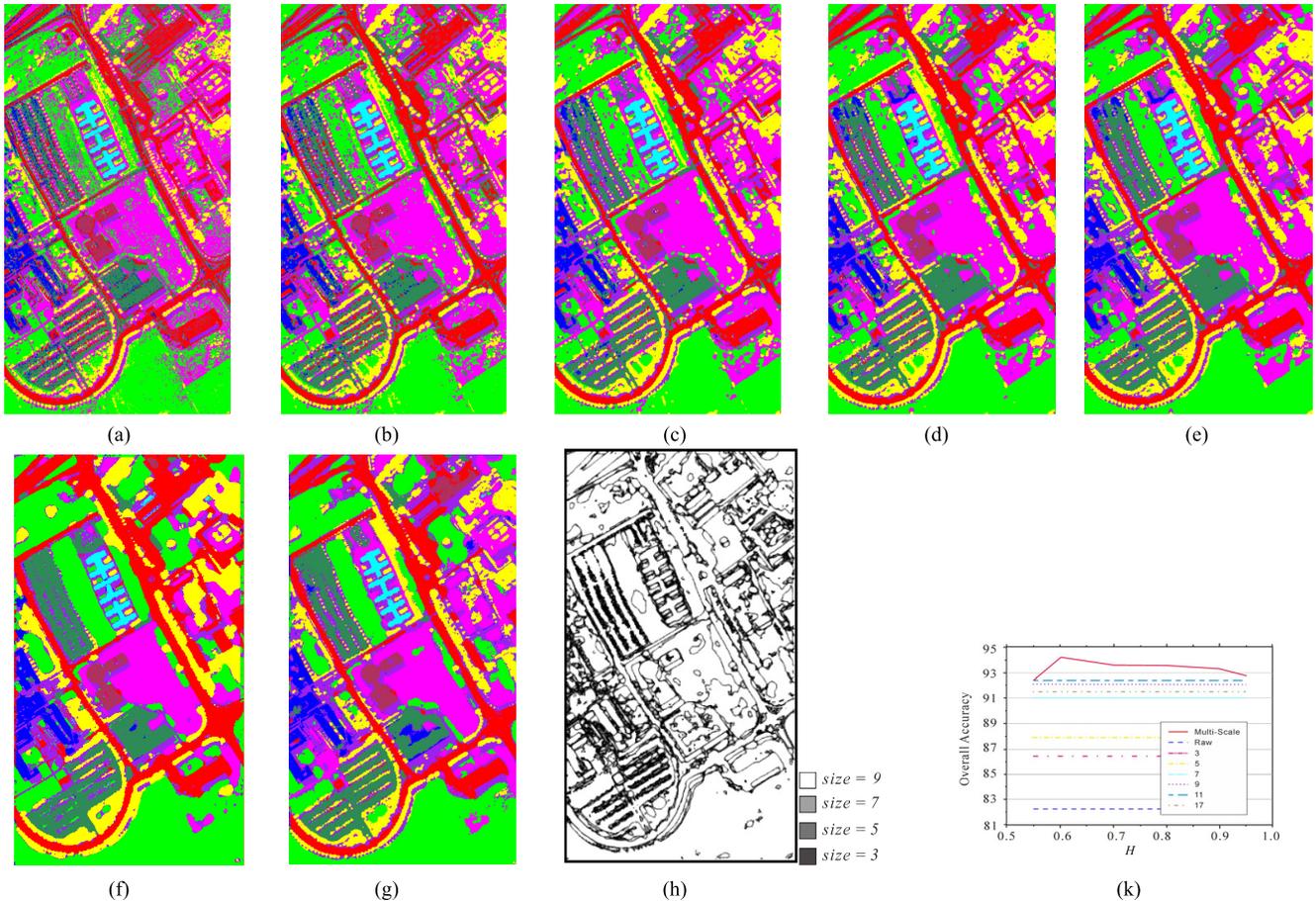


Fig. 21. Classification maps for the relearning algorithms with different window sizes for the Pavia University data set. (a) Raw. (b) $Size_i = 3$. (c) $Size_i = 5$. (d) $Size_i = 7$. (e) $Size_i = 9$. (f) $Size_i = 11$. (g) Multiwindow. (h) $SizeImage (H = 0.7)$. (k) Quantitative comparison.

size used for each pixel. Clearly, $size_i = 3$ and $size_i = 5$ are suitable for the border between two classes, and $size_i = 7$ is suited to the homogeneous regions.

To further investigate the separability between the two classes with the PCM feature, we selected border regions [the red and blue regions in Fig. 19(a)] between the two classes as the ROIs. As shown in Fig. 19(c), when $size_i = 5$, the PCM features of the blue region and the red region are seriously confused. As shown in Fig. 19(b) and (d), when $size_i = 3$ and the multiwindow process is used, the PCM features of the blue region and the red region can be well separated. This intuitive performance can also be proved by numerical calculation. We calculated the Euclidean distance between the mean vectors of the two classes. When $size_i = 3$, the Euclidean distance is 0.4696, which is close to the distance under the case when the multiwindow process is used. When $size_i = 5$, the Euclidean distance is 0.1375, which is clearly smaller than the cases above.

2) *Real Data Experiments With the Adaptive Multiwindow (Multiscale) Window ARL*: This section describes the real data experiments undertaken with the adaptive multiwindow ARL model. Two typical high-resolution images were used to validate the effectiveness of the adaptive multiwindow ARL model.

The SMLR classifier was used for the Pavia University data set and BT-Spa2-ECBD was used as the uncertainty criterion.

Meanwhile, the SVM classifier was used for the WorldView-2 Hainan data set and MCLU-Spa2-ECBD was used as the uncertainty criterion. We compared the adaptive multiwindow ARL model with different fixed window sizes ($size_i = 3, 5, 7, 9, 11$, and 17). For the adaptive multiwindow ARL model, parameter H was experimentally set as 0.7 for each iteration. We discuss below the influence of H on the experimental results.

For both data sets, starting from five samples per class and adding 10 samples at each iteration, the ARL algorithm was run until the learning sample set achieved full accuracy. To show a more detailed comparison for each data set, we also present the classification map in the 30th iteration for both data sets.

For the Pavia University data set, as shown in Fig. 20, the learning curves of the adaptive multiwindow ARL present a faster convergence rate than the fixed window size cases. In particular, when a larger window size is used, the model shows a smoothing effect on the classification map. However, when the window size is increased to 17, the performance of ARL decreases.

A visual comparison of the classification maps is shown in Fig. 21. The raw classification map shown in Fig. 21(a) was generated by 345 samples in the 10th iteration. Fig. 21(b)–(f) shows the classification maps generated by the relearning process. When a smaller window size is used, there is still

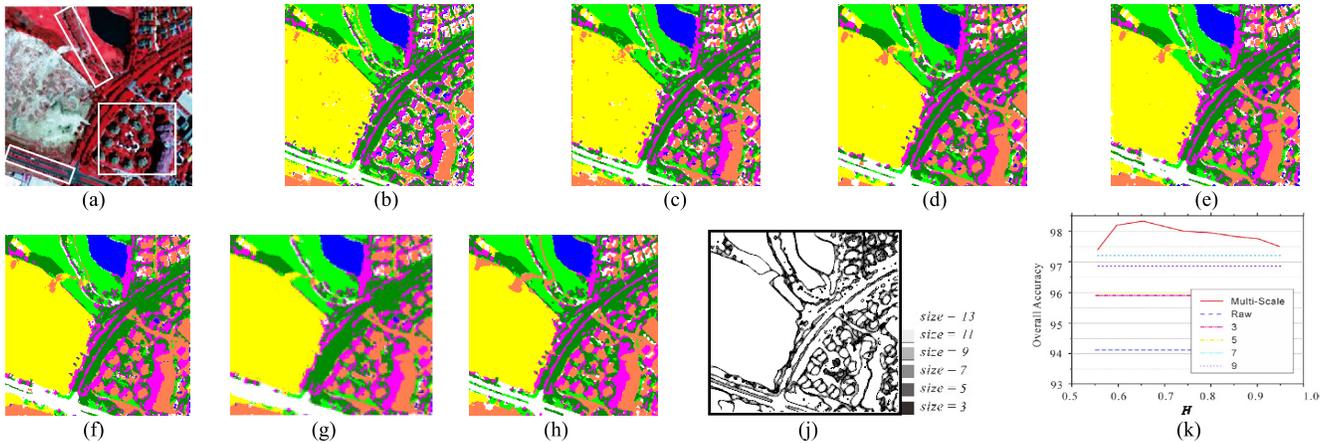


Fig. 22. Classification maps for the relearning algorithms with different window sizes for the WorldView-2 Hainan data set. (a) Original image. (b) Raw. (c) $Size_i = 3$. (d) $Size_i = 5$. (e) $Size_i = 7$. (f) $Size_i = 9$. (g) $Size_i = 11$. (h) Multiwindow. (i) SizeImage. (k) Quantitative comparison.

extensive noise in the classification map. However, when a larger window size is used, some detailed information disappears. As shown in Fig. 21(g), since the proposed adaptive relearning model can effectively determine the window size for each pixel, it can give a more satisfactory classification map, with less noise and richer ground detail. The sensitivity of parameter H in the adaptive multiwindow ARL model was also investigated, and is shown in Fig. 21(k). By setting different values of H from 0.55 to 0.95, it can be seen that the accuracy of the adaptive multiwindow ARL model is consistently higher than the accuracy of the fixed window size.

Considering the WorldView Hainan data set, as shown in Fig. 22, the classification maps produced by the adaptive multiwindow ARL model can not only preserve the detailed structures of the objects but also increase the separability between classes, and hence improve the classification accuracy. To show the details in the image, three rectangles are marked on the images. In the rectangle in the top of the image in Fig.22(a), two small paths stretch onto the grass. In the rectangle in the right of the image, there is a narrow green belt in the middle of the road. Meanwhile, in the rectangle in the left of the image, the shadow class is misclassified as the water class on the raw classification map. From the results shown in Fig. 22(b)–(g), the detailed information is still obvious when the window size is smaller than 7. When the window size is larger than 7, the road and the green belt are not very clear on the classification map, but the misclassification phenomenon between the water and shadow classes is reduced. However, when the adaptive multiwindow ARL model is used, the green belt and the two small paths are still clear, and the misclassification is reduced. From this performance, we can conclude that the adaptive multiwindow ARL model can obtain a smoother and more accurate classification map than when a fixed window size is used. As shown in Fig. 22(h), it can be seen that the accuracy of the adaptive multiwindow ARL model is consistently higher than the accuracy of the fixed window size.

Considering the Worldview Hainan data set, as shown in Fig. 23, the performance of multisize ARL is better than the model with fixed window size. When the fixed window size is getting larger, the overall accuracy is getting higher. However,

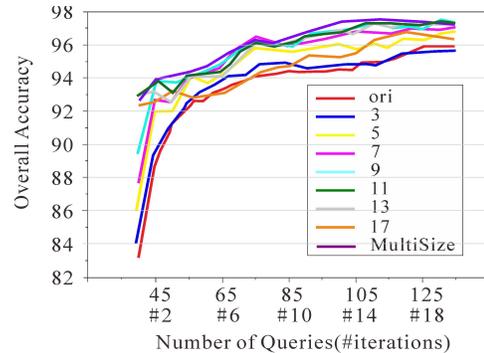


Fig. 23. Learning curves of the relearning algorithm with different window sizes for the WorldView-2 Hainan data set.

when the window size was approaching 13, the accuracy began to decrease.

V. CONCLUSION

Many researchers have paid close attention to improving AL performance by incorporating spatial information for remote sensing image classification. Most of the studies to date have focused on how to select the most informative samples when utilizing the spatial information, which is very important to increase the AL rate. However, with the addition of spatial information, the separability between similar classes can also be enhanced. In this paper, we have proposed a novel ARL model that embeds the relearning model into an AL framework. On the one hand, the relearning model can optimize the classification result by adding the PCM features, which are generated by heuristically exploiting the spatial arrangement between the class labels on the current classification map. However, this improvement is limited by the quality of the initial classification map. To improve the classification even further, AL is utilized to update the training sample set iteratively. Thus, AL improves the classification accuracy in terms of increasing the training samples; meanwhile, it also improves the classification result in terms of optimizing the data representation. It should be underlined that AL and the relearning model interact to promote each other. The experimental results prove that the ARL model shows a faster convergence speed with fewer samples.

Under this framework, we have further improved the performance of ARL in two ways. First, we have proposed new spatial–spectral criteria to select new samples under the ARL framework. The experimental results confirm that the proposed spatial–spectral selection criteria can provide faster learning rates than only considering the spectral uncertainty. Second, we have proposed a strategy to adaptively calculate the window size to enhance the stability and adaptability of the relearning model. The experimental results also confirm the effectiveness of this approach.

REFERENCES

- [1] P. Mitra, B. U. Shankar, and S. K. Pal, “Segmentation of multispectral remote sensing images using active support vector machines,” *Pattern Recognit. Lett.*, vol. 25, no. 9, pp. 1067–1074, 2004.
- [2] J. Zou, W. Li, X. Huang, and Q. Du, “Classification of hyperspectral urban data using adaptive simultaneous orthogonal matching pursuit,” *J. Appl. Remote Sens.*, vol. 8, no. 1, p. 085099, 2014.
- [3] X. Huang *et al.*, “Multiple morphological profiles from multicomponent-base images for hyperspectral image classification,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4653–4669, Dec. 2014.
- [4] X. Huang, Q. Lu, and L. Zhang, “A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas,” *ISPRS J. Photogramm. Remote Sens.*, vol. 90, no. 11, pp. 36–48, 2014.
- [5] D. Tuia, F. Pacifici, M. Kanevski, and W. J. Emery, “Classification of very high spatial resolution imagery using mathematical morphology and support vector machines,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3866–3879, Nov. 2009.
- [6] M. Pal, “Random forest classifier for remote sensing classification,” *Int. J. Remote Sens.*, vol. 26, no. 1, pp. 217–222, 2007.
- [7] X. Liu *et al.*, “A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects,” *Landscape Urban Planning*, vol. 168, pp. 94–116, Dec. 2017.
- [8] X. Li *et al.*, “A new global land-use and land-cover change product at a 1-km resolution for 2010 to 2100 based on human–environment interactions,” *Ann. Assoc. Amer. Geogr.*, vol. 107, no. 5, pp. 1040–1059, 2017.
- [9] Y. Fu, X. Zhu, and B. Li, “A survey on instance selection for active learning,” *Knowl. Inf. Syst.*, vol. 35, no. 2, pp. 249–283, 2013.
- [10] M. I. Mandel, G. E. Poliner, and D. P. W. Ellis, “Support vector machine active learning for music retrieval,” *Multimedia Syst.*, vol. 12, no. 1, pp. 3–13, 2006.
- [11] M. Prince, “Does active learning work? A review of the research,” *J. Eng. Educ.*, vol. 93, no. 3, pp. 223–231, 2004.
- [12] D. Tuia, E. Pasolli, and W. J. Emery, “Using active learning to adapt remote sensing image classifiers,” *Remote Sens. Environ.*, vol. 115, no. 9, pp. 2232–2242, 2011.
- [13] M. Volpi, D. Tuia, and M. Kanevski, “Memory-based cluster sampling for remote sensing image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3096–3106, Aug. 2012.
- [14] S. Tong and D. Koller, “Support vector machine active learning with applications to text classification,” *J. Mach. Learn. Res.*, vol. 2, pp. 45–66, Nov. 2002.
- [15] C. Campbell, N. Cristianini, and A. Smola, “Query learning with large margin classifiers,” in *Proc. 17th Int. Conf. Mach. Learn. (ICML)*, 2000, pp. 111–118.
- [16] G. Schohn and D. Cohn, “Less is more: Active learning with support vector machines,” in *Proc. 17th Int. Conf. Mach. Learn. (ICML)*, 2000, pp. 839–846.
- [17] J. Li, J. Bioucas-Dias, and A. Plaza, “Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.
- [18] T. Luo *et al.*, “Active learning to recognize multiple types of plankton,” *J. Mach. Learn. Res.*, vol. 6, pp. 589–613, Apr. 2005.
- [19] S. Rajan, J. Ghosh, and M. M. Crawford, “An active learning approach to hyperspectral data classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1231–1242, Apr. 2008.
- [20] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, “Active learning methods for remote sensing image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2218–2232, Jul. 2009.
- [21] W. Di and M. M. Crawford, “View generation for multiview maximum disagreement based active learning for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1942–1954, May 2012.
- [22] K. Brinker, “Incorporating diversity in active learning with support vector machines,” in *Proc. 17th Int. Conf. Mach. Learn. (ICML)*, 2003, pp. 59–66.
- [23] B. Demir, C. Persello, and L. Bruzzone, “Batch-mode active-learning methods for the interactive classification of remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1014–1031, Mar. 2011.
- [24] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-Mari, “A survey of active learning algorithms for supervised remote sensing image classification,” *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 606–617, Jun. 2011.
- [25] E. Pasolli, F. Melgani, D. Tuia, F. Pacifici, and W. J. Emery, “SVM active learning approach for image classification using spatial information,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2217–2233, Apr. 2014.
- [26] E. Pasolli, F. Melgani, D. Tuia, F. Pacifici, and W. J. Emery, “Improving active learning methods using spatial information,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2011, pp. 3923–3926.
- [27] S. Sun, P. Zhong, H. Xiao, and R. Wang, “An MRF model-based active learning framework for the spectral-spatial classification of hyperspectral imagery,” *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1074–1088, Sep. 2015.
- [28] J. Xu, R. Hang, and Q. Liu, “Patch-based active learning (PTAL) for spectral-spatial classification on hyperspectral data,” *Int. J. Remote Sens.*, vol. 35, no. 5, pp. 1846–1875, 2014.
- [29] M. Crawford, D. Tuia, and H. L. Yang, “Active learning: Any value for classification of remotely sensed data?” *Proc. IEEE*, vol. 100, no. 3, pp. 593–608, Mar. 2013.
- [30] Q. Shi, B. Du, and L. Zhang, “Spatial coherence-based batch-mode active learning for remote sensing image classification,” *IEEE Trans. Image Process.*, vol. 24, no. 7, pp. 2037–2050, Jul. 2015.
- [31] J. Munoz-Mari, D. Tuia, and G. Camps-Valls, “Semisupervised classification of remote sensing images with active queries,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3751–3763, Oct. 2012.
- [32] J. Li, H. Zhang, and L. Zhang, “Supervised segmentation of very high resolution images by the use of extended morphological attribute profiles and a sparse transform,” *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 8, pp. 1409–1413, Aug. 2014.
- [33] L. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, “Compression of hyperspectral remote sensing images by tensor approach,” *Neurocomputing*, vol. 147, pp. 358–363, Jan. 2015.
- [34] J. Li, H. Zhang, and L. Zhang, “A nonlinear multiple feature learning classifier for hyperspectral images with limited training samples,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2728–2738, Jun. 2015.
- [35] X. Huang, X. Liu, and L. Zhang, “A multichannel gray level co-occurrence matrix for multi/hyperspectral image texture representation,” *Remote Sens.*, vol. 6, no. 9, pp. 8424–8445, 2014.
- [36] Q. Zhang, L. Zhang, and X. Huang, “Classification of high-spatial resolution imagery based on distance-weighted Markov random field with an improved iterated conditional mode method,” *Int. J. Remote Sens.*, vol. 32, no. 24, pp. 9843–9868, 2011.
- [37] Y. Zhong, B. Zhao, and L. Zhang, “A hybrid object-oriented conditional random field classification framework for high spatial resolution remote sensing imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7023–7037, Nov. 2014.
- [38] X. Huang, C. Xie, X. Fang, and L. Zhang, “Combining pixel- and object-based machine learning for identification of water-body types from urban high-resolution remote-sensing imagery,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 2097–2110, May 2015.
- [39] Y. Zhong, B. Zhao, and L. Zhang, “Multiagent object-based classifier for high spatial resolution imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 841–857, Feb. 2014.
- [40] C. Persello and L. Bruzzone, “Active and semisupervised learning for the classification of remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6937–6956, Nov. 2014.
- [41] X. Huang, Q. Lu, L. Zhang, and A. Plaza, “New postprocessing methods for remote sensing image classification: A systematic study,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7140–7159, Nov. 2014.
- [42] T. Joachims, “Making large-scale SVM learning practical,” in *Advances in Kernel Methods—Support Vector Learning*. Dortmund, Germany: Univ. Dortmund, 1999, pp. 41–56.

- [43] D. Böhning, "Multinomial logistic regression algorithm," *Ann. Inst. Stat. Math.*, vol. 44, no. 1, pp. 197–200, Oct. 1992.
- [44] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple-kernel classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3780–3791, Oct. 2010.



Qian Shi received the B.S. degree in sciences and techniques of remote sensing from Wuhan University, Wuhan, China, in 2010, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, in 2015.

She is currently a Post-Doctoral Fellow with the School of Geography and Planning, Sun Yat-sen University, Guangzhou, China. Her research interests include remote sensing image classification, including active learning, transfer learning, and dimension reduction for hyperspectral image.



Xiaoping Liu (M'13) received the B.S. degree in geography and the Ph.D. degree in remote sensing and geographical information sciences from Sun Yat-sen University, Guangzhou, China, in 2002 and 2008, respectively.

He is currently a Professor with the School of Geography and Planning, Sun Yat-sen University. He has authored two books and over 100 articles. His research interests include image processing, artificial intelligence, and geographical simulation.



Xin Huang (M'13–SM'14) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009.

He was with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He is currently a LuoJia Distinguished Professor with Wuhan University, where he teaches remote sensing, photogrammetry, and image interpretation. He is the Founder and the Director of the Institute of Remote Sensing Information Processing, School of Remote Sensing and Information Engineering, Wuhan University. He has been supported by the New Century Excellent Talents in University from the Ministry of Education of China in 2011, the China National Science Fund for Excellent Young Scholars in 2015, and the National Program for Support of Top-notch Young Professionals in 2017. He has authored or co-authored over 100 peer-reviewed articles (SCI papers) in the international journals. His research interests include remote sensing image processing methods and applications.

Dr. Huang was a recipient of the Boeing Award for the Best Paper in Image Analysis and Interpretation from the American Society for Photogrammetry and Remote Sensing in 2010 and the National Excellent Doctoral Dissertation Award of China in 2012. In 2011, he was recognized by the IEEE Geoscience and Remote Sensing Society (GRSS), as a Best Reviewer for IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He was the Winner of the IEEE GRSS 2014 Data Fusion Contest. He was the Lead Guest Editor of the special issue on information extraction from high spatial resolution optical remotely sensed imagery for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (vol. 8, no.5, May 2015) and the Lead Guest Editor of the special issue on Sparsity-Driven High-Dimensional Remote Sensing Image Processing and Analysis for the *Journal of Applied Remote Sensing* (vol.10, no.4, Oct. 2016). Since 2014, he has been an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. Since 2016, he has been an Associate Editor of *Photogrammetric Engineering and Remote Sensing*.