IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 15, 2021, accepted April 8, 2021, date of publication April 20, 2021, date of current version April 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3074405

Spectral-Spatial Active Learning With Structure Density for Hyperspectral Classification

QIANMING LI[1], (Graduate Student Member, IEEE), BOHONG ZHENG[1], AND YUSHENG YANG[2], (Member, IEEE)

[1]School of Architecture and Art, Central South University, Changsha 410083, China
[2]School of Landscape Architecture, Central South University of Forestry and Technology, Changsha 410004, China

Corresponding author: Yusheng Yang (lini_zn@csu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51478470, in part by the Hunan Province Postgraduate Research and Innovation Project under Grant CX2018B075, and in part by the Key Project of Central South University Postgraduate Independent Innovation Project under Grant 502221802.

ABSTRACT In this paper, a spectral-spatial active learning (AL) method is proposed based on an up-to-date unlabeled samples sampling strategy concentrated on the structure density supported by breaking ties. The proposed sampling criterion in AL is used for hyperspectral image classification, which involves several steps: First, superpixel segmentation algorithm is conducted on the HSI to cluster pixels with similar spectral-spatial signature into the same superpixel block. Then, density peak clustering technique is performed on the each superpixel block to obtain structure density of the pixels. Meanwhile, probability-based classifier is employed to achieve the probability distributions of pixel. Next, breaking ties (BT) score of each pixel can be calculated by exploiting the probabilities. Additionally, a fusion mechanism is introduced to select the unlabeled samples with representativeness and informativeness advantages by employing the BT-assisted structure density (SD sampling criterion) of each pixel. Finally, the samples with manual labeled class labels are put into the training set to retrain the classifier. Experimental results manifest that the proposed SD-based sampling criterion in active learning can significantly improve the classification accuracy in few labor costs. Thus, it has certain feasibility in practical application.

INDEX TERMS Hyperspectral imagery, active learning, structure density, breaking ties, sampling criterion.

I. INTRODUCTION

With the advance of spectroscopy sensor technology, hyperspectral image (HSI) with high spectral dimensionality and spatial resolution has been constantly becoming more available. Considering the abundance of spatial and spectral information, numerous classification algorithms exploiting remote sensing images have played a primary role in a variety of applications, such as precision agriculture [1], [2], environmental monitoring [3], land cover [4], and urban expansion [5].

For HSI classification [6]–[9], traditional supervised classifiers such as kernel-based techniques [10]–[12], sparse representation [13]–[15], genetic algorithm [16], support vector machines (SVMs) [17], neural networks [18]–[20], and Bayesian estimation method [21] have given evidence of rock-solid properties in aspects of high classification accuracies. However, when the number of labeled samples is limited, most of the classifiers mentioned above cannot achieve a well-pleasing classification performance due to the constraints of sample conditions. Basically, there is a challenging task that obtaining the labeled samples of land cover classification is costly and time-consuming in practical applications as some districts may be relatively inaccessible. Meanwhile, a larger number of unlabeled samples with rich feature information did not play its role improving performance of classifier in supervised learning. In order to overcome the problem, many scholars and researchers in hyperspectral domain are willing to devote oneself to study advanced machine learning and classification methods, such as semi-supervised learning (SSL) [22], active learning (AL) [23], [24], semi-supervised active learning (SSAL) [25]–[28], and spectral-spatial classification [29]–[32].

The focus of AL is that try to conquer the labeling choke point by asking queries in the form of unlabeled samples to be annotated by an oracle [33]. In particular, AL is successfully applied to deal with classification problem in the HSI applications as a new machine learning

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao.

method [34], [35], its main goal is to effectively find high-information samples in the unlabeled sample dataset and then retrain the pixelwise classifier efficiently by iteratively expanding the labeled samples with an iterative manual labeling processing. By choosing unlabeled samples in a smart active query strategy, the amount of initial labeled samples (training samples) required for training a robust classifier can be observably reduced, consequently, lessening the labeling costs and time [36]. For instance, Rajan *et al.* introduced active learning technological means to classify hyperspectral data [34]. Di *et al.* proposed an AL method with multi-view adaptive inconsistent and conducted an in-depth study on view generation [37]. Sun *et al.* investigated AL method with Gaussian process classifier for the HSI classification [38]. In general, compared with traditional HSI classifiers, the above method has the unique advantages of automatically selecting unlabeled samples under the condition of labeled samples, and has strong discriminative ability and higher classification accuracy.

In addition to research on how to solve the HSI classification with AL, another popular research topic for robust sampling criterion of AL is how to get the utmost out of the informativeness information of unlabeled samples to find the samples enhancing the generalization ability of the model and further improve the classification accuracy. Specifically, uncertainty sampling is widely used as a simplest and most currently employed query framework for a probabilistic classifier in AL [39]. In uncertainty sampling framework, the concern of an active learner is how to label the unlabeled samples about which it is least certain. Recently, the some uncertainty sampling criterion are concerned and introduced into AL methods, such as mutual information (MI) [40], breaking ties (BT) [41], and modified breaking ties (MBT) [42]. As the Fig.1 seen, above three sampling criterion and SD sampling criterion have been graphical description with a toy example. 1) Due to the MI criterion focuses on the most complex area, the sample selected by MI criterion lies in the boundary of the four regions. But it is easy to generate bias sampling [40]. 2) The samples which come from the boundary of the two regions or four regions have been selected by BT criterion. Because the BT criterion aims to correct aforementioned disadvantage of the BT sampling tactics, by adding the posterior of the second most likely label. The BT sampling strategy mainly pay attention to the boundary district between two classes, which aims to composition of training set more diversified, but it may be trapped within the boundary of a single class [41]. 3) For the MBT criterion, the samples are located at the boundary of the two regions. The MBT criterion considering all the class boundaries was proposed in order to promote more diversity of the sampling process [42]. However, for these sampling criterion, these methods only pay attention to the diversity of samples but ignore the local representativeness of samples.

In this paper, an new AL strategy is proposed based on a novel unlabeled samples sampling strategy concentrated on the BT-assisted structure density (SD). Unlike the above

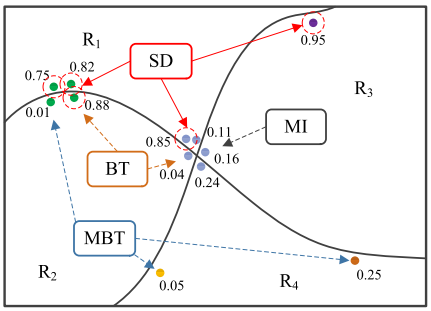


FIGURE 1. Graphical description of the MI, BT, MBT, and SD sampling criterion with a toy example.

traditional methods, the proposed sampling criterion introduces the density information to constrain the selection of unlabeled samples, so that the selected unlabeled samples can not only improve the diversity of training samples but also ensure their own local representation (see the graphical description of SD in Fig.1). The proposed SD sampling criterion in active learning is used for the HSI classification, and its contributions are as follows: First, superpixel segmentation is conducted on 2-D false-color image constructed by first three principal components of the original HSI, and the pixels with similar spectral-spatial structure are clustered into the same homogeneous region. Then, density peak clustering algorithm is performed on the each superpixel block to acquire structure density of the pixels with representativeness. Meanwhile, probability-based classifier, i.e., SVM, EPF, and LORSAL-ERW, is employed to achieve the probabilities of pixels in the hyperspectral image. Next, according to the probability distribution, the BT score of each pixel is used to find unlabeled samples with information content. Additionally, a fusion mechanism is introduced that combines the structural density and BT value of each pixel to select unlabeled samples with representative and information advantages. Please note that, LORSAL-ERW is employed to satisfy active learning strategy in this paper. Finally, the samples with manual labeled class labels are put into the training set to retrain LORSAL-ERW classifier. Experiments indicate that the spectral-spatial AL method can select valuable unlabeled samples with much less artificial cost needed and promote the classification accuracy of classifier obviously. Therefore, the proposed method will have a certain value in practical application.

Section II describes the superpixel segmentation, and breaking ties sampling. Section III introduces the proposed structural density sampling criterion based active learning method for HSI classification in detail. Experiment results are presented in Section IV. Finally, conclusions are given in Section V.

II. RELATED WORK

A. SUPERPIXEL SEGMENTATION

Graph theory based entropy rate superpixel (ERS) [43] has been widely applied in remote sensing image processing. Its advantage compared to simple linear iterative

clustering (SLIC) algorithm [44] is that superpixel blocks generated by the ERS algorithm can obtain adaptive size and shape according to various spatial information. So assume that the total number of superpixel blocks is K and the ith superpixel block is referred to as S_i, an image I can be compactly represented as follows:

$$I = \bigcup_{i=0}^{K} S_i, \text{ and } S_i \cap S_j = \varnothing, \quad (i \neq j) \tag{1}$$

As shown in the superpixel generation process, the ERS algorithm that graph-based clustering firstly maps the base image I to a graph $G = (V, E)$, where V is the vertex set corresponding pixels of the image, and E is the vertex set, which can measure the pairwise similarities between nearby pixels. Then, a subset of edges $Q \subseteq E$ is picked to distribute the base image I into K linked subgraphs. Next, an entropy rate term $H(Q)$ and a balancing term $B(Q)$ are introduced into the superpixel segmentation in order to achieve the homogeneous superpixel block as follows:

$$H(Q) = -\sum_i \mu_i \sum_j p_{i,j}(Q) \log(p_{i,j}(Q)) \tag{2}$$

$$\begin{aligned} B(Q) &\equiv H(Z_Q) - N_Q \\ &= -\sum_i p_{Z_Q}(i) \log\left(p_{Z_Q}(i)\right) - N_Q \end{aligned} \tag{3}$$

where $p_{i,j}$ refers to the transition probability function and μ_i is a stationary distribution, N_Q refers to the number of connected components in the graph G, and Z_Q be the distribution of the cluster membership. Finally, $H(\cdot)$ and $B(\cdot)$ are linearly combined into a objective function to generate superpixel blocks as follows:

$$\max_Q \{H(Q) + \eta B(Q)\} \quad \text{subject to} \quad Q \subseteq E \tag{4}$$

where the parameter, $\eta > 0$, is the weight controlling the balancing term in the objective function. And the greedy algorithm can effectively solve optimization problems [45], [46].

B. BREAKING TIES SAMPLING

To increase the sample diversity of the training set, BT sampling in active learning [41] focuses on more information samples, which are usually located in the boundary area between the two classes. The BT score for sample x_t is defined as:

$$\widehat{\boldsymbol{x}}_t^{BT} = \arg\min_{x_t \in \mathcal{C}} \left\{ \max_{c \in \kappa} = \; p(y_t = c | x_t, \widehat{\boldsymbol{\omega}}) - \max_{c \in \mathcal{L} \backslash \{c^+\}} p(y_t = c | x_t, \widehat{\boldsymbol{\omega}}) \right\} \tag{5}$$

where $c^+ = \arg\max_{c \in \mathcal{L}} p\left(y_t = c | x_t, \widehat{\boldsymbol{\omega}}\right)$ is the largest posterior probability based the most probable class label for sample x_t, $\mathcal{C}$ represents the sample candidate set, $\widehat{\boldsymbol{\omega}}$ refers to supervised learning of logistic regressor set $\boldsymbol{\omega}$, y_t refers to the class label of tth sample, c represents cth class in class label set $\mathcal{L}$, and $\mathcal{L}\backslash\{c^+\}$ indicates a class label that does not contain c^+ in $\mathcal{L}$.

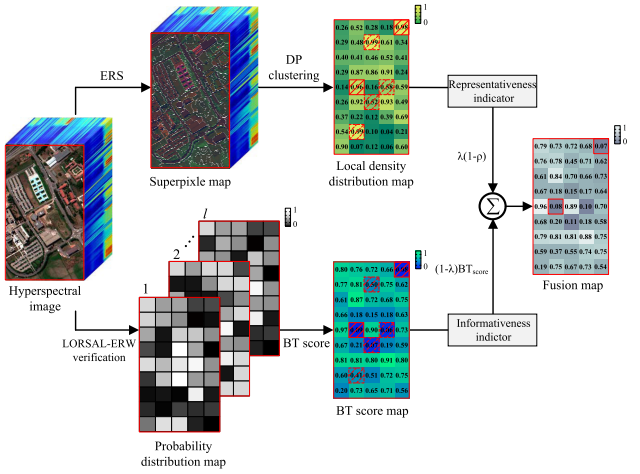


FIGURE 2. Outline of the proposed SD sampling criterion in hyperspectral classification.

III. PROPOSED METHOD

The proposed SD sampling criterion-based AL method for HSI classification is given in detail, which consists of the two main steps, i.e., SD sampling strategy and active learning strategy using LORSAL-ERW classifier. Note that the other probability-based classifier is acceptable here, but this paper sets LORSAL-ERW as the default classifier. In addition, Fig. 2 shows outline of the proposed SD sampling criterion in hyperspectral classification and the pseudo-code of SD sampling criterion based AL strategy is illustrated in algorithm I to clearly describe application of the SD sampling criterion in hyperspectral image classification.

A. STRUCTURAL DENSITY SAMPLING STRATEGY

For HSIs, the same type of pixels usually represent the same ground materials and have the same spectral characteristics. Given an original HSI $I \in \mathbb{R}^{M \times N}$, M and N respectively refer to the amount of samples and dimension of a sample in the I, the first three principal components I_{3d} of the hyperspectral image I can be initially achieved based on PCA algorithm as follows:

$$I_{3d} = \text{PCA}(I) \tag{6}$$

The I_{3d} based 2-D false-color map I_{pc} then is exploited as the input of ERS algorithm to segment I_{pc} into K 2-D nonoverlapping superpixel blocks $S_{block} = \{S_i\}_{i=1}^K$. Next, $S'_{block} = \{S'_i\}_{i=1}^K$ can be determined by mapping S_{block} of I_{pc} to the original HSI I, namely, the boundaries of superpixel blocks are migrated from I_{pc} to I as follows:

$$S_{block} = f_{\text{ERS}}(I_{pc}); \quad f_{\text{ERS}} : I_{pc} \to I \tag{7}$$

After that, density $\boldsymbol{\rho} = [\rho^1, \rho^2, \ldots, \rho^K]$ of pixel within superpixel blocks can be acquired by employing the DP clustering algorithm. The details are as follows:

1) *Compute Correlation Coefficient*: The correlation coefficient is a spectral metric and widely used in HSI

supervised classification [47]. In the SD sampling criteria, the Euclidean metric is replaced by the correlation coefficient, which is formulated by:

$$Cor^i = \frac{cov\left(x_u^i, x_v^i\right)}{\sqrt{var(x_u^i)} \cdot \sqrt{var(x_v^i)}} \tag{8}$$

where Cor^i is a symmetric matrices of correlation coefficient for ith superpixel block, x_u^i and x_v^i respectively refer to uth and vth in the ith superpixel block, $cov(\cdot)$ and $var(\cdot)$ respectively represent the covariance and variance of the pixel vector.

2) *Obtain density of pixel*: Calculating the density of pixels within each superpixel block based on the Cor^i includes major two steps: Firstly, the cut-off correlation coefficient Λ_{dc}^i is defined by:

$$\Lambda_{dc}^i = \Omega^i(\xi) \quad s.t. \quad \xi = < \frac{N_i \cdot (N_i - 1)}{100} \cdot \beta > \tag{9}$$

where $\Omega^i(\xi)$ is a vector that consist of ascending non-zero elements of upper triangular in the symmetric matrices Cor^i, β represents a default free parameter in the DP clustering algorithm [48], N_i on behalf of the amount of pixels of the ith superpixel block, and $< \cdot >$ refers to the round operation. Secondly, the structure densities $\boldsymbol{\rho}^i$ of the samples of tth superpixel block can be calculated based on Λ_{dc}^i as follows:

$$\boldsymbol{\rho}^i = \sum e^{-\left(\frac{Cor^i}{\Lambda_{dc}^i}\right)} \tag{10}$$

Finally, structure density distribution map $\boldsymbol{\rho} = [\boldsymbol{\rho}^1, \boldsymbol{\rho}^2, \ldots, \boldsymbol{\rho}^K]$ will be obtained by incorporating $\boldsymbol{\rho}^i$ into the corresponding position in HSI $\mathbf{I}$.

B. ACTIVE LEARNING STRATEGY

Here, many typical spectral classifiers, such as SVM, EPF, can be exploited to estimate the probability distribution. However, considering the performance of AL classification based on SD sampling criteria, we set the LORSAL-ERW classifier as the default classifier to achieve probability. Assuming that $T_\tau = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \in \left(\mathbb{R}^M \times \mathcal{L}\right)^n$ is initial training set, $T_e = \varnothing$ is initial extended training set, n is the total number of samples in the initial training set, $\mathcal{L} = \{1, 2, \ldots, l\}$ represents the label set corresponding to training set. The probability map $\mathbf{O} = \{\mathbf{O}_1^t, \mathbf{O}_2^t, \ldots, \mathbf{O}_l^t\}_{t=1}^N$ is first achieved by exploiting the LORSAL-ERW classifier to infer the class distributions. Then, the classification map will be obtained by selecting the maximum of probability.

Next, the global distribute map of braking ties scores BT_{score} can be easily achieved by using the solution of (9) based on the probability map $\mathbf{O}$. Here, taking into account some of the aforementioned limitations BT sampling, the BT_{score} is first incorporated structural density distribution $\boldsymbol{\rho}$ to construct a comprehensive consideration [see Fig. 2] $\mathcal{F}_t$ as follows:

$$\mathcal{F}_t = (1 - \lambda) \cdot BT_{score} + \lambda \cdot (1 - \hat{\boldsymbol{\rho}}) \tag{11}$$

Algorithm 1 SD sampling criterion based AL strategy.

Input:
Initial training samples:
$T_\tau \equiv \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \in \left(\mathbb{R}^M \times \mathcal{L}\right)^n$.
Initial extended training samples: $T_e = \varnothing$.
Original hyperspectral image: $I \in \mathbb{R}^{M \times N}$.
Active query size: u.
Stopping criterion for the total number of samples: U.

1: Construct 2-D false-color I_{pc} with PCA based first three principal components.
2: Segment I_{pc} with ERS to obtain the 2-D nonoverlapping superpixel blocks $\mathrm{S}_{block} = \{\mathrm{S}_i\}_{i=1}^K$.
3: Map the S_{block} to I to obtain $\mathrm{S}'_{block} = \{\mathrm{S}'_i\}_{i=1}^K$.
4: Compute pixel density ρ for S'_{block} based on DP.
5: **Repeat**
6: Train classifier with $T_\tau \cup T_e$ to obtain initial probability map $\mathbf{O}$.
7: Calculate the global BT scores of I based on the $\mathbf{O}$.
8: Introduce the structural density of superpixels and perform decision fusion $\mathcal{F}_m = (1-\lambda) \cdot BT_{score} + \lambda \cdot (1 - \hat{\boldsymbol{\rho}})$.
9: Select u data $T_e = \{x_i\}_{i=1}^u$ from candidate set $\mathcal{C}$ based on the $\mathcal{F}_m$ active query strategy.
10: Label T_e by human experts, and updata $T_\tau \equiv T_\tau \cup T_e$.
11: **Until** Stopping criteria is met, i.e., $sizeof\,(T_\tau) > U$.
12: **Return** Classifier

where $\hat{\boldsymbol{\rho}} \in [0, 1]$ is the normalized expression of $\boldsymbol{\rho}$, λ is a tradeoff parameter that control the balance between informativeness and confidence information. From the solution of (9), it can be seen that BT sampling favors samples with a lower BT_{score}. Therefore, the introduction of $(1 - \hat{\boldsymbol{\rho}})$ is to comprehensively consider the sample selection method of BT sampling criteria. In addition, through comprehensive consideration score, the active query size of extended sample set T_e is given human experts to label its class label. Finally, this labeled extending samples is put into T_τ to train classifier until satisfied stopping criterion of active learning as follows:

$$T_\tau \equiv T_\tau \cup T_e \quad \text{subject to} \quad sizeof\,(T_\tau) \leq U \tag{12}$$

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTAL DATASETS

In order to verify the effectiveness of the proposed SD sampling criterion, experiments are performed on four hyperspectral datasets[1], i.e., Indian Pines scene, Salinas scene, University of Pavia scene, and Center of Pavia scene. The details of these scenarios are as follows:

1) The Indian Pines image is over the Indian Pines test site in the northwest of Indian, which was acquired by the airborne visible/infrared imaging spectrometer (AVIRIS) sensor. The image has a spatial dimension of 145 × 145 and 220 spectral bands with a spatial

[1]Datasets can be downloaded at: http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

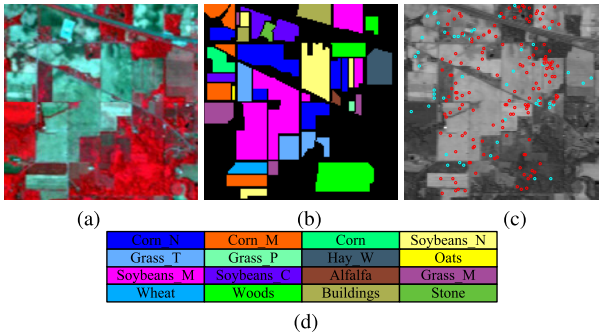


FIGURE 3. Indian Pines dataset. (a) False-color composite. (b) Reference data. (c) Distribution of 48 initial training samples (cyan) and 200 new labeled samples (red) for a single experiment. (d) Color code.

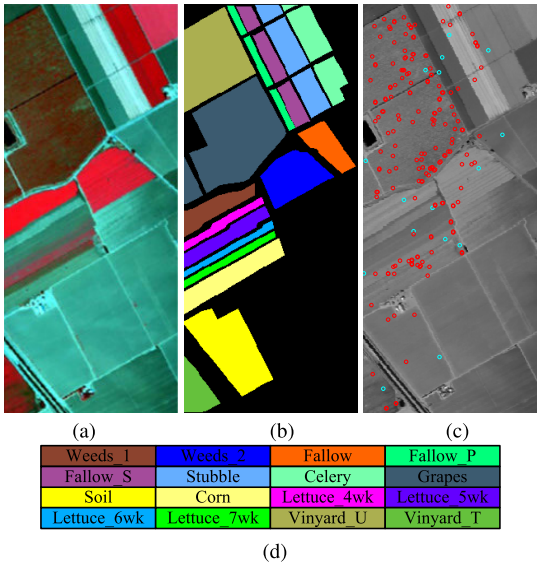


FIGURE 4. Salinas dataset. (a) False-color composite. (b) Reference data. (c) Distribution of 16 initial training samples (cyan) and 200 new labeled samples (red) for a single experiment. (d) Color code.

resolution of 20 *m* per pixel (20 water absorption channels were removed before experiments). As the scene was captured in June, some crops, such as corn and soybean, are still in the early stage of growth. Fig. 3 shows the false-color composite of the Indian Pines image, the corresponding reference data, and distribution of samples in a single experiment for comparative analysis. In the reference classification map, the scene is divided into 16 different classes.

2) The Salinas image was collected by the AVIRIS sensor over the Salinas Valley, California; it has 224 bands of size 512 × 217 pixels. In the experiments, 20 water absorption and noisy bands (no.108-112, 145-167, and 224) have been removed. The false color composite of the Salinas image and the reference classification map are shown in Fig. 4(a) and (b), which includes 16 different classes.
3) The University of Pavia image was acquired with the ROSIS 03 sensor over the campus at the University of Pavia, Italy. The image is of size 610 × 340×120,

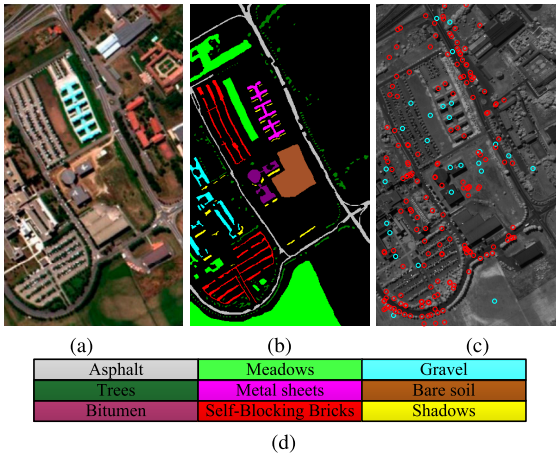


FIGURE 5. University of Pavia dataset. (a) False-color composite. (b) Reference data. (c) Distribution of 27 initial training samples (cyan) and 200 new labeled samples (red) for a single experiment. (d) Color code.

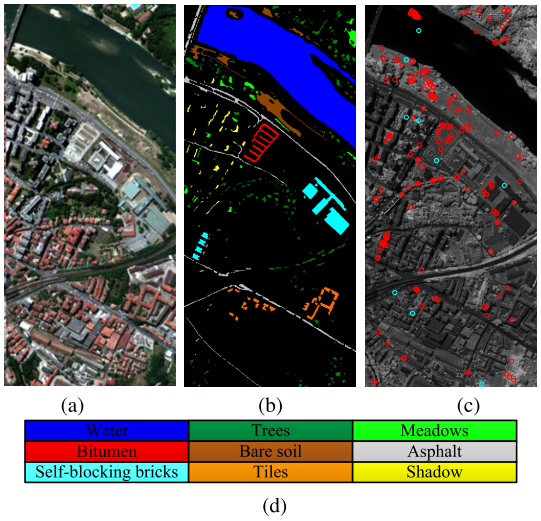


FIGURE 6. Center of Pavia dataset. (a) False-color composite. (b) Reference data. (c) Distribution of 9 initial training samples (cyan) and 200 new labeled samples (red) for a single experiment. (d) Color code.

with a spatial resolution of 1.3 *m* per pixel and a spectral coverage in the range 0.43-0.86 μm. Twelve spectral bands were removed before the classification due to high noise. Fig. 5(a) and (b) show the false-color composite of the University of Pavia image and the corresponding reference data, which includes 9 classes.

4) The Center of Pavia image was captured by the ROSIS sensor. After discarding 13 water noisy bands, the image contains 102 bands of size 1096×492, which contains 9 classes different objects. The false-color composite and the corresponding reference data are shown in Fig. 6.

B. QUANTITATIVE METRICS

In order to evaluate the performance of the proposed SDP-AL method, three widely used objective indicators are used, namely overall accuracy (OA), average accuracy (AA) and

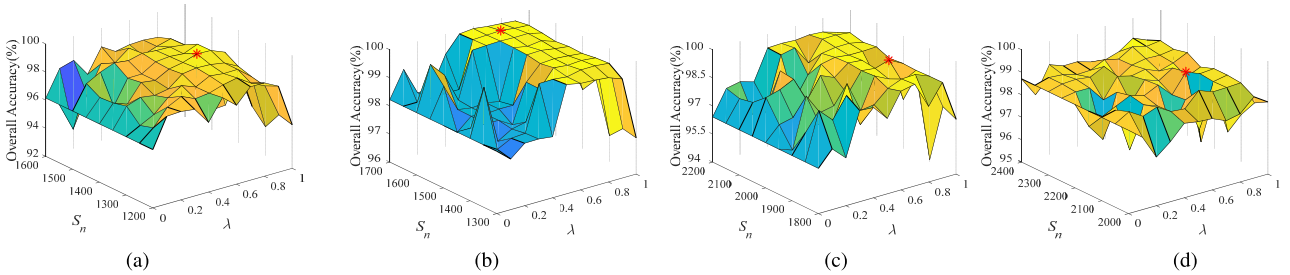


FIGURE 7. Effect of parameters S_n and λ on the proposed method on the (a) Indian Pines dataset (the initial training samples is set to 48, U=200, u=20), (b) Salinas dataset (the initial training samples is set to 16, U=200, u=20), (c) University of Pavia dataset (the initial training samples is set to 27, U=200, u=20), (d) Center of Pavia dataset, (the initial training samples is set to 9, U=200, u=20).

Kappa coefficient (Kappa) statistics. The OA index represents the percentage of pixels that are correctly classified. The AA index refers to the average percentage of correctly classified pixels for each class. The Kappa is introduced to estimate the percentage of classified pixels suffering from uncertainty factors.

C. PARAMETER TUNING

In this part, our experiments were performed on four hyperspectral datasets (Indian Pines, Salinas Valley, University of Pavia, and Center of Pavia) to discuss the influence of parameters in our method respectively. The initial training samples are randomly selected, and the number of samples per class is respectively set as 3, 1, 3, and 1 on the four datasets. Total number of new labeled samples U = 200 and the size of active query per iteration u = 20 are the same on each dataset.

In the first experiment, the number of superpixel blocks (S_n) and the size of tradeoff parameter (λ) will be determined for the proposed SD sampling criterion based active learning method. The influence of the two parameters S_n and λ are analyzed by evaluating the classification accuracy OAs of LORASL-ERW on the four datasets. As shown in Fig. 7 that OA generally show first increased and then decreased trend as the two parameters increase. For Indian Pines and Salinas datasets [see Fig. 7(a)-(b)], the tradeoff parameters λ for both of them are set to $\{0, 0.1, \ldots, 1\}$ and the number of superpixel blocks S_n are set to $\{1200, 1300, \ldots, 1600\}$ and $\{1000, 1100, \ldots, 1400\}$, respectively. Specifically, the proposed SD based LORSAL-ERW can achieve the highest overall classification accuracy when $S_n = 1600$, $\lambda = 0.7$ on the Indian Pines dataset and $S_n = 2000$, $\lambda = 0.7$ on the Salinas dataset. As can be observed that the Indian Pines dataset (image size: 145×145) requires less superpixel parting than the Salinas dataset (image size: 512×217), when the classifier obtain a optimal performance. One explanation is that a small S_n is more likely to generate an incomplete uniform area on the Salinas data set, thereby affecting the pixel structure density distribution. Therefore, the number of superpixel changes with the size of the image in order to obtain a more reliable homogeneous region for different images. Furthermore, the parameter analysis is conducted on

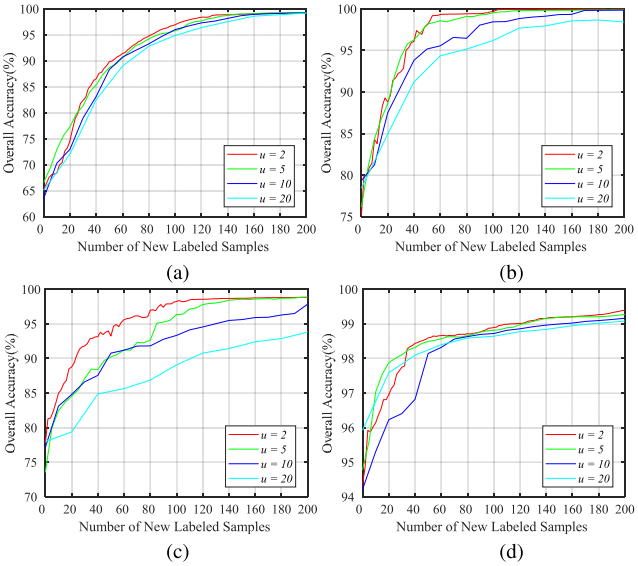


FIGURE 8. Effect of the number of new labeled samples on the proposed method on the India Pines (the initial training samples is set to 48), Salinas (the initial training samples is set to 16), University of Pavia (the initial training samples is set to 27), and Center of Pavia (the initial training samples is set to 9) datasets, respectively.

the University of Pavia and Center of Pavia urban datasets. In Fig. 7(c)-(d), the proposed method can achieve outperformance on the University of Pavia (OA=99.40%) and Center of Pavia (OA=99.15%), respectively. Therefore, the ($S_n = 2100, \lambda = 0.8$) and ($S_n = 2300, \lambda = 0.7$) is set to fault parameter corresponding to the two datasets, respectively. Through experiments on four data sets, it is found that the trade-off parameter a is greater than 0.5, that is, the contribution of structure density in the process of selecting unlabeled samples is greater than BT.

In the second experiment, the effect of diverse active query u values $\{2, 5, 10, 20\}$ to the performance of the SD sampling criterion based LORSAL-ERW-AL is analyzed on the before-mentioned four dataset. The parameter u controls the number of unlabeled pixels that are manually labeled (exploiting ground truth) in each iteration. As can be seen in Fig. 8, for diverse active query u values, the OA obtained by the LORSAL-ERW classifier. The initial training sets respectively contains 48, 16, 27, and 9 samples on the different

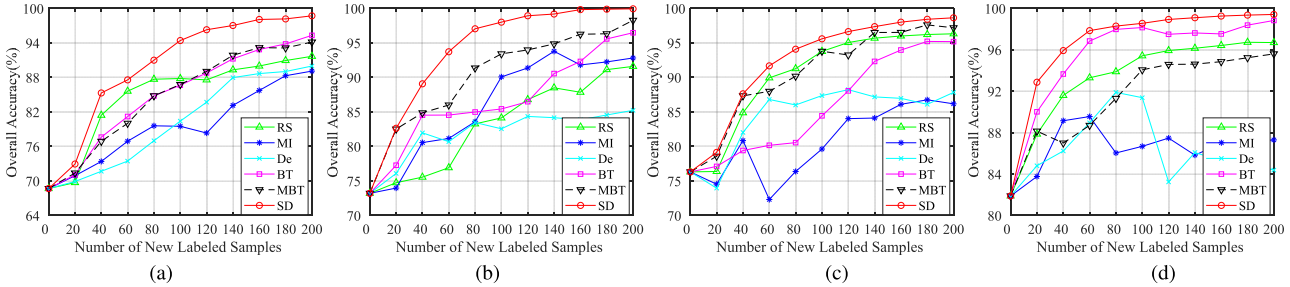


FIGURE 9. Influence of the different sampling criterion on classification accuracy. (a) Indian Pines dataset (the initial training samples is set 48), Salinas dataset (the initial training samples is set 16), University of Pavia dataset (the initial training samples is set 27), and Center of Pavia dataset (the initial training samples is set 9), respectively.

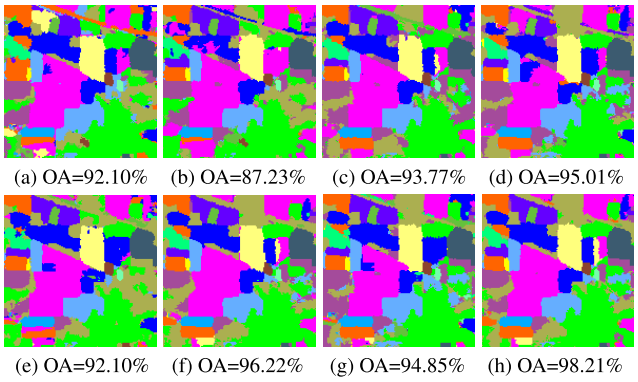


FIGURE 10. Classification results obtained by the (a) RS, (b) MI, (c) BT, (d) MBT, (e) De, (f) DeMI, (g) DeMBT, and (h) SD on the Indian Pines dataset (The training set contains 48 initial randomly selected training samples and 200 new labeled samples).

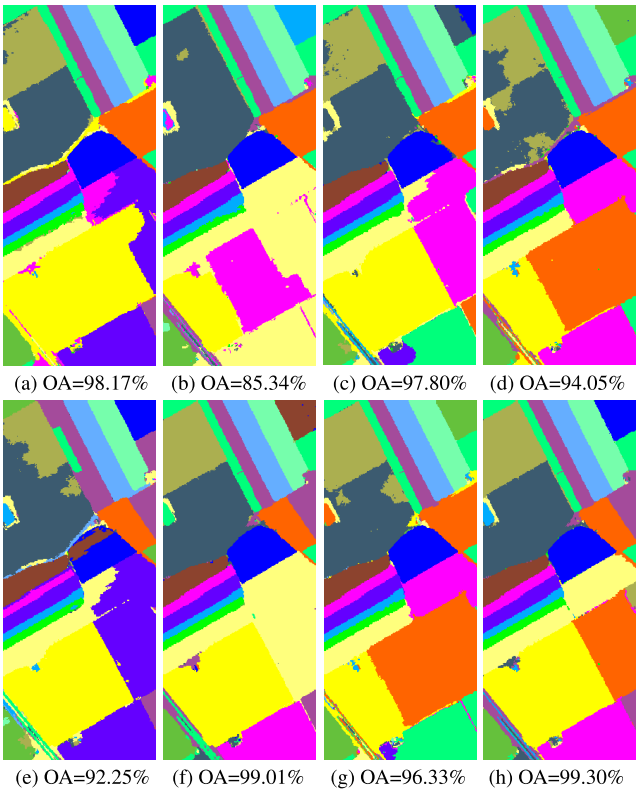


FIGURE 11. Classification results obtained by the (a) RS, (b) MI, (c) BT, (d) MBT, (e) De, (f) DeMI, (g) DeMBT, and (h) SD on the Salinas dataset (The training set contains 16 initial randomly selected training samples and 200 new labeled samples).

dataset. First, at the center of the Indian Pines dataset and Pavia dataset, when the number of newly labeled samples is between 0 and 200, different values of u can often achieve the same OA convergence speed[see Fig. 8(a), (d)]. Therefore, taking time complexity into account, the active query size $u = 20$ is set to fault parameter on the Indian Pines dataset and Center of Pavia dataset, respectively. Furthermore, in Fig. 8(b)-(c), query $u = 5$ and $u = 10$ values can obtain better performance for the classifier on the Salinas dataset and University of Pavia dataset, when the number of new labeled samples are ranged from 140 to 200. Therefore, taking into consideration balance of classification accuracy and time consumption, the size of active query $u = 10$ is set to fault parameter on the Salinas dataset and University of Pavia, respectively.

D. COMPARISON OF DIFFERENT SAMPLING CRITERION

In this subsection, the effect of six diverse sampling criterion of active learning to the performance of the classification performance is analyzed, such as, random selection (RS), MI [40], density (De), BT [41], MBT [42], and the proposed SD sampling criterion. The total number of new labeled samples U is set to 200 and the size of active queries u is set to 20 on the Indian Pines data, Salinas dataset, University of Pavia dataset, and Center of Pavia dataset respectively. The number of initial training samples of each dataset are the same IV.C. As diagrammatized in Fig. 9, the LORSAL-ERW invariably achieve optimal classification accuracy by active learning using the proposed SD sampling criterion. Moreover, through this experiment, the effectiveness of the proposed sampling criterion in active learning will be verified for hyperspectral image classification. It can be seen in Fig. 9(a)-(d) that the classification accuracy of LORSAL-ERW classifier using the proposed SD sampling criterion finally 98.68%, 99.91%, 98.61%, and 99.40% on the four real hyperspectral datasets respectively. Specially, it can be found that the accuracy of MI criterion decrease as the number of new labeld samples

TABLE 1. Classification accuracy of the RS, MI, BT, MBT,De, DeMI, DeMBT, and SD sampling criterion on the Indian Pines dataset, the training samples is set to 3 per class, the total number of new labeled samples is set to 200, the size of active learning is set to 20, and the highest accuracy are presented in bold.

Class	Different of sampling machanisms in active learing for HSI classification							
	RS	MI	BT	MBT	De	DeMI	DeMBT	SD
1	98.14(0.93)	99.07(1.14)	99.30(1.07)	**100.0(0.00)**	**100.0(0.00)**	99.07(1.14)	**100.0(0.00)**	97.91(0.70)
2	89.96(6.17)	74.55(13.0)	92.26(4.13)	88.51(4.49)	84.82(8.27)	92.77(1.56)	90.78(6.14)	**94.17(2.17)**
3	93.25(6.42)	85.79(9.99)	93.53(9.40)	98.56(1.72)	68.54(11.4)	99.60(0.18)	96.25(5.57)	**99.54(0.22)**
4	85.34(15.7)	85.04(13.3)	94.27(6.63)	98.68(0.89)	93.55(4.83)	94.53(2.95)	**99.27(0.33)**	94.44(2.80)
5	90.63(8.84)	86.88(12.0)	85.71(26.5)	85.98(26.2)	82.85(9.86)	97.27(2.41)	98.00(1.76)	**98.10(1.42)**
6	98.97(1.68)	95.82(3.96)	99.20(1.83)	**100.0(0.00)**	**100.0(0.00)**	99.86(0.00)	**100.0(0.00)**	99.86(0.00)
7	98.00(2.00)	97.60(2.65)	98.80(1.83)	**100.0(0.00)**	99.20(1.60)	98.80(1.83)	**100.0(0.00)**	97.60(1.96)
8	**100.0(0.00)**	99.62(1.07)	97.28(8.08)	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	99.96(0.08)
9	**100.0(0.00)**	99.41(1.76)	**100.0(0.00)**	**100.0(0.00)**	98.82(2.35)	**100.0(0.00)**	**100.0(0.00)**	**100.00(0.00)**
10	89.04(4.88)	75.57(18.3)	81.84(21.4)	88.37(3.34)	68.92(20.3)	88.00(2.62)	85.90(6.50)	**93.47(3.22)**
11	98.10(1.84)	90.47(14.1)	96.24(7.93)	88.74(9.80)	96.98(0.91)	99.85(0.10)	92.55(4.67)	99.80(0.14)
12	86.20(16.9)	86.53(12.7)	93.75(14.7)	95.88(4.35)	83.56(11.9)	**98.93(0.66)**	95.49(7.24)	98.86(0.55)
13	99.85(0.23)	99.90(0.20)	99.90(0.30)	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.00(0.00)**
14	99.06(1.25)	91.84(7.12)	97.63(5.90)	99.83(0.27)	99.72(0.22)	99.94(0.07)	**99.90(0.09)**	99.94(0.06)
15	71.64(18.7)	75.82(11.6)	92.58(8.07)	99.58(0.71)	86.61(6.22)	**100.0(0.00)**	99.90(0.13)	**100.00(0.00)**
16	95.44(2.25)	98.22(1.81)	96.44(4.41)	**100.0(0.00)**	98.89(0.70)	98.56(1.00)	**100.0(0.00)**	97.44(1.58)
OA	93.67(2.05)	86.68(5.93)	93.81(3.89)	93.50(2.48)	89.19(2.89)	97.43(0.29)	94.89(1.08)	**98.14(0.34)**
AA	93.35(2.32)	90.13(3.55)	94.92(3.21)	96.51(1.74)	91.40(2.13)	97.95(0.24)	97.38(0.58)	**98.19(0.37)**
Kappa	92.74(2.35)	84.74(6.74)	92.89(4.52)	92.61(2.80)	87.54(3.39)	97.06(0.34)	94.18(1.22)	**97.88(0.39)**

raises. It is because that the MI criterion only consider the multual information, but it is easy to generate bias sampling. This will lead the sample imbalance problem. The characteristic of most labeled samples (belong to one class) are very concentrated, causing a small number of remaining labeled samples (belong to other classes) to be ignored, and it will decrease the overall accuracy. By contrast, our SD sampling criterion integrates the structural density and BT score, and it can be found that the roubustness of proposed sampling critirion is significantly better than other.

In addition, the classification performance of different of sampling criterion are represented in detail on the four datasets. the different sampling criterion include RS, MI, De, BT, MBT, fusion structure density and MI sampling criterion (DeMI), MBT sampling criterion (DeMBT), and BT-assisted structure density (SD), respectively. In this experiment, the number of initial training samples is set to one, three, one, three per class on the Indian pines, Salinas, University of Pavia, and Center of Pavia, respectively. the parameter settings of U and u are the same as described above. The quantitative results are shown in Table 1-4, the best results of each quantitative index are marked in bold, and the classification diagrams corresponding to different sampling standards are shown in Fig.10-13.

1) AVIRIS INDIAN PINES DATASET

As can be seen in Table 1, the classification accuracy of the different sampling criterion based LORSAL-ERW is analyzed to demonstrated the effectiveness of the proposed SD sampling criterion for HSI classification on the Indian Pines dataset. In this experiment, the training set contains the initial randomly selected 3 samples per class and the 200 new labeled samples. As can be observed that the proposed sampling criterions, i.e., DeMI, DeMBT, and SD can basically yield better classification performance than another typical sampling criterion, such as RS, MI, BT, and MBT. For the proposed DeMBT and SD sampling criteria, the DeMBT sampling criteria has the highest level category with the highest accuracy, but the huge difference between the high-precision category and the low-precision category causes the overall accuracy to decrease, SD sampling criteria-based on each The classification accuracy rate of the category exceeds 97%, which is stable compared with the DeMBT sampling standard. From the Table 1, it can also be found that the BT-assisted structure density based SD sampling criterion can achieve the highest classification performance, i.e., OA, AA, and Kappa, compared to BT and De sampling criterion. In addition, Fig. 10 illustrates the classification results obtained by diverse sampling criterion of the Indian Pines image associated with the corresponding OA. Similarly, compare with other sampling criterions, the proposed SD sampling criterion has the highest classification accuracy. The LORSAL-ERW classifier based on SD sampling criterion can obtain classification results more similar to the reference data [see Fig. 3(b)].

2) AVIRIS SALINAS DATASET

An experiment is conducted on the Salinas image which includes 16 diverse classes. Fig. 11 illustrates the classification results acquired by the different sampling criterion, including the RS, MI, BT, MBT, De, DeMI, DeMBT, and SD sampling criterions. It can be observed that, compared with the different sampling criterion based LORSAL-ERW, the classification result achieved by the SD sampling criterion looks more similar to the reference data in Fig. 4(b). Table 2 records the average and standard variances of OA, AA, and Kappa of classification result using LORSAL-ERW classifier under different sampling criterion, when the quantity of new

TABLE 2. Classification accuracy of the RS, MI, BT, MBT,De, DeMI, DeMBT, and SD sampling criterion on the Salinas dataset, the training samples is set to 1 per class, the total number of new labeled samples is set to 200, the size of active learning is set to 20, and the highest accuracy are presented in bold.

Class	Different of sampling machanisms in active learing for HSI classification							
	RS	MI	BT	MBT	De	DeMI	DeMBT	SD
1	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	99.99(0.03)	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**
2	99.86(0.10)	94.45(11.5)	99.81(0.50)	**100.0(0.00)**	97.49(7.52)	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**
3	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	100.0(0.00)	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**
4	96.47(10.3)	94.19(17.4)	96.44(10.4)	99.95(0.03)	56.07(24.6)	**99.99(0.03)**	99.96(0.04)	99.97(0.04)
5	99.29(0.22)	89.59(20.2)	95.57(11.5)	99.84(0.02)	99.67(0.13)	99.17(0.10)	**99.82(0.07)**	99.28(0.16)
6	99.96(0.03)	94.57(16.3)	99.28(2.08)	**100.0(0.00)**	**100.0(0.00)**	99.97(0.03)	**100.0(0.00)**	99.98(0.03)
7	99.97(0.02)	**100.0(0.00)**	99.96(0.05)	**100.0(0.00)**	**100.0(0.00)**	99.95(0.04)	**100.0(0.00)**	99.94(0.03)
8	97.38(2.46)	86.56(25.9)	95.23(6.79)	85.58(18.0)	95.42(1.39)	99.88(0.09)	93.19(2.43)	**99.99(0.02)**
9	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**
10	95.62(3.32)	99.92(0.10)	99.40(0.54)	**99.97(0.09)**	99.85(0.03)	99.89(0.08)	99.58(1.17)	99.87(0.12)
11	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	77.84(38.7)	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**
12	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**
13	98.70(0.86)	91.25(15.5)	98.89(0.48)	**100.0(0.00)**	99.40(0.20)	97.96(0.70)	99.99(0.03)	98.25(0.41)
14	95.90(3.51)	98.65(1.91)	98.48(2.38)	**99.84(0.12)**	85.31(28.4)	98.08(0.21)	99.64(0.33)	99.16(0.61)
15	97.31(1.57)	28.09(34.3)	96.47(6.38)	76.84(25.7)	63.87(3.84)	99.87(0.08)	88.20(11.6)	**99.90(0.07)**
16	94.38(4.70)	90.14(11.1)	97.48(3.49)	**100.0(0.00)**	86.83(5.02)	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**
OA	98.40(0.72)	85.59(2.88)	97.98(1.53)	93.87(3.85)	91.69(1.25)	99.83(0.03)	96.96(1.25)	**99.89(0.02)**
AA	98.43(0.91)	91.71(1.82)	98.56(1.00)	97.63(1.51)	91.36(3.00)	99.67(0.05)	98.77(0.63)	**99.77(0.06)**
Kappa	98.21(0.80)	83.87(3.17)	97.76(1.70)	93.18(4.27)	90.71(1.40)	99.81(0.03)	96.61(1.40)	**99.88(0.03)**

TABLE 3. Classification accuracy of the RS, MI, BT, MBT,De, DeMI, DeMBT, and SD sampling criterion on the university of Pavia dataset, the training samples is set to 3 per class, the total number of new labeled samples is set to 200, the size of active learning is set to 20, and the highest accuracy are presented in bold.

Class	Different of sampling machanisms in active learing for HSI classification							
	RS	MI	BT	MBT	De	DeMI	DeMBT	SD
1	93.53(5.44)	91.91(5.69)	98.90(1.87)	95.20(3.89)	69.58(12.0)	92.87(0.86)	94.00(4.26)	**99.27(0.49)**
2	**99.89(0.30)**	78.93(21.4)	95.39(8.30)	94.79(5.09)	92.90(8.19)	99.17(0.25)	98.06(3.22)	99.63(0.47)
3	99.54(0.66)	40.96(9.32)	86.28(24.1)	96.39(6.44)	83.77(9.05)	100.00(0.00)	95.15(7.42)	**100.0(0.00)**
4	53.59(7.27)	82.46(5.78)	86.97(3.85)	90.93(1.14)	61.77(12.7)	91.92(0.95)	90.00(2.22)	**91.60(0.96)**
5	94.35(11.39)	90.76(15.9)	99.15(2.28)	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	99.93(0.10)
6	99.64(0.50)	78.91(20.0)	99.74(0.47)	99.88(0.36)	**100.0(0.00)**	93.26(19.5)	99.90(0.30)	**100.0(0.01)**
7	99.01(2.96)	76.41(14.0)	70.05(18.0)	97.87(6.38)	**100.0(0.00)**	**100.0(0.00)**	97.86(6.42)	**100.0(0.00)**
8	98.55(1.31)	87.07(17.4)	98.27(4.24)	**99.77(0.30)**	81.11(13.3)	99.77(0.05)	99.65(0.32)	99.70(0.10)
9	98.29(1.68)	99.23(0.67)	99.35(1.32)	**100.0(0.00)**	**100.0(0.00)**	99.82(0.07)	**100.0(0.00)**	99.84(0.09)
OA	95.19(1.05)	80.77(8.76)	95.06(4.47)	96.05(2.43)	87.03(6.10)	97.14(2.22)	97.16(1.59)	**99.09(0.20)**
AA	92.93(1.52)	80.74(4.61)	92.68(3.32)	97.20(1.21)	87.68(3.74)	97.42(2.14)	97.18(1.25)	**98.88(0.13)**
Kappa	93.55(1.43)	75.52(10.1)	93.57(5.61)	94.84(3.13)	83.17(7.64)	96.17(3.07)	96.25(2.05)	**98.80(0.27)**

labeled samples is set to 200 and the size of initial training samples is set to 1 per class. The proposed sampling criterion demonstrates the optimal performance in all cases. In particular, the classification accuracy of seven covers (Weeds_1, Weeds_2, Fallow, soil, Lettuce_4wk, Lettuce_5wk, and Vinyard_T) reach the highest accuracy 100%. Comparing with other sampling criterion, the classification accuracy of the ground coverings (except for the Lettuce_6wk ground covering) based on the SD sampling criterion are greater than 99%. Therefore, the effectiveness of the AL with the recommended SD sampling standard can be proven in the HSI classification application.

3) ROSIS UNIVERSITY OF PAVIA DATASET

To verify the applicability of AL strategy using SD sampling criteria on urban dataset, an experiment is performed on the University of Pavia image. Fig. 12 demonstrates the classification results acquired by diverse sampling criterions when the total quantity of labeled samples is set to 227. Form this figure, it can be seen that the LORSAL-ERW classifier based on the proposed SD sampling criterion can achieve higher classification accuracy and retain detail information of covers at the same time under the same sample conditions. In other words, the SD sampling criterion can select the unlabeled samples with informativeness and representativeness to retrain classifier and improve its performance. In addition, Table 3 records the mean and standard variance of the individual classification accuracies, i.e., OA, AA, and Kappa, of different sampling criterion. The LORSAL-ERW classifier based on the proposed SD sampling criterion also shows the best performance in terms of OA = 99.09%, AA = 98.88%, Kappa = 98.80%.

4) ROSIS CENTER OF PAVIA DATASET

In order to further prove the availability of the SD sampling standard in active learning and the generalization ability of LORSAL-ERW hyperspectral image classification based on the SD sampling standard, a comparison with other widely

TABLE 4. Classification accuracy of the RS, MI, BT, MBT,De, DeMI, DeMBT, and SD sampling criterion on the center of Pavia dataset, the training samples is set to 3 per class, the total number of new labeled samples is set to 200, the size of active learning is set to 20, and the highest accuracy are presented in bold.

Class	Different of sampling machanisms in active learing for HSI classification							
	RS	MI	BT	MBT	De	DeMI	DeMBT	SD
1	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	**100.0(0.00)**	95.72(0.84)	**100.0(0.00)**	99.96(0.11)	**100.0(0.00)**
2	89.23(6.37)	95.90(2.60)	95.11(1.79)	94.25(3.88)	28.86(13.6)	96.00(0.45)	95.38(2.67)	**96.70(0.43)**
3	66.43(18.4)	42.43(22.1)	83.07(8.97)	82.29(12.4)	23.52(14.2)	84.98(2.05)	73.53(22.4)	**89.41(3.05)**
4	85.75(28.9)	71.96(23.2)	96.88(5.51)	**100.0(0.00)**	98.26(4.15)	99.57(0.33)	**100.0(0.00)**	99.34(0.40)
5	96.35(1.90)	67.01(25.5)	97.72(2.23)	**97.07(1.66)**	93.53(3.10)	96.86(0.83)	93.64(11.6)	96.91(1.79)
6	99.74(0.07)	96.06(6.63)	96.27(6.99)	97.11(2.53)	19.33(15.8)	99.40(0.05)	95.93(8.50)	**99.92(0.06)**
7	93.76(5.48)	84.74(19.8)	99.94(0.16)	**100.0(0.00)**	82.97(5.30)	99.98(0.01)	**100.0(0.01)**	**100.0(0.01)**
8	94.82(1.04)	88.01(12.0)	96.52(1.83)	99.59(0.93)	92.91(11.3)	**99.97(0.00)**	99.78(0.66)	**99.97(0.00)**
9	93.34(1.14)	93.91(0.42)	89.88(11.6)	99.47(0.58)	**99.83(0.06)**	93.72(0.08)	99.64(0.58)	94.00(0.19)
OA	97.10(0.89)	93.61(1.68)	98.42(0.71)	98.72(0.27)	82.92(1.52)	98.94(0.04)	98.23(1.19)	**99.15(0.13)**
AA	91.05(4.17)	82.23(3.03)	95.04(2.28)	96.64(1.03)	70.55(2.35)	96.72(0.14)	95.32(2.29)	**97.36(0.38)**
Kappa	95.01(1.53)	89.00(2.88)	97.27(1.22)	97.80(0.47)	71.52(2.51)	98.18(0.08)	96.96(2.04)	**98.54(0.22)**

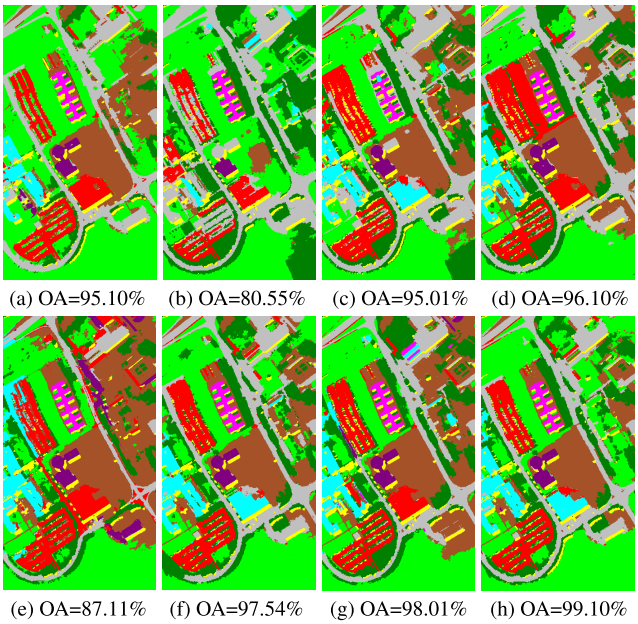


FIGURE 12. Classification results obtained by the (a) RS, (b) MI, (c) BT, (d) MBT, (e) De, (f) DeMI, (g) DeMBT, and (h) SD on the University of Pavia dataset (The training set contains 27 initial randomly selected training samples and 200 new labeled samples).

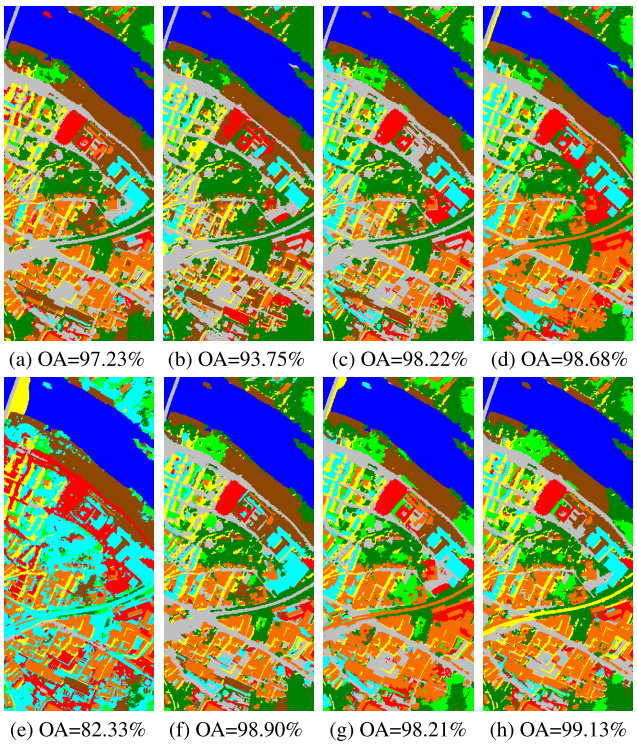


FIGURE 13. Classification results obtained by the (a) RS, (b) MI, (c) BT, (d) MBT, (e) De, (f) DeMI, (g) DeMBT, and (h) SD on the Center of Pavia dataset (The training set contains 9 initial randomly selected training samples and 200 new labeled samples).

used sampling standards was carried out on more complex and larger samples. The experimental data set is the Center of Pavia image.. In this experiment, the number of labeled samples required for different sampling standards is set to 209. Fig. 13 shows classification results obtained by different sampling criterion based LORSAL-ERW classifier, and the mean and standard variance of classification accuracy of different sampling criterion based LORSAL-ERW classifier is recorded in Table 4. As can be observed that the results yielded by the SD sampling criterion based LORSAL-ERW classifier are better than other methods in terms of the highest OA, AA, and Kappa. Similarly, from Fig. 13, it can be observed that the SD sampling criterion based LORSAL-ERW classifier can achieve optimal classification accuracy OA=99.13%, comparing with other different sampling criterions under the same sample condition.

Through comparison experiments on four data sets, i.e., Indian Pines, Salinas, University of Pavia, and University of Center, as can be observed that the proposed SD sampling criterion based active learning method can achieve the best performance classification accuracy for HSI classification. The reason is that SD sampling standards are concentrated on unlabeled samples that have a lot of information and are highly representative. Unlabeled samples with high information content and representativeness usually play a role

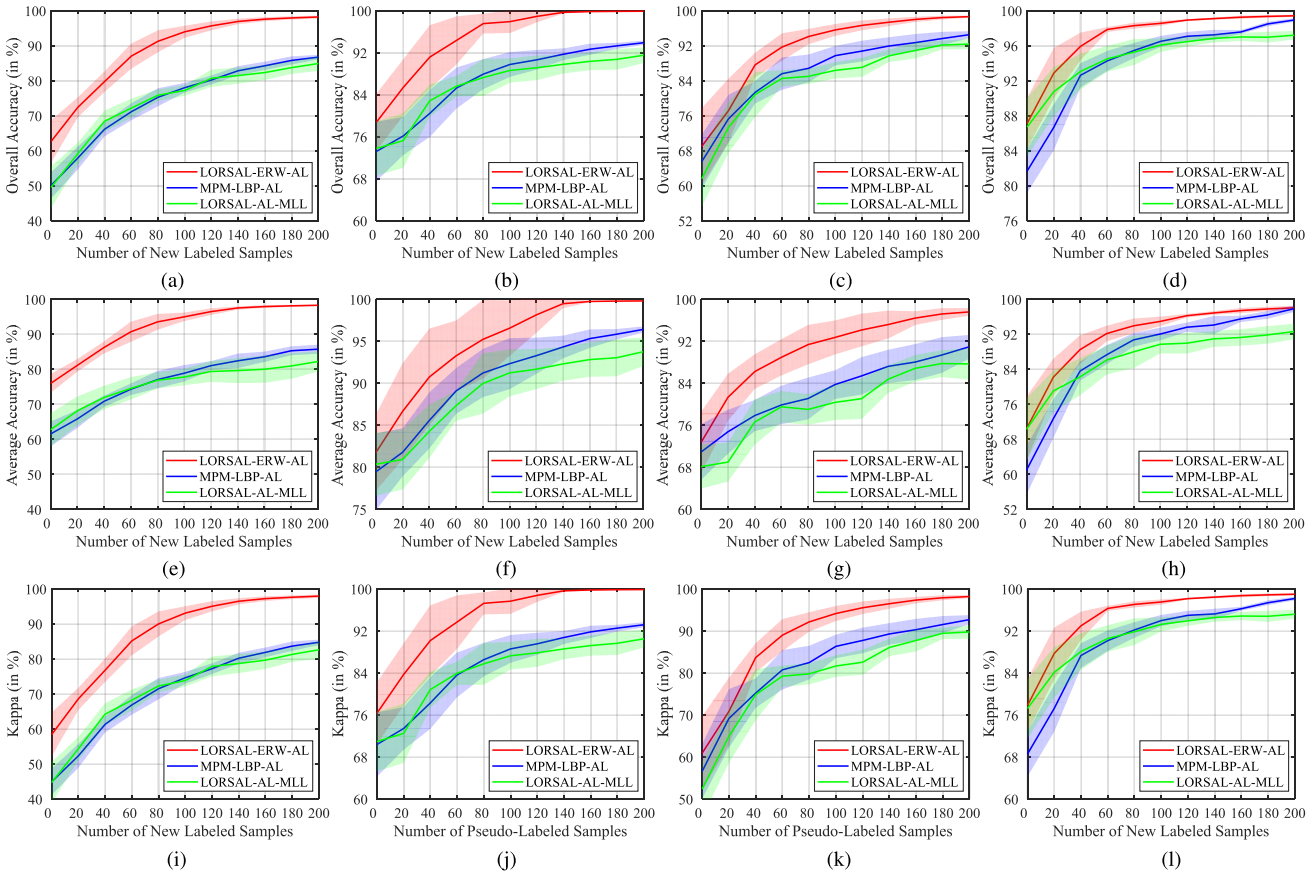


FIGURE 14. OA (The first-row), AA (The second-row), and Kappa (The three-row) achieved by the LORSAL-ERW-AL, MPM-LBP-AL, and LORSAL-AL-MLL on the Indian Pines [(a), (e), and (i)], Salinas [(b), (f), and (j)], University of Pavia [(c), (g), and (k)], and Center of Pavia datasets[(d), (h), and (l)], respectively. The widths of the line areas refer to the standard variances of accuracy obtained in ten independent Monte Carlo runs.

TABLE 5. Classification accuracy (in percent) of the classification methods using RS and SD sampling criterion on the Salinas dataset. Num in parenthesis indicates the standard variance of repeated experiments. Marking 48 initial training samples and 132 new labeled samples as 48(I) + 132(N).

Methods	RS training samples is set to 48(I) + 132(N)				SD training samples is set to 48(I) + 132(N)			
	EMP [49]	EPF [50]	SC-MK [51]	ERW [52]	EMP [49]	EPF [50]	SC-MK [51]	ERW [52]
OA	77.05(2.71)	76.78(2.87)	87.49(1.85)	88.52(2.66)	78.28(3.01)	77.80(3.50)	88.78(1.47)	93.15(1.34)
AA	80.65(2.29)	81.28(2.19)	90.53(2.04)	88.58(3.31)	82.49(2.53)	83.62(3.64)	91.93(0.83)	94.45(1.66)
Kappa	73.93(3.06)	73.24(3.44)	85.78(2.11)	86.79(3.13)	75.30(3.41)	74.52(3.99)	87.19(1.62)	92.15(1.55)

in improving the performance of the classifier. Therefore, the proposed SD sampling criterion can be used as a robust sampling criterion in many machine learning applications (i.e., semi-supervised active learning).

E. COMPARISON WITH DIFFERENT CLASSIFICATION METHODS

In this subsection, to appraise the applicability of SD sampling criterion from the perspective of classification methods, the classification performance of the widely used classification methods, i.e., the extend morphological profiles (EMP) [49], the edge preserving filtering (EPF)-based method [50], the superpixel via multiple kernels (SC-MK)-based method [51], and the extended random walk (ERW)-based method [52] is analyzed under different training sample conditions on the Indian Pines dataset. The training sample condition refers to that the training samples required for the classification method come from diverse sampling criterion (the RS and SD sampling criterion). In this experiment, the number of the initial train sample is set to 48, the number of new labeled sample with the RS sampling criterion and the SD sampling criterion is the same. To avoid any bias, these experiments are repeated 10 times and the average classification results are reported, i.e., OA, AA, and Kappa. In the Table 5, the classification performance of the EMP, EPF, SC-MK, and ERW is achieved under RS sampling criterion (left side of the table) and SD sampling criterion (right side of the table). The OA of the ERW increase form 88.52% (RS) to 93.15% (SD). Similarly, EMP, EPF, and SC-MK can increase the accuracy of training samples using SD sampling criterion

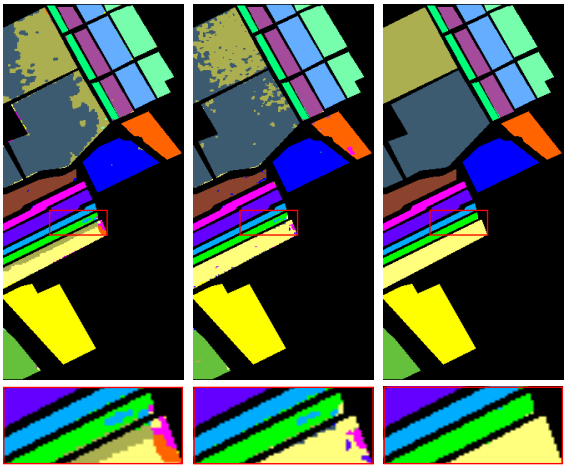


FIGURE 15. Classification results obtained by (a) the AL-LORSAL-MLL (OA = 92.10%), (b) the MPM-LBP-AL (OA = 93.55%), (c) the LORSAL-ERW-AL (OA = 99.94%) on the Salinas data set (The training set contains 16 initial randomly selected samples and 200 new labeled samples).

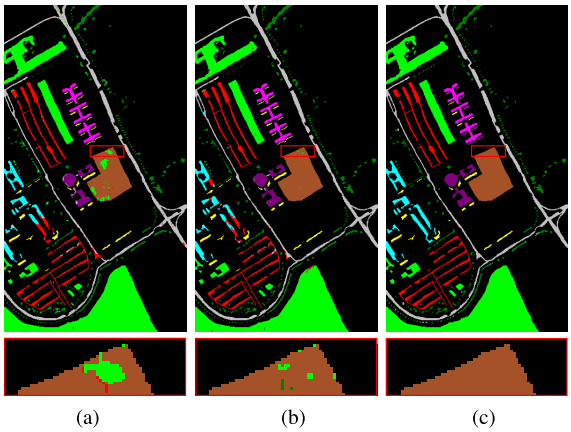


FIGURE 16. Classification results obtained by (a) the AL-LORSAL-MLL (OA = 92.11%), (b) the MPM-LBP-AL (OA = 94.35%), (c) the LORSAL-ERW-AL (OA = 98.53%) on the University of Pavia (The training set contains 27 initial randomly selected samples and 200 new labeled samples).

by approximately 2% compared to the RS sampling criterion in the terms of OA, AA, and Kappa.

F. CLASSIFICATION RESULTS OF ACTIVE LEARNING METHODS

Here, the SD sampling criterion based LORSAL-ERW-AL classification method is compared with other widely used AL-based spectral-spatial classification methods, i.e., LORSAL-AL-MLL [42], MPM-LBP-AL [53], which adopt the same MBT-based active query strategy. The quantity of initial training sample setting is set to 48, 16, 27, 9 on the Indian Pines, Salinas, University of Pavia, and Center of Pavia dataset, respectively. The size of the new labeled samples is all set to 200 on the above four datasets. Fig. 15-17 demonstrates the classification map acquired by the LORSAL-AL-MLL, MPM-LBP-AL, and LORSAL-ERW-AL methods on the Salinas dataset, University of Pavia

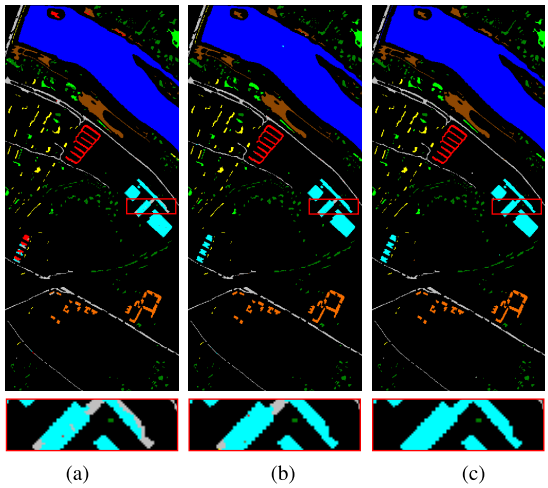


FIGURE 17. Classification results obtained by (a) the AL-LORSAL-MLL (OA = 92.11%), (b) the MPM-LBP-AL (OA = 98.01%), (c) the LORSAL-ERW-AL (OA = 99.66%) on the center of Pavia (The training set contains 9 initial randomly selected samples and 200 new labeled samples).

dataset, and Center of Pavia dataset, respectively. As shown in the figures, the classification maps acquired by the LORSAL-ERW-AL method give best performance in terms of the highest OA on the Salinas image (OA = 99.94%), University of Pavia image (OA = 98.53%), and Center of Pavia image (OA = 99.66%), respectively. The experiments have been repeated 10 times, and initial training samples are randomly selected for each iteration. As shown in Fig. 14, the average and standard variance of OA, AA, and Kappa respectively acquired by the LORSAL-AL-MLL, MPM-LBP-AL, and LORSAL-ERW-AL methods on the different dataset. It can be first found that the SD sampling criterion based LORSAL-ERW-AL clearly achieves the highest classification results for all case. For instance, with 200 new labeled samples, the OA, AA, and Kappa acquired by the LORSAL-ERW-AL method all reach values that are higher than 98% on the different dataset. In particular, on the Indian Pines dataset, for LORSAL-AL-MLL method, the OA reaches 84%, and AA and Kappa reaches around 83%. For MPM-LBP-AL method, the OA reaches around 87%, the AA reaches around 86%, and the Kappa reaches around 85%. In addition, as shown in these figures, the property of the proposed SD sampling criterion based LORSAL-ERW-AL is more stable because its standard deviation expressed by the width of the line area is much smaller.

G. COMPUTATIONAL COMPLEXITY

In this section, Table 6 reports the running time (in seconds) of the active learning strategy method based on the RS, MI, BT, MBT, De, DeMI, DeMBT, and SD sampling criteria to classify four hyperspectral images. All the programs are conducted on a computer with an Intel(R) Core (TM) i7-7800X, CPU 3.50 GHz and 32 GB of RAM, and the software platform is MATLAB R2014a. For the running time of the LORSAL-ERW-AL method based on each sampling

TABLE 6. Computing time of the RS, MI, BT, MBT,De, DeMI, DeMBT, and SD sampling criterion on the Indian Pines, Salinas, university of Pavia and center of Pavia datasets.

Images	Indian Pines	Salinas	University of Pavia	Center of Pavia
RS	5.0058	21.1793	30.6339	79.1089
MI	4.7858	22.5343	30.0793	77.3247
BT	4.6608	22.3831	29.4951	75.7062
MBT	6.3635	25.5221	31.6571	84.0573
De	0.0010	0.0043	0.0048	0.0122
DeMI	5.0235	20.7991	29.0937	75.1573
DeMBT	6.3494	23.2314	32.2456	85.7545
SD	5.0058	21.0893	29.3960	75.3950

criterion, when the number of newly labeled samples U is set to 200 and the size of the active learning query is set to 20, the time taken to complete the classification task is recorded. It can be seen from Table 6 that the MATLAB implementation of the LORSA-ERW-AL method based on the SD sampling criterion proposed in this paper has been very fast (take only 5.0058s for the Indian Pines image). The reason is that the proposed SD sampling criterion only requires BT score computing to be done several times and superpixel segmentation and density peak clustering to be done once. Therefore, it can be concluded that the proposed SD sampling criterion has advantages in improving classification accuracy and computational efficiency, and is indeed an effective candidate active learning method for HSI classification.

V. CONCLUSION

In this paper, a simple yet powerful sampling criterion of active learning has been proposed for spectral-spatial HSI classification. Experimental results demonstrate that the proposed SD-based sampling criterion active learning strategy can improve the classification accuracy significantly in few labor costs. In fact, structural density and BT score can complement each other in the process of selecting unlabeled samples. The reason is that the BT focus on unlabeled data with informativeness but prone to bias sampling and structural density aims at selecting ones with local representativeness but easy to ignore complex areas. Therefore, the SD sampling criterion integrates the advantages of both and focus on unlabeled samples that are both informativeness and local representativeness. However, one limitation of the proposed SD sampling criterion is that it does not consider the noisy label [54] of manual labeling in the classification processing. In our future work, how to solve noisy label problem of active learning and extend the SD sampling criterion based active learning strategy to large-scene the HSI classification will be researched.

REFERENCES

[1] J. Pontius, M. Martin, L. Plourde, and R. Hallett, "Ash decline assessment in emerald ash borer-infested regions: A test of tree-level, hyperspectral technologies," *Remote Sens. Environ.*, vol. 112, no. 5, pp. 2665–2676, May 2008.

[2] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14118–14129, Mar. 2018.

[3] R. D. M. Scafutto, C. R. de Souza Filho, and W. J. de Oliveira, "Hyperspectral remote sensing detection of petroleum hydrocarbons in mixtures with mineral substrates: Implications for onshore exploration and monitoring," *ISPRS J. Photogramm. Remote Sens.*, vol. 128, pp. 146–157, Jun. 2017.

[4] Q. Li, B. Zheng, B. Tu, J. Wang, and C. Zhou, "Ensemble EMD-based spectral-spatial feature extraction for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5134–5148, Aug. 2020.

[5] Q. Li, B. Zheng, B. Tu, Y. Yang, Z. Wang, W. Jiang, K. Yao, and J. Yang, "Refining urban built-up area via multi-source data fusion for the analysis of dongting lake eco-economic zone spatiotemporal expansion," *Remote Sens.*, vol. 12, no. 11, p. 1797, Jun. 2020.

[6] H. Huang, Z. Li, and Y. Pan, "Multi-feature manifold discriminant analysis for hyperspectral image classification," *Remote Sens.*, vol. 11, no. 6, p. 651, Mar. 2019.

[7] G. Shi, H. Huang, and L. Wang, "Unsupervised dimensionality reduction for hyperspectral imagery via local geometric structure feature learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1425–1429, Aug. 2020.

[8] L. Sun, C. Ma, Y. Chen, Y. Zheng, H. J. Shim, Z. Wu, and B. Jeon, "Low rank component induced spatial-spectral kernel method for hyperspectral image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 10, pp. 3829–3842, Oct. 2020.

[9] L. Sun, C. Ma, Y. Chen, H. J. Shim, Z. Wu, and B. Jeon, "Adjacent superpixel-based multiscale spatial-spectral kernel for hyperspectral classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1905–1919, Jun. 2019.

[10] J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, Sep. 2013.

[11] B.-C. Kuo, H.-H. Ho, C.-H. Li, C.-C. Hung, and J.-S. Taur, "A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 317–326, Jan. 2014.

[12] R. Reshma, V. Sowmya, and K. P. Soman, "Dimensionality reduction using band selection technique for kernel based hyperspectral image classification," *Procedia Comput. Sci.*, vol. 93, pp. 396–402, 2016.

[13] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, pp. 490–530, May 2015.

[14] M. Cui and S. Prasad, "Class-dependent sparse representation classifier for robust hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2683–2695, May 2015.

[15] Y. Yuan, J. Lin, and Q. Wang, "Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 2966–2977, Dec. 2016.

[16] S. Li, H. Wu, D. Wan, and J. Zhu, "An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine," *Knowl.-Based Syst.*, vol. 24, no. 1, pp. 40–48, Feb. 2011.

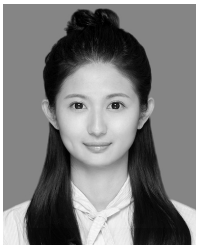
[17] C. B. Fu and A. H. Tian, "Classification of hyperspectral images of small samples based on support vector machine and back propagation neural network," *Sensors Mater.*, vol. 32, no. 1, pp. 447–454, Jan. 2020.

[18] K. Wei, Y. Fu, and H. Huang, "3-D quasi-recurrent neural network for hyperspectral image denoising," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 363–375, Jan. 2021.

[19] P. Kumar Mallick, S. H. Ryu, S. K. Satapathy, S. Mishra, G. N. Nguyen, and P. Tiwari, "Brain MRI image classification for cancer detection using deep wavelet autoencoder-based deep neural network," *IEEE Access*, vol. 7, pp. 46278–46287, Mar. 2019.

[20] H. Huang and K. Xu, "Combing triple-part features of convolutional neural networks for scene classification in remote sensing," *Remote Sens.*, vol. 11, no. 14, p. 1687, Jul. 2019.

[21] Q. Jackson and D. A. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, Nov. 2002.
[22] S. Zhu, X. Sun, and D. Jin, "Multi-view semi-supervised learning for image classification," *Neurocomputing*, vol. 208, pp. 136–142, Oct. 2016.
[23] A. J. Joshi, F. Porikli, and N. P. Papanikolopoulos, "Scalable active learning for multiclass image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2259–2273, Nov. 2012.
[24] W. Di and M. M. Crawford, "Active learning via multi-view and local proximity co-regularization for hyperspectral image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 618–628, Jun. 2011.
[25] 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.
[26] X. Zhou and S. Prasad, "Active and semisupervised learning with morphological component analysis for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1348–1352, Aug. 2017.
[27] Z. Zhang and M. M. Crawford, "Semi-supervised multi-metric active learning for classification of hyperspectral images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 1843–1846.
[28] S. Zhang, J. Wu, J. Li, J. Gu, X. Tang, and X. Xu, "Semi-supervised community detection via constraint matrix construction and active node selection," *IEEE Access*, vol. 8, pp. 39078–39090, Dec. 2020.
[29] X. Hadoux, S. Jay, G. Rabatel, and N. Gorretta, "A spectral–spatial approach for hyperspectral image classification using spatial regularization on supervised score image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2361–2369, Jun. 2015.
[30] K. Bernard, Y. Tarabalka, J. Angulo, J. Chanussot, and J. A. Benediktsson, "Spectral–spatial classification of hyperspectral data based on a stochastic minimum spanning forest approach," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2008–2021, Apr. 2012.
[31] X. Kang, S. Li, and J. A. Benediktsson, "Feature extraction of hyperspectral images with image fusion and recursive filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3742–3752, Jun. 2014.
[32] C. Mu, J. Liu, and Y. Liu, "Hyperspectral image classification based on active learning and spectral-spatial feature fusion using spatial coordinates," *IEEE Access*, vol. 8, pp. 6768–6781, Jan. 2020.
[33] S.-J. Huang, R. Jin, and Z.-H. Zhou, "Active learning by querying informative and representative examples," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 1936–1949, Oct. 2014.
[34] 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.
[35] H. Tang and Z. Hu, "Research on medical image classification based on machine learning," *IEEE Access*, vol. 8, pp. 93145–93154, May 2020.
[36] 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.
[37] 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.
[38] S. Sun, P. Zhong, H. Xiao, and R. Wang, "Active learning with Gaussian process classifier for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1746–1760, Apr. 2015.
[39] B. Settles, "Active learning literature survey," *Comput. Sci. Tech. Rep.*, vol. 39, no. 2, pp. 127–131, Jan. 2009.
[40] B. Krishnapuram, D. Williams, Y. Xue, A. Hartemink, L. Carin, and M. Figueiredo, "On semi-supervised classification," in *Proc. 18th Annu. Conf. Neural Inf. Process. Syst.*, vol. 17, pp. 721–728, Jan. 2004.
[41] T. Luo, K. Kramer, D. B. Goldgof, A. R. S. Samson, T. Hopkins, and D. Cohn, "Active learning to recognize multiple types of plankton," *J. Mach. Learn Res.*, vol. 6, no. 4, pp. 589–613, Sep. 2004.
[42] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947–3960, Oct. 2011.
[43] M. Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy-rate clustering: Cluster analysis via maximizing a submodular function subject to a matroid constraint," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 99–112, Jan. 2014.
[44] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2281, Nov. 2012.
[45] G. Nemhauser, L. Wolsey, and M. Fisher, "An analysis of approximations for maximizing submodular set functions," *Math. Program.*, vol. 14, no. 1, pp. 265–294, Dec. 1978.
[46] G. N. Yesilyurt, A. Erturk, and S. Erturk, "Metric and transform performance analysis for hyperspectral superpixel segmentation," in *Proc. 25th Signal Process. Commun. Appl. Conf. (SIU)*, May 2017, pp. 15–18.
[47] B. Tu, X. Zhang, X. Kang, G. Zhang, and S. Li, "Density peak-based noisy label detection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1573–1584, Mar. 2019.
[48] A. Rodriguez and A. Laio, "Machine learning. Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.
[49] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
[50] X. Kang, S. Li, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.
[51] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.
[52] X. Kang, S. Li, L. Fang, M. Li, and J. A. Benediktsson, "Extended random walker-based classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 144–153, Jan. 2015.
[53] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Feb. 2013.
[54] L. Sun, B. Jeon, B. N. Soomro, Y. Zheng, Z. Wu, and L. Xiao, "Fast superpixel based subspace low rank learning method for hyperspectral denoising," *IEEE Access*, vol. 6, pp. 12031–12043, Feb. 2018.



QIANMING LI (Graduate Student Member, IEEE) received the M.S. degree in urban planning from Central South University, Changsha, China, in 2016, where she is currently pursuing the Ph.D. degree in civil architecture and planning design.

Her research interests include regional spatial structure and urban analysis in remote sensing.

BOHONG ZHENG received the M.S. and Ph.D. degrees in urban and regional planning from East China Normal University, Shanghai, China, in 1999 and 2002, respectively.

He has Postdoctoral experience in urban planning from Southeast University, Nanjing, China, in 2004. He was a Visiting Researcher with Nagoya University, Nagoya, Japan. Since 2011, he has been a Professor with the School of Architecture and Art, Central South University, Changsha, China. His research interests include urban functional structure and spatial form, urban networks and network city theory, eco-city, and low-carbon space planning.

YUSHENG YANG (Member, IEEE) is currently pursuing the B.S. degree in architectural design with the Central South University of Forestry and Technology.

His research interests include architectural design, architectural theory, and architectural analysis in remote sensing.