Neurocomputing 275 (2018) 2512–2524

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Collaborative learning for hyperspectral image classification

Chao Pan[a], Jie Li[a], Ying Wang[a], Xinbo Gao[a,b,*]

[a] School of Electronic Engineering, Xidian University, Xi'an 710071, China
[b] State Key Laboratory of Integrated Services Networks, School of Electronic Engineering, Xidian University, Xi'an 710071, China

ARTICLE INFO

Article history:
Received 31 March 2017
Revised 1 November 2017
Accepted 16 November 2017
Available online 23 November 2017

Communicated by Shiliang Sun

Keywords:
Hyperspectral image classification
Collaborative learning
Active learning
Clustering
Semi-supervised learning

ABSTRACT

Recently, collaborative learning (CL) is introduced to combine active learning (AL) with semi-supervised learning (SSL), and solve the problem of limited training samples. In this paper, we proposed a novel CL framework for hyperspectral image classification, in which AL and SSL are collaboratively integrated using clustering (CLUC). CLUC attempts to obtain more diversity and higher confidence of additional training samples in both AL and SSL. Note that clustering methods, which are used separately to enhance AL or SSL, are utilized to integrate these two learning process in CLUC. First, all unlabeled samples are assigned into clusters. Based on the clustering result, clustering-based query (CBQ) for both AL and SSL, and CBQ-based pseudo-labeling (CBQPL) for SSL are designed for CLUC. Second, the most and secondary uncertain samples in each cluster are selected by CBQ for AL and SSL, respectively, to ensure their informativeness. Third, CBQPL assigns the selected secondary uncertain samples with the same label as the most uncertain one, which is manually-labeled in AL within the same cluster. CBQPL makes the confidence of pseudo-labeling rely on the clustering results. We evaluate the performance of CLUC on three real hyperspectral images. The performance of the proposed method is tested under different numbers of labeled samples and compared with several approaches. We can observe from the experimental results that CLUC have superiority in classification maps and objective metrics with limited training samples.

1. Introduction

Hyperspectral image (HSI) often consists of rich spectral-spatial information. HSI classification is one of the main challenges in the interpretation of remote sensing data, which has been widely used in various applications, such as precision agriculture [1], target detection [2], urban planning [3], and land-cover identification [4,5]. Given a set of observations (i.e., pixel vectors in a hyperspectral image), the goal of classification is to assign a unique label to each pixel vector so that it is well-defined by a given class [6]. Various supervised machine learning algorithms have been widely used in HSI classification, e.g., multinomial logistic regression (MLR) [7], support vector machines (SVMs) [8], artificial neural networks(ANN) [9], multiple feature learning [10,11], sparse representation [12,13], and so on. However, HSI classification often suffers from the problem of limited training samples, and the manually-labeling is time-consuming and expensive. Recently, some advanced machine learning techniques, i.e., active learning (AL) [14], semi-supervised learning (SSL) [15], and spectral-spatial classification [16,17], have been widely developed to solve this problem.

AL involves an iterative learning process of human-machine interaction. It starts with a small-size set of training set, and then iteratively selects the newly-added training samples from the unlabeled samples, which are subsequently labeled by human. The research of AL focuses on the query strategy to rank the most informative training samples for the current classifier [18,19]. These strategies are grouped into three main types: committee-based [14,20,21], large margin-based [22], and posterior probability-based [23] query heuristics. In contrast to AL, SSL leverages the unlabeled data together with the initial training set for the training of the classifier. It attempts to exploit the structural information in the feature space to facilitate the learning process without additional human efforts. The SSL methods are grouped into the following categories: self-learning [24,25], co-training [26], generative probabilistic models [27], semi-supervised SVMs [28], and graph-based SSL [29].

AL and SSL approaches work based on different mechanisms, but both aim to provide promising classification accuracy and alleviate the cost of human efforts. Thus, it is reasonable to combine the AL an SSL learning process. Recently, some methods combining AL and SSL have been developed and demonstrated promising performance [24,30–33]. Very recently, AL and SSL have been

* Corresponding author at: State Key Laboratory of Integrated Services Networks, School of Electronic Engineering, Xidian University, Xi'an 710071, China.
E-mail addresses: lwpanchao@126.com (C. Pan), leejie@xidian.edu.cn (J. Li), yingwang@xidian.edu.cn (Y. Wang), xbgao@mail.xidian.edu.cn (X. Gao).

https://doi.org/10.1016/j.neucom.2017.11.035
0925-2312/

integrated as a collaborative learning process, in which the selected unlabeled samples are collaboratively labeled by the human experts and machines/classifiers. In [34], collaborative active and semi-supervised learning (CASSL) acquires more confidently labeled samples by integrating AL and SSL. In [35], a random-walker-based verification strategy separates the unlabeled samples into the low-confident and high-confident data sets, and manually-labeled and pseudo-labeled samples are used together to train the final classifier.

These collaborative learning methods present promising classification performances with limited manually-labeled samples. However, there still exist several aspects that may be improved. First, the precision of pseudo-labeling relies on the current classification results and the strategy to select the most confident samples. Second, for the SSL learning process, actually the most confident samples are selected and pseudo-labeled, which provides redundant information for the classifiers. Third, the manually-marked labels are just assigned to the selected unlabeled samples in AL, and we wish to make full use of these labels marked by human.

We note that, clustering has been extensively incorporated into AL or SSL separately to improve the performance of HSI classification [29,31,36–42]. Considering only image statistics, clustering algorithms automatically partition the image into homogeneous and coherent groups. In AL, the classifier is retrained with new training set, and it is promised to select two or more samples at each AL iteration. These AL query strategies, such as margin sampling (MS) [43], Multiclass level uncertainty (MCLU) [36], max entropy (ME) [15], breaking ties (BT) [44], and Kullback–Leiber divergence maximization (KL-Max) [45,46], consider only the uncertainty of unlabeled samples. And, the queried samples may be redundant to each other. Therefore, many studies have introduced clustering methods into AL to reduce the redundancy by seeking the most dissimilarity among those ranked samples [36–39]. Different from AL, the unlabeled data in SSL are leveraged with the initial labeled data to train the classifier. SSL attempts to exploit the structural information in feature space to facilitate the learning process without additional manual-labeling. In [31,40], clustering approaches are utilized to improve/construct the SSL learning process. In [29,41], cluster/manifold regularization is included in graph-based SSL methods according to the assumption of labeling smoothness.

In this paper, we propose a novel collaborative learning framework using clustering (CLUC), in which the clustering procedure plays the similar role as a 'bridge' in integrating the AL and SSL learning process. First, all the unlabeled pixels of an HSI image are grouped into a set of clusters using superpixel algorithm. Recently, superpixel techniques have been developed a lot in computer vision [47–49]. And superpixel methods are combined with segmentation-based classifiers for HSI classification [50–55]. Superpixel segmentation is virtually a clustering method considering not only feature similarity but also spatial adjacency. After superpixel segmentation, an image is segmented into non-overlapping regions/clusters. Then, clustering-based query (CBQ) heuristic is designed by combining the breaking BT query strategy with the superpixel clustering result. Second, the most and the secondary uncertain samples in each superpixel are actively selected by the CBQ strategy for the AL and SSL learning process, respectively, to ensure their informativeness for the classifier. Third, CBQ-based pseudo-label (CBQPL) assigns the secondary uncertain unlabeled samples (in SSL) with the same label as the manually-labeled sample (in AL) within the same superpixel. And, both the manually-labeled and CBQPL-labeled samples are used to train the classifier at each iteration of CLUC.

An obvious difference of CLUC from those CL frameworks is that it aims to find the most confident samples that have the same labels with the manually-labeled samples in the CL learning process, instead of pseudo-labeling the unlabeled samples using the current classification result. This will bring several advantages in the following aspects. First, the CBQ strategy selects the most uncertain samples for both AL and SSL, which can provide more information than those aforementioned CL learning framework. Second, the designed CBQPL strategy makes the accuracy of pseudo-labeling rely on the effectiveness of clustering method, instead of the performance of classifiers. In this paper, using effective superpixel algorithm ensures a promising CBQPL labeling process. Additionally, based on the superpixel segmentation, the CBQ heuristic considers the diversity of active sampling process in both feature and spatial domains. Therefore, CBQ promotes more uncertainty of queried samples in both AL and SSL learning processes. Moreover, the CBQ strategy within each cluster is still biased sampling, which partly promotes the accuracy of CBQPL. Mutinomial logistic regression (MLR) is used to model the posterior probability distributions at each CLUC iteration. The variable splitting and augmented Lagrangian (LORSAL) [56] algorithm, followed by a multi-level logistic (MLL) [15] regularization, i.e., the LMLL method, is adopted as the basic classifier in this paper. The performance of CLUC is tested on three real hyperspectral images. Experimental results demonstrate that the proposed CLUC framework can achieve competitive classification performance with limited manually-labeled training samples.

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 presents a detailed description of the proposed CLUC method. The experimental results of CLUC are given in Section 4. Section 5 concludes this paper and suggests some future work.

2. Related work

This section presents a brief review of the existing research about the application of clustering on the AL and SSL learning, and the combination of AL and SSL for HSI classification. For more details about the analysis on AL and SSL (and the comparison between these two algorithms), readers may refer to [57,58].

2.1. Clustering in AL and SSL for HSI classification

The clustering techniques consider the distribution of the sample in feature (and spatial) space and group the data into homogeneous and coherent clusters. Extensive studies have incorporated clustering algorithms to enhance the AL or SSL learning for HSI classification.

In AL, clustering algorithms are used to reduce the redundancy among samples in the active query process. Many query strategies have been studied to measure the uncertainty of unlabeled samples. The classifier is retrained at each iteration with the new training set, and it is promised to select two or more samples per AL iteration. These query strategies aforementioned, however, consider only the uncertainty of pixels, and the queried pixels may be redundant to each other. Therefore, many efforts consider the samples diversity by seeking the most dissimilarity among those highly ranked samples. Clustering-based diversity (CBD) [36] (together with other methods, e.g. angle-based diversity (ABD) [37]) attempts to reduce the sample redundancy in AL. In [37], the standard k-means clustering [38] was used to promote the diversity in AL query for binary SVM. In [59], CBD was incorporated into MCLU to reduce the redundancy in AL query heuristic. In [36], kernel k-means was combined with MCLU to group the selected samples into clusters and obtain divers batches. The BT criterion, a posterior probability-based query heuristic, is improved to modified BT (MBT) [39] to promote more diversity in sampling, which actually achieves a 'clustering' in advance by grouping unlabeled pixels according to the classification result.

In SSL, clustering approaches are also utilized to improve/construct the learning process. Different from AL, the unlabeled data in SSL are leveraged together with the initial labeled data to train the classifier. SSL attempts to exploit the structural information in the feature space to facilitate the learning process without additional manual efforts. Hierarchical clustering methods, which iteratively cluster objects according to their similarity, have been combined with SSL for HSI classification [31,40]. Cluster/manifold regularization is included in graph-based SSL methods according to the assumption of labeling smoothness [29,41]. Especially in [42], a SSL classifier based on hierarchical clustering is combined with AL strategies to promote the uncertainty and confidence of pseudo-labeled samples.

2.2. Combination of AL and SSL for HSI classification

Both AL and SSL aim to alleviate the human effort and improve the classification performance with limited training samples. The combination of these two approaches is reasonable and presents promising results for HSI classification.

Recently, some methods combining AL with SSL have demonstrated promising performance. In [30,31], a family of AL strategies are introduced into a SSL classifier based on hierarchical clustering to select the most informative pixels. First, *k*-means is utilized to iteratively split the biggest cluster and built a top-down hierarchical segmentation. Then, the best cut of dendrogram is obtained by minimizing the global classification error. Finally, node and subnode AL strategies are designed to query the most uncertain sample from the children node with higher classification uncertainty. In [24], posterior probability-based AL query is introduced into semi-supervised self-learning to select the most informative samples from the candidate pool. In [32], an improved progressive semi-supervised SVM (PS^3VM) is presented by exploiting the concepts usually considered in AL. These aforementioned approaches can be regarded as incorporating AL strategies to improve SSL. In addition, SSL can be used to facilitate the AL heuristic. In [33], the most informative samples are selected by combining the multi-view disagreement-based AL with manifold-based SSL.

Very recently, AL and SSL have been integrated as a collaborative process. At each CL iteration the selected unlabeled samples are queried by AL heuristic and collaboratively labeled by human experts and classifiers/machines [60,61]. In [34], collaborative active and semi-supervised learning (CASSL) acquires more confidently labeled samples by integrating AL and SSL. In CASSL, two different classifiers are employed, and in detail, one training with only manually-labeled data in AL and the other trained with additional pseudo-labeled data in SSL. AL strategy is designed to query these unlabeled data that cannot be confidently pseudo-labeled in SSL. In [35], a random walker (RW)-based [62] verification strategy separates the unlabeled samples into low-confident and high-confident data sets. Then, manually-labeled and pseudo-labeled data are obtained in AL and SSL, respectively. Finally, these newly labeled data are use together with the initial data to train the final extended-RW (ERW) classifier [63].

In these CL frameworks, two different classifiers (or classifiers learned using different training sets) constitute the SSL learning process, and the following AL strategy and pseudo-label verification are based on their classification results. Besides, the most uncertain samples are selected for AL as they are usually more informative and labeled by human experts. However, for SSL, the unlabeled samples with the highest confidence are selected and pseudo-labeled according to the current learning algorithm. Thus, these pseudo-labeled samples provide redundant information for the current classifier.

3. The proposed CLUC method

In this work, the proposed CLUC method integrates the AL and SSL learning process using clustering, which attempts to obtain more diversity and higher confidence of the additional training samples in both AL and SSL. Superpixel is utilized to group the unlabeled data into clusters. Based on the clustering results and BT heuristic criterion, CBQ heuristic is used to query the most informative samples for both AL and SSL, and CBQPL strategy is used for SSL to find the most confident samples that have the same labels as the corresponding manually-labeled samples in AL. The manually-labeled and pseudo-labeled samples, including the initial training set, are used for the training of classifier at each CL iteration. Fig. 1 illustrates the flowchart of the proposed method.

3.1. Clustering via superpixel

In the proposed CLUC algorithm, clustering plays a role as a 'bridge' in integrating the AL and SSL learning process. And the CBQ active selection strategy and the CBQPL pseudo-labeling are both based on the clustering results. CBQ attempts to improve the informativeness of the active selected unlabeled samples for AL and SSL. Moreover, CBQPL attempts to obtain higher pseudo-labeling confidence, when assigning the secondary uncertain unlabeled samples (in SSL) with the same label as the manually-labeled sample (in AL) within the same cluster. The above two aspects require the clustering results to provide more diversity among different clusters and more homogeneity within each cluster.

Based on the above analysis, superpixel is utilized in this paper to provide a promising clustering on the samples, which is virtually an over-segmentation on the HSI image. We note that, recently, superpixel techniques have been developed a lot in segmentation-based methods for HSI classification. First, superpixel is different from other clustering techniques applied in remote sensing (e.g., *k*-means clustering and kernel *k*-means clustering [36,38]). Superpixel considers not only the spectral similarity, but also the spatial continuity and divides an image into non-overlapping and homogeneous regions. Second, similar to other hyperspectral segmentation techniques (e.g., watershed-based segmentation [64], stochastic minimum spanning forest [65]), superpixel jointly integrates the spectral-spatial cohesiveness and dissimilarity and provides an efficient segmentation for HSI classification.

Hyperspectral images often consist of hundreds of bands, therefore, principal component analysis (PCA) [66] is always utilized to extract the principal components (PCs) of HSI for an efficient superpixel segmentation [50–55]. In this paper, the first three PCs are extracted and used for superpixel segmentation. Then, entropy rate superpixel (ERS) [48] is employed to segment (cluster) the HSI image containing all samples into a set of non-overlapping regions (clusters). We denote the clustering results as follow.

$$\boldsymbol{C} = \{SP_1, \ldots, SP_N\}, N > 0 \tag{1}$$

$$SP_n = \left\{\mathbf{x}_1^n, \ldots, \mathbf{x}_{M_n}^n\right\}, M_n > 0, n = 1, \ldots, N \tag{2}$$

where N indicates the number of superpixels. See that, ERS realizes an over-segmentation on the image, and the labeled pixels are contained in the initial set of superpixels. In Eq. (2), the labeled pixels have been removed from each superpixels, and therefore in Eq. (1), $\boldsymbol{C}$ is a set of clusters including only the unlabeled samples. Moreover, SP_n is a set of unlabeled samples that contained within itself and M_n is the corresponding number of unlabeled samples. Fig. 2 illustrates this procedure of clustering unlabeled samples into a set of clusters via superpixel. In this paper, ERS is utilized to provide a promising clustering result, which is based on graph partition and favors compact and homogenous non-overlapping clusters [48]. It

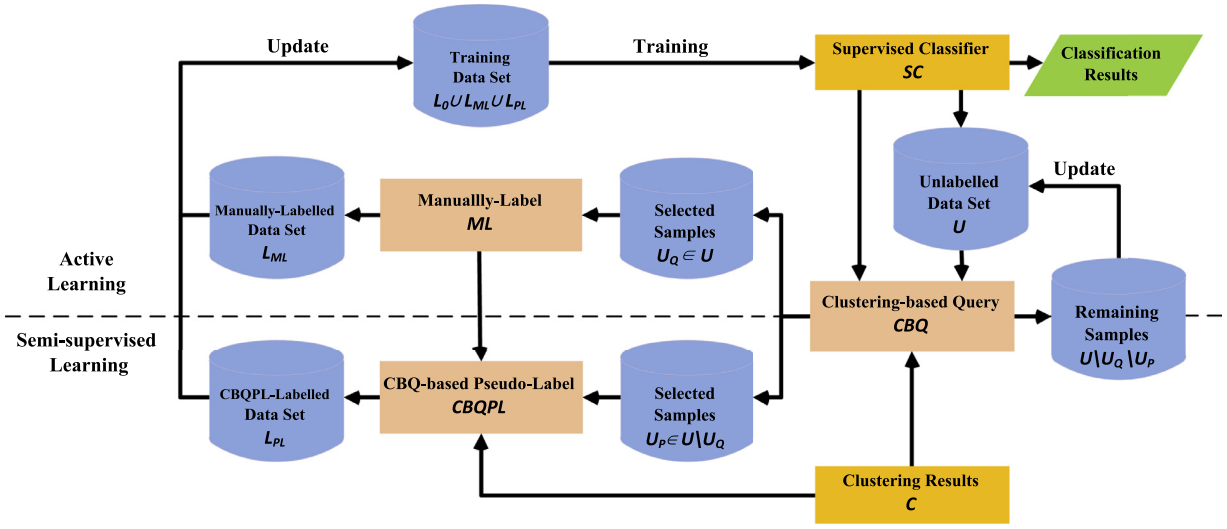


Fig. 1. The flow diagram of the proposed CLUC framework.

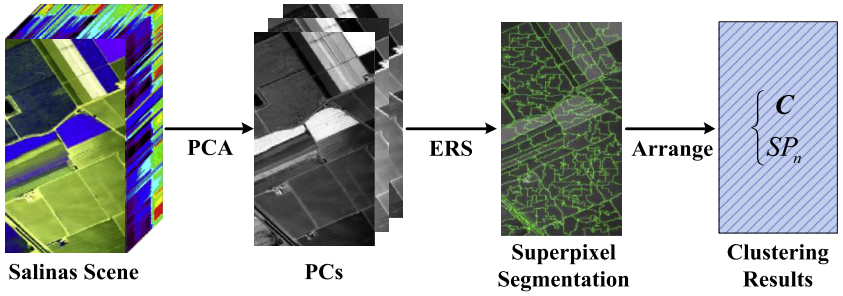


Fig. 2. The procedure for the clustering via superpixel on HSI.

is assumed that these unlabeled samples within each superpixel attach high probability to be the same class, which ensures the performance of the CLUC method for HSI classification.

3.2. CBQ heuristic via SPBT

In this subsection, a modified BT heuristic for HSI classification, i.e., superpixel-based BT (SPBT) is proposed and employed as the query heuristic in CLUC. The BT query strategy [44] was proposed to promote more diversity in the sampling process. The decision criterion of BT is formulated as:

$$\hat{\mathbf{x}}_i^{BT} = \arg\min_{\mathbf{x}_i, i\in U} \left\{ \max_{k\in L} p(y_i = k|\mathbf{x}_i) - \max_{k\in L\backslash\{k^+\}} p(y_i = k|\mathbf{x}_i) \right\} \tag{3}$$

where $k^+ = \arg\max_{k\in L} p(y_i = k|\mathbf{x}_i)$ is the class associating with maximal probability for the sample $\mathbf{x}_i$. The BT heuristic focuses on the boundary regions between two classes to obtain more diversity in sampling process. However, it may get trapped into boundaries containing many samples, while neglecting boundaries with few samples. Then we propose a novel modified strategy SPBT, which promote even more diversity in the sampling process.

For a given superpixel segmentation result $\boldsymbol{C}$ defined in Eqs. (1) and (2), we rank the unlabeled samples in superpixel SP_n according to the BT criterion in Eq. (3), and note the sorted results of all samples within superpixel SP_n as:

$$SP_n^{BT} = \left\{\mathbf{x}_1^n, \ldots, \mathbf{x}_{M_n}^n\right\}^{BT}, n = 1, \ldots, N \tag{4}$$

where $\mathbf{x}_i^n$ indicates the ith most uncertain sample in superpixel SP_n^{BT} based on the BT criterion (e.g., $\mathbf{x}_1^n$ and $\mathbf{x}_2^n$ denote the most and the second uncertain unlabeled samples, respectively). For convenience, we call these samples except the most uncertain one $\mathbf{x}_1^n$ as the secondary uncertain samples, which is represented as:

$$SP_n^{BT}\backslash\{\mathbf{x}_1^n\} = \left\{\mathbf{x}_2^n, \ldots, \mathbf{x}_{M_n}^n\right\}^{BT}, n = 1, \ldots, N \tag{5}$$

where $SP_n^{BT}\backslash\{\mathbf{x}_1^n\}$ indicates the set of the secondary uncertain samples except $\mathbf{x}_1^n$, in which $\mathbf{x}_2^n$ is the most uncertain sample. Then, we use the most uncertain samples $\mathbf{x}_1^n$ in SP_n^{BT} to measure the uncertainty of superpixels, and rank all the superpixels based on the BT criterion in Eq. (3) again. The sorted results of all superpixels are denoted as:

$$\boldsymbol{C}^{BT} = \left\{SP_1^{BT}, \ldots, SP_N^{BT}\right\}^{BT}, N > 0 \tag{6}$$

where SP_1^{BT} is the most uncertain superpixel in $\boldsymbol{C}^{BT}$ based on the SPBT criterion, and the unlabeled samples in SP_n^{BT} have also been arranged according to the formulation in Eq. (4).

The pseudo-code of the SPBT strategy for one iteration in CL is illustrated in Algorithm 1. Given the number of AL and SSL query size q ($q < N$) and p (p is divisible by q), the unlabeled sample $\{\mathbf{x}_1^k\}_{k=1}^q$ and $\{\mathbf{x}_2^k, \ldots, \mathbf{x}_{p/q+1}^k\}_{k=1}^q$ are selected from $\boldsymbol{C}^{BT}$ for AL and SSL, respectively. The adopted superpixel approach, ERS, jointly integrates the spectral features and spatial continuity to segment the HSI image into non-overlapping and homogeneous regions. And therefore, the designed SPBT strategy jointly considers the uncertainty of queried samples in spectral-spatial features. By SPBT, the queried samples are selected from different superpixels, which promotes more diversity in spatial domain.

3.3. CBQPL pseudo-labeling

In this subsection, we propose a pseudo-labeling strategy called CBQPL. According to the CBQ query strategy described in

Algorithm 1
CBQ-CL (for one iteration in CLUC via SPBT).

Inputs:
Initial unlabeled data set: $U_Q = \emptyset$;
Initial pseudo-labeled data set: $U_P = \emptyset$;
Current unlabeled data set: $U = \{\mathbf{x}_j\}_{j=1}^{u}$;
Current classification results: $p(y_i = k|\mathbf{x}_i), \mathbf{x}_i \in U$;
AL and SSL query size: q and p;
Number of the clusters for clustering: N;
1. Cluster unlabeled samples U into a set of clusters via ERS based on Eqs. (1) and (2):
$\boldsymbol{C} = \{SP_1, \ldots, SP_N\}, SP_n = \{\mathbf{x}_1^n, \ldots, \mathbf{x}_{M_n}^n\}$;
2. Rank the clustering result based on the strategies in Eqs. (3)–(6):
$\boldsymbol{C}^{BT} = \{SP_1^{BT}, \ldots, SP_N^{BT}\}^{BT}, SP_n^{BT} = \{\mathbf{x}_1^n, \ldots, \mathbf{x}_{M_n}^n\}^{BT}$;
3. Select q data $U_Q = \{\mathbf{x}_1^k\}_{k=1}^{q}$ from $\boldsymbol{C}^{BT}$ for AL;
4. Select p data $U_P = \{\mathbf{x}_2^k, \ldots, \mathbf{x}_{p/q+1}^k\}_{k=1}^{q}$ from $\boldsymbol{C}^{BT}$ for SSL;
5. Update the unlabeled data set: $U = U \backslash U_Q \backslash U_P$;

Algorithm 2
CBQPL (for one iteration in CLUC using CBQ).

Inputs:
The q data selected in AL using CBQ: $U_Q = \{\mathbf{x}_1^k\}_{k=1}^{q}$;
The p data selected in SSL based on CBQ: $U_P = \{\mathbf{x}_2^k, \ldots, \mathbf{x}_{p/q+1}^k\}_{k=1}^{q}$;
1. Manually label U_Q by human: $U_Q = \{(\mathbf{x}_1^k, \mathbf{y}_1^k)\}_{k=1}^{q}$;
2. Pseudo-label U_P: $U_P = \{(\mathbf{x}_2^k, \mathbf{y}_1^k), \ldots, (\mathbf{x}_{p/q+1}^k, \mathbf{y}_1^k)\}_{k=1}^{q}$;

Section 3.2, we have obtained the unlabeled samples $U_Q = \{\mathbf{x}_1^k\}_{k=1}^{q}$ and $U_P = \{\mathbf{x}_2^k, \ldots, \mathbf{x}_{p/q+1}^k\}_{k=1}^{q}$, for AL and SSL in the CLUC method, respectively. The CBQPL procedure assigns these secondary uncertain unlabeled samples with the same label as the most uncertain one, which is manually-labeled in AL within the same cluster. Superpixel techniques can provide an efficient segmentation on the HSI image.

The pseudo-code of the CBQPL pseudo-labeling is illustrated in Algorithm 2. In details, given $\{\mathbf{x}_1^k\}_{k=1}^{q}$ and $\{\mathbf{x}_2^k, \ldots, \mathbf{x}_{p/q+1}^k\}_{k=1}^{q}$, the label of $\mathbf{x}_1^k$ marked by human in AL is assigned to the p/q unlabeled samples $\{\mathbf{x}_2^k, \ldots, \mathbf{x}_{p/q+1}^k\}$ in SSL. In total, using CBQ query strategy, q unlabeled samples are selected and labeled by human, and meanwhile the corresponding p unlabeled samples are selected and pseudo-labeled by CBQPL strategy. In [50,51], superpixels are regarded as small local regions with adaptive shapes and sizes. In [53], each superpixel is regarded as set of pixels containing spectrally similar and spatially connected pixels, which are assumed to have the same label. In [55,67], the traditional pixel-level classification and MRF regularization methods are utilized in superpixel-level, respectively, assuming that each superpixel is collectively assigned the same class label. As is analyzed in Section 3.1, it is assumed that, the unlabeled samples within each superpixel attaches high probability to be the same class using efficient superpixel segmentation method.

3.4. Details of the CLUC method

The flowchart and the pseudo-code of the proposed CLUC algorithm are illustrated in Fig. 1 and Algorithm 3, respectively. Given an initial labeled training set L_0 and an initial unlabeled data set U, the CLUC algorithm starts from clustering the unlabeled sample into a set of clusters $\boldsymbol{C}$, and arrange these clusters into $\boldsymbol{C}^{BT}$ and SP_n^{BT}. The LMLL method is utilized as the basic classifier in this paper. With the initial labeled data set $L_{ML} = \emptyset$ and $L_{PL} = \emptyset$, the LMLL classifier is first trained using $L_0 \cup L_{ML} \cup L_{PL}$. Then, based on the CBQ strategy and the classification results obtained by LMLL, the unlabeled data U_Q and U_P are selected for AL and SSL, respectively. The unlabeled data in U_Q are labeled by human in AL. Specially, the unlabeled data in U_P are labeled using the CBQPL pseudo-labeling strategy. These $(q + p)$ newly manually-labeled and pseudo-labeled samples will be removed from U and added into L_{ML} and L_{PL}. After updating U, L_{ML}, and L_{PL}, the LMLL classifier will be trained again using the new training set $L_0 \cup L_{ML} \cup L_{PL}$. The algorithm will be repeated until the size of the new training set achieves a threshold predefined or the set of unlabeled sample becomes empty.

Given the initial limited training samples, the CLUC algorithm collaboratively employs the AL and SSL learning process, aiming to obtain more labeled samples with higher confidence and uncertainty. The designed CBQ strategy is based on SPBT, which can provide more diversity in the active selection process. Specially, in CLUC, the CBQPL strategy makes the accuracy of pseudo-labeling rely on the effectiveness of the adopted clustering method, rather than the classification results of the classifiers in the aforemen-

Algorithm 3
CLUC.

Inputs:
Initial labeled data set: $L_0 = \{(\mathbf{x}_i, \mathbf{y}_i)_{i=1}^{l}\}$;
Initial unlabeled data set: $U = \{\mathbf{x}_j\}_{j=1}^{u}$;
Initial unlabeled data set: $L_{ML} = \emptyset$, and $L_{PL} = \emptyset$;
AL query size: q, and SSL query size: p;
Number of the clusters for clustering: N;
Threshold of the size of the expanded labeled data set: T;
1. Cluster and sort the unlabeled samples into a set of superpixels using SPBT:
$\boldsymbol{C}^{BT} = \{SP_1^{BT}, \ldots, SP_N^{BT}\}^{BT}$;
2. **Repeat**
3. Train *LMLL* using $L_0 \cup L_{ML} \cup L_{PL}$
4. Select q data $U_Q = \{\mathbf{x}_k\}_{k=1}^{q}$, and p data $U_P = \{\mathbf{x}_k\}_{k=1}^{p}$ from Uusing *LMLL*, $\boldsymbol{C}$, and SPBT;
5. Label U_Q by experts, and update the manually-labeled data set: $L_{ML} = L_{ML} \cup U_Q$;
6. Label U_P by human experts, and update the CBQPL-labeled data set: $L_{PL} = L_{PL} \cup U_P$;
7. Update the unlabeled data set: $U = U \backslash U_Q \backslash U_P$;
8. **Until** *sizeof*$(L_0 \cup L_{ML}) \geq T$ or $U == \emptyset$;
9. **Return** *LMLL*;

Algorithm 4
CBQ-AL (for one iteration in AL via SPBT).

Inputs:
Initial unlabeled data set: $U_Q = \emptyset$;
Current unlabeled data set: $U = \{\mathbf{x}_j\}_{j=1}^{u}$;
Current classification results: $p(y_i = k|\mathbf{x}_i), \mathbf{x}_i \in U$;
AL query size: q;
Number of the clusters for clustering: N;
1. Cluster unlabeled samples U into a set of clusters via ERS based on Eqs. (1) and (2):
$\boldsymbol{C} = \{SP_1, \ldots, SP_N\}, SP_n = \{\mathbf{x}_1^n, \ldots, \mathbf{x}_{M_n}^n\}$;
2. Rank the clustering result based on the strategies in Eqs. (3)–(6):
$\boldsymbol{C}^{BT} = \{SP_1^{BT}, \ldots, SP_N^{BT}\}^{BT}, SP_n^{BT} = \{\mathbf{x}_1^n, \ldots, \mathbf{x}_{M_n}^n\}^{BT}$;
3. Select q data $U_Q = \{\mathbf{x}_1^k\}_{k=1}^{q}$ from $\boldsymbol{C}^{BT}$ for AL;
4. Update the unlabeled data set: $U = U \backslash U_Q$;

tioned CL methods (analyzed in Section 2.2). Then, the pseudo-labeled samples are used together with the initial training samples to improve the performance of the final classifier.

4. Experimental results

In this section, we evaluate the performance of the proposed CLUC method using three real hyperspectral image datasets: the AVIRIS Indian Pines image, the AVIRIS Salinas image, and the University of Pavia image acquired by the ROSIS sensor. Using the SPBT and CBQPL and LMLL [39] as active query heuristic and basic classifier, respectively, we term the proposed method as CLUC-SPBT-LMLL in this section. The experiment focuses on verifying the effectiveness of the proposed CLUC algorithm with limited manually-labeled samples in comparison with other AL approaches, such as AL-BT-LMLL, AL-MBT-LMLL [39], AL-SPBT-LMLL and CLUC-SPBT$_{RS}$-LMLL. In addition, two approaches combining AL and SSL learning are also compared in our experiment. For MLR^{Semi}MLL-ME [70], the regressors in multinomial logistic regression are learning using both labeled and unlabeled samples through a graph-based SSL technique, and ME is selected as the AL query strategy. In MPM-LBP-MBT [71], the marginal probability distribution is exploited using the whole information in the hyperspectral data and the AL query criterion is MBT.

Without regard of the SSL query for unlabeled sample, the SPBT strategy can be used as a query heuristic in AL learning. Algorithm 4 illustrates the pseudo-code of the SPBT query strategy for one loop in AL. As we can see, the unlabeled sample $U_Q = \{\mathbf{x}_1^k\}_{k=1}^{q}$ are selected and then labeled by human.

In order to make a fair comparison, the LMLL algorithm is adopted as the basic classifier for these AL and CL learning methods. The Gaussian radius basis function (RBF) kernel is adopted to improve the data separability in the transformed feature space, and the parameters for LMLL are set to default values in [39]. Different active query strategies, BT, MBT, and SPBT, are utilized in the AL learning approaches. The difference of CLUC-SPBT$_{RS}$-LMLL from CLUC-SPBT-LMLL is that for the former the pseudo-labeled samples are randomly selected from the set of samples containing the secondary uncertain samples. Additionally, the objective metrics, overall accuracy (OA), average accuracy (AA), and Kappa statistic are used to evaluate the classification results.

4.1. Data sets

The Indian Pines image was collected by the AVIRIS sensors over North-western Indiana. The image consists of 145 × 145 pixels and 224 spectral bands. The ground truth of Indian Pines contains 16 classes. The number of the bands is reduced from 224 to 200 in our experiment for the 24 bands ([104–108], [150–163], 220) covering the region of water absorption. The color composite of the Indian Pines image and the ground truth are shown in Fig. 3(a) and (b), respectively.

The AVIRIS Salinas Scene image was gathered with 224 bands by AVIRIS sensor over Salinas Valley. This image is characterized by high spatial resolution of 3.7-meter pixels. It consists of vegetables, bare soils, and vineyard fields, and has 512 × 217 samples. The 20 water absorption band ([108–112], [154–167], 224) are removed in the experiment. Fig. 5(a) and (b) show its color composite image and 16 classes of ground references.

The University of Pavia image was acquired by the ROSIS sensor over Pavia, north Italy. The image consists of 610 × 340 pixels and 103 spectral bands after discarding 12 bands for water absorption and noise. The image has 9 classes of ground references, and a geometric resolution of 1.3 m. Its color composite image is shown in Fig. 7(a), and its ground references are shown in Fig. 7(b), respectively.

4.2. Classification results

In the experiments, the parameters for the proposed CLUC-SPBT-LMLL algorithms are empirically selected for the Indian Pines, Salinas Scene and University of Pavia images. The number of superpixels is set the same for the three HSI test images, i.e., $N = 200$. With q unlabeled samples selected for manually-labeling based on the SPBT criterion, $p = 5q$ unlabeled samples are selected for the CBQPL-based labeling. Specifically, the 5 SPBT-based most uncertain samples in the set of the secondary uncertain samples within each superpixel (of the q most uncertain superpixels) are assigned with the same label as the manually-labeling for the most uncertain sample, according to the CBQPL pseudo-labeling method described in Section 3.3.

The first experiment is performed on the Indian Pines image which contains 16 different classes. For this image, five samples of each class are randomly selected as the initial training set and the rest for testing. The numbers of manually-labeled samples are ranged from 80 to 530. The classification maps of the MLR^{Semi}MLL-ME, AL-BT-LMLL, AL-MBT-LMLL, AL-SPBT-LMLL, MPM- LBP-MBT, CLUC-SPBT$_{RS}$-LMLL, and CLUC-SPBT-LMLL approaches are illustrated in Fig. 3(c)–(i), respectively. Fig. 4(a)–(c) shows the mean values and standard deviations of the OA, AA and kappa accuracies when the number of manually-labeled samples increase. All the objective metrics are obtained after twenty independent realizations by the compared algorithms.

We can observe from Fig. 3 and Table 1 that CLUC-SPBT-LMLL obtains the highest accuracies in terms of OA among all compared approaches. From Fig. 4(a)–(c), we can observe that the performance of all the CLUC and AL approaches generally increase when more samples are labeled by human. Moreover, as can be seen that CLUC-SPBT-LMLL and CLUC-SPBT$_{RS}$-LML present obviously better classification results than MLR^{Semi}MLL-ME, AL-BT- LMLL, AL-MBT-LMLL and AL-SPBT-LMLL. This is due to that the CLUC algorithms make full use of the labeled samples by human. In other words, the CBQPL strategy provides more pseudo-labeled sam-

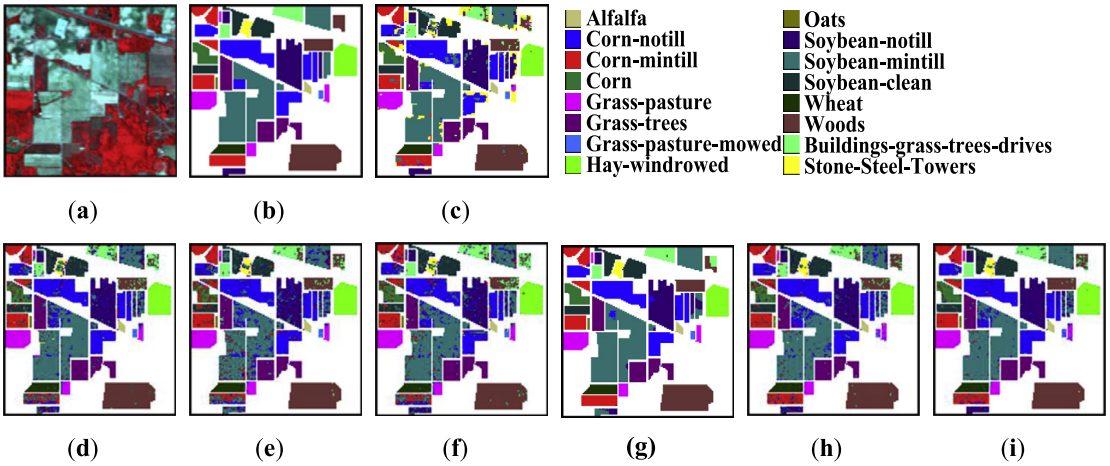


Fig. 3. The Indian Pines image (a) three-band color composite image, (b) ground truth image, and the classification results (OA in %) obtained by the methods: (c) MLR^{Semi}MLL-ME, OA = 88.23%, (d) AL-BT- LMLL, OA = 92.24%, (e) AL-MBT-LMLL, OA = 91.46%, (f) AL-SPBT-LMLL, OA = 93.22%, (g) MPM- LBP-MBT, OA = 96.56%, (h) CLUC-SPBT$_{RS}$-LMLL, OA = 96.21%, and (i) CLUC-SPBT-LMLL, OA = 97.66%.

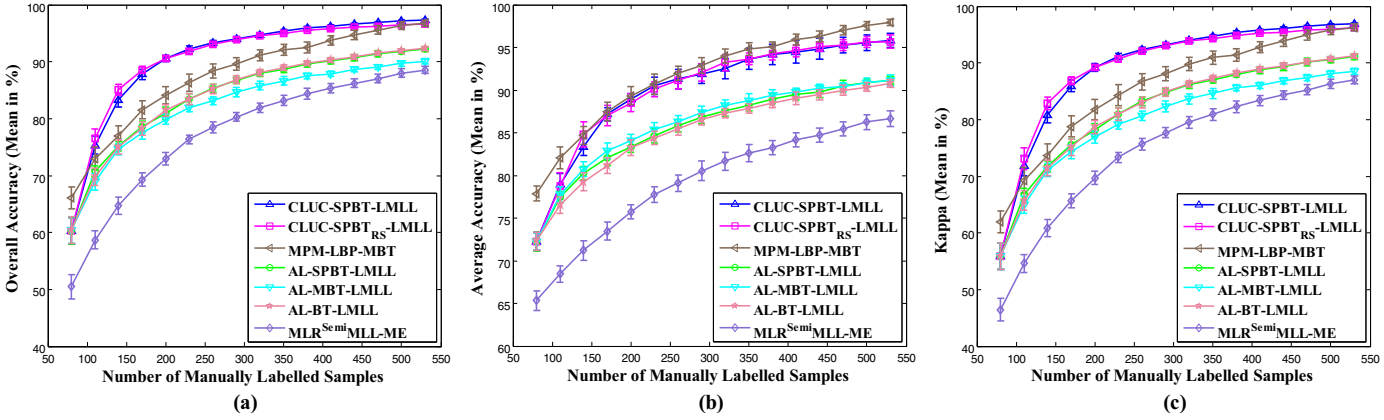


Fig. 4. (a) OA, (b) AA, and (c) Kappa coefficient obtained by the MLR^{Semi}MLL-ME, AL-BT-LMLL, AL-MBT-LMLL, AL-SPBT-LMLL, MPM-LBP-MBT, CLUC-SPBT$_{RS}$-LMLL, and CLUC-SPBT-LMLL methods on the Indian Pines image. The height of the error bars refer to the standard deviations over twenty independent random runs.

Table 1
OA, AA, and Kappa coefficient obtained by the MLR^{Semi}MLL-ME, AL-BT-LMLL, AL-MBT-LMLL, AL-SPBT-LMLL, MPM-LBP-MBT, CLUC-SPBT$_{RS}$-LMLL, and CLUC-SPBT-LMLL methods on the three images. The quantitative metrics are obtained over twenty independent random runs.

Method	Indian Pines			Salinas Scene			University of Pavia		
	OA (%)	AA (%)	Kappa (%)	OA (%)	AA (%)	Kappa (%)	OA (%)	AA (%)	Kappa (%)
MLR^{Semi}MLL-ME	88.57 ± 0.64	86.67 ± 0.89	86.98 ± 0.73	90.75 ± 0.73	95.59 ± 0.37	89.71 ± 0.80	86.70 ± 1.12	81.03 ± 2.40	82.30 ± 1.45
AL-BT-LMLL	92.32 ± 0.32	90.77 ± 0.48	91.19 ± 0.37	92.78 ± 0.65	95.89 ± 0.36	91.95 ± 0.72	91.31 ± 0.77	87.83 ± 0.34	88.33 ± 1.04
AL-MBT-LMLL	90.06 ± 0.49	91.11 ± 0.63	88.54 ± 0.58	92.98 ± 0.53	96.68 ± 0.33	92.17 ± 0.59	92.35c0.59	90.07 ± 0.54	89.70 ± 0.80
AL-SPBT-LMLL	92.21 ± 0.37	91.15 ± 0.48	91.06 ± 0.43	93.45 ± 0.31	96.45 ± 0.23	92.69 ± 0.34	92.01 ± 0.83	88.72 ± 1.00	89.26 ± 1.13
MPM-LBP-MBT	96.72 ± 0.56	**97.95 ± 0.38**	96.65 ± 0.65	95.15 ± 0.62	**97.86 ± 0.31**	94.59 ± 0.69	95.75 ± 0.88	**94.79 ± 0.36**	94.37 ± 1.14
CLUC-SPBT$_{RS}$-LMLL	96.69 ± 0.14	95.68 ± 0.76	96.22 ± 0.16	**95.24 ± 0.14**	97.35 ± 0.16	**94.69 ± 0.16**	95.40 ± 0.37	91.67 ± 1.11	93.87 ± 0.49
CLUC-SPBT-LMLL	**97.25 ± 0.20**	95.83 ± 0.81	**96.86 ± 0.23**	94.58 ± 0.19	97.24 ± 0.14	93.95 ± 0.21	**95.79 ± 0.18**	93.03 ± 0.58	**94.39 ± 0.25**

ples to train the classifier at each iterative learning process. In addition, compared with CLUC-SPBT$_{RS}$-LML, CLUC-SPBT-LMLL obtains the best results in all objective metrics. For example, with 530 samples labeled by human, the OA, AA and kappa accuracies obtained by CLUC-SPBT-LMLL reach 97%, 95% and 96%, respectively. In detail, with an initial increase of the manually-labeled samples, CLUC-SPBT$_{RS}$-LML presents a very small advantage over the CLUC-SPBT-LMLL method. Then, as the number of manually-labeled samples continues increasing, CLUC-SPBT-LMLL outperforms CLUC-SPBT$_{RS}$-LML in classification accuracies. This is due to that for the CLUC-SPBT-LMLL, the CBQ query within clusters is inter-cluster biased, which partly increase the accuracies of CBQPL labeling. The CLUC-SPBT-LMLL sacrifices some diversity in the CBQ query strategy within each cluster to ensure the accuracy of CBQPL pseudo-labeling.

The second and the third experiments are conducted on the Salinas Scene and University of Pavia images, respectively. For each class in these two images, five samples are randomly selected to comprise the initial training sets and the remaining data are used for test. For the Salinas Scene image with 16 classes, the MLR^{Semi}MLL-ME, AL-BT-LMLL, AL-MBT-LMLL, AL-SPBT-LMLL, MPM-LBP-MBT, CLUC-SPBT$_{RS}$-LMLL, and CLUC-SPBT-LMLL algorithms start with 80 training samples, and the numbers of AL and SSL query samples are set as $q = 20$ and $p = 100$ at each iteration, and totally 20 iterations are executed. For the University of Pavia image with 9 classes, 45 samples are selected to com-

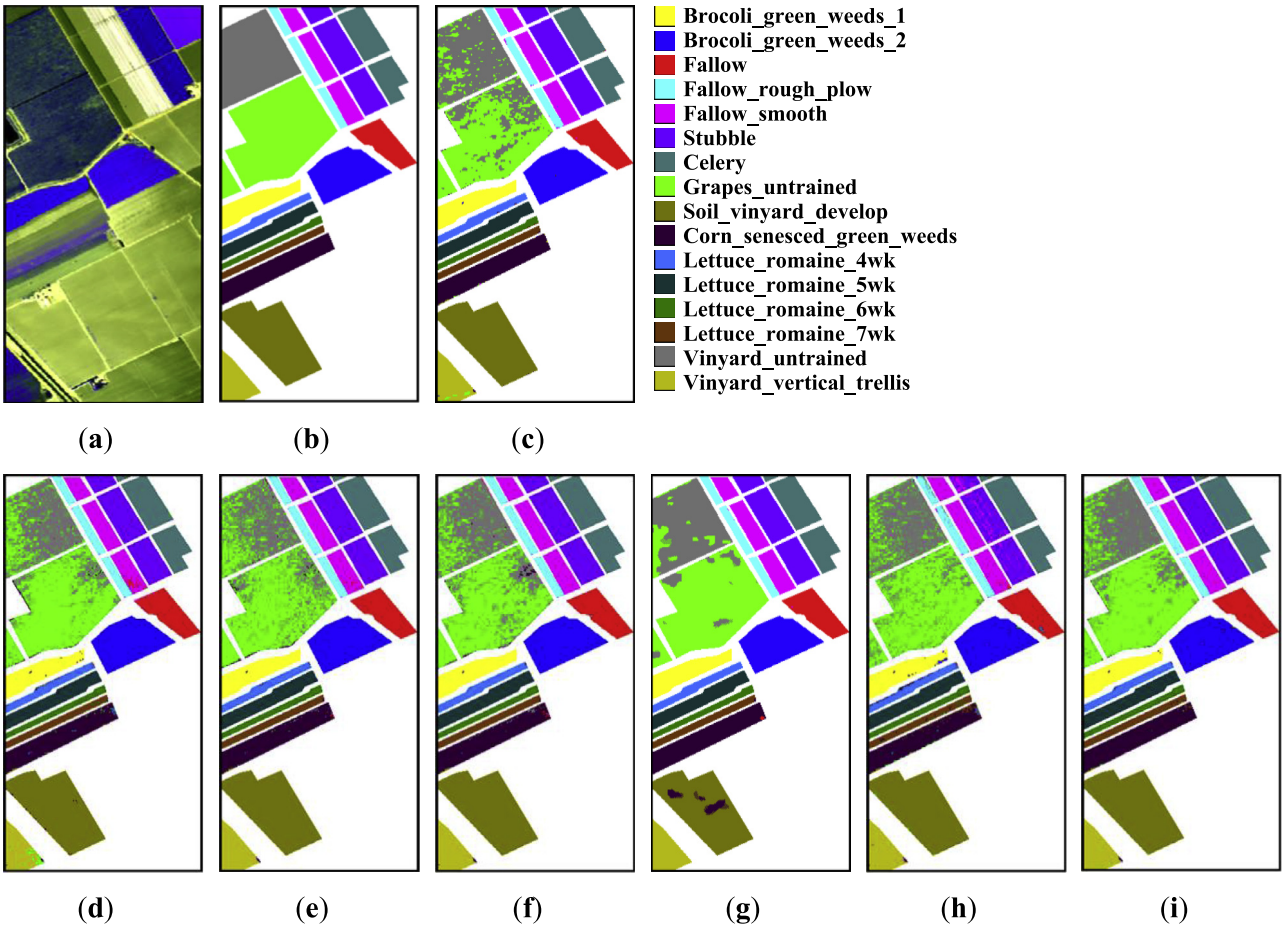


Fig. 5. The Salina Scene image (a) three-band color composite image, (b) ground truth image, and the classification results (OA in %) obtained by the methods: (c) MLR^{Semi}MLL-ME, OA = 92.31%, (d) AL-BT-LMLL, OA = 93.12%, (e) AL-MBT- LMLL, OA = 93.14%, (f) AL-SPBT-LMLL, OA = 93.43%, (g) MPM-LBP-MBT, OA = 95.13%, (h) CLUC-$SPBT_{RS}$-LMLL, OA = 95.27%, and (i) CLUC-SPBT-LMLL, OA = 94.61%.

prise the initial training data for the compared approaches, and the number of AL and SSL query samples are set as $q = 15$ and $p = 75$, respectively, and the classification results are obtained after the loop executes 10 times. The classification accuracies (OA, AA, and kappa) are the mean values over twenty independent runs.

The classification maps obtained using different algorithms on these two images are shown in Fig. 5(c)–(i) and Fig. 7(c)–(g), respectively. The performance of the CLUC methods generally outperform the compared AL approaches in terms of the OA accuracies. Fig. 6(a)–(c) and Fig. 8(a)–(c) illustrate the OA, AA, and kappa accuracies obtained by different approaches when the numbers of the manually-labeled samples increase (from 80 to 380 for the Salinas image, and from 45 to 195 for the University of Pavia image, respectively). We can observe that the classification accuracies of all the algorithms generally improve as more samples are selected and labeled by human. First, the CLUC-$SPBT_{RS}$-LMLL and CLUC-SPBT-LMLL methods outperform the compared MLR^{Semi}MLL-ME, AL-BT-LMLL, AL-MBT-LMLL, and AL-SPBT- LMLL, and MPM-LBP-MBT approaches in terms of OA and Kappa (mean in %) with different number of manually-labeled samples. And, the standard deviations (indicated by the height of the error bars) obtained by the CLUC methods are also smaller than those obtained by the AL algorithms. These two experiments demonstrate that the proposed CLUC methods can provide better and more stable performance than the compared approaches. Moreover, compared with AL-BT-LMLL and AL-MBT-LMLL, AL-SPBT-LMLL obtains higher classification accuracies in terms of OA and Kappa. This demonstrates that the SPBT query strategy promotes more sampling diversity in feature and spatial domains than the BT and MBT criterions.

4.3. Parameter analysis

In this subsection, we will analyze the effect of the number of superpixels (clusters) N and the number of SSL query samples within each superpixel p/q on the performance of the proposed CLUC-SPBT-LMLL framework. The Indian Pines, Salinas Scene, and University of Pavia images are used in this analysis. And the numbers of training and testing data, including other parameters related, are set the same as those described in Section 4.2. The objective metric, i.e., OA, used here, is the average over twenty independent runs.

The number of superpixels ranges from 50 to 1000 for these three HSI images. Fig. 9(a)–(c) shows the OA accuracies obtained by CLUC-SPBT-LMLL with different numbers of superpixels when the manually-labeled samples increase. We take Fig. 9(a) as an example to explain the experimental results. The horizontal coordinates of the curves in Fig. 9(a) denotes the number of manually labeled samples, which are changed from 80 to 530 and 50 samples are labeled at each iteration. It can be observed that the OA accuracies obtained by the proposed method on the Indian Pines image generally increase when more manually labeled samples are used. And "N = 50–1000″ at the bottom right of the curves refers to the number of superpixels on the classification. The performance of the proposed method has an obvious improve-

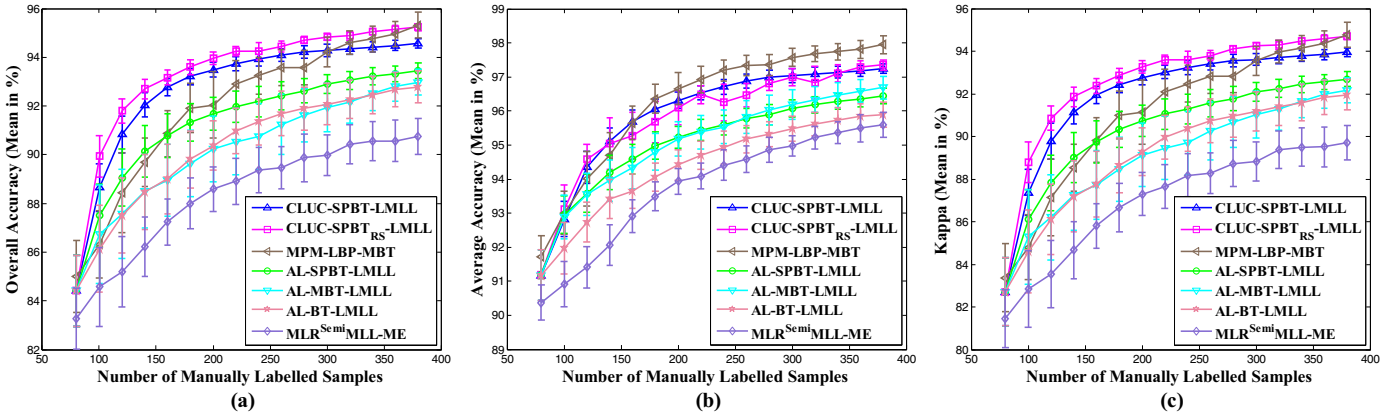


Fig. 6. (a) OA, (b) AA, and (c) Kappa coefficient obtained by the MLR^{Semi}MLL-ME, AL-BT-LMLL, AL-MBT-LMLL, AL-SPBT-LMLL, MPM-LBP-MBT, CLUC-SPBT$_{RS}$-LMLL, and CLUC-SPBT-LMLL methods on the Indian Pines image. The height of the error bars refer to the standard deviations over twenty independent random runs.

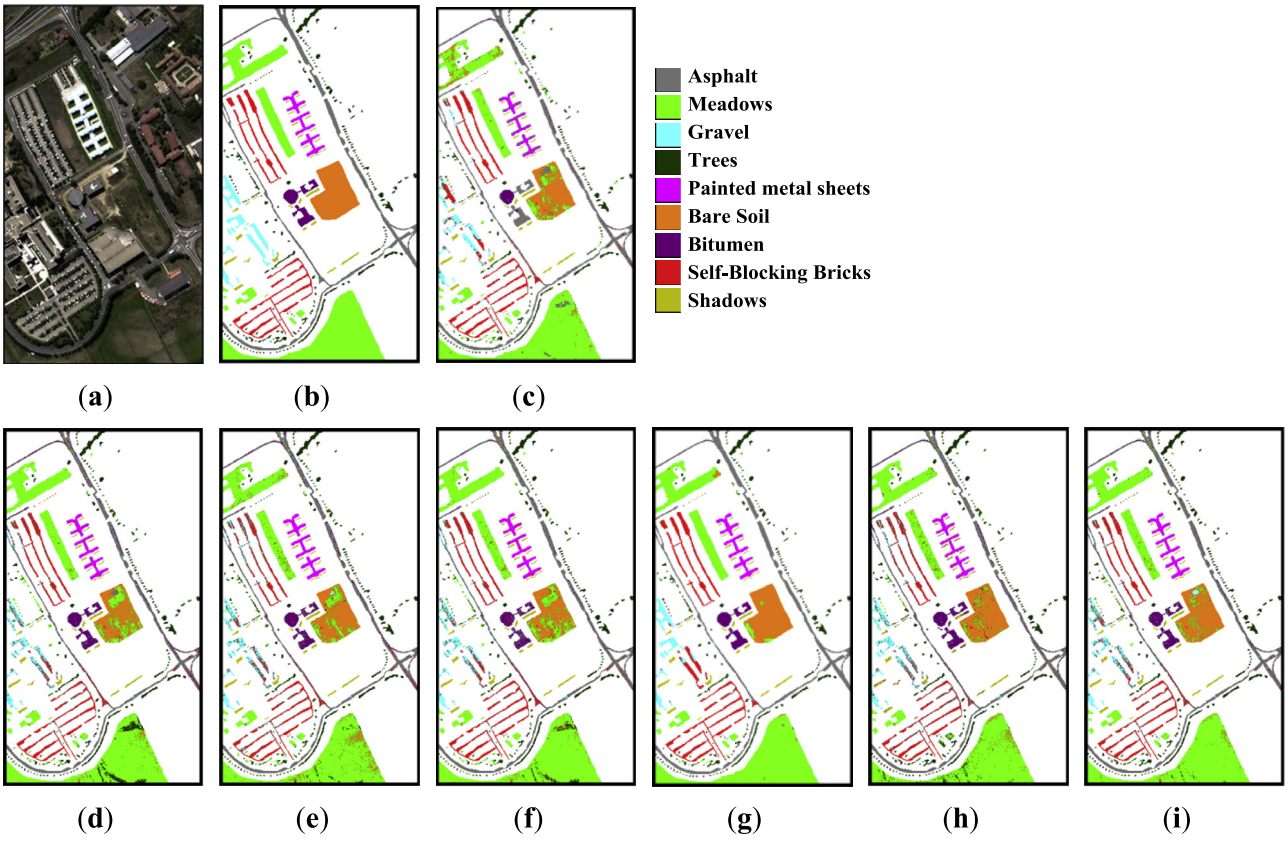


Fig. 7. The University of Pavia image (a) three-band color composite image, (b) ground truth image and the classification results (OA in %) obtained by the methods (c) MLR^{Semi}MLL-ME, OA = 86.57% (d) AL-BT-LMLL, OA = 90.67%, (e) AL-MBT-LMLL, OA = 92.81%, (f) AL-SPBT-LMLL, OA = 90.76%, (g) MPM-LBP-MBT, OA = 96.37% (h) CLUC-SPBT$_{RS}$-LMLL, OA = 95.97%, and (i) CLUC-SPBT-LMLL, OA = 96.23%.

ment when N changes from 50 to 200 for the Indian Pines images. When N continues increasing from 200 to 1000, the OA values is almost no change. We can observe similar results on the other two testing HSI images. For larger number of superpixels, the HSI image is more over-segmented. And this will improve the confidence of CBQPL labeling and avoid unreliable pseudo-labels in the SSL learning process. On the other hand, however, with more clusters, the newly-added training samples at each active query process tend to be the same class. This will reduce the diversity of manually-labeled samples in AL learning process. This experimental results show that the proper value of N, e.g., $N = 200$, for three HSI images, ensures a promising balance between the sampling diversity of CBQ and the accuracy of CBQPL. In this case, CLUC presents a promising performance for HSI classification.

The number of SSL query samples within each superpixel p/q is set from 1 to 9 to analyze its effect on the performance of CLUC-SPBT-LMLL. The obtained OA metrics on the three images are illustrated in Fig. 10(a)–(c), respectively. As can be seen, the values of OA improve significantly as p/q increases from 1 to 5. With p/q larger than 5, there is nearly no improvement on the classification result. However, too large p/q will increase the time consumption of CLUC. In addition, the effect of p/q on the run time of the proposed method will be analyzed in detail in the next subsection.

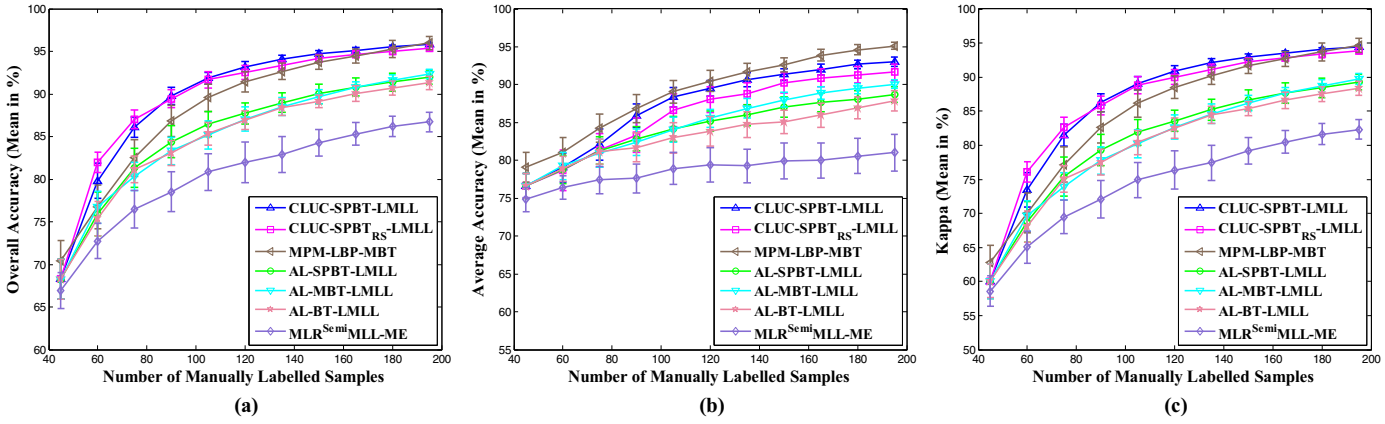


Fig. 8. (a) OA, (b) AA, and (c) Kappa coefficient obtained by the MLR^{Semi}MLL-ME, AL-BT-LMLL, AL-MBT-LMLL, AL-SPBT-LMLL, MPM-LBP-MBT, CLUC-SPBT$_{RS}$-LMLL, and CLUC-SPBT-LMLL methods on the Indian Pines image. The height of the error bars refer to the standard deviations over twenty independent random runs.

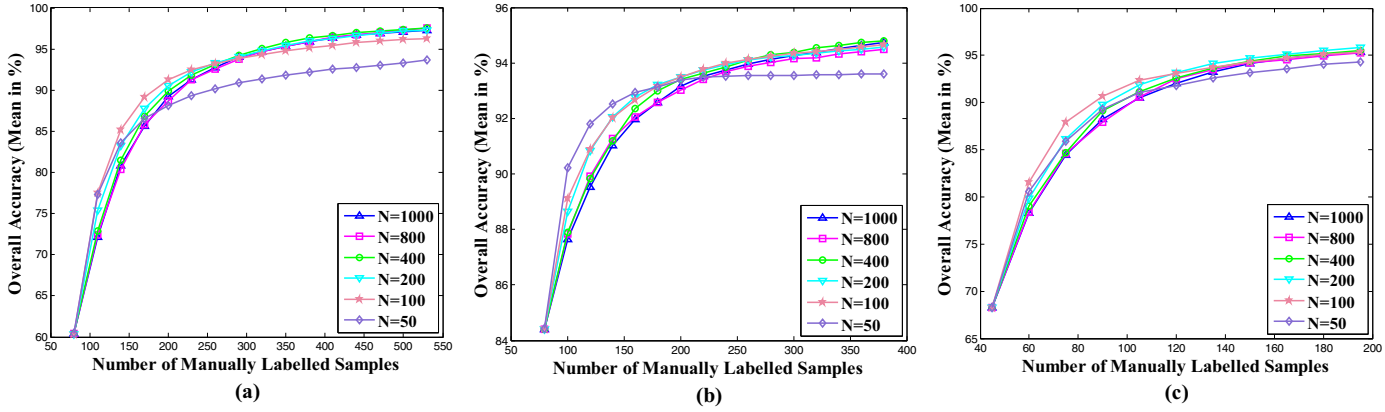


Fig. 9. Effect of the numbers of superpixels on the classification results obtained by the CLUC-SPBT- LMLL method on the (a) Indian Pines, (b) Salinas Scene, and (c) University of Pavia images, respectively. N indicates the number of superpixels for different HSI images.

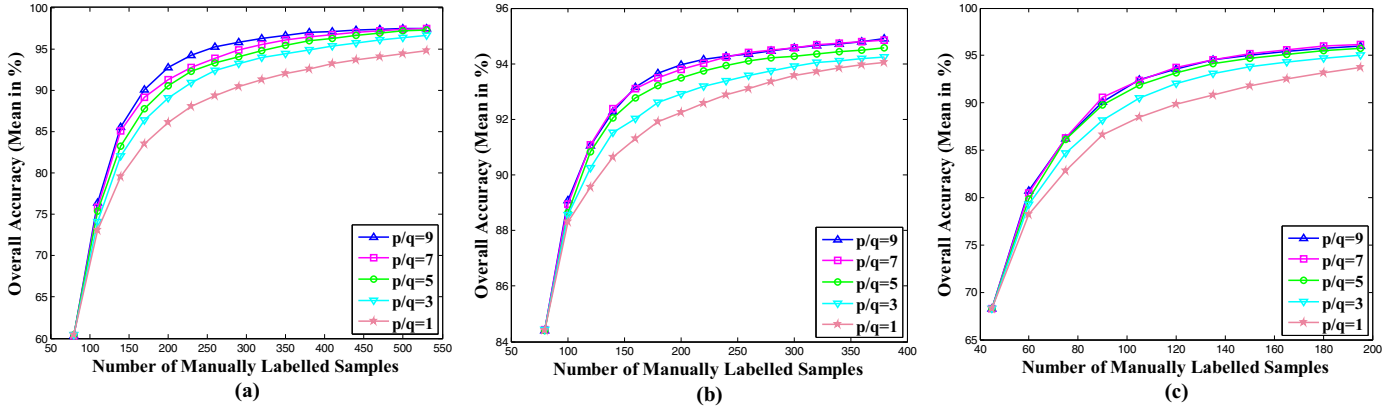


Fig. 10. Effect of the numbers of CBQPL-labeled samples on the classification results obtained by the CLUC-SPBT-LMLL method on the (a) Indian Pines, (b) Salinas Scene, and (c) University of Pavia images, respectively. p/q indicates the number of samples selected for CBQPL-labeling within each superpixel.

4.4. Computational complexity analysis

This subsection gives an analysis on the computational complexity and CPU time of the proposed CLUC framework. We will analyze the overall time complexity of CLUC. Then we will point out the effect of the number of SSL query samples p/q on the run time of CLUC. The parameters are set the same as the experiment described in Section 4.2. And, all the algorithms are conducted on a personal computer with an Intel Core i5-4590 3.30-GHz CPU and 8.00-GB RAM.

Let n, K, and d indicate the number of pixels, the number of classes, and the dimension of kernel function, respectively. In CLUC, ERS is used to group the samples into clusters, which has the complexity approximating $O(nlogn)$ [48]. For the basic classifier LMLL adopted in this paper, the posterior probability is learned through the LORSAL algorithm, following by a MLL regularization, which is solved by the α-expansion min-cut algorithm [68]. The LORSAL and MLL methods have the complexity of $O(d^2K)$ and $O(n)$, respectively. The ERS algorithm is executed one time before the iterative learning process. The overall com-

Table 2
The CPU time (in second) of the CLUC-SPBT$_{RS}$-LMLL, and CLUC-SPBT-LMLL methods on the three images. The values are obtained over twenty independent random runs.

Method	CPU Time (s)		
	Indian Pines	Salinas Scene	University of Pavia
CLUC-SPBT$_{RS}$-LMLL	131.52	149.66	83.81
CLUC-SPBT-LMLL	147.35	159.85	80.28

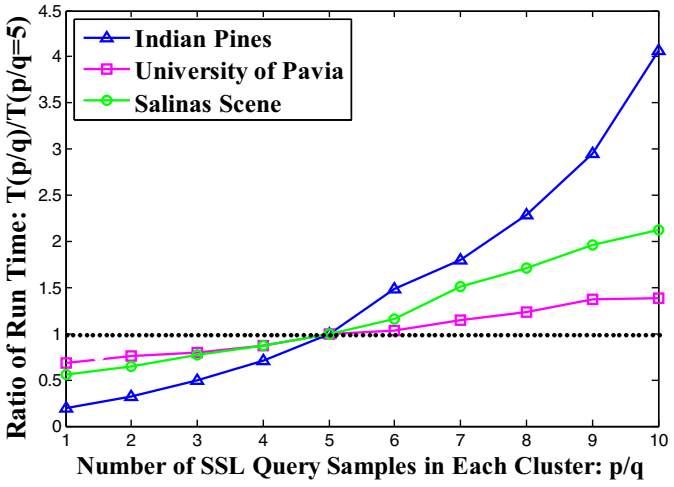


Fig. 11. Effect of the number of SSL query samples on the run time of the CLUC-SPBT-LMLL method on the Indian Pines, Salinas Scene, and University of Pavia images, respectively. p/q indicates the number of samples selected for CBQPL labeling within each superpixel.

plexity of CLUC is dominated by LMLL, which is utilized at each iteration.

Table 2 presents the run time of the CLUC-SPBT$_{RS}$-LMLL and CLUC-SPBT-LMLL methods on the three images over twenty independent realizations. And the effect of the number of SSL query samples on the run time of CLUC-SPBT-LMLL is illustrated in Fig. 11. Note that in Fig. 11, $T(p/q)$ indicates the run time of CLUC-SPBT-LMLL when p/q samples within each cluster are selected for SSL. It can be found that, the run time of CLUC-SPBT-LMLL on the three images increases as p/q increases. Considering the effect of p/q on classification accuracy of the CLUC-SPBT-LMLL (illustrated in Fig. 9), the performance of the proposed method has nearly no change when p/q is bigger than 5. Therefore, five unlabeled samples within each superpixel are selected in SSL and pseudo-labeled by CBQPL in our experiment. Again, the overall complexity of CLUC is dominated by the basic classifier adopted at each CL iteration. With more efficient classifiers, CLUC can be more efficiency in time consumption.

5. Conclusions

In this paper, we present a novel CLUC framework for hyperspectral images classification, which integrates AL and SSL collaboratively using clustering. The proposed CLUC method aims to find the most confident samples that have the same labels with the manually-labeled samples in the collaborative learning process, instead of pseudo-labeling the unlabeled samples using the current classification result. Moreover, the CBQ query heuristic promotes more sampling diversity in AL and SSL query procedure. In addition, the CBQPL strategy provides an effective pseudo-labeling scheme, which makes full use of the human labeling efforts. The LMLL algorithm is adopted as the basic classifier at each CL iteration. The superiority of CLUC is verified by the comparison with several HSI classification algorithms with limited manually-labeled samples.

As is shown that the CBQPL pseudo-labeling plays an important role in the CLUC framework. Therefore, a crucial issue is to ensure the confidence of pseudo-labeling procedure and avoid unreliable pseudo-labels. See that the proposed method still has its room for improvement. The classification results (used in [34,35] for pseudo-labeling verification) at each CL iteration can be further exploited in CLUC, which can be integrated with CBQPL to improve the pseudo-label assignment. Moreover, superpixel is used for the clustering of the unlabeled samples, and other efficient clustering algorithm, e.g. hierarchical-clustering can be developed in CLUC. In addition, some validation strategies (e.g. in [69]) could be researched to evaluate the prediction of pseudo-labels. Finally, alternative methods for the basic classifiers and the AL query heuristic, e.g. SVM [8] and MCLU [36], will also be researched in our future work.

Acknowledgments

The authors would like to thank Dr. M.-Y. Liu and Dr. J. Li for providing the source codes of superpixel and HSI classification, respectively, on their websites (http://www.mingyuliu.net/ and http://www.lx.it.pt/~jun/). This work was supported in part by the National Natural Science Foundation of China under Grants 61432014, U1605252, 61772402 and 61671339, in part by the National Key Research and Development Program of China under Grant 2016QY01W0200, in part by Key Industrial Innovation Chain in Industrial Domain under Grant 2016KTZDGY04-02, and in part by National High-Level Talents Special Support Program of China under Grant CS31117200001.

References

[1] R. Moreno, F. Corona, A. Lendasse, M. Graña, L.S. Galvão, Extreme learning machines for soybean classification in remote sensing hyperspectral images, Neurocomputing 128 (5) (2014) 207–216.
[2] B. Du, L. Zhang, D. Tao, D. Zhang, Unsupervised transfer learning for target detection from hyperspectral images, Neurocomputing 120 (10) (2013) 72–82.
[3] W. Heldens, U. Heiden, T. Esch, E. Stein, A. Muller, Can the future EnMAP mission contribute to urban applications? a literature survey, Remote Sens. 3 (9) (2011) 1817–1846.
[4] M.L. Clark, D.A. Roberts, Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier, Remote Sens. 4 (2012) 1820–1855.
[5] L. Zhang, X. Huang, Object-oriented subspace analysis for airborne hyperspectral remote sensing imagery, Neurocomputing 73 (4) (2010) 927–936.
[6] J.M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N.M. Nasrabadi, J. Chanussot, Hyperspectral remote sensing data analysis and future challenges, IEEE Geosci. Remote Sens. Mag. 1 (2) (2013) 6–36.
[7] J. Li, J.M. Bioucas-Dias, A. Plaza, Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression, IEEE Geosci. Remote Sens. Lett. 10 (2) (2013) 318–322.
[8] F. Melgani, L. Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, IEEE Trans. Geosci. Remote Sens. 42 (8) (2014) 1778–1790.
[9] Y. Chen, H. Jiang, C. Li, X. Jia, Deep feature extraction and classification of hyperspectral images based on convolutional neural networks, IEEE Trans. Geosci. Remote Sens. 54 (10) (2016) 1–20.
[10] J. Li, X. Huang, P. Gamba, J.M. Bioucas-Dias, L. Zhang, J.A. Benediksson, A. Plaza, Multiple feature learning for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 53 (3) (2015) 1592–1606.
[11] C. Zhao, X. Gao, Y. Wang, J. Li, Efficient multiple-feature learning-based hyperspectral image classification with limited training samples, IEEE Trans. Geosci. Remote Sens. 54 (7) (2016) 4052–4062.
[12] Y. Chen, N.M. Nasrabadi, T.D. Tran, Hyperspectral image classification using dictionary-based sparse representation, IEEE Trans. Geosci. Remote Sens. 49 (10) (2011) 3973–3985.
[13] E. Zhang, X. Zhang, L. Jiao, H. Liu, S. Wang, B. Hou, Weighted multifeature hyperspectral image classification via kernel joint sparse representation, Neurocomputing 178 (C) (2015) 71–86.
[14] D. Tuia, F. Ratle, F. Pacifici, M.F. Kanevski, W.J. Emery, Active learning methods for remote sensing image classification, IEEE Trans. Geosci. Remote Sens. 47 (1) (2009) 2218–2232.
[15] J. Li, J.M. Bioucas-Dias, A. Plaza, Semi-supervised hyperspectral image segmentation using multinomial logistic regression with active learning, IEEE Trans. Geosci. Remote Sens. 48 (11) (2010) 4085–4098.

[16] J. Li, J.M. Bioucas-Dias, A. Plaza, Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields, IEEE Trans. Geosci. Remote Sens. 50 (3) (2012) 809–823.

[17] Z. Cui, Y. Wang, X. Gao, Y. Zheng, Multispectral image classification based on improved weighted MRF bayesian, Neurocomputing 212 (5) (2016) 75–87.

[18] D. Tuia, M. Volpi, L. Copa, M. Kanevski, J. Muñoz-Marí, A survey of active learning algorithms for supervised remote sensing image classification, IEEE J. Sel. Topics Signal Process. 5 (3) (2011) 606–617.

[19] M.M. Crawford, D. Tuia, H.L. Yang, Active learning: any value for classification of remotely sensed data? Proc. IEEE 100 (3) (2013) 593–608.

[20] W. Di, M.M. Crawford, Active learning via multi-view and local proximity co-regularization for hyperspectral image classification, IEEE J. Sel. Topics Signal Process. 5 (3) (2011) 618–628.

[21] W. Di, M.M Crawford, View generation for multiview maximum disagreement based active learning for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 50 (5) (2012) 1942–1954.

[22] P. Mitra, B.U. Shankar, S.K. Pal, Segmentation of multispectral remote sensing images using active support vector machines, Pattern Recognit. Lett. 25 (9) (2004) 1067–1074.

[23] D. Tuia, E. Pasolli, W.J. Emery, Using active learning to adapt remote sensing image classifiers, Remote Sens. Environ. 115 (9) (2011) 2232–2242.

[24] I. Dopido, J. Li, P.R. Marpu, A. Plaza, J.M. Bioucas-Dias, J.A. Benediktsson, Semisupervised self-learning for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 51 (7) (2013) 4032–4044.

[25] X. Lu, J. Zhang, T. Li, Y. Zhang, Incorporating diversity into self-learning for synergetic classification of hyperspectral and panchromatic images, Remote Sens. 8 (10) (2016) 804.

[26] A. Blum, T. Mitchell, Combining labeled and unlabeled data with co-training, in: Proceedings of the 11th Annual Conference on Computational Learning Theory, 1998, pp. 92–100.

[27] Q. Jackson, D. Landgrebe, An adaptive method for combined covariance estimation and classification, IEEE Trans. Geosci. Remote Sens. 40 (5) (2002) 1082–1087.

[28] M. Chi, L. Bruzzone, Semisupervised classification of hyperspectral images by SVMs optimized in the primal, IEEE Trans. Geosci. Remote Sens. 45 (6) (2007) 1870–1880.

[29] L. Gomez-Chova, G. Camps-Valls, J. Munoz-Mari, J. Calpe, Semisupervised image classification with Laplacian support vector machines, IEEE Geosci. Remote Sens. Lett. 5 (3) (2008) 336–340.

[30] D. Tuia, J. Munoz-Mari, G. Camps-Valls, Large scale semi-supervised image segmentation with active queries, in: Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, 2011, pp. 2653–2656.

[31] J. Munoz-Mari, D. Tuia, G. Camps-Valls, Semisupervised classification of remote sensing images with active queries, IEEE Trans. Geosci. Remote Sens. 50 (10) (2011) 3751–3763.

[32] L. Bruzzone, M. Chi, M. Marconcini, A novel transductive SVM for semisupervised classification of remote-sensing images, IEEE Trans. Geosci. Remote Sens. 44 (11) (2006) 3363–3373.

[33] W. Di, M.M. Crawford, Active learning via multi-view and local proximity co-regularization for hyperspectral image classification, IEEE J. Sel. Topics Signal Process. 5 (3) (2011) 618–628.

[34] L. Wan, K. Tang, M. Li, Y. Zhong, A.K. Qin, Collaborative active and semisupervised learning for hyperspectral remote sensing image classification, IEEE Trans. Geosci. Remote Sens. 53 (5) (2015) 2384–2396.

[35] B. Sun, X. Kang, S. Li, J.A. Benediktsson, Random-walker-based collaborative learning for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 55 (1) (2017) 212–222.

[36] B. Demir, C. Persello, L. Bruzzone, Batch-mode active-learning methods for the interactive classification of remote sensing images, IEEE Trans. Geosci. Remote Sens. 49 (3) (2011) 1014–1031.

[37] K. Brinker, Incorporating diversity in active learning with support vector machines, in: Proceedings of the 20th International Conference on Machine Learning, 2003, pp. 59–66.

[38] Z. Xu, K. Yu, V. Tresp, X. Xu, J. Wang, Representative sampling for text classification using support vector machines, in: Proceedings of the 25th European Conference on Information Retrieval Research, 2003, pp. 393–407.

[39] J. Li, J.M. Bioucas-Dias, A. Plaza, Hyperspectral image segmentation using a new Bayesian approach with active learning, IEEE Trans. Geosci. Remote Sens. 49 (10) (2011) 3947–3960.

[40] D. Tuia, J. Muñoz-Marí, G. Camps-Valls, Remote sensing image segmentation by active queries, Pattern Recognit. 45 (6) (2012) 2180–2192.

[41] M. Belkin, P. Niyogi, Semi-supervised learning on riemannian manifolds, Mach. Learn. 56 (1) (2004) 209–239.

[42] Z. Zhang, E. Pasolli, M.M. Crawford, J.C. Tilton, An active learning framework for hyperspectral image classification using hierarchical segmentation, IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens. 9 (2) (2016) 640–654.

[43] Z. Wang, B. Du, L. Zhang, L. Zhang, A batch-mode active learning framework by querying discriminative and representative samples for hyperspectral image classification, Neurocomputing 179 (C) (2016) 88–100.

[44] D. Tuia, E. Pasolli, W.J. Emery, Using active learning to adapt remote sensing image classifiers, Remote Sens. Environ. 115 (9) (2011) 2232–2242.

[45] N. Roy, A. McCallum, Toward optimal active learning through sampling estimation of error reduction, in: Proceedings of the 20th International Conference on Machine Learning, 2001, pp. 441–448.

[46] S. Rajan, J. Ghosh, M.M. Crawford, An active learning approach to hyperspectral data classification, IEEE Trans. Geosci. Remote Sens. 46 (4) (2008) 1231–1242.

[47] S. Wang, H. Lu, F. Yang, M.H. Yang, Superpixel tracking, in: Proceedings of the 13th IEEE International, Conference on Computer Vision, 2011, pp. 1323–1330.

[48] M.Y. Liu, O. Tuzel, S. Ramalingam, R. Chellappa, Entropy-rate clustering analysis via maximizing a submodular function subject to matroid constraint, IEEE Trans. Pattern Anal. Mach. Intell. 36 (1) (2014) 99–112.

[49] C. Peng, X. Gao, N. Wang, J. Li, Superpixel-based face sketch-photo synthesis, IEEE Trans. Circuits Syst. Video Technol. 27 (2) (2017) 288–299.

[50] L. Fang, S. Li, X. Kang, J.A. Benediktsson, Spectral-spatial classification of hyperspectral images with a superpixel-based discriminative sparse model, IEEE Trans. Geosci. Remote Sens. 53 (8) (2015) 4186–4201.

[51] J. Li, H. Zhang, L. Zhang, Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 53 (10) (2015) 1–14.

[52] L. Fang, S. Li, W. Duan, J. Ren, J.A. Benediktsson, Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels, IEEE Trans. Geosci. Remote Sens. 53 (12) (2015) 6663–6674.

[53] T. Lu, S. Li, L. Fang, L. Bruzzone, J.A. Benediktsson, Set-to-set distance-based spectral-spatial classification of hyperspectral images, IEEE Trans. Geosci. Remote Sens. 54 (12) (2016) 7122–7134.

[54] S. Li, T. Lu, L. Fang, X. Jia, J.A. Benediktsson, Probabilistic fusion of pixel-level and superpixel-level hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 54 (12) (2016) 7416–7430.

[55] J. Guo, X. Zhou, J. Li, A. Plaza, S. Prasad, Superpixel-based active learning and online feature importance learning for hyperspectral image analysis, IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens. 10 (1) (2017) 347–359.

[56] J.M. Bioucas-Dias, M. Figueiredo, Logistic Regression Via Variable Splitting and Augmented Lagrangian Tools, Instituto Superior Técnico, TULisbon, 2009.

[57] B. Settles, Active Learning Literature Survey, 52, University of Wisconsin, Madison, USA, 2010, pp. 127–131.

[58] X. Zhu, Semi-supervised Learning Literature Survey, University of Wisconsin, Madison, USA, 2007.

[59] C. Persello, Interactive domain adaptation for the classification of remote sensing images using active learning, IEEE Geosci. Remote Sens. Lett. 10 (4) (2013) 736–740.

[60] M. Li, H. Zhang, R. Wu, Z.H. Zhou, Sample-based software defect prediction with active and semi-supervised learning, Autom. Softw. Eng. 19 (2) (2012) 201–230.

[61] S.A. Goldman, Y. Zhou, Enhancing supervised learning with unlabeled data, in: Proceedings of the 17th International Conference on Machine Learning, 2000, pp. 327–334.

[62] L Grady, Random walks for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 28 (1) (2006) 1768–1783.

[63] X. Kang, S. Li, L. Fang, M. Li, J.A. Benediktsson, Extended random walker-based classification of hyperspectral images, IEEE Trans. Geosci. Remote Sens. 53 (1) (2015) 144–153.

[64] Y. Tarabalka, J. Chanussot, J.A. Benediktsson, Segmentation and classification of hyperspectral images using watershed transformation, Pattern Recognit. 43 (7) (2010) 2367–2379.

[65] K. Bernard, Y. Tarabalka, J. Angulo, J. Chanussot, J.A. Benediktsson, Spectral-spatial classification of hyperspectral data based on a stochastic minimum spanning forest approach, IEEE Trans. Image Process. 21 (4) (2012) 2008–2021.

[66] I.T. Jolliffe, Principal Component Analysis, Wiley, New York, NY, USA, 2005.

[67] S. Li, X. Jia, B. Zhang, Superpixel-based Markov random field for classification of hyperspectral images, in: Proceedings of the 2013 IEEE International Geoscience and Remote Sensing Symposium, 2013, pp. 3491–3494.

[68] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans. Pattern Anal. Mach. Intell. 23 (11) (2011) 1222–1239.

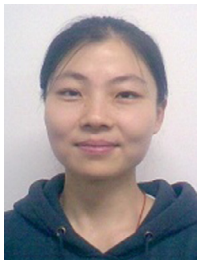
[69] L. Bruzzone, M. Marconcini, Domain adaptation problems: a DASVM classification technique and a circular validation strategy, IEEE Trans. Pattern Anal. Mach. Intell. 32 (5) (2010) 770–787.

[70] J. Li, J.M. Bioucas-Dias, A. Plaza, Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning, IEEE Trans. Geosci. Remote Sens 48 (11) (2010) 4085–4098.

[71] J. Li, J. M.Bioucas-Dias, A. Plaza, Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning, IEEE Trans. Geosci. Remote Sens. 51 (2) (2013) 844–856.

Chao Pan received the B.Sc. degree in software engineering and the M.Sc. degree in computer application technology from Xidian University, Xi'an, China, in 2009 and 2012, respectively. He is currently working toward the Ph.D. degree in intelligent information processing with the school of Electronic Engineering, Xidian University. His research interests include hyperspectral image analysis, image segmentation and signal processing.

Jie Li received the B.Sc. degree in electronic engineering, the M.Sc. degree in signal and information processing, and the Ph.D. degree in circuit and systems from Xidian University, Xi'an, China, in 1995, 1998, and 2004, respectively. She is currently a Professor with the School of Electronic Engineering, Xidian University. She is the author of around 50 technical articles in refereed journals and proceedings, including the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and *Information Sciences*. Her research interests include image processing and machine learning.



Ying Wang received the B.Sc., M.Sc., and Doctoral degrees in signal and information processing from Xidian University, Xi'an, China, in 2003, 2006, and 2010, respectively. She is currently an Associate Professor of signal and information processing with Xidian University. Her research interests include medical image analysis, pattern recognition, and computer-aided diagnosis.

Xinbo Gao (M'02-SM'07) received the B.Eng., M.Sc., and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively. From 1997 to 1998, he was a Research Fellow at the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Postdoctoral Research Fellow at the Department of Information Engineering, the Chinese University of Hong Kong, Hong Kong. Since 2001, he has been at the School of Electronic Engineering, Xidian University. He is currently a Cheung Kong Professor of Ministry of Education, a Professor of Pattern Recognition and Intelligent System, and the Director of the State Key Laboratory of Integrated Services Networks, Xi'an, China. His current research interests include multimedia analysis, computer vision, pattern recognition, machine learning, and wireless communications. He has published six books and around 200 technical articles in refereed journals and proceedings. Prof. Gao is on the Editorial Boards of several journals, including *Signal Processing* (Elsevier) and *Neurocomputing* (Elsevier). He served as the General Chair/Co-Chair, Program Committee Chair/Co-Chair, or PC Member for around 30 major international conferences. He is a Fellow of the Institute of Engineering and Technology and a Fellow of the Chinese Institute of Electronics.