

Superpixel-Based Semisupervised Active Learning for Hyperspectral Image Classification

Chenying Liu , Student Member, IEEE, Jun Li , Senior Member, IEEE, and Lin He , Member, IEEE

Abstract—In this work, we propose a new semisupervised active learning approach for hyperspectral image classification. The proposed method aims at improving machine generalization by using pseudolabeled samples, both confident and informative, which are automatically and actively selected, via semisupervised learning. The learning is performed under two assumptions: a local one for the labeling via a superpixel-based constraint dedicated to the spatial homogeneity and adaptivity into the pseudolabels, and a global one modeling the data density by a multinomial logistic regressor with a Markov random field regularizer. Furthermore, we propose a density-peak-based augmentation strategy for pseudolabels, due to the fact that the samples without manual labels in their superpixel neighborhoods are out of reach for the automatic sampling. Three real hyperspectral datasets were used in our experiments to evaluate the effectiveness of the proposed superpixel-based semisupervised learning approach. The obtained results indicate that the proposed approach can greatly improve the potential for semisupervised learning in hyperspectral image classification.

Index Terms—Active learning, clustering, hyperspectral image classification, semisupervised learning, superpixel.

I. INTRODUCTION

OVER the past several decades, hyperspectral imaging (also called imaging spectrometry [1]) has experienced a tremendous interest in the remote sensing community, due to its capability to achieve more detailed spectral measures than 3-D pictorial data [2]. Specifically, each pixel in a hyperspectral image is characterized by a wealth of spectral information as it comprises hundreds of narrow spectral bands, which opens a wide range of possibilities for distinguishing objects that exhibit similar spectral signatures in other kinds of data (e.g.,

multispectral images), thus offering great potential for classification. While such high data dimensionality brings about abundant spectral information, it incurs a problem that a large number of labeled samples are required for supervised classification (where the classifier is trained with some *a priori* label information) in order to avoid the curse of dimensionality, i.e., the Hughes phenomenon [3], [4]. Recently, the multinomial logistic regression (MLR) algorithm has been proven to be effective in solving this problem to some extent [5], wherein the demand for training samples is still critical. More recently, other information sources such as GIS data have been used to automatically label training samples, as suggested in [6]. However, such label annotations based on experienced human labor is generally difficult, expensive, and time-consuming. It is noteworthy that unlabeled samples are much easier and cheaper to obtain [7]. This has fostered the idea of using semisupervised learning techniques, in which the classifier utilizes unlabeled information (hereinafter, we refer to such information as *pseudolabels*), derived from a limited labeled training set to improve machine generalization ability [8].

As a very active research topic, semisupervised learning has attracted great attention for remote sensing image interpretation [9]–[14]. For instance, Li *et al.* [15] employed soft labels to exploit unlabeled information in mixed pixels for hyperspectral image classification. Zhang *et al.* [10] utilized different criteria associated with different classes for unlabeled sampling, where the classes of interest were separated into good, medium, or poor groups in accordance with training results. In [16], the task-driven dictionary learning was conducted in a semisupervised fashion, thus making unlabeled samples contribute to both minimizing the reconstruction loss and promoting the discrimination ability. In general, these approaches iteratively learn the classifier jointly with the truly labeled and pseudolabeled samples. Since most of the samples in a hyperspectral image are unlabeled, and due to the fact that it is impossible to use all the unlabeled samples in the image (as the complexity is far beyond the reach of most machines), a first question for a semisupervised approach comes as “which samples should be used?” In other words, semisupervised approaches generally only exploit part of the unlabeled samples, and determining “which ones” are better is one of the major concerns in a semisupervised algorithm.

Therefore, most of the semisupervised methods focus on finding a limited number of pseudolabeled samples with high confidence and informativeness [17]. On the one hand, to secure the confidence of the selected pseudolabeled samples, there

Manuscript received May 21, 2018; revised September 6, 2018; accepted October 27, 2018. Date of publication December 10, 2018; date of current version January 21, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61771496 and Grant 61571195, in part by the Guangdong Provincial Natural Science Foundation under Grant 2016A030313254, Grant 2016A030313516, and Grant 2017A030313382, and in part by the National Key Research and Development Program of China under Grant 2017YFB0502900. (*Corresponding authors:* Jun Li and Lin He.)

C. Liu is with the Guangdong Provincial Key Laboratory of Urbanization and Geo-Simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China (e-mail: sysuluchy@163.com).

Jun Li is with the Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province, College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: lijun8206@hnu.edu.cn).

L. He is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: helin@scut.edu.cn).

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

Digital Object Identifier 10.1109/JSTARS.2018.2880562

are some general spatial assumptions in the aforementioned semisupervised methods. A widely used assumption is the local similarity, which was originally introduced via the local spatial neighborhood system, i.e., assuming that neighboring pixels share the same class labels [18]. This is reasonable since in natural scenarios, two neighboring pixels likely belong to the same class. In [19], majority voting-based sampling was conducted among the unlabeled pixels, which are adjacent to the labeled ones. In [20], semisupervised sampling was implemented by averaging regular windows surrounding each pixel. Furthermore, Ma *et al.* presented a sparse-graph-based semisupervised method by inferring the weights between two points with local manifold learning to capture the geometric properties of local neighborhoods [21]. Under the local similarity spatial assumption, which usually comes in the form of a small regular window, these methods exploiting unlabeled information have shown to be effective in the task of generating pseudolabels for semisupervised classification of hyperspectral images.

On the other hand, to further constrain the size of pseudolabels, considering that the sample size from the neighboring system can increase dramatically, a recent trend is to perform active learning to select the most informative ones. Active learning was originally introduced for supervised learning [22]–[25], where the classifier actively selects the most informative unlabeled samples, i.e., the queries, which are formulated according to certain sampling criteria, and then requires experienced oracles to annotate them [26]–[28]. For example, in [29], the breaking ties (BT) sampling criterion is used along with a subspace MLR algorithm to select the most informative samples. In [30], multiclass-level uncertainty (MCLU) and BT are adopted on support vector machines (SVMs) and sparse MLR, respectively, to select the samples with high uncertainty. As aforementioned, under the supervised framework, these methods require human experts to label the newly selected samples. Nevertheless, the concept of active learning is based on minimizing the training size, which exactly matches the natural demand of semisupervised learning. Therefore, the integration of semisupervised learning and active learning has been quite attractive for hyperspectral classification. In [5], unlabeled samples were actively selected through a maximum-entropy-based criterion. In [31], an MCLU criterion was introduced for unlabeled sampling. In [32], active learning was adopted in a self-learning context for synergistic classification of hyperspectral and panchromatic images. In [33], an active-learning-based heuristics was utilized for the selection of unlabeled samples.

More recently, the BT criterion was used in combination with MLR to actively select pseudolabeled samples among the four-connected neighboring pixels [34]. In [35], a similar strategy to [34] was implemented for pseudosampling. Tan *et al.* [36] used both rough classification results and spatial information to actively label unlabeled samples. In [37], the pseudolabeled samples were selected from the most uncertain ones according to the spatial assumption. This new tendency, which integrates the local spatial assumption and active learning into semisupervised learning—forming a new framework called semisupervised active learning (SSAL)—is able to find confident and informative unlabeled samples, exhibiting great potential for

boosting the learning performance [34]. As mentioned before, the local assumption secures the confidence of the pseudolabels, but, however, limits the capacity of SSAL due to the fact that it generally performs over a regular and unified window, which conflicts with the spatial nature of real scenarios, that is, the fact that spatial structures are generally adaptive [38].

Therefore, to tackle the inconsistency between the use of a window-based neighborhood system and the irregular spatial structures found in real images, in this paper, we propose a superpixel-based SSAL approach for hyperspectral image classification. Similar to other neighborhood-based semisupervised methods, the proposed approach aims at searching for the most confident and informative unlabeled samples under a spatial assumption and active learning, which involves two parts. First, a candidate set of unlabeled samples with high confidence is generated by considering the superpixel-based similarity assumption. Then, a BT-based active learning strategy is performed on the obtained candidates to pick out the most informative ones. Furthermore, we introduce an augmentation strategy that relies on the use of a density peak (DP) clustering algorithm to assist the pseudolabeling in the search for unlabeled superpixels, in which all pixels are unlabeled. Finally, the classification is performed via an MLR classifier along with a Markov random field (MRF) regularizer to incorporate the spatial context information [23], [39]. The main innovative contributions of this work can be summarized as follows.

- 1) First and foremost, we introduce a new superpixel-guided SSAL technique, specifically designed for hyperspectral image classification with very limited training samples.
- 2) Second, in the newly developed technique, the local similarity criterion makes use of superpixels, therefore introducing spatial adaptivity into the pseudolabeling stage.
- 3) Third, in order to assign pseudolabels to unlabeled superpixels, a DP augmentation strategy is also introduced, which is based on the intrinsic structural information in superpixels.

The remainder of this paper is organized as follows. Section II introduces the general framework of our SSAL method. Then, we present our newly proposed superpixel-based SSAL method with DP augmentation in detail in Section III. Next, three real hyperspectral datasets, i.e., the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) Indian Pines dataset, the Reflective Optics Imaging Spectrometer System (ROSIS) Pavia University dataset, and the Operational Modular Imaging Spectrometer (OMIS) Zaoyuan scene, were used for extensive experimental evaluation in Section IV. Finally, Section V concludes this paper with some remarks and hints at plausible future research lines.

II. SEMISUPERVISED ACTIVE LEARNING

Let $\mathbf{x} \equiv \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathbb{R}^{d \times n}$ be a hyperspectral image, where n is the number of pixels and d is the number of bands; let $\mathbf{y} \equiv \{y_1, \dots, y_n\} \in \mathbb{R}^n$ be an image of labels, of which each element belongs to one of the K predefined labels, i.e., $y_i \in \mathcal{K} \equiv \{1, \dots, K\}$. With the aforementioned definitions in mind, in this paper, we model the class densities using the MLR

algorithm as follows:

$$p(y_i = k | \mathbf{x}_i, \boldsymbol{\omega}) = \frac{\exp(\boldsymbol{\omega}^{(k)^T} \mathbf{h}(\mathbf{x}_i))}{\sum_{c=1}^K \exp(\boldsymbol{\omega}^{(c)^T} \mathbf{h}(\mathbf{x}_i))} \quad (1)$$

where $\boldsymbol{\omega} \equiv [\boldsymbol{\omega}^{(1)}, \dots, \boldsymbol{\omega}^{(K-1)}]$ is the logistic regressor set with $\boldsymbol{\omega}^{(K)} = \mathbf{0}$, since the density in (1) does not depend on translations on the regressors $\boldsymbol{\omega}^{(k)}$; $\mathbf{h}(\mathbf{x}_i) \equiv [h_1(\mathbf{x}_i), \dots, h_l(\mathbf{x}_i)]^T$ is a feature vector composed of l fixed functions of the input, which can be either linear or nonlinear. In this paper, we use a nonlinear kernel function, Gaussian radial basis function (RBF), in the form of $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)$, where σ denotes the scale parameter of the RBF kernel, to build the transformed space. The RBF kernel has been a widely used solution to hyperspectral image classification problems [40].

A. Supervised Learning

Suppose $\mathcal{D}_L \equiv \{(y_1, \mathbf{x}_1), \dots, (y_l, \mathbf{x}_l)\}$ is a labeled set with l samples. We can formulate the supervised learning of $\boldsymbol{\omega}$ as follows:

$$\hat{\boldsymbol{\omega}} = \arg \max_{\boldsymbol{\omega}} \ell(\boldsymbol{\omega} | \mathcal{D}_L) p(\boldsymbol{\omega}) \quad (2)$$

where $p(\boldsymbol{\omega})$ is *a priori* distribution of $\boldsymbol{\omega}$. Following [41], we adopt $p(\boldsymbol{\omega}) \propto \exp(-\lambda \|\boldsymbol{\omega}\|_1)$, a Laplacian *priori* controlling the sparsity, with λ being the regularization parameter.

In the literature, there are many algorithms that solve the optimization problem in (2), achieving very good performance in hyperspectral classification [5], [41], [42]. One of the most remarkable algorithms is the MLR via variable splitting and augmented Lagrangian algorithm (LORSAL) [43], which is shown to be effective and efficient, especially for the problem with large quantities of training samples. Following [23] and [43], in this work, we use the LORSAL algorithm to learn the class densities.

B. Semisupervised Learning

In semisupervised learning, the pseudolabeled samples, i.e., the unlabeled samples with predicted labels, are used to infer the regressors together with the labeled ones. Let $\mathcal{D}_U \equiv \{(\hat{y}_1, \mathbf{x}_1), \dots, (\hat{y}_u, \mathbf{x}_u)\}$ be an unlabeled set with u pseudolabeled samples; we can learn $\boldsymbol{\omega}$ as follows:

$$\hat{\boldsymbol{\omega}} = \arg \max_{\boldsymbol{\omega}} \ell(\boldsymbol{\omega} | \mathcal{D}_L \cup \mathcal{D}_U) p(\boldsymbol{\omega}). \quad (3)$$

As shown in (3), the computation burden depends on the total number of training samples, i.e., $l + u$. It is intractable to adopt all the unlabeled samples for training, under which the computation burden is usually beyond the reach of any optimization algorithm. In practice, only a small part of the unlabeled samples could be used, i.e., $u \ll n$. To effectively select the u unlabeled samples, two criteria (namely, the confidence and the informativeness) are jointly performed.

1) *Confidence Criterion*: The confidence is generally guaranteed by some spatial assumptions in the duration of candidate collecting for pseudolabeled samples. A widely used one is the local similarity, assuming that neighboring pixels share the same class labels.

Algorithm 1: Semisupervised Active Learning.

Input: $\mathbf{x}, \mathcal{D}_T = \mathcal{D}_L$
1: **repeat**
2: $\hat{\mathbf{p}} := \text{MLR}(\mathcal{D}_T, \mathbf{x})$
3: $\mathcal{D}_C := \text{Confidence}(\mathcal{D}_T, \hat{\mathbf{p}}, \text{Local Similarity})$
4: $\hat{\mathbf{x}}' := \text{Informativeness}(\mathcal{D}_C, \hat{\mathbf{p}}, BT)$
5: $\mathcal{D}_T := [\mathcal{D}_T, (\hat{y}', \hat{\mathbf{x}}')]$
6: **until** some stopping criterion is satisfied

2) *Informativeness Criterion*: To reduce the computation burden of training, we perform active learning over the confident candidate set to select the most informative samples. For the MLR classifier, the BT active sampling strategy is effective and easy to implement, which is formulated as

$$\hat{\mathbf{x}}_i^{BT} = \arg \min_{\mathbf{x}_i \in \mathbf{X}_U} \left[\max_{k \in \mathcal{K}} p(y_i = k | \mathbf{x}_i, \boldsymbol{\omega}) - \max_{k \in \mathcal{K} \setminus \{k^+\}} p(y_i = k | \mathbf{x}_i, \boldsymbol{\omega}) \right] \quad (4)$$

where $k^+ = \arg \max_{k \in \mathcal{K}} p(y_i = k | \mathbf{x}_i, \boldsymbol{\omega})$ is the class label corresponding to the largest posterior probability for \mathbf{x}_i , and $\mathcal{K} \setminus \{k^+\}$ is the class labels of interest excluding k^+ . From (4), it can be observed that BT favors samples located near the class boundaries, which are considered to be more informative [28].

Finally, we conclude the general SSAL framework in Algorithm 1, where \mathcal{D}_C denotes the candidate set obtained by the Confidence criterion; $\mathcal{D}_T = \mathcal{D}_L \cup \mathcal{D}_U$ is the union of labeled and unlabeled training sets with $\hat{\mathbf{x}}'$ being the newly generated unlabeled training samples under the Informativeness criterion such as BT. The stopping criterion can be set in many forms, such as a given number of iterations, the total number of pseudolabeled samples, the convergence of the classification performance, etc.

III. PROPOSED METHOD

In this section, we present the proposed superpixel-based SSAL method. As shown in Fig. 1, the proposed approach automatically generates unlabeled training samples via the DP-based augmentation, which has three main steps. First, we implement the mode augmentation to pseudolabel superpixels without labels and then collect candidates. Second, the BT scores are calculated as an informativeness indicator for each sample. Finally, the new pseudolabeled samples are derived from the candidate set using their BT scores to retrain the classifier.

The remainder of this section is organized as follows. Prior to the description of DP augmentation, in Section III-A, we first introduce the superpixel-based spatial assumption aiming to replace the local similarity assumption as a confidence criterion. We also revisit the simple linear iterative clustering (SLIC) method for practical purposes in this section. Then, in Section III-B, the mode augmentation procedure is presented in detail. In Section III-C, we describe the informativeness indicator, i.e., the BT score. Next, the pseudolabeling is presented in

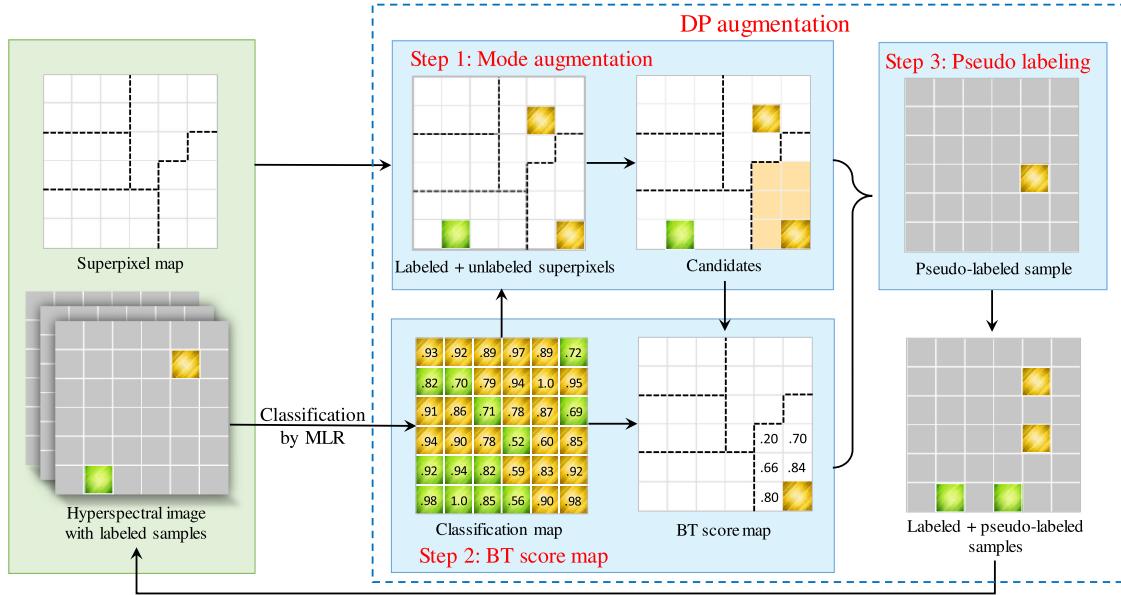


Fig. 1. Flowchart of the proposed superpixel-based SSAL method under a two-class circumstance, which has three main steps. First, the mode augmentation is implemented for the pseudolabeling of unlabeled superpixels and the generation of candidates. Then, the BT scores of candidates are calculated as an informativeness indicator. Finally, the pseudolabeled samples are obtained based on their BT scores.

Section III-D to select pseudolabeled training samples. Finally, the MLR algorithm with an MRF regularizer, i.e., MLR-MRF, is briefly described in Section III-E.

In order to integrate the spatial-contextual information into the pixelwise results, in this work, we adopt a widely used MRF regularizer to obtain the final classification [39], [51]. The MRF scheme exploits the fact that adjacent pixels are likely to have the same class in a scene, thus promoting performance with a piecewise smooth. This model, namely MLR-MRF, has obtained great success in hyperspectral image classification [39], [52].

A. Superpixel Assumption

We utilize the superpixel-based spatial assumption to serve as the confidence criterion in our SSAL method. For a given (truly or pseudo-) labeled sample, the neighborhood is defined as the superpixel it belongs to, based on the fact that superpixels are homogeneous regions of a segmented image [44]. Therefore, the pixels within a superpixel are supposed to share the same label under the local (spatial) assumption. Furthermore, from a global point of view, similar features or spectral signatures generally associate with the same class. In other words, the neighboring samples whose predicted labels are consistent with the local assumptions are more confident than the others. Hence, the automatically labeled samples can still be effective even though a human supervisor is absent.

Among various superpixel algorithms, the SLIC method is one of the most remarkable due to its capability to generate high-quality, compact, and nearly uniform superpixels [45]. Inspired by [46], Xu *et al.* [47] have recently developed a hyperspectral SLIC method, which is concisely akin to a classic k -means algorithm [48]. Initially, the image is evenly split into a required number of rectangular subregions with geometric centers being

cluster centers. Afterwards, the cluster centers are repeatedly updated until some stopping criterion, for instance, a given number of iterations, is satisfied, according to a spectral–spatial distance formulated as follows:

$$D = (1 - \lambda) \times SD + \lambda \times ED \quad (5)$$

where SD and ED are spectral information divergence (SID) distance [49] and Euclidean distance defined in the spectral and spatial spaces, respectively, with λ being a tradeoff parameter. After the main loop, the isolated pixels, i.e., those in the vicinity of a larger cluster having the same label but not connected to it [46], are divided into their largest neighboring clusters for better compactness.

At this point, we introduce two new definitions, i.e., labeled and unlabeled superpixels, where the *labeled superpixels* refer to the ones with at least one truly labeled pixel, while the *unlabeled superpixels* refer to the ones in which all pixels located inside are actually unlabeled. Since the labeled training set is very small, most of the superpixels are unlabeled. Due to the fact that the sample augmentation strongly depends on the available labeled samples, it is expected that the pseudolabeled samples may be trapped in the labeled superpixels.

B. Mode Augmentation

In this work, we introduce a new clustering-based augmentation strategy to deal with unlabeled superpixels.

Since superpixels can be regarded as a series of clusters with spatial constraints [44], clustering is first implemented within unlabeled superpixels. The cluster centers, termed *modes* hereinafter, are locally the most representative ones from an unsupervised perspective. In this regard, we augment unlabeled superpixels in terms of their modes, thus naming this strategy as mode augmentation. In practice, we choose the DP clustering

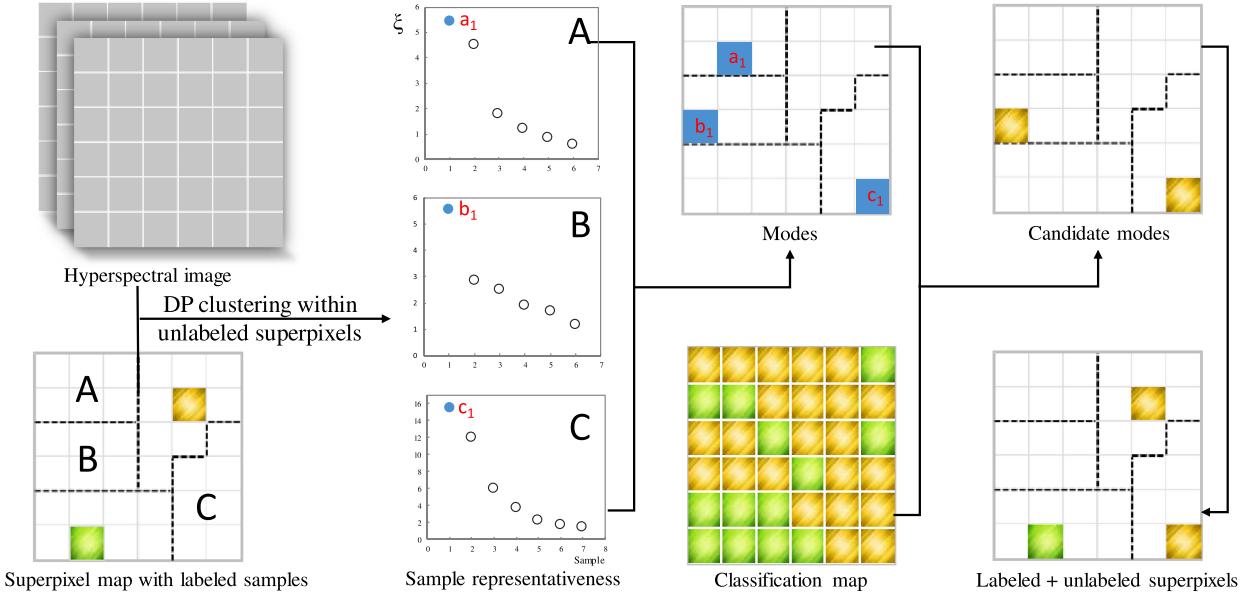


Fig. 2. Graphic illustration for pseudolabeling of unlabeled superpixels in the mode augmentation, which comprises four steps. First, DP clustering within superpixels is implemented to estimate the sample representativeness in (6). Then, the modes are generated according to (9). Later, the candidate modes are selected based on the criterion defined in (10). Finally, the pseudolabeled modes are recruited according to the location-based criterion of (11).

algorithm [50] for mode generation. The advantages are twofold. First, we do not need to indicate a predetermined cluster number. Second, the algorithm provides a quantitative indicator for representativeness, thus feasibly coping with the presence of mixed superpixels and outliers in the data.

Let ρ denote the local density, and δ denote the minimum distance to any other samples with a higher local density. Given a pixel \mathbf{x}_i , the representativeness indicator ξ is written as

$$\xi_i = \rho_i \delta_i \quad (6)$$

where

$$\rho_i = \sum_j \sigma(\text{SD}_t - \text{SD}_{i,j}) \quad (7)$$

and

$$\delta_i = \min_{j: \rho_j > \rho_i} \text{SD}_{i,j} \quad (8)$$

with $\text{SD}_{i,j}$ being the SID distance between \mathbf{x}_i and \mathbf{x}_j , SD_t being a cutoff distance, i.e., a given threshold value, and $\sigma(\cdot)$ being a quasi-signal function in which $\sigma(x) = 1$ if $x > 0$, otherwise $\sigma(x) = 0$. Especially, we set $\delta_i = \max_j (\text{SD}_{i,j})$ if \mathbf{x}_i has the largest local density. Intuitively, within a superpixel $\mathbf{S} \equiv \{\mathbf{x}_1, \dots, \mathbf{x}_s\}$, the mode \mathbf{x}_i^* is identified as

$$\mathbf{x}_i^* = \arg \max_{\mathbf{x}_i \in \mathbf{S}} \xi_i. \quad (9)$$

On the other hand, taking the homogeneity of superpixels into consideration, the modes are simultaneously consistent with the global assumption mentioned above. In this case, \mathbf{x}_i^* serves as a candidate only when it satisfies

$$\hat{y}_i^* = \arg \max_{k \in \mathcal{K}} \sum_{j: \mathbf{x}_j \in \mathbf{S}} \sigma(\hat{y}_j - k) \quad (10)$$

where $\hat{y}_i^* = \max_{k \in \mathcal{K}} p(y_i^* = k | \mathbf{x}_i^*, \boldsymbol{\omega})$ is the predicted label for \mathbf{x}_i^* . Therefore, the candidate mode set $\mathbf{x}^* \equiv \{\mathbf{x}_1^*, \dots, \mathbf{x}_m^*\}$ is at a

low risk of error despite the absence of prior labeled information. As suggested in [28], diversity sampling is crucial to embed the spectral signature drift to the training set. Based on this observation, we introduce a location-based sampling criterion for unlabeled modes that works as follows:

$$\hat{\mathbf{x}}_i^* = \arg \max_{\mathbf{x}_i^* \in \mathbf{x}^*} \min_{\mathbf{x}_j^* \in \mathbf{x}_{(l)}^*} \text{ED}_{i,j} \quad (11)$$

where $\mathbf{x}_{(l)}^*$ is the mode set of the labeled and unlabeled superpixels. As in (11), the modes far away from the labeled ones are favored for selection. In this work, five unlabeled superpixels were augmented in each iteration in our experiments.

Accordingly, a graphic illustration for pseudolabeling of unlabeled superpixels is given in Fig. 2. As shown in this figure, first of all, DP clustering within unlabeled superpixels is implemented to estimate the sample representativeness for their associated superpixels. Then, a set of modes are generated based on the representativeness indicator. After that, the candidate modes are selected based on the criterion defined in (10). Finally, the pseudolabeled modes are recruited according to the location-based criterion of (11). Consequently, as illustrated in Fig. 1, both labeled and unlabeled superpixels can be involved in the final candidate collection.

C. BT Scores

In order to identify the most valuable training samples, in our proposed method, we adopt a BT score to quantify the informativeness of each sample, which is defined as

$$\text{BT}_{\text{score}} = \max_{k \in \mathcal{K}} p(y_i = k | \mathbf{x}_i, \boldsymbol{\omega}) - \max_{k \in \mathcal{K} \setminus \{k^+\}} p(y_i = k | \mathbf{x}_i, \boldsymbol{\omega}). \quad (12)$$

In accordance with (4), BT minimizes the BT_{score} in (12). That means, BT favors samples with a lower BT_{score} . In other words, the BT_{score} in (12) can be considered as an indicator

Algorithm 2: SDP.**Input:** $\mathbf{x}, \mathcal{D}_T = \mathcal{D}_L, u$

- 1: **repeat**
- 2: $\hat{\mathbf{p}} := \text{MLR-MRF}(\mathcal{D}_T, \mathbf{x})$
- 3: $\mathcal{D}_M := \text{Augment}(\mathbf{x}, \hat{\mathbf{p}}, DP)$
- 4: $\mathcal{D}_C := \text{Confidence}(\mathcal{D}_T \cup \mathcal{D}_M, \hat{\mathbf{p}}, \text{Superpixel})$
- 5: $\hat{\mathbf{x}}' := \text{Informativeness}(\mathcal{D}_C, \hat{\mathbf{p}}, BT)$
- 6: $\mathcal{D}_T := [\mathcal{D}_T, (\hat{y}', \hat{\mathbf{x}}')]$
- 7: **until** the classification performance is converged

of the informativeness for samples. Particularly, samples with smaller BT_{score} are more informative than the ones with larger BT_{score}.

D. Pseudo Labeling

Using the mode augmentation procedure to collect confident candidates and BT scores to quantify the informativeness of samples, and according to (4), the pseudo sampling is finally given by

$$\hat{\mathbf{x}}'_i = \arg \min_{\mathbf{x}_i \in \mathcal{D}_C} \text{BT}_{\text{score}} \quad (13)$$

where \mathcal{D}_C is the candidate set obtained by the mode augmentation under the superpixel-based assumption. Taken into consideration the computational complexity, we select more than one sample at a time. Hereinafter, we employ u_i to denote the number of pseudolabeled samples that are newly added per iteration. From (13), it can be observed that the selected pseudolabeled samples are both confident and informative, thus providing potential for improving the machine generalization.

E. MLR-MRF Classification

In order to integrate the spatial-contextual information into the pixelwise results, here, we adopt a widely used MRF regularizer to obtain the final classification [39], [51]. The MRF scheme exploits the fact that adjacent pixels are likely to have the same class in a scene, thus improving performance with a piecewise smooth. This model, namely MLR-MRF, has obtained great success in hyperspectral image classification [39], [52].

Finally, to conclude this section, the proposed superpixel with DP augmentation SSAL method, called SDP in short, is summarized in Algorithm 2, where \mathcal{D}_M denotes the unlabeled mode set selected by the Augment strategy, i.e., the mode augmentation based on the DP clustering algorithm. Regarding the stopping criterion, the algorithm stops when the classification performance converges, i.e., when there is no significant difference between the classification maps obtained in two consecutive iterations.

IV. EXPERIMENTAL RESULTS

In this section, three real hyperspectral datasets are used to evaluate the effectiveness of the proposed method. As the main contribution of the proposed approach relies on the superpixel-based similarity assumption, the local neighborhood-based as-

TABLE I
SSAL METHODS CONSIDERED FOR EXPERIMENTS

N	The 4-connected neighborhood based SSAL method
S	The superpixel based SSAL method
SN	N with superpixel assumption based SSAL method
SDP	S with DP augmentation based SSAL method

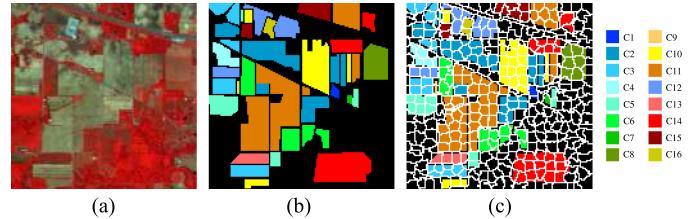


Fig. 3. AVIRIS Indian Pines scene with 16 classes along with the superpixel map. (a) False color map. (b) Ground truth. (c) Superpixel map.

sumption (which demonstrated great success in semisupervised learning [34], [35]) is considered for comparison. Furthermore, the SVM with an MRF regularizer, namely SVM-MRF, is also considered for comparison purposed, while the parameters are set through tenfold cross validation. As listed in Table I, for simplicity, we use N, S, SN, and SDP to denote the considered SSAL methods, where SN is an extension of N, which adopts the superpixel assumption in the beginning and the neighborhood one later.

In this work, we use the overall accuracy (OA), average accuracy (AA), class individual accuracy (CA), and kappa statistic (κ) for quantitative evaluation [2], [53]. In order to guarantee the statistical significance, the results are reported by averaging ten Monte Carlo runs corresponding to independent initial labeled sets.

The overall organization of this section is as follows. Then, the detailed classification results obtained for each of these images: AVIRIS Indian Pines (see Section IV-B), ROSIS Pavia University (see Section IV-C), and OMIS Zaoyuan scene (see Section IV-D) are, respectively, reported.

A. Hyperspectral Datasets

Three hyperspectral datasets collected by different instruments are considered in our experiments.

- 1) The first hyperspectral image used in our experiments is the well-known Indian Pines dataset collected over a mixed agricultural/forest area in North-western Indiana, America, by the AVIRIS sensor in 1992 [54]. This scene is composed of 220 spectral bands with the wavelength varying from 0.4 to 2.5 μm . It contains 145×145 pixels, of which each one has an actual spatial coverage of 20 m \times 20 m. The available ground-truth map contains only 10 366 labeled samples. Fig. 3(a) and (b) shows the false color map and the available 16-class ground-truth map, respectively. The 16 classes marked by C1-C16 in Fig. 3(b) represent Alfalfa, Corn-no till, Corn-min till, Corn, Grass/Pasture, Grass/Trees, Grass/Pasture-mowed, Hay-windrowed, Oats, Soybeans-no till, Soybeans-min

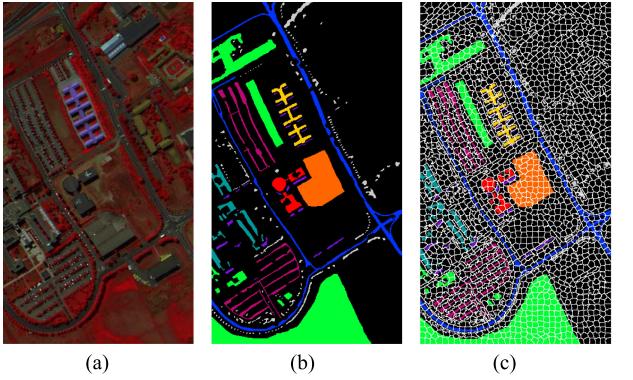


Fig. 4. ROSIS Pavia University scene with nine classes along with the superpixel map. (a) False color. (b) Ground truth. (c) Superpixel map.

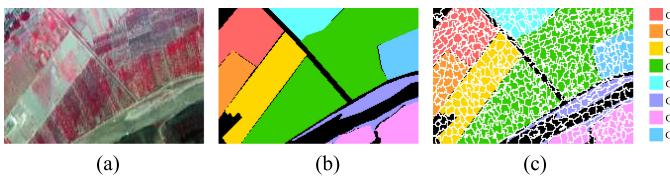


Fig. 5. OMIS Zaoyuan scene with eight classes along with the superpixel map. (a) False color map. (b) Ground-truth map. (c) Superpixel map.

till, Soybeans-clean till, Wheat, Woods, Buildings-Grass-Tree-Drives, and Stone-steel-towers, respectively. The scene is challenging due to the fact that most of its pixels are highly mixed, and the number of samples in each class is quite unbalanced.

- 2) The second dataset is the ROSIS Pavia University scene. This scene [see Fig. 4(a)] is composed of 610 lines and 340 samples, which was collected by the ROSIS sensor in 2003. It contains 103 spectral bands ranging from 0.43 to $0.86 \mu\text{m}$ with a spatial resolution of 1.3 m. The scene is atmospherically corrected and has an available ground-truth map containing 42 776 labeled samples belonging to nine classes of interest. The classes remarked by C1–C9 in Fig. 4(b) represent Asphalt, Meadows, Gravel, Trees, Painted metal sheets, Bare Soil, Bitumen, Self-Blocking Bricks, and Shadows, respectively.
- 3) The last hyperspectral image was collected by the OMIS over Zaoyuan region, China, in 2001. The Zaoyuan scene [see Fig. 5(a)] contains 137×202 pixels and 80 spectral bands with the first 64 spectral bands covering the region of $0.4\text{--}1.1 \mu\text{m}$ and the last 16 within the region of $1.06\text{--}1.7 \mu\text{m}$. The ground-truth map used as a reference for classification contains eight landcover classes. As shown in Fig. 5(b), the classes denoted by C1–C8 represent Vegetable, Grape, Dry vegetable, Pear, Corn, Terrace/Grass, Bush-Lespedeza, and Peach, respectively.

For the SLIC superpixel segmentation on the three considered hyperspectral datasets, the tradeoff parameter λ is set to 0.1, 0.3, and 0.05, respectively, as suggested in [47]. As shown in Figs. 3–5(c), the number of superpixels is set to 400, 500, and 2000, respectively.

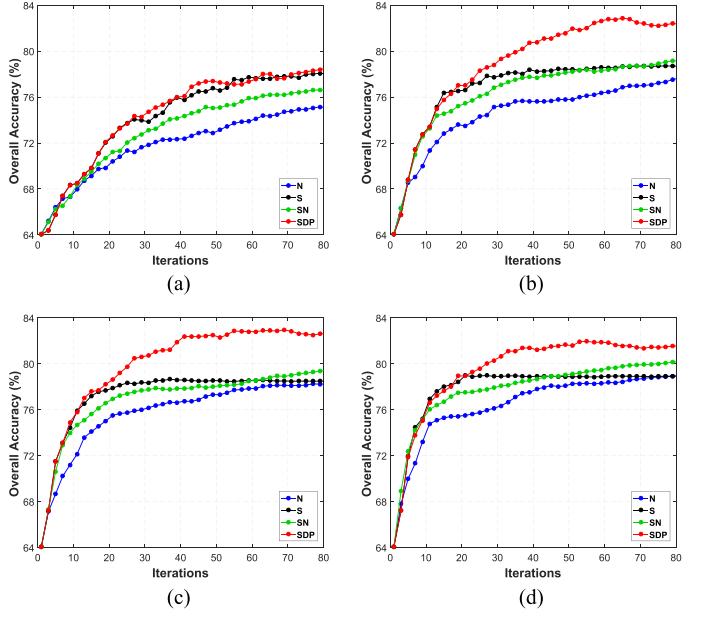


Fig. 6. OAs as a function of the number of pseudolabeled samples obtained using different values of u_i for the AVIRIS Indian Pines scene. (a) $u_i = 5$. (b) $u_i = 15$. (c) $u_i = 25$. (d) $u_i = 35$.

Finally, in order to evaluate the proposed SDP for scenarios with limited labeled information in our different experiments, a small number of labeled samples, i.e., five per class were randomly selected from the ground truth for training, while the remaining ones were used for validation.

B. Experiments With the AVIRIS Indian Pines Scene

For the AVIRIS Indian Pines scene, we designed four experiments to analyze different relevant aspects of the proposed approach.

1) *Impact of u_i :* Recall that u_i is the new pseudolabeled samples included at each iteration. In our first experiment, we investigate the impact of u_i . Fig. 6 reports the obtained OAs with $u_i = 5, 15, 25, 35$ for 80 iterations. It is worth noting that, in all cases, the proposed SDP yielded competitive results with the other ones, while the performances of the proposed SDP are more remarkable with a media size of u_i , i.e., $u_i = 15, 25$. The advantage increases as the number of iterations increases, which indicates that the pseudolabeled samples selected by SDP are more effective. Moreover, all the considered methods can significantly improve the initial supervised classification performances. Another interesting observation is that all the SSAL methods resulted in similar performance with the same number of pseudolabeled samples. For instance, by using around 400 pseudolabeled samples, the proposed SDP obtained an OA around 78% in all cases. Finally, taking into account the classification performance and the computational complexity, we set $u_i = 25$ in the following experiments.

2) *Impact of Pseudolabels:* In the second experiment, we perform a quantitative comparison with an *optimal case* [34], where we effectively label the pseudolabeled samples using their true labels (obtained from the ground-truth map) instead of the

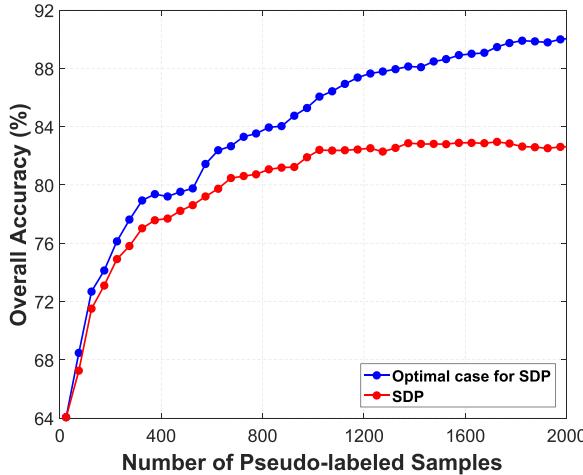


Fig. 7. OAs as a function of the number of pseudolabeled samples for the AVIRIS Indian Pines scene. Recall that u is the total number of pseudolabeled samples. Two cases are displayed: the proposed SDP with u pseudolabels and its optimal case, with the u_l samples being truly labeled.

pseudo ones. Typically, true labels will be only available for part of the samples as the considered hyperspectral datasets do not contain ground-truth information for all pixels. Hereinafter, we employ u_l to denote the number of pseudolabeled samples for which a true label is available in the ground-truth map. Clearly, these u_l true labels will be favored over their pseudo counterparts. In the optimal case, the u_l true labels are used to replace their pseudo counterparts. Using this strategy, we can substantiate the deviation of our proposed approach. For more details about the optimal validation framework, we refer to [34].

Fig. 7 presents the classification results obtained by the proposed SDP as a function of the number of pseudolabeled samples, along with its optimal case. As can be seen in Fig. 7, the proposed method achieved a good result in comparison to the optimal case, with a difference of about 7% in OA when 2000 pseudolabeled samples were included.

3) Informativeness of Pseudolabels: In this experiment, we evaluate the informativeness of pseudolabels in terms of the BT score in (12). Notice that BT scores are computed for all the samples in the considered image. However, for illustrative purposes, we only report the BT scores for the pseudolabeled samples. Fig. 8 demonstrates the BT scores as a function of the number of pseudolabeled samples. A first observation we can conduct from this figure is that the BT scores of the proposed SDP vary significantly along with the pseudolabeled samples, while those of the competitors mandatorily increase as the number of pseudolabeled samples increases. Recall that more informative samples are generally identified by lower BT scores. This reveals the fact that the pseudolabeled samples selected by the proposed SDP are more informative than the others. In Fig. 8, we notice that, in the first iterations, for SDP, the BT score increases when there is no unlabeled superpixel included. This phenomenon is similar to those of the other methods. As far as some unlabeled superpixels are generated via the mode argumentation, the BT score decreases. Later, the local minimum

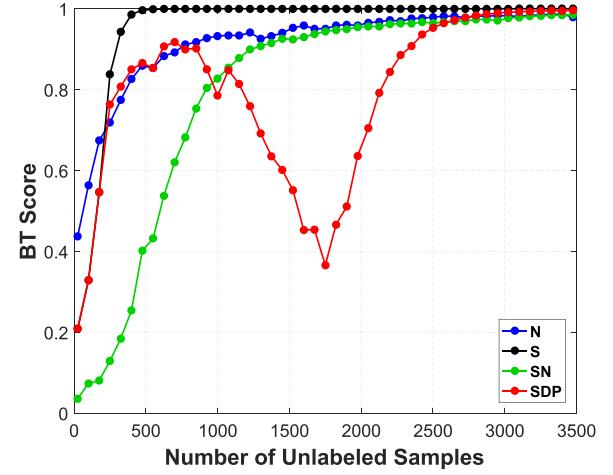


Fig. 8. Maximum BT scores of pseudolabeled samples as a function of the number of pseudolabeled samples for the AVIRIS Indian Pines scene.

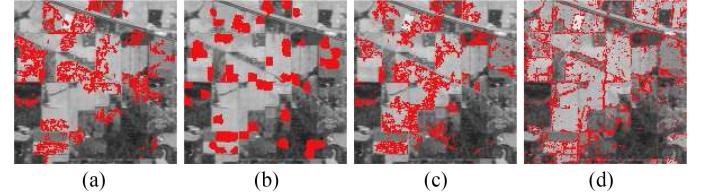


Fig. 9. Training samples at the 140th iteration for the AVIRIS Indian Pines scene. (a) N. (b) S. (c) SN. (d) SDP.

happens when all the potential unlabeled superpixels are included. Based on this observation, we can conduct that the unlabeled superpixels generated via mode argumentation can result in pseudolabeled samples with high informativeness. Specifically, in the beginning, the BT score decreases as the number of unlabeled superpixels increases, which means that the new samples are more informative than the former ones. In the end, when there is no more potential unlabeled superpixel, the BT score increases, which means the informativeness of pseudolabels is decreasing.

Finally, Fig. 9 graphically demonstrates the distribution of training samples with 3500 pseudolabeled samples for SDP, N, and SN. While for S, only around 3000 pseudolabeled samples are finally involved due to the trap by labeled superpixels. It is noticeable that the training samples of SDP are well distributed in the whole images. However, for S (as expected), the training samples are trapped in a few labeled superpixels. For the neighborhood-based sampling methods, N and SN, the distributions of the training samples are subject to labeled samples. This observation indicates that the proposed SDP has great potential for large-scale problems with limited labeled samples located in limited regions, as it is able to introduce pseudosampling in the whole image.

4) Classification Performance: In this set of experiments, we evaluate the classification performance of the proposed method. Fig. 10 presents the obtained OA results as a function of the number of pseudolabeled samples. It can be observed that the proposed SDP obtained very competitive results to the

TABLE II

CAS [%], OAs [%], AAs [%], KAPPA (k) [%] STATISTICS (STANDARD DEVIATIONS IN BRACKETS), AND RUNNING TIME (IN SECONDS) OBTAINED FROM TEN MONTE CARLO RUNS WITH 80 LABELED (FIVE PER CLASS) AND $u=3000$ PSEUDOLABELED SAMPLES FOR THE AVIRIS INDIAN PINES SCENE, WHERE THE BEST RESULTS AMONG SSAL METHODS ARE HIGHLIGHTED IN BOLD

Class	#Test	Supervised		N		S		SN		SDP	
		SVM-MRF	MLR-MRF	$u_l = 0$	$u_l = 2175$	$u_l = 0$	$u_l = 2442$	$u_l = 0$	$u_l = 2104$	$u_l = 0$	$u_l = 998$
C1	49	87.76	85.51	91.02	91.63	92.45	92.65	90.00	90.82	91.84	93.47
C2	1429	38.54	49.90	72.68	81.83	71.48	71.39	74.93	82.30	81.64	91.39
C3	829	23.57	50.59	60.75	70.62	64.45	68.56	64.67	80.51	63.15	81.66
C4	229	78.30	71.83	95.24	96.51	96.42	94.67	96.42	96.77	90.13	95.81
C5	492	66.14	78.17	79.74	91.50	83.05	90.26	79.25	90.75	78.19	85.33
C6	742	77.06	95.11	96.00	98.27	96.43	98.50	95.31	96.64	88.32	97.56
C7	21	94.29	92.86	99.52	99.52	99.52	100.0	98.05	99.52	95.71	98.10
C8	484	66.26	86.76	99.21	99.30	99.71	99.73	99.69	99.71	99.48	99.79
C9	15	90.00	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
C10	963	48.41	57.61	78.63	85.82	77.46	79.70	81.64	90.46	76.18	91.21
C11	2463	45.59	49.50	68.25	76.93	63.76	67.71	72.00	78.83	78.07	92.93
C12	609	29.62	51.22	77.93	89.61	86.12	88.67	80.54	88.60	88.13	94.35
C13	207	98.94	99.61	99.18	99.23	99.66	99.66	99.28	99.32	98.16	99.57
C14	1289	81.18	87.28	86.49	86.46	88.01	88.01	88.05	88.00	90.85	93.09
C15	375	48.08	56.64	93.41	93.97	93.36	93.25	93.79	93.87	92.03	95.95
C16	90	89.44	84.33	92.33	93.78	98.22	98.00	96.44	98.00	98.33	
OA		53.57	64.06	78.67	84.97	78.60	80.67	80.85	86.75	82.66	90.08
		(6.99)	(3.12)	(2.68)	(2.55)	(2.47)	(3.07)	(3.75)	(2.04)	(4.61)	(2.64)
AA		66.47	74.81	86.90	90.94	88.13	89.42	88.19	92.03	88.12	93.66
		(2.67)	(2.11)	(1.46)	(1.39)	(1.54)	(1.75)	(1.69)	(0.98)	(2.28)	(1.70)
k		47.98	59.66	75.92	83.00	75.93	78.24	78.38	85.00	80.27	90.27
Time		2.53	23.95	7.37×10^3		6.11×10^3		7.31×10^3		7.28×10^3	

The optimal cases are also presented along with the values of u_l , i.e., the number of the used true labels.

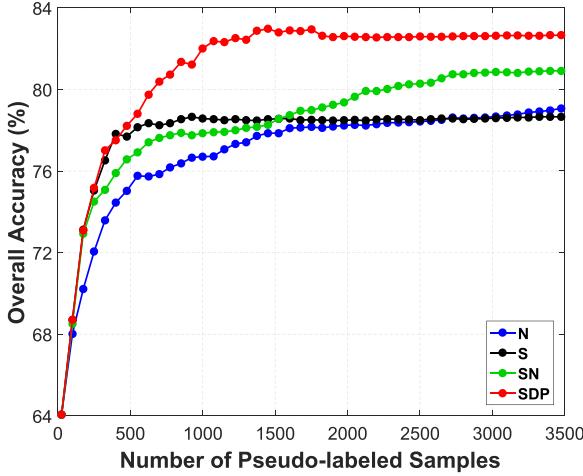


Fig. 10. OAs as a function of the number of pseudolabeled samples for the AVIRIS Indian Pines scene.

others. The improvements obtained by the superpixel-based pseudosampling are significant, especially in the beginning, when compared with the neighborhood-based methods. This is expected as the searching ranges of superpixels are wider than those of neighborhoods. Furthermore, the S converges quickly. As mentioned before, the sampling in S is trapped in the labeled superpixels. In all cases, the SSAL methods can significantly improve the initial supervised classification performance via unlabeled sampling.

For further analysis, we report the CA, OA, AA, k results, and running time (in seconds) for the case of 3000 pseudolabels from Fig. 10. Table II lists the detailed results, where $u_l = 0$ corresponds to the results in Fig. 10, while $u_l > 0$ indicates the aforementioned optimal cases. The supervised results obtained by SVM-MRF and MLR-MRF with the initial 80 labeled samples are included as reference. Several conclusions can be drawn from Table II. First of all, the proposed SDP method yielded better performance in comparison with the other methods. The advantages were particularly significant for classes with large sizes such as C2, C11, and C14. This is expected as the training samples by SDP are well distributed on the whole image, as shown in Fig. 9. Furthermore, the optimal case for SDP achieved the best results with the smallest size of true labels. Based on this observation, we can infer that the pseudolabeled samples by SDP are more informative than those by the competitors. This observation is consistent with that in Fig. 8. It also should be noted that the supervised results were significantly improved by all the SSAL methods, due to the exploration of pseudolabeled information.

For further comparison, Table III reports the number of labeled samples required by SVM-MRF and MLR-MRF, respectively, in order to achieve a comparable OA, which is obtained by our proposed SDP by using $l = 80$ initial labeled training samples and $u = 3000$ pseudolabeled training samples, respectively. This table suggests the excellent ability of our newly proposed SDP. It can be observed that, for an OA around 82%, the proposed SDP just needed about one-third of the labeled

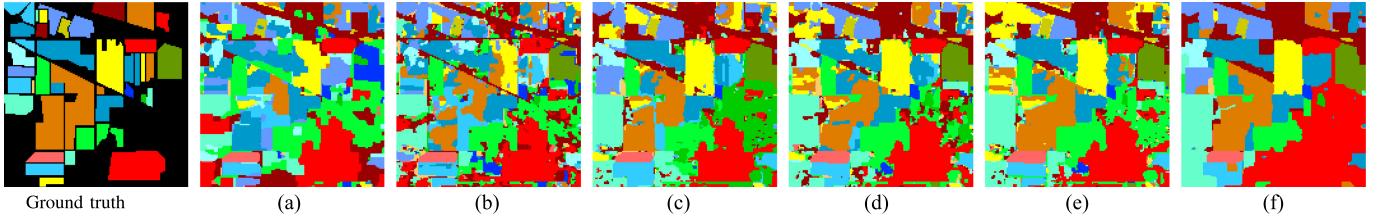


Fig. 11. Classification maps along with their corresponding OAs obtained from one of the Monte Carlo runs with 80 labeled samples and 3000 pseudolabeled samples for the AVIRIS Indian Pines scene. (a) SVM-MRF (52.68%). (b) MLR-MRF (64.56%). (c) N (78.16%). (d) S (77.01%). (e) SN (81.85%). (f) SDP (84.42%).

TABLE III
NUMBER OF LABELED AND PSEUDOLABELED SAMPLES, I.E., l AND u ,
RESPECTIVELY, REQUIRED BY THE PROPOSED SDP,
MLR-MRF, AND SVM-MRF, FOR A COMPARABLE OA
RESULT FOR THE AVIRIS INDIAN PINES SCENE

	SDP	MLR-MRF	SVM-MRF
u	3000	0	0
l	80	227	330
OA (%)	82.63	82.25	82.12

samples that MLR-MRF required and one-fourth of those required SVM-MRF, respectively.

Finally, for illustrative purposes, Fig. 11 displays the classification maps obtained by one of the Monte Carlo runs, along with the corresponding OAs. Effective results can be observed from these maps. Especially, for SDP, the classification map obtained is quite smooth, with well-delineated edges.

C. Experiments With the ROSIS Pavia University Scene

For the ROSIS Pavia University scene, we implemented three sets of experiments to evaluate the proposed approach.

1) *Impact of Pseudolabels*: In the first set of experiments, we conducted the optimal case for SDP to substantiate the deviation of our proposed approach. Fig. 12 plots the obtained OAs by SDP and its optimal case as a function of the number of pseudolabeled samples. Remarkably, the results obtained by the proposed method are very close to those obtained in the optimal case, indicating that the pseudolabels obtained by SDP are reliable.

2) *Informativeness of Pseudolabels*: In the second experiment, we evaluate the informativeness of pseudolabeled samples according to BT scores in (12). Fig. 13 plots the BT scores versus the number of pseudolabeled samples. The first observation is that, similar to that of the AVIRIS Indian Pines, the BT scores of the proposed SDP vary significantly along with the number of pseudolabeled samples, while those of the competitors mandatorily increase as the number of pseudolabeled samples increases, which means that the pseudolabeled samples are more informative than those by the other methods. However, in comparison with that of the AVIRIS Indian Pines, there is only one maximum and no minimum in the figure. This is most likely due to the difference of the image size. There are 2000 superpixels in this image, while the number is only 400 for the AVIRIS Indian Pines image. Therefore, there are still

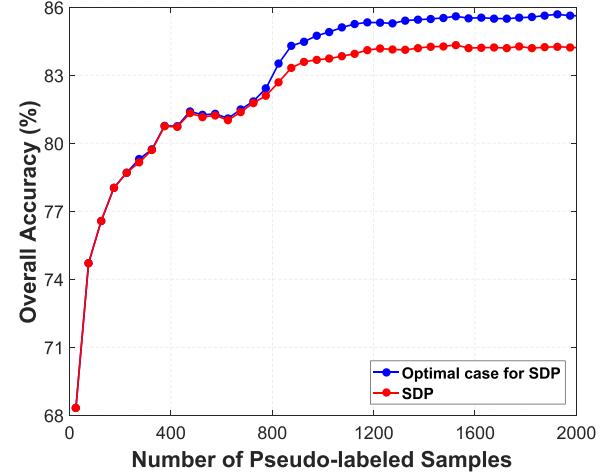


Fig. 12. OAs as a function of the number of pseudolabeled samples obtained by SDP for the ROSIS Pavia University scene. Two cases are displayed: the proposed SDP with u pseudolabels and its optimal case with u_l samples being truly labeled.

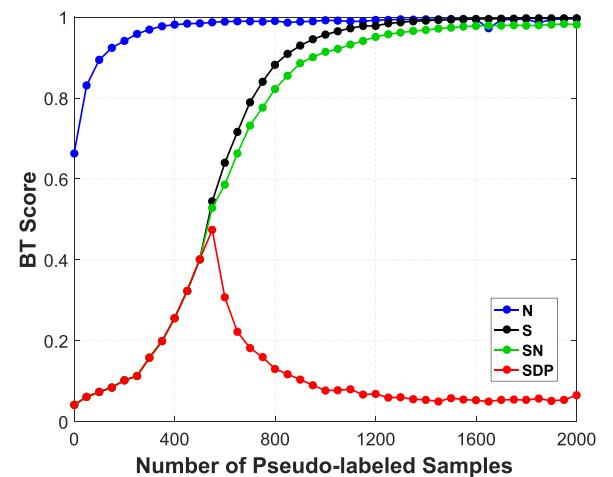


Fig. 13. Maximum BT scores of pseudolabeled samples as a function of the number of pseudolabeled samples for the ROSIS Pavia University scene.

potential unlabeled superpixels possibly to be selected for the ROSIS Pavia University, while all the potential ones were exhausted for the AVIRIS Indian Pines image. Notice that, here, we only experimented with 2000 unlabeled samples due to the good classification performance and the consideration of time

TABLE IV
 CAS [%], OAs [%], AAs [%], AND KAPPA (k) [%] STATISTICS (STANDARD DEVIATIONS IN BRACKETS) OBTAINED FROM
 TEN MONTE CARLO RUNS WITH 45 LABELED AND $u = 2000$ PSEUDOLABELED SAMPLES FOR THE ROSIS PAVIA UNIVERSITY SCENE,
 WHERE THE BEST RESULTS AMONG SSAL METHODS ARE HIGHLIGHTED IN BOLD

Class	#Test	Supervised		N		S		SN		SDP	
		SVM	MLR	$u_l = 0$	$u_l=1220$	$u_l = 0$	$u_l=1166$	$u_l = 0$	$u_l=1069$	$u_l = 0$	$u_l=539$
C1	6626	84.12	69.14	71.44	71.54	81.88	81.88	82.17	82.17	91.56	91.76
C2	18644	59.76	59.30	76.89	76.93	74.15	74.30	76.24	76.22	77.95	80.67
C3	2094	62.14	65.89	76.44	76.47	82.85	68.86	82.84	82.87	79.05	79.27
C4	3059	75.23	85.02	86.21	86.71	91.40	91.39	89.41	90.08	93.21	93.51
C5	1340	99.75	82.80	95.23	95.24	99.65	99.65	99.63	99.63	99.80	99.80
C6	5024	70.16	64.58	94.89	94.88	88.54	88.52	92.63	92.61	83.70	84.67
C7	1325	50.08	96.75	98.54	98.51	98.21	98.21	97.88	97.88	95.58	95.57
C8	3677	61.31	82.32	79.91	80.15	83.43	83.44	83.97	83.97	86.21	86.42
C9	942	96.53	96.55	98.51	98.51	97.62	97.78	97.77	97.87	95.17	96.54
OA		67.88	68.31	80.79	80.88	81.56	81.63	82.90	82.94	84.20	85.61
		(6.32)	(7.79)	(5.02)	(4.98)	(3.61)	(3.64)	(4.06)	(3.96)	(3.79)	(3.20)
AA		73.23	78.04	86.45	86.55	88.64	88.67	89.17	89.26	89.14	89.80
		(6.15)	(5.04)	(2.63)	(2.57)	(2.13)	(2.14)	(1.94)	(1.83)	(1.72)	(1.51)
k		59.78	60.84	75.73	75.84	76.75	76.83	78.39	78.44	79.86	81.56
Time		3.22	196.13	1.95×10^4		1.87×10^4		1.91×10^4		1.93×10^4	

The optimal cases are also presented along with the values of u_l , i.e., the number of the used true labels.

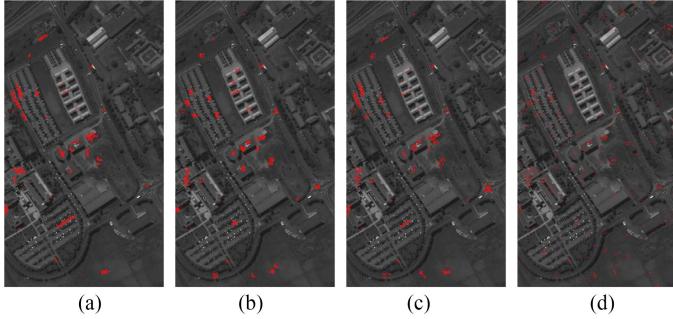


Fig. 14. Training samples at the 80th iteration for the ROSIS Pavia University scene. (a) N. (b) S. (c) SN. (d) SDP.

complexity. Nevertheless, it is expected that, if we increase the number of iterations, i.e., when all the potential unlabeled superpixels are included, the BT score in Fig. 13 will increase again.

Finally, Fig. 14 graphically illustrates the distribution of the training samples with 2000 pseudolabeled samples. It can be seen that the proposed SDP can select pseudolabeled samples that are well distributed on the whole image, resulting in very good spatial diversity. However, the pseudolabels of S are trapped in a few labeled superpixels. With respect to N and SN, the training samples are obviously gathered around the labeled ones.

3) *Classification Performance*: In the final set of experiments, we evaluate the classification performance obtained by the proposed method. Fig. 15 illustrates the OA results as a function of the number of pseudolabeled samples. It can be seen from Fig. 15 that the proposed SDP method exhibits competitive performance when compared with the other methods. Furthermore, the superpixel-based methods showed a significant advantage in the early stage, due to the fact that the pseudolabel-

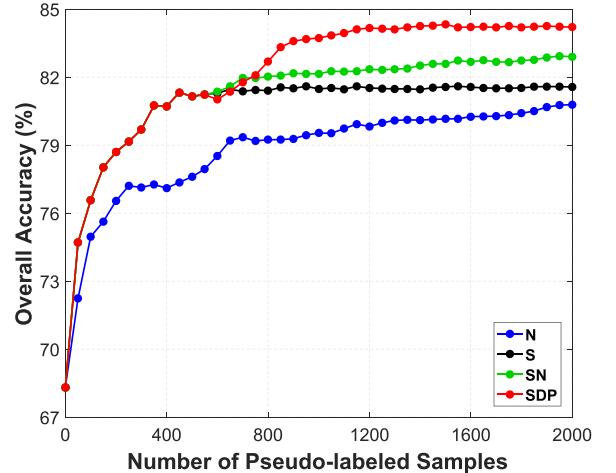


Fig. 15. OAs as a function of the number of pseudolabeled samples for the ROSIS Pavia University scene.

ing (under the superpixel-based similarity region) can search in wider ranges than the neighborhood-based one. However, the S reached the limit quickly due to the trap of labeled superpixels. On this scene, all the SSAL methods are also proven effective in improving the initial supervised classification results.

For further comparison, we list the classification results and running time (in seconds) obtained with 45 initial labeled samples and 2000 pseudolabeled ones in Table IV. The conclusions drawn from this table are in accordance with the good performance exhibited by the proposed SDP, as reported in previous experiments for the AVIRIS Indian Pines scene. It can be observed that SDP outperformed the other methods, especially on the classes with large sizes such as C1, C2, and C8. Additionally, the optimal case for SDP demonstrated a clear advantage over the competitors, with a smaller size of truly labeled

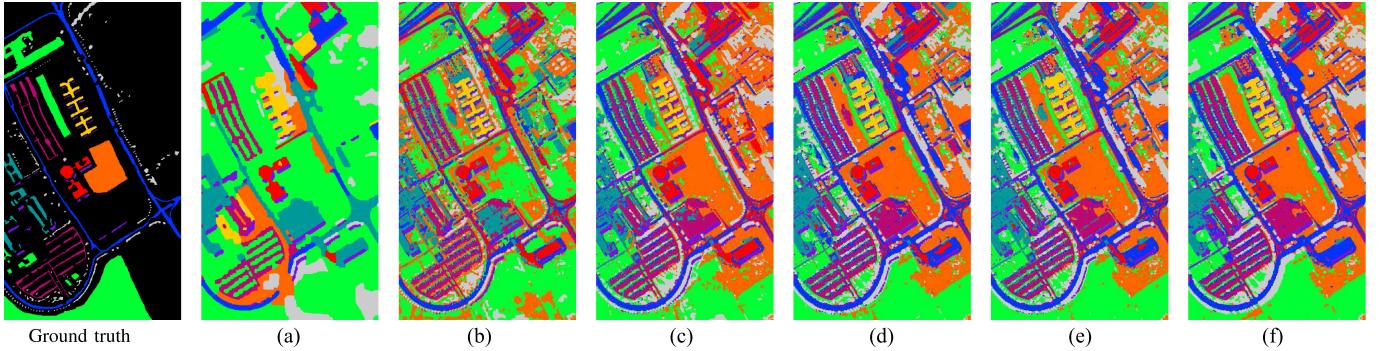


Fig. 16. Classification maps along with the corresponding OAs obtained from one of the Monte Carlo runs with 45 labeled and 2000 pseudolabeled samples for the ROSIS Pavia University scene.

TABLE V

NUMBER OF LABELED SAMPLES, I.E., l , AND THAT OF PSEUDOLABELED SAMPLES, I.E., u , USED BY SDP, MLR-MRF, AND SVM-MRF, RESPECTIVELY, IN ORDER TO OBTAIN COMPARABLE OA RESULTS FOR THE ROSIS PAVIA UNIVERSITY SCENE

	SDP	MLR-MRF	SVM-MRF
u	2000	0	0
l	45	88	146
OA (%)	84.20	83.86	83.50

samples, from which it can be inferred that the pseudolabeled samples by SDP are more informative than those obtained by the other methods. In this context, all the SSAL methods can significantly improve the initial supervised classification performances.

Furthermore, Table V presents the number of labeled samples required by MLR-MRF and SVM-MRF, respectively, for achieving an OA that is comparable to the one obtained by our proposed SDP in Table IV. The excellent ability of the proposed SDP can also be well demonstrated in this table. Notice that, for an OA around 84%, the proposed SDP only used about half of the labeled samples required by MLR-MRF, and one-third of those by SVM-MRF.

Finally, Fig. 16 shows the obtained classification maps from one of the Monte Carlo runs, along with the corresponding OAs. The good performance of the proposed SDP can also be seen from these maps.

D. Experiments With the OMIS Zaoyuan Scene

In this experiment, we use the OMIS Zaoyuan scene to further evaluate the classification performance of the proposed approach. This image constitutes a challenging problem due to the significant presence of mixed pixels in all available classes. Fig. 17 illustrates the obtained OAs as a function of the number of pseudolabeled samples. It is worthwhile to note that the proposed SDP exhibits a significant advantage when compared with the other considered methods. Its advantages increase with the increase of the number of pseudolabeled samples. This observation is consistent with those reported for the two previously considered datasets, which suggests that the newly included

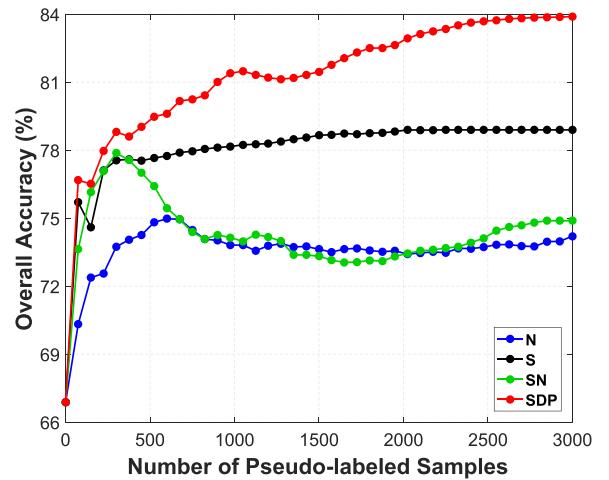


Fig. 17. OAs as a function of the number of pseudolabeled samples for the OMIS Zaoyuan scene.

samples should be confident and informative. Furthermore, the performance of S converged quickly, which allows us to infer that the new samples might be confident enough (although without enough informativeness). Finally, the performance of the two neighborhood-based methods decreases when the number of pseudolabeled samples increases, which suggests that the pseudolabeled samples may be poorly labeled. Therefore, from these observations, it can be revealed that our newly proposed SDP is quite robust in different scenarios. Specifically, its advantages are remarkable in challenging classification scenarios dominated by the presence of highly mixed pixels.

V. CONCLUSION

In this paper, we have presented a new superpixel-based SSAL method, which introduces the concept of spatial adaptivity into pseudolabeling to improve the machine generalization. Our newly developed method also includes a DP-based augmentation strategy, which aims at solving the problem that pseudolabeling may be trapped in a few labeled superpixels. Our experimental results with three benchmark hyperspectral datasets indicate the robustness of our proposed approach when compared with other SSAL methods. Our future work will be

focused on investigating a more accurate DP-based strategy for augmentation, thus allowing our newly proposed superpixel-based method to benefit more from such a strategy (especially in large-scale cases). Besides, we will also investigate the possible development a pixelwise solution to generate superpixel-based neighborhoods.

REFERENCES

- [1] A. F. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, 1985.
- [2] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, 2009.
- [3] Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, Jan. 1968.
- [4] X. Guo, X. Huang, L. Zhang, L. Zhang, A. Plaza, and J. A. Benediktsson, "Support tensor machines for classification of hyperspectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3248–3264, Jun. 2016.
- [5] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.
- [6] X. Huang, C. Weng, Q. Lu, T. Feng, and L. Zhang, "Automatic labelling and selection of training samples for high-resolution remote sensing image classification over urban areas," *Remote Sens.*, vol. 7, no. 12, pp. 16024–16044, 2015. [Online]. Available: <http://www.mdpi.com/2072-4292/7/12/15819>
- [7] O. Chapelle, B. Schlkopf, and A. Zien, *Semi-Supervised Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 2010.
- [8] X. Zhu, "Semi-supervised learning literature survey," Univ. Wisconsin-Madison, Madison, WI, USA, Comput. Sci. Tech. Rep. 1530, 2005.
- [9] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. 11th Annu. Conf. Comput. Learn. Theory*, 1998, pp. 92–100.
- [10] X. Zhang, Q. Song, R. Liu, W. Wang, and L. Jiao, "Modified co-training with spectral and spatial views for semisupervised hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2044–2055, Jun. 2014.
- [11] U. Maulik and D. Chakraborty, "Remote sensing image classification: A survey of support-vector-machine-based advanced techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 33–52, Mar. 2017.
- [12] S. Velasco-Forero and V. Manian, "Improving hyperspectral image classification using spatial preprocessing," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 297–301, Apr. 2009.
- [13] Q. Shi, L. Zhang, and B. Du, "Semisupervised discriminative locally enhanced alignment for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4800–4815, Sep. 2013.
- [14] M. S. Aydemir and G. Bilgin, "Semisupervised hyperspectral image classification using small sample sizes," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 621–625, May 2017.
- [15] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 318–322, Mar. 2013.
- [16] Z. Wang, N. M. Nasrabadi, and T. S. Huang, "Semisupervised hyperspectral classification using task-driven dictionary learning with laplacian regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1161–1173, Mar. 2015.
- [17] J. Li, P. Gamba, and A. Plaza, "A novel semi-supervised method for obtaining finer resolution urban extents exploiting coarser resolution maps," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 10, pp. 4276–4287, Oct 2014.
- [18] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Econ. Geography*, vol. 46, pp. 234–240, 1970.
- [19] D. Han, Q. Du, and N. H. Younan, "Semisupervised classification of hyperspectral remote sensing images with spatial majority voting," in *Proc. 9th IAPR Workshop Pattern Recognit. Remote Sens.*, Dec. 2016, pp. 1–4.
- [20] G. Camps-Valls, T. V. B. Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.
- [21] L. Ma, M. M. Crawford, X. Yang, and Y. Guo, "Local-manifold-learning-based graph construction for semisupervised hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2832–2844, May 2015.
- [22] T. Luo *et al.*, "Active learning to recognize multiple types of plankton," in *Proc. 17th Int. Conf. Pattern Recognit.*, Aug. 2004, vol. 3, pp. 478–481.
- [23] 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.
- [24] 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.
- [25] D. Tuia, E. Pasolini, and W. Emery, "Using active learning to adapt remote sensing image classifiers," *Remote Sens. Environ.*, vol. 115, no. 9, pp. 2232–2242, 2011.
- [26] B. Settles, "Active learning literature survey," Univ. Wisconsin–Madison, Madison, WI, USA, Comput. Sci. Tech. Rep. 1648, Jan. 2016.
- [27] A. Krishnakumar, "Active learning literature survey," Jun. 2007.
- [28] M. M. Crawford, D. Tuia, and H. L. Yang, "Active learning: Any value for classification of remotely sensed data?" *Proc. IEEE*, vol. 101, no. 3, pp. 593–608, Mar. 2013.
- [29] J. Guo, X. Zhou, J. Li, A. Plaza, and S. Prasad, "Superpixel-based active learning and online feature importance learning for hyperspectral image analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 1, pp. 347–359, Jan. 2017.
- [30] Q. Shi, X. Liu, and X. Huang, "An active relearning framework for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3468–3486, Jun. 2018.
- [31] 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.
- [32] X. Lu, J. Zhang, T. Li, and Y. Zhang, "A novel synergetic classification approach for hyperspectral and panchromatic images based on self-learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4917–4928, Aug. 2016.
- [33] Z. Xue, P. Du, J. Li, and H. Su, "Sparse graph regularization for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2351–2366, Apr. 2017.
- [34] I. Dópido, J. Li, P. R. Marpu, A. Plaza, J. M. B. Dias, and J. A. Benediktsson, "Semisupervised self-learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4032–4044, Jul. 2013.
- [35] L. Wang, S. Hao, Y. Wang, Y. Lin, and Q. Wang, "Spatial-spectral information-based semisupervised classification algorithm for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3577–3585, Aug. 2014.
- [36] K. Tan, J. Zhu, Q. Du, L. Wu, and P. Du, "A novel tri-training Technique for semi-supervised classification of hyperspectral images based on diversity measurement," *Remote Sens.*, vol. 8, no. 9, Sep. 2016, Art. no. 749.
- [37] X. Lu, J. Zhang, T. Li, and Y. Zhang, "Incorporating diversity into self-learning for synergetic classification of hyperspectral and panchromatic images," *Remote Sens.*, vol. 8, no. 10, Oct. 2016, Art. no. 804.
- [38] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, vol. 1, pp. 10–17.
- [39] 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.
- [40] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.
- [41] B. Krishnapuram, L. Carin, M. A. T. Figueiredo, and A. J. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 957–968, Jun. 2005.
- [42] B. Krishnapuram, D. Williams, Y. Xue, A. Hartemink, L. Carin, and M. A. T. Figueiredo, "On semi-supervised classification," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2005, pp. 721–728.
- [43] J. Bioucas-Dias and M. Figueiredo, "Logistic regression via variable splitting and augmented Lagrangian tools," Inst. Superior Técnico, Tech. Univ. Lisbon, Lisbon, Portugal, Tech. Rep., 2009.
- [44] A. Schick, M. Fischer, and R. Stiefelhagen, "An evaluation of the compactness of superpixels," *Pattern Recognit. Lett.*, vol. 43, pp. 71–80, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865513003498>

- [45] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [46] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC Superpixels," École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, Tech. Rep. 149300, 2010.
- [47] X. Xu, J. Li, C. Wu, and A. Plaza, "Regional clustering-based spatial pre-processing for hyperspectral unmixing," *Remote Sens. Environ.*, vol. 204, pp. 333–346, Jan. 2018.
- [48] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010, award winning papers from the Proc. 19th Int. Conf. Pattern Recognit., [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865509002323>
- [49] Y. Du, C. Chang, H. Ren, C. Chang, J. Jensen, and F. D'Amico, "New hyperspectral discrimination measure for spectral characterization," *Opt. Eng.*, vol. 43, no. 8, pp. 1777–1786, Aug. 2004.
- [50] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.
- [51] S. Li, *Markov Random Field Modeling in Computer Vision*. Tokyo, Japan: Springer, 1995.
- [52] M. Khodadadzadeh, J. Li, A. Plaza, H. Ghassemian, J. M. Bioucas-Dias, and X. Li, "Spectral spatial classification of hyperspectral data using local and global probabilities for mixed pixel characterization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6298–6314, Oct. 2014.
- [53] J. A. Richards and X. Jia, "Assessment of classification accuracy," in *Remote Sensing Digital Image Analysis*, 4th ed. Berlin, Germany: Springer, 2006, ch. 11.5, pp. 303–324.
- [54] R. O. Green *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, 1998.



Chenying Liu (S'16) received the B.S. degree in geographic information science from Sun Yat-sen University, Guangzhou, China, in 2017. She is currently pursuing the M.S. degree at the School of Geography and Planning, Sun Yat-sen University.

Her research interests include pattern recognition and remote sensing image processing, especially hyperspectral image processing.



Jun Li (M'12–SM'16) received the B.S. degree in geographic information systems from Hunan Normal University, Changsha, China, in 2004, the M.E. degree in remote sensing from Peking University, Beijing, China, in 2007, and the Ph.D. degree in electrical engineering from the Instituto de Telecomunicações, Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Lisbon, Portugal, in 2011.

From 2007 to 2011, she was a Marie Curie Research Fellow with the Departamento de Engenharia Electrotécnica e de Computadores and the Instituto de Telecomunicações, IST, Universidade Técnica de Lisboa, in the framework of the European Doctorate for Signal Processing. She has also been actively involved in the Hyperspectral Imaging Network, a Marie Curie Research Training Network involving 15 partners in 12 countries and intended to foster research, training, and cooperation on hyperspectral imaging at the European level. Since 2011, she has been a Postdoctoral Researcher with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, Spain. From 2014–2018, she was a Professor with Sun Yat-sen University, Guangzhou, China. She is currently a Professor with Human University, Changsha, China. Her current research interests include hyperspectral image classification and segmentation, spectral unmixing, signal processing, and remote sensing.

Dr. Li is an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING and a Reviewer for several journals.



Lin He (S'05–M'12) received the B.S. degree from the Xi'an Institute of Technology, China, in 1995, the M.S. degree from Chongqing University, Chongqing, China, in 2003, both in instrumentation engineering, and the Ph.D. degree in pattern recognition and intelligent systems from Northwestern Polytechnical University, Xi'an, China, in 2007.

Since 2007, he has been with the School of Automation Science and Engineering, South China University of Technology, Guangzhou, China, where he is currently an Associate Professor.

His current research interests include statistical pattern recognition, hyperspectral image processing, and high-dimensional signal processing.