Hypergraph p-Laplacian Regularization for Remotely Sensed Image Recognition

Xueqi Ma, Weifeng Liu, *Senior Member, IEEE*, Shuying Li, Dapeng Tao, and Yicong Zhou, *Senior Member, IEEE*

***Abstract*— Graph-based and manifold-regularization (MR)-based semisupervised learning, including Laplacian regularization (LapR) and hypergraph LapR (HLapR), have achieved prominent performance in preserving locality and similarity information. However, it is still a great challenge to exactly explore and exploit the local structure of the data distribution. In this paper, we present an efficient and effective approximation algorithm of hypergraph p-Laplacian and then propose hypergraph p-LapR (HpLapR) to preserve the geometry of the probability distribution. In particular, hypergraph is a generalization of a standard graph while hypergraph p-Laplacian is a nonlinear generalization of the standard graph Laplacian. The proposed HpLapR shows great potential to exploit the local structures. We integrate HpLapR with logistic regression for remote sensing image recognition. Experiments on UC-Merced data set demonstrate that the proposed HpLapR has superior performance compared with several popular MR methods including LapR and HLapR.**

***Index Terms*— Hypergraph, manifold learning, remote sensing, semisupervised learning (SSL), p-Laplacian.**

Nomenclature

V	Finite set of vertices.
$w(e)$	Weight associated with each hyperedge e.
$\delta(e)$	Degree of a hyperedge $e \in E$, $\delta(e) = \lvert e \rvert$.
L^{hp}	Hypergraph Laplacian.
$\mathcal{F}^{*\text{hp}}$	Eigenvectors of hypergraph p-Laplacian $\mathcal{F}^{*\text{hp}} = (f^{*\text{hp}1}, f^{*\text{hp}2}, \ldots, f^{*\text{hp}n})$.
f^{hp}	Eigenvector of hypergraph p-Laplacian.
p	p of p-Laplacian.
K	Embedding dimension.
N	Number of training images.
u	Number of unlabeled images.
$\boldsymbol{K}$	Kernel function $\boldsymbol{K}_{ij} = \boldsymbol{K}(x_i, x_j)$.
$\lVert f \rVert_K^2$	Penalty term used to control the complexity of the classification model.
α	Coefficients.
Υ_A	Parameter corresponds with $\lVert f \rVert_K^2$.
E	Family of subsets of V, $\bigcup_{e \in E} = V$.
$d(v)$	Degree of a vertex $v \in V$, $d(v) = \sum_{\{e \in E \mid v \in e\}} w(e)$.
H	$h(v, e) = 1$ if $v \in e$, and $h(v, e) = 0$ otherwise.
L_p^{hp}	Hypergraph p-Laplacian.
$\lambda^{*\text{hp}}$	Eigenvalue of hypergraph p-Laplacian $\lambda^{*\text{hp}} = (\lambda_1^{*hp}, \lambda_2^{*\text{hp}}, \ldots, \lambda_n^{*\text{hp}})$.
λ^{hp}	Eigenvalue of hypergraph p-Laplacian.
k	Number of nearest neighbors.
W^{hp}	Adjacency matrix of hypergraph.
l	Number of labeled images.
Y	Class labels of labeled images $Y = \{y_i\}_{i=1}^{l}$, $y_i \in \{\pm 1\}$.
$\lVert f \rVert_I^2$	Appropriate penalty term corresponding to the probability distribution.
$\mathbf{f}$	$\mathbf{f} = [f(x_1), f(x_2), \ldots, f(x_{l+u})]^T$.
Υ_I	Parameter corresponds with $\lVert f \rVert_I^2$.
$\text{pre}(r)$	Precision at recall r.

Manuscript received June 2, 2018; revised August 5, 2018 and August 18, 2018; accepted August 21, 2018. Date of publication September 20, 2018; date of current version February 25, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61671480, in part by the Foundation of Shandong province under Grant ZR2018MF017, in part by the Fundamental Research Funds for the Central Universities, China University of Petroleum (East China) under Grant 18CX07011A and Grant YCX2017059, in part by the National Natural Science Foundation of China under Grant 61772455 and Grant U1713213, in part by the Yunnan Natural Science Funds under Grant 2016FB105, in part by the Program for Excellent Young Talents of Yunnan University under Grant WX069051, in part by the Macau Science and Technology Development Fund under Grant FDCT/189/2017/A3, in part by the Research Committee at University of Macau under Grant MYRG2016-00123-FST and Grant MYRG2018-00136-FST, and in part by the National Natural Science Foundation of China under Grant 61701387. *(Corresponding authors: Weifeng Liu; Yicong Zhou.)*

X. Ma and W. Liu are with the College of Information and Control Engineering, China University of Petroleum, Qingdao 266580, China (e-mail: liuwf@upc.edu.cn).

S. Li is with the School of Automation, Xi'an University of Posts and Telecommunications, Xi'an 710121, China (e-mail: angle_lisy@163.com).

D. Tao is with the School of Information Science and Engineering, Yunnan University, Kunming 650091, China (e-mail: dapeng.tao@gmail.com).

Y. Zhou is with the Faculty of Science and Technology, University of Macau, Macau 999078, China (e-mail: yicongzhou@umac.mo).

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

Digital Object Identifier 10.1109/TGRS.2018.2867570

I. Introduction

NOWADAYS, billions of images (e.g., action images, human face images, object images, and scene images) are uploaded to social media platform such as YouTube, Facebook, and Twitter. Image recognition including scene recognition [1]–[3], human face recognition [4], [5], action recognition [6]–[8], and remotely sensed image recognition [9]–[12] has become a quite significant topic in machine vision. With the fast development of space technologies, remotely sensed techniques have been utilized in many applications including targeting, environment monitoring, surveillance, and military systems. In the remotely sensed community, it is particularly important to effectively assign the categories of high-resolution remotely sensed imagery data, e.g., aerial images and radar images. Aerial scene classification [13], [14] has received growing attention due to the drastically increasing number of aerial images and the highly complex geometrical structures and spatial patterns.

However, because annotating images is costly and time consuming, a small number of labeled samples are available in practical applications, whereas a lot of unlabeled samples are easy to collect. Semisupervised learning (SSL) which can make use of labeled and unlabeled data has been investigated to solve this problem. One successful work is manifold regularization (MR), which has attracted considerable attention due to its rich theoretical studies [15]–[17] and its excellent performance in multimedia data (e.g., text, image, video, and audio) processing [18]–[24]. The main idea of MR is to explore the geometry of the intrinsic probability distribution of data to leverage the learning performance. Another successful work is graph-based SSL [25], [26], which constructs a similarity graph over data to exploit the local geometry of labeled and unlabeled data and has achieved appealing performance due to its flexibility and low computation complexity in practice.

The MR framework [15] exploits the geometry of the probability distribution of the data and incorporates it as a regularization term. Laplacian regularization (LapR) is one prominent MR-based SSL algorithm, which determines the underlying manifold by using the graph Laplacian. Wang *et al.* [27] presented a manifold-regularized (MR) multiview subspace clustering method to better incorporate the correlated and complementary information from different views. The graph Laplacian is constructed to maintain the local data manifold of each view. Luo *et al.* [19] employed MR to smooth the functions along the data manifold for multitask learning. Jiang *et al.* [18] presented a mutimanifold method for recognition by exploring the local geometric structure of samples. Liu and Tao [28] proposed multiview Hessian-regularized logistic regression which combines multiple Hessian regularizations to leverage the local geometry. Lu and Mou [29] built a model of sparse-feature-selection-based MR to select the optimal information and preserve the underlying manifold structure of data for scene recognition.

Typically, in graph-based SSL, it is assumed that there is a graph over the data lying on data manifolds. In the graph, vertices represent samples and edge weights indicate the similarity between samples. For example, Zhou *et al.* [30] constructed a directed graph learned from labeled and unlabeled data for web categorization, in which each vertex represents a web page, and each edge represents a hyperlink between two web pages. For graph-based SSL, it is essential to construct an effective graph over data with complex distribution. Compared with the existing simple graph that only models the pairwise relationship of images, hypergraph learning that uses a hyperedge to link multiple samples can model the high-order relationship of samples.

In [31], the hypergraph idea was first introduced to the field of computer vision. It is a generalization of a simple graph. Unlike a simple graph that considers the relationship between two vertices, a set of vertices is connected by a hyperedge in a hypergraph. Thus, the hypergraph contains more local grouping information than does the simple graph. Hypergraph has been widely used in image classification [32], [4], ranking [33], [34], and video segmentation [35]. Sun *et al.* [36] constructed a hypergraph to exploit the correlation information among different labels for multilabel learning. Zass and Shashua [37] presented a hypergraph-based image matching problem in a probabilistic setting represented by a convex optimization problem. Huang *et al.* [34] proposed a hypergraph-based transductive algorithm to the field of image retrieval. Yu *et al.* [4] proposed an adaptive hypergraph learning method for transductive image classification.

In this paper, we propose a hypergraph p-Laplacian-regularized (HpLapR) method for remote sensing image recognition. The hypergraph and p-Laplacian [38], [39], [41] both provide convincing theoretical evidences to better preserve the local structure of data. However, the computation of hypergraph p-Laplacian is a strenuous task. We provide an effective and efficient approximation algorithm of hypergraph p-Laplacian. Considering the high-order relationship of samples, the HpLapR is built for preserving local structures. HpLapR is also introduced to logistic regression for remote sensing image recognition. Experiments on the UC-Merced data set [42] compare the proposed method with the popular algorithms including LapR, hypergraph LapR (HLapR), and p-LapR (pLapR). The contributions of this paper can be summarized as follows.

1) We present an efficient approximation algorithm of hypergraph p-Laplacian, significantly improving computation efficiency.
2) We propose HpLapR to preserve the local similarity of data.
3) We integrate HpLapR into logistic regression and conduct comprehensive experiments to empirically analyze our method on UC-Merced data set. The experimental results validate the effectiveness of our method.

The rest of this paper is organized as follows. Section II briefly reviews related work on MR and hypergraph learning. Section III introduces an approximate computation of the hypergraph p-Laplacian. Section IV proposes HpLapR. Section V presents the HpLapR logistic regression. Section VI provides the experimental results and analysis on UC-Merced data set. Finally, Section VII gives the conclusion.

II. Related Work

In this section, we briefly review MR and hypergraph.

A. Manifold Regularization

Assume that the estimated function is generated from the probability distribution on samples (labeled and unlabeled samples). The labeled samples are (x, y) pairs generated according to probability distribution, and lie on the estimation curve in the ideal case. The unlabeled samples are drawn according to the marginal distribution. Based on the manifold assumption that if two samples are close in the intrinsic geometry, they have the similar labels. It is important to exploit the knowledge of the marginal distribution for better function learning.

Introducing an additional regularizer for preserving local structures, the MR framework can be interpreted as regularization algorithms that consist of different empirical cost

functions, complexity measures in an appropriately chosen reproducing kernel Hilbert space (RKHS) and additional information about the geometric structure of the margins. Consider l labeled samples and u unlabeled samples, the objective function can be written as

$$f^* = \arg\min_{f\in H_K} \frac{1}{l}\sum_{i=1}^{l} V(x_i, y_i, f) + \Upsilon_A \|f\|_K^2 + \Upsilon_I \|f\|_I^2 \quad (1)$$

where H_K is the RKHS, V is a loss function, such as the hinge loss function $\max[0, 1 - y_i f(x_i)]$ for support vector machines. There is an associated RKHS H_K of functions $X \to \mathbb{R}$ with the corresponding norm $\|\|_k$. $\|f\|_K^2$ is used to control the complexity of the classification model, while $\|f\|_I^2$ is an appropriate penalty term corresponding to the probability distribution. Parameters Υ_A and Υ_I control the complexity of the function in the ambient space and the intrinsic geometry, respectively.

The classical Representer theorem states that f exists in H_K and can be written as

$$f^*(x) = \sum_{i=1}^{l+u} \alpha_i^* \boldsymbol{K}(x_i, x) \quad (2)$$

where α^i is the coefficient and $\boldsymbol{K}$ is a kernel function.

Graph Laplacian has been widely used to explore and exploit the local geometry of data distribution. As a nonlinear generalization of the standard graph Laplacian, graph p-Laplacian has attracted attentions from the machine-learning community. Zhou and Schölkopf [39] proposed a general discrete regularization framework of p-Laplacian for the classification problem, and its objective function can be computed as follows:

$$f^* = \underset{f\in\mathcal{H}(V)}{\operatorname{argmin}} \{\mathcal{S}_p(f) + \mu\|f - y\|^2\} \quad (3)$$

where $\mathcal{S}_p(f) := (1/2)\sum_{v\in V} \|\nabla_v f\|^p$ is the p-Dirichlet form of the function f, μ is a parameter balancing the two competing terms, and $y \in \{-1, 0, 1\}$ depends on the label of the sample.

Bühler and Hein [38] used the graph p-Laplacian for spectral clustering and demonstrated the relationship between the second eigenvalue of the graph p-Laplacian and the optimal Cheeger cut as follows:

$$\mathrm{RCC} \le \mathrm{RCC}^* \le p\Big(\max_{i\in V} d_i\Big)^{\frac{p-1}{p}} \mathrm{RCC}^{\frac{1}{p}} \quad (4)$$

or

$$\mathrm{NCC} \le \mathrm{NCC}^* \le p\mathrm{NCC}^{\frac{1}{p}} \quad (5)$$

where RCC^* and NCC^* are the ratio/normalized Cheeger cut values obtained by thresholding the second eigenvector of the unnormalized/normalized p-Laplacian, d_i is the degree of vertex i, and RCC and NCC are the optimal ratio/normalized Cheeger cut values.

Luo *et al.* [43] used the p-Laplacian for multiclass clustering and provided an approximation of the whole eigenvectors by solving the tractable optimization problem

$$\min_{\mathcal{F}} \ J_E(\mathcal{F}) = \sum_k \frac{\sum_{ij} w_{ij}\left|f_i^k - f_j^k\right|^p}{\|f^k\|_p^p}$$
$$\text{s.t. } \mathcal{F}^T\mathcal{F} = I \quad (6)$$

where w_{ij} is the edge weight, f^k is an eigenvector of p-Laplacian, and $\mathcal{F} = (f^1, f^2, \ldots, f^n)$ are whole eigenvectors.

Liu *et al.* [6] proposed pLapR sparse coding for preserving the manifold structure.

B. Hypergraph

In machine-learning issues, we generally assume pairwise relationship among the object set. An object set endowed with pairwise relationship can be considered as a graph. The graph can be undirected or directed. However, in a number of questions, it is not complete to represent the relations among samples only using simple graphs. Hypergraph learning [34] addresses this problem. Compared with traditional graph, a hypergraph illustrates the complex relationship by hyperedges which connect three or more vertices (see in Fig. 1).

Let V denote a finite set of vertices and E be a family of subsets of V such that $\bigcup_{e\in E} = V$. A hypergraph $G = (V, E)$ corresponding to the vertex set V and the hyperedge set E. Denote the weight associated with each hyperedge e as $w(e)$. The degree of a vertex $v \in V$ is defined by $d(v) = \sum_{\{e\in E|v\in e\}} w(e)$. The degree of a hyperedge $e \in E$ is denoted by $\delta(e) = |e|$. Denote the incident matrix H by a $|V| \times |E|$ matrix, whose entry $h(v, e) = 1$ if $v \in e$, and $h(v, e) = 0$ otherwise. Then

$$d(v) = \sum_{e\in E} w(e)h(v, e) \quad (7)$$

$$\delta(e) = \sum_{v\in V} h(v, e). \quad (8)$$

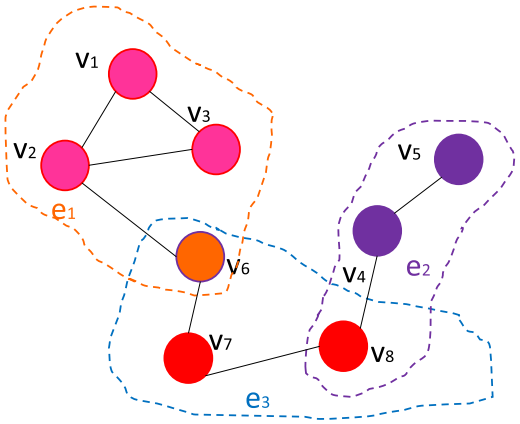
Let Dv denote the diagonal matrices containing the degree of vertex, D_e denote the diagonal degree matrices of each hyperedge, and W is the diagonal matrix of edge weights. Then, the hypergraph Laplacian can be defined.

There have been many methods for building the graph Laplacian of hypergraphs across the literature. The first category includes star expansion [44], clique expansion [44], and Rodriquez's [45] Laplacian. These methods aim to construct a simple graph from the original hypergraph, and then partition the vertices using spectral clustering techniques. The second category of approaches defines a hypergraph Laplacian using analogies from the simple graph Laplacian. Representative methods in this category include Bolla's [46] Laplacian and Zhou *et al.*'s [47] normalized Laplacian. In [47], the normalized hypergraph Laplacian is defined as

$$L^{\mathrm{hp}} = I - Dv^{-1/2}\mathrm{HW}D_e^{-1}H^T Dv^{-1/2}. \quad (9)$$

Note that L^{hp} is the positive semidefinite. The adjacency matrix of hypergraph can be formulated as follows:

$$W^{\mathrm{hp}} = \mathrm{HW}H^T - Dv. \quad (10)$$



	e1	e2	e3
v1	1	0	0
v2	1	0	0
v3	1	0	0
v4	0	1	0
v5	0	1	0
v6	1	0	1
v7	0	0	1
v8	0	1	1

Fig. 1. Block scheme of hypergraph. (Left) Simple graph in which two points are joined together by an edge if they are highly similarity. A hypergraph completely illustrates the complex relationship among points by hyperedges. (Right) H matrix of the hypergraph. The entry (v_i, e_j) is set to 1 if a hyperedge e_j contains v_i, or 0 otherwise.

For a simple graph, the edge degree matrix D_e is replaced by $2I$. Thus, the standard graph Laplacian is

$$L = I - \frac{1}{2} Dv^{-\frac{1}{2}} \mathrm{HW} H^T Dv^{-\frac{1}{2}} = \frac{1}{2}(I - Dv^{-1/2} W^{hp} Dv^{-1/2}). \tag{11}$$

III. Approximation of Hypergraph p-Laplacian

In this section, we introduce the approximation algorithm of hypergraph p-Laplacian L_p^{hp}.

Assume that hypergraph p-Laplacian has n eigenvectors $\mathcal{F}^{*\mathrm{hp}} = (f^{*\mathrm{hp}1}, f^{*\mathrm{hp}2}, \ldots, f^{*\mathrm{hp}n})$ associated with unique eigenvalues $\lambda^{*\mathrm{hp}} = (\lambda_1^{*\mathrm{hp}}, \lambda_2^{*\mathrm{hp}}, \ldots, \lambda_n^{*\mathrm{hp}})$, we compute the approximation of L_p^{hp} by $L_p^{\mathrm{hp}} = \mathcal{F}^{*\mathrm{hp}} \lambda^{*\mathrm{hp}} \mathcal{F}^{*\mathrm{hp}T}$. Thus, it is important to obtain all eigenvectors and eigenvalues of hypergraph p-Laplacian.

Although a complete analysis of hypergraph p-Laplacian is challenging, we can easily generate a hypergraph with a group of hyperedges [47]. In detail, we construct hypergraph Laplacian L^{hp} and compute adjacency matrix W^{hp} by (8) and (9), respectively.

Then, we introduce the basic definition of p-Laplacian Δ_p^w including its eigenvalue and eigenvector.

The real number λ_p is called an eigenvalue of Δ_p^w, if there exists a function $f : V \to \mathcal{R}$ satisfying the relationship as follows:

$$\left(\Delta_p^w f\right)_i = \lambda_p \phi_p(f_i), \quad i \in V. \tag{12}$$

Function f is called a p-eigenfunction (also called eigenvector) associated with λ_p. ϕ_p is defined by $\phi_p(x) = |x|^{p-1}\mathrm{sign}(x)$. Note that the operator $\Delta_2^w = L$ becomes the regular graph Laplacian.

Following previous studies on p-Laplacian [38], eigenvalue and the corresponding eigenvector on nonlinear operator Δ_p^w can be computed by the theorem.

Function F_p has a critical point at f if and only if f is an eigenvector of Δ_p^w. F_p is defined as

$$F_p(f) = \frac{\sum_{ij} w_{ij} |f_i - f_j|^p}{2\|f\|_p^p} \tag{13}$$

where

$$\|f\|_p^p = \sum_i |f_i|^p.$$

Here, w_{ij} is the edge weight and the corresponding eigenvalue λ_p is given by $\lambda_p = F_p(f)$. The above theorem serves as the foundational analysis of eigenvectors and eigenvalues. Moreover, $F_p(\alpha f) = F_p(f)$ applies for any real value of α.

Naturally, we can extend the above theorem to the hypergraph p-Laplacian as follows.

f^{hp} is an eigenvector of hypergraph p-Laplacian, if and only if the following function F_p^{hp} has a critical point at f^{hp}:

$$F_p^{\mathrm{hp}}(f^{\mathrm{hp}}) = \frac{\sum_{ij} w_{ij}^{\mathrm{hp}} \left| f_i^{\mathrm{hp}} - f_j^{\mathrm{hp}} \right|^p}{2\|f^{\mathrm{hp}}\|_p^p} \tag{14}$$

where

$$\|f^{\mathrm{hp}}\|_p^p = \sum_i \left| f_i^{\mathrm{hp}} \right|^p.$$

The eigenvalue λ^{hp} associated with f^{hp} is given by $\lambda^{\mathrm{hp}} = F_p^{\mathrm{hp}}(f^{\mathrm{hp}})$.

If we want to obtain all eigenvectors and eigenvalues of hypergraph p-Laplacian, we have to find all critical points of function F_p^{hp}. Following this idea, we can obtain the full eigenvector space by solving local solution of the following optimization problem:

$$\min_{\mathcal{F}^{\mathrm{hp}}} J(\mathcal{F}^{\mathrm{hp}}) = \sum_k F_p^{\mathrm{hp}}(f^{\mathrm{hp}k}) \quad \text{s.t.} \sum_i \phi_p\left(f_i^{\mathrm{hp}k}\right) \phi_p\left(f_i^{\mathrm{hp}l}\right) = 0, \quad k \neq l \tag{15}$$

where $\mathcal{F}^{\mathrm{hp}} = (f^{\mathrm{hp}1}, f^{\mathrm{hp}2}, \ldots, f^{\mathrm{hp}n})$.

We analyze the full eigenvectors by solving the following hypergraph p-Laplacian embedding problem instead of (15):

$$\min_{\mathcal{F}^{\rm hp}} J_E(\mathcal{F}^{\rm hp}) = \sum_k \frac{\sum_{ij} w_{ij}^{\rm hp} \left| f_i^{{\rm hp}k} - f_j^{{\rm hp}k} \right|^p}{\| f^{{\rm hp}k} \|_p^p}$$
$$\text{s.t. } \mathcal{F}^{{\rm hp}T} \mathcal{F}^{\rm hp} = I. \tag{16}$$

Differentiating with respect to $f_i^{{\rm hp}k}$ yields the following equation:

$$\frac{\partial J_E}{\partial f_i^{{\rm hp}k}} = \frac{1}{\| f^{{\rm hp}k} \|_p^p} \left[\sum_j w_{ij}^{\rm hp} \phi_p \left(f_i^{{\rm hp}k} - f_j^{{\rm hp}k} \right) - \frac{\phi_p \left(f_i^{{\rm hp}k} \right)}{\| f^{{\rm hp}k} \|_p^p} \right]. \tag{17}$$

Solving problem (16) with the gradient descend optimization, the gradient is defined in the following way:

$$G^{\rm hp} = \frac{\partial J_E}{\partial \mathcal{F}^{\rm hp}} - \mathcal{F}^{\rm hp} \left(\frac{\partial J_E}{\partial \mathcal{F}^{\rm hp}} \right)^T \mathcal{F}^{\rm hp}. \tag{18}$$

Meanwhile, the full eigenvalue $\lambda^{\rm hp} = (\lambda_1^{\rm hp}, \lambda_2^{\rm hp}, \ldots, \lambda_n^{\rm hp})$ can be computed by

$$\lambda_k^{\rm hp} = \frac{\sum_{ij} w_{ij}^{\rm hp} \left| f_i^{{\rm hp}k} - f_j^{{\rm hp}k} \right|^p}{\| f^{{\rm hp}k} \|_p^p}. \tag{19}$$

Finally, the approximation of $L_p^{\rm hp}$ can be solved by the full eigenvectors and eigenvalues of hypergraph p-Laplacian in this paper. We summarize the approximation of hypergraph p-Laplacian in Algorithm 1. In the algorithm, the step length α is set to $\alpha = 0.01((\sum_{ik} |\mathcal{F}_{ik}^{\rm hp}|)/(\sum_{ik} |G_{ik}^{\rm hp}|))$.

Algorithm 1 Approximation of Hypergraph p-Laplacian

Input: Training samples X; Embedding dimension K; p
output: hypergraph p-Laplacian: L_p^{hp}
Step1: Construct hypergraph Laplacian matrix L^{hp} and data adjacency matrix W^{hp}.
Step2: Decompose graph Laplacian: $L^{hp} = USU^T$.
Initialize: $\mathcal{F}^{hp} = U(:, 1:K)$
Step3: **While** not converged **do:**
$G^{hp} = \frac{\partial J_E}{\partial \mathcal{F}^{hp}} - \mathcal{F}^{hp} \left(\frac{\partial J_E}{\partial \mathcal{F}^{hp}} \right)^T \mathcal{F}^{hp}$, where $\frac{\partial J_E}{\partial \mathcal{F}^{hp}}$ is given by Equation (16)
$\mathcal{F}^{hp} = \mathcal{F}^{hp} - \alpha G^{hp}$
End
Step4: $\lambda_k^{hp} = \frac{\sum_{ij} w_{ij}^{hp} \left| f_i^{hpk} - f_j^{hpk} \right|^p}{\| f^{hpk} \|_p^p}$
return: $L_p^{hp} = \mathcal{F}^{hp} \lambda^{hp} \mathcal{F}^{hpT}$

IV. HpLapR

In SSL, we are given N training samples including l labeled samples $\{(x_i, y_i)\}_{i=1}^l$ and u unlabeled samples $\{(x_j)\}_{j=l+1}^{l+u}$. Class labels are given in $Y = \{y_i\}_{i=1}^l$, where $y_i \in \{\pm 1\}$. Typically, $l \ll u$ and the goal is to predict the labels of unseen samples.

According to the MR framework, the proposed HpLapR can be written as the following optimization problem:

$$f^* = \arg\min_{f \in H_K} \frac{1}{l} \sum_{i=1}^l V(x_i, y_i, f) + \Upsilon_A \| f \|_K^2 + \frac{\Upsilon_I}{(l+u)^2} \mathbf{f}^T L_p^{\rm hp} \mathbf{f}. \tag{20}$$

Here, $\mathbf{f}$ is given as $\mathbf{f} = [f(x_1), f(x_2), \ldots, f(x_{l+u})]^T$, and $L_p^{\rm hp}$ is the hypergraph p-Laplacian.

Next, to solve the optimization problem (20), we discuss the Representer theorem of HpLapR.

Lemma 1: Suppose an arbitrary finite set of points $\{x_1, \ldots, x_n\}$, the kernel function $\boldsymbol{K}(x, y)$ satisfying $\boldsymbol{K}(x, y) = \boldsymbol{K}(y, x)$ is a positive semidefinite kernel, and kernel matrix $\boldsymbol{K}$ with $\boldsymbol{K_{ij}} = \boldsymbol{K}(x_i, x_j)$ is symmetric positive definite. If the penalty term of the optimization problem (1) is a strictly monotonically increasing real-valued function on f, the minimizer of the problem admits an expansion in terms of both labeled and unlabeled samples as

$$f^*(x) = \sum_{i=1}^{l+u} \alpha_i^* \boldsymbol{K}(x_i, x). \tag{21}$$

Lemma 2: The hypergraph p-Laplacian $L_p^{\rm hp}$ is positive semidefinite [48].

Theorem 3: The minimization of the object problem (20) with respect to f has the representation

$$f^*(x) = \sum_{i=1}^{l+u} \alpha_i^* \boldsymbol{K}(x_i, x). \tag{22}$$

Proof: Kernel $\boldsymbol{K}$ and hypergraph p-Laplacian $L_p^{\rm hp}$ are both positive semidefinite. Thus, the regularization term $\Upsilon_A \| f \|_K^2 + (\Upsilon_I/(l+u)^2) \mathbf{f}^T L_p^{\rm hp} \mathbf{f}$ is a monotonically increasing real-value function with respect to f. According to Lemma 1, the proof of Theorem 3 is complete.

The Representer theorem demonstrates that the solution of (20) exists and has the general form of (22).

V. HpLapR Logistic Regression

The proposed HpLapR can be applied to variant applications by integrating different choices of loss function $V(x_i, y_i, f)$ into MR framework. In this section, we apply HpLapR to logistic regression and give its complexity analysis.

Substitute logistic loss function into (20), the HpLapR can be rewritten as

$$f^* = \arg\min_{f \in H_K} \frac{1}{l} \sum_{i=1}^l (\log(1 + e^{-y_i f(x_i)})) + \Upsilon_A \| f \|_K^2 + \frac{\Upsilon_I}{(l+u)^2} \mathbf{f}^T L_p^{\rm hp} \mathbf{f}^T. \tag{23}$$

According to the Representer theorem, the solution of (23) with respect to f exists and can be expressed by (22). Thus, we finally construct the HpLapR as the following optimization

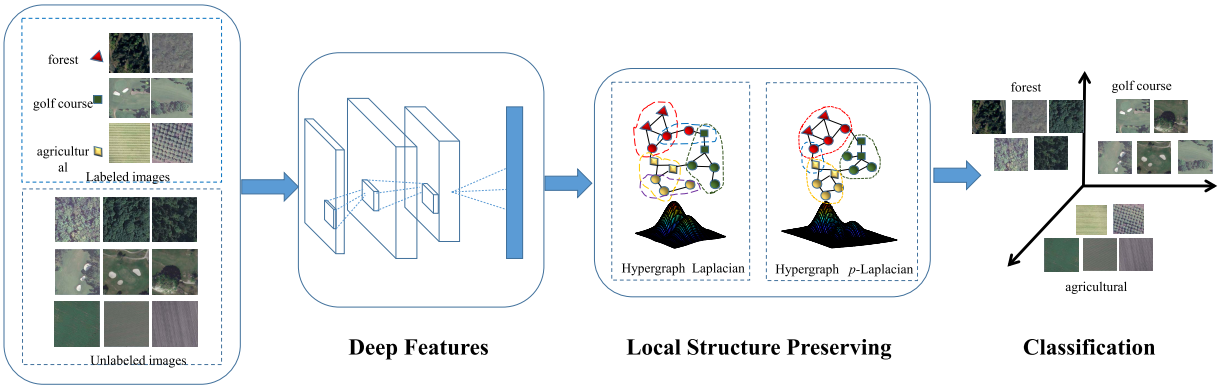


Fig. 2. Framework of HpLapR for remote sensing image classification.

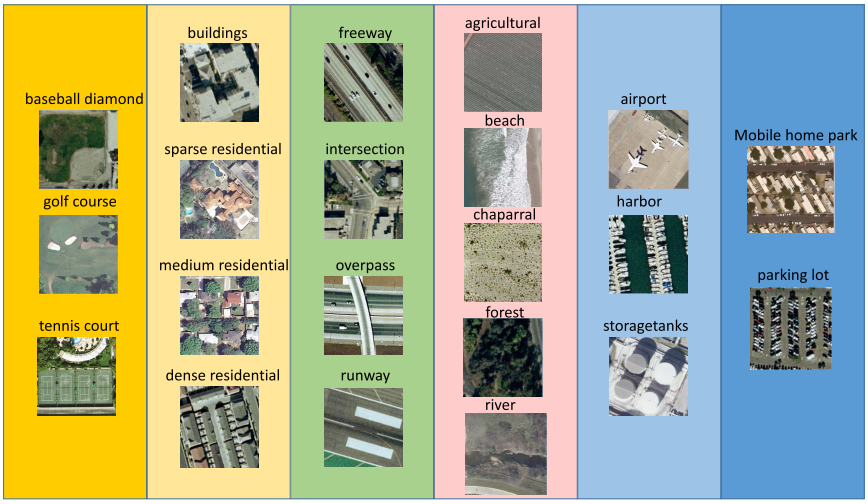


Fig. 3. Class examples of UC-Merced data set. The data set totally has 21 remote sensing categories that can be simply grouped into six groups according to the distinction of land use. Each column represents one group.

problem:

$$f^* = \arg\min_{f\in H_K} \frac{1}{l}\sum_{i=1}^{l}(\log(1+e^{-y_i \boldsymbol{K}(x_i,x)\alpha})) + \Upsilon_A \alpha^T \boldsymbol{K}\alpha + \frac{\Upsilon_I}{(l+u)^2}\alpha^T \boldsymbol{K} L_p^{\mathrm{hp}} \boldsymbol{K}\alpha. \quad (24)$$

To solve the optimization problem in (24), we can employ the conjugate gradient algorithm. We take derivative of the objective function as

$$\nabla f(\alpha) = -\frac{\log(e)}{l}\sum_{i=1}^{l}\left(\frac{y_i}{1+e^{y_i \boldsymbol{K}(x_i,x)\alpha}}\boldsymbol{K}^T(x_i,x)\right) + \Upsilon_A(\boldsymbol{K}+\boldsymbol{K}^T)\alpha + \frac{\Upsilon_I}{(l+u)^2}\left(\boldsymbol{K}L_p^{\mathrm{hp}}\boldsymbol{K} + \left(\boldsymbol{K}L_p^{\mathrm{hp}}\boldsymbol{K}\right)^T\right)\alpha. \quad (25)$$

The optimization procedure of conjugate gradient algorithm for HpLapR logistic regression is described in Algorithm 2.

Suppose that we are given N samples. Denote the embedding dimension as K and the number of iterations as η_1 for approximation of hypergraph p-Laplacian. The time cost for constructing hypergraph p-Laplacian is $\mathrm{O}(\eta_1(N^2K + nK^2))$. When K is much smaller than N, the time cost is around $\mathrm{O}(\eta_1 N^2)$. Denote the number of iterations as η_2 for HpLapR logistic regression and the number of candidate parameters that need the m-fold cross-validation as r. The time cost for HpLapR logistic regression is $\mathrm{O}(\eta_2 r N^3)$.

Algorithm 2 HpLapR Logistic Regression

Input: l labeled samples $\{(x_i, y_i)\}_{i=1}^{l}$,
u unlabeled samples $\{(x_j)\}_{j=l+1}^{l+u}$.
output: Estimated function: $f^*(x) = \sum_{i=1}^{n}\alpha_i^* \boldsymbol{K}(x_i, x)$.
Step1: Construct approximate Hypergraph p-Lapalcian L_p^{hp}.
Step2: Choose a kernel function and compute the Gram matrix $\boldsymbol{K}_{ij} = \boldsymbol{K}(x_i, x_j)$.
Step3: Compute α^*:
Initialize: $\alpha^0 \in \boldsymbol{R}^N$, $d^0 = -\nabla f(\alpha)$, δ,$0 < \varepsilon \ll 1$, $m = 0$
while $\left|f\left(\alpha^{m+1}\right) - f\left(\alpha^m\right)\right| > \varepsilon$
do:
$\alpha^{m+1} = \alpha^m + \delta d^m$
$d^{m+1} = -\nabla f\left(\alpha^{m+1}\right) + \frac{\|\nabla f(\alpha^{m+1})\|^2}{\|\nabla f(\alpha^m)\|^2} d^m$
$m = m + 1$
return: $\alpha^* = \alpha^{m+1}$

VI. Experiments

In this section, to evaluate the effectiveness of the proposed HpLapR, we compare HpLapR with other local structure preserving algorithms including LapR, HLapR, and pLapR. We apply the logistic regression for remote sensing image

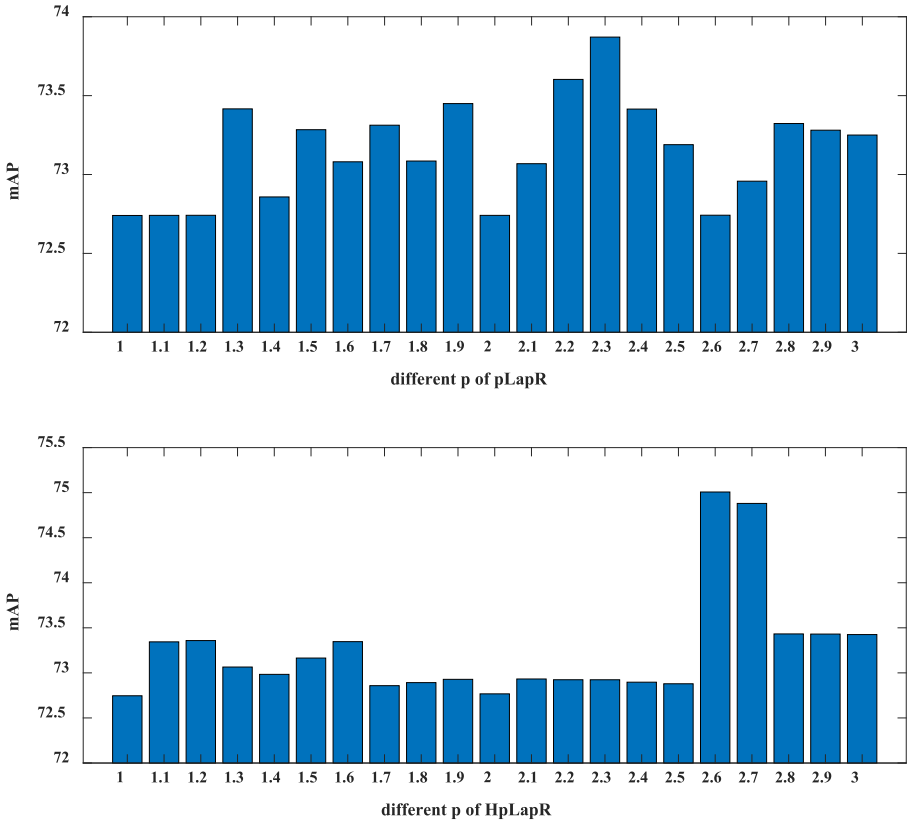


Fig. 4. Performance of mAP with different p values on validation set.

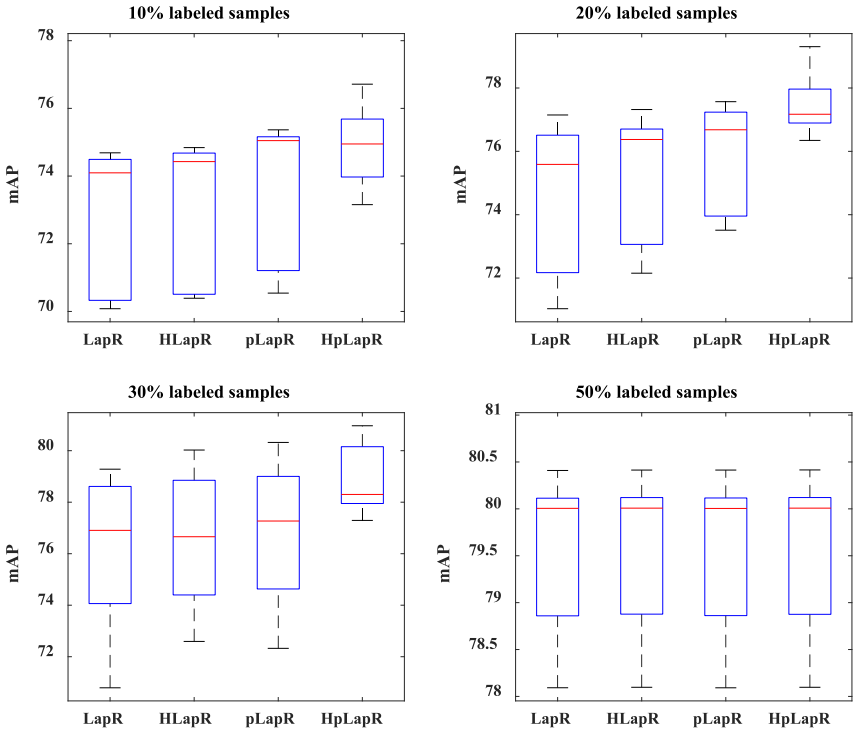


Fig. 5. mAP performance of different algorithms.

classification. Fig. 2 illustrates the framework of HpLapR for UC-Merced data set.

UC-Merced data set [42] consists of totally 2100 land-use images collected from aerial orthoimage with the pixel resolution of one foot. The original images were downloaded from the United States Geological Survey National Map of 20 U.S. regions. These images were manually selected into 21 classes: agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium density residential,

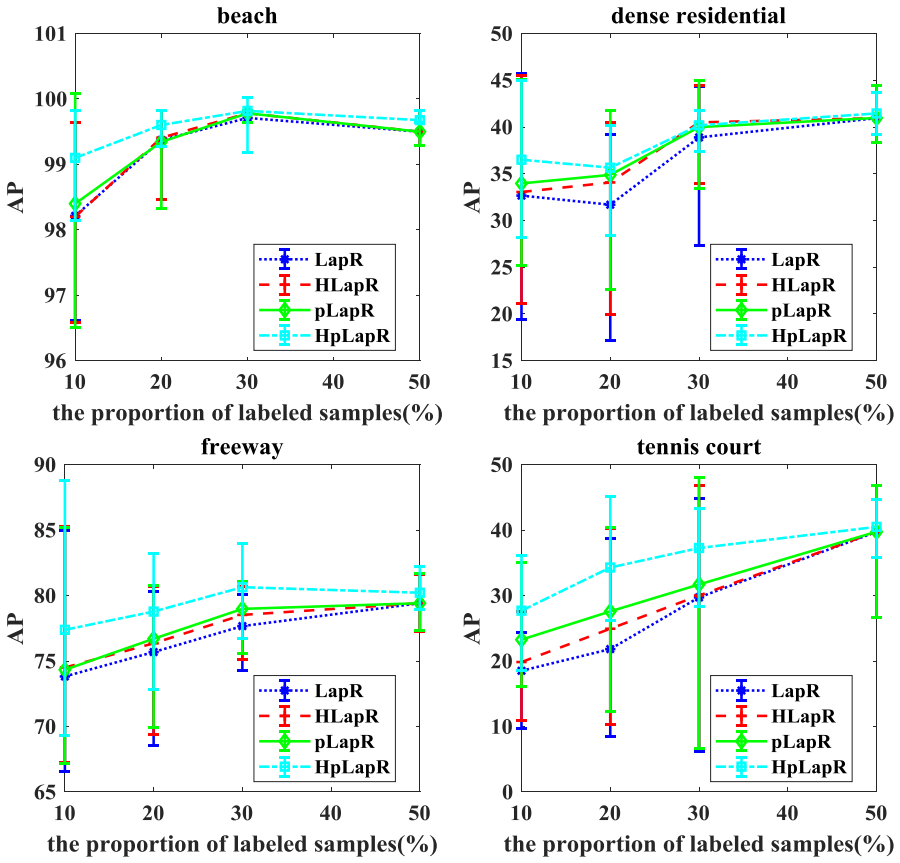


Fig. 6. AP performance of different methods on several classes.

mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis courts. In this paper, we organize these 21 classes into six groups (see Fig. 3). Note that UC-Merced data set contains a variety of land-use classes, which make the data set quite challenging, especially several highly overlapped classes, e.g., sparse residential, medium density residential, and dense residential. They mainly differ in the density.

In our experiments, we extract high-level visual features using the deep convolutional neural network [49]. We randomly choose 50 images per class as training samples and the rest as testing samples. For hypergraph construction, we regard each sample in the training set as a vertex, and generate a hyperedge for each vertex with its k nearest neighbors (so the hyperedge connects $k+1$ samples) [34]. It is worth noting that for our experiments, the kNN-based hyperedges generating method is implemented only in six groups, not in the overall training samples. For example, for a sample of baseball diamond, the vertices of the corresponding hyperedge are chosen from the first group (baseball diamond, golf course, and tennis courts) of Fig. 3.

In semisupervised classification experiments, we assign 10%, 20%, 30%, and 50% samples of training data as labeled data; the rest are used as unlabeled data. The process is repeated five times independently to avoid any bias introduced by the random partition of data.

We conduct the experiments on the data set to obtain the proper modal parameters. The neighborhood size k of a hypergraph varies in a range $\{5, 6, 7, \ldots, 15\}$ through

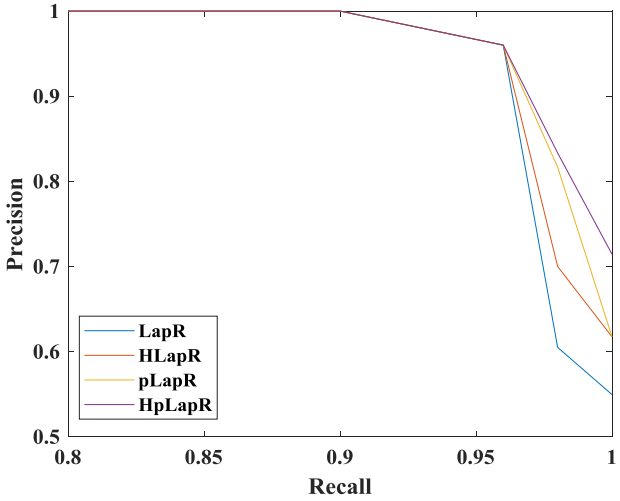
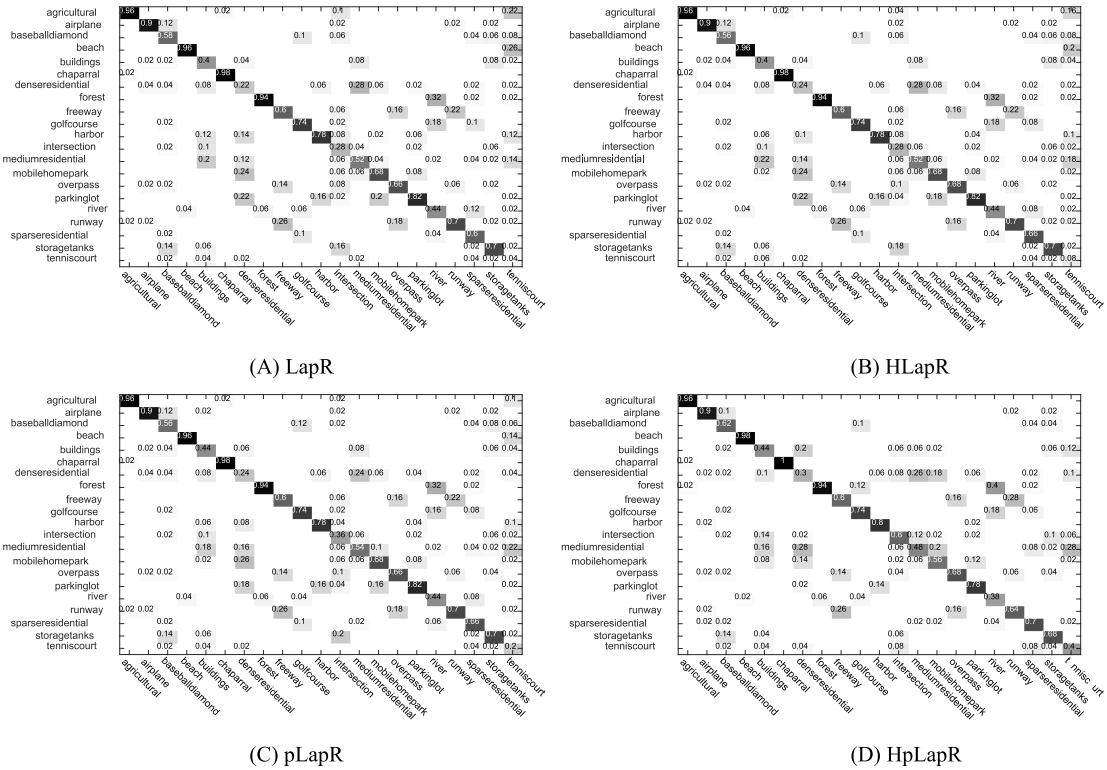


Fig. 7. PR curve of different methods with 10% labeled samples.

cross-validation. Regularization parameters γ_A and γ_I are selected from the candidate set $\{10^i \mid i = -10, -9, -8, \ldots, 10\}$ through cross-validation, and parameter p for pLapR and HpLapR is chosen from $\{1, 1.1, 1.2, \ldots, 3\}$ through cross-validation with 10% labeled samples on the training data, respectively. We verify the classification performance by average precision (AP) performance for single class and mean AP (mAP) [50] for overall classes. The AP is defined as the mean precision at a set of 11 equally



(A) LapR (B) HLapR

(C) pLapR (D) HpLapR

Fig. 8. Confusion matrices of different methods with 10% labeled samples. (a) LapR. (b) HLapR. (c) pLapR. (d) HpLapR.

spaced recall levels and can be expressed as follows:

$$\mathrm{AP} = \frac{1}{11}\sum_{t}\left[\max_{r\geq t}\mathrm{pre}(r)\right], \quad t \in \{0, 0.1, 0.2, \ldots, 1.0\} \tag{26}$$

where $\mathrm{pre}(r)$ is the precision at recall r. The mAP is the mean AP over all remote sensing image classes and can be written as

$$\mathrm{mAP} = \frac{\sum_{i=1}^{c}\mathrm{AP}_i}{c} \tag{27}$$

where c is the number of remote sensing image classes.

Fig. 4 illustrates the mAP performance of pLapR and HpLapR on the validation set when p varies. The x-axis is parameter p and the y-axis is mAP for performance measure. We can see that the best mAP performance for pLapR can be obtained when $p = 2.3$, while the best performance of HpLapR is achieved when p is equal to 2.6.

We compare our proposed HpLapR with the representative LapR, HLapR, and pLapR. From Fig. 5, we can observe that HpLapR outperforms other methods especially when only a small number of samples are labeled. This suggests that our proposed method has the superiority to preserve the local structure of the data because it integrates hypergraph learning with p-Laplacian.

To evaluate the effectiveness of HpLapR for single class, Fig. 6 shows the AP results of different methods on several selected land-use classes including beach, dense residential, freeway, and tennis court. From Fig. 6, we can find that in most cases, HpLapR performs better than do both pLapR and HLapR, while pLapR and HLapR consistently outperforms than LapR.

Moreover, we show the precision–recall (PR) curves and the confusion matrices of different methods with the 10% labeled samples in Figs. 7 and 8, respectively. We can find that our proposed HpLapR can obtain the better performance for the most classes in comparison to other methods. We compute the f-measure values of LapR, HLapR, pLapR, and HpLapR. The HpLapR can get the higher f-measure value (0.695859) while those of LapR, HLapR, and pLapR are 0.674872, 0.677027, and 0.679709, respectively.

VII. Conclusion

Existing SSL algorithms have achieved great performance in computer vision applications including classification, clustering, and ranking. However, it is still challenging on how to obtain the high-order relationship while exploiting the local geometry of the data distribution. Therefore, after introducing a full approximation algorithm of hypergraph p-Laplacian to significantly low down its computation complexity, this paper has proposed an HpLapR method to preserve the geometry of the probability distribution. This is because hypergraph and p-Laplacian have the advantage of preserving local structures. Furthermore, we proposed HpLapR logistic regression for remote sensing recognition. The experimental results on UC-Merced data set have demonstrated the effectiveness of our proposed method in comparison to other regularized methods including LapR, HLapR, and pLapR.

References

[1] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1452–1464, Jun. 2018.

[2] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. CVPR*, New York, NY, USA, Jun. 2006, pp. 2169–2178.

[3] W. Liu, X. Ma, Y. Zhou, D. Tao, and J. Cheng, "p-Laplacian regularization for scene recognition," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2018.2833843.

[4] J. Yu, D. Tao, and M. Wang, "Adaptive hypergraph learning and its application in image classification," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3262–3272, Jul. 2012.
[5] M. Alkanhal and G. Muhammad, "Polynomial correlation filters for human face recognition," in *Proc. 11th Int. Conf. Mach. Learn. Appl.*, Boca Raton, FL, USA, Dec. 2012, pp. 646–650.
[6] W. Liu, Z.-J. Zha, Y. Wang, K. Lu, and D. Tao, "p-Laplacian regularized sparse coding for human activity recognition," *IEEE Trans. Ind. Electron.*, vol. 63, no. 8, pp. 5120–5129, Aug. 2016.
[7] Y. Guo, D. Tao, W. Liu, and J. Cheng, "Multiview cauchy estimator feature embedding for depth and inertial sensor-based human action recognition," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 4, pp. 617–627, Apr. 2017.
[8] P. Wang, W. Li, Z. Gao, C. Tang, and P. O. Ogunbona, "Depth pooling based large-scale 3-D action recognition with convolutional neural networks," *IEEE Trans. Multimedia*, vol. 20, no. 5, pp. 1051–1061, May 2018.
[9] X. Lu, X. Zheng, and Y. Yuan, "Remote sensing scene classification by unsupervised representation learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5148–5157, Sep. 2017.
[10] G.-S. Xia *et al.*, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.
[11] S. Niazmardi, B. Demir, L. Bruzzone, A. Safari, and S. Homayouni, "Multiple kernel learning for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1425–1443, Mar. 2018.
[12] W. Li and Q. Guo, "A new accuracy assessment method for one-class remote sensing classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4621–4632, Aug. 2014.
[13] G. Cheng, J. Han, L. Guo, Z. Liu, S. Bu, and J. Ren, "Effective and efficient midlevel visual elements-oriented land-use classification using VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4238–4249, Aug. 2015.
[14] F. Hu, G.-S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14680–14707, 2015.
[15] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Nov. 2006.
[16] F. Nie, D. Xu, I. W.-H. Tsang, and C. Zhang, "Flexible manifold embedding: A framework for semi-supervised and unsupervised dimension reduction," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1921–1932, Jul. 2010.
[17] M. Belkin, I. Matveeva, and P. Niyogi, "Regularization and semi-supervised learning on large graphs," in *Proc. COLT*, 2004, pp. 624–638.
[18] J. Jiang, R. Hu, Z. Wang, and Z. Cai, "CDMMA: Coupled discriminant multi-manifold analysis for matching low-resolution face images," *Signal Process.*, vol. 124, pp. 162–172, 2016.
[19] Y. Luo, D. Tao, B. Geng, C. Xu, and S. J. Maybank, "Manifold regularized multitask learning for semi-supervised multilabel image classification," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 523–536, Feb. 2013.
[20] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.
[21] K. N. Ramamurthy, J. J. Thiagarajan, and A. Spanias, "Improved sparse coding using manifold projections," in *Proc. IEEE Int. Conf. Image Process*, Brussels, Belgium, Sep. 2011, pp. 1237–1240.
[22] X. Lu, Y. Wang, and Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4009–4018, Jul. 2013.
[23] S. Zhang, Y. Sui, S. Zhao, and L. Zhang, "Graph-regularized structured support vector machine for object tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 6, pp. 1249–1262, Jun. 2017.
[24] B. Chaudhuri, B. Demir, S. Chaudhuri, and L. Bruzzone, "Multilabel remote sensing image retrieval using a semisupervised graph-theoretic method," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 1144–1158, Feb. 2018.
[25] H. H. Shin, N. J. Hill, and G. Rätsch, "Graph based semi-supervised learning with sharper edges," in *Proc. ECML*. Berlin, Germany, 2006, pp. 401–412.
[26] X. Pei, C. Chen, and Y. Guan, "Joint sparse representation and embedding propagation learning: A framework for graph-based semisupervised learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2949–2960, Dec. 2017.
[27] L. Wang, D. Li, T. He, and Z. Xue, "Manifold regularized multi-view subspace clustering for image representation," in *Proc. 23rd ICPR*, Cancún, Mexico, Dec. 2016, pp. 283–288.
[28] W. Liu and D. Tao, "Multiview Hessian regularization for image annotation," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2676–2687, Jul. 2013.
[29] X. Lu, X. Li, and L. Mou, "Semi-supervised multitask learning for scene recognition," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1967–1976, Sep. 2015.
[30] D. Zhou, J. Huang, and B. Schölkopf, "Learning from labeled and unlabeled data on a directed graph," in *Proc. ICML*, 2005, pp. 1036–1043.
[31] S. Agarwal, J. Lim, L. Zelnik-Manor, P. Perona, D. Kriegman, and S. Belongie, "Beyond pairwise clustering," in *Proc. CVPR*, Jun. 2005, pp. 838–845.
[32] R. Ji, Y. Gao, R. Hong, Q. Liu, D. Tao, and X. Li, "Spectral-spatial constraint hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1811–1824, Mar. 2014.
[33] J. Yu, Y. Rui, and D. Tao, "Click prediction for Web image reranking using multimodal sparse coding," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2019–2032, May 2014.
[34] Y. Huang, Q. Liu, S. Zhang, and D. N. Metaxas, "Image retrieval via probabilistic hypergraph ranking," in *Proc. CVPR*, San Francisco, CA, USA, Jun. 2010, pp. 3376–3383.
[35] Y. Huang, Q. Liu, and D. Metaxas, "Video object segmentation by hypergraph cut," in *Proc. CVPR*, Miami, FL, USA, Jun. 2009, pp. 1738–1745.
[36] L. Sun, S. Ji, and J. Ye, "Hypergraph spectral learning for multi-label classification," in *Proc. KDD*, Las Vegas, NV, USA, 2008, pp. 668–676.
[37] R. Zass and A. Shashua, "Probabilistic graph and hypergraph matching," in *Proc. CVPR*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.
[38] T. Bühler and M. Hein, "Spectral clustering based on the graph p-Laplacian," in *Proc. ICML*, Montreal, QC, Canada, 2009, pp. 81–88.
[39] D. Zhou and B. Schölkopf, "Regularization on discrete spaces," in *Proc. Pattern Recognit.*, 2005, pp. 361–368.
[40] H. Takeuchi, "The spectrum of the p-Laplacian and p-harmonic morphisms on graph," *Illinois J. Math.*, vol. 47, no. 3, pp. 939–955, 2003.
[41] S. Amghibech, "Bounds for the largest p-Laplacian eigenvalue for graphs," *Discrete Math.*, vol. 306, no. 21, pp. 2762–2771, 2006.
[42] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. GIS*, San Jose, CA, USA, 2010, pp. 270–279.
[43] D. Luo, H. Huang, C. Ding, and F. Nie, "On the eigenvectors of p-Laplacian," *Mach. Learn.*, vol. 81, no. 1, pp. 37–51, 2010.
[44] J. Y. Zien, M. D. F. Schlag, and P. K. Chan, "Multilevel spectral hypergraph partitioning with arbitrary vertex sizes," in *Proc. ICCAD*, San Jose, CA, USA, Nov. 1996, pp. 201–204.
[45] J. A. Rodriguez, "On the Laplacian spectrum and walk-regular hypergraphs," *Linear Multilinear Algebra*, vol. 51, no. 3, pp. 285–297, 2003.
[46] M. Bolla, "Spectra, Euclidean representations and clusterings of hypergraphs," *Discrete Math.*, vol. 117, pp. 19–39, Jul. 1993.
[47] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Proc. NIPS*, Vancouver, BC, Canada, 2006, pp. 1601–1608.
[48] S. Saito, D. P. Mandic, and H. Suzuki. (2017). "Hypergraph p-Laplacian: A differential geometry view." [Online]. Available: https://arxiv.org/abs/1711.08171
[49] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[50] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Sep. 2009.

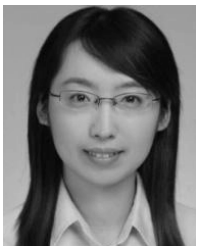
Xueqi Ma is currently pursuing the master's degree with the College of Information and Control Engineering, China University of Petroleum, Qingdao, China.

Her research interests include pattern recognition and computer vision.

Weifeng Liu (M'12–SM'17) received the double B.S. degrees in automation and business administration and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2002 and 2007, respectively.

He was a Visiting Scholar with the Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW, Australia, from 2011 to 2012. He is currently a Full Professor with the College of Information and Control Engineering, China University of Petroleum, Qingdao, China. He has authored or co-authored a dozen papers in top journals and prestigious conferences, including four Essential Science Indicators (ESI) highly cited papers and two ESI hot papers. His research interests include computer vision, pattern recognition, and machine learning.

Dr. Liu serves as an Associate Editor for the *Neural Processing Letters*, the Co-Chair for the IEEE SMC Technical Committee on Cognitive Computing, and a Guest Editor for special issue of the *Signal Processing*, the*IET Computer Vision*, the *Neurocomputing*, and the *Remote Sensing*. He also serves over 20 journals and over 40 conferences.



Shuying Li received the B.E. degree from the University of Science and Technology of China, Hefei, China, and the Ph.D. degree from the Chinese Academy of Sciences, Beijing, China.

She is currently a Professor with the School of Automation, Xi'an University of Posts and Telecommunications, Xi'an, China. Her research interests include remote sensing, computer vision, and pattern recognition.

Dapeng Tao received the B.E. degree from Northwestern Polytechnical University, Xi'an, China, and the Ph.D. degree from the South China University of Technology, Guangzhou, China.

He is currently an Engineer with the School of Information Science and Engineering, Yunnan University, Kunming, China. He has authored or co-authored over 30 scientific articles. His research interests include machine learning, computer vision, and cloud computing.

Dr. Tao has served over 10 international journals, including the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE SIGNAL PROCESSING LETTERS, and the *PLoS ONE*.

Yicong Zhou (M'07–SM'14) received the B.S. degree in electrical engineering from Hunan University, Changsha, China, and the M.S. and Ph.D. degrees in electrical engineering from Tufts University, Medford, MA, USA.

He is currently an Associate Professor and the Director with the Vision and Image Processing Laboratory, Department of Computer and Information Science, University of Macau, Macau, China. His research interests include chaotic systems, multimedia security, image processing and understanding, and machine learning.

Dr. Zhou was a recipient of the Third Price of Macau Natural Science Award in 2014. He served as an Associate Editor for the *Neurocomputing*, the *Journal of Visual Communication and Image Representation*, and the *Signal Processing: Image Communication*. He is a Co-Chair of Technical Committee on Cognitive Computing in the IEEE Systems, Man, and Cybernetics Society.