We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Recent Advances of Manifold Regularization

Xueqi Ma and Weifeng Liu

Additional information is available at the end of the chapter

Abstract

Semi-supervised learning (SSL) that can make use of a small number of labeled data with a large number of unlabeled data to produce significant improvement in learning performance has been received considerable attention. Manifold regularization is one of the most popular works that exploits the geometry of the probability distribution that generates the data and incorporates them as regularization terms. There are many representative works of manifold regularization including Laplacian regularization (LapR), Hessian regularization (HesR) and p-Laplacian regularization (pLapR). Based on the manifold regularization framework, many extensions and applications have been reported. In the chapter, we review the LapR and HesR, and we introduce an approximation algorithm of graph p-Laplacian. We study several extensions of this framework for pairwise constraint, p-Laplacian learning, hypergraph learning, etc.

Keywords: Laplacian regularization, Hessian regularization, p-Laplacian regularization, semi-supervised learning, manifold learning

1. Introduction

In practical applications, it is generally laborious to obtain the labeled samples, though vast amounts of unlabeled samples are easily achieved and provide auxiliary information. Semi-supervised learning (SSL), which takes the full advantages of unlabeled data, is specifically designed to improve learning performance. In representative semi-supervised learning algorithms, it is usually assumed that the intrinsic geometry of the data distribution is supported on the low-dimensional manifold.

The popular manifold learning methods include principal components analysis (PCA), multidimensional scaling (MDS) [1, 2], generative topological mapping (GTM) [3], locally linear embedding (LLE) [4], ISOMAP [5], Laplacian eigenmaps (LE) [6], Hessian eigenmaps

(HLLE) [7], and local tangent space alignment (LTSA) [8]. PCA aims to find the low-dimensional linear subspace which captures the maximum proportion of the variation within the data. MDS aims to place each object in N-dimensional space such that the between-object distances are preserved as well as possible. GTM can be seen as a nonlinear form of principal component analysis or factor analysis. LLE assumes a given sample can be reconstructed by its neighbors, represents the local geometry and then seeks a low-dimensional embedding. ISOMAP incorporates the geodesic distances imposed by a weighted graph. LE preserves neighbor relations of pairwise samples by manipulations on an undirected weighted graph. HLLE obtains the final low-dimensional representations by applying eigenanalysis to a matrix, which is built by estimating the Hessian over neighborhood. LTSA [8] exploits the local tangent information as a representation of the local geometry, and this local tangent information is then aligned to provide a global coordinate. Regularization is a key idea in the theory of splines [9] and is widely used in machine learning [10] (e.g., support vector machines). In 2006, Belkin et al. [11] proposed the manifold regularization framework by introducing a new regularization term to exploit the geometry of the probability distribution. Based on this framework, many successful manifold regularized semi-supervised learning (MRSSL) algorithms have been reported.

Laplacian regularization (LapR) [11, 12] is one prominent manifold regularization-based SSL algorithm, which approximates the manifold by using the graph Laplacian. Putting the simple calculation and prominent performance together, the LapR-based SSL algorithms have been widely used in many applications. Liu et al. [13] introduced Laplacian regularization for local structure preserving and proposed manifold regularized kernel logistic regression (KLR) for web image annotation. Luo et al. [14] employed manifold regularization to smooth the functions along the data manifold for multitask learning. Ma et al. [15] proposed a local structure preserving method that effectively integrates Laplacian regularization and pairwise constraints for human action recognition. Hu et al. [16] introduced graph Laplacian regularization for joint denoising and superresolution of generalized piecewise smooth images.

Hessian regularization [17] (HesR) has attracted considerable attentions and has shown empirically to perform well in practical problems [18–26]. Liu et al. [27] incorporated both Hessian regularization and sparsity constraints into auto-encoders and proposed a new auto-encoder algorithm called Hessian regularized sparse auto-encoders (HSAE). Liu et al. [28] proposed multi-view Hessian regularized logistic regression for action recognition. While the null space of the graph Laplacian along the underlying manifold is a constant function, HesR steers the learned function varying linearly in reference to the geodesic distance. In result, HesR can be more accurate to describe the underlying manifold of data and achieves the better learning performance than LapR-based ones [18]. However, the stability of Hessian estimation depends mostly on the quality of the local fit for each data point, which leads to inaccurate estimation particularly when the function is heavily oscillating [17].

As a nonlinear generalization of the standard graph Laplacian, discrete p-Laplacian has been well studied in mathematics community and solid properties have been investigated by previous work [29, 30]. Meanwhile, graph p-Laplacian has been proved having the advantages for exploiting the manifold of data distribution. Bühler et al. [31] provided a rigorous proof of the approximation of the second eigenvector of p-Laplacian to the Cheeger cut which indicates the

superiority of graph p-Laplacian in local geometry exploiting. Luo et al. [32] proposed full eigenvector analysis of p-Laplacian and obtain a natural global embedding for multi-class clustering problems, instead of using greedy search strategy implemented by previous researchers. Liu et al. [33] proposed p-Laplacian regularized sparse coding for human activity recognition.

In this chapter, we first present some related work, and then introduce several extensions based on the manifold regularization framework. Specifically, we present the approximation of graph p-Laplacian and the p-Laplacian regularization framework.

Notations: We present some notations that will be used throughout this chapter. We use $L^{''}$ as the novel graph Laplacian constructed by the traditional graph Laplacian L and the side information. L_p, L_p^{hp} and L represent the graph p-Laplacian, hypergraph p-Laplacian and ensemble graph p-Laplacian, respectively.

2. Related works

This section reviews some related works on manifold regularization, pairwise constraints and hypergraph learning.

2.1. Manifold regularization framework

In semi-supervised learning, assume that N training samples X containing l labeled samples $\left\{ \left(x_i, y_i \right) \right\}_{i=1}^{l}$ and u unlabeled samples $\left\{ \left(x_j \right) \right\}_{j=l+1}^{l+u}$ are available. The labeled samples are pairs generated from probability distribution, while unlabeled samples are simply drawn according to the marginal distribution. To utilize marginal distribution induced by unlabeled samples, we assume that if two points x_1, x_2 are close in the intrinsic geometry of marginal distribution, then the labels of x_1 and x_2 are similar.

Manifold regularized method introduces appropriate penalty term $\left(\|f\|_I^2 \right)$ and reproducing kernel Hilbert spaces (RKHS) norm $\left(\|f\|_K^2 \right)$ that is used to control the complexity of the intrinsic geometric structure of the function and the complexity of the classification model, respectively. By incorporating two regularization terms, the standard framework aims to minimize the following function:

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

where V is some loss function, such as the hinge loss function $max\left[0, 1 - y_i f(x_i) \right]$ for support vector machines (SVM). The parameters Υ_A and Υ_I balance the loss function and two regularization terms. For semi-supervised learning, the manifold regularization term $\|f\|_I^2$ is a key to smooth function along the manifold estimated from the unlabeled samples.

2.2. Pairwise constraints

Pairwise constraints (side information) [34, 35] is a type of supervised information that specify whether a pair of data samples belong to the same class (must-link constraints) or different classes (cannot-link constraints). Compared with class labels, pairwise constraints can provide us weak and more general supervised information. Currently, it has been widely used in semi-supervised clustering [36, 37], distance metric learning [38], feature selection [39] and dimension reduction [40, 41].

Donate $X = \{x_i\}_{i=1}^{n}$ as data set with $Y = \{y_i\}_{i=1}^{n}$ as class labels. Let $M = \{(x_i, x_j)\}$ be the pairwise must-link constraints set and $C = \{(x_i, x_j)\}$ be the pairwise cannot-link constraints set, that is,

$$M = \{(x_i, x_j) | x_i \text{ and } x_j \text{ belong to the same class}\}$$

$$C = \{(x_i, x_j) | x_i \text{ and } x_j \text{ belong to different classes}\}.$$

Defined on the pairwise must-link constraint set and the cannot-link constraint set, we construct similarity matrices S^M and S^C, respectively:

$$S_{ij}^M = \begin{cases} 1, & if (x_i, x_j) \in M \\ 0, & otherwise \end{cases} \tag{2}$$

$$S_{ij}^C = \begin{cases} 1, & if (x_i, x_j) \in C \\ 0, & otherwise \end{cases}. \tag{3}$$

Then, the must-link Laplacian matrix L^M is given by $L^M = D^M - S^M$, and the cannot-link Laplacian matrix L^C is given by $L^C = D^C - S^C$. Where D^M and D^C are two diagonal matrices with $D_{ii}^M = \sum_{j=1}^{n} S_{ij}^M$ and $D_{ii}^C = \sum_{j=1}^{n} S_{ij}^C$, respectively.

Ding et al. [42] introduced pairwise constraints into spectral clustering algorithm. Especially, they revised the distances between sample points by the distance matrix D, where $D_{ij} = \begin{cases} 0 & if (x_i, x_j) \in M \\ \infty & if (x_i, x_j) \in C \end{cases}$.

Kalakech et al. [43] developed a semi-supervised constraint score by using both pairwise constraints and local properties of the unlabeled data.

Luo et al. [44] denoted the training set with side information by $\{x_i, x_j, y_{ij}\}_{i,j=1}^{N}$, where $y_{ij} = \pm 1$ indicates x_i and x_j are similar or dissimilar. The side information was utilized by denoting the loss function $y_{ij} \left[1 - \|x_i - x_j\|_{A_m}^2\right]$, where A_m is the metric in the m'th heterogeneous domain.

2.3. Hypergraph learning

Hypergraph [45] is a generalization of a simple graph. Compared with simple graphs, a hypergraph illustrates the complex relationship by hyperedges that connect three or more

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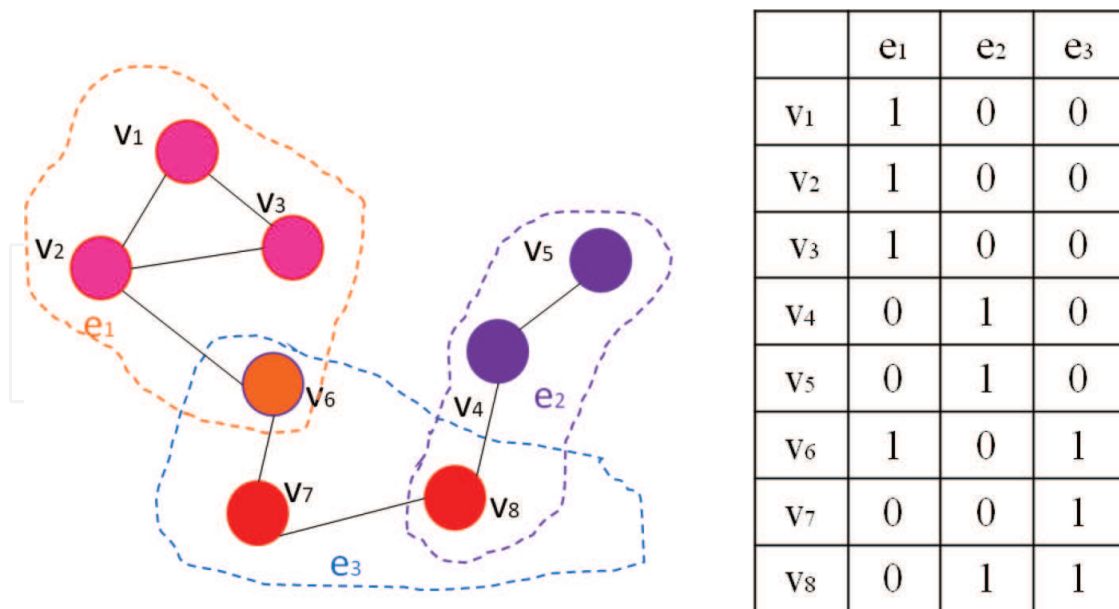
Figure 1. The block scheme of hypergraph. Left: A simple graph in which two points are joined together by an edge if they are highly similar. A hypergraph completely illustrates the complex relationship among points by hyperedges. Right: The H matrix of the hypergraph. The entry (v_i, e_j) is set to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

vertices (see in **Figure 1**). Thus, the hypergraph contains more local structure information in comparison to simple graph. Hypergraph has been widely used in image classification [46], ranking [47] and video segmentation [48].

Let $G = (V, E)$ denote a hypergraph with the vertex set V and the hyperedge set E. Denote the weight associated with each hyperedge e as $w(e)$. The degree $d(v)$ of a vertex is defined by $d(v) = \sum_{\{e \in E | v \in e\}} w(e)$. The degree of a hyperedge e is denoted as $\delta(e) = |e|$. Denote the vertex-edge incident matrix H by a $|V| \times |E|$ matrix, where entry $h(v, e) = 1$ if $v \in e$, and $h(v, e) = 0$ otherwise. By these definitions, we have:

$$d(v) = \sum_{e \in E} w(e)h(v, e), \delta(e) = \sum_{v \in V} h(v, e). \tag{4}$$

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

A number of different methods have been used in the literature to build the graph Laplacian of hypergraphs. The first category includes star expansion [49], clique expansion [49], Rodriquez's Laplacian [50], etc. These methods aim to construct a simple graph from the original hypergraph, and then partitioning the vertices by 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 Laplacian [51], Zhou' normalized Laplacian [52], etc. According to [52], the normalized hypergraph Laplacian L^{hp} is defined as

$$L^{hp} = I - Dv^{-1/2}HWD_e^{-1}H^TDv^{-1/2}. \tag{5}$$

It is worth noting that L^{hp} is positive semi-definite. The adjacency matrix of hypergraph W^{hp} can be formulated as follows:

$$W^{hp} = HWH^T - Dv. \tag{6}$$

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

$$\begin{aligned}L &= I - \frac{1}{2}Dv^{-\frac{1}{2}}HWH^TDv^{-\frac{1}{2}} \\ &= \frac{1}{2}\left(I - Dv^{-1/2}W^{hp}Dv^{-1/2}\right).\end{aligned} \tag{7}$$

3. LapR-based SSL

Laplacian regularization is one of most prominent manifold regularization methods that utilizes the graph Laplacian matrix to characterize the manifold structure. In this section, we introduce the traditional Laplacian support vector machines (LapSVM) and Laplacian kernel least squares (LapKLS) as examples of Laplacian regularization algorithms. Then, we extend the algorithms by building the novel graph Laplacian $L^{''}$ which combines the traditional graph Laplacian L with the side information to boost locality preservation.

3.1. LapSVM and LapKLS

As previously mentioned, the manifold regularization framework is built by Eq. (1). The traditional LapSVM solves this optimization problem with the hinge loss function

$$f(*) = arg \min_{f \epsilon H_K} \frac{1}{l}\sum_{i=1}^{l}\left(1 - y_if(x_i)\right)_+ + \Upsilon_A\|f\|_K^2 + \frac{\Upsilon_I}{(l+u)^2}\mathbf{f}^TL\mathbf{f}. \tag{8}$$

where $\mathbf{f}$ is given as $\mathbf{f} = [f(x_1), f(x_2), \cdots, f(x_{l+u})]^T$, L is the graph Laplacian with $L = D - W$, where W_{ij} is weight vector, the diagonal matrix D is given by $D_{ii} = \sum_{j=1}^{n}W_{ij}$.

According to the representer theorem, the solution of the above problem can be expressed as below:

$$f^*(x) = \sum_{i=1}^{l+u}\alpha_i^*K(x_i, x). \tag{9}$$

where K is the kernel function. Therefore, we rewrite the objective function as

$$f(*) = arg \min_{f \epsilon H_K} \frac{1}{l}\sum_{i=1}^{l}\left(1 - y_if(x_i)\right)_+ + \Upsilon_A\alpha^TK\alpha + \frac{\Upsilon_I}{(l+u)^2}\alpha^TKLK\alpha. \tag{10}$$

By employing the least square loss in Eq. (10), we can present the locality preserved kernel least squares model defined in Eq. (11) as follows

$$f(*) = \min_{f \in H_K} \frac{1}{l} \sum_{i=1}^{l} (y_i - f(x_i))^2 + \Upsilon_A \alpha^T K \alpha + \frac{\Upsilon_I}{(l+u)^2} \alpha^T K L K \alpha. \tag{11}$$

Taking the derivation to the objective functions, we can get the solution of α.

3.2. Pairwise constraints-combined manifold regularization

Assume that samples with the similar features tend to have the similar class labels, combining the Laplacian regularization and pairwise constraints is a good way to exploit the local structure and boost the classification results. Therefore, we introduce the pairwise constraints into traditional LapR. Particularly, we introduce three combination strategies based on experiences. Finally, we present the locality preserved support vector machines and kernel least squares respectively.

According to the definition, we can compute the must-link Laplacian matrix L^M and the cannot-link Laplacian matrix L^C. The first two forms of the combination are defined on the traditional graph Laplacian L and must-link constraints and can be written as

$$L'' = L(L^M + \alpha I) \tag{12}$$

and

$$L'' = L + \alpha L^M \tag{13}$$

respectively, where α is the parameter to balance the weight between the two types of Laplacian matrices.

Based on the cannot-link constraints C, we can compute the similarity matrix S as $S_{ij} = \begin{cases} -1, & if (x_i, x_j) \in C \\ 1, & otherwise \end{cases}$. The third form of the combination is defined on the traditional graph Laplacian and pairwise cannot-link constraints and can be written as

$$L'' = L.*S. \tag{14}$$

Actually, there are other combination strategies using both the must-link and cannot-link constraints to get a better result than traditional methods. However, the performance is no better than the result using one only from the experiences. Therefore, we just put these three proposed graph Laplacian into practice.

Introducing the novel graph Laplacian L'' to SVM, we rewrite the learning model as follows:

$$f(*) = \arg \min_{f \in H_K} \frac{1}{l} \sum_{i=1}^{l} (1 - y_i f(x_i))_+ + \Upsilon_A \|f\|_K^2 + \frac{\Upsilon_I}{(l+u)^2} \mathbf{f}^T L'' \mathbf{f}. \tag{15}$$

According to the representer theorem, the solution of the above problem can be expressed as below:

$$f^*(x) = \sum_{i=1}^{l+u} \alpha_i^* \mathbf{K}(x_i, x). \tag{16}$$

Therefore, we rewrite the objective function as

$$f(*) = \arg \min_{f \epsilon H_K} \frac{1}{l} \sum_{i=1}^{l} \left(1 - y_i f(x_i)\right)_+ + \Upsilon_A \alpha^T \mathbf{K} \alpha + \frac{\Upsilon_I}{(l+u)^2} \alpha^T \mathbf{K} L'' \mathbf{K} \alpha. \tag{17}$$

By employing the least square loss in Eq. (17), we can present the locality preserved kernel least squares model defined in Eq. (18) as follows

$$f(*) = \min_{f \epsilon H_K} \frac{1}{l} \sum_{i=1}^{l} \left(y_i - f(x_i)\right)^2 + \Upsilon_A \Upsilon_A \alpha^T \mathbf{K} \alpha + \frac{\Upsilon_I}{(l+u)^2} \alpha^T \mathbf{K} L'' \mathbf{K} \alpha. \tag{18}$$

We compare our proposed local structure preserving algorithms with the traditional well-known Laplacian algorithms on CAS-YNU-MHAD dataset [53]. CAS-YNU-MHAD dataset contains 10 human actions including jumping up, jumping forward, running, walking S, walking quickly, walking, standing up, sitting down, lying down and typing. **Figure 2** shows the examples. In experiments, we choose the data from four sensors (be placed in the right shoulder, left forearm, left hand and spine) to construct multi-view features. Ninety percent data of per action are randomly selected as the training data, and the rest for testing.

In semi-supervised classification experiments, we randomly select a certain percentage (10, 20, 30, 50%) samples of training data as labeled data. All the classification methods are measured by the average precision (AP) [54] based on the testing data. Note that the supervised information

Figure 2. Three examples from 10 actions, jumping up, walking S and sitting down (up to bottom).
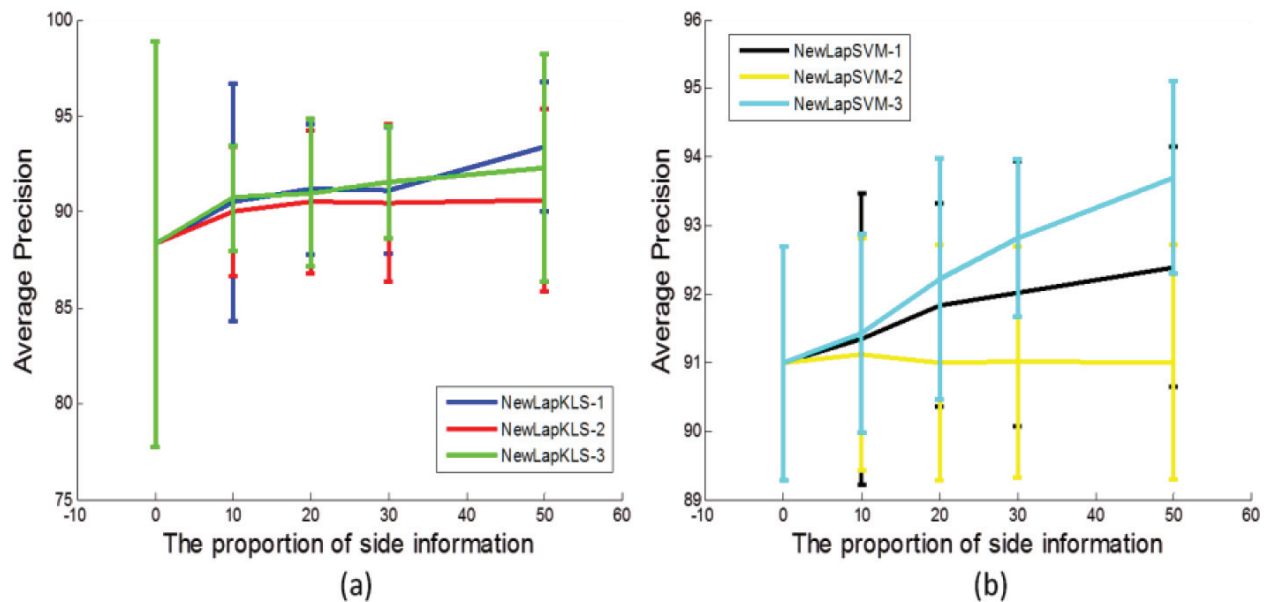
Figure 3. The total classification result under 10% labels (a) KLS, (b) SVM.
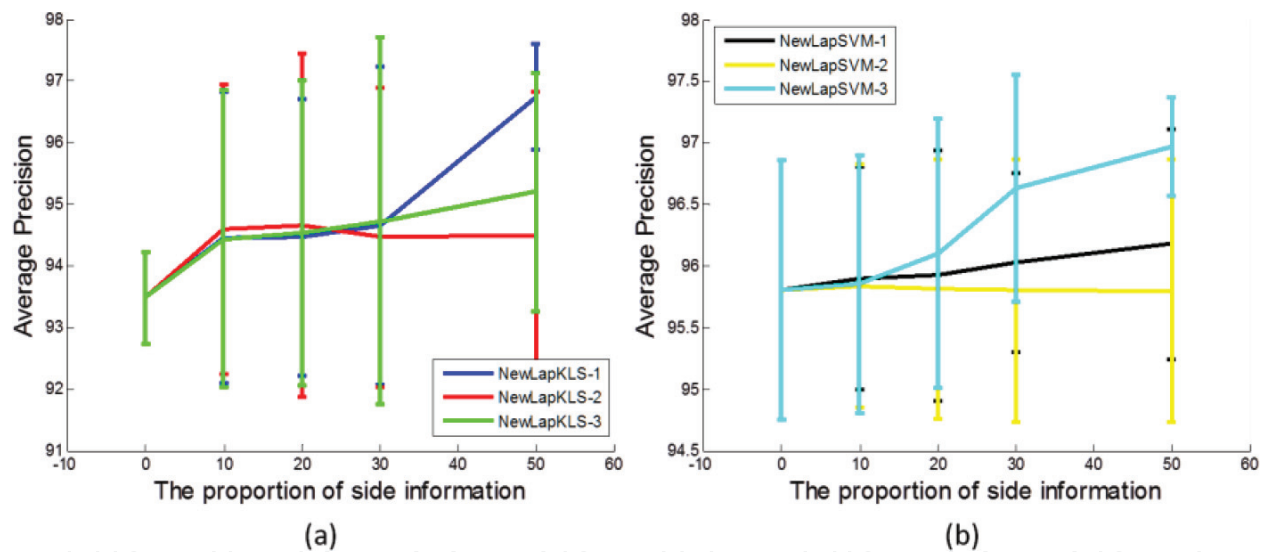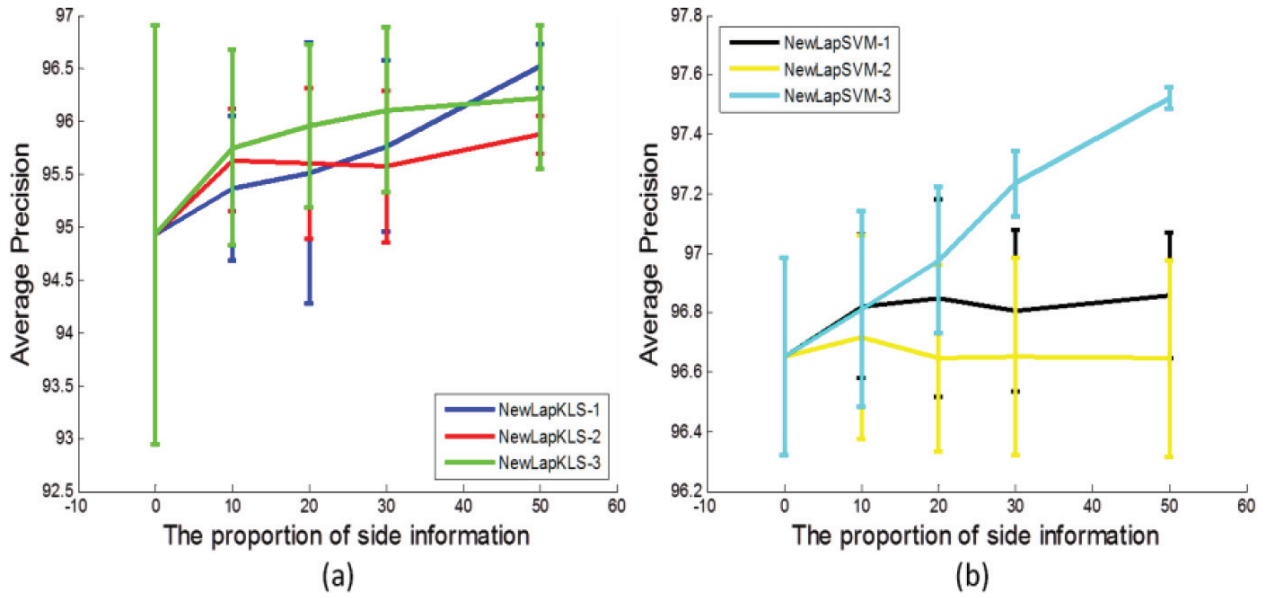

Figure 4. The total classification result under 20% labels (a) KLS, (b) SVM.

(labeled information and side information) are randomly selected from training set. To avoid any potential bias induced by data selecting, the above process is repeated for five times.

For the first two proposed algorithms using the must-link constraints, we first determine the parameter α which balances the traditional graph Laplacian and the must-link Laplacian matrix. The parameter α of novel methods is tuned from the candidate set $\{e^i | i = -10, -9, -8, \cdots, 10\}$ through cross-validation. In addition, the regularization parameters Υ_A, Υ_I are chosen from $\{10^{-8}, , 10^{-7}, , 10^{-6}, \cdots, , 10^6, , 10^7, , 10^8\}$ through cross-validation on the training data. We verify the AP performance to select the proper parameters. Note that the parameter α may be different for the same classifier to get the best performance under the different proportion of side

information. In results, the legend NewLapKLS-1 represents the kernel least squares classifier using algorithm $L'' = L(L^M + \alpha I)$, NewLapSVM-2 stands for the support vector machines classifier using algorithm $L'' = L + \alpha L^M$, and so on.

Figure 3 shows the classification results achieved by KLS and SVM classifiers under the 10% labeled samples. We can see two main points. First, our proposed three local structure preserving algorithms with pairwise constraints usually get the overall better performances than the well-known semi-supervised methods (LapKLS and LapSVM) without side information. Second, we can clearly see, in most cases, the results gradually become better with the increase of side information. From **Figures 4–6**, we can get the analogous observations for our proposed

Figure 5. The total classification result under 30% labels (a) KLS, (b) SVM.
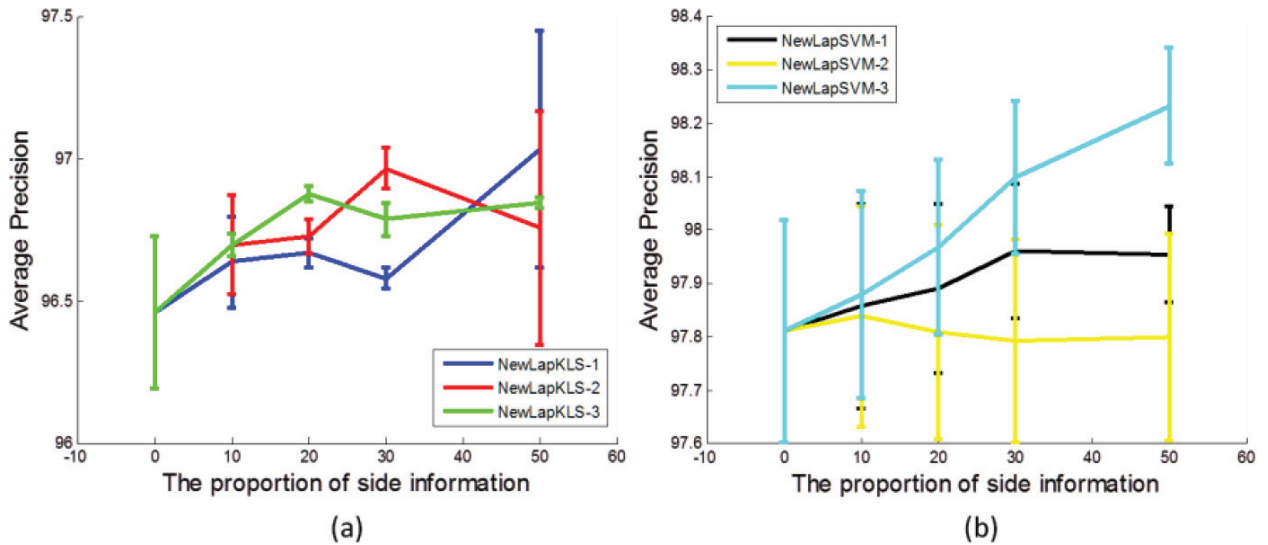
Figure 6. The total classification result under 50% labels (a) KLS, (b) SVM.
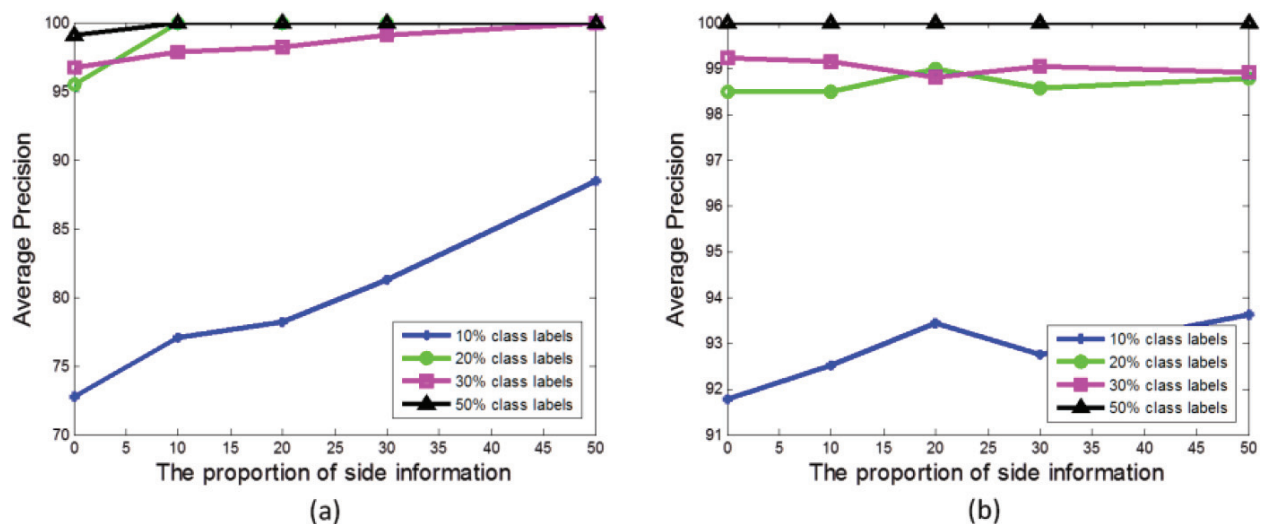
Figure 7. The result of jumping up with the different proportion of side information by LapKLS, LapSVM, NewLapKLS-1 and NewLapSVM-1.

methods compared with their counterparts. These observations indicate that our proposed learning model can better explore and exploit the local structure by taking advantage of the geometrical structure information in the pairwise constraints and manifold regularization. What we can note is that the classification results have slight fluctuation with more side information when the number of class labels is large. These observations suggest it is critical to select parameters for our proposed methods.

To investigate whether the single action of CAS-YNU-MHAD can get the outperformance, we choose jumping up as an example in **Figure 7**. We can find that, our proposed algorithm consistently performs better than the previous algorithm without side information. Especially, we can see, the classification result can get a significant development when the number of labeled samples is limited.

4. HesR-based SSL

Although LapR has received extensive attention, it is observed that the null space of the graph Laplacian along the underlying manifold is a constant function that possibly results in poor generalization. In contrast to Laplacian, Hessian can properly exploit the intrinsic local geometry of the data manifold. In recent works [23–26, 28], HesR based SSL algorithms have been proved to achieve better performance than LapR based ones.

Hessian matrix can be computed by the following four steps.

Step 1: Neighborhood construction. Using k-neighborhood to define neighbors in Euclidean distance for each input point x_i, we get neighborhood matrix N_i.

Step 2: Create local tangent coordinates. Conduct singular value decomposition on neighborhood matrix $N_i = UDV$. The first d columns of V ($V_i = [v_1, v_2, \ldots, v_d]$) mean the tangent coordinates of data points x_i.

Step 3: Build local Hessian estimator. Apply Gram-Schmidt procedure on the matrix $[1, V_i, Q_i]$ with the first column is a vector of ones, $Q_i = \left[v_i \boxtimes v_j\right]_{1 \le i \le j \le d}$ is a matrix of $m(m+1)/2$ columns to get $\widehat{M}_i^k$. Then taking the last $m(m+1)/2$ columns of $\widehat{M}_i^k$ as H^i.

Step 4: Construct Hessian matrix H. A symmetric matrix H is constructed with the entry $H_{ij} = \sum_l \sum_r H_{r,i}^l H_{r,j}^l$.

The HesR model can be expressed in:

$$f(*) = arg \min_{f \epsilon H_K} \frac{1}{l} \sum_{i=1}^{l} V(x_i, y_i, f) + \Upsilon_A \|f\|_K^2 + \frac{\Upsilon_I}{(l+u)^2} \mathbf{f}^T H \mathbf{f}. \tag{19}$$

Hessian has been widely utilized in improving the SSL classification performance. Liu et al. [18] present multi-view Hessian discriminative sparse coding (mHDSC) which seamlessly integrates Hessian regularization with discriminative sparse coding for multi-view learning problems. In [24], HesR was employed into support vector machine to boost the classifier. In [19], HesR was integrated into multi-view learning for image annotation, extensive experiments on the PASCAL VOC'07 dataset validate the effectiveness of HesR by comparing it with LapR.

5. pLapR-based SSL

Although the p-Laplacian has nice theoretical foundations, it is still a strenuous work to approximate graph p-Laplacian, which extremely limits the applications of p-Laplacian regularization. In this section, we provide an effect and efficient fully approximation of graph p-Laplacian, which significantly lows down the computation cost. Then we integrate the approximated graph p-Laplacian into manifold regularization framework and develop p-Laplacian regularization. Based on the pLapR, several extended algorithms were proposed.

5.1. pLapR

The graph p-Laplacian is approximated by getting all eigenvectors and eigenvalues of p-Laplacian [55]. Assume that $f^{*1}, f^{*2}, \cdots, f^{*K}$ are K eigenvectors of p-Laplacian Δ_p^w associated with unique eigenvalues $\lambda_1^*, \lambda_2^*, \cdots, \lambda_K^*$. Luo et al. [32] introduced an approximation for full eigenvectors of p-Laplacian by solving the following p-Laplacian embedding problem:

$$min \, \mathcal{F} J_E(\mathcal{F}) = \sum_k \frac{\sum_{ij} w_{ij} \left| f_i^k - f_j^k \right|^p}{\|f^k\|_p^p} \tag{20}$$

$$s.t. \quad \mathcal{F}^T \mathcal{F} = I.$$

Solving the Eq. (20) with the gradient descend optimization, we can then obtain the full eigenvalues $\Lambda^* = \left(\lambda_1^*, \lambda_2^*, \cdots, \lambda_K^*\right)$ of p-Laplacian associated with the eigenvectors $\mathcal{F}^* = \left(f^{*1}, f^{*2}, \cdots, f^{*K}\right)$ by $\lambda_p = \frac{\sum_{ij} w_{ij} |f_i - f_j|^p}{2\|f\|_p^p}$. Finally, the graph p-Laplacian approximated by $L_p = \mathcal{F}^* \Lambda \mathcal{F}^{*T}$.

We introduce the approximation graph p-Laplacian into a regularizer to exploit the intrinsic local geometry of the data manifold. Therefore, in p-Laplacian regularization framework, the optimization problem in Eq. (1) becomes

$$f^* = \arg \min_{f \in H_K} \frac{1}{l} \sum_{i=1}^{l} V(x_i, y_i, f) + \Upsilon_A \|f\|_K^2 + \frac{\Upsilon_I}{(l+u)^2} \mathbf{f}^T L_p \mathbf{f}. \tag{21}$$

Here, L_p is the graph p-Laplacian.

The proposed pLapR can be applied to variant MRSSL-based applications with different choices of loss function. Here, we apply pLapR to support vector machines (SVM) and kernel least squares (KLS) as examples.

Applying the hinge loss function in p-Laplacian learning, the p-Laplacian support vector machines (pLapSVM) solves the following optimization problem:

$$f^* = \arg \min_{f \in H_K} \frac{1}{l} \sum_{i=1}^{l} \left(1 - y_i f(x_i)\right)_+ + \Upsilon_A \|f\|_K^2 + \frac{\Upsilon_I}{(l+u)^2} \mathbf{f}^T L_p \mathbf{f}. \tag{22}$$

The represular theorem has been proved exist and has the general form in Eq. (16). Hence the optimization problem (21) can be expressed as

$$f^* = \arg \min_{f \in H_K} \frac{1}{l} \sum_{i=1}^{l} \left(1 - y_i f(x_i)\right)_+ + \Upsilon_A \alpha^T \mathbf{K} \alpha + \frac{\Upsilon_I}{(l+u)^2} \alpha^T \mathbf{K} L_p \mathbf{K} \alpha. \tag{23}$$

We outline the KLS with p-Laplacian regularization. For p-Laplacian kernel least squares (pLapKLS), it solves the following optimization problem

$$f^* = \min_{f \in H_K} \frac{1}{l} \sum_{i=1}^{l} \left(y_i - f(x_i)\right)^2 + \Upsilon_A \alpha^T \mathbf{K} \alpha + \frac{\Upsilon_I}{(l+u)^2} \alpha^T \mathbf{K} L_p \mathbf{K} \alpha. \tag{24}$$

To evaluate the effectiveness of the proposed pLapR, we apply pLapSVM and pLapKLS to scene recognition on the Scene 67 database [56] and Scene 15 data set [57]. **Figure 8** illustrates the framework of pLapR for scene recognition.

The Scene 67 data set contains 15,620 indoor scene images collected from different sources including online image search tools, online photo sharing sites and the LabelMe dataset. Particularly, these images can be categorized into 67 classes covering 5 big scene groups (i.e., stores, home, public spaces, leisure and working place). Some example images are shown in **Figure 9**.

Scene 15 data set is composed of 15 scene categories, totally 4485 images. Each category has 200–400 images. The images contain not only indoor scenes, such as living room, kitchen, and store, but also outdoor scenes, such as forest, mountain, tall building, open country, and so on (see in **Figure 10**).

For Scene 67 dataset, we randomly select 80 images of each class to form the training set and the rest as testing set. For Scene 15 dataset, 100 images per class are randomly selected as the
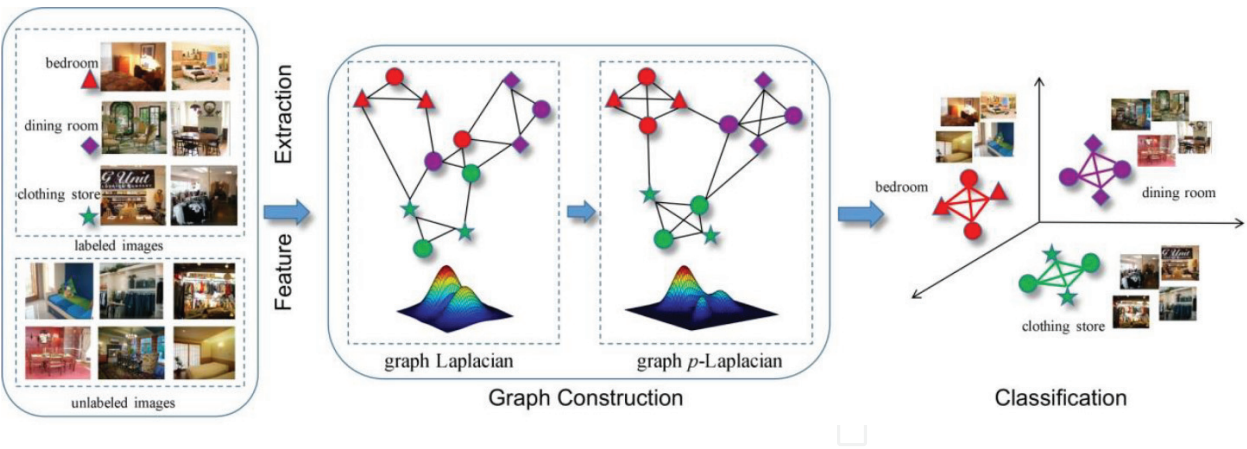
Figure 8. The framework of pLapR for indoor scene recognition.
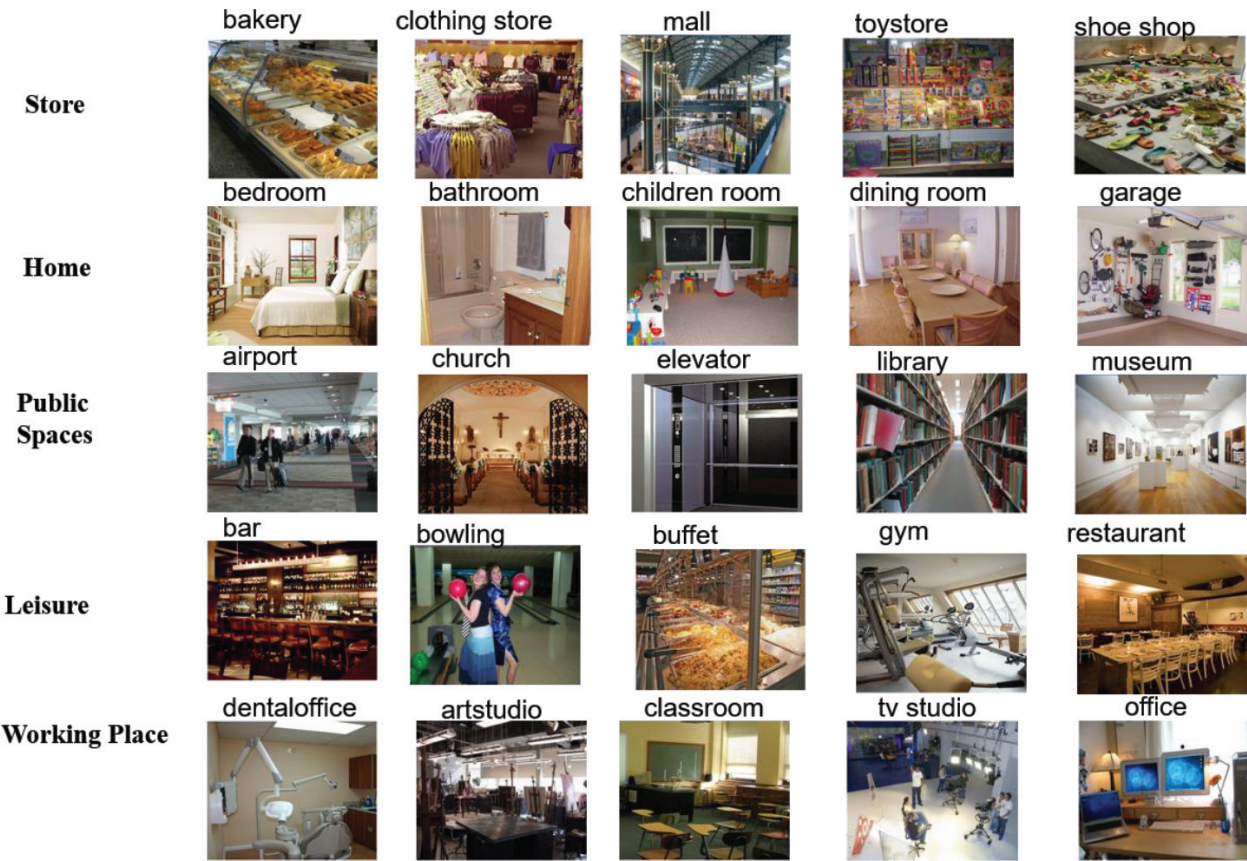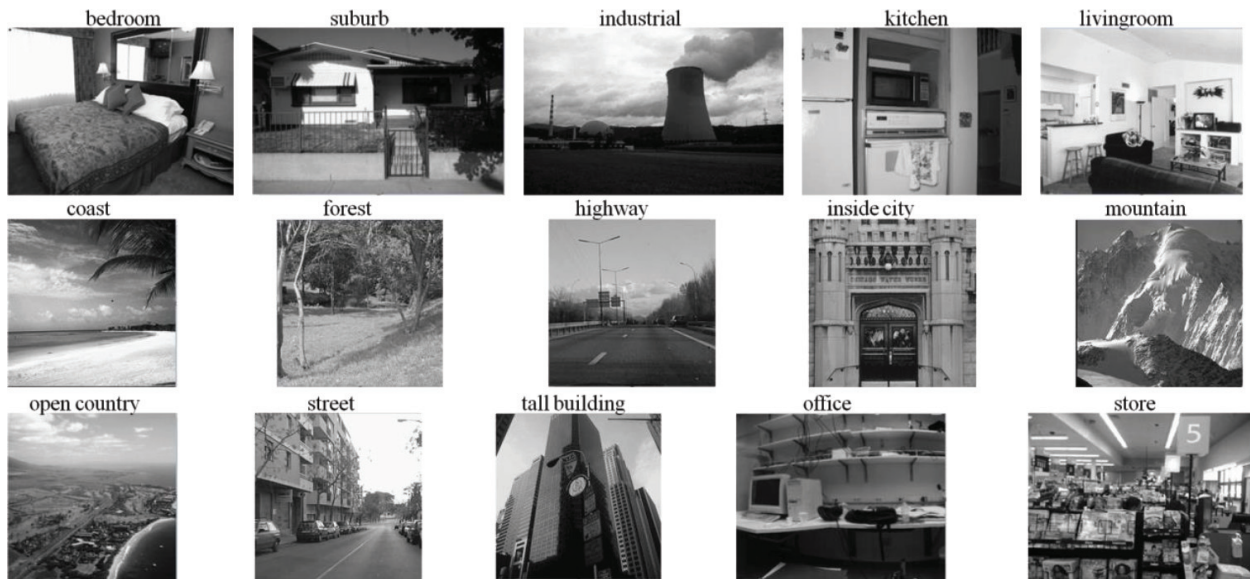

Figure 9. Some example images of Scene 67 database. The dataset totally has 67 indoor scene categories that can be grouped into 5 big scene groups. Each row demonstrates one big scene group.

training set, and the rest for testing. In semi-supervised experiments, a certain percentage (10, 20, 30, 50%) samples of training set are randomly assigned as labeled data. To avoid any bias introduced by the random partitioning of samples, the above assignment is carried out for five times independently.

Figure 10. Some example images of Scene 15 data set. The dataset totally has 15 scene categories.

The regularization parameters that is, γ_A and γ_I are tuned from the candidate set $\{10^i | i = -10, -9, -8, \cdots, 10\}$ and the parameter p for pLapR from the candidate set $\{1, 1.1, 1.2, \cdots, 3\}$ through cross-validation on the training data with 10% labeled sample, respectively. The performance is measured by the average precision (AP) for single class and mean average precision (mAP) for overall classes. Firstly, we show the mAP boxplot of the pLapR on Scene 67 dataset when $p = 2$ and the standard LapR for comparison in **Figure 11**. We can clearly see that the performance of pLapR with $p = 2$ is similar to standard LapR, which demonstrates that the graph p-Laplacian with $p = 2$ becomes the standard graph Laplacian.

Figure 12 illustrates the performance of pLapKLS with different p values. The upper subfigure is the performance of the Scene 67 database. We observe that the best performance of indoor scene classification on the Scene 67 dataset can be obtained with $p = 1.1$. The lower subfigure is the performance of the Scene 15 database and the best performance is achieved when $p = 1$.
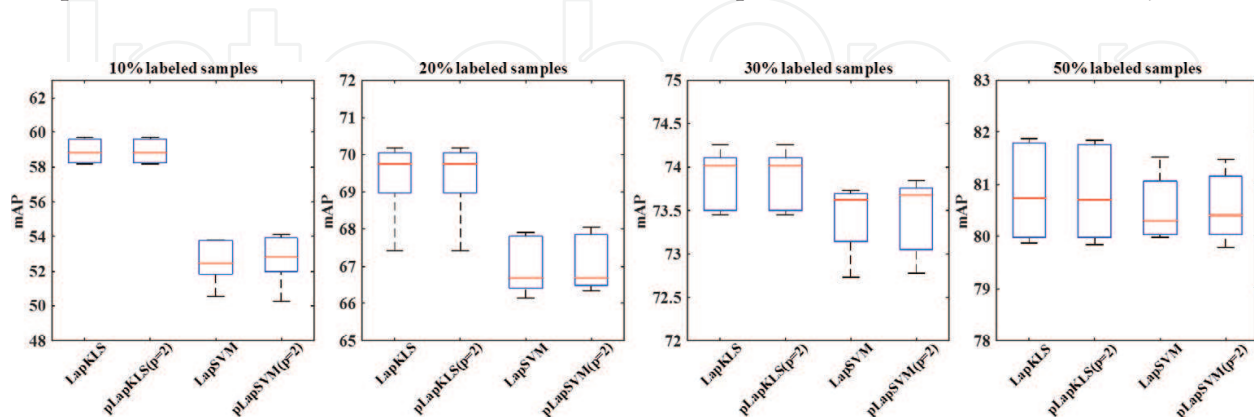
Figure 11. mAP of pLapR($p = 2$) and LapR on Scene 67 dataset. Each subfigure reports the results under different labeled samples. In each subfigure, the y-axis is the mAP over all scene classes, and the x-axis is different classifiers.
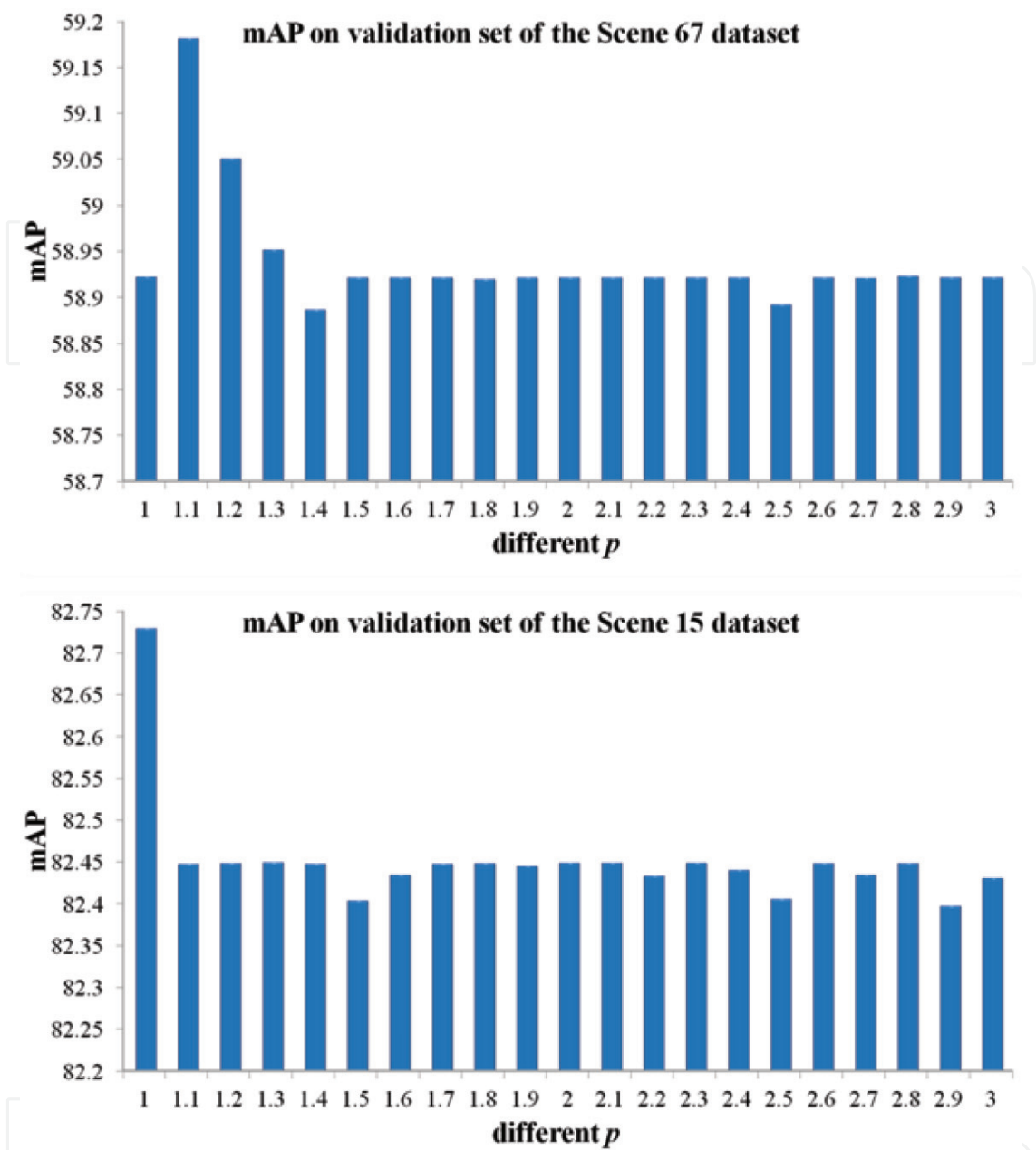
Figure 12. mAP results of pLapKLS under different p with 10% labeled sample. The y-axis is the mAP over all classes, and the x-axis is the parameter p.

Then we evaluate the performance of the pLapR with the representative LapR and HesR. **Figure 13** and **Figure 14** show the mAP performance on Scene 67 data set and Scene 15 data set, respectively. The four subfigures of upper row are KLS methods, and the lower four ones are SVM methods. From the results of two data sets, we can see that the pLapR outperforms both LapR and HesR especially when only a small number of samples labeled.
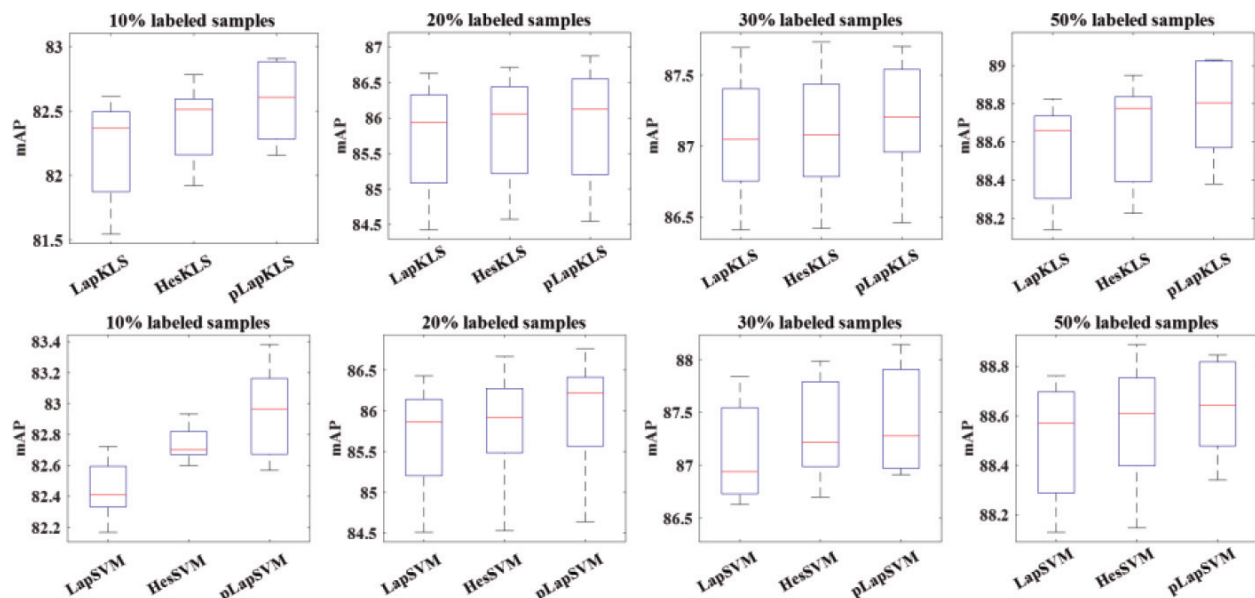
To discuss the AP performance of different algorithms for single class, we show the results of several classes of Scene 15 data set including mountain, open country, tall building and industrial. Each subfigure corresponds on single scene class. The upper four subfigures are
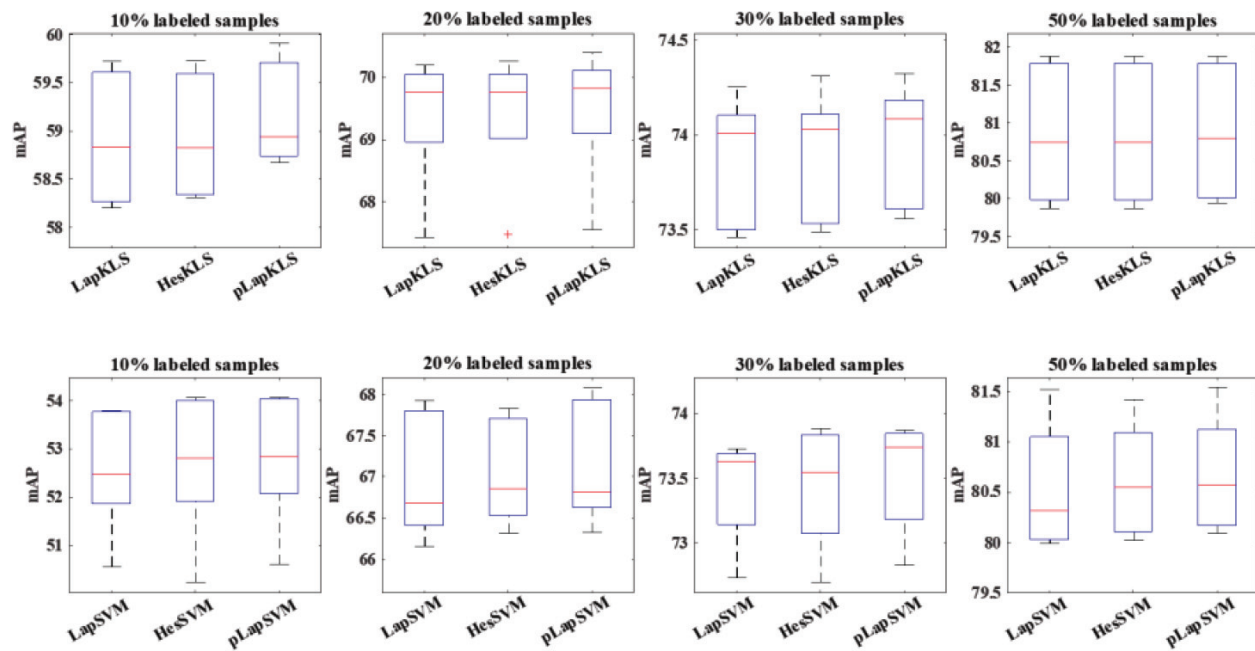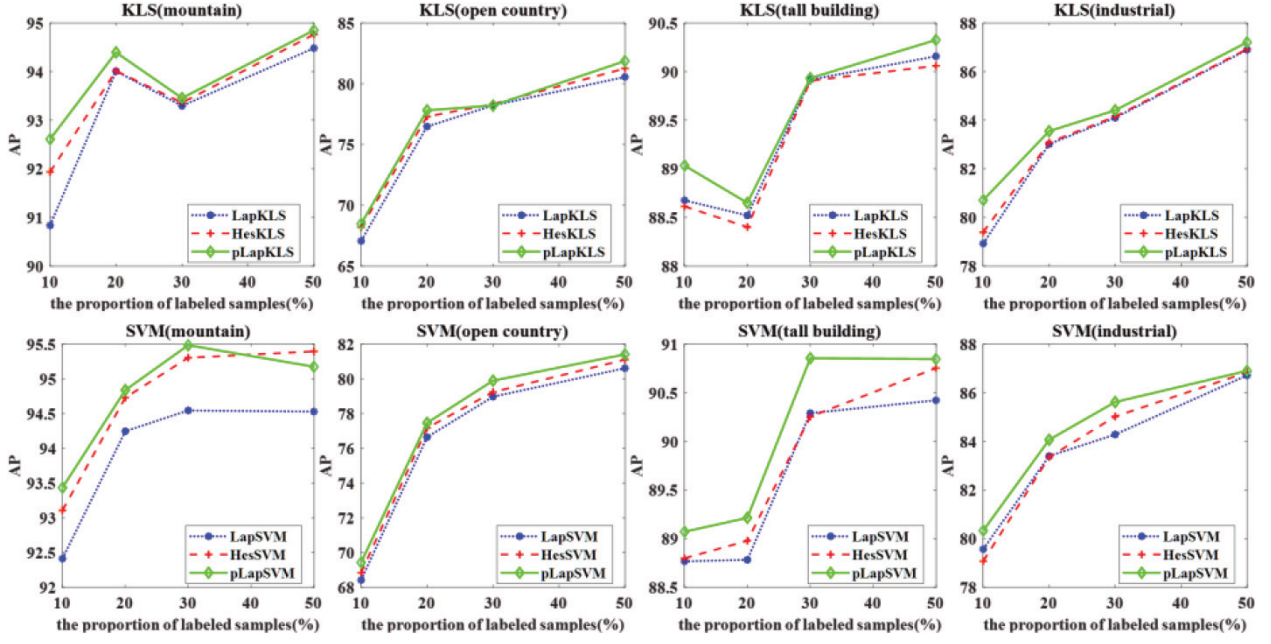
Figure 13. mAP of different algorithms on Scene 67 data set. The four subfigures of upper row are KLS methods, and the lower four ones are SVM methods.

KLS methods, and the lower four ones are SVM methods. In each subfigure, the y-axis is the AP results and the x-axis is the number of labeled samples. From the AP results, we can find that, in most cases, the pLapR performs better than the traditional methods including LapR and HesR (**Figure 15**).

Figure 14. mAP of different algorithms on Scene 15 data set. The four subfigures of upper row are KLS methods, and the lower four ones are SVM methods.

Figure 15. AP of different methods on several classes of Scene 15 data set including mountain, open country, tall building and industrial. Each subfigure corresponds on single scene class. The upper four subfigures are KLS methods, and the lower four ones are SVM methods. In each subfigure, the y-axis is the AP results and the x-axis is the number of labeled samples.

5.2. Hypergraph p-Laplacian (HpLapR)

In this subsection, we propose a hypergraph p-Laplacian regularized method for image recognition. The hypergraph and p-Laplacian [31, 58, 59] both provide convincing theoretical evidence to better preserve the local structure of data. However, the computation of hypergraph p-Laplacian is difficult. We provide an effect and efficient approximation algorithm of hypergraph p-Laplacian. Considering the higher order relationship of samples, the hypergraph p-Laplacian regularizer is built for preserving local structures. Hypergraph p-Laplacian regularization (HpLapR) is also introduced to logistic regression for remote sensing image recognition.

Assume that hypergraph p-Laplacian has n eigenvectors $\mathcal{F}^{*hp} = \left(f^{*hp1}, , f^{*hp2}, \cdots, f^{*hpn} \right)$ associated with unique eigenvalues $\lambda^{*hp} = \left(\lambda_1^{*hp}, \lambda_2^{*hp}, \cdots, \lambda_n^{*hp} \right)$, we compute the approximation of hypergraph p-Laplacian L_p^{hp} by $L_p^{hp} = \mathcal{F}^{*hp} \lambda^{*hp} \mathcal{F}^{*hp^T}$. 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 [52]. In detail, we construct hypergraph Laplacian L^{hp} and compute adjacency matrix W^{hp} by Eq. (5) and Eq. (6), respectively.

Following the study on plapR [31, 55], eigenvalue and the corresponding eigenvector on hypergraph p-Laplacian can be computed by the following hypergraph p-Laplacian embedding problem:

$$min\,\mathcal{F}^{hp}J_E\left(\mathcal{F}^{hp}\right) = \sum_k \frac{\sum_{ij} w_{ij}^{hp}\left|f_i^{hpk} - f_j^{hpk}\right|^p}{\|f^{hpk}\|_p^p} \tag{25}$$

$$s.t.\quad \mathcal{F}^{hp^T}\mathcal{F}^{hp} = I$$

Solving the problem of Eq. (25) with the gradient descend optimization. We can also get the full eigenvalue $\lambda^{hp} = \left(\lambda_1^{hp}, \lambda_2^{hp}, \cdots, \lambda_n^{hp}\right)$ by $\lambda_k^{hp} = \frac{\sum_{ij} w_{ij}^{hp}\left|f_i^{hpk} - f_j^{hpk}\right|^p}{\|f^{hpk}\|_p^p}$.

Finally, the approximation of L_p^{hp} can be solved by $L_p^{hp} = \mathcal{F}^{hp}\lambda^{hp}\mathcal{F}^{hp^T}$.

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

$$f^* = \arg\min_{f \in H_K} \frac{1}{l}\sum_{i=1}^l V(x_i, y_i, f) + \Upsilon_A\|f\|_K^2 + \frac{\Upsilon_I}{(l+u)^2}\mathbf{f}^T L_p^{hp}\mathbf{f}. \tag{26}$$

Here, L_p^{hp} is hypergraph p-Laplacian. We employ the proposed HpLapR with logistic regression.

Substitute logistic loss function in Eq. (26), the HpLapR can be rewritten as

$$f^* = \arg\min_{f \in H_K} \frac{1}{l}\sum_{i=1}^l \left(log\left(1 + e^{-y_i f(x_i)}\right)\right) + \Upsilon_A\|f\|_K^2 + \frac{\Upsilon_I}{(l+u)^2}\mathbf{f}^T L_p^{hp}\mathbf{f}^T. \tag{27}$$

According to the representer theorem, the solution of (27) w.r.t. f exists and can be expressed by Eq. (16). Thus, we finally construct the HpLapR as the following optimization problem:

$$f^* = \arg\min_{f \in H_K} \frac{1}{l}\sum_{i=1}^l \left(log\left(1 + e^{-y_i K(x_i, x)\alpha}\right)\right) + \Upsilon_A\alpha^T K\alpha + \frac{\Upsilon_I}{(l+u)^2}\alpha^T KL_p^{hp}K\alpha. \tag{28}$$

Apply the conjugate gradient algorithm, we can get the solution of the optimized f.

To evaluate the effectiveness of the proposed HpLapR, we compare HpLapR with other local structure preserving algorithms including LapR, HLapR and pLapR. **Figure 16** illustrates the framework of HpLapR for UC-Merced data set.

UC-Merced data set [60] consists of totally 2100 land-use images collected from aerial orthoimage with the pixel resolution of one foot. 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, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis courts (see in **Figure 17**).

In our experiments, we extract high-level visual features using the deep convolution neural network (CNN) [61]. We randomly choose 50 images per class as training samples and the rest
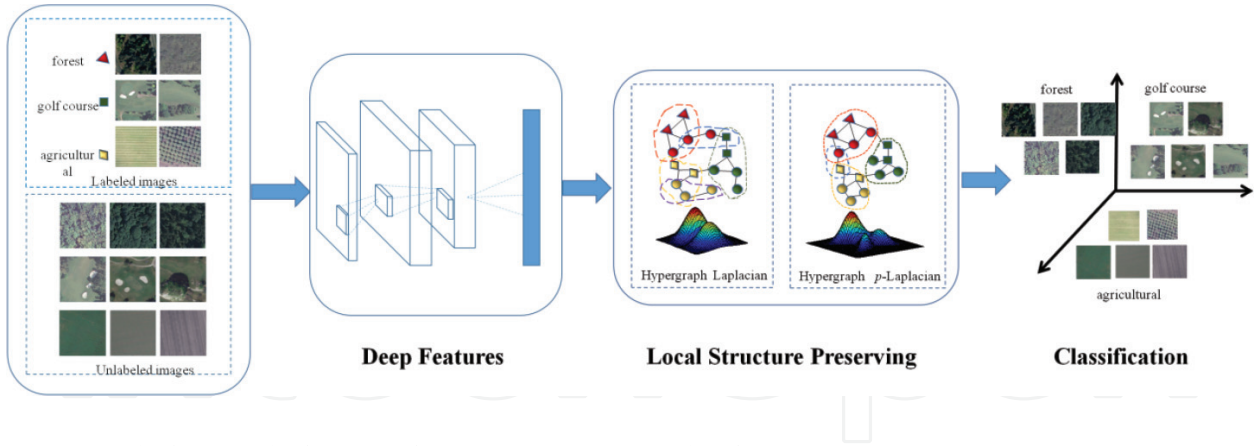
Figure 16. The framework of HpLapR for remote sensing image classification.
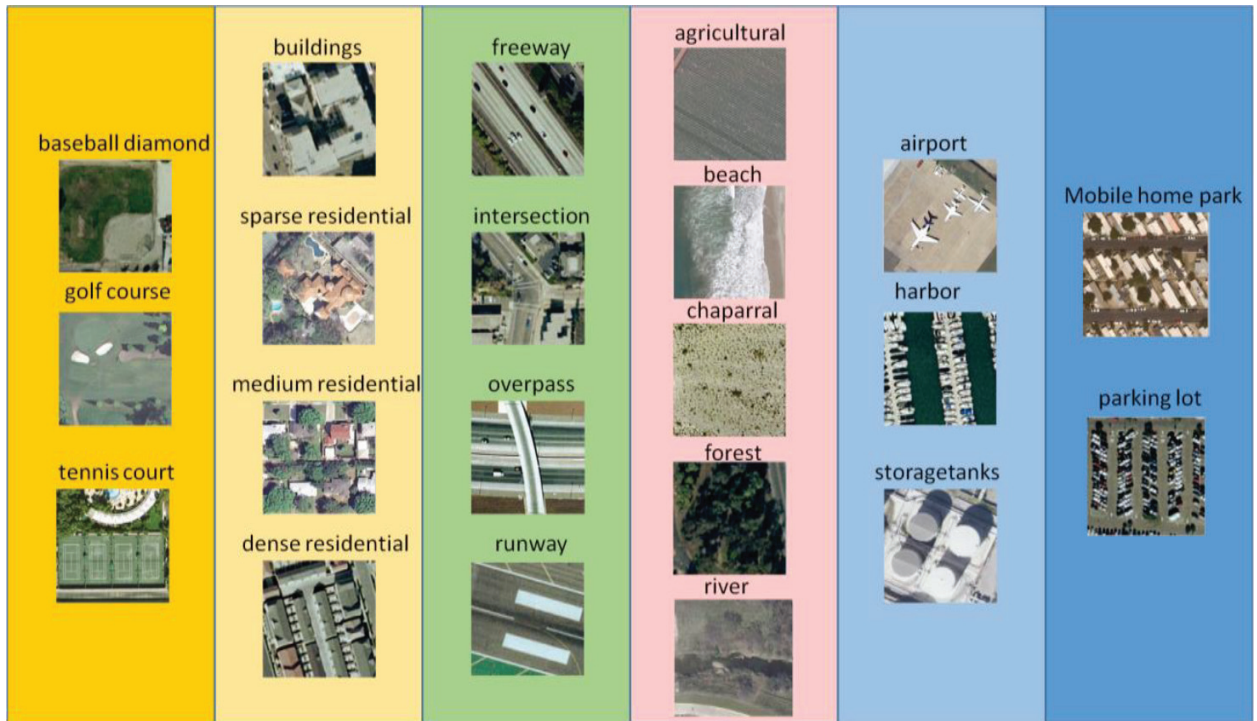
Figure 17. Class examples of UC-Merced data set. The dataset totally has 21 remote sensing categories that can be simply grouped into six groups according to the distinction of land use. Each column represents one group.

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) [62]. It is worthy to notice 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 **Figure 17**. The setting of class labels is as same as pLapR.

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 $\{5, 6, 7, \cdots, 15\}$ through cross-validation. The setting of regularization parameters γ_A, γ_I and p are as same as pLapR experiments.
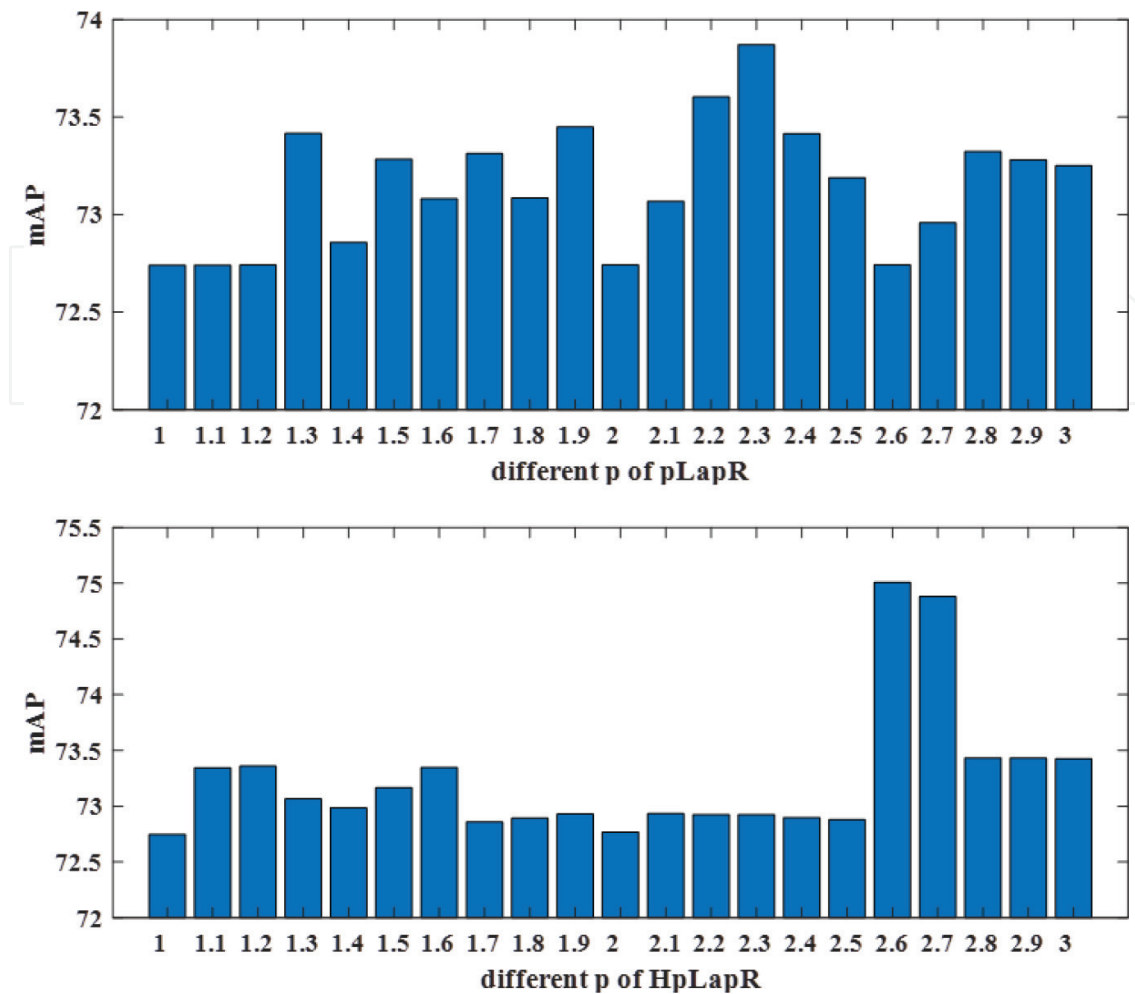
Figure 18. Performance of mAP with different p values on validation set.

Figure 18 illustrates the mAP performance of pLapR and HpLapR on the validation set when p varies. The x-axis is the 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 $p = 2.6$.

We compare our proposed HpLapR with the representative LapR, HLapR and pLapR. From **Figure 19**, 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 graph p-Laplacian. To evaluate the effectiveness of HpLapR for single class, **Figure 20** shows the AP results of different methods on several land-use classes including beach, dense residential, freeway and tennis court. From **Figure 20**, we can find that, in most cases, HpLapR performs better than both pLapR and HLapR, while pLapR and HLapR consistently outperforms than LapR.

5.3. Ensemble p-Laplacian regularization (EpLapR)

As a natural nonlinear generalization of graph Laplacian, p-Laplacian has been proved having the rich theoretical foundations to better preserve the local structure. However, it is difficult to
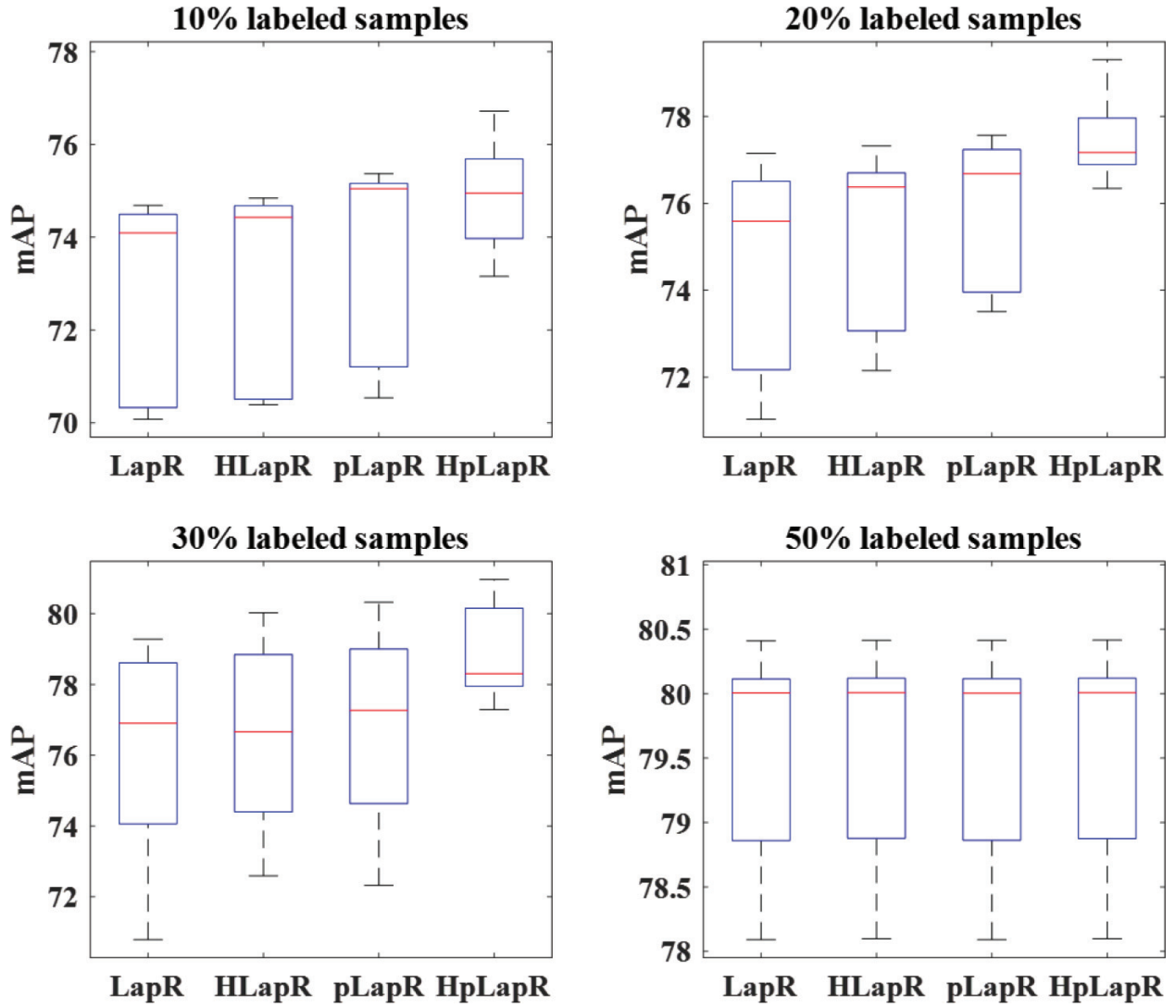
Figure 19. mAP performance of different algorithms.

determine the fitting graph p-Lapalcian, that is, the parameter p that is a critical factor for the performance of graph p-Laplacian. In this section, we develop an ensemble p-Laplacian regularization to fully approximate the intrinsic manifold of the data distribution. EpLapR incorporates multiple graphs into a regularization term in order to sufficiently explore the complementation of graph p-Laplacian. Specifically, we construct a fused graph by introducing an optimization approach to assign suitable weights on different p-value graphs. Then, we conduct semi-supervised learning framework on the fused graph.

Assume a set of candidate graph p-Laplacian $\{L_1^p, \cdots, L_m^p\}$, according to the manifold regularization framework, the proposed EpLapR can be written as the following optimization problem:

$$f^* = \arg\min_{f \in H_K} \frac{1}{l} \sum_{i=1}^{l} V(x_i, y_i, f) + \Upsilon_A \|f\|_K^2 + \frac{\Upsilon_I}{n^2} \mathbf{f}^T \mathbf{L} \mathbf{f}. \tag{29}$$

where L is the optimal fused graph with $L = \sum_{k=1}^{m} \mu_k L_k^p$, $s.t. \sum_{k=1}^{m} \mu_k = 1$, $\mu_k \geq 0$, $for\ k = 1, \cdots, m$.
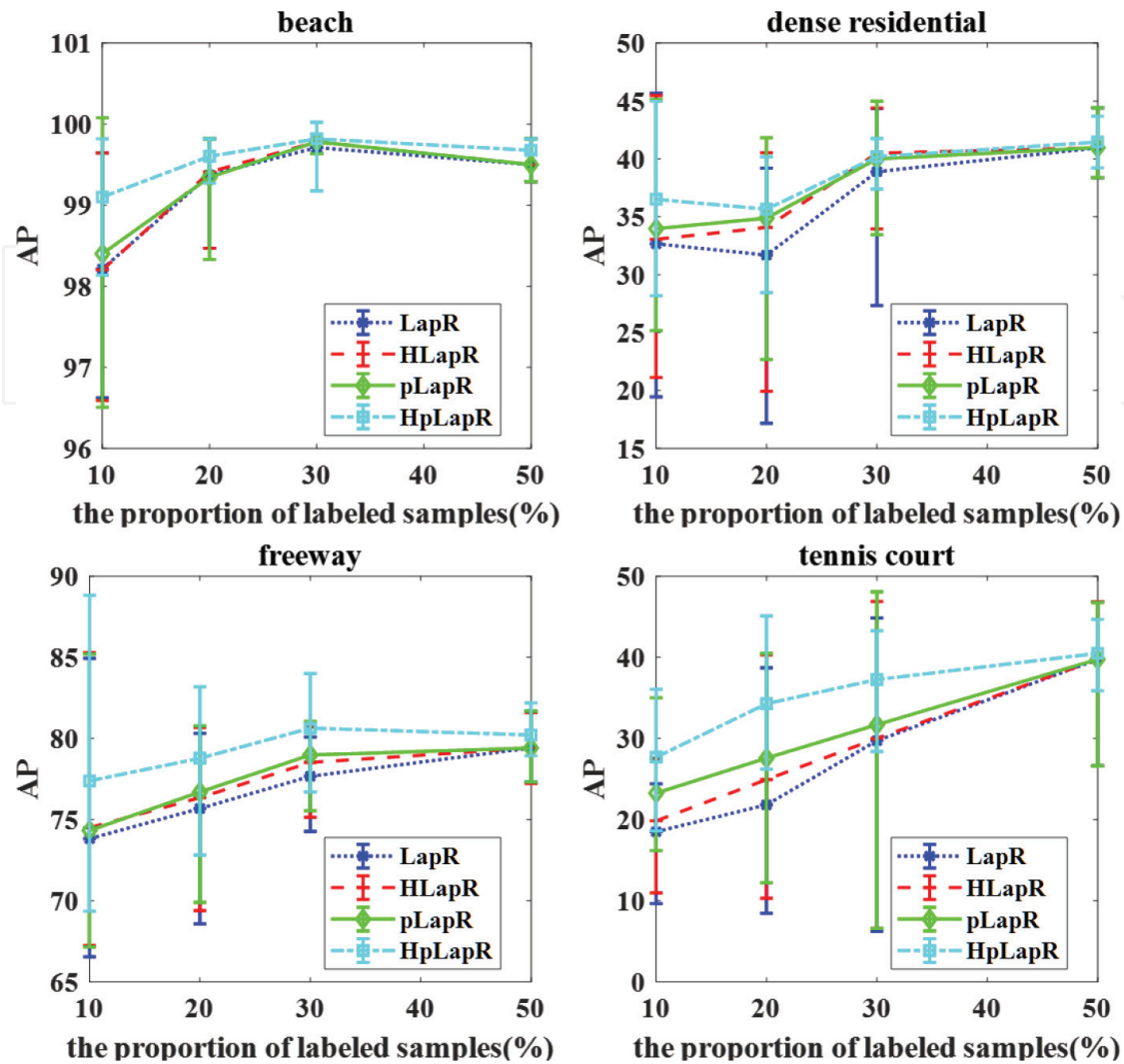
Figure 20. AP performance of different methods on several classes.

To avoid the parameter μ_k overfitting to one graph [63], we make a relaxation by changing μ_k to μ_k^γ, and obtain the optimization problem as:.

$$f^* = arg\,min_{f \in H_K} \frac{1}{l} \sum_{i=1}^{l} V(x_i, y_i, f) + \Upsilon_A \|f\|_K^2 + \frac{\Upsilon_I}{n^2} \mathbf{f}^T \left(\sum_{k=1}^{m} \mu_k^\gamma L_k^p \right) \mathbf{f}.$$

$$\text{s.t.} \sum_{k=1}^{m} \mu_k = 1, \quad \mu_k \geq 0, \quad \text{for } k = 1, \cdots, m$$

(30)

The representor theorem presents us with the existence and the general form of Eq. (16) under a fixed μ. Therefore, we rewrite the objective function as

$$f^* = arg\,min_{f \in H_K} \frac{1}{l} \sum_{i=1}^{l} V(x_i, y_i, f) + \Upsilon_A \alpha^T K\alpha + \frac{\Upsilon_I}{(l+u)^2} \alpha^T K \left(\sum_{k=1}^{m} \mu_k^\gamma L_k^p \right) K\alpha.$$

$$\text{s.t.} \sum_{k=1}^{m} \mu_k = 1, \quad \mu_k \geq 0, \quad \text{for } k = 1, \cdots, m$$
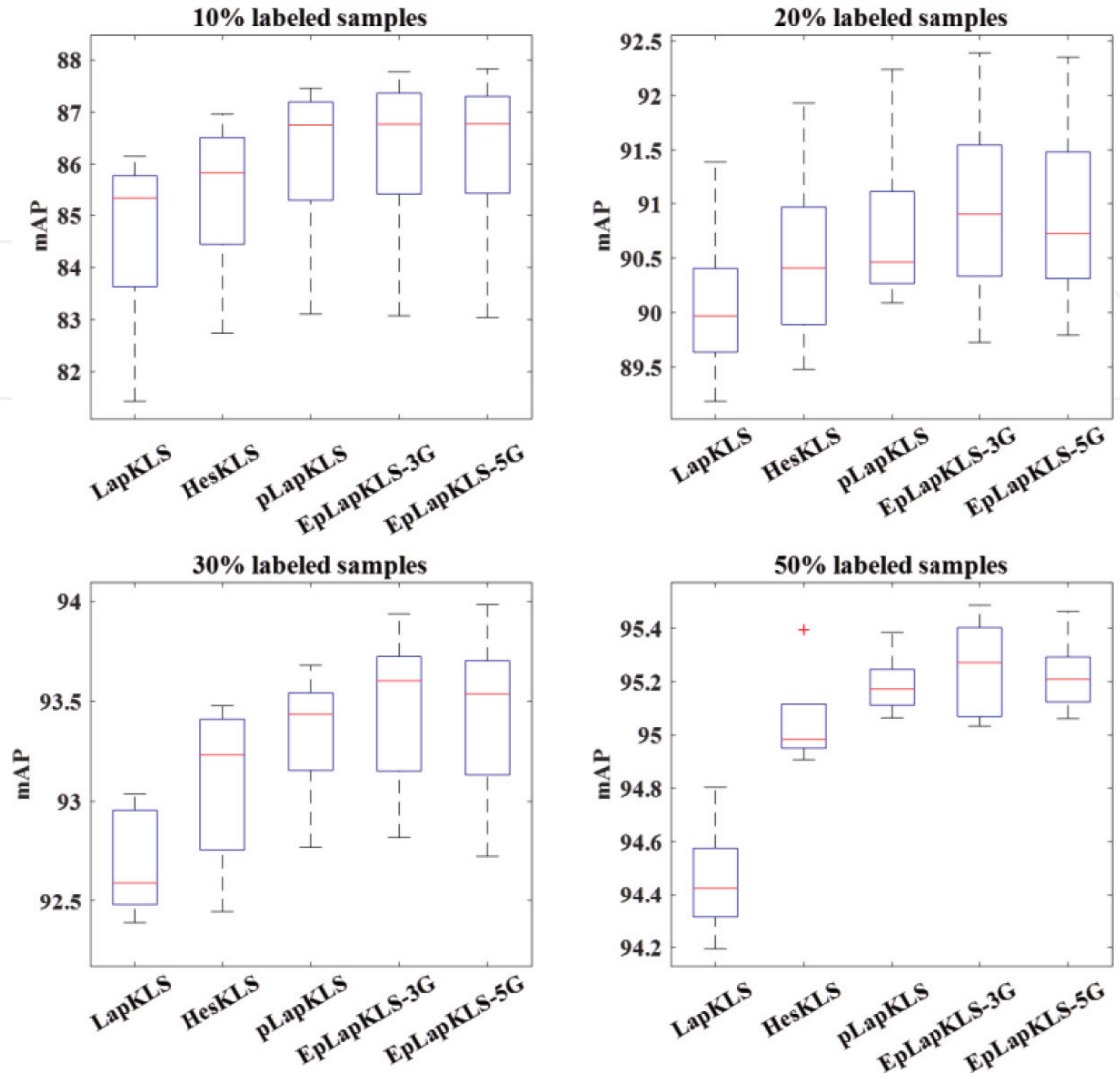
(31)

Figure 21. mAP performance of different algorithms on KLS method.

Here, an alternating optimization procedure is utilized to minimize f^*.

We compare EpLapR with other local structure preserving algorithms including LapR, HesR and pLapRon UC-Merced data set. We apply the support vector machines and kernel least squares for remote sensing image classification.

In the experiments, we apply the parameter setting as the same as pLapR, and the experiment of pLapR is conducted with $p = 2.8$. For EpLapR, we created two graph p-Laplacian sets. For the first set (EpLapR-3G), we choose $p = \{2.5, 2.7, 2.8\}$, which led to 3 graphs. For another one (EpLapR-5G), with 5 graphs where $p = \{2.4, 2.5, 2.6, 2.7, 2.8\}$.

We compare our proposed EpLapR with the representative LapR, HesR and pLapR. **Figures 21** and **22** demonstrate the mAP results of different algorithms on KLS methods and SVM methods, respectively. We can see that, in most cases, the EpLapR outperforms LapR, HesR and pLapR, which shows the advantages of EpLapR in local structure of preserving.
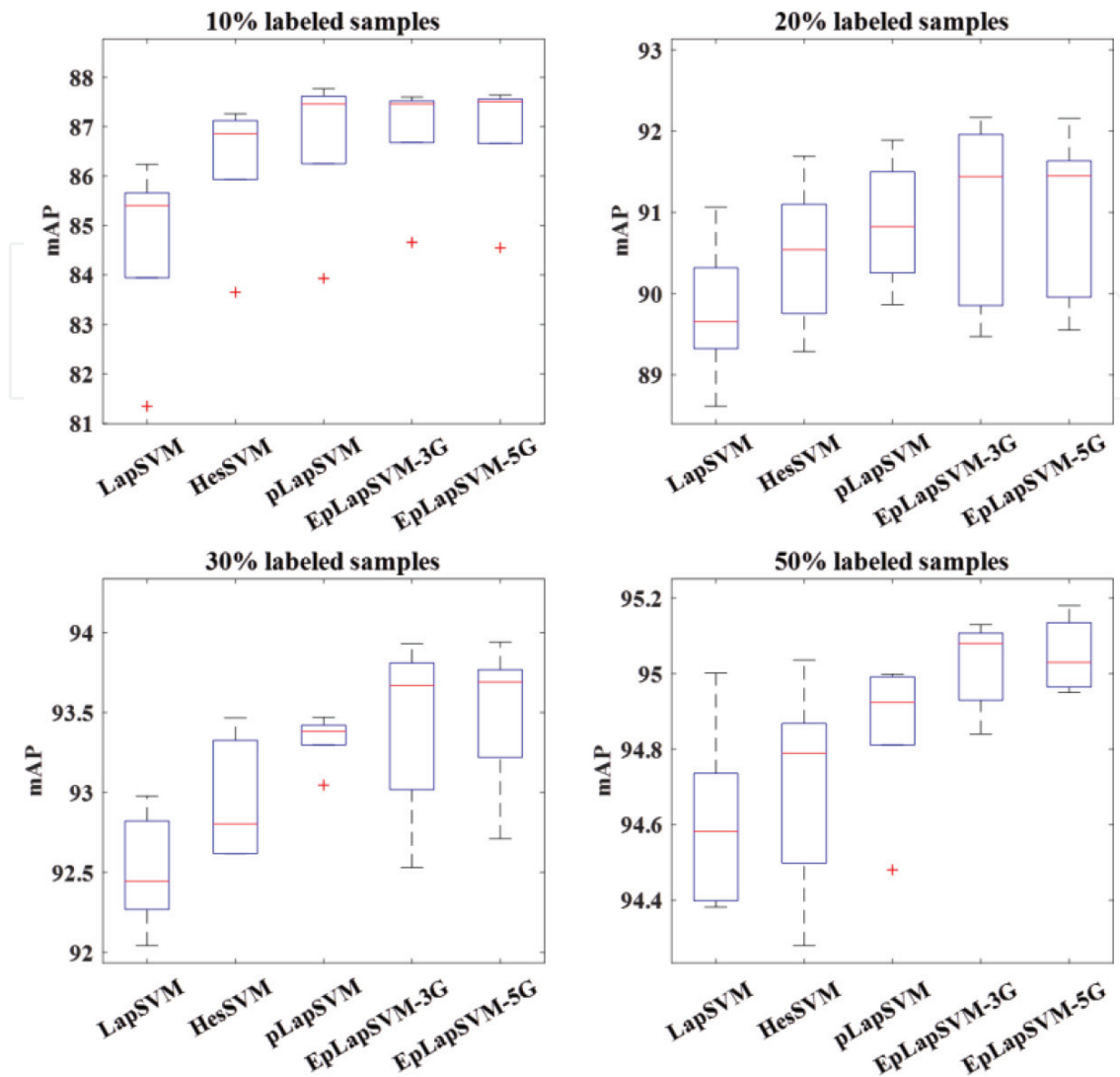
Figure 22. mAP performance of different algorithms on SVM method.

6. Conclusions

In this chapter, we show the LapR, HesR, pLapR and present several extensions based on the manifold regularization framework. We propose a local structure preserving method that effectively integrates manifold regularization and pairwise constraints. We develop an efficient approximation algorithm of graph p-Laplacian and propose p-Laplacian regularization to preserve the local geometry. Considering the hypergraph contains more local grouping information in comparison to simple graph, we propose hypergraph p-Laplacian regularization to preserve the geometry of the probability distribution. In practical application of p-Laplacian regularization model, it is difficult to determine the optimal graph p-Lapalcian because the parameter p usually chose by cross validation method which lacks the ability to approximate the optimal solution. Therefore, we propose an ensemble p-Laplacian regularization to better approximate the geometry of the data distribution.

7. Expectations

In the general image recognition, images are naturally represented by multi-view features, such as color, shape and texture. Each view of a feature summarizes a specific characteristic of the image, and features for different views are complementary to one another. Therefore, in the future work, we will study the multi-view p-Laplacian regularization to effectively explore the complementary properties of different features from different views. Meanwhile, we will try to combine the p-Laplacian learning with the deep learning to get a more effective p-Laplacian learning algorithm.

Author details

Xueqi Ma and Weifeng Liu*

*Address all correspondence to: liuwf@upc.edu.cn

China University of Petroleum (East China), Qingdao, China

References

[1] Kruskal JB. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika. 1964;**29**(1):1-27. DOI: 10.1007/bf02289565

[2] Borg I, Groenen P. Modern multidimensional scaling: Theory and applications. Journal of Educational Measurement. 2003;**40**(3):277-280

[3] Bishop CM, Svensén M, Williams CKI. GTM: The generative topographic mapping. Neural Computation. 1998;**10**(1):215-234. DOI: 10.1162/089976698300017953

[4] Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science. 2000;**290**(5500):2323-2326. DOI: 10.1126/science.290.5500.2323

[5] Bernstein M, Silva VD, Langford JC, et al. Graph approximations to geodesics on embedded manifolds. Stanford University. 2000;**24**(9):153-158

[6] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering. Advances in Neural Information Processing Systems. 2001;**14**(6):585-591

[7] Donoho DL, Grimes C. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. Proceedings of the National Academy of Sciences of the United States of America. 2003;**100**(10):5591. DOI: 10.1073/pnas.1031596100

[8] Zhang Z, Zha H. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. Society for Industrial and Applied Mathematics. 2005. DOI: 10.1137/s1064827502419154

[9] Wahba G. Spline models for observational data. Society for Industrial and Applied Mathematics. 1990. DOI: 10.1137/1.9781611970128

[10] Smola AJ, Bartlett P, Schölkopf B, et al. Regularization networks and support vector machines. Advances in Computational Mathematics. 2000;**13**(1):1-50

[11] Belkin M, Niyogi P, Sindhwani V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. JMLR. 2006

[12] Geng B, Tao D, Xu C, et al. Ensemble manifold regularization. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012;**34**(6):1227-1233. DOI: 10.1109/TPAMI.2012.57

[13] Liu W, Liu H, Tao D, Wang Y, Lu K. Manifold regularized kernel logistic regression for web image annotation. Neurocomputing. Elsevier BV. 2016 Jan;**172**:3-8. DOI: 10.1016/j.neucom.2014.06.096

[14] Luo Y, Tao D, Geng B, Xu C, Maybank SJ. Manifold regularized multitask learning for semi-supervised multilabel image classification. IEEE Transactions on Image Processing. 2013;**22**(2):523-536. DOI: 10.1109/tip.2012.2218825

[15] Ma X, Tao D, Liu W. Effective human action recognition by combining manifold regularization and pairwise constraints. Multimedia Tools and Applications. 2017. DOI: 10.1007/s11042-017-5172-1

[16] Hu W, Cheung G, Li X, Au OC. Graph-based joint denoising and super-resolution of generalized piecewise smooth images. In: 2014 IEEE International Conference on Image Processing (ICIP). IEEE; 2014. DOI: 10.1109/icip.2014.7025412

[17] Kim KI, Steinke F, Hein M. Semi-supervised regression using hessian energy with an application to semi-supervised dimensionality reduction. Advances in Neural Information Processing Systems. 2009:979-987

[18] Liu W, Tao D, Cheng J, Tang Y. Multiview hessian discriminative sparse coding for image annotation. Computer Vision and Image Understanding. 2014;**118**:50-60. DOI: 10.1016/j.cviu.2013.03.007

[19] Liu W, Tao D. Multiview hessian regularization for image annotation. IEEE Transactions on Image Processing. 2013;**22**(7):2676-2687. DOI: 10.1109/tip.2013.2255302

[20] Liu X, Shi J, Wang C. Hessian regularization based non-negative matrix factorization for gene expression data clustering. In: Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE. IEEE; 2015. pp. 4130-4133. DOI: 10.1109/embc.2015.7319303

[21] Zhu J, Shi J. Hessian regularization based semi-supervised dimensionality reduction for neuroimaging data of Alzheimer's disease. In: Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on. IEEE; 2014. pp. 165-168. DOI: 10.1109/ISBI.2014.6867835

[22] Liu W, Yang X, Tao D, Cheng J, Tang Y. Multiview dimension reduction via hessian multiset canonical correlations. Information Fusion. 2018;**41**:119-128. DOI: 10.1016/j.inffus.2017.09.001

[23] Liu W, Liu H, Tao D. Hessian regularization by patch alignment framework. Neurocomputing. 2016;**204**:183-188. DOI: 10.1016/j.neucom.2015.07.152

[24] Liu W, Zhang L, Tao D, Cheng J. Support vector machine active learning by hessian regularization. Journal of Visual Communication and Image Representation. 2017;**49**:47-56. DOI: 10.1016/j.jvcir.2017.08.001

[25] Liu W, Li Y, Lin X, Tao D, Wang Y. Hessian-regularized co-training for social activity recognition. Chen K, editor. PLoS One. 2014;**9**(9):e108474. DOI: 10.1371/journal.pone.0108474

[26] Tao D, Jin L, Liu W, et al. Hessian regularized support vector machines for mobile image annotation on the cloud. IEEE Transactions on Multimedia. 2013;**15**(4):833-844. DOI: 10.1109/tmm.2013.2238909

[27] Liu W, Ma T, Tao D, You J. HSAE: A hessian regularized sparse auto-encoders. Neurocomputing. 2016;**187**:59-65. DOI: 10.1016/j.neucom.2015.07.119

[28] Liu W, Liu H, Tao D, Wang Y, Lu K. Multiview hessian regularized logistic regression for action recognition. Signal Processing. 2015;**110**:101-107. DOI: 10.1016/j.sigpro.2014.08.002

[29] Amghibech S. Eigenvalues of the discrete p-Laplacian for graphs. Ars Combinatoria-Waterloo then Winnipeg. 2003;**67**:283-302

[30] Bouchala J. Resonance problems for p-Laplacian. Mathematics and Computers in Simulation. 2003;**61**(3–6):599-604. DOI: 10.1016/s0378-4754(02)00139-8

[31] Bühler T, Hein M. Spectral clustering based on the graph p-Laplacian. In: Proceedings of the 26th Annual International Conference on Machine Learning (ICML'09). ACM Press; 2009. DOI: 10.1145/1553374.1553385

[32] Luo D, Huang H, Ding C, Nie F. On the eigenvectors of p-Laplacian. Machine Learning. . Springer Nature. 2010;**81**(1):37-51. DOI: 10.1007/s10994-010-5201-z

[33] Liu W, Zha Z-J, Wang Y, Lu K, Tao D. p-Laplacian regularized sparse coding for human activity recognition. IEEE Transactions on Industrial Electronics. 2016:1-1. DOI: 10.1109/tie.2016.2552147

[34] Wagstaff K, Cardie C. Clustering with instance-level constraints. International Conference on Machine Learning. 2000:1103-1110

[35] Davidson I, Basu S. A survey of clustering with instance level constraints. ACM Transactions on Knowledge Discovery from Data. 2007:1-41

[36] Fu Z, Lu Z, Ip HHS, Lu H, Wang Y. Local similarity learning for pairwise constraint propagation. Multimedia Tools and Applications. 2014;**74**(11):3739-3758. DOI: 10.1007/s11042-013-1796-y

[37] Bar-Hillel A, Hertz T, Shental N, et al. Learning a mahalanobis metric from equivalence constraints. Journal of Machine Learning Research. 2005;**6**(6):937-965

[38] Mignon A, Jurie F. PCCA: A new approach for distance learning from sparse pairwise constraints. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2012. DOI: 10.1109/cvpr.2012.6247987

[39] Zhang D, Chen S, Zhou Z-H. Constraint score: A new filter method for feature selection with pairwise constraints. Pattern Recognition. 2008;**41**(5):1440-1451. DOI: 10.1016/j.patcog.2007.10.009

[40] Zhang D, Zhou ZH, Chen S. Semi-supervised dimensionality reduction. In: Siam International Conference on Data Mining, April 26–28, 2007, Minneapolis, Minnesota, USA. DBLP; 2007. pp. 11-393

[41] Cevikalp H, Verbeek JJ, Jurie F, et al. Semi-supervised dimensionality reduction using pairwise equivalence constraints. International Journal of System Dynamics Applications. 2008;**1**(3):489-496. DOI: 10.5220/0001070304890496

[42] Ding S, Jia H, Zhang L, Jin F. Research of semi-supervised spectral clustering algorithm based on pairwise constraints. Neural Computing and Applications. 2012;**24**(1):211-219. DOI: 10.1007 /s00521-012-1207-8

[43] Kalakech M, Biela P, Macaire L, Hamad D. Constraint scores for semi-supervised feature selection: A comparative study. Pattern Recognition Letters. 2011 Apr;**32**(5):656-665. DOI: 10.1016/j.patrec.2010.12.014

[44] Luo Y, Wen Y, Tao D. On combining side information and unlabeled data for heterogeneous multi-task metric learning. In: International Joint Conference on Artificial Intelligence. 2016. pp. 1809-1815

[45] Agarwal S, Lim J, Zelnik-Manor L, Perona P, Kriegman D, Belongie S. Beyond pairwise clustering. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE; 2005. DOI: 10.1109/cvpr.2005.89

[46] Ji R, Gao Y, Hong R, Liu Q, Tao D, Li X. Spectral-spatial constraint Hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing. 2014 Mar;**52**(3):1811-1824. DOI: 10.1109/tgrs.2013.2255297

[47] Yu J, Rui Y, Tao D. Click prediction for web image reranking using multimodal sparse coding. IEEE Transactions on Image Processing. 2014;**23**(5):2019-2032. DOI: 10.1109/tip.2014.2311377

[48] Huang Y, Liu Q, Metaxas D. Video object segmentation by hypergraph cut. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2009. DOI: 10.1109/cvprw.2009.5206795

[49] Zien JY, Schlag MDF, Chan PK. Multilevel spectral hypergraph partitioning with arbitrary vertex sizes. Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on, 1997;**18**(9):1389-1399. DOI:10.1109/iccad.1996.569592

[50] Rodríguez JA. On the Laplacian spectrum and walk-regular hypergraphs. Linear and Multilinear Algebra. 2003;**51**(3):1-1. DOI: 10.1080/03081080306587

[51] Bolla M. Spectra, Euclidean representations and clusterings of hypergraphs. Discrete Mathematics. 1993;**117**(1–3):19-39. DOI: 10.1016/0012-365x(93)90322-k

[52] Zhou D, Huang J. Learning with hypergraphs: Clustering, classification, and embedding. In: International Conference on Neural Information Processing Systems. MIT Press; 2006. pp. 1601-1608

[53] Guo Y, Tao D, Liu W, Cheng J. Multiview Cauchy estimator feature embedding for depth and inertial sensor-based human action recognition. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2017;**47**(4):617-627. DOI: 10.1109/tsmc.2016.2617465

[54] Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The pascal visual object classes (VOC) challenge. International Journal of Computer Vision. 2009;**88**(2):303-338. DOI: 10.1007/s11263-009-0275-4

[55] Liu W, Ma X, Zhou Y, Tao D, Cheng J. P-Laplacian regularization for scene recognition. IEEE Transactions on Cybernetics. 2018. DOI: 10.1109/TCYB.2018.2833843

[56] Quattoni A, Torralba A. Recognizing indoor scenes. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2009. DOI: 10.1109/cvprw.2009.5206537

[57] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). IEEE; 2006. DOI: 10.1109/cvpr.2006.68

[58] Zhou D, Schölkopf B. Regularization on discrete spaces. In: Joint Pattern Recognition Symposium. Berlin, Heidelberg: Springer; 2005. pp. 361-368

[59] Takeuchi H. The spectrum of the p-Laplacian and p-harmonic morphisms on graphs. Illinois Journal of Mathematics. 2003;**47**(2003):939-955

[60] Yang Y, Newsam S. Bag-of-visual-words and spatial extensions for land-use classification. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '10. 2010. DOI: 10.1145/1869790.1869829

[61] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Computer Science. 2014

[62] Huang Y, Liu Q, Zhang S, Metaxas DN. Image retrieval via probabilistic hypergraph ranking. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE; 2010. DOI: 10.1109/cvpr.2010.5540012

[63] Wang M, Hua XS, Hong R, Tang J, et al. Unified video annotation via multigraph learning. IEEE Transactions on Circuits and Systems for Video Technology. 2009;**19**(5):733-746. DOI: 10.1109/tcsvt.2009.2017400