Cognitive Computation (2019) 11:841 854
https://doi.org/10.1007/s12559-019-09637-z

Ensemble *p*-Laplacian Regularization for Scene Image Recognition

Xueqi Ma[1] · Weifeng Liu[1] · Dapeng Tao[2] · Yicong Zhou[3]

Received: 28 March 2018 / Accepted: 10 March 2019 / Published online: 22 March 2019

Abstract
Recently, manifold regularized semi-supervised learning (MRSSL) received considerable attention, because it successfully exploits the geometry of the intrinsic data probability distribution to leverage the performance of a learning model. 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 determine the fitting graph *p*-Lapalcian, i.e., the parameter *p*, which is a critical factor for the performance of graph *p*-Laplacian. Therefore, we develop an ensemble *p*-Laplacian regularization (EpLapR) 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. And then, we conduct semi-supervised learning framework on the fused graph. Extensive experiments on UC-Merced dataset and Scene 15 dataset demonstrate the effectiveness and efficiency of the proposed method.

Keywords Manifold regularization · Semi-supervised learning · Ensemble *p*-Laplacian regularization

Introduction

With rapid advances in storage devices and mobile networks, large-scale multimedia data have become available to ordinary users. However, in practical applications, e.g., human action recognition [19, 20], scene classification [21], text categorization [1], and video annotation and retrieval [2], the labeled samples are always insufficient, though vast amounts of unlabeled samples are readily accessible and provide auxiliary information. Semi-supervised learning (SSL) aiming to exploit both labeled data and unlabeled data is designed to address such problem. In SSL [3–8], it is assumed that nearby samples are likely to share the same label. The manifold regularization [9] is one of the most representative works, which assumes that the geometry of the intrinsic data probability distribution is supported on the low-dimensional manifold.

In order to build better classifiers, the typical MRSSL algorithms exploit the intrinsic geometry of the labeled and unlabeled samples, then naturally captured by a graph. Therefore, researchers paid their attention to build a good graph to capture the essential data structure. Laplacian regularization (LapR) [9, 17] is one prominent manifold regularization based SSL algorithm, which explores the geometry of the probability distribution by using the graph Laplacian. LapR-based SSL algorithms have been widely used in many applications. Luo et al. [23] employed manifold regularization to smooth the functions along the data manifold for multitask learning. Hu et al. [22] introduced graph Laplacian regularization for joint denoising and superresolution of generalized piecewise smooth images. Jiang et al. [24] presented a multi-manifold method for recognition by exploring the local geometric structure of samples. Another relatively new prior is the Hessian regularization (HesR), which has been shown empirically to perform well in a wide range of inverse problems [10–12]. In comparison with LapR, HesR steers the values of function varying linearly in reference to the geodesic distance. In result, HesR can be more accurate to describe the underlying manifold of data. However, Hessian estimation will be inaccurate, while it has poor quality of the local fit for each data point [14]. The *p*-Laplacian [15] [16] is nonlinear generalization of general graph Laplacian and has tighter isoperimetric inequality. In particular, Bühler et al. [13] 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

✉ Weifeng Liu
liuwf@upc.edu.cn

[1] China University of Petroleum (East China), Qingdao 266580, China

[2] Yunnan University, Kunming 650091, China

[3] Faculty of Science and Technology, University of Macau, Macau, China

Table 1 Iterative solution method for EpLapR

Step1: initialize $\mu_k \in \boldsymbol{R}^N$.
Step2: update α according to Eq. (21).
Step3: based on the updated α, re-calculate μ_k according to Eq. (22)
Step4: repeat from step 2 until convergence.

exploiting. An efficient approximation method of graph p-Laplacian has been proposed in [30]. However, the parameter p of graph p-Laplacian is difficult to determine.

In manifold regularized SSL, the manifold is determined by the graph with the predefined hyperparameters. Unfortunately, it cannot define an objective function to choose graph hyperparameters for intrinsic manifold estimation. In general, the cross-validation [25] has been widely utilized for parameter selection. But, to the best of our knowledge, this method that selects parameters in a discrete and limited parameter space lacks the ability to approximate the optimal solution tries. Furthermore, the performance of the classification model is weakly relevant to the difference between the intrinsic and approximated manifolds. Thus, the pure cross-validation-based parameter selection cannot perform well on model learning. An automatic and full approximation of the intrinsic manifold of the data distribution will be valuable for the SSL methods.

In this paper, we propose an ensemble p-Laplacian regularized method, which combines a series of graphs p-Laplacian. By a conditionally optimal way, the proposed method learns to assign suitable weights on graphs and finally construct an optimal fused graph. The fused graph can sufficiently approximate the intrinsic manifold by the complementation of graphs p-Laplacian. We build the graph-regularized classifiers, including support vector machines (SVM) and kernel least squares (KLS) as special cases for scene image recognition. Experiments on the UC-Merced dataset [27] and Scene 15 dataset [34] validate the effect of proposed method compared with the popular algorithms, including Laplacian regularization (LapR), Hessian regularization (HLapR), and p-Laplacian regularization (pLapR).

The rest of the paper is organized as follows. "Related Work" briefly reviews related work on manifold regularization and p-Laplacian learning. "Method" proposes EpLapR. "Example Algorithms" presents the EpLapR for KLS and SVM. "Experiments" provides the experimental results and analysis on UC-Merced dataset and Scene 15 dataset. Finally, "Conclusion" gives the conclusions.

Related Work

The proposed EpLapR is motivated by MRSSL and p-Laplacian learning. This section briefly describes the related works for better understanding.

Manifold Regularization

Suppose, the labeled samples are (x, y) pairs drawn from a probability distribution P, and unlabeled samples x are drawn according to the marginal distribution P_x of P. It assumes that the probability distribution of data is supported on a submanifold of the ambient space. In general, the manifold regularization defines a similarity graph over labeled and unlabeled examples and incorporates it as an additional regularization term. Hence, the manifold regularization framework has two regularization terms: one controlling the complexity measures in an appropriately chosen Reproducing Kernel Hilbert Space (RKHS) and the other controlling the additional information about the geometric structure of the marginal. The objective function of the framework is defined as:

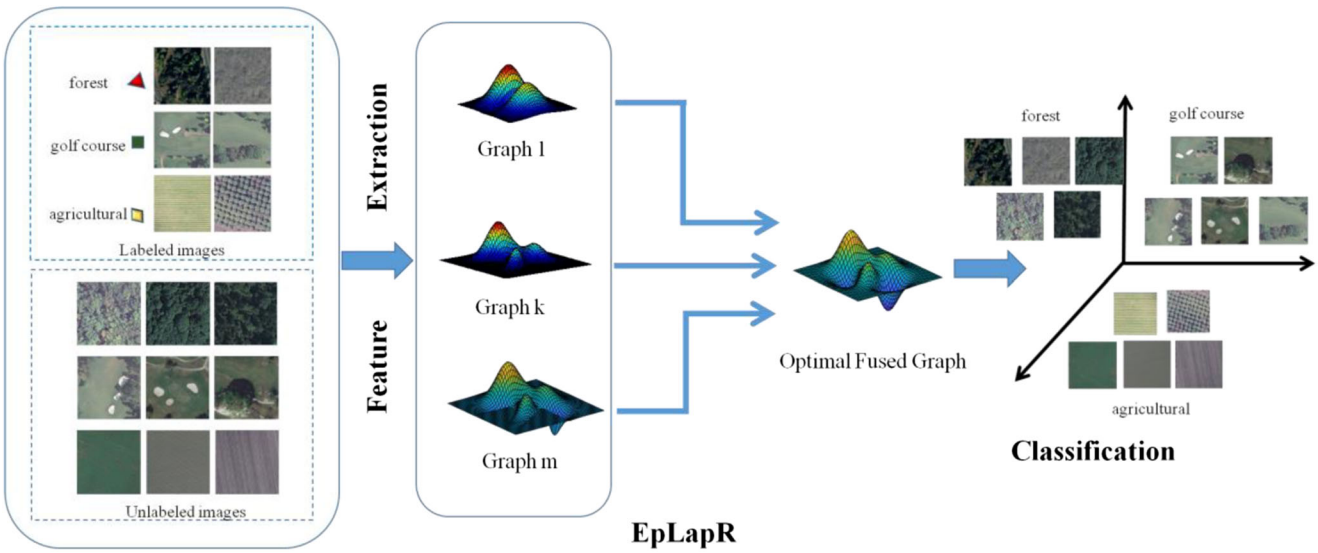

Fig. 1 The framework of EpLapR for remote sensing image classification

Fig. 2 Some example images of UC-Merced dataset. The dataset totally has 21 remote sensing categories

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

where V is a loss function, such as the hinge loss function $\max[0, 1 - y_i f(x_i)]$ for SVM. $\|f\|_K^2$ is used to control the complexity of the classification model, while $\|f\|_I^2$ is an appropriate penalty term, which is approximated by the graph matrix (e.g., graph Laplacian L, $L = D - W$, where W_{ij} is the weight vector, the diagonal matrix D is given by $D_{ii} = \sum_{j=1}^{n} W_{ij}$) and the function prediction. The parameters Υ_A and Υ_I control the complexity of the function in the ambient space and the intrinsic geometry, respectively.

p-Laplacian Regularization

As a nonlinear generalization of the standard graph Laplacian, graph p-Laplacian has the superiority on local structure preserving.

As we all know, the standard graph Laplacian Δ_2 can be defined as the operator by inducing the following quadratic form for a function f

$$\langle f, \Delta_2 f \rangle = 1/2 \sum_{i,j \in V} w_{ij} \left(f_i - f_j \right)^2 \qquad (2)$$

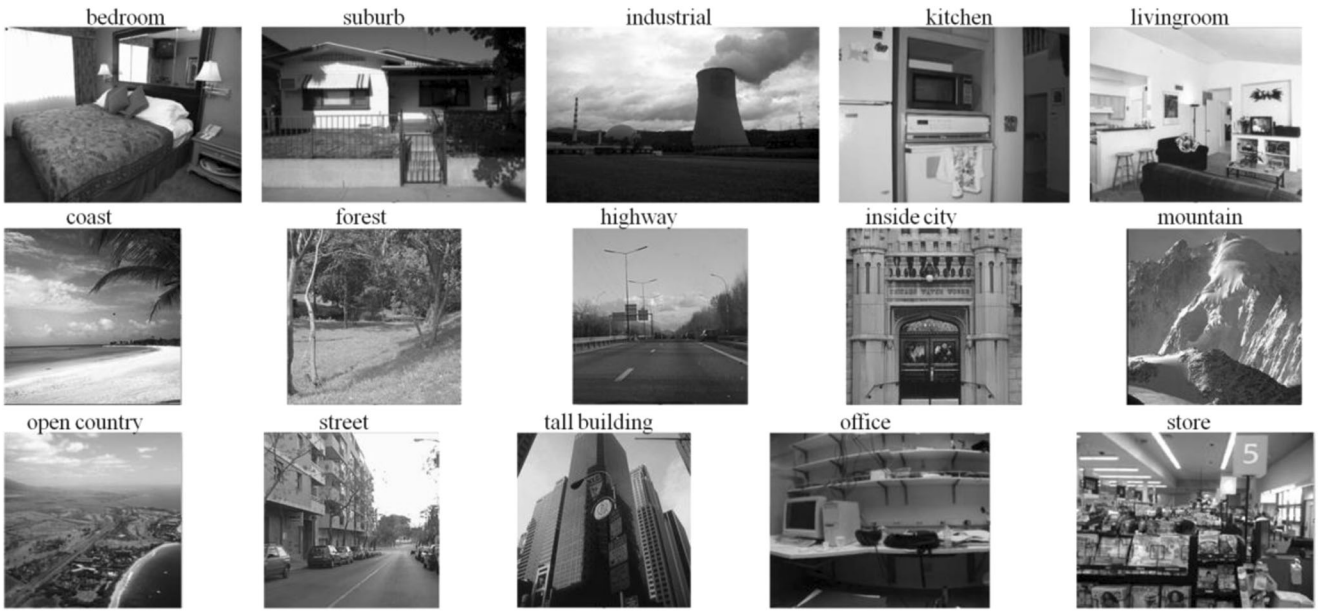

Fig. 3 Some example images of Scene 15 dataset. The dataset totally has 15 scene classes

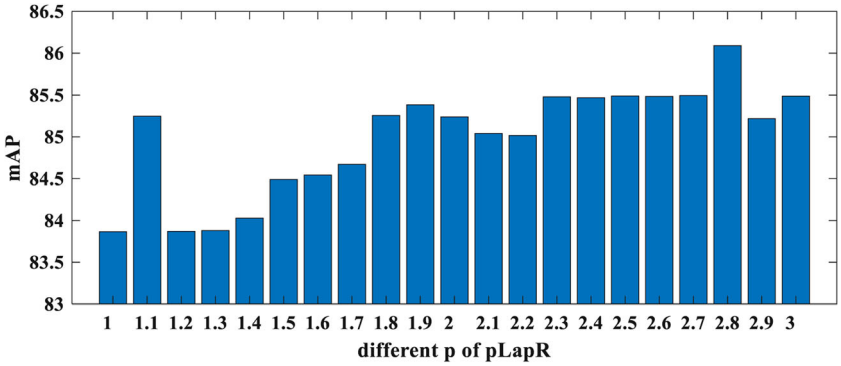

Fig. 4 Performance of mAP with different p values on validation set of UC-Merced dataset

where the vertices V represent the points in the feature space, and the positive edge weights w_{ij} encode the similarity of points (i, j).

Similar to the graph Laplacian [13], the unnormalized p-Laplacian Δ_p (for $p > 1$) can be defined by:

$$\langle f, \Delta_p f \rangle = 1/2 \sum_{i,j \in V} w_{ij} \left(f_i - f_j \right)^p \tag{3}$$

or

$$\left(\Delta_p f \right)_i = \sum_{j \in V} w_{ij} \phi_p \left(f_i - f_j \right) \tag{4}$$

where ϕ_p is defined as $\phi_p(x) = |x|^{p-1} \mathrm{sig}(x)$. With $p = 2$, the p-Laplacian becomes the standard graph Laplacian.

Bühler and Hein [13] 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: for value $p > 1$,

$$RCC \leq RCC^* \leq p \left(\max_{i \in V} d_i \right)^{\frac{p-1}{p}} RCC^{\frac{1}{p}} \tag{5}$$

or

$$NCC \leq NCC^* \leq p NCC^{\frac{1}{p}} \tag{6}$$

where RCC^* and NCC^* are the ratio/normalized Cheeger cut values obtained by tresholding the second eigenvector of the unnormalized/normalized p-Laplacian; d_i is the degree of vertex i; RCC and NCC are the optimal ratio/normalized Cheeger cut values. This proves that in the limit as $p \to 1$, the cut found by thresholding the second eigenvector of the graph p-Laplacian converges to the optimal Cheeger cut.

In mathematic community, discrete p-Laplacian has been studied in a general regularization framework. In [28], the objective function of a general discrete p-Laplacian regularization framework can be computed as follows:

$$f^* = \mathrm{argmin}_{f \in \mathcal{H}(V)} \left\{ \mathcal{S}_p(f) + \mu \| f - y \|^2 \right\} \tag{7}$$

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

In [26], Luo et al. proposed full eigenvector analysis of p-Laplacian and obtained a natural global embedding for multiclass clustering problems, the whole eigenvector analysis of p-Laplacian was achieved by an efficient gradient descend optimization approach as:

$$\min_{\mathcal{F}} J_E(\mathcal{F}) = \sum_k \frac{\sum_{ij} w_{ij} \left| f_i^k - f_j^k \right|^p}{\| f^k \|_p^p} \quad s.t. \quad \mathcal{F}^T \mathcal{F} = I \tag{8}$$

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

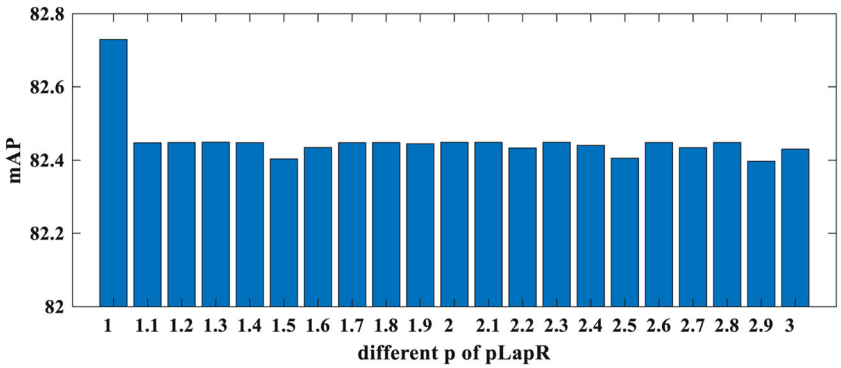

Fig. 5 Performance of mAP with different p values on validation set of Scene 15 dataset

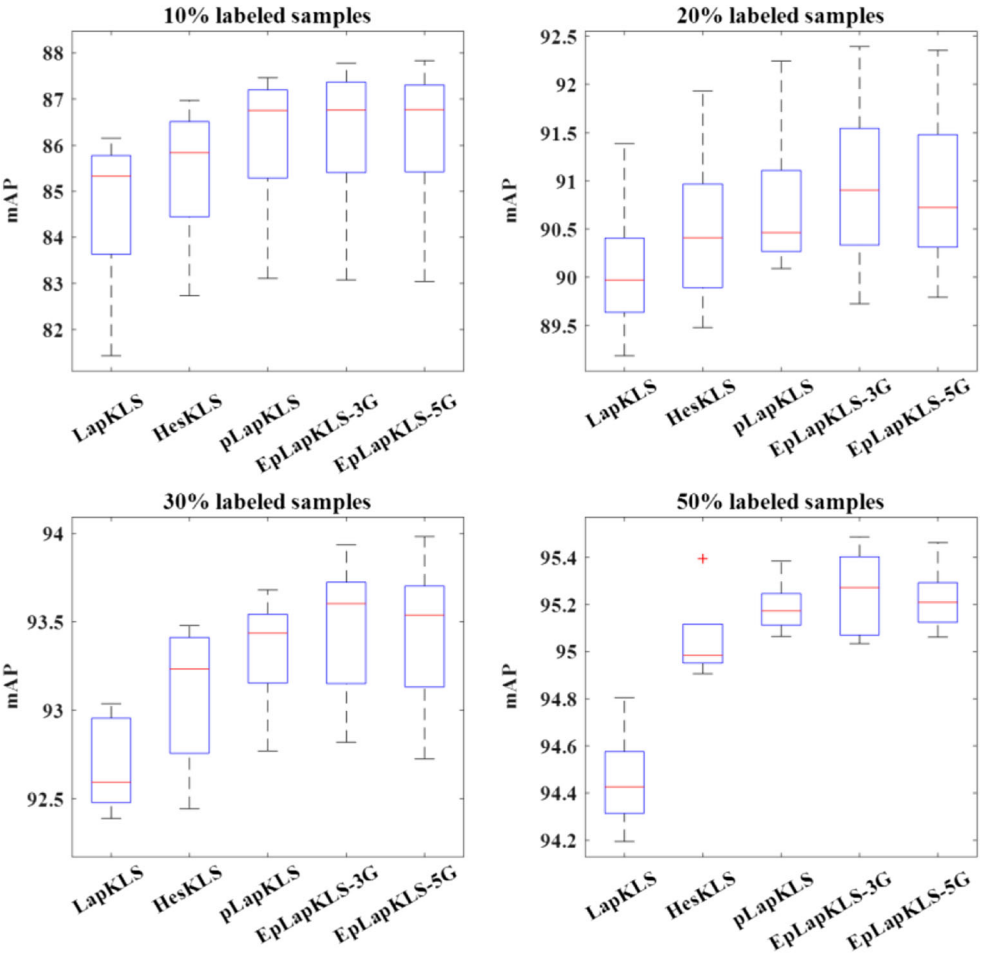

Fig. 6 mAP performance of different algorithms on KLS method of UC-Merced dataset

Liu et al. [29] proposed *p*-Laplacian regularized sparse coding for human activity recognition.

Liu et al. [30] illustrated the differences of semi-supervised regression by using LapR, HesR, and pLapR for fitting two points on the 1-D spiral. The results show the pLapR can fit the data exactly and extrapolates smoothly to unseen data with the geodesic distance. Specially, they proposed an efficient approximation method of graph *p*-Laplacian and built *p*-Laplacian regularization framework.

Ma et al. [31] presented an efficient and effective approximation algorithm of hypergraph *p*-Laplacian and then proposed hypergraph *p*-Laplacian regularization (HpLapR) to preserve the geometry of the probability distribution.

Method

The proposed EpLapR approximates the manifold of the data distribution by fusing a set of graph *p*-Laplacian. First, we show the fully approximation of the graph *p*-Laplacian. Then, we propose the EpLapR.

Approximation of Graph *p*-Laplacian

We approximate the graph *p*-Laplacian (L^{p}) by the fully analysis of the eigenvalues and eigenvectors. In [13], the computation of eigenvalue and the corresponding eigenvector on nonlinear operator $\Delta_{\mathrm{p}}^{\mathrm{w}}$ can be solved by the theorem:

The functional F_{p} has a critical point at f if and only if f is an eigenvector of $\Delta_{\mathrm{p}}^{\mathrm{w}}$; the corresponding eigenvalue λ_{p} is given by $\lambda_{\mathrm{p}} = F_{\mathrm{p}}(f)$. The definition of F_{p} is given as:

$$F_{\mathrm{p}}(f) = \frac{\sum_{\mathrm{ij}} w_{\mathrm{ij}} \left| f_{\mathrm{i}} - f_{\mathrm{j}} \right|^{\mathrm{p}}}{2 \| f \|_{\mathrm{p}}^{\mathrm{p}}} \tag{9}$$

where

$$\| f \|_{\mathrm{p}}^{\mathrm{p}} = \sum_{\mathrm{i}} |f_{\mathrm{i}}|^{\mathrm{p}}.$$

The above theorem serves as the foundational analysis of eigenvectors and eigenvalues. Moreover, we have $F_{\mathrm{p}}(\alpha f) = F_{\mathrm{p}}(f)$ for all real value α.

Suppose that the graph *p*-Laplacian has *n* eigenvectors (f^{*1}, f^{*2}, $\cdots$, f^{*n}) associated with unique eigenvalues $\left(\lambda_1^*, \lambda_2^*, \cdots, \lambda_{\mathrm{n}}^*\right)$. According to the above theorem, if we want

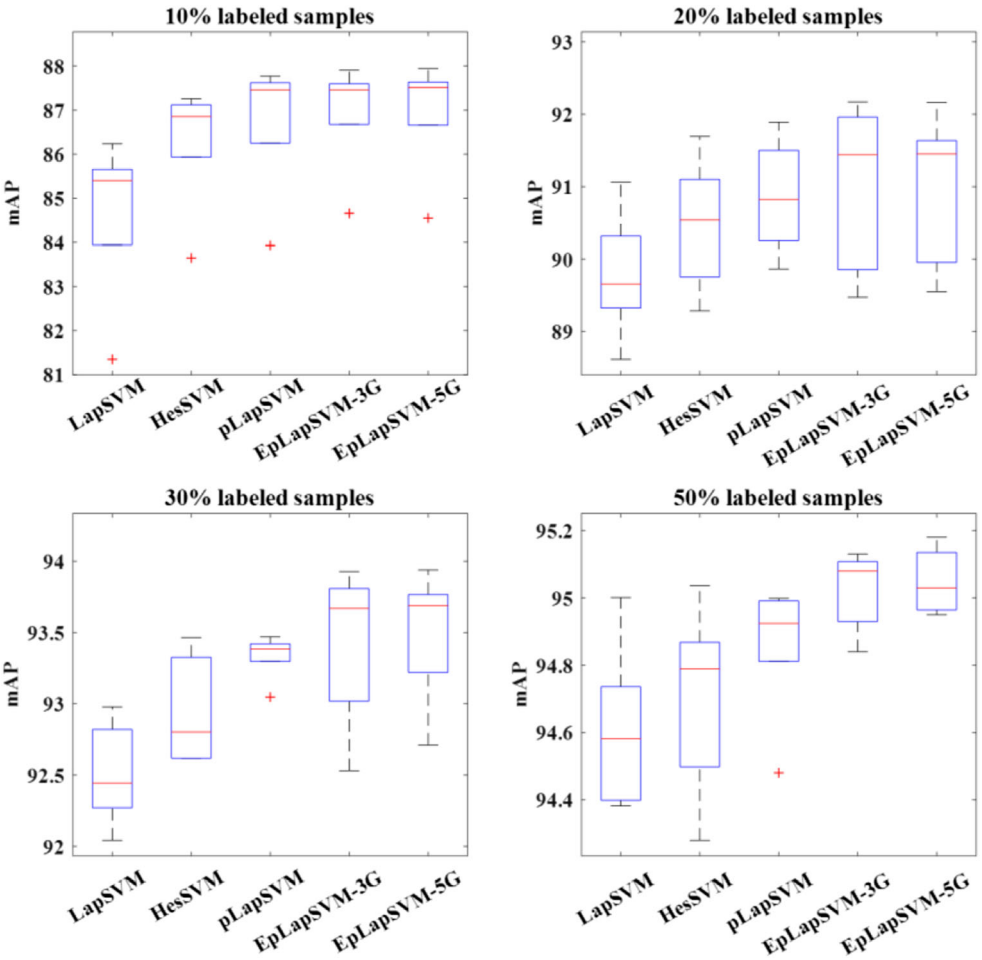

Fig. 7 mAP performance of different algorithms on SVM method of UC-Merced dataset

to get all eigenvectors and eigenvalues of graph p-Laplacian, we have to find all critical points of the functional F_p. Therefore, we exploit the full eigenvectors space by solving local solution of the following minimization problem:

$$\min_{\mathcal{F}} J(\mathcal{F}) = \sum_k F_p(f^k) \; s.t. \quad \sum_i \phi_p(f_i^k)\phi_p(f_i^l) = 0, k \neq l \quad (10)$$

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

Rewrite the optimization problem (10), we analyze the full eigenvectors by solving the following graph p-Laplacian embedding problem:

$$\min_{\mathcal{F}} J_E(\mathcal{F}) = \sum_k \frac{\sum_{ij} w_{ij} \left| f_i^k - f_j^k \right|^p}{\|f^k\|_p^p} \quad s.t. \quad \mathcal{F}^T\mathcal{F} = I \quad (11)$$

The gradient of J_E with respect to f_i^k yields the following equation:

$$\frac{\partial J_E}{\partial f_i^k} = \frac{1}{\|f^k\|_p^p}\left[\sum_j w_{ij}\phi_p\left(f_i^k - f_j^k\right) - \frac{\phi_p(f_i^k)}{\|f^k\|_p^p}\right] \quad (12)$$

The problem (11) can be solved with the gradient descend optimization. However, if we simply use the gradient descend approach, the solution f^k might not be orthogonal [26]. So, the gradient is modified in Eq. (13) in order to enforce the orthogonality.

$$G = \frac{\partial J_E}{\partial \mathcal{F}} - \mathcal{F}\left(\frac{\partial J_E}{\partial \mathcal{F}}\right)^T \mathcal{F} \quad (13)$$

Meanwhile, the full eigenvalue $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_n)$ can be computed by $\lambda_k = \frac{\sum_{ij} w_{ij} \left| f_i^k - f_j^k \right|^p}{\|f^k\|_p^p}$. Finally, the approximate L^p can be computed by $L^p = \mathcal{F}\lambda\mathcal{F}^T$. We summarize the approximation of graph p-Laplacian in Algorithm 1. In the algorithm, the step length α is set to be $\alpha = 0.01\frac{\sum_{ik}|\mathcal{F}_{ik}|}{\sum_{ik}|G_{ik}|}$.

We can get that if $\mathcal{F}^T\mathcal{F} = I$, then using the simple gradient descend approach can guarantee to give a feasible solution. Since Laplacian L is symmetric, we have $\mathcal{F}^T\mathcal{F} = I$ for initialization, and from Algorithm 1, we have $\mathcal{F}^{t+1} = \mathcal{F}^t - \alpha G$, thus, $\left(\mathcal{F}^{t+1}\right)^T\mathcal{F}^{t+1} = (\mathcal{F}^t - \alpha G)^T (\mathcal{F}^t - \alpha G) = (\mathcal{F}^t)^T\mathcal{F}^t - \eta\left[G^T\mathcal{F}^t + (\mathcal{F}^t)^T G\right]$. Here,

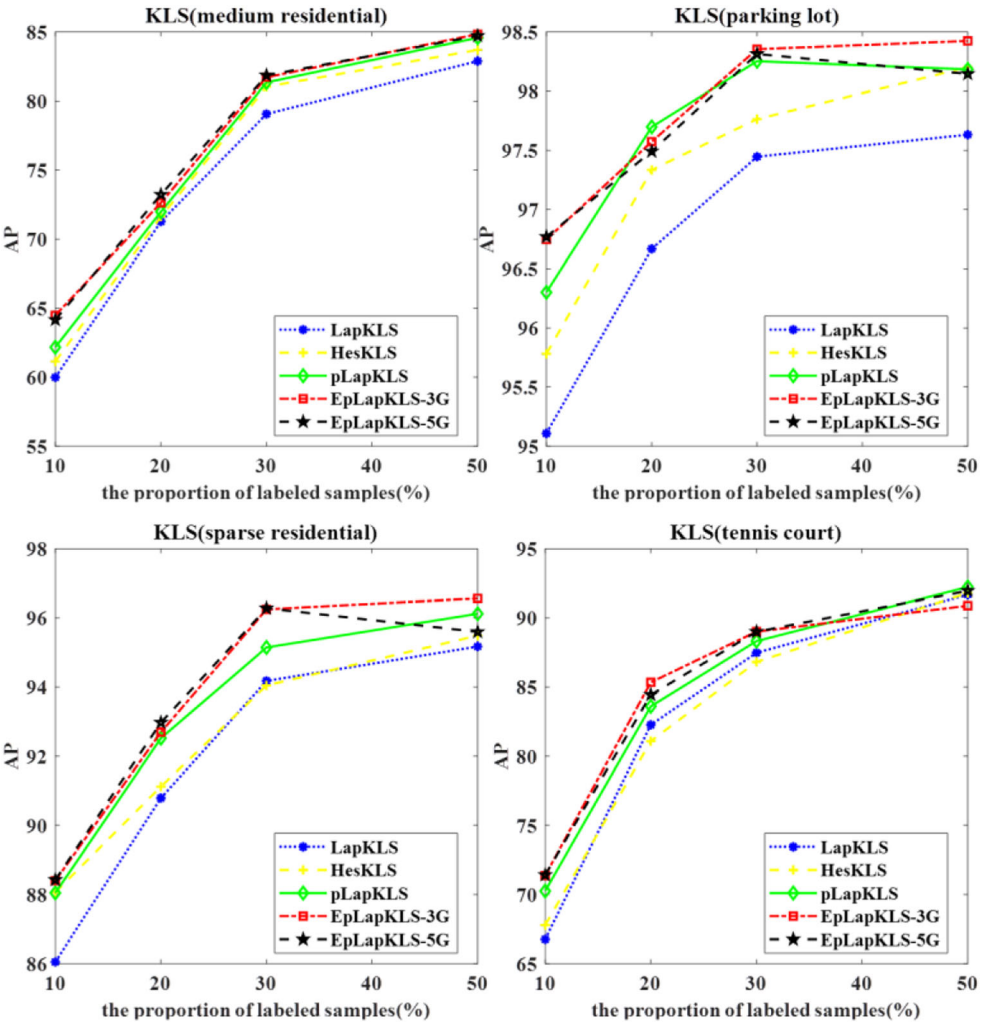

Fig. 8 AP performance of different KLS methods on several classes of UC-Merced dataset

$$G^{\mathrm{T}}\mathcal{F}^{\mathrm{t}}+(\mathcal{F}^{\mathrm{t}})^{\mathrm{T}}G$$
$$=\left(\frac{\partial J_{\mathrm{E}}}{\partial\mathcal{F}}-\mathcal{F}\left(\frac{\partial J_{\mathrm{E}}}{\partial\mathcal{F}}\right)^{\mathrm{T}}\mathcal{F}\right)^{\mathrm{T}}\mathcal{F}^{\mathrm{t}}+(\mathcal{F}^{\mathrm{t}})^{\mathrm{T}}\left[\frac{\partial J_{\mathrm{E}}}{\partial\mathcal{F}}-\mathcal{F}\left(\frac{\partial J_{\mathrm{E}}}{\partial\mathcal{F}}\right)^{\mathrm{T}}\mathcal{F}\right]$$
$$=\left(\frac{\partial J_{\mathrm{E}}}{\partial\mathcal{F}}\right)^{\mathrm{T}}\mathcal{F}^{\mathrm{t}}-(\mathcal{F}^{\mathrm{t}})^{\mathrm{T}}\frac{\partial J_{\mathrm{E}}}{\partial\mathcal{F}}-\left(\frac{\partial J_{\mathrm{E}}}{\partial\mathcal{F}}\right)^{\mathrm{T}}\mathcal{F}^{\mathrm{t}}+(\mathcal{F}^{\mathrm{t}})^{\mathrm{T}}\frac{\partial J_{\mathrm{E}}}{\partial\mathcal{F}}=0$$

So $\left(\mathcal{F}^{\mathrm{t+1}}\right)^{\mathrm{T}}\mathcal{F}^{\mathrm{t+1}}=(\mathcal{F}^{\mathrm{t}})^{\mathrm{T}}\mathcal{F}^{\mathrm{t}}=\mathrm{I}$. The orthogonality of solution $\mathcal{F}$ is enhanced by G, and solution $\mathcal{F}$ is the feasible solution.

Assume the t^{th} iteration $J_{\mathrm{E}}(\mathcal{F}^{\mathrm{t}})$ and $(t-1)^{\mathrm{th}}$ iteration $J_{\mathrm{E}}\left(\mathcal{F}^{\mathrm{t-1}}\right)$. For this degenerated problem, we have $J_{\mathrm{E}}(\mathcal{F}^{\mathrm{t}})=J_{\mathrm{E}}\left(\mathcal{F}^{\mathrm{t-1}}-\alpha G\right)\leq J_{\mathrm{E}}\left(\mathcal{F}^{\mathrm{t-1}}\right)$. Since $J_{\mathrm{E}}(\mathcal{F})\geq 0$, thus, our algorithm is guaranteed to converge.

EpLapR

Consider the MRSSL setting, where two sets of samples X are available, i.e., l labeled samples $\{(x_{\mathrm{i}},y_{\mathrm{i}})\}_{\mathrm{i}=1}^{\mathrm{l}}$ and u unlabeled samples $\{(x_{\mathrm{j}})\}_{\mathrm{j}=\mathrm{l}+1}^{\mathrm{l}+\mathrm{u}}$, for a total of $n=l+u$ samples. Class labels are given in $Y=\{y_{\mathrm{i}}\}_{\mathrm{i}=1}^{\mathrm{l}}$, where $y_{\mathrm{i}}\in\{\pm 1\}$. Typically, $l\ll u$ and we focus on predicting the labels of unseen examples.

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

$$f^{*}=\arg\min_{\mathrm{f}\in\mathrm{H_K}}\frac{1}{\mathrm{l}}\sum_{\mathrm{i}=1}^{\mathrm{l}}V(x_{\mathrm{i}},y_{\mathrm{i}},f)+\Upsilon_{\mathrm{A}}\|f\|_{\mathrm{K}}^{2}+\frac{\Upsilon_{I}}{n^{2}}\mathbf{f}^{\mathrm{T}}\boldsymbol{L}\mathbf{f}\quad(14)$$

Here, $\mathbf{f}$ is given as $\mathbf{f}=[f(x_1),f(x_2),\cdots,f(x_{\mathrm{l+u}})]^{\mathrm{T}}$, $\boldsymbol{L}$ is the optimal fused graph with $\boldsymbol{L}=\sum_{\mathrm{k}=1}^{\mathrm{m}}\mu_{\mathrm{k}}L_{\mathrm{k}}^{\mathrm{p}}$, s.t. $\sum_{\mathrm{k}=1}^{\mathrm{m}}\mu_{\mathrm{k}}=1,\mu_{\mathrm{k}}\geq 0$, for $k=1,\cdots,m$. Where we define a set of candidate graph p-Laplacian $C=\left\{L_1^{\mathrm{p}},\cdots,L_{\mathrm{m}}^{\mathrm{p}}\right\}$ and denote the convex hull of set A as: $\mathbf{conv}\,A=\{\psi_1x_1+\cdots+\psi_{\mathrm{m}}x_{\mathrm{m}}|\,\psi_1+\cdots+\psi_{\mathrm{m}}=1,x_{\mathrm{i}}\in A,\psi_{\mathrm{i}}\geq 0,i=1,\cdots,m\}$, where $A=\{x_1,\cdots,x_{\mathrm{m}}\}$. Therefore, we have $\boldsymbol{L}\in\mathbf{conv}C$, where $\mathbf{conv}C=\left\{\mu_1L_1^{\mathrm{p}}+\cdots+\mu_{\mathrm{m}}L_{\mathrm{m}}^{\mathrm{p}}|\,\mu_1+\cdots+\mu_{\mathrm{m}}=1,L_{\mathrm{i}}^{\mathrm{p}}\in C,\mu_{\mathrm{i}}\geq 0,i=1,\cdots,m\right\}$.

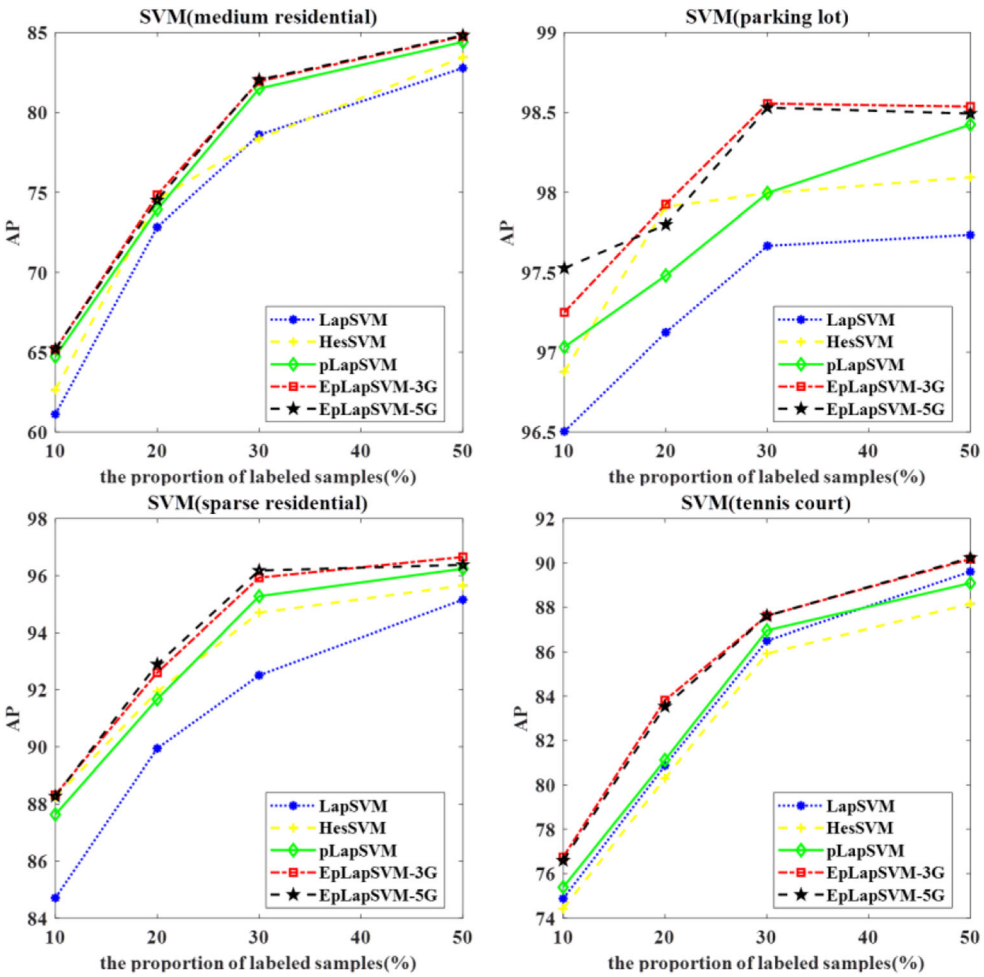

Fig. 9 AP performance of different SVM methods on several classes of UC-Merced dataset

To avoid the parameter μ_k overfitting to one graph [18], we make a relaxation by changing μ_k to μ_k^{γ} and obtain the optimization problem as:

$$f^* = \arg\min_{f\in H_K} \frac{1}{l}\sum_{i=1}^{l} V(x_i, y_i, f) + \Upsilon_A \|f\|_K^2 + \frac{\Upsilon_I}{n^2}\mathbf{f}^T\left(\sum_{k=1}^{m}\mu_k^{\gamma}L_k^p\right)\mathbf{f} \text{ s.t. } \sum_{k=1}^{m}\mu_k = 1,\ \mu_k \geq 0,\ \text{for } k = 1,\cdots,m \tag{15}$$

Next, we present theoretical analysis for EpLapR.

Theorem 1: For $\boldsymbol{L} \in \mathbf{conv}C$, the solution of the problem (15) exists and admits the following representation:

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

which is an expansion in terms of the labeled and unlabeled examples. Where the kernel matrix $\boldsymbol{K}$ with $\boldsymbol{K_{ij}} = \boldsymbol{K}(x_i, x_j)$ is symmetric positive definite; α^* is the coefficient.

The represented theorem shows that the solution of Eq. (15) exists and has the general form of Eq. (16) under a fixed μ.

So, we rewrite the objective function as

$$f^* = \arg\min_{f\in H_K} \frac{1}{l}\sum_{i=1}^{l} V(x_i, y_i, f) + \Upsilon_A \|f\|_K^2 + \frac{\Upsilon_I}{n^2}\alpha^T K\left(\sum_{k=1}^{m}\mu_k^{\gamma}L_k^p\right)K\alpha \text{ s.t.} \sum_{k=1}^{m}\mu_k = 1,\ \mu_k \geq 0,\ \text{for } k = 1,\cdots,m \tag{17}$$

Example Algorithms

Generally, the proposed EpLapR can be applied to variant MRSSL-based applications with different choices of loss function $V(x_i, y_i, f)$. In this section, we apply EpLapR to KLS and SVM.

EpLapR Kernel Least Squares (EpLapKLS)

By employing the least square loss in optimization problem (14), we can present the EpLapKLS model defined in Eq. (18) as follows:

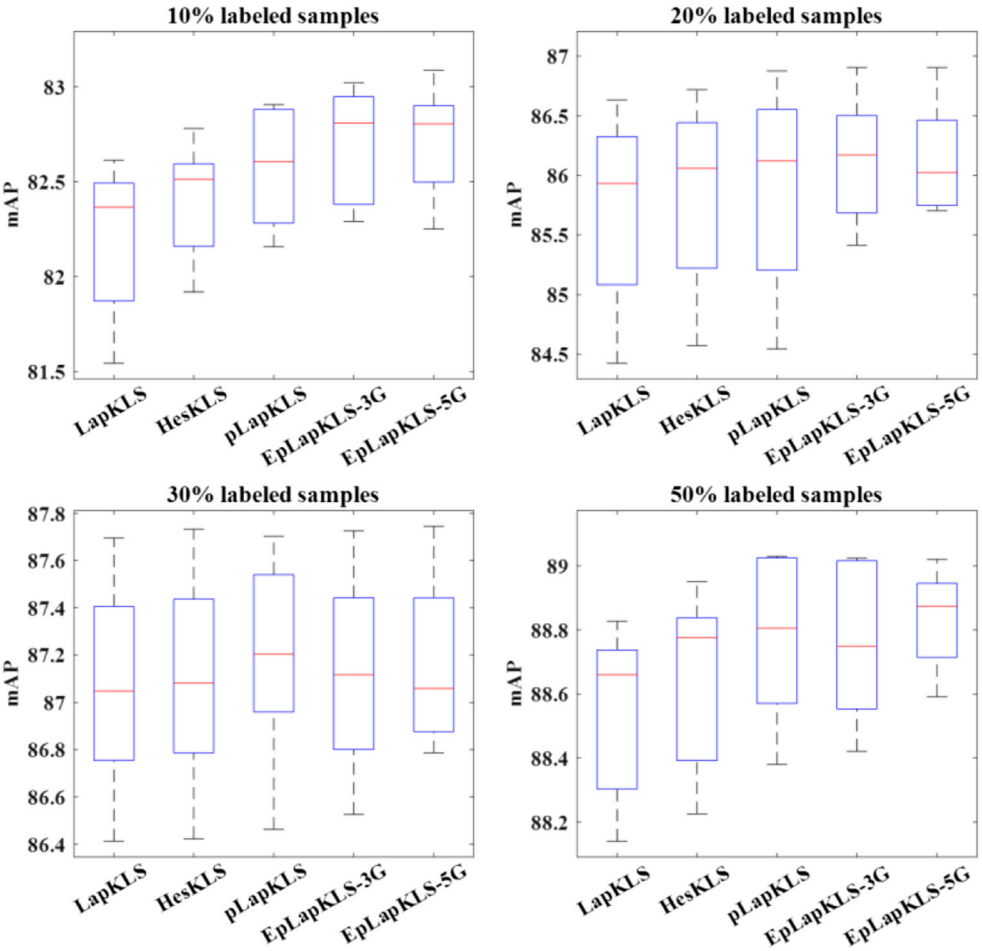

Fig. 10 mAP performance of different algorithms on KLS method of Scene 15 dataset

$$f^* = \arg\min_{f\in H_K} \frac{1}{l}\sum_{i=1}^{l} V(y_i - f(x_i))^2 + \Upsilon_A \alpha^T K\alpha + \frac{\Upsilon_I}{n^2}\alpha^T K\left(\sum_{k=1}^{m}\mu_k^{\gamma} L_k^p\right)K\alpha \quad \text{s.t.} \sum_{k=1}^{m}\mu_k = 1,\ \ \mu_k \geq 0,\ \ \text{for } k = 1,\cdots,m \tag{18}$$

Then, we obtain the partial derivative of the objective function with respect to μ_k and α as follows:

$$\frac{\partial f}{\partial \mu_k} = \frac{\Upsilon_I}{n^2}\alpha^T \boldsymbol{K}\left(\gamma {\mu_k}^{\gamma-1}\right)L_k^p \boldsymbol{K}\alpha \tag{19}$$

$$\frac{\partial f}{\partial \alpha} = \frac{1}{l}(Y - J\boldsymbol{K}\alpha)^T(-J\boldsymbol{K}) + \left(\Upsilon_A \boldsymbol{K} + \frac{\Upsilon_I}{n^2}\boldsymbol{K}\boldsymbol{L}\boldsymbol{K}\right)\alpha \tag{20}$$

Here, we adopt a process that iteratively updates α and μ_k to minimize f^*. Firstly, when μ_k is fixed, Eq. (20) turns to $\text{argmin}_{\mu_k} f$, from which we can derive that

$$\alpha^* = \left(J\boldsymbol{K} + \Upsilon_A l I + \frac{\Upsilon_I l}{n^2}\boldsymbol{L}\boldsymbol{K}\right)^{-1} Y \tag{21}$$

where I is $n * n$ diagonal matrix; J is an $n * n$ diagonal matrix with the labeled points diagonal entries as 1 and the rest 0; Y is an n dimensional label vector given by $Y = \text{diag}(y_1, \ldots, y_l, 0, \cdots, 0)$.

Given fixed α, we obtain the solution of μ_k with s.t. $\sum_{k=1}^{m}\mu_k = 1$:

$$\mu_k = \frac{\left(\frac{n^2}{\Upsilon_I \alpha^T \boldsymbol{K} L_k^p \boldsymbol{K}\alpha}\right)^{\frac{1}{\gamma-1}}}{\sum_{k=1}^{m}\left(\frac{n^2}{\Upsilon_I \alpha^T \boldsymbol{K} L_k^p \boldsymbol{K}\alpha}\right)^{\frac{1}{\gamma-1}}} \tag{22}$$

The iterative solution procedure of EpLapR is described in Table 1.

Now, we prove the convergence of this iterative solution. Intuitively, the update criteria in Eq. (22) tend to assign larger value to μ_k with smaller $\frac{\Upsilon_I}{n^2}\alpha^T K L_k^p K\alpha$. Denoted by α^t and μ^t, the values of α and μ in t^{th} iteration of the process, we have $f(\alpha^t, \mu^{t+1}) \leq f(\alpha^t, \mu^t)$, and according to Eq. (21), the solution of α is updated to minimize f; thus, we can get $f(\alpha^{t+1}, \mu^{t+1}) \leq f(\alpha^t, \mu^{t+1})$. Therefore, $f(\alpha^{t+1}, \mu^{t+1}) \leq f(\alpha^t, \mu^{t+1}) \leq f(\alpha^t, \mu^t)$.

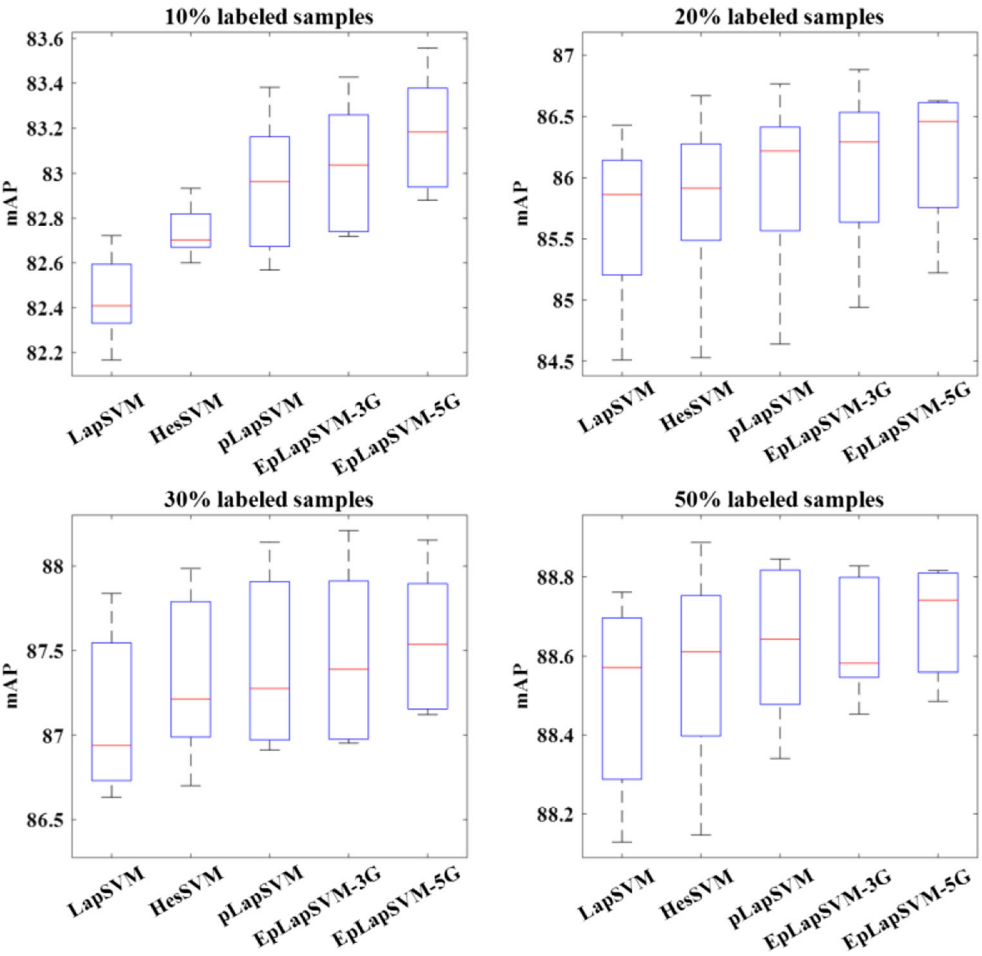

Fig. 11. mAP performance of different algorithms on SVM method of Scene 15 dataset

EpLapR Support Vector Machines (EpLapSVM)

The EpLapSVM solves the optimization problem (17) with the hinge loss function as

$$f(*) = \arg\min_{f\epsilon H_K} \frac{1}{1}\sum_{i=1}^{l}(1-y_i f(x_i))_+ + \Upsilon_A \alpha^T \boldsymbol{K}\alpha + \frac{\Upsilon_I}{n^2}\alpha^T \boldsymbol{K}\left(\sum_{k=1}^{m}\mu_k^{\gamma} L_k^p\right)\boldsymbol{K}\alpha \quad (23)$$

The partial derivative of the objective function with respect to μ_k is same as Eq. (19).

Introducing Lagrange multiplier method with β_i and η_i and add an unregularized bias term b, we arrive at a convex differentiable objective function:

$$L(\alpha,\xi,b,\beta,\eta) = \frac{1}{l}\sum_{i=1}^{l}\xi_i + \frac{1}{2}\alpha^T\left(2\Upsilon_A\boldsymbol{K} + 2\frac{\Upsilon_A}{n^2}\boldsymbol{K}\left(\sum_{k=1}^{m}\mu_k^{\gamma}L_k^p\right)\boldsymbol{K}\right)\alpha - \sum_{i=1}^{l}\beta_i\left(y_i\left(\sum_{j=1}^{l+u}\alpha_j\boldsymbol{K}(x_i,x_j)+b\right)-1+\xi_i\right) - \sum_{i=1}^{l}\eta_i\xi_i$$

We reduce the Lagrangian using $\frac{\partial L}{\partial b} = 0$ and $\frac{\partial L}{\partial \xi_i} = 0$ and take partial derivative with respect to α

$$\alpha^* = \left(2\Upsilon_A I + 2\frac{\Upsilon_I}{n^2}\boldsymbol{L}\boldsymbol{K}\right)^{-1} J^T Y\beta^* \quad (24)$$

where I is $l * l$ diagonal matrix; J is an $l * n$ diagonal matrix given by $J = [I, 0]$; Y is an l dimensional label vector given by: $Y = \text{diag}(y_1, \ldots, y_l)$. β^* is the n-dimensional variable given by

$$\beta^* = \max_{\beta\in\mathbb{R}^l}\sum_{i=1}^{l}\beta_i - \frac{1}{2}\beta^T Q\beta \text{ subject to : } \sum_{i=1}^{l}\beta_i y_i = 00 \le \beta_i \le \frac{1}{l}, i = 1,\cdots,l$$

where

$$Q = YJK\left(2\Upsilon_A I + 2\frac{\Upsilon_I}{n^2}\boldsymbol{L}\boldsymbol{K}\right)^{-1} J^T Y$$

Then, the problem (23) can also be solved by the iterative solution process in Table 1.

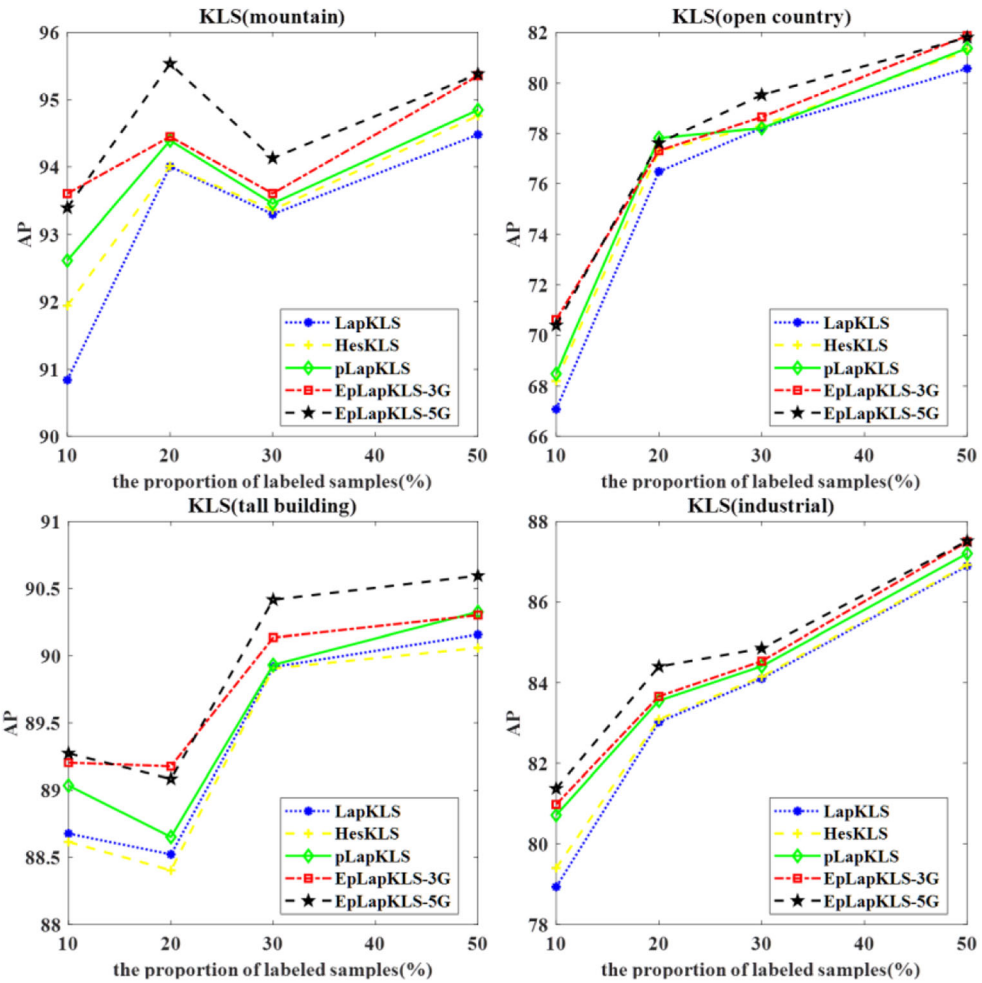

Fig. 12 AP performance of different KLS methods on several classes of Scene 15 dataset

Experiments

In this section, to evaluate the effectiveness of the proposed EpLapR, we compare EpLapR with other local structure preserving algorithms including LapR, HesR, and pLapR on UC-Merced database and Scene 15 dataset. We apply the SVM and KLS for image classification. Figure 1 illustrates the framework of EpLapR for UC-Merced dataset.

UC-Merced dataset [27] consists of 2100 remote sensing images collected from aerial orthoimage with the pixel resolution of 1 ft. The original images were downloaded from the United States Geological Survey National Map of different U.S. regions. There are totally 21 classes, including chaparral, dense residential, medium residential, sparse residential, forest, freeway, agricultural, airplane, baseball diamond, mobile home park, overpass, parking lot, river, runway, beach, buildings, golf course, harbor, intersection, storage tanks, and tennis courts (see in Fig. 2). It is worth noticing that this dataset has some highly overlapped classes, e.g., sparse residential, medium-density residential, and dense residential; so, it is difficult to get a successful classification.

Scene 15 dataset is composed of 15 scene categories, totally 4485 images. Each class has 200 to 400 images. The images contain indoor scenes and outdoor scenes, such as living room, kitchen, forest, mountain, and tall building (see in Fig. 3).

In our experiments, we extract high-level visual features using the deep convolution neural network (CNN) [32] for UC-Merced dataset and extract SIFT feature for Scene 15 dataset. For UC-Merced dataset, we randomly choose 50 images per class as training samples and the rest as testing samples. For Scene 15 dataset, 100 images per class are randomly selected as the training data, and the rest for testing. In semi-supervised classification experiments, in particular, we select 10, 20, 30, and 50% samples of training data as labeled data, and the rest as unlabeled data. To avoid any bias introduced by the random partition of samples, the process is repeated five times independently.

We conduct the experiments on the dataset to choose suitable model parameters. The regularization parameters γ_A and γ_I are selected from the candidate set $\{10^i | i = -10, -9, -8, \cdots, 10\}$ through cross-validation. For pLapR, the parameter p is chosen from $\{1, 1.1, 1.2, \cdots, 3\}$ through cross-validation with 10% labeled samples on the training data. Figures 4

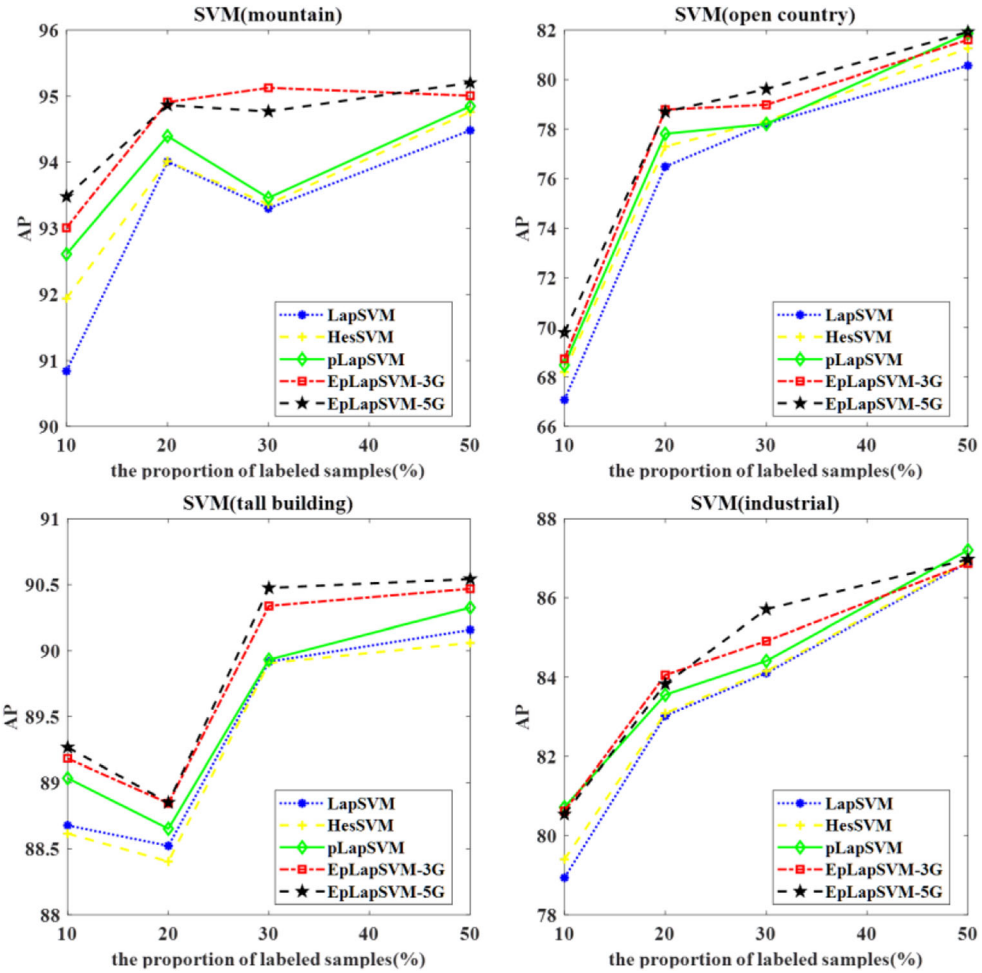

Fig. 13 AP performance of different SVM methods on several classes of Scene 15 dataset

and 5 illustrate the mAP performance of pLapR on the validation set when p varies. The x axis is the parameter p, and the y axis is mAP for performance measure. From Fig. 4, we can see that, for UC-Merced data set, the best mAP performance for pLapR can be obtained when $p = 2.8$. Figure 5 is the performance of the Scene 15 database, and the best performance is achieved when p is equal to 1. For EpLapR, we created two graph p-Laplacian sets on UC-Merced dataset. For the first set (EpLapR-3G), we choose $p = \{2.5, 2.7, 2.8\}$, which led to three graphs. For another one (EpLapR-5G), with five graphs where $p = \{2.4, 2.5, 2.6, 2.7, 2.8\}$. For Scene 15 dataset, the parameters p of EpLapR-3G is $\{1, 1.4, 2.1\}$, and the p of EpLapR-5G is $\{1, 1.1, 1.4, 2, 2.1\}$. We verify classification performance by average precision (AP) performance for single class and mean average precision (mAP) [33] for overall classes. The AP is defined as the mean precision at a set of 11 equally spaced recall levels and can be expressed as follows:

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

where pre(r) is the precision at recall r. The mAP is the mean AP over all image classes and can be written as:

$$\mathrm{mAP} = \frac{\sum_{i=1}^{c} \mathrm{AP_i}}{c}$$

where c is the number of image classes.

We compare our proposed EpLapR with the representative LapR, HesR and pLapR. Figures 6 and 7 demonstrate the mAP results of different algorithms on KLS methods and SVM methods on UC-Merced data set, respectively. Figures 8 and 9 show the mAP performance on Scene 15 dataset. 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.

To evaluate the effectiveness of EpLapR for single class, Figs. 10 and 11 show the AP results of different methods on several selected remote sensing classes, including medium residential, parking lot, sparse residential, and tennis court of UC-Merced dataset. Figure 8 reveals the KLS method, while Fig. 9 represents the SVM method. Figures 12 and 13 show the AP results of Scene 15 dataset about several classes, including mountain, open country, tall building, and industrial. We can find that EpLapR performs better than LapR, HesR, and pLapR by sufficiently explore the complementation of

graph p-Laplacian. Moreover, what we can note is that the classification results are not always improved with a larger proportion of labels or a fused graph of much more graph p-Laplacian. These observations suggest that it is critical to select parameters for our proposed method.

Conclusion

As a nonlinear generalization of graph Laplacian, the p-Laplacian regularization precisely exploits the geometry of the probability distribution to leverage the learning performance. However, in practical, it is difficult to determine the optimal graph p-Lapalcian because the parameter p usually is chosen 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. EpLapR incorporates multiple graphs into a fused graph by an optimization approach to assign suitable weights on different p value graphs. And then, we introduce the optimal fused graph as a regularizer for SSL. Finally, we construct the ensemble graph p-Laplacian regularized classifiers, including EpLapKLS and EpLapSVM for scene image recognition. Experimental results on the UC-Merced dataset show that our proposed EpLapR learner can generalize well than traditional ones.

Funding Information 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, the Fundamental Research Funds for the Central Universities, China University of Petroleum (East China) under Grant 18CX07011A and Grant YCX2017059, the National Natural Science Foundation of China under Grant 61772455 and Grant U1713213, the Yunnan Natural Science Funds under Grant 2016FB105, the Program for Excellent Young Talents of Yunnan University under Grant WX069051, the Macau Science and Technology Development Fund under Grant FDCT/189/2017/A3, and the Research Committee at University of Macau under Grant MYRG2016-00123-FST and Grant MYRG2018-00136-FST.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Subramanya, Bilmes J. Soft-supervised learning for text classification. EMNLP. 2008:1090–9.
2. Scardapane S, Uncini A. Semi-supervised echo state networks for audio classification. Cogn Comput. 2017;9(1):125–35.
3. Zhou D, Bousquet O, Lal TN, et al. Learning with local and global consistency. Adv Neural Inf Proces Syst. 2004:321–8.
4. Zhao M, Zhang Z, Chow TWS. Trace ratio criterion based generalized discriminative learning for semi-supervised dimensionality reduction. Pattern Recogn. 2012;45(4):1482–99.
5. Oneto L, Bisio F, Cambria E, Anguita D. Semi-supervised learning for affective common-sense reasoning. Cogn Comput. 2017;9(1):18–42.
6. Ding S, Xi X, Liu Z, Qiao H, Zhang B. A novel manifold regularized online semi-supervised learning model. Cogn Comput. 2018;10(1):49–61.
7. Zhao M, Chow TWS, Zhang Z, Li B. Automatic image annotation via compact graph based semi-supervised learning. Knowl-Based Syst. 2015;76:148–65.
8. Khan FH, Qamar U, Bashir S. Multi-objective model selection (MOMS)-based semi-supervised framework for sentiment analysis. Cogn Comput. 2016;8(4):614–28.
9. Belkin M, Niyogi P, Sindhwani V. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. J Mach Learn Res. 2006;7(Nov):2399–434.
10. Liu W, Tao D. Multiview Hessian regularization for image annotation. IEEE Trans Image Process. 2013;22(7):2676–87.
11. Liu X, Shi J, Wang C. Hessian regularization based non-negative matrix factorization for gene expression data clustering. Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, IEEE, 2015: 4130–4133.
12. Zhu J, Shi J. Hessian regularization based semi-supervised dimensionality reduction for neuroimaging data of Alzheimer's disease. Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium, IEEE, 2014: 165–168.
13. Bühler T, Hein M. Spectral clustering based on the graph p-Laplacian. In: Proceedings of the 26th Annual International Conference on Machine Learning: ACM; 2009. p. 81–8.
14. Kim KI, Steinke F, Hein M. Semi-supervised regression using Hessian energy with an application to semi-supervised dimensionality reduction. Adv Neural Inf Proces Syst. 2009:979–87.
15. Takeuchi H. The spectrum of the p-Laplacian and p-harmonic morphisms on graphs. Ill J Math. 2003;47(3):939–55.
16. Allegretto W, Xi HY. A Picone's identity for the p-Laplacian and applications. Nonlinear Anal: Theory, Methods Appl. 1998;32(7):819–30.
17. Geng B, Tao D, Xu C, et al. Ensemble manifold regularization. IEEE Trans Pattern Anal Mach Intell. 2012;34(6):1227–33.
18. Wang M, Hua XS, Hong R, et al. Unified video annotation via multigraph learning. IEEE Trans Circuits Syst Video Technol. 2009;19(5):733–46.
19. Hong C, Yu J, Tao D, Wang M. Image-based three-dimensional human pose recovery by multiview locality-sensitive sparse retrieval. IEEE Trans Ind Electron. 2015;62(6):3742–51.
20. Zhang J, Han Y, Tang J, Hu Q, Jiang J. Semi-supervised image-to-video adaptation for video action recognition. IEEE Trans Cybern. 2017;47(4):960–73.
21. Bian X, Zhang T, Zhang X, Yan LX, Li B. Clustering-based extraction of near border data samples for remote sensing image classification. Cogn Comput. 2013;5(1):19–31.
22. Hu W, Cheung G, Li X, et al. Graph-based joint denoising and super-resolution of generalized piecewise smooth images. Image Processing (ICIP), 2014 IEEE International Conference. IEEE, 2014: 2056–2060.
23. Luo Y, Tao D, Geng B, Xu C, Maybank SJ. Manifold regularized multitask learning for semi-supervised multilabel image classification. IEEE Trans Image Process. 2013;22(2):523–36.
24. Jiang J, Hu R, Wang Z, Cai Z. CDMMA: coupled discriminant multi-manifold analysis for matching low-resolution face images. Signal Process. 2016;124:162–72.
25. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. IJCAI. 1995;14(2):1137–45.
26. Luo D, Huang H, Ding C, Nie F. On the eigenvectors of p-Laplacian. Mach Learn. 2010;81(1):37–51.

27. Yang Y, Newsam S. Bag-of-visual-words and spatial extensions for land-use classification. Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. ACM, 2010: 270–279.
28. Zhou D, Schölkopf B. Regularization on discrete spaces. Joint Pattern Recognition Symposium. Berlin, Heidelberg: Springer; 2005. p. 361–8.
29. Liu W, Zha ZJ, Wang Y, et al. *p*-Laplacian regularized sparse coding for human activity recognition. IEEE Trans Ind Electron. 2016;63(8):5120–9.
30. Liu W, Ma X, Zhou Y, Tao D, and Cheng J.. *p*-Laplacian regularization for scene recognition. IEEE Transactions on Cybernetics, **to be published**, doi: https://doi.org/10.1109/TCYB.2018.2833843, 2018.
31. Ma X, Liu W, Li S, Tao D, and Zhou Y. Hypergraph *p*-Laplacian regularization for remotely sensed image recognition. IEEE Trans Geosci Remote Sens, **to be published**, doi: https://doi.org/10.1109/TGRS.2018.2867570, 2018.
32. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv. 2014:1409.1556.
33. Everingham M, Van Gool L, Williams CKI, et al. The Pascal visual object classes (VOC) challenge. Int J Comput Vis. 2010;88(2):303–38.
34. Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. New York, NY, USA: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06); 2006. p. 2169–78.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com