p-Laplacian Regularization for Scene Recognition

Weifeng Liu, *Senior Member, IEEE*, Xueqi Ma, Yicong Zhou, *Senior Member, IEEE*, Dapeng Tao, and Jun Cheng

***Abstract*—The explosive growth of multimedia data on the Internet makes it essential to develop innovative machine learning algorithms for practical applications especially where only a small number of labeled samples are available. Manifold regularized semi-supervised learning (MRSSL) thus received intensive attention recently because it successfully exploits the local structure of data distribution including both labeled and unlabeled samples to leverage the generalization ability of a learning model. Although there are many representative works in MRSSL, including Laplacian regularization (LapR) and Hessian regularization, how to explore and exploit the local geometry of data manifold is still a challenging problem. In this paper, we introduce a fully efficient approximation algorithm of graph *p*-Laplacian, which significantly saving the computing cost. And then we propose *p*-LapR (pLapR) to preserve the local geometry. Specifically, *p*-Laplacian is a natural generalization of the standard graph Laplacian and provides convincing theoretical evidence to better preserve the local structure. We apply pLapR to support vector machines and kernel least squares and conduct the implementations for scene recognition. Extensive experiments on the Scene 67 dataset, Scene 15 dataset, and UC-Merced dataset validate the effectiveness of pLapR in comparison to the conventional manifold regularization methods.**

***Index Terms*—Laplacian regularization (LapR), manifold learning, *p*-Laplacian, scene recognition, semi-supervised learning (SSL).**

Manuscript received September 11, 2017; revised February 25, 2018 and April 25, 2018; accepted May 3, 2018. Date of publication May 22, 2018; date of current version May 7, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61671480, in part by the Foundation of Shandong province under Grant ZR2018MF017, in part by the Fundamental Research Funds for the Central Universities, China University of Petroleum (East China) under Grant 18CX07011A and Grant YCX2017059, in part by the Yunnan Natural Science Funds under Grant 2016FB105, in part by the Program for Excellent Young Talents of Yunnan University under Grant WX069051, in part by the Macau Science and Technology Development Fund under Grant FDCT/189/2017/A3, and in part by the Research Committee at University of Macau under Grant MYRG2016-00123-FST and Grant MYRG2018-00136-FST. This paper was recommended by Associate Editor Y. Yuan. *(Corresponding authors: Weifeng Liu; Yicong Zhou.)*

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

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

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

J. Cheng is with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, and also with the Department of Mechanical and Automation Engineering, Chinese University of Hong Kong, Hong Kong (e-mail: jun.cheng@siat.ac.cn).

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

Digital Object Identifier 10.1109/TCYB.2018.2833843

I. Introduction

With the prodigious development of the mobile device, social media network, and Internet technologies, the multimedia data on the Internet is explosively growing everyday. Usually, there are only a small number of data online is annotated in practical. Hence, it is essential to build innovative learning models to effectively utilize the increasing data where most of the data are unlabeled for practical applications, such as object or scene recognition [1], [2], Web image annotation and classification [3], [4], human behavior analysis for abnormal detection [5], [6], large scale video retrieval [7], intelligent surveillance [8], and so on. One of the successful solutions is manifold regularized semi-supervised learning (MRSSL) that exploits the local structure of data distribution including both labeled and unlabeled samples to leverage the generalization ability of a learning model [9].

In MRSSL, it assumes that the conditional probability distribution of class labels varies smoothly along the geodesics in the intrinsic geometry of the marginal distribution according to which examples are drawn [10]. In other words, two samples with closer representation generally have the same class labels. Manifold regularization tries to explore the geometry of the probability distribution that generates the data and incorporates it as a regularization term to penalizing the regression function along the potential manifold. There are many successful MRSSL algorithms have been reported recently which tries to explore the geometry of the intrinsic data probability distribution to leverage the learning performance. The prominent MRSSL algorithms are Laplacian regularization (LapR) [10] and Hessian regularization (HesR) [11], [12].

LapR determines the underlying manifold by using the graph Laplacian [10], [13]. With the merits of simple calculation and promising performance, LapR-based semi-supervised learning (SSL) has received extensive attention and many algorithms have been developed, including Laplacian regularized support vector machines (LapSVMs), Laplacian regularized kernel least squares (LapKLSs) [10], [14], and Laplacian regularized nonnegative matrix factorization (NMF) [15]. However, the null space of the graph Laplacian along the underlying manifold is a constant function, which results in that the solution beyond the domain of training samples is always a constant function. Thus, the learning performance of LapR-based models may sharply decrease especially when there are only few labeled examples.

In comparison to LapR, HesR steers the learned function varying linearly along the underlying manifold benefitting from the advantages of Hessian that has a richer nullspace than Laplacian [11]. In result, HesR can extrapolates linearly even

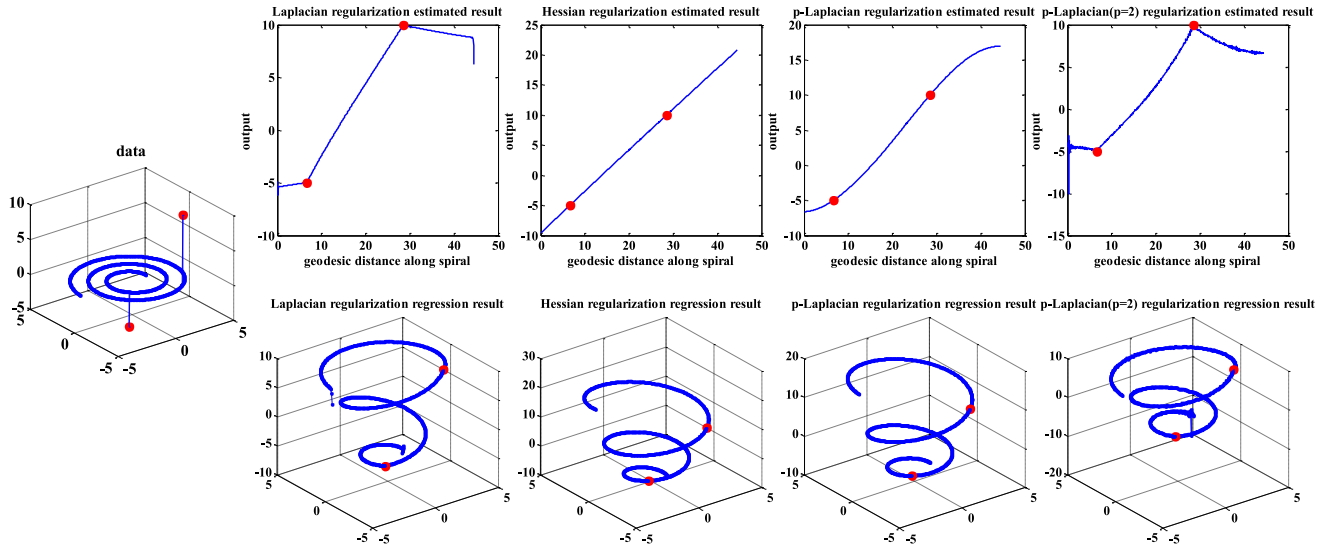


Fig. 1. Differences of semi-supervised regression by using LapR, HesR, and pLapR for fitting two points on the 1-D spiral.The LapR has always a bias toward the constant function which cannot fit the data properly. The HesR fits the data better than LapR and the extrapolation functions vary linearly along spiral. The pLapR can fit the data exactly and extrapolates smoothly to unseen data with the geodesic distance. In addition, with $p = 2$, the pLapR get the similar results with the standard LapR.

when the solution is beyond the training data domain. Thus, HesR-based SSL algorithms can achieve better performance than LapR-based ones [12]. 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 [11]. On the other hand, for real-world datasets the sampling is usually not dense enough to accurately fit the local structure. Therefore, how to explore and exploit the local geometry of data distribution is still a challenging problem.

To tackle the above-mentioned problems, we propose p-LapR (pLapR) in this paper. The p-Laplacian [16]–[18] is nonlinear generalization of the graph Laplacian and provides convincing theoretical evidence to better preserve the local structure. In particular, p-Laplacian has tighter isoperimetric inequality and the second eigenvalue of p-Laplacian approximates the optimal Cheeger cut well [19], which indicates the superiority of pLapR in local geometry exploiting. Fig. 1 illustrates the differences of semi-supervised regression by using LapR, HesR, and pLapR for fitting two points on the 1-D spiral.

On the other hand, although there are many efforts on the study of p-Laplacian for learning tasks, the efficient approximation of graph p-Laplacian is still unsolved, which extremely limits the applications of pLapR. We provide an effect and efficient fully approximation algorithm of p-Laplacian in this paper, which significantly improves computation efficiency.

Finally, we introduce pLapR to support vector machines (SVMs) and kernel least squares (KLSs) for scene recognition. We conduct experiments on the Scene 67 dataset [20], Scene 15 dataset [21], and UC-Merced dataset [22] by comparing with the conventional manifold regularization algorithms including LapR and HesR.

In summary, the contributions of this paper are as follows.

1) We develop an efficient approximation algorithm of graph p-Laplacian, significantly slowing down its computation cost.
2) We propose pLapR to preserve the local similarity of data.
3) Based on pLapR, we further propose the pLapSVMs and pLapKLSs.
4) To investigate the application of pLapR and evaluate its performance, we propose a pLapR-based framework for scene recognition, and test pLapSVM and pLapKLS on three datasets.

The rest of this paper is organized as follows. Section II briefly reviews related work on MRSSL and p-Laplacian. Section III describes the approximation of graph p-Laplacian. Section IV presents the proposed pLapR framework and Section V details the implementation of pLapR for SVMs and KLSs. Section VI demonstrates the experimental results. And finally Section VII is the conclusion.

II. Related Work

Scene recognition [23]–[25] has been a hot technology because of its significance in real life. Many classification methods have been proposed to get a well performance in scene recognition. Cakir *et al.* [26] categorized the scenes by employing a metric function based on the nearest-neighbor classification procedure using the bag-of-visual words scheme. However, it is not helpful for the case of small labeled-sample.

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

A. Manifold Regularized Semi-Supervised Learning

MRSSL assumes that the data distribution is supported on a submanifold of the ambient space and thus aims to find

proper graph to encode the local geometry between examples. Many prominent MRSSL algorithms have been reported these years.

Belkin *et al.* [10] presented the manifold regularization framework in which the objective function includes two terms: 1) a loss function and 2) a regularizer. This framework allows manifold regularization to be integrated into many classifiers. Luo *et al.* [27] proposed a manifold regularized multitask learning algorithm to find a discriminative subspace shared by multiple classification tasks. They employed manifold regularization to steer the functions in the shared hypothesis space varying smoothly along the data manifold. Guan *et al.* [15] and Cai *et al.* [28] applied manifold regularization to NMF and simultaneously learned NFM subspace and preserved the data geometric structure. Yu *et al.* [29] modeled the high-order relationship of samples by using hypergraph and proposed an adaptive hypergraph learning method for transductive image classification. Geng *et al.* [14] developed an ensemble manifold regularization framework to approximate the intrinsic manifold. Wang *et al.* [30] introduced multimodal graph-based learning scheme for Web image search. And then Liu and Tao [12] proposed multiview HesR by exploiting the local structure of multiple features with multiview Hessian.

B. p-Laplacian

Suppose we are given a weighted, undirected graph $G = (V, E)$, where V is the vertex set and E is the edge set of the graph. Let $\mathcal{H}(V)$ and $\mathcal{H}(E)$ denote the Hilbert space of real-valued functions on each vertex and edge, respectively.

The graph gradient $\nabla : \mathcal{H}(V) \to \mathcal{H}(E)$ is defined as

$$\nabla f_{uv} \coloneqq \sqrt{\frac{w_{uv}}{d_v}} f_v - \sqrt{\frac{w_{uv}}{d_u}} f_u \tag{1}$$

where f is a function on V, $uv \in E$, w_{uv} is the edge weight on uv and d_v is the degree of vertex v.

And the graph divergence $\mathrm{div} : \mathcal{H}(E) \to \mathcal{H}(V)$ for $\psi \in \mathcal{H}(E)$ on vertex v is defined as

$$\mathrm{div}(\psi)_v = \sum_{u \sim v} \sqrt{\frac{w_{uv}}{d_v}} (\psi_{vu} - \psi_{uv}). \tag{2}$$

Then we have the graph p-Laplacian $\Delta_p : \mathcal{H}(V) \to \mathcal{H}(V)$ as

$$\Delta_p f \coloneqq -\frac{1}{2} \mathrm{div}\Big(\|\nabla f\|^{p-2} \nabla f \Big). \tag{3}$$

Clearly, when $p = 2$, $\Delta_p f = \Delta_2 f = -(1/2)\mathrm{div}(\nabla f)$, i.e., the graph p-Laplacian becomes the standard graph Laplacian.

On the other hand, it is well known that the standard graph Laplacian Δ_2 can be defined as the operator by introducing the quadratic form for a function f

$$\langle f, \Delta_2 f \rangle = \frac{1}{2} \sum_{u,v \in V} w_{uv} (f_u - f_v)^2. \tag{4}$$

Similar to the graph Laplacian, the unnormalized p-Laplacian operator can be also defined by the choice of the inner product [18]

$$\left(\Delta_p^w \right)_v = \sum_{u \in V} w_{uv} \phi_p (f_v - f_u), v \in V \tag{5}$$

where ϕ_p is defined by $\phi_p(x) = |x|^{p-1}\mathrm{sig}(x)$. Also when $p = 2$, $\phi_2(x) = x$, $\Delta_p^w f$ becomes the standard graph Laplacian.

The p-Laplacian attracts many attentions due to its advantages in graph representation.

Amghibech [31] discussed the bounds for the largest p-Laplacian eigenvalue for graphs including lower bound and upper bound and showed that the p-Laplacian eigenvalue has the property as

$$2^{p-1} \min \{d_v, v \in V\} \le \lambda \le 2^{p-1} \max \{d_v, v \in V\}. \tag{6}$$

Bühler and Hein [19] then proved the second eigenvector of the p-Laplacian approximates the optimal Cheeger cut arbitrarily well as

$$RCC \le RCC^* \le p \left(\max_{i \in V} d_i \right)^{\frac{p-1}{p}} RCC^{\frac{1}{p}} \tag{7}$$

or

$$NCC \le NCC^* \le p NCC^{\frac{1}{p}} \tag{8}$$

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

Zhou and Schölkopf [32] presented a discrete regularization framework of p-Laplacian as

$$f^* = \mathrm{argmin}_{f \in \mathcal{H}(V)} \left\{ \mathcal{S}_p(f) + \mu \|f - y\|^2 \right\} \tag{9}$$

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

Liu *et al.* [6] developed p-Laplacian regularized sparse coding for human activity recognition.

Luo *et al.* [33] proposed an efficient way to obtain the full eigenvectors of p-Laplacian for multiclass clustering problems.

Slepčev and Thorpe [34] provided some analysis on pLapR for SSL.

III. Approximation of Graph p-Laplacian

In this section, we briefly describe the approximating of graph p-Laplacian. And then we integrate the approximated graph p-Laplacian into manifold regularization framework and develop pLapR in the next section.

First, we define the eigenvector and eigenvalue of p-Laplacian Δ_p^w.

Definition 1: The function $f : V \to R$ is a p-eigenfunction or an eigenvector of p-Laplacian Δ_p^w, if there exists a real number λ_p and such that

$$\left(\Delta_p^w f \right)_i = \lambda_p \phi_p(f_i), i \in V \tag{10}$$

where $\phi_p(x) = |x|^{p-1}\mathrm{sign}(x)$, and λ_p is called as an eigenvalue of Δ_p^w associated with eigenvector f.

By using the Rayleigh quotient formula, one can compute the eigenvalue and the corresponding eigenvector for a symmetric matrix $A \in R^{n \times n}$. Carrying out the variational characterization of Rayleigh–Ritz principle on p-Laplacian, we have the following theorem.

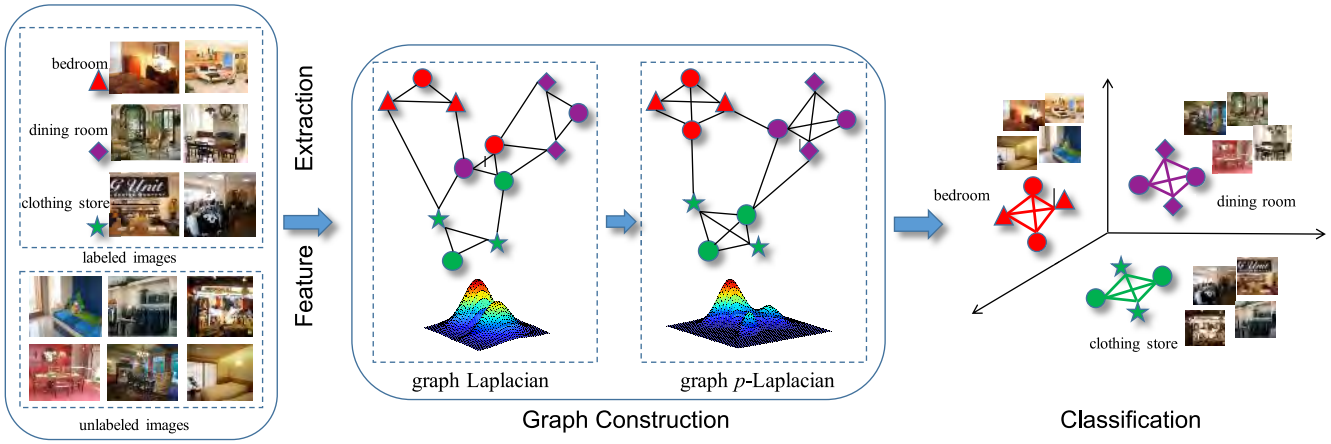


Fig. 2. Framework of pLapR for indoor scene recognition.

Theorem 1 [19]: f is an eigenvector of p-Laplacian, if and only if the following function F_p has a critical point at f:

$$F_p(f) = \frac{\sum_{ij} w_{ij}|f_i - f_j|^p}{2\|f\|_p^p} \tag{11}$$

where $\|f\|_p^p = \sum_i |f_i|^p$, $i,j \in V$ and the corresponding eigenvalue λ_p is given by $\lambda_p = F_p(f)$.

Theorem 1 provides the foundational analysis of eigenvectors and eigenvalues for p-Laplacian. And moreover it indicates that $F_p(\alpha f) = F_p(f)$ for all real value $\alpha \neq 0$.

Corollary 1: If we want to get all eigenvectors and eigenvalues of p-Laplacian, we have to find all critical points of the function F_p.

The full eigenvectors of p-Laplacian can be obtained by exploring the relationship of the full eigenvectors space as the following Theorem 2.

Theorem 2 [33]: If $f^{*1}, f^{*2}, \ldots, f^{*K}$ are K eigenvectors of p-Laplacian Δ_p^w associated with unique eigenvalues $\lambda_1^*, \lambda_2^*, \ldots, \lambda_K^*$, then $f^{*1}, f^{*2}, \ldots, f^{*K}$ are local solution of the following optimization problem:

$$\min_{\mathcal{F}} J(\mathcal{F}) = \sum_k F_p\left(f^k\right)$$
$$\text{s.t.} \sum_i \phi_p\left(f_i^k\right)\phi_p\left(f_i^l\right) = 0, \forall k \neq l \tag{12}$$

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

Theorem 2 provides an optimization solution for full eigenvectors of p-Laplacian. However, the optimization problem (12) remains unsolvable. Luo *et al.* [33] introduced an approximation for full eigenvectors of p-Laplacian by solving the following p-Laplacian embedding problem:

$$\min_{\mathcal{F}} J_E(\mathcal{F}) = \sum_k \frac{\sum_{ij} w_{ij}\left|f_i^k - f_j^k\right|^p}{\left\|f^k\right\|_p^p}$$
$$\text{s.t. } \mathcal{F}^T\mathcal{F} = I. \tag{13}$$

Solving the problem (13) with the gradient descend optimization, we can then obtain the full eigenvalues $\Lambda^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_K^*)$ of p-Laplacian associated with the eigenvectors $\mathcal{F}^* = (f^{*1}, f^{*2}, \ldots, f^{*K})$ by $\lambda_p = F_p(f)$. Finally, we approximate the graph p-Laplacian by $L_p = \mathcal{F}^*\Lambda^*\mathcal{F}^{*T}$ in this paper. We briefly describe the approximation of the graph p-Laplacian in Algorithm 1. Note the step length $\eta = 0.01(\sum_{ik}|\mathcal{F}_{ik}|/\sum_{ik}|\mathcal{G}_{ik}|)$ in this paper.

Algorithm 1 Approximation of Graph p-Laplacian

Input: Training sample set X, p
Output: graph p-Laplacian L_p
Step 1: Construct adjacency matrix W, standard graph Laplacian $L = D - W$, where D is diagonal matrix with $D_{ii} = \sum_{j=1}^n W_{ij}$.
Step 2: Compute eigenvector decomposition of graph Laplacian $L = USU^T$.
Step 3: Initialize $\mathcal{F} = U(:, 1{:}K)$
repeat
$G = \frac{\partial J_E}{\partial \mathcal{F}} - \mathcal{F}\left(\frac{\partial J_E}{\partial \mathcal{F}}\right)^T\mathcal{F}$,
where $\frac{\partial J_E}{\partial f_i^k} = \frac{1}{\|f^k\|_p^p}\left[\sum_j w_{ij}\phi_p\left(f_i^k - f_j^k\right) - \frac{\phi_p(f_i^k)}{\|f^k\|_p^p}\right]$.
$\mathcal{F} = \mathcal{F} - \eta G$
until convergence
Step 4: $\lambda_k = \frac{\sum_{ij} w_{ij}|f_i^k - f_j^k|^p}{\|f^k\|_p^p}$
return $L_p = \mathcal{F}\Lambda\mathcal{F}^T$

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 $\mathcal{F}^t + 1 = \mathcal{F}^t - \eta G$, thus $(\mathcal{F}^{t+1})^T\mathcal{F}^{t+1} = (\mathcal{F}^t - \eta G)^T(\mathcal{F}^t - \eta G) = (\mathcal{F}^t)^T\mathcal{F}^t - \eta[G^T\mathcal{F}^t + (\mathcal{F}^t)^T G]$.

Here,

$$\begin{aligned} &G^T\mathcal{F}^t + \left(\mathcal{F}^t\right)^T G \\ &= \left[\frac{\partial J_E}{\partial \mathcal{F}} - \mathcal{F}\left(\frac{\partial J_E}{\partial \mathcal{F}}\right)^T\mathcal{F}\right]^T\mathcal{F}^t + \left(\mathcal{F}^t\right)^T\left[\frac{\partial J_E}{\partial \mathcal{F}} - \mathcal{F}\left(\frac{\partial J_E}{\partial \mathcal{F}}\right)^T\mathcal{F}\right] \\ &= \left(\frac{\partial J_E}{\partial \mathcal{F}}\right)^T\mathcal{F}^t - \left(\mathcal{F}^t\right)^T\frac{\partial J_E}{\partial \mathcal{F}} - \left(\frac{\partial J_E}{\partial \mathcal{F}}\right)^T\mathcal{F}^t + \left(\mathcal{F}^t\right)^T\frac{\partial J_E}{\partial \mathcal{F}} \\ &= 0. \end{aligned}$$

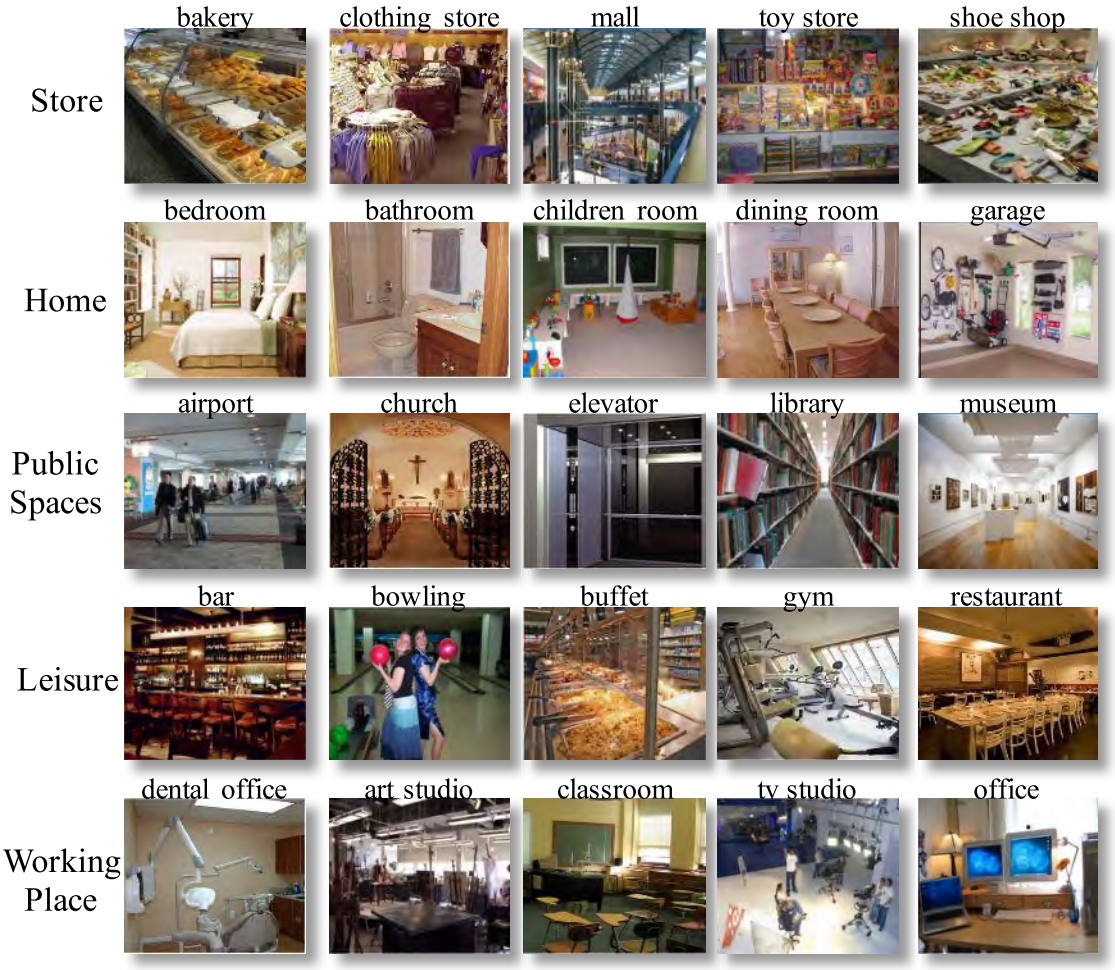


Fig. 3. Some example images of Scene 67 database. The dataset totally has 67 indoor scene categories that can be grouped into five big scene groups. Each row demonstrates one big scene group.

So $(\mathcal{F}^{t+1})^T\mathcal{F}^{t+1} = (\mathcal{F}^t)^T\mathcal{F}^t = I$. The solution $\mathcal{F}$ obtained from Algorithm 1 satisfies the constraint in problem (13) and it is the feasible solution.

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

IV. *p*-Laplacian Regularization

In MRSSL, we are given N training samples $X = \{X_L, X_U\}$ containing l labeled examples $X_L = \{(x_i, y_i)\}_{i=1}^{l}$ and u unlabeled samples $X_U = \{(x_j)\}_{j=l+1}^{l+u}$, where $N = l + u$, and $y_i \in \{+1, -1\}$ is the label of x_i. The labeled samples are pairs generated from probability distribution, while unlabeled samples are drawn according to the marginal distribution. The learning tasks of MRSSL aim to predict the labels of the unseen samples according to the manifold assumption that if two points x_1 and x_2 are close in the intrinsic geometry of marginal distribution, then the labels of x_1 and x_2 are tend to be similar. Considering the two regularization terms of classifier complexity in an appropriate reproducing kernel Hilbert spaces (RKHSs) and along the underlying manifold, the problem of MRSSL can be written as the following optimization problem:

$$\min_{f\in\mathcal{H}_K} \frac{1}{l}\sum_{i=1}^{l} V(x_i, y_i, f) + \gamma_A\|f\|_K^2 + \gamma_I\|f\|_I^2 \tag{14}$$

where V is a general loss function, $\|f\|_K^2$ controls the classifier complexity in an RKHS, $\|f\|_I^2$ steers the function varying smooth along the underlying manifold, and the parameters γ_A and γ_I balance the loss function and two regularization terms, respectively.

It transpires that precisely exploring the local geometry of the marginal distribution plays a critical role in MRSSL according to the underlying manifold assumption. In this paper, we introduce pLapR to MRSSL for indoor scene recognition. In comparison with the traditional choices for $\|f\|_I^2$ including LapR and HesR, pLapR can superiorly preserve the local geometry in theory with tighter isoperimetric inequality. Thus, the optimization problem of (14) can be expressed as

$$\min_{f\in\mathcal{H}_K} \frac{1}{l}\sum_{i=1}^{l} V(x_i, y_i, f) + \gamma_A\|f\|_K^2 + \gamma_I\mathbf{f}^T L_p\mathbf{f} \tag{15}$$

where $\mathbf{f} = [f(x_1), f(x_2), \ldots, f(x_{l+u})]^T$, L_p denotes the graph p-Laplacian. Although Zhou and Schölkopf [32] discussed the

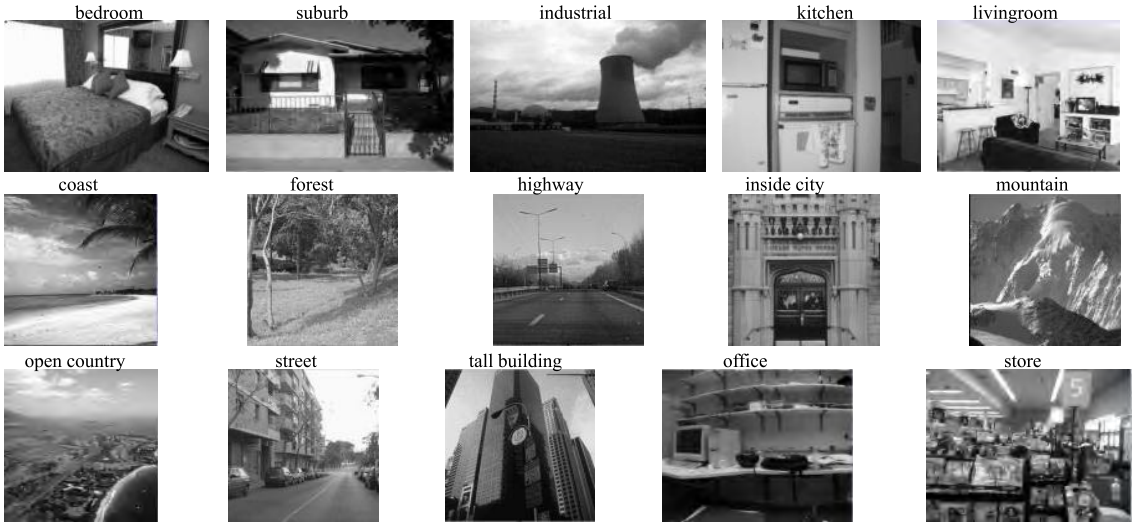


Fig. 4. Some example images of Scene 15 dataset. The dataset totally has 15 scene categories.

Fig. 5. Some example images of UC-Merced dataset. The dataset totally has 21 remote sensing categories.

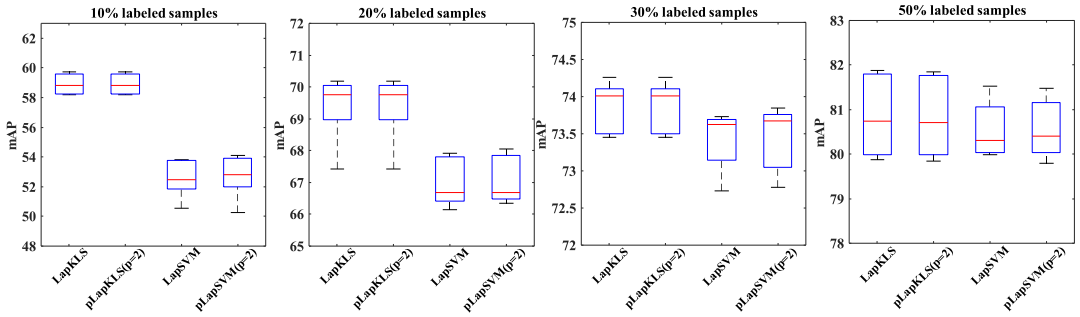


Fig. 6. mAP of pLapR($p = 2$) and LapR on Scene 67 dataset. Each subfigure reports the results under different labeled samples. In each subfigure, the y-axis is the mAP over all scene classes, and the x-axis is different classifiers.

solution for pLapR on a specific loss function, there has no generic and efficient computing method for graph p-Laplacian. We provide an approximation method of graph p-Laplacian in this paper. The proposed approximation facilitates the integrating of p-Laplacian to broader applications just as LapR and HesR.

To solve the optimization problem (15), we now discuss the representer theorem of pLapR. Proofs are detailed in the Appendix section.

Lemma 1: Suppose $\boldsymbol{K}$ is a valid (symmetric positive definite) kernel. If the penalty term of the optimization problem (14) is a strictly monotonically increasing real-valued function with respect to $\|f\|$, the minimizer of the problem admits an expansion

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

in terms of both labeled and unlabeled examples.

Lemma 2: L_p is semi-definite positive.

Theorem 3: The minimization of the object problem (15) with respect to f exits and has the representation

$$f^*(x) = \sum_{i=1}^{l+u} \alpha_i^* \boldsymbol{K}(x_i, x) \tag{17}$$

that is an expansion in terms of both labeled and unlabeled examples.

The representer theorem demonstrates the solution of (15) exists and has the general form of (17). Hence, learning the classifier f in the optimization problem (15) can be expressed as learning the $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_{l+u}]^T$ of the following problem:

$$\min_{\alpha \in R^{l+u}} \frac{1}{l} \sum_{i=1}^{l} V(x_i, y_i, f) + \gamma_A \alpha^T \boldsymbol{K} \alpha + \gamma_I \alpha^T \boldsymbol{K} L_p \boldsymbol{K} \alpha. \tag{18}$$

V. Example Algorithms

Generally, the proposed pLapR 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 pLapR to SVMs and KLSs.

A. pLapR Support Vector Machines

By selecting the hinge loss function, i.e., $V(x_i, y_i, f) = (1 - y_i f(x_i))_+ = \max(0, 1 - y_i f(x_i))$, we can obtain pLapSVM

$$\min_{f \in H_K} \frac{1}{l} \sum_{i=1}^{l} (1 - y_i f(x_i))_+ + \gamma_A \|f\|_K^2 + \gamma_I \mathbf{f}^T L_p \mathbf{f}. \tag{19}$$

According to the representer theorem, substitute (17) into (19), the pLapSVM can be expressed as

$$\min_{\alpha \in R^{l+u}} \frac{1}{l} \sum_{i=1}^{l} (1 - y_i \boldsymbol{K}(x_i, x)\alpha)_+ + \gamma_A \alpha^T \boldsymbol{K} \alpha + \gamma_I \alpha^T \boldsymbol{K} L_p \boldsymbol{K} \alpha. \tag{20}$$

Adding an unregularized bias term b, the primal problem of (20) leads to

$$\min_{\alpha \in R^{l+u}, \xi \in R^l} \frac{1}{l} \sum_{i=1}^{l} \xi_i + \gamma_A \alpha^T \boldsymbol{K} \alpha + \gamma_I \alpha^T \boldsymbol{K} L_p \boldsymbol{K} \alpha$$
$$\text{s.t. } y_i(\boldsymbol{K}(x_i, x)\alpha + b) \geq 1 - \xi_i, \xi \geq 0, i = 1, \ldots, l. \tag{21}$$

Introducing Lagrange multipliers β_i and η_i, we have

$$\mathcal{L}(\alpha, \xi, b, \beta, \eta) = \frac{1}{l} \sum_{i=1}^{l} \xi_i + \frac{1}{2} \alpha^T \big(2\gamma_A \boldsymbol{K} + 2\gamma_I \boldsymbol{K} L_p \boldsymbol{K}\big)\alpha - \sum_{i=1}^{l} \beta_i (y_i(\boldsymbol{K}(x_i, x)\alpha + b) - 1 + \xi_i) - \sum_{i=1}^{l} \eta_i \xi_i. \tag{22}$$

Reducing the Lagrangian by using $(\partial \mathcal{L}/\partial b) = 0$ and $(\partial \mathcal{L}/\partial \xi_i) = 0$, we get the following form:

$$\mathcal{L}^R(\alpha, \beta) = \frac{1}{2} \alpha^T \big(2\gamma_A \boldsymbol{K} + 2\gamma_I \boldsymbol{K} L_p \boldsymbol{K}\big)\alpha - \alpha^T \boldsymbol{K} J^T Y \beta + \sum_{i=1}^{l} \beta_i \tag{23}$$

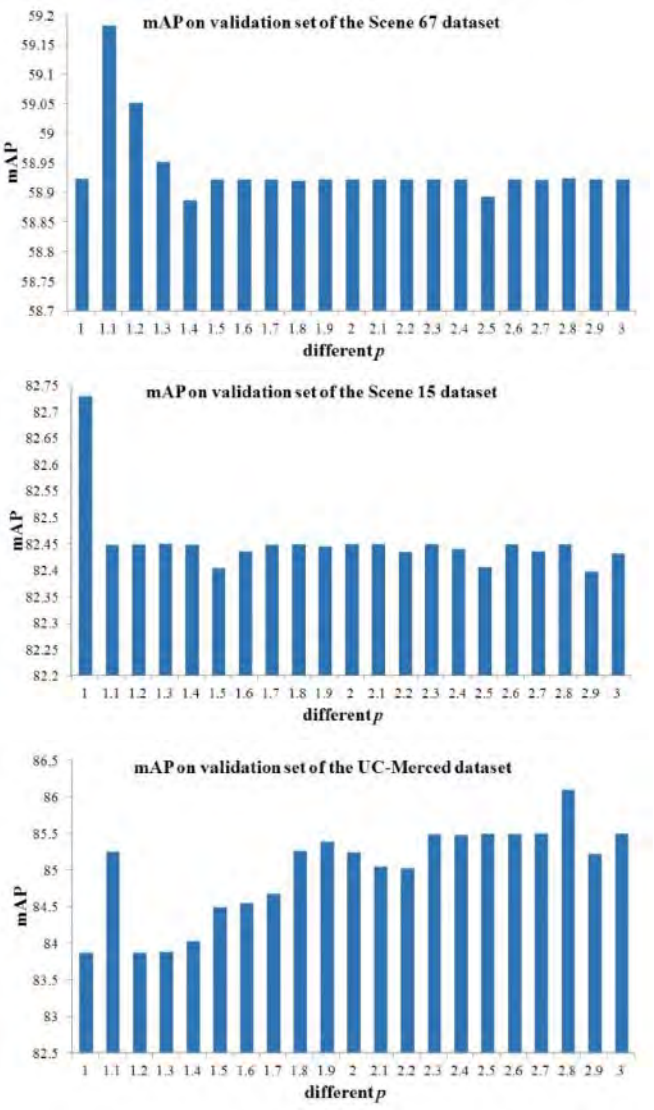


Fig. 7. mAP results of pLapKLS under different p with 10% labeled sample. The y-axis is the mAP over all classes, and the x-axis is the parameter p.

where J is an $l \times (l+u)$ matrix given by $J = [I, 0]$ with I is the $l \times l$ diagonal matrix, and Y is given by $Y = \text{diag}(y_1, \ldots, y_l)$.

By taking the derivative of the reduced Lagrangian $\mathcal{L}^R$ with respect to α, the solution can be obtained as

$$\alpha^* = \big(2\gamma_A I + 2\gamma_I L_p \boldsymbol{K}\big)^{-1} J^T Y \beta^*. \tag{24}$$

Substituting (24) back into (23) we obtain

$$\beta^* = \arg\max_{\beta \in R^l} \sum_{i=1}^{l} \beta_i - \frac{1}{2} \beta^T Q \beta$$
$$\text{s.t. } \sum_{i=1}^{l} \beta_i y_i = 0, 0 \leq \beta_i \leq \frac{1}{l}, i = 1, \ldots, l \tag{25}$$

where $Q = YJ\boldsymbol{K}(2\gamma_A I + 2\gamma_I L_p \boldsymbol{K})^{-1} J^T Y$.

The optimization (25) is a standard SVM problem and can be easily solved with the quadratic programming.

B. pLapR Kernel Least Squares

By selecting the squared loss function, i.e., $V(x_i, y_i, f) = (y_i - f(x_i))^2$, we can obtain

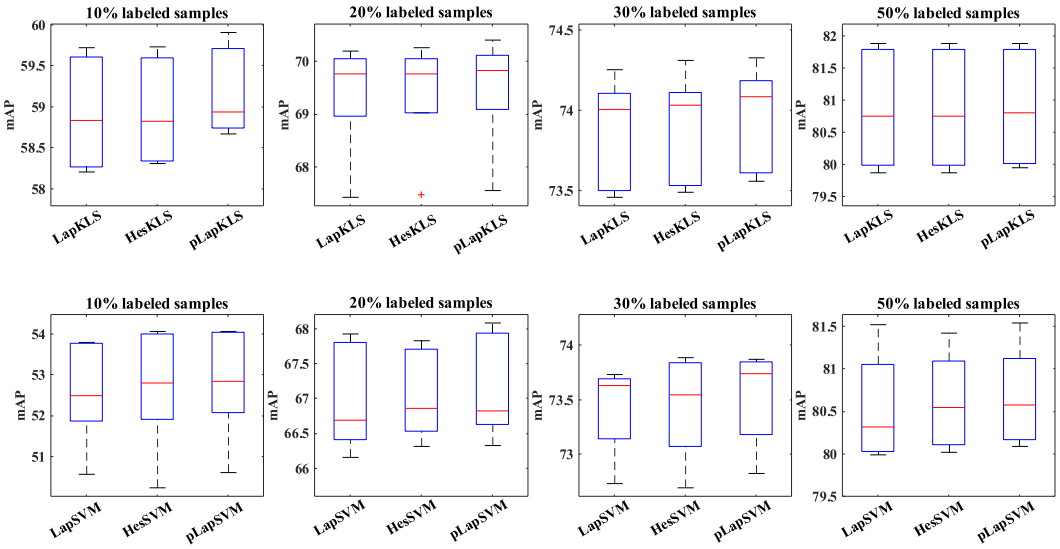


Fig. 8. mAP of different algorithms on Scene 67 dataset. The four subfigures of upper row are KLS methods, and the lower four ones are SVM methods.

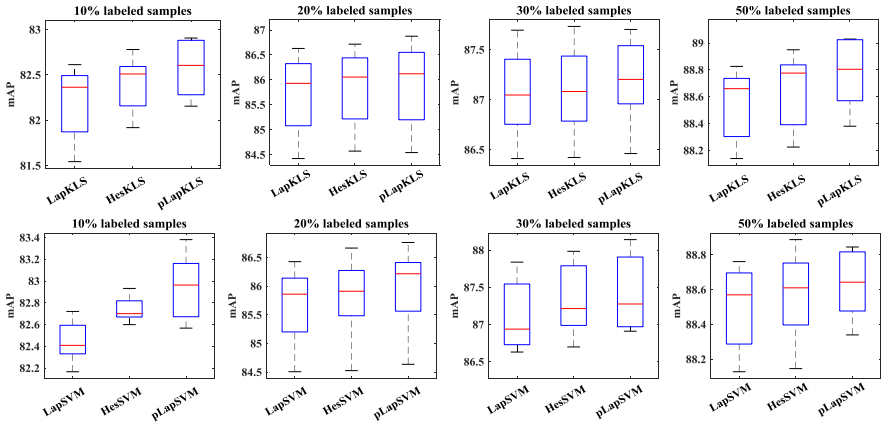


Fig. 9. mAP of different algorithms on Scene 15 dataset. The four subfigures of upper row are KLS methods, and the lower four ones are SVM methods.

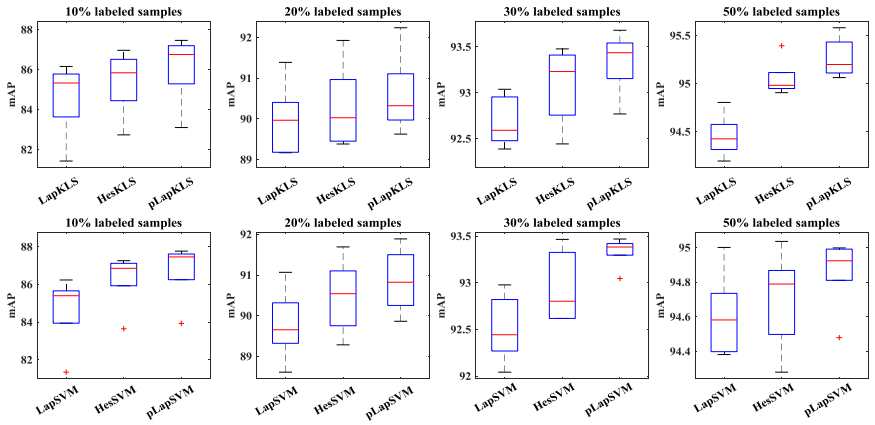


Fig. 10. mAP of different algorithms on UC-Merced dataset. The four subfigures of upper row are KLS methods, and the lower four ones are SVM methods.

pLapKLS

$$\min_{f\in H_K} \frac{1}{l}\sum_{i=1}^{l}(y_i - \boldsymbol{K}(x_i,x)\alpha)^2 + \gamma_A\alpha^T\boldsymbol{K}\alpha + \gamma_I\alpha^T\boldsymbol{K}L_p\boldsymbol{K}\alpha. \tag{26}$$

According to the representer theorem, substituting (17) into (26), the pLapKLS can be written as

$$\min_{\alpha\in R^{l+u}} \frac{1}{l}\sum_{i=1}^{l}(y_i - \boldsymbol{K}(x_i,x)\alpha)^2 + \gamma_A\alpha^T\boldsymbol{K}\alpha + \gamma_I\alpha^T\boldsymbol{K}L_p\boldsymbol{K}\alpha. \tag{27}$$

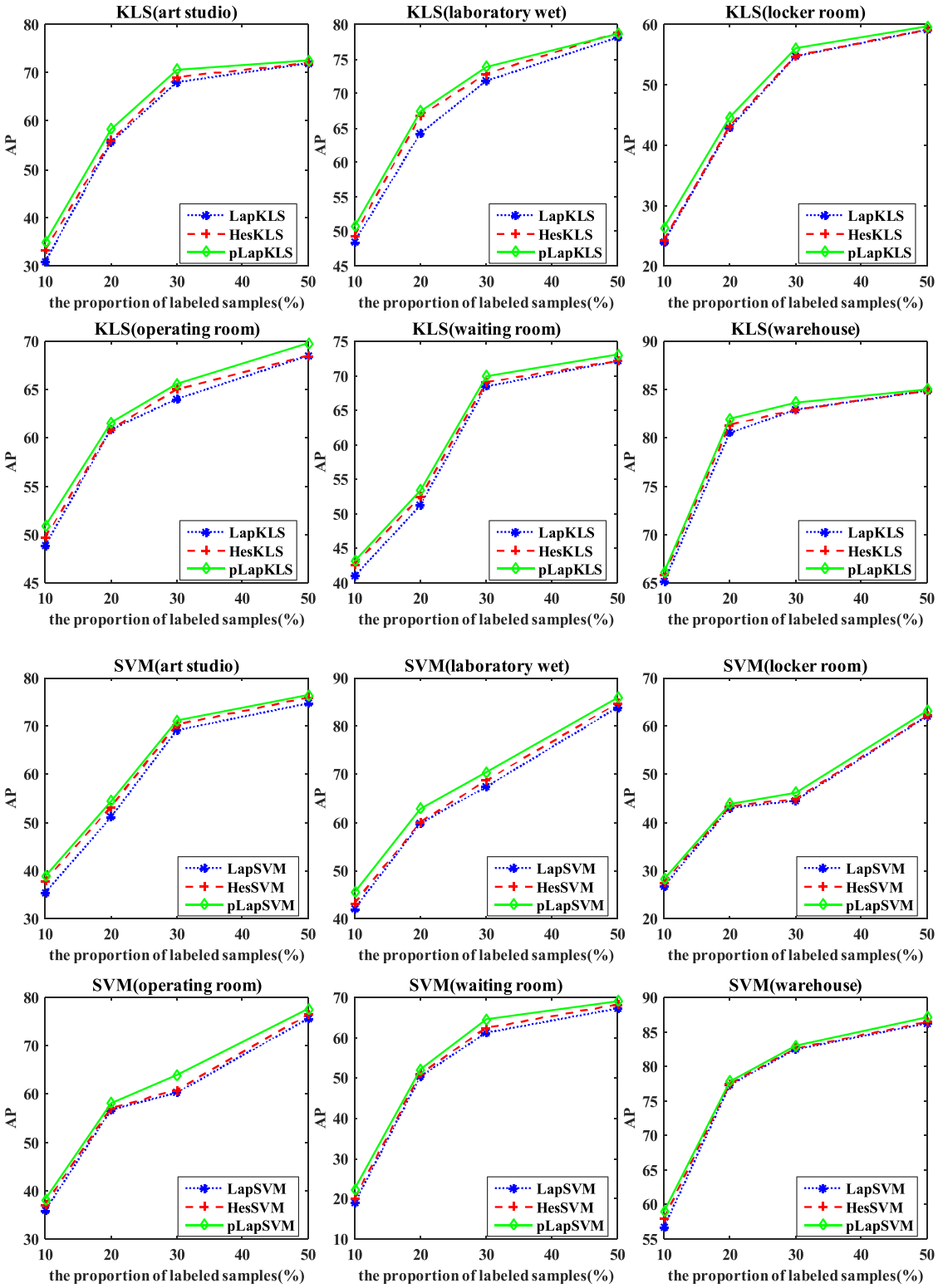


Fig. 11. AP of different methods on several classes of Scene 67 dataset, including art studio, clothing store, laboratory wet, locker room, operating room, and studio music. Each subfigure corresponds on single scene class. The upper six subfigures are KLS methods, and the lower six ones are SVM methods. In each subfigure, the y-axis is the AP results and the x-axis is the number of labeled samples.

The problem can be equally expressed as

$$\min_{\alpha \in R^{l+u}} (Y - J\boldsymbol{K}\alpha)^T (Y - J\boldsymbol{K}\alpha) + \gamma_A \alpha^T \boldsymbol{K}\alpha + \gamma_I \alpha^T \boldsymbol{K} L_p \boldsymbol{K}\alpha \tag{28}$$

where $J = \text{diag}(1, \ldots, 1, 0, \ldots, 0)$ is a diagonal matrix with the first l diagonal entries as 1 and the rest 0, and $Y = [y_1, \ldots, y_l, 0, \ldots, 0]$ is an $(l+u)$-dimensional vector.

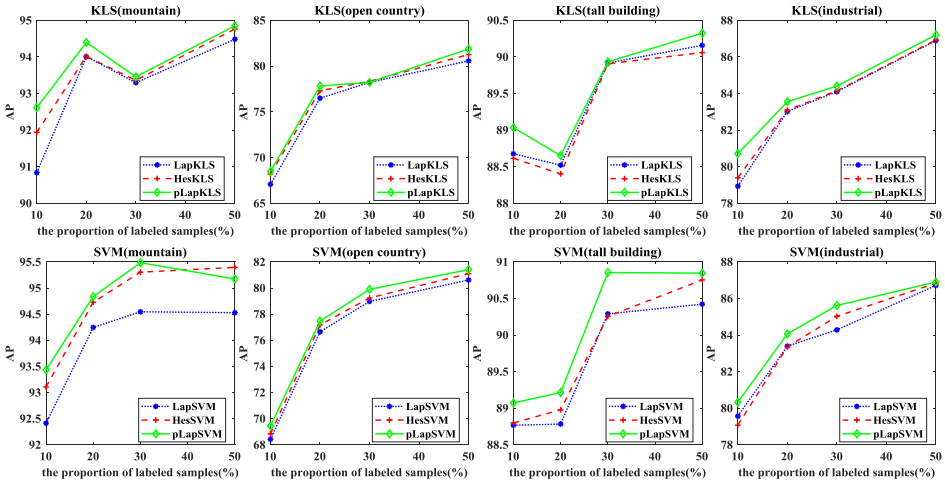


Fig. 12. AP of different methods on several classes of Scene 15 dataset, including mountain, open country, tall building, and industrial. Each subfigure corresponds on single scene class. The upper four subfigures are KLS methods, and the lower four ones are SVM methods. In each subfigure, the y-axis is the AP results and the x-axis is the number of labeled samples.

Then, we take derivative of the objective function with respect to α and set it to zero, which leads to the solution

$$\alpha^* = \left(J\boldsymbol{K} + \gamma_A lI + \gamma_I lL_p\boldsymbol{K}\right)^{-1} Y \tag{29}$$

where I is an $(l+u) \times (l+u)$ identity matrix.

VI. Experiments

To evaluate the effectiveness of the proposed pLapR, we apply pLapSVM and pLapKLS to indoor scene recognition on the Scene 67 database [20], Scene 15 dataset [21], and UC-Merced dataset [22]. Fig. 2 illustrates the framework of pLapR for scene recognition.

The Scene 67 database contains 15 620 indoor scene images collected from different sources, including online image search tools, online photograph sharing sites, and the LabelMe dataset [20]. Particularly, these images can be categorized into 67 classes covering five big scene groups, i.e., stores, home, public spaces, leisure, and working place. Some example images are shown in Fig. 3.

Scene 15 dataset is composed of 15 scene categories, totally 4485 images. Each category has 200 to 400 images. The images contain indoor scenes, such as living room, kitchen, and store, and outdoor scenes, such as forest, mountain, tall building, highway, open country, and so on (see in Fig. 4).

UC-Merced dataset consists of totally 2100 land-use images collected from aerial orthoimage with the pixel resolution of one foot. The original images were downloaded from the United States Geological Survey National Map of 20 U.S. regions. These images were manually selected into 21 classes (see in Fig. 5): 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.

In our experiments, we use the visual features extracted by convolutional neural network with 19 weight layers [35] for Scene 67 dataset and UC-Merced dataset, and extracted SIFT feature for Scene 15 dataset. For Scene 67 dataset, we randomly select 80 images of each scene 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 training data, and the rest for testing. For UC-Merced dataset, we randomly choose 50 images per class as training samples and the rest as testing samples. In semi-supervised classification experiments, we randomly assign a certain percentage (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 partitioning of samples, the above assignment is carried out for five times independently.

The regularization parameters, i.e., γ_A and γ_I are tuned from the candidate set $\{10^i | i = -10, -9, -8, \ldots, 9, 10\}$ and the parameter p for pLapR from the candidate set $\{1, 1.1, 1.2, \ldots, 3\}$ 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 AP (mAP) [36] for overall classes.

First, we show the mAP boxplot of the pLapR on Scene 67 dataset when $p = 2$ and the standard LapR for comparison in Fig. 6. 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. We can clearly see that the performance of pLapR with $p = 2$ is similarly to standard LapR, which demonstrates that when $p = 2$, the graph p-Laplacian becomes the standard graph Laplacian.

Fig. 7 illustrates the mAP performance of pLapKLS with different p values. The y-axis is the mAP over all classes, and the x-axis is the parameter p. The upper subfigure is the performance of the Scene 67 database. We observe that the best performance of indoor scene recognition on the Scene 67 dataset can be obtained with $p = 1.1$. The middle subfigure is the performance of the Scene 15 database and the best performance is achieved when p is equal to 1. The lower subfigure is the performance of the UC-Merced database and the best performance is achieved when p is equal to 2.8.

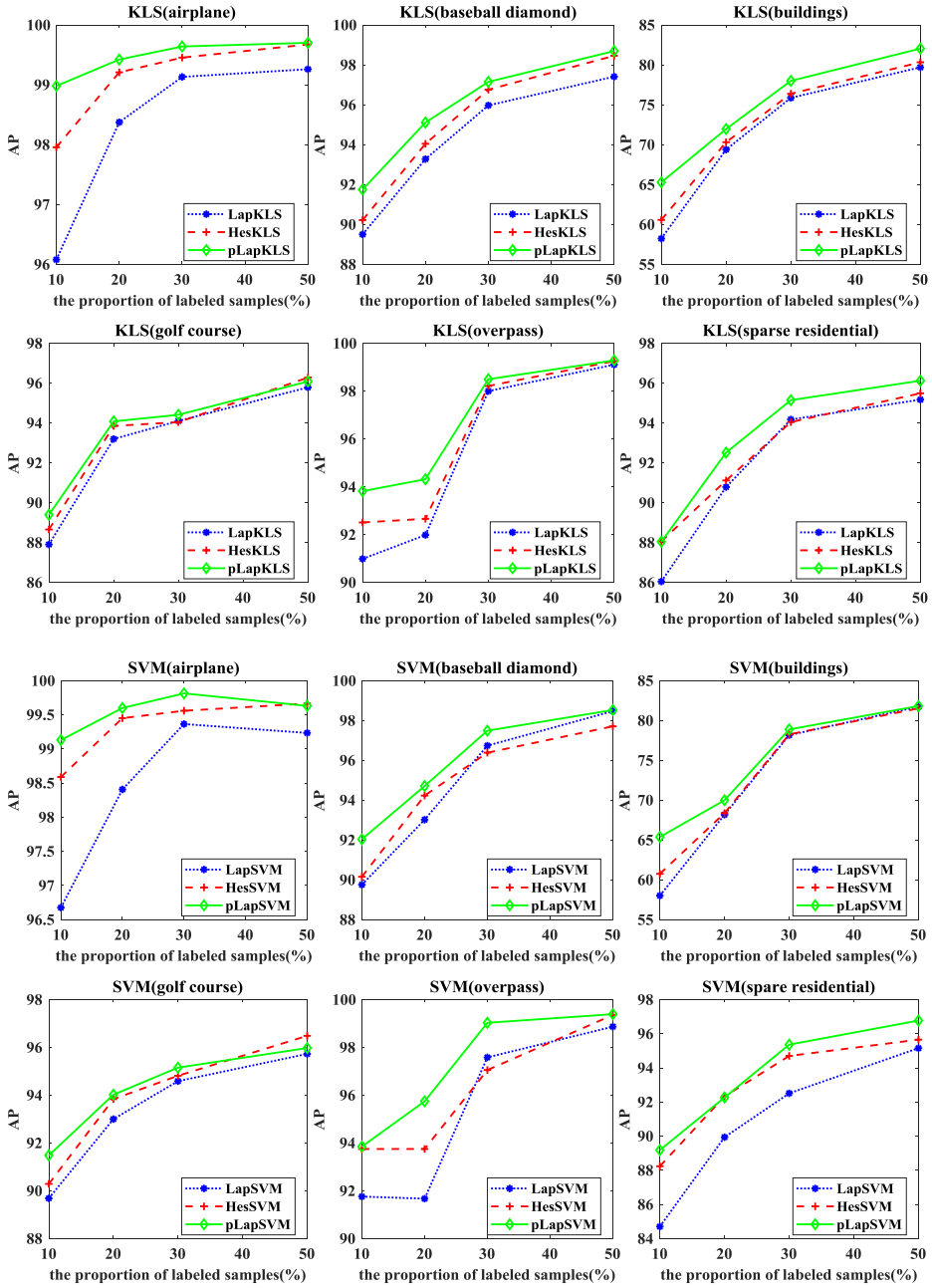


Fig. 13. AP of different methods on several classes of UC-Merced dataset, including airplane, baseball diamond, buildings, golf course, overpass, and sparse residential. Each subfigure corresponds on single class. The upper six subfigures are KLS methods, and the lower six ones are SVM methods. In each subfigure, the *y*-axis is the AP results and the *x*-axis is the number of labeled samples.

Then we evaluate the performance of the pLapR with the representative LapR and HesR. Fig. 8 demonstrates the mAP results of different algorithms on Scene 67 dataset. Figs. 9 and 10 show the mAP performance on Scene 15 dataset and UC-Merced dataset, respectively. The four subfigures of upper row are KLS methods, and the lower four ones are SVM methods. From the results of three datasets, We can see that the pLapR outperforms both LapR and HesR especially when only a small number of samples labeled.

We further compare the AP performance over single scene class of different methods. Fig. 11 shows the AP results of Scene 67 dataset on several selected scene classes, including art studio, clothing store, laboratory wet, locker room, operating room, and studio music. Fig. 12 shows the AP results of Scene 15 dataset about several classes, including mountain, open country, tall building, and industrial. Fig. 13 is the AP results of UC-Merced dataset on several classes, including airplane, baseball diamond, buildings, golf course, overpass, and sparse residential. Each subfigure corresponds on single scene class. The upper six subfigures are KLS methods, and the lower six ones are SVM methods for Scene 67 dataset and UC-Merced dataset. The upper four subfigures are KLS methods, and the lower four ones are SVM methods for Scene 15 dataset. In each subfigure, the *y*-axis is the AP results and the *x*-axis is the number of labeled samples. From the AP results of three datasets, we can find that, in most

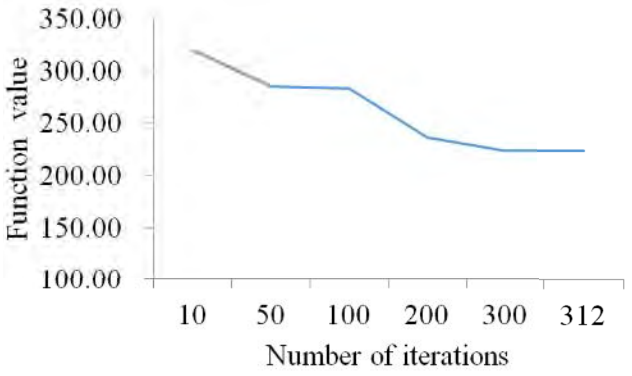


Fig. 14. Convergence analysis of Algorithm 1, the y-axis is the function value $J_E(\mathcal{F})$ and the x-axis is the number of iterations.

TABLE I
P PERFORMANCE OF T-TEST ON SCENE 67 DATASET

Method1/Method2	10%	20%	30%	50%
LapKLS/HesKLS	0.277651	0.17194	0.117367	0.474191
LapKLS/pLapKLS	0.041969	0.035067	0.000269	0.152193
HesKLS/pLapKLS	0.025288	0.023861	0.016305	0.15217
LapSVM/HesSVM	0.342444	0.61985	0.786384	0.274386
LapSVM/pLapSVM	0.01224	0.001167	0.011769	0.035385
HesSVM/pLapSVM	0.183967	0.195357	0.105284	0.054894

TABLE II
P PERFORMANCE OF T-TEST ON SCENE 15 DATASET

Method1/Method2	10%	20%	30%	50%
LapKLS/HesKLS	0.097312	0.00028	0.004491	0.000188
LapKLS/pLapKLS	0.005134	0.002257	0.011592	0.001565
HesKLS/pLapKLS	0.160543	0.18708	0.02098	0.021678
LapSVM/HesSVM	0.012722	0.076196	0.032097	0.048368
LapSVM/pLapSVM	0.003725	0.009646	0.000297	0.007181
HesSVM/pLapSVM	0.10622	0.070104	0.334243	0.172735

TABLE III
P PERFORMANCE OF T-TEST ON UC-MERCED DATASET

Method1/Method2	10%	20%	30%	50%
LapKLS/HesKLS	0.004341	0.025824	0.015218	0.000577
LapKLS/pLapKLS	0.0003	0.003796	0.001381	0.000361
HesKLS/pLapKLS	0.017959	0.022595	0.009905	0.005206
LapSVM/HesSVM	0.003308	0.061032	0.058988	0.462276
LapSVM/pLapSVM	0.000954	0.029988	0.007422	0.122538
HesSVM/pLapSVM	0.021009	0.091766	0.058691	0.06163

cases, the pLapR performs better than the traditional methods including LapR and HesR.

To verify that the convergence of the algorithm, we choose 30% labeled samples of UC-Merced dataset as an example. Fig. 14 shows the performance of the algorithm with $p = 2$ and $k = 10$. The y-axis is the function value $J_E(\mathcal{F})$ and the x-axis is the number of iterations. We can get the algorithm is converge.

To verify that the results are statistically significant, we perform t-test of the experimental results on these three datasets. The performance is measured by similar probability (P) of the results of two methods. From Tables I–III, we can see that, in most cases, $P< 0.05$, which demonstrate the statistically significant of the results and the effectiveness of our proposed method.

VII. CONCLUSION

There are many successful MRSSL algorithms been applied to scene recognition. However, how to explore and exploit the local geometry of data manifold is still a challenging problem. Therefore, we presented pLapR to advance the research of this problem for scene recognition. The proposed pLapR provides convincing theoretical evidence to better preserve the local structure. Furthermore, we introduced a fully approximation algorithm of p-Laplacian to improve computation efficiency significantly. We applied pLapR to SVMs and KLSs as two example implementations for scene recognition. Extensive experiments on the Scene 67 dataset, Scene 15 dataset, and UC-Merced dataset demonstrated that the proposed pLapR outperforms the conventional manifold regularization methods including LapR and HesR.

APPENDIX

Proof of Theorem 1: We briefly describe the proof of Theorem 1 here for better understanding although details can be found in [19].

Proof: We can find that a critical point of F_p at f and get

$$\Delta_p^w f - \frac{\sum_{ij} w_{ij}|f_i - f_j|^p}{2\|f\|_p^p}\phi_p(f_i) = 0.$$

With Definition 1, $(\Delta_p^w f)_i = \lambda_p \phi_p(f_i)$, so $\Delta_p^w f - \lambda_p \phi_p(f) = 0$. Here, f is an eigenvector of Δ_p^w. Moreover, the equation implies that a given eigenvector f to the eigenvalue λ_p is a critical point of F_p if $\lambda_p = F_p(f)$. ■

Proof of Theorem 2: We briefly describe the proof of Theorem 2 here for better understanding although details can be found in [33].

Proof: By taking the derivative of (13) with respect to f^k, we can get

$$\frac{\partial J_E}{\partial f^k} = \frac{1}{\|f^k\|_p^p}\left[\Delta_p^w f^k - \frac{\sum_{ij} w_{ij}|f_i^k - f_j^k|^p}{\|f^k\|_p^p}\phi_p\left(f^k\right)\right].$$

According to Theorem 1

$$\lambda_k^* = \frac{\sum_{ij} w_{ij}|f_i^{*k} - f_j^{*k}|^p}{2\|f^{*k}\|_p^p}.$$

So

$$\frac{\partial J_E}{\partial f^{*k}} = \frac{1}{\|f^{*k}\|_p^p}\left[\Delta_p^w f^{*k} - \lambda_k^* \phi_p\left(f^{*k}\right)\right] = 0.$$

Thus, $f^{*1}, f^{*2}, \ldots, f^{*n}$ are local solutions for (12). ■

Proof of Lemma 1: Any function $f \in H_K$ can be expressed by a component $f_\|$ in the linear subspace and a component $f_\perp$ orthogonal to it. Thus, $f = f_\| + f_\perp$. By the reproducing property, the evaluation of f on any data point x_i, $1 \le i \le l+u$ is independent of the orthogonal component $f_\perp$

$$f(x_i) = \langle f, \boldsymbol{K}(x_i, \cdot)\rangle = \left\langle f_\|, \boldsymbol{K}(x_i, \cdot)\right\rangle + \langle f_\perp, \boldsymbol{K}(x_i, \cdot)\rangle.$$

On the other hand, $\|f\|^2 = \|f_\|\|^2 + \|f_\perp\|^2$ and therefore $\|f\|^2 \ge \|f_\|\|^2$. It follows that the penalty term $\gamma_A\|f\|_K^2 + \gamma_I\|f\|_I^2$ is a strictly monotonically increasing real-valued function on f, and the minimizer f^* is in subspace of H_K with $f_\perp = 0$. As the

component $f_{\|}$ along the subspace spanned by the kernel functions $\{\boldsymbol{K}(x_i, \cdot)\}_{i=1}^{l+u}$, the optimization problem in (14) is given by $f^*(x) = \sum_{i=1}^{l+u} \alpha_i^* \boldsymbol{K}(x_i, x)$.

Proof of Lemma 2: According to Algorithm 1, the eigenvalues of the p-Laplacian is positive, then $L_p = \mathcal{F}\Lambda\mathcal{F}^T$ is positive definite.

Proof of Theorem 3: Since $\boldsymbol{K}(x, y)$ satisfying $\boldsymbol{K}(x, y) = \boldsymbol{K}(y, x)$ is a positive semi-definite kernel, the corresponding matrix $\boldsymbol{K}^{n\times n}$ with $\boldsymbol{K}_{ij} = \boldsymbol{K}(x_i, x_j)$ is positive semi-definite. Meanwhile, L_p is semi-definite positive. Thus, $\gamma_A\|f\|_K^2 + \gamma_I \mathbf{f}^T L_p \mathbf{f}$ is a monotonically increasing real-valued function with respect to f. According to Lemma 1, the proof of Theorem 3 is complete.

REFERENCES

[1] X. Lu, X. Li, and L. Mou, "Semi-supervised multitask learning for scene recognition," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1967–1976, Sep. 2015.

[2] G.-S. Xie, X.-Y. Zhang, S. Yan, and C.-L. Liu, "Hybrid CNN and dictionary-based models for scene recognition and domain adaptation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 6, pp. 1263–1274, Jun. 2017.

[3] S. Yan, X. Xu, D. Xu, S. Lin, and X. Li, "Image classification with densely sampled image windows and generalized adaptive multiple kernel learning," *IEEE Trans. Cybern.*, vol. 45, no. 3, pp. 381–390, Mar. 2015.

[4] D. Tao, D. Tao, X. Li, and X. Gao, "Large sparse cone non-negative matrix factorization for image annotation," *ACM Trans. Intell. Syst. Technol. (TIST)*, vol. 8, no. 3, p. 37, 2017.

[5] L. Zhang, X. Li, L. Nie, Y. Yang, and Y. Xia, "Weakly supervised human fixations prediction," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 258–269, Jan. 2016.

[6] W. Liu, Z.-J. Zha, Y. Wang, K. Lu, and D. Tao, "p-Laplacian regularized sparse coding for human activity recognition," *IEEE Trans. Ind. Electron.*, vol. 63, no. 8, pp. 5120–5129, Aug. 2016.

[7] A. Araujo and B. Girod, "Large-scale video retrieval using image queries," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/TCSVT.2017.2667710.

[8] P. Liu, P. Yang, C. Wang, K. Huang, and T. Tan, "A semi-supervised method for surveillance-based visual location recognition," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3719–3732, Nov. 2016.

[9] M. Fan, X. Zhang, L. Du, L. Chen, and D. Tao, "Semi-supervised learning through label propagation on geodesics," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1486–1499, May 2018.

[10] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Nov. 2006.

[11] K. I. Kim, F. Steinke, and M. Hein, "Semi-supervised regression using hessian energy with an application to semi-supervised dimensionality reduction," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 979–987.

[12] W. Liu and D. Tao, "Multiview hessian regularization for image annotation," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2676–2687, Jul. 2013.

[13] J. Tang, L. Shao, X. Li, and K. Lu, "A local structural descriptor for image matching via normalized graph Laplacian embedding," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 410–420, Feb. 2016.

[14] B. Geng, D. Tao, C. Xu, L. Yang, and X.-S. Hua, "Ensemble manifold regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1227–1233, Jun. 2012.

[15] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2030–2048, Jul. 2011.

[16] W. Allegretto and Y. X. Huang, "A Picone's identity for the p-Laplacian and applications," *Nonlin. Anal. Theory Methods Appl.*, vol. 32, no. 7, pp. 819–830, 1998.

[17] H. Takeuchi *et al.*, "The spectrum of the p-Laplacian and p-harmonic morphisms on graphs," *Illinois J. Math.*, vol. 47, no. 3, pp. 939–955, 2003.

[18] S. Amghibech, "Eigenvalues of the discrete p-Laplacian for graphs," *Ars Combinatoria*, vol. 67, pp. 283–302, Apr. 2003.

[19] T. Bühler and M. Hein, "Spectral clustering based on the graph p-Laplacian," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 81–88.

[20] A. Quattoni and A. Torralba, "Recognizing indoor scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Miami, FL, USA, 2009, pp. 413–420.

[21] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. New York, NY, USA, 2006, pp. 2169–2178.

[22] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2010, pp. 270–279.

[23] Y. Yuan, J. Wan, and Q. Wang, "Congested scene classification via efficient unsupervised feature learning and density estimation," *Pattern Recognit.*, vol. 56, pp. 159–169, Aug. 2016.

[24] Q. Wang, J. Wan, and Y. Yuan, "Deep metric learning for crowdedness regression," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/TCSVT.2017.2703920.

[25] Q. Wang, J. Gao, and Y. Yuan, "A joint convolutional neural networks and context transfer for street scenes labeling," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 5, pp. 1457–1470, May 2018.

[26] F. Cakir, U. Güdükbay, and Ö. Ulusoy, "Nearest-neighbor based metric functions for indoor scene recognition," *Comput. Vis. Image Understand.*, vol. 115, no. 11, pp. 1483–1492, 2011.

[27] Y. Luo, D. Tao, B. Geng, C. Xu, and S. J. Maybank, "Manifold regularized multitask learning for semi-supervised multilabel image classification," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 523–536, Feb. 2013.

[28] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.

[29] J. Yu, D. Tao, and M. Wang, "Adaptive hypergraph learning and its application in image classification," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3262–3272, Jul. 2012.

[30] M. Wang, H. Li, D. Tao, K. Lu, and X. Wu, "Multimodal graph-based reranking for Web image search," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4649–4661, Nov. 2012.

[31] S. Amghibech, "Bounds for the largest p-Laplacian eigenvalue for graphs," *Discrete Math.*, vol. 306, no. 21, pp. 2762–2771, 2006.

[32] D. Zhou and B. Schölkopf, "Regularization on discrete spaces," in *Proc. DAGM Symp.*, vol. 3663, 2005, pp. 361–368.

[33] D. Luo, H. Huang, C. Ding, and F. Nie, "On the eigenvectors of p-Laplacian," *Mach. Learn.*, vol. 81, no. 1, pp. 37–51, 2010.

[34] D. Slepčev and M. Thorpe, "Analysis of p-Laplacian regularization in semi-supervised learning," *CoRR*, vol. abs/1707.06213, 2017.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[36] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010.

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

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

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

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

Her current research interests include pattern recognition and computer vision.

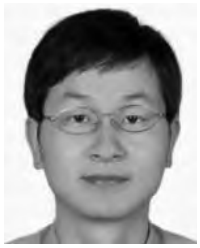
Yicong Zhou (M'07–SM'14) received the B.S. degree in electrical engineering from Hunan University, Changsha, China, and the M.S. and Ph.D. degrees in electrical engineering from Tufts University, Medford, MA, USA.

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

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

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

He is currently an Engineer with the School of Information Science and Engineering, Yunnan University, Kunming, China. He has authored and co-authored over 30 scientific articles. He has served over ten international journals, including the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE SIGNAL PROCESSING LETTERS, and *PLoS-ONE*. His current research interests include machine learning, computer vision, and cloud computing.



Jun Cheng received the B.Eng. and M.Eng. degrees from the University of Science and Technology of China, Hefei, China, in 1999 and 2002, respectively, and the Ph.D. degree from the Chinese University of Hong Kong, Hong Kong, in 2006.

He is currently with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, as a Professor and the Director of the Laboratory for Human Machine Control. His current research interests include computer vision, robotics, machine intelligence, and control.