Prior Knowledge Regularized Multiview Self-Representation and Its Applications

Xiaolin Xiao, Yongyong Chen, Yue-Jiao Gong, *Member, IEEE*, and Yicong Zhou, *Senior Member, IEEE*

***Abstract*— To learn the self-representation matrices/tensor that encodes the intrinsic structure of the data, existing multiview self-representation models consider only the multiview features and, thus, impose equal membership preference across samples. However, this is inappropriate in real scenarios since the prior knowledge, e.g., explicit labels, semantic similarities, and weak-domain cues, can provide useful insights into the underlying relationship of samples. Based on this observation, this article proposes a prior knowledge regularized multiview self-representation (P-MVSR) model, in which the prior knowledge, multiview features, and high-order cross-view correlation are jointly considered to obtain an accurate self-representation tensor. The general concept of "prior knowledge" is defined as the complement of multiview features, and the core of P-MVSR is to take advantage of the membership preference, which is derived from the prior knowledge, to purify and refine the discovered membership of the data. Moreover, P-MVSR adopts the same optimization procedure to handle different prior knowledge and, thus, provides a unified framework for weakly supervised clustering and semisupervised classification. Extensive experiments on real-world databases demonstrate the effectiveness of the proposed P-MVSR model.**

***Index Terms*— Low-rank tensor representation, multiview, prior knowledge, self-representation, semisupervised classification, tensor Singular Value Decomposition (t-SVD), weakly supervised clustering.**

I. Introduction

MANY real-world applications are confronted with multiview data as a single-view feature cannot reveal the structure of the data in most cases. For example, in computer vision, multiple heterogeneous features, such as color, texture, and shape, are used to characterize the images. Usually, each view portrays a specific relationship and captures only partial information of the data. Thus, it is necessary to integrate the information from multiple views to explore the underlying relationship of samples [1], [2]. Taking advantage of the complementary information beneath different views, the multiview models have witnessed the performance enhancement over their single-view counterparts in applications of clustering [3]–[15], semisupervised classification [8], [10], [16], [17], and so on.

Among the extensive studies on multiview learning, the self-representation models [3], [5]–[7], [9], [11]–[13], [18] have become the mainstream. Existing models can be characterized by their assumptions on the cross-view correlation and are roughly classified into two categories.

1) *Pairwise Correlation Based:* For example, Cheng *et al.* [3] concatenated the representation matrices of different views along the column direction as a large representation matrix and then ensured the sparsity-consistency among all views by imposing the $l_{2,1}$-norm on the concatenated representation matrix. Gao *et al.* [7] devised an indicator to uncover a common cluster structure agreed by all views. Meanwhile, the models in [5], [9], and [13] emphasized the complementarity of multiview features by exploring the diversity, exclusivity, and nonlinear local manifold structures from different views, respectively.
2) *High-Order Correlation Based:* To well capture the high-order relationship among samples and across different views, the third-order tensor representation is exploited. Zhang *et al.* [6] stacked the representation matrices into a third-order tensor and imposed a low-rank constraint on this representation tensor. Relying on the unfolding-based tensor nuclear norm (u-TNN) [19], the low-rank constraint in [6] suffers from the loss of representation optimality [11], [12]. To overcome this limitation, the works in [11] and [12] proposed two optimal strategies to measure the low-rank property of the self-representation tensor.

While effective, previous methods consider only the multiview features and their cross-view relationship. In particular, when calculating the representation coefficients of one sample, existing algorithms assign equal preference for all the other samples. This is inappropriate in the setting of weakly supervised clustering where the membership of samples is constrained via prior knowledge, i.e., weak-domain cues. The domain cues are of great importance in correctly detecting the cluster membership especially when the features are not enough discriminant [3]. Taking region-based image segmentation as an example, images are segmented into superpixels, and the superpixels are clustered into homogeneous segments

Manuscript received March 28, 2019; revised October 21, 2019 and March 13, 2020; accepted March 28, 2020. Date of publication April 17, 2020; date of current version March 1, 2021. This work was supported in part by China Postdoctoral Science Foundation under Grant 2019M662913, by The Science and Technology Development Fund, Macau SAR (File no. 189/2017/A3), and by the Research Committee at University of Macau under Grant MYRG2018-00136-FST. *(Corresponding author: Yicong Zhou.)*

Xiaolin Xiao and Yue-Jiao Gong are with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China (e-mail: shellyxiaolin@gmail.com; gongyuejiao@gmail.com).

Yongyong Chen and Yicong Zhou are with the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: yongyongchen.cn@gmail.com; yicongzhou@um.edu.mo).

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

Digital Object Identifier 10.1109/TNNLS.2020.2984625

considering multiview features [3]. According to the Gestalt principle of perceptual organization [20], [21], superpixels within one segment should be spatially connected and compact. However, this requirement cannot be satisfied by existing multiview clustering algorithms since they overlooked this valuable prior knowledge. Similar problems may arise in many real-world applications in which the membership preference derived from prior knowledge is informative.

Moreover, in the application of semisupervised classification, it is also critical to associate the available data labels with multiview features. In addition, fine-grained semantic similarities between samples could also provide rich information about the underlying class membership [22]. In these scenarios, the prior information is drawn from explicit data labels or semantic similarities. To exploit the structure information beneath the data labels, Cai *et al.* [16] proposed a multimodal model to propagate class labels from the labeled samples to the unlabeled ones. By investigating the difficulty of classifying challenging samples, Gong *et al.* [17] resorted to curriculum learning to boost the performance of multimodal learning. Besides, it is feasible to adapt existing multiview clustering models [4] to the semisupervised scenario [16], [18]. Nevertheless, the adaption always requires tedious works to define new strategies for label propagation, with the exceptions of [8] and [10].

Based on these observations, we propose a novel prior knowledge regularized multiview self-representation (P-MVSR) model that seamlessly integrates the comprehensive information from prior knowledge, multiview features, and the high-order cross-view correlation for multiview learning. As a complement of the multiview features, prior knowledge is used to purify and refine the learned self-representation tensor. In real-world applications, the prior knowledge may come from implicit domain cues, explicit data labels, or semantic similarities, with applications to weakly supervised clustering and semisupervised classification. The novelty and contributions of this article lie in the following aspects.

1) We propose a novel P-MVSR model that takes advantage of the prior knowledge for learning the self-representation tensor. To the best of our knowledge, this is the first model to improve the performance of multiview learning using extra information, which is complementary to the multiview features.
2) P-MVSR employs the same optimization procedure for different prior knowledge and, thus, provides a unified framework for semisupervised classification and weakly supervised clustering, relieving the burden of designing new models for different applications. Besides, the multiview clustering algorithm in [12] falls into our special case when no prior knowledge is imposed.
3) We devise an efficient optimization algorithm to solve P-MVSR via the augmented Lagrangian method with theoretical convergence analysis. Besides, the introduction of prior knowledge will not increase the computation complexity.
4) Within the P-MVSR framework, two illustrative applications are investigated by exploiting the spatial cues in region-based image segmentation and the explicit labels/implicit semantic similarities in semisupervised classification. Extensive experiments have demonstrated the effectiveness and the generalization ability of the proposed P-MVSR framework.

In the rest of this article, Section II reviews the related works on multiview learning. Section III introduces the notations and background knowledge. Section IV elaborates the proposed P-MVSR model. The applications of P-MVSR are demonstrated in Section V. Finally, conclusions are drawn in Section VI.

II. Related Work

Multiview learning [1], [2] aims to learn the intrinsic structure of data from diverse views, from which the consensus and/or the complementary information are well-considered. According to the strategies to integrate multiview features, we classify multiview learning methods into two general groups: 1) multiview representation fusion and 2) multiview representation alignment [2].

The core of multiview representation fusion is to blend the multiview inputs into a single common representation shared by all views. Models belonging to this category differ in the common representation measures and/or fusion schemes. Karasuyama and Mamitsuka [23] proposed to learn a common class indicator matrix by unifying the multiple graph Laplacian matrices with a sparse weighting scheme. To learn the weights of multiple graphs automatically, Nie *et al.* [8] devised an autoweighting multiple graph learning (AMGL) algorithm to obtain the common class indicator matrix. Afterward, a multiview learning model with adaptive neighbors (MLAN) [10] was introduced to learn the common cluster indicator matrix and the sample similarity matrix simultaneously. Considering each view as one modal, the multimodal learning techniques were employed to find out the hidden pattern from data, e.g., [16] (AMMSS) and [17] (MMCL).

On the other hand, multiview representation alignment captures the relationship among different views via feature alignment. That is, the individual representations of different views are aligned through predefined metrics, such as the correlation measurement [24], the similarity and distance measurements [25]–[27], meaningful customized structures [3], [9], manifold structures [5], [7], [13], [14], [28], and the low-rank tensor constraints [6], [11], [12], [18]. Representatives of multiview alignment are the canonical correlation analysis-based methods [24], [29] and the co-training-/co-regularization-based methods [25]–[27]. The abovementioned methods suffer limitations in capturing the high-order cross-view correlation [2]. Inspired by the wide applicability of sparse representation [30], [31] and low-rank representation [32], [33], the self-representation-based methods were extended to the multiview scenarios. Increasing research works were developed in this direction, and a large number of multiview learning algorithms have been proposed based on multiview self-representation [3], [5]–[7], [9], [11]–[13], [18]. According to the assumptions of the cross-view correlation, the multiview self-representation models can be sorted into the pairwise correlation-based methods MLAP [3],

DiMSC [5], ECMSC [9], and the high-order correlation-based ones [6], [11], [12].

To exploit the valuable high-order relationship among the multiview representations, the idea of exploiting low-rank tensor representation for multiview subspace clustering (LT-MSC) was introduced in [6]. However, due to the limitation of the u-TNN, LT-MSC suffers losses in capturing the high-order cross-view correlation. To remedy this situation, Yin *et al.* [11] devised a new model by directing using the self-expressiveness of the third-order tensor (3rdT-MSC). Alternatively, Xie *et al.* [12] imposed the tensor singular value decomposition-based tensor nuclear norm (t-SVD-TNN) on the rotated representation tensor to thoroughly explore the high-order cross-view correlation (t-SVD-MSC).

III. Notations and Preliminaries

In this section, we first clarify the notations and then briefly review some preliminary works. Throughout this article, we use the calligraphy letters (e.g., $\mathcal{Z}$) to represent tensors. A blackboard bold letter $\mathbb{Z}$ is used as a counterpart of tensor $\mathcal{Z}$, and they are exploited to represent the original representation tensor and the rotated representation tensor, respectively. The two symbols $\|\cdot\|_*$ and $\|\cdot\|_\circledast$ denote the u-TNN [19] and the t-SVD-TNN [34], [35], respectively. The uppercase letters (e.g., Z) are used to indicate matrices. For the detailed introduction on the tensor algebra, please refer to [12], [35], [49].

A. Low-Rank Tensor Representation-Based Multiview Learning

Traditional multiview learning models [3], [5], [7], [9] can only capture the pairwise correlation of different views. To avoid this limitation, a third-order self-representation tensor can be constructed to simultaneously utilize all views, and its tensor rank is exploited as a metric for multiview representation alignment.

A general formulation of low-rank tensor representation-based multiview learning [**?**], [6], [11], [12] is given by

$$\begin{aligned} &\min_{\mathbb{Z},E} \mathcal{R}(\mathbb{Z}) + \lambda\|E\|_r \\ &\text{s.t. } \mathbb{Z} \in \Omega,\ E \in \Xi \end{aligned} \tag{1}$$

where $\mathbb{Z} \in \mathbb{R}^{n\times n\times V}$ is a third-order tensor constructed by stacking V representation matrices $\{Z^{(v)}\}$ along the third direction; $Z^{(v)} \in \mathbb{R}^{n\times n}$ measures the self-representation coefficients among n samples according to the vth view feature; Ω and Ξ are two compact convex sets; $\mathcal{R}(\mathbb{Z})$ is used to induce the low-rankness of $\mathbb{Z}$; and E is the concatenation of error matrices, while $\|\cdot\|_r$ indicates the regularization strategy, e.g., the squared Frobenius norm ($\|\cdot\|_F^2$) can be used to model the Gaussian noise and the $l_{2,1}$-norm is commonly adopted to deal with sample-specific corruptions and outliers [32]. Within this framework, Zhang *et al.* [6] exploited the u-TNN $\mathcal{R}(\mathbb{Z}) = \|\mathbb{Z}\|_*$, which is defined as the sum of the nuclear norms of the matrices unfolded along all modes [19] to capture the high-order cross-view correlation.

B. t-SVD-TNN

While effective in many applications, minimizing u-TNN essentially imposes the low-rank constraint in the matrix SVD-based vector space, resulting in an inadequate representation of the tensor low-rankness [12]. Besides, u-TNN lacks a clear physical meaning. To remedy these problems, the work in [12] exploited the t-SVD-TNN $\mathcal{R}(\mathbb{Z}) = \|\mathbb{Z}\|_\circledast$ to provide a more accurate low-rank constraint. As shown in Fig. 1, given a tensor $\mathbb{Z} \in \mathbb{R}^{n_1\times n_2\times n_3}$, let its t-SVD be $\mathbb{Z} = \mathcal{U} * \mathcal{S} * \mathcal{V}'$, where $'$ is the transpose operator. Then, $\mathcal{U} \in \mathbb{R}^{n_1\times n_1\times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2\times n_2\times n_3}$ are orthogonal tensors, and $\mathcal{S} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is an f-diagonal tensor [34], [35].

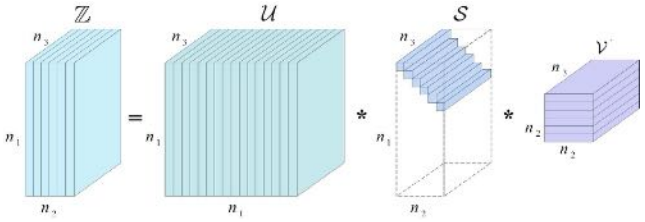

Fig. 1. t-SVD of an $n_1 \times n_2 \times n_3$ tensor $\mathbb{Z}$.

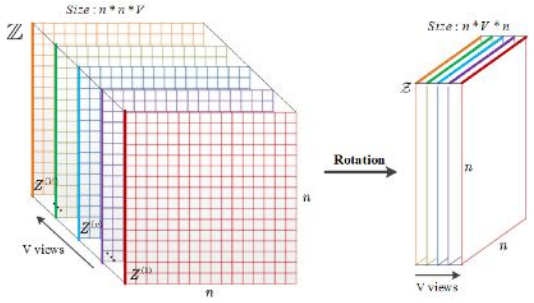

Fig. 2. Rotation of the representation tensor.

The t-SVD-TNN is defined as the sum of the singular values of all frontal slices of $\mathbb{Z}_f$

$$\|\mathbb{Z}\|_\circledast = \sum_{k=1}^{n_3} \|\mathbb{Z}_f(:,:,k)\|_* = \sum_{i=1}^{\min\{n_1,n_2\}} \sum_{k=1}^{n_3} |\mathcal{S}(i,i,k)| \tag{2}$$

where $\mathbb{Z}_f = fft(\mathbb{Z}, [], 3)$ is obtained by applying fast Fourier transformation (FFT) on $\mathbb{Z}$ along the third dimension, and $\mathbb{Z}$ and $\mathbb{Z}_f$ are of the same size.

As reported in [12], it is inappropriate to directly impose t-SVD-TNN on $\mathbb{Z}$ due to the intrinsic circulant algebra underlying t-SVD-TNN. To fix this problem, the dimensionality of $\mathbb{Z}$ is shifted to obtain an $n \times V \times n$ rotated representation tensor $\mathcal{Z}$, as illustrated in Fig. 2.

Remark: The single-view self-representation matrix uncovers the pairwise affinities of samples [32], [36]. By stacking the self-representation matrices from multiple views, the third-order self-representation tensor then captures the high-order relationship among samples and across different views [12]. Conceptually, the t-SVD-TNN imposes a structural constraint on this self-representation tensor to encourage a consensus low-rank tensor structure beneath the third-order affinities of samples. More specifically, the low-rankness of the self-representation tensor can be used to reveal the underlying relationship of samples, somewhat analogous to the low-rank matrix representation [32]. The only difference is that the former employs the high-order relationship, while the latter uses the pairwise relationship.

From this viewpoint, both u-TNN and t-SVD-TNN can be used to explore the high-order cross-view correlation. However, t-SVD-TNN possesses representation optimality due to the following facts.

1) The rank of the tensor is computationally intractable, and t-SVD-TNN has been proven to be the tightest convex relaxation to the l_1-norm of the tensor multirank [37, Th. 2.4.1].
2) Adopting the rotation operation, each frontal slice of the rotated representation tensor considers the information among different samples and different views in the Fourier domain. This way, t-SVD-TNN well depicts the complicated relationship between samples and views [12].

Throughout this article, we resort to t-SVD-TNN as a surrogate to replace the rank of the representation tensor for computational tractability.

IV. Proposed P-MVSR

In this section, we introduce our P-MVSR model in which the prior knowledge, for the first time, is used to optimize the representation tensor in the process of multiview learning.

A. Model of P-MVSR

Suppose that $X^{(v)} \in \mathbb{R}^{d_v \times n}$ $(v = 1, 2, \ldots, V)$ is the feature matrix of the vth view, d_v denotes the dimension of feature, and n is the number of samples. Our P-MVSR model is presented as

$$\begin{aligned} &\min_{Z^{(v)},E} \ \|\mathcal{Z}\|_{\circledast} + \lambda\|E\|_{2,1} \\ &\text{s.t. } X^{(v)} = X^{(v)}\Big(\underbrace{P^{(v)}}_{\textbf{Prior}} \odot Z^{(v)}\Big) + E^{(v)}, \quad v = 1, 2, \ldots, V \\ &\qquad \mathcal{Z} = \phi(Z^{(1)}, Z^{(2)}, \ldots, Z^{(V)}) \\ &\qquad E = [E^{(1)}; E^{(2)}; \ldots; E^{(V)}] \end{aligned} \tag{3}$$

where $\|\cdot\|_{\circledast}$ is the t-SVD-TNN defined in (2); $P^{(v)}$ is the prior knowledge matrix that describes the membership preference of the vth view, and practically, we can use either coarse-grained prior knowledge by assuming $P^{(1)} = \cdots = P^{(V)}$ or fine-grained prior similarity for each specific view; $\odot$ is the Hadamard product operator; $\phi(\cdot)$ stacks the representation matrices $\{Z^{(v)}\}$ into a third-order tensor and then shifts it to obtain an $n*V*n$ rotated representation tensor $\mathcal{Z}$, as illustrated in Fig. 2; and E is obtained by vertically concatenating the error matrices $\{E^{(v)}\}$. The underlying assumption beneath this concatenation operation is that natural corruptions are always sample-specific, i.e., some data are corrupted, while others are clean [6]. In this viewpoint, we stack $\{E^{(v)}\}$ vertically to enforce the columns of $E^{(1)}, \ldots, E^{(V)}$ to have jointly consistent magnitude values across views. Accordingly, the $l_{2,1}$-norm is used to overwhelm the effect of sample-specific corruptions.

In real-world applications, the prior knowledge may come from any complementary information of the multiview features, e.g., labels, semantic similarities, and domain cues. It provides a meaningful way to assign specific membership preferences across samples when calculating the self-representation coefficients. As a result, explicit or implicit cues will have a direct influence on the discovered relationship of the samples. Within the P-MVSR framework, we can adapt the prior knowledge matrices $\{P^{(v)}\}$ for different applications:

1) For semisupervised classification, the vth prior knowledge matrix $P^{(v)}$ is defined as

$$P^{(v)}_{i,j} = \begin{cases} \text{sim}(i, j), & \text{if} \quad l_i = l_j \\ 0, & \text{if} \quad l_i \neq l_j \\ \tau, & \text{otherwise} \end{cases} \tag{4}$$

where $\text{sim}(i, j)$ is the similarity of samples i and j; l_i $(i = 1, \ldots, n)$ represents the label of the ith sample when it is known; and $\tau \in (0, 1)$ is a constant, and it is imposed on the sample pairs when at least one label is unavailable. In practice, different similarity measures will be depicted in Section V-A [see (17) and (18)].
2) In weakly supervised clustering, $\{P^{(v)}\}$ are specified by domain cues; two examples of the domain-specific priors in the application of region-based image segmentation will be introduced in Section V-B [see (19) and (20)].
3) Without prior knowledge, all entries of $\{P^{(v)}\}$ are set to one, and thus, the clustering model in [12] can be considered as a special case of P-MVSR.

B. Solution of P-MVSR

Due to the Hadamard product of $\{P^{(v)}\}$ and $\{Z^{(v)}\}$, it is difficult to solve (3) and is intractable to simultaneously update variables $\{Z^{(v)}\}$ and E. We exploit the augmented Lagrange multiplier with an alternating direction minimization scheme [38] to solve (3) for its efficiency and effectiveness. By introducing V auxiliary variables $\{D^{(v)}\}$ and an auxiliary tensor $\mathcal{G}$, the optimization of (3) can be equivalently transformed into the following optimization problem:

$$\begin{aligned} &\min_{Z^{(v)},D^{(v)},E,\mathcal{G}} \|\mathcal{G}\|_{\circledast} + \lambda\|E\|_{2,1} \\ &\quad \text{s.t. } X^{(v)} = X^{(v)}D^{(v)} + E^{(v)} \\ &\qquad D^{(v)} = P^{(v)} \odot Z^{(v)}, \quad v = 1, 2, \ldots, V \\ &\qquad \mathcal{Z} = \phi(Z^{(1)}, Z^{(2)}, \ldots, Z^{(V)}) \\ &\qquad E = [E^{(1)}; E^{(2)}; \ldots; E^{(V)}] \\ &\qquad \mathcal{Z} = \mathcal{G}. \end{aligned} \tag{5}$$

The optimal variables $\{Z^{(v)}\}, \{D^{(v)}\}, E$, and $\mathcal{G}$ can be alternately obtained by minimizing the augmented Lagrange function of (5) as

$$\begin{aligned} \mathcal{L}(Z^{(v)}, D^{(v)}, E, \mathcal{G}; \Theta_1, \Theta_2, \Theta_3) &= \|\mathcal{G}\|_{\circledast} + \lambda\|E\|_{2,1} \\ &+ \frac{\rho}{2}\left[\sum_{v=1}^{V}\left(\left\|X^{(v)} - X^{(v)}D^{(v)} - E^{(v)} + \frac{\Theta_1^{(v)}}{\rho}\right\|_F^2\right.\right. \\ &\left.\left. + \left\|D^{(v)} - P^{(v)} \odot Z^{(v)} + \frac{\Theta_2^{(v)}}{\rho}\right\|_F^2\right) + \left\|\mathcal{Z} - \mathcal{G} + \frac{\Theta_3}{\rho}\right\|_F^2\right] \end{aligned} \tag{6}$$

where Θ_1, Θ_2, and Θ_3 are the Lagrange multipliers and $\rho > 0$ is the penalty parameter. More specifically, the optimization of (6) is composed of the following subproblems.

1) $Z^{(v)}$-Subproblem: Fixing other variables except $Z^{(v)}$, the problem reduces to

$$\min_{Z^{(v)}} \|P^{(v)} \odot Z^{(v)} - B^{(v)}\|_F^2 + \|Z^{(v)} - C^{(v)}\|_F^2 \tag{7}$$

where $B^{(v)} = D^{(v)} + (\Theta_2^{(v)}/\rho)$ and $C^{(v)} = G^{(v)} - (\Theta_3^{(v)}/\rho)$. Due to the Hadamard product, the optimization procedure is composed of elementwise operations. Without loss of generality, we take the optimization of the (i, j)th entry of $Z^{(v)}$ as an example

$$\min_{Z_{i,j}^{(v)}} \left(P_{i,j}^{(v)} * Z_{i,j}^{(v)} - B_{i,j}^{(v)}\right)^2 + \left(Z_{i,j}^{(v)} - C_{i,j}^{(v)}\right)^2. \tag{8}$$

Then, $Z_{i,j}^{(v)*} = (P_{i,j}^{(v)} * B_{i,j}^{(v)} + C_{i,j}^{(v)})/(1 + P_{i,j}^{(v)^2})$ is obtained by setting derivative of (8) to zero.

2) $D^{(v)}$-Subproblem: Fixing other variables except $D^{(v)}$, (6) reduces to

$$\min_{D^{(v)}} \left\|X^{(v)} D^{(v)} - T_1^{(v)}\right\|_F^2 + \left\|D^{(v)} - T_2^{(v)}\right\|_F^2 \tag{9}$$

where $T_1^{(v)} = X^{(v)} - E^{(v)} + \Theta_1^{(v)}/\rho$ and $T_2^{(v)} = P^{(v)} \odot Z^{(v)} - \Theta_2^{(v)}/\rho$. By setting the derivative of (9) to zero, the closed-form solution is given by

$$D^{(v)^*} = (X^{(v)'} X^{(v)} + \mathbf{I})^{-1} \left(X^{(v)'} T_1^{(v)} + T_2^{(v)}\right). \tag{10}$$

In practice, we can precalculate $(X^{(v)'} X^{(v)} + \mathbf{I})^{-1}$ to avoid extra computation cost.

3) E-Subproblem: Fixing other variables except E and concatenating V matrices $\{X^{(v)} - X^{(v)} D^{(v)} + \Theta_1^{(v)}/\rho\}$ along the column direction as a temporary matrix W, E can be obtained by minimizing

$$\min_E \frac{1}{2}\|E - W\|_F^2 + \frac{\lambda}{\rho}\|E\|_{2,1}. \tag{11}$$

Equation (11) is a group Lasso problem, and we use the Lemma 1 (see [39, Lemma 3.1]) to find the solution.

Lemma 1: Given a matrix $W \in \mathbb{R}^{m*n}$ and a positive scalar σ, the optimal solution of

$$\min_E \frac{1}{2}\|E - W\|_F^2 + \sigma\|E\|_{2,1} \tag{12}$$

is obtained at

$$E^*(:, j) = \begin{cases} \dfrac{\|W(:, j)\|_2 - \sigma}{\|W(:, j)\|_2} W(:, j), & \text{if } \sigma < \|W(:, j)\|_2 \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

4) $\mathcal{G}$-Subproblem: Fixing other variables except $\mathcal{G}$, the closed-form solution of $\mathcal{G}$ can be calculated by optimizing

$$\min_{\mathcal{G}} \frac{1}{\rho}\|\mathcal{G}\|_{\circledast} + \frac{1}{2}\|\mathcal{G} - \mathcal{F}\|_F^2 \tag{14}$$

where $\mathcal{F} = \mathcal{Z} + (\Theta_3/\rho)$. Equation (14) is the t-SVD-TNN minimization problem, and it can be solved by using the tensor tubal-shrinkage operator [12], [40] as

$$\mathcal{G}^* = \mathcal{C}_{\frac{V}{\rho}}(\mathcal{F}) = \mathcal{U} * \mathcal{C}_{\frac{V}{\rho}}(\mathcal{S}) * \mathcal{V}' \tag{15}$$

where $\mathcal{F} = \mathcal{U} * \mathcal{S} * \mathcal{V}'$ and $\mathcal{C}_{\frac{V}{\rho}}(\mathcal{S}) = \mathcal{S} * \mathcal{J}$, in which $\mathcal{J} \in \mathbb{R}^{n \times n \times V}$ is an f-diagonal tensor whose diagonal element in the Fourier domain is $\mathcal{J}(i, i, j) = \max\{1 - V/(\rho^* S(i, i, j)), 0\}$.

5) Multipliers and Penalty Parameter: The multipliers and penalty parameter can be updated by

$$\begin{aligned} \Theta_1^{(v)*} &= \Theta_1^{(v)} + \rho(X^{(v)} - X^{(v)} D^{(v)} - E^{(v)}) \\ \Theta_2^{(v)*} &= \Theta_2^{(v)} + \rho(D^{(v)} - P^{(v)} \odot Z^{(v)}) \\ \Theta_3^* &= \Theta_3 + \rho(\mathcal{Z} - \mathcal{G}) \\ \rho^* &= \min\{\beta * \rho,\ \rho_{\max}\}. \end{aligned} \tag{16}$$

To accelerate the convergence, we employ a continuation scheme [38] to iteratively update the penalty parameter ρ until a maximum value ρ_{max} is achieved, and β is empirically set to 2.

Afterward, the affinity matrix of samples can be formulated as $A = (1/V)\sum_{v=1}^{V}(|Z^{(v)}| + |Z^{(v)'}|)$. The whole procedure of P-MVSR is summarized in Algorithm 1.

Algorithm 1: P-MVSR

Input: multiview feature matrices: $\{X^{(v)}\}$; parameter: λ; prior knowledge matrices: $\{P^{(v)}\}$;
Initialize: $\{Z^{(v)}\}$, $\{D^{(v)}\}$, E, $\mathcal{G}$, $\{\Theta_1^{(v)}\}$, $\{\Theta_2^{(v)}\}$, Θ_3 are initialized to $\mathbf{0}$, ρ is initialized to 10^{-3}, $\rho_{max} = 10^{10}$, $\beta = 2$, $\epsilon = 10^{-7}$;
while not converged **do**
  for $v = 1$ to V **do**
    Update $Z^{(v)}$ by Eq. (7);
    Update $D^{(v)}$ by Eq. (10);
  end for
  Update E by Eq. (13);
  Update $\mathcal{G}$ by Eq. (15);
  Update $\{\Theta_1^{(v)}\}$, $\{\Theta_2^{(v)}\}$, Θ_3, and ρ by Eq. (16);
  Check the convergence conditions
    $\max_{v=1}^{V}\{\|X^{(v)} - X^{(v)} D^{(v)} - E^{(v)}\|_\infty\} \le \epsilon$,
    $\max_{v=1}^{V}\{\|D^{(v)} - P^{(v)} \odot Z^{(v)}\|_\infty\} \le \epsilon$,
    $\|\mathcal{Z} - \mathcal{G}\|_\infty \le \epsilon$,
end while
Output: affinity matrix A.

C. Discussion

1) Convergence Analysis: The objective function of P-MVSR is coupled with respect to the representation tensor $\mathcal{Z}$ since $\mathcal{Z}$ is constrained by the t-SVD-TNN, prior knowledge, and the self-expressive property of each view. We introduce $\mathcal{G}$ and $\{D^{(v)}\}$ to make $\mathcal{Z}$ separable, resulting in four block variables. It has been proven that the direct extension of the augmented Lagrangian multipliers for multiblock convex optimization is not necessarily convergent [41]. Therefore, it is infeasible to strictly prove the convergence properties of P-MVSR. Nevertheless, as verified in pioneer work [6], [12], the convexity of the Lagrange function could guarantee the empirical validity of self-representation-based subspace learning methods to some extent. We will show the empirical convergence curves of P-MVSR in Section V-C2.

2) Computation Complexity: The computation cost of P-MVSR is examined by considering the following subproblems.

1) The $Z^{(v)}$-subproblem for all views costs $\mathcal{O}(Vn^2)$ operations.
2) When solving the $D^{(v)}$-subproblem, the inverse of matrix can be calculate in advance, and thus, its cost can be ignored, while the bottleneck lies in the cost of matrix multiplication, which is of order $\mathcal{O}(Vn^3)$.
3) The E-subproblem costs $\mathcal{O}(Vn^2)$ operations per iteration.
4) Considering the $\mathcal{G}$-subproblem, it takes $\mathcal{O}(Vn^2\log(n))$ operations to calculate 3-D FFT and inverse FFT and $\mathcal{O}(V^2n^2)$ operations for SVD; since $n \gg V$ and $\log(n) > V$, $\mathcal{O}(V^2n^2)$ is negligible compared with $\mathcal{O}(Vn^2\log(n))$.

Therefore, the total computation complexity of P-MVSR is $\mathcal{O}(iteVn^3)$, where *ite* is the number of iterations. Compared with the computation complexity of the method [12], the use of prior knowledge will not lead to extra computation complexity since it only contains elementwise operations.

3) Comparison With Other Low-Rank Representation-Based Multiview Learning Methods: The models in [6] and [12] are most relevant to our work because all of them first stack the multiple self-representation matrices into a third-order tensor and then exploit the low-rank constraint on the representation tensor to capture the high-order cross-view correlation. However, they differ from motivations, mathematical formulations, and the applications.

1) P-MVSR exploits the prior knowledge that is a complement to multiview features to purify and refine the accuracy of the self-representation tensor, whereas [6] and [12] focus on designing strategies to encode the high-order cross-view correlation.
2) In [6] and [12], the low-rank constraints (with different definitions) are used to optimize the self-representation tensor. By contrast, the self-representation tensor in P-MVSR is jointly constrained by the low-rank constraint and the prior knowledge. Since the prior knowledge can be considered as a complement of the multiview features, the proposed P-MVSR will take advantage of this complementary information for performance enhancement.
3) P-MVSR adopts the same procedure to handle different prior knowledge and provides a unified framework for weakly supervised clustering and semisupervised classification. Meanwhile, [6] and [12] are designed especially for the clustering tasks.

V. Experiments

As the proposed P-MVSR incorporates the prior knowledge for learning the self-representation tensor, it is expected to take advantage of the membership preference derived from the prior knowledge to find an accurate class/cluster membership from the data. In this section, two illustrative applications are presented by exploiting different prior knowledge, resulting in the settings of semisupervised classification and weakly supervised clustering, respectively.

A. P-MVSR for Semisupervised Classification

1) Experimental Settings:

a) Setups for P-MVSR: For semisupervised classification, a straightforward way is to exploit data labels as a hard similarity measure. In this setting, (4) results in a coarse-grained prior $P_c^{(1)} = \cdots P_c^{(v)} = \cdots P_c^{(V)}$, and

$$P_c^{(v)}(i,j) = \begin{cases} 1, & \text{if} \quad l_i = l_j \\ 0, & \text{if} \quad l_i \neq l_j \\ \tau, & \text{otherwise.} \end{cases} \tag{17}$$

Practically, the semantic similarities between samples could be quite complex to provide rich information about the underlying class relationship. When being applied to multiview hashing, the fine-grained ranking technique has shown promising performance at different semantic levels [22]. Inspired by this, we introduce a fine-grained prior $P_f^{(v)}$ as

$$P_f^{(v)}(i,j) = \begin{cases} \exp\left(-\dfrac{\left\|x_i^{(v)} - x_j^{(v)}\right\|_2^2}{\sigma_f^2}\right), & \text{if} \quad l_i = l_j \\ 0, & \text{if} \quad l_i \neq l_j \\ \tau, & \text{otherwise} \end{cases} \tag{18}$$

where $x_i^{(v)}$ denotes the feature vector of the ith sample from the vth view, and σ_f is set to the average of all feature distance pairs. Compared with P_c_MVSR, the core of P_f_MVSR is to extract membership preference from fine-grained semantic similarities instead of the coarse-grained labels. The selection of other two parameters λ and τ will be examined in Section V-C1. Once the affinity matrix A is obtained using Algorithm 1, we set $-A$ as the distance matrix of samples, and the unlabeled samples are classified using the nearest neighbor classifier.

b) Databases and multiview features: Ten benchmark databases on handwritten digits (HW), scene (MSRC-v1, MITIndoor-67, Scene-15), web images (NUS-WIDE), face (ORL, Yale), animals (AwA), and genetic objects (Caltech101-20, Caltech101) are chosen for experiments. The summary of databases and the corresponding multiview features is reported in Table I. For detail descriptions of the multiview features extracted from these databases, please refer to [10], [12], and [42].

c) Competing algorithms: We compare the performance of P_c_MVSR and P_f_MVSR with four state-of-the-arts, i.e., AMMSS [16], AMGL [8], MMCL [17], and MLAN [10]. Among them, AMGL, MLAN, P_c_MVSR, and P_f_MVSR can be applied to both clustering and semisupervised classification, while AMMSS and MMCL are specialized for the task of semisupervised classification. For a fair comparison, we test the performance of competing algorithms over the recommended parameters on all data sets, respectively. Besides, the ratios of available labeled samples are fixed to the first 10%, 20%, and 30% of all samples, respectively.

2) Classification Accuracy: The best classification results of all algorithms are reported in Tables II and III. It can be observed that P_c_MVSR and P_f_MVSR obtain obvious advantages over its competitors. In particular, the performances of AMMSS, MLAN, P_c_MVSR, and P_f_MVSR are

TABLE I

SUMMARY OF THE MULTIVIEW DATABASES USED FOR SEMISUPERVISED CLASSIFICATION

	HW	**MSRC-v1**	**Caltech101-20**	**ORL**	**Yale**
Content	handwritten digits	scene	generic objects	face	face
Classes	10	7	20	40	15
Total number	2000	210	2386	400	165
View 1	Fourier of shape(76 d)	color moment(24 d)	Gabor(48 d)	intensity(4096 d)	intensity(4096 d)
View 2	profile correlations(216 d)	GIST(512 d)	wavelet moments(40 d)	LBP(3304 d)	LBP(3304 d)
View 3	Karhunen-love(64 d)	CENTRIST(254 d)	CENTRIST(254 d)	Gabor(6750 d)	Gabor(6750 d)
View 4	pixel averages(240 d)	LBP(256 d)	HOG(1984 d)	-	-
View 5	Zernike moment(47 d)	-	GIST(512 d)	-	-
View 6	morphologic(6 d)	-	LBP(928 d)	-	-
	AwA	**NUS-WIDE**	**Caltech101**	**MITIndoor-67**	**Scene-15**
Content	animals	web images	generic objects	scene	scene
Classes	50	25	101	67	15
Total number	30475	3000	8677	5360	4485
View 1	color histogram(2688 d)	color histogram(64 d)	Gabor(48 d)	PHOW(3600 d)	PHOW(1800 d)
View 2	local self-similarity(2000 d)	color correlation(144 d)	wavelet moments(1770 d)	PRI-CoLBP(216 d)	PRI-CoLBP(1180 d)
View 3	pyramid HOG(252 d)	edge direction(73 d)	CENTRIST(254 d)	CENTRIST(1240 d)	CENTRIST(1240 d)
View 4	SIFT(2000 d)	wavelet texture(128 d)	HOG(1984 d)	VGG19(4096 d)	-
View 5	color SIFT(2000 d)	color moment(225 d)	GIST(512 d)	-	-
View 6	SURF(2000 d)	-	LBP(928 d)	-	-

TABLE II

CLASSIFICATION RESULTS OVER DIFFERENT PERCENTS OF LABELED SAMPLES ON SMALL DATABASES

	HW			MSRC-v1			Caltech101-20			ORL			Yale		
	10%	20%	30%	10%	20%	30%	10%	20%	30%	10%	20%	30%	10%	20%	30%
AMMSS	0.9733	0.975	0.9756	0.8354	0.8691	0.8956	0.6834	0.7084	0.742	0.8583	0.8938	0.9071	0.5464	0.7481	0.8917
AMGL	0.9065	0.9345	0.9511	0.8039	0.8515	0.8697	0.7696	0.8252	0.8546	0.8722	0.9156	0.9179	0.5133	0.8444	0.8417
MMCL	0.8965	0.9405	0.9705	0.7524	0.7762	0.8714	-	-	-	0.8475	0.9075	0.9275	0.5394	0.8424	0.8909
MLAN	0.9759	0.9788	0.9789	0.8258	0.8783	0.8913	0.847	0.8602	0.882	0.8056	0.85	0.8464	0.62	0.7556	0.8083
P_c_MVSR	0.9783	0.9844	0.9871	0.8783	0.9464	0.9524	0.886	0.9291	0.9321	0.9389	0.9656	0.9821	0.8867	0.9778	0.9833
P_f_MVSR	0.9789	0.9856	0.9871	0.8942	0.9583	0.9592	0.8902	0.9334	0.9291	0.9611	0.9938	0.9857	0.94	0.9778	0.9833

1. Red number indicates the best performance; blue number denotes the second best performance.
2. MMCL run out of memory when being applied to Caltech101-20 on a server with 128 GB memory for constructing V matrices of sizes $d_v * d_v * n$ simultaneously: $(48^2 + 40^2 + 254^2 + 1984^2 + 512^2 + 928^2) * 2386 * 8./(2^{30}) = 91.16$ GB.

TABLE III

CLASSIFICATION RESULTS OVER DIFFERENT PERCENTS OF LABELED SAMPLES ON LARGE DATABASES

	AwA			NUS-WIDE			Caltech101			MITIndoor-67			Scene-15		
	10%	20%	30%	10%	20%	30%	10%	20%	30%	10%	20%	30%	10%	20%	30%
AMMSS	0.0422	0.0983	0.1281	0.2047	0.2371	0.2605	0.4274	0.4661	0.5073	0.2764	0.4388	0.5356	0.3639	0.5117	0.5717
AMGL	0.0653	0.0819	0.0985	0.1603	0.2032	0.23	0.4393	0.5374	0.5743	0.2913	0.3496	0.4072	0.6336	0.6785	0.7005
MMCL	-	-	-	0.2413	0.3583	0.461	-	-	-	-	-	-	-	-	-
MLAN	0.0807	0.1019	0.1212	0.2335	0.2679	0.3011	0.5557	0.6087	0.6279	0.515	0.5514	0.5804	0.6846	0.711	0.7281
P_c_MVSR	0.2515	0.2512	0.25	0.4989	0.4983	0.5071	0.6859	0.7322	0.7594	0.6815	0.6966	0.7134	0.7489	0.7856	0.7984
P_f_MVSR	0.2035	0.1924	0.2146	0.5037	0.5075	0.501	0.6372	0.6746	0.6892	0.6438	0.6672	0.6529	0.7437	0.7822	0.793

MMCL run out of memory when being applied to AwA, Caltech101, MITIndoor-67, and Scene-15.

comparable on HW, and they have much better performance than the other two methods; P_c_MVSR and P_f_MVSR outperform the best peer algorithms by the margins of 5%–10% on MSRV-v1, ORL, and Caltech101-20; and they also achieve more than 10% improvements over the second-best algorithms on the other six databases. Please also note that the improvements of P_c_MVSR and P_f_MVSR are more obvious and clearly significant on the large and complicated databases, e.g., AwA and NUS-WIDE, showing a potential ability to process challenging scenarios. The superiorities of the proposed models own much to the efficient usage of the available data labels in discovering the underlying affinity of the data set.

Comparing the performance of P_f_MVSR to that of P_c_MVSR, we can see that using fine-grained semantic similarities can improve the classification accuracy when the data set is small and/or the available data labels are limited. The performance of P_f_MVSR decreases when there are plenty of labeled data. The reason is that with the increase of data labels, directly measuring the fine-grained similarities from view features may not be distinguishable. Several well-designed semantic similarity measures may overcome this limitation. Our future work will investigate this.

B. P-MVSR for Weakly Supervised Clustering

In terms of weakly supervised clustering, region-based image segmentation is chosen as an application example, in which images are preprocessed to generate superpixels, and the superpixels are clustered into several segments. The reasons to employ this application example lie in two folds: 1) the superpixels of natural images usually lie in a low-rank

subspace [3], [43] and 2) the domain cues can be easily observed in image segmentation.

1) Experimental Settings:

a) Setups for P-MVSR: Following the Gestalt principle of perceptual organization [21], we adopt two priors to regularize the self-representation tensor within the P-MVSR framework.

1) *Adjacent Prior:* Superpixels that are adjacent within certain layers have high cluster preference, and we empirically set the number of adjacent layers to 4 in all experiments.
2) *Spatial Prior:* The cluster preference is calculated according to the spatial distance of two superpixels.

These two priors result in the settings of P_a_MVSR and P_s_MVSR, respectively. They are calculated from domain cues rather than specific view features and, thus, can be shared among all views, i.e., $P^{(1)} = \cdots = P^{(V)}$. Mathematically, the abovementioned two priors are defined as

$$P_a(i, j) = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are adjacent} \\ \tau, & \text{otherwise} \end{cases} \tag{19}$$

where $0 < \tau < 1$, and

$$P_s(i, j) = \begin{cases} \exp\left(-\dfrac{\text{dist}(i, j)}{\sigma_s^2}\right), & \text{if } i \text{ and } j \text{ are adjacent} \\ \tau, & \text{otherwise} \end{cases} \tag{20}$$

where $\text{dist}(i, j)$ measures the normalized feature distance between i and j, σ_s is fixed to 1 empirically, and $0 < \tau < 1$. Afterward, we set $P_s(P_s < \tau) = \tau$ to avoid artifacts induced by stratification. Once the affinity matrix A is formed, we use the spectral clustering algorithm to obtain the final clustering result.

b) Databases and competing algorithms: Four segmentation databases are used in the experiments, namely, Berkeley segmentation data set 500 (BSDS500), the segmentation subset of PASCAL visual object classes 2007 (VOC2007), and Weizmann segmentation data sets with one and two objects (WSD 1obj and WSD 2obj). The corresponding sample numbers are 500, 632, 100, and 100. In addition, eight state-of-the-art multiview clustering algorithms are chosen for comparison, including MLAP [3], DiMSC [5], LT-MSC [6], AMGL [8], ECMSC [9], MLAN [10], 3rdT-MSC [11], and t-SVD-MSC [12].

c) Superpixels and multiview features: Since superpixels generated by different algorithms may capture different visual patterns, we adopt two typical superpixel segmentation methods from different categories, i.e., FH [44] with parameters [0.8, 200, 100] and SLIC [45] with 100 superpixels per image, to show the generalization ability of our P-MVSR model. FH tends to produce irregular superpixels, while SLIC always generates superpixels with similar sizes. To take advantage of multiview features in capturing human perception, three kinds of image features (color, texture, and shape) are extracted, i.e., RGB histogram ($8 * 8 * 8 = 512$ d), uniform color LBP feature (177 d) [46], and the Bag-of-Visual-Words (BoW) feature by calculating SIFT [47] at each pixel and then applying the k-means algorithm to generate 100 visual words (100 d).

2) Quantitative Evaluation: Following the literature [3], four standard evaluation metrics are used for comparison: the probabilistic rand index (PRI), the variation of information (VoI), the global consistency error (GCE), and the boundary displacement error (BDE). For PRI, the higher, the better performance; for the other three metrics, the lower, the better.

Considering BSDS500 and VOC2007, since the number of ground-truth segments is unknown, we set the segmentation scale (namely, the number of clusters) to $2, 3, \ldots, 40$, respectively. Then, the segmentation results are compared at an optimal data set scale (ODS) for the entire data set and an optimal image scale (OIS), following [48]. The quantitative results of all competing algorithms on BSDS500 and VOC2007 are reported in Table IV. Overall, our P_a_MVSR obtains consistently best performance in most cases, followed by P_s_MVSR. As images in BSDS500 and VOC2007 usually have complex structures and multiple objects, the proposed priors have a natural advantage in obtaining spatially compact segments and, thus, can eliminate the false positive clustering results.

Since images in WSD 1obj and WSD 2obj capture only one and two objects, we fix their cluster numbers as two and three (object segments plus one background segment), respectively. The ODS values are, therefore, the same as the OIS ones, and we only report one result for clarity. Table V presents the quantitative segmentation results of different methods on the WSD data sets. P_a_MVSR and P_s_MVSR obtain relatively better performance. It can be observed that compared with the results on BSDS500 and VOC2007, the advantage of our P_a_MVSR and P_s_MVSR on WSD is not so significant. This is because images in the WSD databases contain only one or two object segments, and accordingly, the background segment is likely to spread over all image borders. Thus, the proposed (four-layer) adjacent or spatially compact priors become less effective. By contrast, images in BSDS500 and VOC2007 contain multiple objects, and thus, the priors show superiority. This observation coincides with human perception, and more importantly, it confirms the necessity of an appropriate prior according to specific domain knowledge.

3) Visual Comparison: Examples of the segmented images are shown in Fig. 3, in which superpixels within one segment are labeled with the same color. Specifically, we assign each clustered segment to the label of the most overlapped ground-truth segment. The visual results demonstrate the consistency between the clustered segments and the ground truth. P_a_MVSR and P_s_MVSR obtain better visual results compared with their competitors. More importantly, the positive influence of prior knowledge can be observed. For instance, in P_a_MVSR and P_s_MVSR, the superpixels on the branches of the tree (Image 1), the flowers behind the butterfly (Image 2), and the face and neck of the lady (Image 6) are better grouped compared with peer algorithms. This owns much to the high cluster preference of spatially adjacent/compact superpixels using the proposed priors.

C. Model Analysis

1) Parameter Selection: There are two parameters λ and τ that should be tuned for P-MVSR. For semisupervised classification, we empirically set the parameter λ in the

TABLE IV

QUANTITATIVE SEGMENTATION RESULTS ON BSDS500 AND VOC2007

	BSDS500 dataset															
	FH superpixel								**SLIC superpixel**							
	PRI ↑		VoI ↓		GCE ↓		BDE ↓		PRI ↑		VoI ↓		GCE ↓		BDE ↓	
	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS
MLAP	0.7688	0.8142	2.2595	2.0041	0.1538	0.1364	13.2855	10.7898	0.7272	0.7774	2.0579	2.2317	0.1571	0.1473	14.818	12.1608
DiMSC	0.7541	0.789	2.3357	2.1006	0.1481	0.1335	14.6877	12.6122	0.7092	0.7458	2.2235	2.3978	0.1891	0.1781	18.3028	15.2243
LT-MSC	0.7699	0.8169	2.2344	1.9826	0.1496	0.1312	13.1805	10.7273	0.7311	0.7809	2.0398	2.232	0.1551	0.1421	14.613	12.1204
AMGL	0.7652	0.812	2.244	2.0354	0.1442	0.1285	13.359	11.864	0.7386	0.7732	2.1325	2.2473	0.1664	0.1526	14.216	13.54
ECMSC	0.758	0.7777	2.3799	2.1949	0.1697	0.1472	14.4466	13.4589	0.6932	0.7117	2.4523	2.5233	0.21	0.2049	20.263	19.4771
MLAN	0.7697	0.816	2.2575	2.0472	0.1263	0.1167	12.8813	10.5753	0.7278	0.7773	2.0927	2.2503	0.1402	0.1322	13.9612	11.7039
3rdT-MSC	0.7622	0.8112	2.241	2.0573	0.1493	0.133	13.285	11.044	0.7321	0.7794	2.0252	2.2116	0.1421	0.1309	14.732	12.208
t-SVD-MSC	0.7721	0.8178	2.2204	1.9691	0.1481	0.1302	13.1484	10.6587	0.7363	0.784	1.9856	2.1868	0.1399	0.1287	14.4242	11.9035
P_a_MVSR	0.7922	0.8339	2.0896	1.8139	0.1262	0.1085	12.6249	9.9104	0.7556	0.8151	1.8929	2.1922	0.131	0.1125	13.6725	10.4663
P_s_MVSR	0.7762	0.8214	2.0986	1.8221	0.1443	0.1198	13.1257	10.5546	0.7544	0.8045	1.9496	2.248	0.1489	0.1231	13.713	11.0676
	VOC2007 dataset															
	FH superpixel								**SLIC superpixel**							
	PRI ↑		VoI ↓		GCE ↓		BDE ↓		PRI ↑		VoI ↓		GCE ↓		BDE ↓	
	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS
MLAP	0.5839	0.6441	1.6856	1.659	0.19	0.1629	26.1116	21.5948	0.5605	0.6069	1.7481	1.7274	0.248	0.2105	27.7894	23.7012
DiMSC	0.5679	0.6358	1.6949	1.6962	0.2135	0.18	26.3435	22.1248	0.5585	0.611	1.727	1.7102	0.2589	0.2138	29.1868	24.6845
LT-MSC	0.5809	0.6472	1.6813	1.6516	0.1871	0.1586	25.967	21.6521	0.5669	0.6163	1.7223	1.6964	0.236	0.2008	27.6794	23.3177
AMGL	0.5473	0.6185	1.7018	1.7728	0.1982	0.1835	27.8274	22.6147	0.5562	0.6065	1.902	1.8264	0.24431	0.2208	28.2352	24.56
ECMSC	0.5571	0.609	1.7189	1.8044	0.1916	0.1937	26.9058	22.6427	0.5492	0.5985	1.7063	1.6537	0.255	0.2364	28.3218	25.5448
MLAN	0.5666	0.6389	1.7001	1.6536	0.1754	0.1441	25.8629	20.6554	0.5675	0.6277	1.6924	1.6493	0.2371	0.1967	26.995	23.0864
3rdT-MSC	0.5629	0.6392	1.7422	1.6883	0.1828	0.1534	26.0732	21.092	0.5702	0.6297	1.7102	1.6932	0.2267	0.1874	27.01	23.1661
t-SVD-MSC	0.5693	0.6483	1.7207	1.6749	0.1744	0.147	25.7832	20.9522	0.5692	0.6269	1.705	1.692	0.2212	0.1828	27.0032	23.0417
P_a_MVSR	0.5893	0.652	1.6807	1.6249	0.1744	0.143	25.0788	19.9522	0.5744	0.6301	1.6863	1.6448	0.2135	0.1797	26.8476	22.1326
P_s_MVSR	0.5877	0.6559	1.665	1.614	0.1932	0.1508	25.4342	20.9417	0.5723	0.629	1.6931	1.652	0.2142	0.1809	26.9573	22.298

ODS: optimal dataset scale; OIS: optimal image scale.

TABLE V

QUANTITATIVE SEGMENTATION RESULTS ON WSD DATABASES

	WSD 1obj dataset								**WSD 2obj dataset**							
	FH superpixel				**SLIC superpixel**				**FH superpixel**				**SLIC superpixel**			
	PRI ↑	VoI ↓	GCE ↓	BDE ↓	PRI ↑	VoI ↓	GCE ↓	BDE ↓	PRI ↑	VoI ↓	GCE ↓	BDE ↓	PRI ↑	VoI ↓	GCE ↓	BDE ↓
MLAP	0.7071	1.0708	0.193	18.6651	0.6862	1.1072	0.1889	20.1497	0.7461	1.0712	0.1371	12.0187	0.5821	1.5437	0.1694	15.846
DiMSC	0.6893	1.0998	0.19	21.5995	0.6247	1.294	0.2326	27.1605	0.7277	1.0962	0.1392	14.8961	0.5902	1.5484	0.1848	17.4758
LT-MSC	0.7189	1.0235	0.1785	18.3171	0.5965	1.4895	0.1628	16.765	0.7458	1.0634	0.1344	11.9898	0.5965	1.4895	0.1628	15.765
AMGL	0.7317	0.9767	0.1662	19.9213	0.6451	1.1116	0.1956	24.7861	0.7446	1.0798	0.1436	13.8742	0.6017	1.4884	0.1712	18.2203
ECMSC	0.675	1.1883	0.2043	22.9323	0.5964	1.4557	0.2321	30.6716	0.7244	1.1046	0.1609	18.1	0.5507	1.7047	0.205	24.3847
MLAN	0.7296	0.9606	0.1677	15.5204	0.6016	1.4255	0.1587	17.6974	0.7456	1.0162	0.122	11.5636	0.6016	1.4255	0.1487	17.6974
3rdT-MSC	0.7301	0.9822	0.1795	19.244	0.6214	1.4352	0.1801	17.8922	0.7478	1.0515	0.1241	14.775	0.6122	1.438	0.1628	17.1225
t-SVD-MSC	0.7291	0.9873	0.1625	20.2782	0.6151	1.4296	0.1604	16.0054	0.7431	1.0214	0.1233	14.0257	0.6151	1.4296	0.1604	16.0054
P_a_MVSR	0.7396	0.9795	0.1779	16.0238	0.6891	1.1055	0.1988	20.8693	0.7521	1.0129	0.1264	11.2104	0.6311	1.3546	0.1512	14.4838
P_s_MVSR	0.7322	0.9845	0.1788	15.97	0.6635	1.1124	0.1887	19.2445	0.7535	1.0408	0.1204	11.0922	0.6288	1.3652	0.1526	14.6038

range [0.01, 0.03, 0.05, 0.07, 0.1, 0.3, 0.5, 0.7], and τ is selected from $(0, 1)$ with step 0.1, and P_c_MVSR is chosen for analysis. As two illustrative examples, the classification performance of P_c_MVSR over parameters λ and τ is reported on the Caltech101-20 and Yale data sets with 10% labeled samples, respectively. We can observe from Fig. 4 that although the parameters λ and τ play important roles on the classification accuracy, the results of P_c_MVSR are still stable over local ranges of parameters. In terms of Caltech101-20, the recommended values of λ and τ locate within the ranges [0,01, 0.05] and [0.4, 0.9], respectively; for Yale, we suggest to choose λ from [0.3, 0.7] and τ from the range [0.2, 0.8]. The optimal values of λ on Yale are larger than those on Caltech101-20. This is because the Yale data set contains complicated variations, such as illumination changes and occlusions. By setting a relatively large value of λ, our P_c_MVSR model will greatly penalize the error term to achieve better performance. This observation is helpful to tune the parameters of P_c_MVSR.

In the application of weakly supervised clustering, λ is empirically selected from [1.1, 1.3, . . . , 2.3], and τ is chosen from [0, 0.2, 0.4, 0.8] for P_a_MVSR and P_s_MVSR. One example of the performance of P_a_MVSR over model parameters on BSDS500 is given in Fig. 5. As can been seen, for image segmentation, the performance of P-MVSR is relatively insensitive to the choice of λ in the range [1.1, 1.3, . . . , 2.3], while τ is recommended to set to 0.2.

2) Empirical Convergence: To examine the convergence of P-MVSR in real scenarios, we define three residuals (i.e., the stopping criterion in lines 10–12 in Algorithm 1)

$$\text{Reconstruction error}: \ r1 = \max_{v=1}^{V}\{\|X^{(v)} - X^{(v)}D^{(v)} - E^{(v)}\|_\infty\}$$

$$\text{Match error 1}: \ r2 = \max_{v=1}^{V}\{\|D^{(v)} - P^{(v)} \odot Z^{(v)}\|_\infty\}$$

$$\text{Match error 2}: \ r3 = \|\mathcal{Z} - \mathcal{G}\|_\infty.$$

Then, we plot the empirical convergence curves of P_c_MVSR on one large database Caltech101 and one small

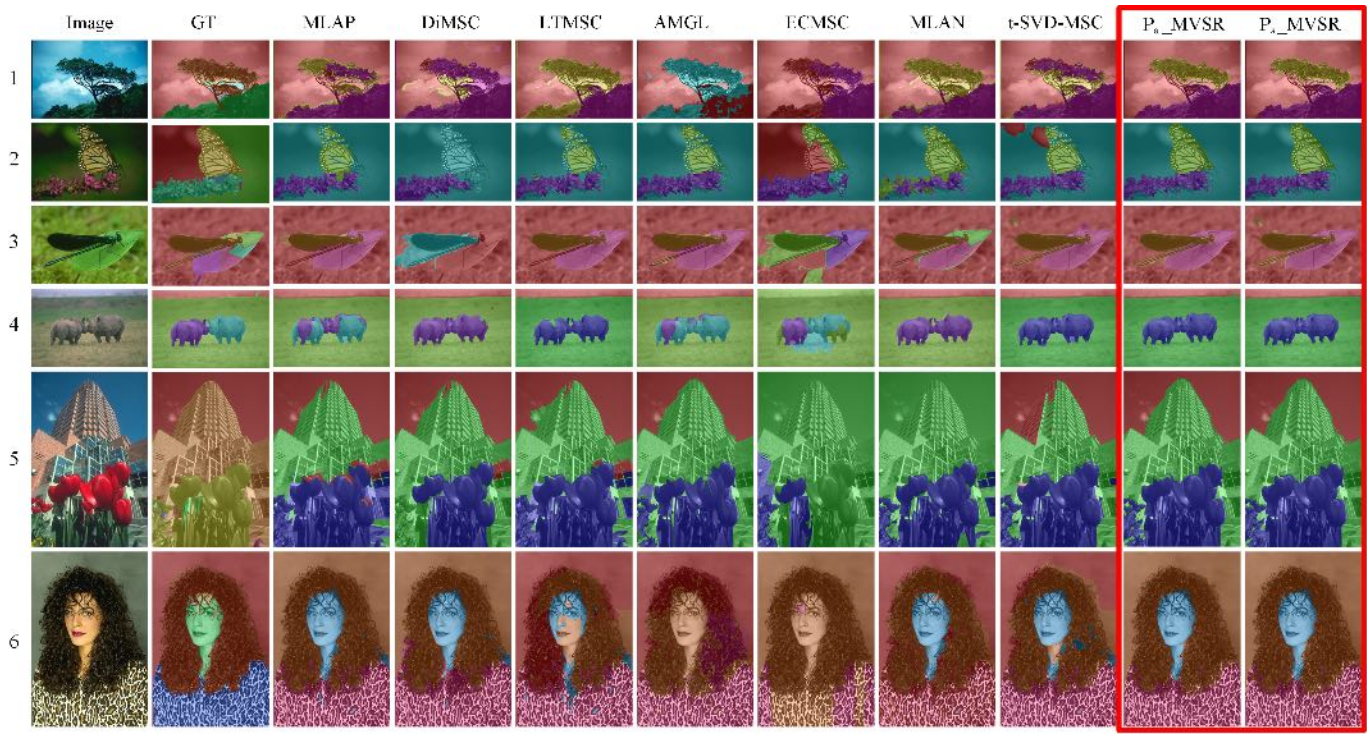

Fig. 3. Visual comparisons of different algorithms on BSDS500 with five clustered segments per image using FH superpixels.

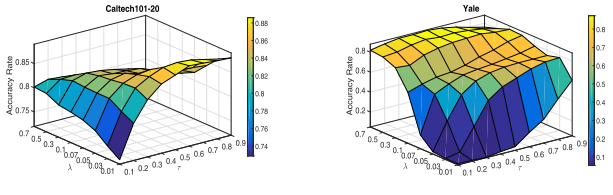

Fig. 4. Sensitivity analysis of parameters in P_c_MVSR on Caltech101-20 and Yale data sets with 10% labeled samples.

database Yale as two illustrative examples in Fig. 6. The three residuals yield stable values after 50 iterations, showing that P_c_MVSR converges well in real applications. The average convergence curves of P_c_MVSR and competing algorithms are plotted in Fig. 7. As can be seen, AMMSS and AMGL generally converge within ten iterations, while MLAN and P_c_MVSR obtain stable solutions after 30–50 iterations. MMCL converges within 20 iterations on the Yale database and runs out of memory on the Caltech101 database. It can be observed that P_c_MVSR costs most iterations before convergence. This is because the optimization function of P_c_MVSR has relatively more constraints compared with the competitors.

3) Runtime Comparison: Theoretically, the computation complexity of P-MVSR is $\mathcal{O}(iteVn^3)$. That is, the runtime of P-MVSR relates to the number of samples, views, and real iteration numbers. To compare the empirical runtime, all algorithms are implemented using MATLAB and tested on a workstation with Intel Core i9-9920X processor and Ubuntu system. The empirical runtime is reported in Tables VI and VIII, and the theoretical complexity is presented for reference. For semisupervised classification, AMMSS is the most efficient algorithm, and MMCL is

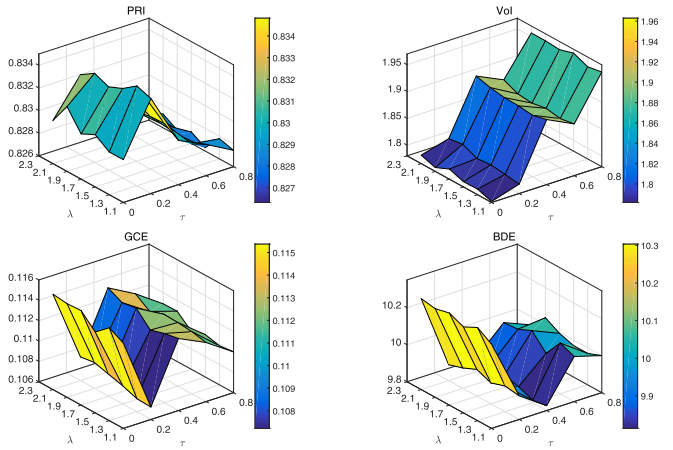

Fig. 5. Sensitivity analysis of parameters in P_a_MVSR on BSDS500.

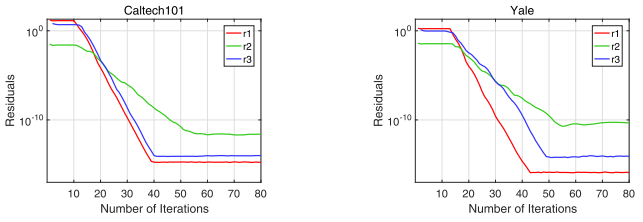

Fig. 6. Empirical convergence curves of P_c_MVSR on Caltech101 and Yale databases.

the most time-consuming one. The main cost of P-MVSR lies in the operations in the matrix multiplication and the t-SVD-TNN minimization subproblem. To solve computation bottleneck, fast tensor learning methods are needed.

TABLE VI

EMPIRICAL RUNTIME (SECONDS) AND COMPUTATION COMPLEXITY OF DIFFERENT ALGORITHMS ON SEMISUPERVISED CLASSIFICATION

	HW	MSRC-v1	Caltech101-20	ORL	Yale	AwA	NUE-WIDE	Caltech101	MITIndoor-67	Scene-15	Complexity
AMMSS	8.54	0.08	22.96	1.08	0.34	398.67	19.06	277.63	166.56	142.68	$\mathcal{O}(iteVn^3)$
AMGL	18.75	0.19	48.62	2.53	0.97	985.67	46.91	557.92	367.92	331.42	$\mathcal{O}(iteVn^3)$
MMCL	1123.57	9.32	2672.1	88.95	47.39	33167.3	2129.62	17889.4	12365.8	8869.52	$\mathcal{O}(iteVn^3)$
MLAN	47.83	0.39	89.76	4.97	2.21	2012.86	108.76	1224.83	872.75	642.27	$\mathcal{O}(iten^3)$
P-MVSR	199.82	1.67	451.74	15.22	8.26	6227.68	406.234	4789.65	3785.78	1878.59	$\mathcal{O}(iteVn^3)$

TABLE VII

ABLATION STUDY ON THE PRIOR KNOWLEDGE AND T-SVD-TNN MODULES FOR SEMISUPERVISED CLASSIFICATION

	HW	MSRC-v1	Caltech101-20	ORL	Yale	AwA	NUE-WIDE	Caltech101	MITIndoor-67	Scene-15
no Prior	0.965	0.836	0.8632	0.8944	0.8533	0.1726	0.3224	0.3321	0.3427	0.5972
concatenation	0.9606	0.8413	0.8288	0.9021	0.8056	0.1022	0.3663	0.4127	0.3725	0.5763
u-TNN	0.9661	0.8677	0.8688	0.9139	0.825	0.1642	0.3852	0.4664	0.4333	0.6228
P_c_MVSR	0.9783	0.8783	0.886	0.9389	0.8867	0.2515	0.4989	0.6859	0.6815	0.7489
P_f_MVSR	0.9789	0.8942	0.8902	0.9611	0.94	0.2035	0.5037	0.6372	0.6438	0.7437

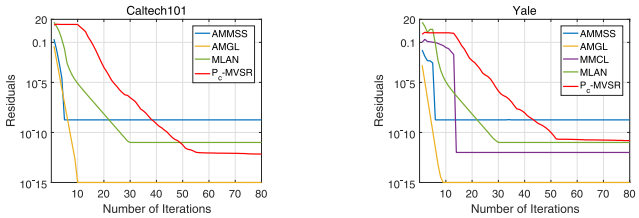

Fig. 7. Average empirical convergence curves of different algorithms on Caltech101 and Yale databases.

TABLE VIII

AVERAGE EMPIRICAL RUNTIME (SECONDS) AND COMPUTATION COMPLEXITY OF DIFFERENT ALGORITHMS ON IMAGE SEGMENTATION

	FH			SLIC			Complexity
	BSDS	VOC	WSD	BSDS	VOC	WSD	
MLAP	3.9227	3.8924	1.5998	0.5151	0.515	0.5311	$\mathcal{O}(iteVn^3)$
DiMSC	0.7992	0.8286	0.307	0.1059	0.1366	0.1232	$\mathcal{O}(iteVn^3)$
LT-MSC	2.1872	2.1085	0.6827	0.3086	0.2932	0.3184	$\mathcal{O}(iteVn^3)$
AMGL	0.6947	0.7835	0.251	0.0919	0.0993	0.0944	$\mathcal{O}(iteVn^3)$
ECMSC	6.8223	6.7076	2.6611	0.8429	0.8384	0.8456	$\mathcal{O}(iteVn^3)$
MLAN	0.3257	0.2188	0.0198	0.1296	0.1519	0.1342	$\mathcal{O}(iten^3)$
3rdT-MSC	1.9087	1.8278	0.7607	0.28	0.2582	0.2951	$\mathcal{O}(iteVn^3)$
t-SVD-MSC	1.6387	1.7443	0.6783	0.2579	0.2586	0.2625	$\mathcal{O}(iteVn^3)$
P-MVSR	1.6452	1.789	0.7743	0.2594	0.2705	0.2741	$\mathcal{O}(iteVn^3)$

For example, Wu *et al.* [49] proposed an essential tensor learning method for multiview clustering to avoid matrix multiplication. To further reduce the computation cost of low-rank tensor minimization, we can resort to tensor factorization [50] instead of t-SVD-TNN.

When being applied to image segmentation, the average execution time over the whole database is recorded for each algorithm. Since the runtime is closely related to the number of samples, different superpixel segmentation methods will have a direct influence on the empirical runtime. Specifically, FH generates 200–450 superpixels on BSDS200/VOC2007 and around 150–300 superpixels on WSD databases. Meanwhile, SLIC consistently produces 100 superpixels on all databases. As a result, the runtime of competing algorithms varies dramatically when using the FH superpixel, and it is relatively stable when the SLIC superpixel is adopted.

4) Ablation Study:

a) Benefits from prior knowledge and t-SVD-TNN: To start with, we present an ablation study to show the influence of the prior knowledge module and the effectiveness of t-SVD-TNN. Specifically, when no prior knowledge is imposed, the method in [12] is subsumed into the P-MVSR framework; if the t-SVD-TNN module is ablated, we use two schemes to deal with the cross-view relationship: 1) concatenating all view features into a large feature matrix and 2) stacking the representation coefficient matrices and seeking their consensus via u-TNN. These two strategies are denoted by "concatenation" and "u-TNN" in Tables VII and IX. For semisupervised classification, we conduct experiments using 10% labeled samples on all databases. As to image segmentation, the results on the BSDS500 are reported. In both tests, the performance enhancement of P-MVSR models benefits from the joint consideration of the prior knowledge and t-SVD-TNN, showing the superiority of the proposed P-MVSR framework.

To further provide an intuitive illustration of the effectiveness of the two modules, we visualize the discovered affinity of samples using every single view and all views without and with prior knowledge, respectively. Following the method in [51], we adopt t-distributed stochastic neighbor embedding (t-SNE) [52] to show the data structure of MSRC-v1 in Fig. 8. We can see that: 1) the underlying class structures cannot be well discovered using only a single view [see Fig. 8(a)–(d)]; 2) integrating the information from multiview features via t-SVD-TNN, the affinity matrices can well preserve the local structures of different classes [see Fig. 8(e) and (f)]; and 3) compared with Fig. 8(e), a more compact and accurate class structure is observed in Fig. 8(f), demonstrating the advantage of using prior knowledge in learning the affinity matrix.

b) Multiview features integration: Given features extracted from multiple sources, P-MVSR engages in designing an integration scheme to obtain enhanced performance. Nowadays, the development of deep learning models makes it possible to extract enhanced features by taking advantage

TABLE IX

ABLATION STUDY ON THE PRIOR KNOWLEDGE AND T-SVD-TNN MODULES FOR CLUSTERING

	PRI ↑		VoI ↓		GCE ↓		BDE ↓	
	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS
no Prior	0.7363	0.784	1.9856	2.1868	0.1399	0.1287	14.4242	11.9035
concatenation	0.7375	0.7902	1.9722	2.2089	0.1399	0.1224	14.082	11.5422
u-TNN	0.7445	0.7951	1.9652	2.2012	0.1376	0.1184	13.823	11.224
P_a_MVSR	0.7556	0.8151	1.8929	2.1922	0.131	0.1125	13.6725	10.4663
P_s_MVSR	0.7544	0.8045	1.9496	2.248	0.1489	0.1231	13.713	11.0676

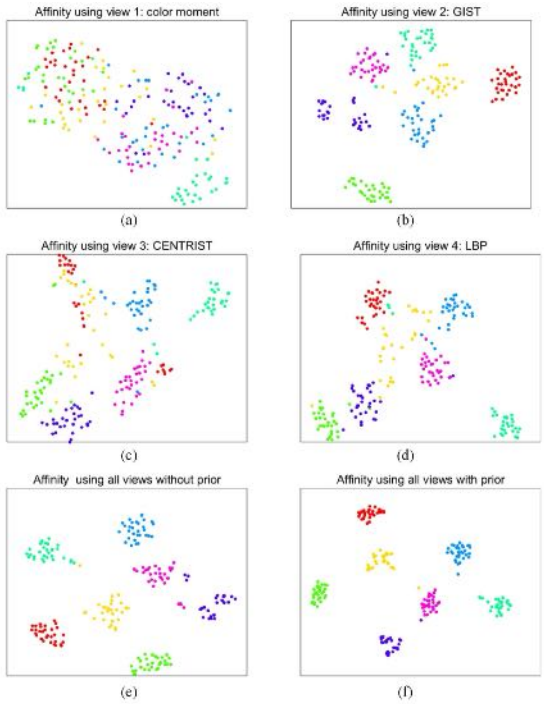

Fig. 8. Visualization of the affinity matrices of MSRC-v1 using each single view and all views without and with prior knowledge via t-SNE.

TABLE X

CLASSIFICATION RESULTS OVER DIFFERENT KINDS OF FEATURES ON THE CALTECH101-20 DATABASE

	Hand-crafted feature			Deep feature		
	Best-view	P-MVSR	Gain	Best-view	P-MVSR	Gain
10%	0.6769	0.886	30.89%	0.8962	0.9847	9.88%
20%	0.7611	0.9291	22.07%	0.9364	0.9953	6.29%
30%	0.7974	0.9321	16.89%	0.944	0.9917	5.05%

of the network structure and plenty of samples [53], [54]. A question naturally arises as follows. Will the proposed P-MVSR model work when features from single views are already well-performed? To answer this question, we extract deep features from three representative networks,[1] i.e., InceptionV3 (mixed10 layer), ResNet-50 (activation-49 layer), and VGG19 (fc2 layer). All deep models are trained on ImageNet. Since these models have different structures, we consider the extracted features as three heterogeneous features. We use the Caltech101-20 database as an illustrative example and record the results from the best-performing single view,[2] P-MVSR, and their performance gains, respectively, in Table X. Although a single deep feature shows promising performance, considerable improvements are witnessed by integrating the features from multiple deep models. This validates the general assumption on integrating multiview features for performance enhancement.

[1]https://keras.io/zh/applications/

[2]By extending the single-view self-representation model in [32] to the semisupervised scenario.

VI. CONCLUSION

This article proposed a P-MVSR model to take advantage of the prior knowledge in discovering the underlying relationship of samples. Specifically, we considered the prior knowledge, multiview features, and the high-order cross-view correlation simultaneously to learn an accurate self-representation tensor. Taking the valuable prior information as a complement to the multiview features, P-MVSR has shown the superiority compared with existing multiview learning models. Extensive experiments on semisupervised classification and region-based image segmentation have demonstrated the effectiveness and the generalization ability of our P-MVSR model by investigating different settings of prior knowledge.

As P-MVSR generalizes different kinds of side information, e.g., explicit labels, semantic similarities, weak-domain cues, as the prior knowledge, and uses the prior knowledge to guide the learning procedure, it opens up new opportunities of developing powerful multiview self-representation models. In the future, we can investigate the application of P-MVSR in constrained clustering where the task-dependent constraints provide valuable information on the relationship of data. In addition, inspired by the work in [55]–[57], it is worth exploring the fine-grained prior knowledge from the initial coarse priors for performance enhancement.

REFERENCES

[1] J. Zhao, X. Xie, X. Xu, and S. Sun, "Multi-view learning overview: Recent progress and new challenges," *Inf. Fusion*, vol. 38, pp. 43–54, Nov. 2017.

[2] Y. Li, M. Yang, and Z. Zhang, "A survey of multi-view representation learning," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 10, pp. 1863–1883, Oct. 2019.

[3] B. Cheng, G. Liu, J. Wang, Z. Huang, and S. Yan, "Multi-task low-rank affinity pursuit for image segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2439–2446.

[4] X. Cai, F. Nie, H. Huang, and F. Kamangar, "Heterogeneous image feature integration via multi-modal spectral clustering," in *Proc. CVPR*, Jun. 2011, pp. 1977–1984.

[5] X. Cao, C. Zhang, H. Fu, S. Liu, and H. Zhang, "Diversity-induced multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 586–594.

[6] C. Zhang, H. Fu, S. Liu, G. Liu, and X. Cao, "Low-rank tensor constrained multiview subspace clustering," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1582–1590.

[7] H. Gao, F. Nie, X. Li, and H. Huang, "Multi-view subspace clustering," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4238–4246.

[8] Z. Kang, X. Lu, J. Yi, and Z. Xu, "Self-weighted multiple kernel learning for graph-based clustering and semi-supervised classification," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 1881–1887.

[9] X. Wang, X. Guo, Z. Lei, C. Zhang, and S. Z. Li, "Exclusivity-consistency regularized multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 923–931.

[10] F. Nie, G. Cai, and X. Li, "Multi-view clustering and semi-supervised classification with adaptive neighbours," in *Proc. Nat. Conf. Artif. Intell.*, 2017, pp. 2408–2414.

[11] M. Yin, J. Gao, S. Xie, and Y. Guo, "Multiview subspace clustering via tensorial t-product representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 851–864, Mar. 2019.

[12] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang, and Y. Qu, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *Int. J. Comput. Vis.*, vol. 126, no. 11, pp. 1157–1179, Nov. 2018.

[13] Y. Wang, L. Wu, X. Lin, and J. Gao, "Multiview spectral clustering via structured low-rank matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4833–4843, Oct. 2018.

[14] Q. Yin, S. Wu, and L. Wang, "Multiview clustering via unified and view-specific embeddings learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5541–5553, Nov. 2018.

[15] Y. Chen, X. Xiao, and Y. Zhou, "Jointly learning kernel representation tensor and affinity matrix for multi-view clustering," *IEEE Trans. Multimedia*, early access, doi: 10.1109/TMM.2019.2952984.

[16] X. Cai, F. Nie, W. Cai, and H. Huang, "Heterogeneous image features integration via multi-modal semi-supervised learning model," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1737–1744.

[17] C. Gong, D. Tao, S. J. Maybank, W. Liu, G. Kang, and J. Yang, "Multi-modal curriculum learning for semi-supervised image classification," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3249–3260, Jul. 2016.

[18] Y. Xie, W. Zhang, Y. Qu, L. Dai, and D. Tao, "Hyper-Laplacian regularized multilinear multiview self-representations for clustering and semi-supervised learning," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 572–586, Feb. 2020.

[19] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.

[20] M. Wertheimer, "Laws of organization in perceptual forms," in *A Source Book of Gestalt Psychology*, W. D. Ellis, Ed. Whitelackington, U.K.: Kegan Paul, Trench, Trubner & Company, 1938, pp. 71–88, doi: 10.1037/11496-005.

[21] X. Wang, Y. Tang, S. Masnou, and L. Chen, "A global/local affinity graph for image segmentation," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1399–1411, Apr. 2015.

[22] X. Liu, L. Huang, C. Deng, B. Lang, and D. Tao, "Query-adaptive hash code ranking for large-scale multi-view visual search," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4514–4524, Oct. 2016.

[23] M. Karasuyama and H. Mamitsuka, "Multiple graph label propagation by sparse integration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 12, pp. 1999–2012, Dec. 2013.

[24] Y. Luo, D. Tao, K. Ramamohanarao, C. Xu, and Y. Wen, "Tensor canonical correlation analysis for multi-view dimension reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 11, pp. 3111–3124, Nov. 2015.

[25] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. Conf. Comput. Learn. Theory*, 1998, pp. 92–100.

[26] A. Kumar and H. Daumé, "A co-training approach for multi-view spectral clustering," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 393–400.

[27] A. Kumar, P. Rai, and H. Daume, "Co-regularized multi-view spectral clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 1413–1421.

[28] Y. Chen, S. Wang, F. Zheng, and Y. Cen, "Graph-regularized least squares regression for multi-view subspace clustering," *Knowl.-Based Syst.*, Jan. 2020, Art. no. 105482, doi: 10.1016/j.knosys.2020.105482.

[29] K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan, "Multi-view clustering via canonical correlation analysis," in *Proc. Int. Conf. Mach. Learn.*, 2009, pp. 129–136.

[30] X. Xiao and Y. Zhou, "Two-dimensional quaternion PCA and sparse PCA," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 7, pp. 2028–2042, Jul. 2019.

[31] X. Xiao, Y. Chen, Y.-J. Gong, and Y. Zhou, "Two-dimensional quaternion sparse discriminant analysis," *IEEE Trans. Image Process.*, vol. 29, pp. 2271–2286, Oct. 2019, doi: 10.1109/TIP.2019.2947775.

[32] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 663–670.

[33] Y. Chen, X. Xiao, and Y. Zhou, "Low-rank quaternion approximation for color image processing," *IEEE Trans. Image Process.*, vol. 29, pp. 1426–1439, Sep. 2019, doi: 10.1109/TIP.2019.2941319.

[34] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra Appl.*, vol. 435, no. 3, pp. 641–658, Aug. 2011.

[35] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5249–5257.

[36] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2790–2797.

[37] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-SVD," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3842–3849.

[38] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 612–620.

[39] J. Huang, F. Nie, H. Huang, and C. Ding, "Robust manifold nonnegative matrix factorization," *ACM Trans. Knowl. Discovery Data*, vol. 8, no. 3, pp. 1–21, Jun. 2014.

[40] W. Hu, D. Tao, W. Zhang, Y. Xie, and Y. Yang, "The twist tensor nuclear norm for video completion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2961–2973, Dec. 2017.

[41] C. Chen, B. He, Y. Ye, and X. Yuan, "The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent," *Math. Program.*, vol. 155, nos. 1–2, pp. 57–79, Jan. 2016.

[42] Y. Li, F. Nie, H. Huang, and J. Huang, "Large-scale multi-view spectral clustering via bipartite graph," in *Proc. Nat. Conf. Artif. Intell.*, vol. 2015, pp. 2750–2756.

[43] Y. Ma, A. Y. Yang, H. Derksen, and R. Fossum, "Estimation of subspace arrangements with applications in modeling and segmenting mixed data," *SIAM Rev.*, vol. 50, no. 3, pp. 413–458, Jan. 2008.

[44] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.

[45] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[46] C. Zhu, C.-E. Bichot, and L. Chen, "Multi-scale color local binary patterns for visual object classes recognition," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 3065–3068.

[47] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, vol. 2, Sep. 1999, pp. 1150–1157.

[48] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.

[49] J. Wu, Z. Lin, and H. Zha, "Essential tensor learning for multi-view spectral clustering," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 5910–5922, Dec. 2019.

[50] P. Zhou, C. Lu, Z. Lin, and C. Zhang, "Tensor factorization for low-rank tensor completion," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1152–1163, Oct. 2017.

[51] C. Zhang, Q. Hu, H. Fu, P. Zhu, and X. Cao, "Latent multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4279–4287.

[52] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[53] E. Yang, C. Deng, C. Li, W. Liu, J. Li, and D. Tao, "Shared predictive cross-modal deep quantization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5292–5303, Nov. 2018.

[54] C. Deng, E. Yang, T. Liu, and D. Tao, "Two-stream deep hashing with class-specific centers for supervised image search," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, doi: 10.1109/TNNLS.2019.2929068.

[55] X. Liu, L. Huang, C. Deng, J. Lu, and B. Lang, "Multi-view complementary hash tables for nearest neighbor search," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1107–1115.

[56] X. Liu, J. He, and B. Lang, "Multiple feature kernel hashing for large-scale visual search," *Pattern Recognit.*, vol. 47, no. 2, pp. 748–757, Feb. 2014.

[57] C. Deng, E. Yang, T. Liu, J. Li, W. Liu, and D. Tao, "Unsupervised semantic-preserving adversarial hashing for image search," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4032–4044, Aug. 2019.

Xiaolin Xiao received the B.E. degree from Wuhan University, Wuhan, China, in 2013, and the Ph.D. degree from the University of Macau, Macau, China, in 2019.

She is currently a Post-Doctoral Fellow with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. Her research interests include multiview learning and color image processing and understanding.

Yongyong Chen received the B.S. and M.S. degrees from the College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao, China. He is currently pursuing the Ph.D. degree with the Department of Computer and Information Science, University of Macau, Macau, China.

He visited the National Key Lab for Novel Software Technology, Nanjing University, Nanjing, China, as an Exchange Student, in 2017. His research interests include (nonconvex) low-rank and sparse matrix/tensor decomposition models, with applications to image processing, data mining, and computer vision.

Yue-Jiao Gong (Member, IEEE) received the B.S. and Ph.D. degrees in computer science from Sun Yat-sen University, Guangzhou, China, in 2010 and 2014, respectively.

She is currently a Full Professor with the School of Computer Science and Engineering, South China University of Technology, Guangzhou. Her research interests include evolutionary computation, swarm intelligence, and their applications to intelligent transportation and smart city scheduling. She has published over 80 articles, including more than the 30 IEEE TRANSACTIONS, in these areas.

Yicong Zhou (Senior Member, IEEE) received the B.S. degree from Hunan University, Changsha, China, and the M.S. and Ph.D. degrees from Tufts University, Medford, MA, USA, all in electrical engineering.

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 research interests include image processing, computer vision, machine learning, and multimedia security.

Dr. Zhou is a Senior Member of the International Society for Optical Engineering (SPIE). He was a recipient of the Third Price of Macau Natural Science Award in 2014. He is also the Co-Chair of the Technical Committee on Cognitive Computing in the IEEE Systems, Man, and Cybernetics Society. He also serves as an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and four other journals.