Hyper-Laplacian Regularized Multilinear Multi-View Self-Representations for Clustering and Semi-supervised Learning

Yuan Xie, *Member, IEEE,* Wensheng Zhang, Yanyun Qu*, *Member, IEEE,* Longquan Dai, Dacheng Tao, *Fellow, IEEE,*

***Abstract*—In this paper, we address the multi-view nonlinear subspace representation problem. Traditional multi-view subspace learning methods assume that the heterogeneous features of the data usually lie within the union of multiple linear subspaces. However, instead of linear subspaces, data feature actually resides in multiple nonlinear subspaces in many real-world applications, resulting in unsatisfactory clustering performance. To overcome this, we propose a Hyper-Laplacian Regularized Multilinear Multi-View Self-representation model, which is referred to as HLR-M^{2}VS, to jointly learn multiple views correlation and local geometrical structure in a unified tensor space and view-specific self-representation feature spaces, respectively. In unified tensor space, a well-founded tensor low-rank regularization is adopted to impose on the self-representation coefficient tensor to ensure the global consensus among different views. In view-specific feature space, hypergraph induced hyper-Laplacian regularization is utilized to preserve the local geometrical structure embedded in a high-dimensional ambient space. An efficient algorithm is then derived to solve the optimization problem of the established model with theoretical convergence guarantee. Furthermore, the proposed model can be extended to semi-supervised classification without introducing any additional parameter. Extensive experiment of our method is conducted on many challenging datasets, where a clear advance over state-of-the-art multi-view clustering and multi-view semi-supervised classification approaches is achieved.**

***Index Terms*—t-SVD, multilinear, multi-view features, manifold regularization, nonlinear subspace clustering.**

I. Introduction

Multi-view clustering is becoming one of researching hotspots in unsupervised learning currently. In particular, given multiple heterogeneous features of data sampled from a union of subspaces, multi-view subspace clustering aims to partition data into several clusters, so that each cluster corresponds to one subspace. The success of some recently proposed multi-view subspace clustering methods [21], [22], [54] attributes to the use of self-representation, *e.g.,* Sparse Subspace Clustering (SSC) [19] or low-rank representation (LRR) [20]. The most representative approaches among them are [21], [22], which can achieve the state-of-the-art performance.

Inherited from the LRR and SSC, most multi-view subspace clustering methods, including [21], [22], are originally proposed to deal with the data that lies within multiple linear subspaces from multi-view features perspective. However, in real world applications, this assumption might be violated, leading to unsatisfactory results when dealing with the data from *nonlinear subspaces* [1]. Commonly, there are two ways to handle the nonlinear subspaces: 1) Since data points drawn on a nonlinear low-dimensional manifold are usually hidden in a high-dimensional ambient space [2], using kernel-induced mapping to map the data from the original input space to a high-dimensional feature space may have the mapped data resided in multiple linear subspaces [1], [3]; 2) Assuming that the whole data points reside on a nonlinear manifold while the local neighbors are linearly related, one can utilize manifold regularization to preserve the local geometrical structure embedded in a high-dimensional space [4], [5], [9]. In this paper, we will focus on the second way, *i.e.,* utilizing the manifold constraint to figure out the nonlinear problem in multi-view self-representation modeling.

The proposed method is inspired by the tensor multi-view self-representation model (t-SVD-MSC) [22], and the high-order local geometrical regularization, namely hyper-Laplacian regularization [24]. Even with relatively desirable clustering performance, the method [22] only emphasizes *global* consensus among multiple views. Since view-specific *local* geometrical structure is ignored, t-SVD-MSC may fail to discover the discriminative structure of the nonlinear feature spaces of the data [20], [53], which is essential to the actual applications. On the other hand, beyond pairwise connectivity, the hypergraph [24] can capture high-order relationship of the data locality. Due to this merit, its derived hyper-Laplacian regularization has been widely introduced into low-rank representation [9] and sparse coding [5]. However, [9] only considers the single view feature, resulting in the loss of complementary information from multiple heterogeneous feature spaces.

In this paper, by taking advantage of these two works aforementioned, we propose a new Hyper-Laplacian Regularized Multilinear Multi-View Self-representation model, namely HLR-M^{2}VS for short, for clustering task and semi-

*indicates corresponding author.

Y. Xie, and W. Zhang are with the Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China; E-mail: {yuan.xie, wensheng.zhang}@ia.ac.cn

Y. Qu is with School of Information Science and Technology, Xiamen University, Fujian, 361005, China; E-mail: yyqu@xmu.edu.cn

L. Dai is with School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China; E-mail: dailongquan@njust.edu.cn

D. Tao is with the UBTECH Sydney Artificial Intelligence Centre and the School of Information Technologies, the Faculty of Engineering and Information Technologies, the University of Sydney, 6 Cleveland St, Darlington, NSW 2008, Australia; E-mail: dacheng.tao@sydney.edu.au

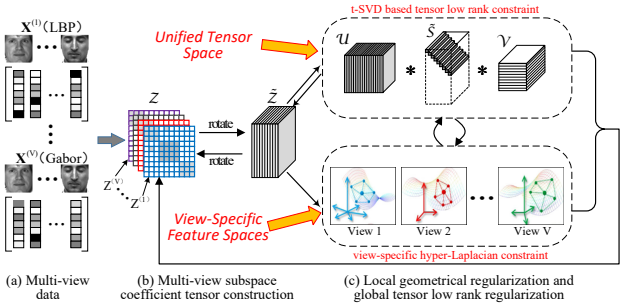


Figure 1: The Flowchart of HLR-M^{2}VS. (a) multi-view features $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(V)}$ of a collection of data points; (b) tensor $\tilde{\boldsymbol{\mathcal{Z}}}$ is formed by firstly stacking all the subspace representations $\{\mathbf{Z}^{(v)}\}_{v=1}^{V}$ to construct tensor $\boldsymbol{\mathcal{Z}}$, and then rotating it to $\tilde{\boldsymbol{\mathcal{Z}}}$; (c) tensor $\tilde{\boldsymbol{\mathcal{Z}}}$ would be optimized alternatively in unified tensor space by using t-SVD based tensor low-rank norm, and in view-specific feature spaces through hypergraph regularized local geometrical constraint.

supervised learning. Fig. 1 shows the pipline of the proposed HLR-M^{2}VS method. Suppose we have a collection of data points with multiple features $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(V)}$ (Fig. 1 (a)), HLR-M^{2}VS firstly obtains the subspace representation matrices $\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(V)}$ (Fig. 1 (b)). Next, those subspace coefficients will be alternatively optimized both in a unified tensor space and view-specific feature spaces (Fig. 1 (c)). In unified tensor space, the rotated subspace coefficient tensor will be optimized by using t-SVD based tensor multi-rank minimization, which can be considered as a *global constraint* to ensure the consensus principle. In view-specific feature spaces, by constructing one hypergraph from view-specific subspace coefficient matrix, the local high-order geometrical structure would be discovered by using the hyper-Laplacian regularization. This can be regarded as *local constraint* in each view independently. Those constraints interact with one another, and the process runs iteratively until convergence is arrived. Only in this way can the intrinsic nonlinear property in original feature subspaces be clustered successfully.

Since the label information of a subset of the samples could be effectively and efficiently propagated to the remaining unlabeled data over a well-constructed graph, the proposed model is highly suitable for being extended to a semi-supervised classification model. In this paper, built upon HLR-M^{2}VS model, we additional propose an effective semi-supervised HLR-M^{2}VS classification model (semi-HLR-M^{2}VS) by utilizing the hyper-graph regularization. It is a parameter-free method that automatically learn a set of weights for all the graph Laplacian. By fusing the optimized view-specific hyper-Laplacian matrices into a final hyper-Laplacian, the optimal weights of edges are obtained by encoding the high-order correlations not only among different views but also within different instances. The established semi-supervised model can capture both the global mixture of subspaces structure (by the tensor low rankness) and the locally high-order linear structure (by the hypergraph Laplacian) of the data.

In a word, for the first time to our knowledge, we design an effective multi-view self-representation model for handling nonlinear feature subspace, which can be confirmed by our excellent clustering and semi-supervised classification performance presented in Section VI. The major contributions can be listed as follows:

- We design a new hyper-Laplacian regularized multilinear multi-view self-representation model, *i.e.*, HLR-M^{2}VS, to simultaneously consider global consensus constraint and local view-specific geometrical regularization for nonlinear subspace learning.
- An efficient optimization procedure is presented to solve the HLR-M^{2}VS optimization problem with empirically fast convergence.
- The HLR-M^{2}VS is further extended to semi-supervised classification, where the fused hyper-Laplacian matrix is learned without using explicit weight parameters.
- Our experimental results show that, on several challenging datasets, the proposed models are effectively better than many the state-of-the-art methods in the multi-view clustering and multi-view semi-supervised classification.

The remainder of this paper is organized as follows. In Section II, we give a brief introduction of related works. Section III is dedicated to some preliminaries on hypergraph and tensor. In Section IV, we describe the proposed model formally, and then present an algorithm to solve it. The semi-supervised extension is presented in Section V. Experimental results are provided in Section VI. Finally, we give the conclusions in Section VII.

II. Related Work

Before introducing the proposed model, in this section, we give a brief review of the recent progress on multi-view subspace learning and manifold regularization, which are the two most related topics to the proposed model.

Supposing that all the views are generated from a latent subspace, the aim of subspace learning approaches is to discover the shared latent subspace first and then conduct clustering. The representative methods in this stream are proposed in [11], [12], which applied canonical correlation analysis (CCA) and kernel CCA to project the multi-view high-dimensional data onto a low-dimensional subspace, respectively. [15] provided a convex reformulation by replacing the squared loss used in CCA, enforcing conditional independence between views. By taking the advantage of deep representation, [14] proposed a deep model which integrates autoencoder-based terms and CCA to capture the deep information from both views. As the CCA based approaches can only handle two views simultaneously, tensor CCA [13] extended them to do with arbitrary number of views.

Alternatively, the success of recent multi-view subspace clustering methods [21], [22], [54] can be attributed to uncover the relationship between samples either by using sparse subspace clustering (SSC) [19] or by using low-rank representation (LRR) [20]). To constrain the subspace coefficient tensor, Zhang *et al.* proposed a method called LTMSC [21] to extend the LRR to multi-view setting by employing the unfolding based tensor norm. LTMSC tried to achieve the tensor low rank in vector space such that the optimality in

the representation might be ignored. On the contrary, in [22], by introducing a new tensor decomposition scheme (t-SVD) [16], [29], Xie *et al.* designed to impose a new type of low-rank tensor constraint on the rotated subspaces coefficient tensor to ensure the consensus among multiple views. While reasonably effective, focusing on global constraint among different views and ignoring the local geometrical structure will lead to performance degeneration in the face of nonlinear subspaces.

Numerous manifold learning approaches have been proposed to preserve the local geometrical structure embedded in a high-dimensional space, for example, the Locally Linear Embedding (LLE) [6], Locality Preserving Projection (LPP) [2], and Neighborhood Preserving Embedding (NPE) [7]. Instead of finding embedding or projection function directly, it is more convenient to resort to manifold regularization [8] to impose local geometrical constraint on new feature space. Specifically, it is reasonable to assume that if two data points are sufficiently close on intrinsic manifold of the data distribution, then the new representations of those two points through a certain basis are also close to each other. This assumption was adopted by [4], which proposed a graph Laplacian regularized sparse coding. In the new coding space, learned sparse representations could reflect the local manifold structures of the original data. Beyond pairwise relationship, Gao *et al.* [5] even introduced hypergraph based high-order geometrical constraint to propose the hyper-Laplacian sparse coding. Similarly, in [9], authors designed a non-negative sparse hyper-Laplacian regularized low-rank representation model, termed NSHLRR, for subspace clustering. Nevertheless, NSHLRR is only a single-view approach which can not take advantage of complementarity among heterogeneous features.

III. Notations and Background

In this section, some notations and preliminaries used throughout this paper will be provided. The bold calligraphy letters (*e.g.*, $\boldsymbol{\mathcal{X}}$), the bold upper case letters (*e.g.*, $\mathbf{X}$), the bold lower case letters (*e.g.*, $\mathbf{x}$), and the lower case letters (*e.g.*, x_{ij}) are used to represent tensors, matrices, vectors, and entries, respectively. The matrix Frobenius norm is defined as $||\mathbf{X}||_F := (\sum_{i,j} |x_{ij}|^2)^{\frac{1}{2}}$. Let $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$ be the SVD of $\mathbf{X}$, and $\sigma_i(\mathbf{X})$ represent the ith largest singular value, its matrix nuclear norm is defined as $||\mathbf{X}||_* := \sum_i \sigma_i(\mathbf{X})$. The corresponding singular-value thresholding (SVT) function with threshold τ is $\mathcal{D}_\tau(\mathbf{X}) = \mathbf{U}\mathbf{\Sigma}_\tau\mathbf{V}^{\mathrm{T}}$, in which $\Sigma_\tau = \mathrm{diag}\{\max(\sigma_i(\mathbf{X}) - \tau, 0)\}$.

An N-way (or N-mode) tensor is a multi-linear structure in $\mathbb{R}^{n_1 \times n_2 \times \ldots \times n_N}$. A **slice** of an tensor is a 2D section defined by fixing all but two indices, and a **fiber** is a 1D section defined by fixing all indices but one [27]. For a 3-way tensor $\boldsymbol{\mathcal{X}}$, we use the Matlab notation $\boldsymbol{\mathcal{X}}(k,:,:)$, $\boldsymbol{\mathcal{X}}(:,k,:)$ and $\boldsymbol{\mathcal{X}}(:,:,k)$ to denote the kth horizontal, lateral and frontal slices, respectively; $\boldsymbol{\mathcal{X}}(:,i,j)$, $\boldsymbol{\mathcal{X}}(i,:,j)$ and $\boldsymbol{\mathcal{X}}(i,j,:)$ to denote the mode-1, mode-2 and mode-3 fibers, and $\boldsymbol{\mathcal{X}}_f = \mathrm{fft}(\boldsymbol{\mathcal{X}}, [\,], 3)$ to denote the Fourier transform along the third dimension. In particular, $\boldsymbol{\mathcal{X}}^{(k)}$ is used to represent $\boldsymbol{\mathcal{X}}(:,:,k)$. Unfolding the tensor $\boldsymbol{\mathcal{X}}$ along the lth mode defined as $\mathrm{unfold}_l(\boldsymbol{\mathcal{X}}) = \mathbf{X}_{(l)} \in \mathbb{R}^{n_l \times \prod_{l' \neq l} n_{l'}}$, which is a matrix whose columns are mode-l fibers [27]. The opposite operation "fold" of the unfolding is defined as $\mathrm{fold}_l(\mathbf{X}_{(l)}) = \boldsymbol{\mathcal{X}}$. The Frobenius norm of $\boldsymbol{\mathcal{X}}$ is $||\boldsymbol{\mathcal{X}}||_F := (\sum_{i,j,k} |x_{ijk}|^2)^{\frac{1}{2}}$, and the ℓ_1 norm of $\boldsymbol{\mathcal{X}}$ is $||\boldsymbol{\mathcal{X}}||_1 := \sum_{i,j,k} |x_{ijk}|$.

The following five block-based operators, *i.e.*, bcirc, bvec, bvfold, bdiag and bdfold [16], are related to t-SVD. For $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the block circulant matrix can be constructed by:

$$\mathrm{bcirc}(\boldsymbol{\mathcal{X}}) := \begin{bmatrix} \boldsymbol{\mathcal{X}}^{(1)} & \boldsymbol{\mathcal{X}}^{(n_3)} & \cdots & \boldsymbol{\mathcal{X}}^{(2)} \\ \boldsymbol{\mathcal{X}}^{(2)} & \boldsymbol{\mathcal{X}}^{(1)} & \cdots & \boldsymbol{\mathcal{X}}^{(3)} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{\mathcal{X}}^{(n_3)} & \boldsymbol{\mathcal{X}}^{(n_3-1)} & \cdots & \boldsymbol{\mathcal{X}}^{(1)} \end{bmatrix}, \quad (1)$$

the block vectorizing and its opposite operation

$$\mathrm{bvec}(\boldsymbol{\mathcal{X}}) := \begin{bmatrix} \boldsymbol{\mathcal{X}}^{(1)} \\ \boldsymbol{\mathcal{X}}^{(2)} \\ \vdots \\ \boldsymbol{\mathcal{X}}^{(n_3)} \end{bmatrix}, \quad \mathrm{bvfold}(\mathrm{bvec}(\boldsymbol{\mathcal{X}})) = \boldsymbol{\mathcal{X}}, \quad (2)$$

and the block diag matrix and its opposite operation

$$\mathrm{bdiag}(\mathcal{X}) := \begin{bmatrix} \boldsymbol{\mathcal{X}}^{(1)} & & \\ & \ddots & \\ & & \boldsymbol{\mathcal{X}}^{(n_3)} \end{bmatrix}, \quad \mathrm{bdfold}(\mathrm{bdiag}(\boldsymbol{\mathcal{X}})) = \boldsymbol{\mathcal{X}}. \quad (3)$$

A. Tensor Singular Value Decomposition (t-SVD)

Before understanding the t-SVD, we need to introduce the follow definitions [16]:

Definition 1 (**t-product**). *Let $\boldsymbol{\mathcal{X}}$ be $n_1 \times n_2 \times n_3$, and $\boldsymbol{\mathcal{Y}}$ be $n_2 \times n_4 \times n_3$. The t-product $\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{Y}}$ is an $n_1 \times n_4 \times n_3$ tensor*

$$\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{Y}} =: \mathrm{bvfold}\{\mathrm{bcirc}(\boldsymbol{\mathcal{X}})\mathrm{bvec}(\boldsymbol{\mathcal{Y}})\}. \quad (4)$$

Definition 2 (**Tensor Transpose**). *Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the transpose tensor $\boldsymbol{\mathcal{X}}^{\mathrm{T}}$ is an $n_2 \times n_1 \times n_3$ tensor obtained by transposing each frontal slice of $\boldsymbol{\mathcal{X}}$ and then reversing the order of the transposed frontal slices 2 through n_3.*

Definition 3 (**Identity Tensor**). *The identity tensor $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ is a tensor whose first frontal slice is the $n_1 \times n_1$ identity matrix and all other frontal slices are zero.*

Definition 4 (**Orthogonal Tensor**). *A tensor $\boldsymbol{\mathcal{Q}} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ is orthogonal if*

$$\boldsymbol{\mathcal{Q}}^{\mathrm{T}} * \boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{Q}} * \boldsymbol{\mathcal{Q}}^{\mathrm{T}} = \boldsymbol{\mathcal{I}}, \quad (5)$$

where $*$ is the t-product.

Definition 5 (**f-diagonal Tensor**). *A tensor is called f-diagonal if each of its frontal slices is diagonal matrix.*

Given the above definitions, we can define the t-SVD of $\boldsymbol{\mathcal{X}}$:

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^{\mathrm{T}}, \quad (6)$$

where $\boldsymbol{\mathcal{U}}$ and $\boldsymbol{\mathcal{V}}$ are orthogonal tensors of size $n_1 \times n_1 \times n_3$ and $n_2 \times n_2 \times n_3$ respectively. $\boldsymbol{\mathcal{S}}$ denotes an f-diagonal tensor with the size of $n_1 \times n_2 \times n_3$, and $*$ denotes the t-product.

B. Tensor Nuclear Norm via t-SVD

The tensor t-SVD can be reformulated as the following [29]:

$$\boldsymbol{\mathcal{X}} = \sum_{i=1}^{\min(n_1,n_2)} \boldsymbol{\mathcal{U}}(:,i,:) * \boldsymbol{\mathcal{S}}(i,i,:) * \boldsymbol{\mathcal{V}}(:,i,:)^{\mathrm{T}}. \quad (7)$$

The tensor multi-rank can be defined as follows [16]–[18] :

Definition 6 (**Tensor multi-rank**). *The multi-rank of* $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *is a vector* $\mathbf{r} \in \mathbb{R}^{n_3 \times 1}$ *with the* i*-th element equal to the rank of the* i*-th frontal slice of* $\boldsymbol{\mathcal{X}}_f$.

Now, the tensor nuclear norm (t-TNN) is given by

$$||\boldsymbol{\mathcal{X}}||_{\circledast} := \sum_{i=1}^{\min(n_1,n_2)} \sum_{k=1}^{n_3} |\boldsymbol{\mathcal{S}}_f(i,i,k)|. \quad (8)$$

Note that the t-TNN is a valid norm, and is also the tightest convex relaxation to ℓ_1 norm of the tensor multi-rank in [17], [18].

C. Hypergraph Preliminaries

Given a hypergraph $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{W})$, $\mathbf{V}$ represents a finite set of vertices, and $\mathbf{E}$ is a family of hyperedge $\boldsymbol{e}$ of $\mathbf{V}$ such that $\cup_{\boldsymbol{e}\in\mathbf{E}} = \mathbf{V}$, and a positive number $\boldsymbol{w}(\boldsymbol{e})$, which is the element of weight matrix $\mathbf{W}$, is associated with each hyperedge $\boldsymbol{e}$. An incidence matrix $\mathbf{H}$ with a size of $|\mathbf{V}| \times |\mathbf{E}|$ denotes the relationship between the vertices and the hyperedges, with entries defined as:

$$h(\boldsymbol{v}_i, \boldsymbol{e}_j) = \begin{cases} 1, & \text{if } v_i \in \boldsymbol{e}_j \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

In fact, $\mathbf{H}$ indicates to which hyperedge a vertex belongs. Given incidence matrix $\mathbf{H}$, the vertex degree of each vertex $\boldsymbol{v}_i \in \mathbf{V}$ and the edge degree of hyperedge $\boldsymbol{e}_j \in \mathbf{E}$ can be calculated as:

$$d(\boldsymbol{v}_i) = \sum_{\boldsymbol{e}_j \in \mathbf{E}} \boldsymbol{w}(\boldsymbol{e}_j) h(\boldsymbol{v}_i, \boldsymbol{e}_j), \quad (10)$$

$$d(\boldsymbol{e}_j) = \sum_{\boldsymbol{v}_i \in \mathbf{V}} h(\boldsymbol{v}_i, \boldsymbol{e}_j). \quad (11)$$

Let $\mathbf{D}_V$ and $\mathbf{D}_E$ denote the diagonal matrices whose elements correspond to the degree of each vertex $d(\boldsymbol{v}_i)$ and the degree of a hyperedge $d(\boldsymbol{e}_j)$, respectively, then the unnormalized hyper-Laplacian matrix [24] can be defined as:

$$\mathbf{L}_h = \mathbf{D}_V - \mathbf{H}\mathbf{W}\mathbf{D}_E^{-1}\mathbf{H}^{\mathrm{T}}. \quad (12)$$

From the above definition, the **main difference** between a hypergraph and a traditional pairwise graph lies in that a hyperedge can link more than two vertices. Therefore, the hypergraph can be considered as a good model to represent local group information and high order relationship among samples. For more details about the hypergraph, please refer to [24].

IV. Hyper-Laplacian Regularized M^{2}VS

Let $\mathbf{X}^{(v)} = [\mathbf{x}_1^{(v)}, \mathbf{x}_2^{(v)}, \ldots, \mathbf{x}_N^{(v)}] \in \mathbb{R}^{d^{(v)} \times N}$ denote the feature matrix corresponding to the v-th view, and $\mathbf{Z}^{(v)} = [\mathbf{z}_1^{(v)}, \mathbf{z}_2^{(v)}, \ldots, \mathbf{z}_N^{(v)}] \in \mathbb{R}^{N \times N}$ is subspace coefficient for the v-th view. The objective function of t-SVD-MSC proposed in [22] is defined as follows:

$$\begin{aligned} &\min_{\mathbf{Z}^{(v)}, \mathbf{E}^{(v)}} \lambda||\mathbf{E}||_{2,1} + ||\boldsymbol{\mathcal{Z}}||_{\circledast}, \\ &\text{s.t.} \quad \mathbf{X}^{(v)} = \mathbf{X}^{(v)}\mathbf{Z}^{(v)} + \mathbf{E}^{(v)}, v = 1, \ldots, V, \\ &\qquad \boldsymbol{\mathcal{Z}} = \Phi(\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \ldots, \mathbf{Z}^{(V)}), \\ &\qquad \mathbf{E} = [\mathbf{E}^{(1)}; \mathbf{E}^{(2)}; \ldots, \mathbf{E}^{(V)}], \end{aligned} \quad (13)$$

where $\Phi(\cdot)$ denotes how to constructs the tensor $\boldsymbol{\mathcal{Z}}$, *i.e.,* merge different $\mathbf{Z}^{(v)}$ to a tensor, and rotate it to change its dimensionality to $N \times V \times N$, as illustrated in Fig. 2. The benefit of tensor rotation primarily comes from the following [22]: the structure of self-representation coefficient can be preserved in Fourier domain, as well as the computational complexity can be significantly reduced. Also, the following relationship can be found: $\Phi_{(v)}^{-1}(\boldsymbol{\mathcal{Z}}) = \mathbf{Z}^{(v)}$, where $\Phi^{-1}(\cdot)$ represents the inverse functions of $\Phi(\cdot)$, and its subscript (v) denotes the v-th frontal slice. The term $\mathbf{E} = [\mathbf{E}^{(1)}; \mathbf{E}^{(2)}; \ldots, \mathbf{E}^{(V)}]$ can restrict the column of $\mathbf{E}^{(v)}$ from each view to have similar magnitude. Compared with unfolding based tensor low-rank norm [21], the t-TNN is a better way to extend the low rank property in matrix space into tensor space, such that the block-diagonal structure can be effectively discovered in tensor space. Consequently, the objective function in Eq. (13) can be able to capture the optimal self-representations via ensuring the consensus principle among different views.

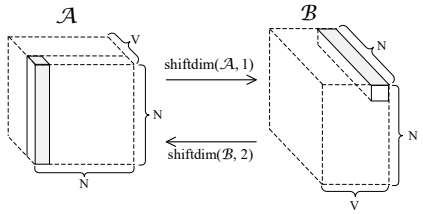


Figure 2: The rotated coefficient tensor in our approach.

While the global low rank property is captured, the intrinsic geometrical structure within each view is not taken into account by the objective function (13), leading to the lost of the locality and similarity information within individual view. To improve the t-SVD-MSC in this regard, we propose a hyper-laplacian regularized multilinear multi-view self-representation model (HLR-M^{2}VS) to model the low rank constraint in unified self-representation tensor space and geometrical constraint in individual self-representation coefficient matrix simultaneously.

A. Problem Formulation

The objective function of the proposed method can be formulated as:

$$\begin{aligned}&\min_{\mathbf{Z}^{(v)},\mathbf{E}^{(v)}} \lambda_1||\mathbf{E}||_{2,1}+\lambda_2\sum_{v=1}^{V}\mathrm{tr}(\mathbf{Z}^{(v)}\mathbf{L}_h^{(v)}\mathbf{Z}^{(v)^{\mathbf{T}}})+||\boldsymbol{\mathcal{Z}}||_{\circledast},\\ &\text{s.t.}\quad \mathbf{X}^{(v)}=\mathbf{X}^{(v)}\mathbf{Z}^{(v)}+\mathbf{E}^{(v)}, v=1,\ldots,V,\\ &\qquad \boldsymbol{\mathcal{Z}}=\Phi(\mathbf{Z}^{(1)},\mathbf{Z}^{(2)},\ldots,\mathbf{Z}^{(V)}),\\ &\qquad \mathbf{E}=[\mathbf{E}^{(1)};\mathbf{E}^{(2)};\ldots,\mathbf{E}^{(V)}].\end{aligned} \tag{14}$$

where $\mathbf{L}_h^{(v)}$ denotes the view-specific hyper-Laplacian matrix built on the optimized subspace representation $\mathbf{Z}^{(v)}$. The hyper-Laplacian regularized term is constructed based on the natural assumption that if several feature points $\mathbf{x}_{i_1}^{(v)},\ldots,\mathbf{x}_{i_k}^{(v)}$ are close in the intrinsic geometry of the feature distribution, their view-specific mappings in self-representation feature space are also close to each other. When the $\mathbf{L}_h^{(v)}$ is replaced by $\mathbf{L}^{(v)}$, which is a normal Laplacian matrix built on $\mathbf{Z}^{(v)}$, this hyper-Laplacian regularized term will be reduced to traditional graph Laplacian constraint that can only capture the pairwise relationship.

The optimization problem (14) could be solved via the inexact Augmented Lagrange Multiplier (ALM) [23]. By introducing the auxiliary matrix variable $\mathbf{Q}^{(v)}$ and the auxiliary tensor variable $\boldsymbol{\mathcal{G}}$ to replace $\mathbf{Z}^{(v)}$ in the trace term and $\boldsymbol{\mathcal{Z}}$ in the t-TNN norm respectively, the original problem can be reformulated as the following unconstrained one:

$$\begin{aligned}&\boldsymbol{\mathcal{L}}(\mathbf{Z}^{(1)},\ldots,\mathbf{Z}^{(V)};\mathbf{E}^{(1)},\ldots,\mathbf{E}^{(V)};\mathbf{Q}^{(1)},\ldots,\mathbf{Q}^{(V)};\boldsymbol{\mathcal{G}})\\ &=\lambda_1||\mathbf{E}||_{2,1}+\lambda_2\sum_{v=1}^{V}\mathrm{tr}(\mathbf{Q}^{(v)}\mathbf{L}_h^{(v)}\mathbf{Q}^{(v)^{\mathbf{T}}})+\Big(\langle\mathbf{B}_v,\mathbf{Q}^{(v)}-\mathbf{Z}^{(v)}\rangle\\ &+\frac{\mu_2}{2}||\mathbf{Q}^{(v)}-\mathbf{Z}^{(v)}||_F^2\Big)+||\boldsymbol{\mathcal{G}}||_{\circledast}+\sum_{v=1}^{V}\Big(\langle\mathbf{Y}_v,\mathbf{X}^{(v)}-\\ &\mathbf{X}^{(v)}\mathbf{Z}^{(v)}-\mathbf{E}^{(v)}\rangle+\frac{\mu_1}{2}||\mathbf{X}^{(v)}-\mathbf{X}^{(v)}\mathbf{Z}^{(v)}-\mathbf{E}^{(v)}||_F^2\Big)\\ &+\langle\boldsymbol{\mathcal{W}},\boldsymbol{\mathcal{Z}}-\boldsymbol{\mathcal{G}}\rangle+\frac{\rho}{2}||\boldsymbol{\mathcal{Z}}-\boldsymbol{\mathcal{G}}||_F^2.\end{aligned} \tag{15}$$

where $\mathbf{Y}_v$, $\mathbf{B}_v$, and $\boldsymbol{\mathcal{W}}$ represent Lagrange multipliers, μ_1, μ_2 and ρ denote the penalty parameters.

B. Optimization

The optimization process could be separated into four steps:

$\mathbf{Z}^{(v)}$-subproblem: When $\mathbf{E}$, $\mathbf{Q}^{(v)}$, and $\boldsymbol{\mathcal{G}}$ keep fixed, note that we have $\Phi_{(v)}^{-1}(\boldsymbol{\mathcal{W}})=\mathbf{W}^{(v)}$ and $\Phi_{(v)}^{-1}(\boldsymbol{\mathcal{G}})=\mathbf{G}^{(v)}$, $\mathbf{Z}^{(v)}$ can be achieved by solving the following subproblem:

$$\begin{aligned}\min_{\mathbf{Z}^{(v)}}\ &\langle\mathbf{Y}_v,\mathbf{X}^{(v)}-\mathbf{X}^{(v)}\mathbf{Z}^{(v)}-\mathbf{E}^{(v)}\rangle+\frac{\mu_1}{2}||\mathbf{X}^{(v)}-\mathbf{X}^{(v)}\mathbf{Z}^{(v)}\\ &-\mathbf{E}^{(v)}||_F^2+\langle\mathbf{B}_v,\mathbf{Q}^{(v)}-\mathbf{Z}^{(v)}\rangle+\frac{\mu_2}{2}||\mathbf{Q}^{(v)}-\mathbf{Z}^{(v)}||_F^2\\ &+\langle\mathbf{W}^{(v)},\mathbf{Z}^{(v)}-\mathbf{G}^{(v)}\rangle+\frac{\rho}{2}||\mathbf{Z}^{(v)}-\mathbf{G}^{(v)}||_F^2.\end{aligned} \tag{16}$$

The closed-form of $\mathbf{Z}^{(v)}$ can be calculated by setting the derivative of (16) to zero:

$$\begin{aligned}\mathbf{Z}^{(v)^*}=&(\rho\mathbf{I}+\mu_1\mathbf{X}^{(v)^T}\mathbf{X}^{(v)})^{-1}\Big(\mathbf{X}^{(v)^T}\mathbf{Y}_v+\mu_1\mathbf{X}^{(v)^T}\mathbf{X}^{(v)}\\ &-\mu_1\mathbf{X}^{(v)^T}\mathbf{E}^{(v)}-\mathbf{W}^{(v)}+\rho\mathbf{G}^{(v)}+\mathbf{B}_v+\mu_2\mathbf{Q}^{(v)}\Big).\end{aligned} \tag{17}$$

Note that when ρ and μ_1 are carefully chosen, *i.e.,* $\rho=\mu_1$, the matrix inverse term needs to be pre-calculated only once.

$\mathbf{E}^{(v)}$-subproblem: When $\mathbf{Z}^{(v)}$ is fixed, we have

$$\mathbf{E}^*=\operatorname*{argmin}_{\mathbf{E}}\frac{\lambda_1}{\mu_1}||\mathbf{E}||_{2,1}+\frac{1}{2}||\mathbf{E}-\mathbf{D}||_F^2, \tag{18}$$

where $\mathbf{D}$ is built through vertically concatenating the matrices $\mathbf{X}^{(v)}-\mathbf{X}^{(v)}\mathbf{Z}^{(v)}+(1/\mu_1)\mathbf{Y}_v$ together along column. This subproblem has the following solution,

$$\mathbf{E}_{:,i}^*=\begin{cases}\dfrac{||\mathbf{D}_{:,i}||_2-\frac{\lambda_1}{\mu_1}}{||\mathbf{D}_{:,i}||_2}\mathbf{D}_{:,i}, & ||\mathbf{D}_{:,i}||_2>\dfrac{\lambda_1}{\mu_1}\\ \mathbf{0} \quad \text{otherwise.}\end{cases} \tag{19}$$

where $\mathbf{D}_{:,i}$ denotes the i-th column of the matrix $\mathbf{D}$.

Q-subproblem: When $\mathbf{Z}^{(v)}$, and $\mathbf{L}_h^{(v)}$ are given, the closed-form solution of this subproblem can be calculated by:

$$\begin{aligned}\mathbf{Q}^*=&\operatorname*{argmin}_{\mathbf{Q}}\lambda_3\mathrm{tr}(\mathbf{Q}^{(v)}\mathbf{L}_h^{(v)}\mathbf{Q}^{(v)^{\mathbf{T}}})+\langle\mathbf{B}_v,\mathbf{Q}^{(v)}-\mathbf{Z}^{(v)}\rangle\\ &+\frac{\mu_2}{2}||\mathbf{Q}^{(v)}-\mathbf{Z}^{(v)}||_F^2\\ =&(\mu_2\mathbf{Z}^{(v)}-\mathbf{B}_v)(2\lambda_3\mathbf{L}_h^{(v)}+\mu_2\mathbf{I})^{-1},\end{aligned} \tag{20}$$

Algorithm 1: t-SVD based Tensor Nuclear Norm Minimization

Input: Observed tensor $\boldsymbol{\mathcal{F}}\in\mathbb{R}^{n_1\times n_2\times n_3}$, scalar $\tau>0$
Output: tensor $\boldsymbol{\mathcal{G}}$

$\boldsymbol{\mathcal{F}}_f=\mathrm{fft}(\boldsymbol{\mathcal{F}},[\,],3)$, $\tau'=n_3\tau$;
for $j=1:n_3$ **do**
  $[\boldsymbol{\mathcal{U}}_f^{(j)},\boldsymbol{\mathcal{S}}_f^{(j)},\boldsymbol{\mathcal{V}}_f^{(j)}]=\mathrm{SVD}(\boldsymbol{\mathcal{F}}_f^{(j)})$;
  $\boldsymbol{\mathcal{J}}_f^{(j)}=\mathrm{diag}\{(1-\frac{\tau'}{\boldsymbol{\mathcal{S}}_f^{(j)}(i,i)})_+\},\quad i=1,\ldots,\min(n_1,n_2)$;
  $\boldsymbol{\mathcal{S}}_{f,\tau'}^{(j)}=\boldsymbol{\mathcal{S}}_f^{(j)}\boldsymbol{\mathcal{J}}_f^{(j)}$;
  $\boldsymbol{\mathcal{G}}_f^{(j)}=\boldsymbol{\mathcal{U}}_f^{(j)}\boldsymbol{\mathcal{S}}_{f,\tau'}^{(j)}\boldsymbol{\mathcal{V}}_f^{(j)^{\mathrm{T}}}$;
end
$\boldsymbol{\mathcal{G}}=\mathrm{ifft}(\boldsymbol{\mathcal{G}}_f,[\,],3)$;
Return tensor $\boldsymbol{\mathcal{G}}$.

$\boldsymbol{\mathcal{G}}$-subproblem: Given $\{\mathbf{Z}^{(v)}\}_{v=1}^V$, we solve the following subproblem for $\boldsymbol{\mathcal{G}}$:

$$\boldsymbol{\mathcal{G}}^*=\operatorname*{argmin}_{\boldsymbol{\mathcal{G}}}||\boldsymbol{\mathcal{G}}||_{\circledast}+\frac{\rho}{2}||\boldsymbol{\mathcal{G}}-(\boldsymbol{\mathcal{Z}}+\frac{1}{\rho}\boldsymbol{\mathcal{W}})||_F^2. \tag{21}$$

To obtain the solution of (21), we use the following theorem [22]:

Theorem 1. *[22] For $\tau>0$ and $\boldsymbol{\mathcal{G}},\boldsymbol{\mathcal{F}}\in\mathbb{R}^{n_1\times n_2\times n_3}$, the globally optimal solution to the following problem*

$$\min_{\boldsymbol{\mathcal{G}}}\ \tau||\boldsymbol{\mathcal{G}}||_{\circledast}+\frac{1}{2}||\boldsymbol{\mathcal{G}}-\boldsymbol{\mathcal{F}}||_F^2 \tag{22}$$

can be calculated by the tensor tubal-shrinkage operator

$$\boldsymbol{\mathcal{G}} = \mathcal{C}_{n_3\tau}(\boldsymbol{\mathcal{F}}) = \boldsymbol{\mathcal{U}} * \mathcal{C}_{n_3\tau}(\boldsymbol{\mathcal{S}}) * \boldsymbol{\mathcal{V}}^{\mathrm{T}}, \tag{23}$$

where $\boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^{\mathrm{T}}$ *and* $\mathcal{C}_{n_3\tau}(\boldsymbol{\mathcal{S}}) = \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{J}}$*, herein,* $\boldsymbol{\mathcal{J}}$ *is an* $n_1 \times n_2 \times n_3$ *f-diagonal tensor whose diagonal element in the Fourier domain is* $\boldsymbol{\mathcal{J}}_f(i,i,j) = (1 - \frac{n_3\tau}{\boldsymbol{\mathcal{S}}_f^{(j)}(i,i)})_+$.

We provide the optimization procedure of the problem (21) is Algorithm 1. The $\mathbf{Y}_v$, $\mathbf{B}_v$, and $\boldsymbol{\mathcal{W}}$, as well as the parameters μ_1, μ_2, and ρ could be updated as follows:

$$\mathbf{Y}_v^* = \mathbf{Y}_v + \mu_1(\mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{Z}^{(v)} - \mathbf{E}^{(v)}), \tag{24}$$

$$\mathbf{B}_v^* = \mathbf{Y}_v + \mu_2(\mathbf{Q}^{(v)} - \mathbf{Z}^{(v)}), \tag{25}$$

$$\boldsymbol{\mathcal{W}}^* = \boldsymbol{\mathcal{W}} + \rho(\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{G}}), \tag{26}$$

$$\mu_i = \min(\eta\mu_i, \mu_{\max}), i = 1, 2, \tag{27}$$

$$\rho = \min(\eta\rho, \rho_{\max}). \tag{28}$$

Finally, the optimization procedure of the proposed HLR-M^{2}VS method is summarized in Algorithm 2.

In [23], the authors has proved the convergence of the inexact ALM, whose number of blocks is smaller than 3. However, its convergence properties for the objective function with N ($N \geq 3$) blocks variables, have remained unknown. Since we have blocks $\{\mathbf{Z}^{(v)}\}_{v=1}^V, \{\mathbf{E}^{(v)}\}_{v=1}^V, \{\mathbf{Q}^{(v)}\}_{v=1}^V$, and $\boldsymbol{\mathcal{G}}$ in Algorithm 2, and the objective function of (14) is not smooth, and even with involving the updated hyper-Laplacian in each iteration, it might be difficult to prove the convergence theoretically. Fortunately, the proposed method converges fast in practice, and we would provide empirical convergence analysis in the experimental section VI-E.

Algorithm 2: Hyper-Laplacian Regularized M^{2}VS

Input: Multi-view feature matrices: $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(V)}$, λ, and cluster number K
Output: Clustering results $\mathcal{C}$

Initialize $\mathbf{Z}^{(v)} = \mathbf{E}^{(v)} = \mathbf{Q}^{(v)} = \mathbf{Y}_v = \mathbf{B}_v = \mathbf{0}$, $i = 1, \ldots, V$; $\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{W}} = \mathbf{0}$;
$\mu_1 = \mu_2 = \rho = 10^{-5}$, $\eta = 2$, $\mu_{\max} = \rho_{\max} = 10^{10}$, $\varepsilon = 10^{-7}$;
while *not converge* **do**
  Compute hyper-Laplacian matrices $\{\mathbf{L}_h^{(v)}\}_{v=1}^V$ from $\{\mathbf{Z}^{(v)}\}_{v=1}^V$ by using Eqn. (12);
  `// for` $\mathbf{Z}$
  Update $\{\mathbf{Z}^{(v)}\}_{v=1}^V$ by solving (17);
  `// for` $\mathbf{E}$
  Update $\mathbf{E}$ using (18);
  `// for` $\mathbf{Q}$;
  Update $\{\mathbf{Q}^{(v)}\}_{v=1}^V$ by using (20);
  Obtain $\boldsymbol{\mathcal{Z}} = \Phi(\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \ldots, \mathbf{Z}^{(V)})$;
  `// for` $\boldsymbol{\mathcal{G}}$
  Update $\boldsymbol{\mathcal{G}}$ via Algorithm 1;
  Update $\mathbf{Y}_v$, $\mathbf{B}_v$, $\boldsymbol{\mathcal{W}}$, μ_1, μ_2, and ρ by using (24)$\sim$(28), respectively;
  $(\mathbf{G}^{(1)}, \ldots, \mathbf{G}^{(V)}) = \Phi^{-1}(\boldsymbol{\mathcal{G}})$;
  Check the convergence conditions: $||\mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{Z}^{(v)} - \mathbf{E}^{(v)}||_\infty < \varepsilon$ and $||\mathbf{Z}^{(v)} - \mathbf{G}^{(v)}||_\infty < \varepsilon$;
end
Obtain affinity matrix $\mathbf{A} = \frac{1}{V}\sum_{v=1}^V |\mathbf{Z}^{(v)}| + |\mathbf{Z}^{(v)^{\mathrm{T}}}|$;
Apply spectral clustering on matrix $\mathbf{A}$;
Return Clustering result $\mathcal{C}$.

V. HLR-M^{2}VS Model for Semi-Supervised Classification

The proposed HLR-M^{2}VS model not only can perform well in unsupervised learning, but also can be applied to semi-supervised learning scenario effectively. Since labeled examples are often expensive to obtain while unlabeled ones are easy to acquire, the semi-supervised learning problem has attracted an increasing amount of interest over the past decades [25], [26]. Among these methods is a promising family of techniques that exploit the manifold structure of the data. Under the manifold assumption, labeled and unlabeled samples are treated as vertices of a graph, such that the category information can be propagated from labeled ones to unlabeled ones through edges.

In this section, we present a novel parameter-free semi-supervised learning framework incorporating the above proposed HLR-M^{2}VS with weights adaptive multiple graph fusion. Suppose there are l ($1 \leq l \leq n$) labeled data samples, and u unlabeled ones, the proposed semi-supervised learning model is defined as:

$$\min_{\mathbf{F}\in\mathbb{R}^{n\times K}} \sum_{v=1}^V \sqrt{\mathrm{tr}(\mathbf{F}^T\mathbf{L}_h^{(v)}\mathbf{F})} \qquad s.t. \quad \mathbf{F}_l = \mathbf{Y}_l, \tag{29}$$

where $\mathbf{F} = [\mathbf{F}_l; \mathbf{F}_u] \in \mathbb{R}^{n\times K}$ denotes the class indicator matrix, with $\mathbf{F}_l$ and $\mathbf{F}_u$ including probability vectors (each row) for the labeled and unlabeled samples, respectively; $\mathbf{Y}_l = [\mathbf{y}_1, \ldots, \mathbf{y}_l]^T$ represents the label matrix in which $\mathbf{y}_i \in \mathbb{R}^{K\times 1}$ is one-hot, with $y_{ij} = 1$ indicating that the i-th sample belongs to the j-th class; $\mathbf{L}_h^{(v)}$ is the hyper-Laplacian matrix optimized iteratively by the Algorithm 2, which can be split into four blocks:

$$\mathbf{L}_h^{(v)} = \begin{bmatrix} \mathbf{L}_{ll}^{(v)} & \mathbf{L}_{lu}^{(v)} \\ \mathbf{L}_{ul}^{(v)} & \mathbf{L}_{uu}^{(v)} \end{bmatrix}. \tag{30}$$

The Lagrange function of problem (29) can be written as:

$$\sum_{v=1}^V \sqrt{\mathrm{tr}(\mathbf{F}^T\mathbf{L}_h^{(v)}\mathbf{F})} + \Omega(\boldsymbol{\Lambda}, \mathbf{F}), \tag{31}$$

where $\boldsymbol{\Lambda}$ denotes the Lagrange multiplier, $\Omega(\boldsymbol{\Lambda}, \mathbf{F})$ is the formalized term derived from the constraints. By taking its derivative w.r.t $\mathbf{F}$ and setting the derivative to zero,

$$\sum_{v=1}^V \gamma^{(v)}\partial_{\mathbf{F}}(\mathrm{tr}(\mathbf{F}^T\mathbf{L}_h^{(v)}\mathbf{F})) + \partial_{\mathbf{F}}(\Omega(\boldsymbol{\Lambda}, \mathbf{F})) = 0, \tag{32}$$

which means

$$\gamma^{(v)} = \frac{1}{2\sqrt{\mathrm{tr}(\mathbf{F}^T\mathbf{L}_h^{(v)}\mathbf{F})}}. \tag{33}$$

Since $\gamma^{(v)}$ is dependent on variable $\mathbf{F}$, the problem (32) can not be solved directly. Alternatively, we resort to the following optimization problem:

$$\min_{\mathbf{F}\in\mathbb{R}^{n\times K}} \sum_{v=1}^V \gamma^{(v)}\mathrm{tr}(\mathbf{F}^T\mathbf{L}_h^{(v)}\mathbf{F}), \qquad s.t. \quad \mathbf{F}_l = \mathbf{Y}_l, \tag{34}$$

which can be approximately solved by initializing the weight factor $\gamma^{(v)} = 1/V$, and then minimizing between $\mathbf{F}$ and

$\{\gamma^{(v)}\}_{v=1}^{V}$ alternatively by fixing the other variables. Specifically, given $\{\gamma^{(v)}\}_{v=1}^{V}$, the fused hyper-Laplacian matrix can be calculated by

$$\mathbf{L}_h^* = \sum_{v=1}^{V} \gamma^{(v)} \mathbf{L}_h^{(v)}, \tag{35}$$

then accordingly to [25], the class indicator for the unlabeled data can be achieved by

$$\mathbf{F}_u = -\mathbf{L}_{uu}^* \mathbf{L}_{ul}^* \mathbf{Y}_l. \tag{36}$$

When $\mathbf{F}_u$ is fixed, let $\mathbf{F} = [\mathbf{Y}_l; \mathbf{F}_u]$, $\gamma^{(v)}$ can be updated by Eqn. (33). Once the indicator matrix $\mathbf{F}_u$ is achieved, the predicted label for the unlabeled data can be calculated by the following function:

$$y_i = \underset{j}{\operatorname{argmax}} \mathbf{F}_{ij}, \forall i = l+1, \ldots, n; \forall j = 1, 2, \ldots, K. \tag{37}$$

The proposed semi-supervised classification method can be easily integrated into the aforementioned HLR-M^{2}VS model, and we refer it to as the Semi-HLR-M^{2}VS. Finally, the optimization procedure of the proposed Semi-HLR-M^{2}VS method is summarized in Algorithm 3. It is noteworthy that, if we have matrix A on hand (obtained from Algorithm 2), we can directly use it to construct the Laplacian matrix, and calculate the final results by alternative updating between Eqn. (33) and (36).

Algorithm 3: Semi-supervised HLR-M^{2}VS for Classification

Input: Multi-view feature: $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(V)}$, λ, and cluster number K, label matrix $\mathbf{Y}_l$
Output: The predicted labels for the unlabeled data

Including the initialization step of Alg. 2; Initialize the weight $\gamma^{(v)} = \frac{1}{V}$;
while *not converge* **do**
  Compute hyper-Laplacian matrix $\mathbf{L}_h^{(v)}$ for each view by using (12);
  Calculate $\mathbf{L}_h^*$ using (35), and achieve $\mathbf{F}_u$ by using (36);
  Update $\{\gamma^{(v)}\}_{v=1}^{V}$ by solving (33);
  Include steps $4 \sim 11$ of the Alg. 2;
end
Obtain the predicted labels by using (37);
Return The predicted labels for the unlabeled data.

VI. Experimental Results and Analysis

TABLE I: Statistics of different test datasets

Dataset	Images	Objective	Clusters
Extended YaleB	640	Face	10
ORL	400	Face	40
Notting-Hill	4660	Face	5
Scene-15	4485	Scene	15
MITIndoor-67	5360	Scene	67
COIL-20	1440	Generic Object	20
Caltech-101	8677	Generic Object	101
Caltech-256	30607	Generic Object	256

A. Experimental Setup

In this section, to investigate the performance of the proposed hyper-Laplacian regularized multilinear multi-view self-representation (HLR-M^{2}VS) model, the comprehensive experiments for both unsupervised learning (image clustering) and semi-supervised learning (image classification) are conducted.

For unsupervised learning task, we conduct experiments on eight challenging image datasets to evaluate the proposed clustering method, compared with other related state-of-the-art multi-view clustering approaches. Three applications ae included: face clustering (Extended YaleB[1], ORL[2], Notting-Hill [38] datasets), scene clustering (Scene-15[3], MITIndoor-67 [50]), and generic object clustering (COIL-20[4], Caltech-101 [55], Caltech-256 [56]). The description of all the datasets are shown in Table I. For semi-supervised learning task, we select three image datasets (Scene-15, MITIndoor-67, Caltach-101) to present the classification performance of the proposed semi-supervised method, compared with other related state-of-the-art multi-view semi-supervised classification approaches.

Dataset Descriptions: The *Extended YaleB* dataset consists 38 individuals, each of which has 64 near frontal images captured under different illumination. Following [20], [21], the first 10 individuals (640 images) are used in our experiments.

The *ORL* dataset includes 40 individuals with 10 different images per person. All these 10 images are taken from different lighting, times, facial details, and facial expressions.

The *Notting-Hill* [38] dataset consists of the faces of 5 main casts in the movie "Notting-Hill". This dataset includes 4660 faces obtained from 76 tracks. We reduce each facial image from original size of 120×150 to the size of 40×50.

Scene-15 dataset was provided in [39]–[41] with 15 categories, including kitchen, living room, bedroom, etc. Images are about 250×300 resolution, with 210 to 410 images per category. It contains a wide range of outdoor and indoor scene environments.

MITIndoor-67 dataset, which was introduced by the work [50], has more than 15K indoor pictures with 67 categories. We conduct clustering on its training subset including 5360 images. Note that this dataset is very difficult for traditional clustering methods.

The *COIL-20* dataset has 1440 images of 20 object categories generated from varying angels, and each category includes 72 images. Similar to [21], [47], all the images in our experiment are normalized to 32×32.

The *Caltech-101* dataset [55] contains 8677 image of objects belonging to 101 categories, with about 40 to 800 images per category. Usually, traditional multi-view clustering methods [31], [54] are commonly evaluated under sub-categories of this dataset. Instead of using a small portion, we use the whole dataset to evaluate the proposed methods.

The *Caltech-256* dataset [56] is a challenging set for clustering with 256 object categories containing a total of 30607 images. Compared with Caltech-101, it is collected in a similar

[1]https://cvc.yale.edu/projects/yalefacesB/yalefacesB.html
[2]http://www.uk.research.att.com/facedatabase.html
[3]http://www-cvr.ai.uiuc.edu/ponce_grp/data/
[4]http://www.cs.columbia.edu/CAVE/software/softlib/

way but with some improvements: 1) the number of categories is more than doubled; 2) the minimum number of images in each category is increased from 31 to 80. It is a relatively large dataset for subspace clustering methods.

View/Feature Description: For all the face datasets and COIL-20 datasets, similar to [21], three types of features are extracted: intensity, LBP [48] and Gabor [49]. The dimensionalities of LBP and Gabor are 3304 and 6750, respectively. For more details about these feature extraction, please refer to [22].

TABLE II: Multi-view features of different datasets

Dataset	views/features
Extended YaleB	(Intensity, LBP, Gabor)
ORL	(Intensity, LBP, Gabor)
Notting-Hill	(Intensity, LBP, Gabor)
Scene-15	(PHOW, PRI-CoLBP, CENTRIST)
MITIndoor-67	(PHOW, PRI-CoLBP, CENTRIST, VGG19)
COIL-20	(Intensity, LBP, Gabor)
Caltech-101	(PHOW, PRI-CoLBP, CENTRIST, InceptionV3)
Caltech-256	(PHOW, PRI-CoLBP, CENTRIST, VGG19, InceptionV3)

For Scene-15, MITIndoor-67, Caltech-101, Caltech-256, three types of handcrafted visual features are extracted: (1) Pyramid histograms of visual words (PHOW)[5] [43]; (2) Pairwise rotation invariant co-occurrence local binary pattern (PRI-CoLBP) feature [44]; (3) CENsus TRansform hISTogram (CENTRIST) feature [42]. For more details about these feature extraction, please refer to [22].

Besides the above three types of features, two powerful deep features, *i.e.,* the VGG-VD [51] and Inception-V3 [57] which were pre-trained on ILSVRC12 [52], are imported as new views to complement handcrafted features for MITIndoor-67 and Caltech-101, respectively. Following [22], for VGG-VD feature, we change the smaller size of image to 448 while keeping the aspect ratio, and utilize the activations of the penultimate layer as feature vector. The features are extracted from 5 scales $\{2^s, s = -1, -0.5, 0, 0.5, 1\}$, and all local features are pooling together regardless of scales and locations. As for Inception-V3 feature, it is directly extracted from the activations of the penultimate layer, resulting in a 2048-dimensional feature vector. The employed features for all the datasets are listed in Table II.

Evaluation Measures. For clustering task, we employ six popular metrics to evaluate the performances [30]: Normalized Mutual Information (NMI), Accuracy (ACC), Adjusted Rank index (AR), F-score, Precision and Recall. For the detailed definitions, please refer to [30]. For each of the metrics, the higher value means the better performance.

For semi-supervised classification, the evaluation metric is accuracy (the proportion of the correct-classified data points in all unlabeled data). It is noteworthy that the accuracy used in semi-supervised setting has slight difference from that in clustering, which needs additional permutation mapping function to assign cluster to its corresponding groundtruth. Note that in all datasets, for both clustering and semi-supervised learning tasks, we report the final results by the average of 20 runs. For more details about the means and standard deviations, please refer to the supplemental material.

Two parameters λ_1 and λ_2 need to be tuned in the proposed model, and we find their empirical values are within the range $[0.01, 0.2]$ and $[0.1, 0.9]$, respectively. More details about the parameters will be discussed in Section VI-D. The weight of the hyperedge in the v-th hypergraph is defined on the Euclidean distance between the two columns in $A^{(v)} = \frac{1}{2}(|\mathbf{Z}^{(v)}| + |\mathbf{Z}^{(v)^{\mathrm{T}}}|)$. The k-Nearest Neighbor (kNN) is used to construct the hyperedges with fixing $k = 5$ in all experiments. The parameters settings of all the competitors are set according to their original papers, and we have tuned these parameters to show the best results. The experiments are implemented in Matlab on a workstation with 4.0GHz CPU, 128GB RAM, and TITANX GPU (12GB caches). The source codes and some results of the proposed models can be achieved at https://www.researchgate.net/profile/Yuan_Xie4.

B. Experiments on Clustering

Competitors: The standard spectral clustering algorithm with the most informative view (SPC_{best}), LRR method with the most informative view (LRR_{best}), the non-negative sparse hyper-Laplacian regularized LRR with most informative view ($\text{NSH-LRR}_{\text{best}}$) [9]. All these methods belong to single-view baselines.

The following state-of-the-art methods are also included: the robust multi-view spectral clustering via low-rank and sparse decomposition (RMSC) [10], diversity-induced multi-view subspace clustering (DiMSC) [47], multi-view learning with adaptive neighbours (MLAN) [36], low-rank tensor constrained multi-view subspace clustering (LTMSC) [21], exclusivity-consistency regularized multi-view subspace clustering (ECMSC) [59], the t-SVD based multi-view subspace clustering (t-SVD-MSC), learning and transferring deep ConvNet representations with group-sparse factorization (GSNMF-CNN) [37].

TABLE III: Clustering results on *Extended YaleB*. We set $\lambda_1 = 0.05$ and $\lambda_2 = 0.2$ in proposed HLR-M^{2}VS.

Method	NMI	ACC	AR	F-score	Precision	Recall
SPC_{best}	0.360	0.366	0.225	0.308	0.296	0.310
LRR_{best}	0.627	0.615	0.451	0.508	0.481	0.539
$\text{NSH-LRR}_{\text{best}}$	0.632	0.629	0.431	0.523	0.507	0.545
RMSC	0.157	0.210	0.060	0.155	0.151	0.159
DiMSC	0.636	0.615	0.453	0.504	0.481	0.534
MLAN	0.352	0.346	0.093	0.213	0.159	0.321
LTMSC	0.637	0.626	0.459	0.521	0.485	0.539
ECMSC	**0.759**	**0.783**	**0.544**	**0.597**	**0.513**	**0.718**
t-SVD-MSC	0.667	0.652	0.500	0.550	0.514	0.590
HLR-M^{2}VS	0.703	0.670	0.529	0.577	0.560	0.595

1) Experiments on Face Clustering: Surprisingly, the proposed method obtains a *perfect* result on the ORL dataset, as it is shown Table IV. For Extended YaleB dataset (Table III), it has been pointed out in [21], [22] that LBP and Gabor

[5]This feature was extracted by using vlfeat toolbox [45]

TABLE IV: Clustering results on *ORL*. We set $\lambda_1 = 0.2$ and $\lambda_2 = 0.4$ in proposed HLR-M^{2}VS.

Method	NMI	ACC	AR	F-score	Precision	Recall
SPC$_{\text{best}}$	0.884	0.725	0.655	0.664	0.610	0.728
LRR$_{\text{best}}$	0.895	0.773	0.724	0.731	0.701	0.754
NSH-LRR$_{\text{best}}$	0.913	0.786	0.733	0.740	0.707	0.775
RMSC	0.872	0.723	0.645	0.654	0.607	0.709
DiMSC	0.940	0.838	0.802	0.807	0.764	0.856
MLAN	0.854	0.705	0.384	0.376	0.254	0.721
LTMSC	0.930	0.795	0.750	0.768	0.766	0.837
ECMSC	0.947	0.854	0.810	0.821	0.783	0.859
t-SVD-MSC	0.993	0.970	0.967	0.968	0.946	0.991
HLR-M^{2}VS	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

TABLE V: Clustering results on *Notting-Hill*. We set $\lambda_1 = 0.04$ and $\lambda_2 = 0.9$ in proposed HLR-M^{2}VS.

Method	NMI	ACC	AR	F-score	Precision	Recall
SPC$_{\text{best}}$	0.723	0.816	0.712	0.775	0.780	0.776
LRR$_{\text{best}}$	0.579	0.794	0.558	0.653	0.672	0.636
NSH-LRR$_{\text{best}}$	0.615	0.808	0.579	0.667	0.684	0.651
RMSC	0.585	0.807	0.496	0.603	0.621	0.586
DiMSC	0.799	0.837	0.787	0.834	0.822	0.827
MLAN	0.476	0.584	0.301	0.504	0.380	0.748
LTMSC	0.779	0.868	0.777	0.825	0.830	0.814
ECMSC	0.817	0.767	0.679	0.764	0.637	0.954
t-SVD-MSC	0.900	0.957	0.900	0.922	0.937	0.907
HLR-M^{2}VS	**0.982**	**0.996**	**0.990**	**0.986**	**0.989**	**0.984**

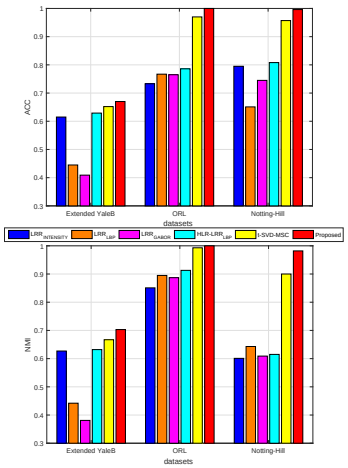


Fig. 3: Comparison among LRR with all the view features, NSH-LRR with the most information view, t-SVD-MSC, and the proposed HLR-M^{2}VS in terms of accuracy and NMI on face clustering datasets.

features will lead to degenerate views, which is illustrated in Fig. 3 (see the first group bars). The t-SVD-MSC and the proposed HLR-M^{2}VS seem to be insusceptible to those two degenerate views, and moreover, the HLR-M^{2}VS performs better due to the intrinsic geometrical structures being captured through hyper-Laplacian constraint.

Table V presents the results on Notting-Hill, in which our method achieves *nearly perfect* result, outperforms all competitors by a large margin. Our result even beat the state-of-the-art result achieved by [46] (with NMI 0.920 and ACC 0.934 in [46], vs NMI 0.982 and ACC 0.996 in proposed method), where two additional constraints regarding the video-specific priori are utilized, *i.e.,* the faces, which appear together in the same frame, are not likely to belong to a same person; while the faces can be considered to be captured from a same person if they are contained in the same track.

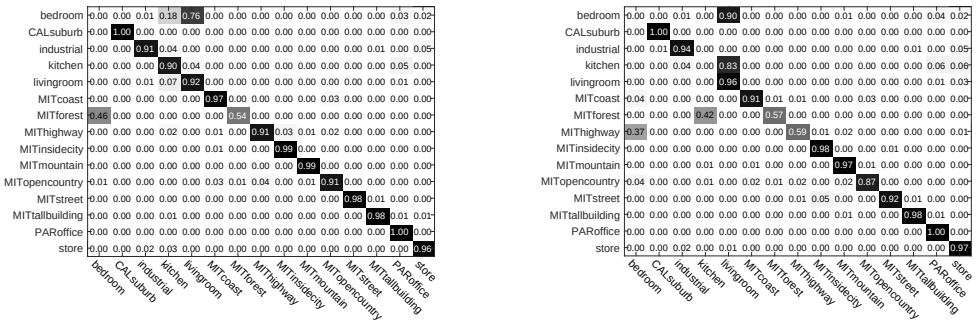
TABLE VI: Clustering results on *Scene-15*. We set $\lambda_1 = 0.01$ and $\lambda_2 = 0.1$ in proposed HLR-M^{2}VS.

Method	NMI	ACC	AR	F-score	Precision	Recall
SPC$_{\text{best}}$	0.421	0.437	0.270	0.321	0.314	0.329
LRR$_{\text{best}}$	0.426	0.445	0.272	0.324	0.316	0.333
NSH-LRR$_{\text{best}}$	0.521	0.515	0.389	0.441	0.436	0.447
RMSC	0.564	0.507	0.394	0.437	0.425	0.450
DiMSC	0.269	0.300	0.117	0.181	0.173	0.190
MLAN	0.475	0.331	0.151	0.248	0.149	0.733
LTMSC	0.571	0.574	0.424	0.465	0.452	0.479
ECMSC	0.463	0.457	0.303	0.357	0.318	0.408
t-SVD-MSC	0.858	0.812	0.771	0.788	0.743	0.839
HLR-M^{2}VS	**0.895**	**0.878**	**0.850**	**0.861**	**0.850**	**0.871**

2) Experiments on Scene Clustering: In Table VI, the results of Scene-15 dataset, we can see a noticeable performance gain by comparing the t-SVD-MSC with the LTMSC method. Although the performance is excellent, the t-SVD-MSC can still be improved significantly by integrating the high-order local structural information derived from hyper-Laplacian regularizer. This can be further supported by the confusion matrices shown in Fig. 4, where column and row names are the predicted labels and the groundtruth, respectively. The way to calculation of predicted labels is the same as the method used in computing the metric ACC [53]. Compared with t-SVD-MSC, accuracy in almost all categories has been improved by the proposed methods. Specially, the accuracy in "kitchen" and "MIThighway" is dramatically improved with 0.90 vs 0.00 and 0.91 vs 0.59, respectively. The biggest confusion occurs between two indoor classes, *i.e.,* "bedroom" and "livingroom", which coincides well with the the confusion distribution in [41].

As for MITIndoor-67 dataset, compared with t-SVD-MSC (see table VII), our method gains significant improvement around 11.6%, 11.8%, 17.5%, 17.2%, 17.0%, and 17.5% in terms of NMI, ACC, AR, F-score, Precision, and Recall, respectively. This is another evidence that the local geometrical information provided by hypergraph can make a complement to the global low rank imposed on unified tensor space.

3) Experiments on Generic Clustering: As shown in Table VIII, on COIL-20 dataset, the proposed methods also outperforms four recently published approaches, *i.e.,* RMSC, DiMSC, LTMSC, and t-SVD-MSC. The results for Caltech-101 dataset are shown in Table IX. By introducing the more powerful deep feature (Inception-V3), the single-view baseline approaches, *i.e.,* SPC and LRR, even perform better than multi-view methods such as RMSC and DiMSC. It might



(a) HLR-M^{2}VS (b) t-SVD-MSC

Fig. 4: Comparison the confusion matrices between t-SVD-MSC and the proposed method on *Scene-15* dataset.

TABLE VII: Clustering results on *MITIndoor-67*. We set $\lambda_1 = 0.02$ and $\lambda_2 = 0.2$ in proposed HLR-M^{2}VS.

Method	NMI	ACC	AR	F-score	Precision	Recall
SPC$^{\text{CNN}}_{\text{best}}$	0.559	0.443	0.304	0.315	0.294	0.340
LRR$^{\text{CNN}}_{\text{best}}$	0.226	0.120	0.031	0.045	0.044	0.047
NSH-LRR$^{\text{CNN}}_{\text{best}}$	0.270	0.202	0.106	0.118	0.117	0.120
RMSC	0.342	0.232	0.110	0.123	0.121	0.125
DiMSC	0.383	0.246	0.128	0.141	0.138	0.144
MLAN	0.408	0.232	0.012	0.041	0.021	0.662
LTMSC	0.546	0.431	0.280	0.290	0.279	0.306
ECMSC	0.489	0.353	0.216	0.228	0.213	0.247
GSNMF-CNN	0.673	0.517	0.264	0.372	0.367	0.381
t-SVD-MSC	0.750	0.684	0.555	0.562	0.543	0.582
HLR-M^{2}VS	**0.866**	**0.802**	**0.730**	**0.734**	**0.713**	**0.757**

TABLE VIII: Clustering results on *COIL-20*. We set $\lambda_1 = 0.1$ and $\lambda_2 = 0.8$ in proposed HLR-M^{2}VS.

Method	NMI	ACC	AR	F-score	Precision	Recall
SPC$_{\text{best}}$	0.806	0.672	0.619	0.640	0.596	0.692
LRR$_{\text{best}}$	0.829	0.761	0.720	0.734	0.717	0.751
NSH-LRR$_{\text{best}}$	0.840	0.785	0.738	0.758	0.726	0.793
RMSC	0.800	0.685	0.637	0.656	0.620	0.698
DiMSC	0.846	0.778	0.732	0.745	0.739	0.751
MLAN	0.945	0.844	0.804	0.815	0.726	0.929
LTMSC	0.860	0.804	0.748	0.760	0.741	0.776
ECMSC	0.942	0.782	0.781	0.794	0.695	0.925
t-SVD-MSC	0.884	0.830	0.786	0.800	**0.785**	0.808
HLR-M^{2}VS	**0.960**	**0.852**	**0.833**	**0.842**	0.757	**0.949**

TABLE IX: Clustering results on *Caltech-101*. We set $\lambda_1 = 0.08$ and $\lambda_2 = 0.2$ in proposed HLR-M^{2}VS.

Method	NMI	ACC	AR	F-score	Precision	Recall
SPC$^{\text{CNN}}_{\text{best}}$	0.723	0.484	0.319	0.340	0.597	0.235
LRR$^{\text{CNN}}_{\text{best}}$	0.728	0.510	0.304	0.339	0.627	0.231
NSH-LRR$^{\text{CNN}}_{\text{best}}$	0.757	0.522	0.338	0.368	0.635	0.260
RMSC	0.573	0.346	0.246	0.258	0.457	0.182
DiMSC	0.589	0.351	0.226	0.253	0.362	0.191
MLAN	0.748	0.579	0.222	0.265	0.173	0.560
LTMSC	0.788	0.559	0.393	0.403	0.670	0.288
ECMSC	0.606	0.359	0.273	0.286	0.433	0.214
GSNMF-CNN	0.775	0.534	0.246	0.275	0.230	**0.347**
t-SVD-MSC	0.858	0.607	0.430	0.440	0.742	0.323
HLR-M^{2}VS	**0.872**	**0.650**	**0.463**	**0.472**	**0.760**	0.343

because that both the DiMSC and RMSC are affected by the less representation capabilities of the handcrafted features. However, this does not mean that the traditional feature is unnecessary. The performance gains of HLR-M^{2}VS over all competitors on all the datasets confirm complementary both in feature level and subspace level. In feature level, each view might has its own knowledge that other views do not possess, which is help the model improve the clustering performance effectively. In subspace level, the local geometrical information from all the views will be considered as a complement to the global low-rank based consensus to upgrade the performance to a higher level. The similar observation can be seen from the results Table X of the Caltech-256.

C. Experiments on Semi-Supervised Classification

To evaluate our proposed model on semi-supervised learning task, we select three challenging image datasets that are used in clustering task for the classification in our experiments, *i.e.,* Scene-15, MITIndoor-67, and Caltech-101. The aim of semi-supervised learning is to reveal more unlabeled information from limited known labeled data, so we only select the percentage of labeled samples ranges from 10% to 60% for training. For each dataset, we run 10 times for each algorithms with randomly labeled data sets, and then report results of classification accuracy by averaging the results corresponding to each labeled percentage. Note that the proposed Semi-HLR-M^{2}VS does not introduce any other additional parameter. Therefore, the parameter setting is the same with clustering version on respective dataset.

Competitors: In semi-supervised learning task, we compare the proposed method with six representative single-view and multi-view semi-supervised classification algorithms: the non-negative sparse hyper-laplacian regularized LRR model (NSH-LRR) [9], the sparse multiple graph integration approach (SMGI) [32], the adaptive multi-model semi-supervised classification method (AMMSS) [33], the multi-modal curriculum learning for semi-supervised classification (MMCL) [34], the parameter-free auto-weighted multiple graph learning (AMGL) [35], the multi-view semi-supervised learning with

TABLE X: Clustering results on *Caltech-256*. We set $\lambda_1 = 0.04$ and $\lambda_2 = 0.1$ in proposed HLR-M^{2}VS.

Method	NMI	ACC	AR	F-score	Precision	Recall
$\text{SPC}^{\text{INS}}_{\text{best}}$	0.192	0.030	0.002	0.006	0.007	0.005
$\text{LRR}^{\text{INS}}_{\text{best}}$	0.661	0.470	0.306	0.310	0.350	0.278
$\text{NSH-LRR}^{\text{INS}}_{\text{best}}$	0.679	0.500	0.323	0.333	0.374	0.306
RMSC	0.507	0.322	0.181	0.159	0.197	0.129
DiMSC*	\	\	\	\	\	\
MLAN	0.784	0.582	0.384	0.389	0.300	0.550
LTMSC	0.761	0.502	0.335	0.345	0.227	0.725
ECMSC	0.521	0.337	0.243	0.330	0.222	0.649
GSNMF-CNN	0.737	0.517	0.345	0.349	0.311	0.398
t-SVD-MSC	0.840	0.570	0.419	0.423	0.286	0.815
HLR-M^{2}VS	**0.885**	**0.616**	**0.441**	**0.460**	**0.323**	**0.849**

*DiMSC runs out of memory in current platform due to the calculation of Sylvester equation.

TABLE XI: Classification accuracy for *Scene-15* dataset based on various method under different percentages of labeled samples.

Methods	10	20	30	40	50	60
NSH-LRR	0.639	0.684	0.697	0.717	0.724	0.722
SMGI	0.617	0.649	0.676	0.692	0.699	0.706
AMMSS	0.599	0.680	0.694	0.709	0.701	0.716
AMGL	0.678	0.712	0.750	0.761	0.770	0.778
MMCL	0.645	0.680	0.725	0.747	0.748	0.740
MLAN	0.600	0.659	0.671	0.677	0.702	0.739
Semi-HLR-M^{2}VS	**0.895**	**0.917**	**0.934**	**0.967**	**0.983**	**0.943**

TABLE XII: Classification accuracy for *MITIndoor-67* dataset based on various method under different percentages of labeled samples.

Methods	10	20	30	40	50	60
NSH-LRR	**0.610**	0.665	0.693	0.717	0.733	0.739
SMGI	0.530	0.577	0.623	0.638	0.660	0.684
AMMSS	0.572	0.632	0.662	0.686	0.699	0.703
AMGL	0.298	0.377	0.461	0.502	0.547	0.594
MMCL	0.524	0.539	0.569	0.580	0.595	0.605
MLAN	0.536	0.552	0.572	0.590	0.603	0.607
Semi-HLR-M^{2}VS	0.534	**0.693**	**0.798**	**0.843**	**0.922**	**0.871**

adaptive neighbors (MLAN) [36]. The first method is the single view hyper-Laplacian regularized baseline, while the rest ones represent the state-of-the-art approaches in multi-view semi-supervised learning.

The semi-supervised classification results are reported in Table XI $\sim$ XIII and Fig. 5. We can see that the Semi-HLR-M^{2}VS method always achieves the best classification accuracy except the case of 10 percentage labeled data on MITIndoor-67 dataset. Numerically, on Scene-15, the proposed method leads AMGL (the second best approach, the magenta curve in Fig. 5 (a)) with the margins approximately at least 17% and up to 22% for different labeled percentages. Similar observation can be seen on the other two datasets. Comparing with the single-view baseline, we can conclude that, merely using single-view feature accompanying with hyper-Laplacian regularization is not enough to handle the nonlinear subspace segmentation, as it is done in NSH-LRR. Moreover, the superiority to other state-of-the-art approaches means that the multi-view consensus (global constraint) and view-specific manifold regularization (local constraint) interact with each other so as to facilitate the supervised information to propagate to unlabeled data through the graph implicitly defined by the fused hyper-Laplacian matrix (Eqn. (35)) more effectively.

TABLE XIII: Classification accuracy for *Caltech-101* dataset based on various method under different percentages of labeled samples.

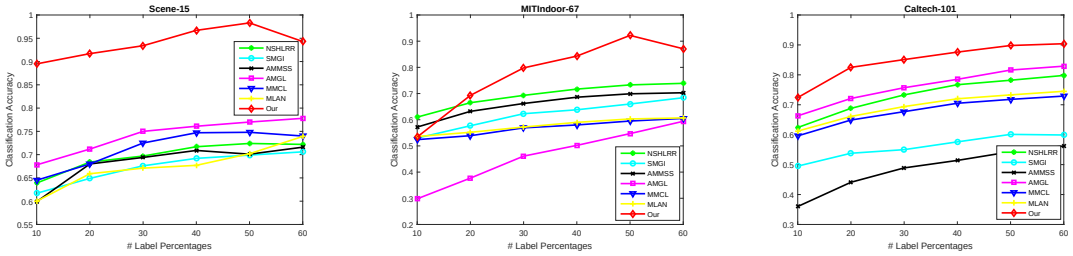
Methods	10	20	30	40	50	60
NSH-LRR	0.624	0.688	0.733	0.767	0.782	0.798
SMGI	0.495	0.538	0.550	0.576	0.601	0.599
AMMSS	0.360	0.441	0.489	0.514	0.543	0.562
AMGL	0.663	0.721	0.757	0.785	0.816	0.829
MMCL	0.596	0.649	0.677	0.705	0.718	0.729
MLAN	0.612	0.661	0.694	0.720	0.733	0.745
Semi-HLR-M^{2}VS	**0.724**	**0.825**	**0.851**	**0.876**	**0.898**	**0.904**

D. Parametric Sensitivity

The weighting factors λ_1 and λ_2 are two tuning parameters in the proposed HLR-M^{2}VS model. The parameter $\lambda_1 > 0$ is used to denote the influence of the noise upon the dataset. Commonly, the choice of λ_1 relies on the error level of the data, *i.e.,* λ_1 will be set to relatively large value when relative large corruption appears in multiple features, and small value would be preferred otherwise. The $\lambda_2 > 0$ is utilized to tradeoff the effects between the view-specific geometrical regularization and the low rank constraint in unified self-representation coefficient tensor space.

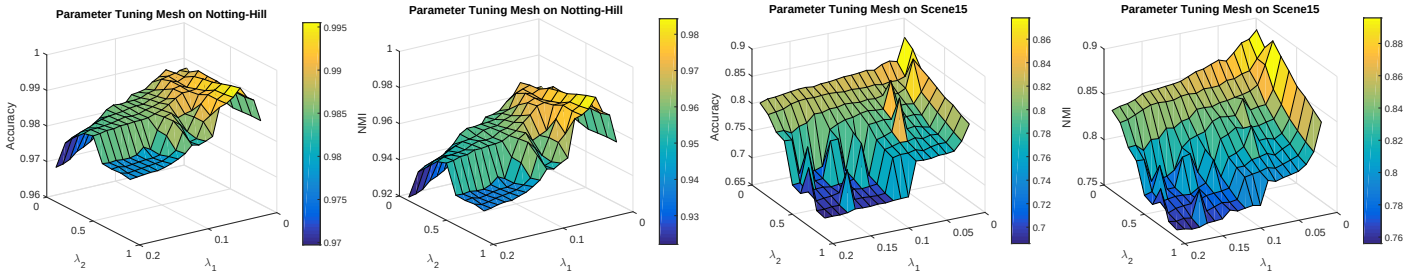
Through intensively parameters tuning, we empirically found that λ_1 and λ_2 locate within the ranges $[0.01, 0.2]$ and $[0.1, 0.9]$, respectively. Specifically, we are more interested in how the variations of those two parameters influence the algorithm output. To this end, the parameters λ_1 and λ_2 are changing from 0.01 to 0.2 and 0.1 to 0.9 with intervals 0.01 and 0.1, respectively, to see the produced accuracy and normalized mutual information. The experimental results are illustrated in Fig. 6, which shows the evaluation results on two typical datasets, *i.e.,* Notting-Hill and Scene-15.

Although the parameters λ_1 and λ_2 play an important role on the performance, as illustrated in Fig. 6 (a) and (b), on Notting-Hill dataset, all results are still better than its most powerful competitor t-SVD-MSC, which indicates the partial stability of the proposed model. As for Scene-15 dataset, not all the outputs of the proposed method outperform the t-SVD-MSC. However, it is noteworthy that the reported results of t-SVD-MSC are obtained by carefully tuning. Since the proposed model can further improve the result of t-SVD-MSC, which obtained by its best parameter configuration, this can demonstrate that the hyper-Laplacian regularization can be



(a) Comparison on Scene-15 (b) Comparison on MITIndoor-67 (c) Comparison on Caltech-101

Fig. 5: Classification accuracy versus the labeled percentages on *Scene-15*, *MITIndoor-67*, and *Caltech-101* datasets by using different semi-supervised learning approaches.



(a) ACC on Notting-Hill (b) NMI on Notting-Hill (c) ACC on Scene15 (d) NMI on Scene15

Fig. 6: Parameters tuning (λ_1 and λ_2) in terms of ACC and NMI on Notting-Hill and Scene15 datasets.

viewed as an efficient local geometrical complement to the global tensor based low rank regularization.

It also can be learned from the above parametric analysis that the relationship between the two critical regularizers, *i.e.,* t-TNN part and hyper-Laplacian part, could be clearly understood. The t-TNN part can be considered as the cornerstone of the proposed method. This part acts as a global constraint to ensure the consensus among multiple views to obtain a refinement pairwise relationship among samples for all the views. On the other hand, hyper-Laplacian part acts as an auxiliary module to supplement to the t-TNN part, since the model (Eqn. (14)) without hyper-graph part, *i.e.,* t-SVD-MSC, can not handle the data sampled from non-linear subspaces. The contributions of the two parts can be judged from their corresponding weights in objective function Eqn. (14), where the weight λ_2 represents the importance of the hyper-Laplacian part in the proposed model, while the weight for t-TNN part is always set to 1. According to the above parametric analysis, λ_2 usually locates within the range $[0.1, 0.9]$ (good performances usually appear when $\lambda_2 < 0.5$, see Fig. 6). Comparing the weights of different parts, we can conclude that the t-TNN part contributes more for the final result.

E. Computational Complexity and Convergence

The construction of $\{\mathbf{L}_h^{(v)}\}_{v=1}^V$ and the optimization procedures for $\mathbf{Q}$ and $\boldsymbol{\mathcal{G}}$ are the three computation-intensive steps. As for the hyper-Laplacian, it takes $\mathcal{O}(VN^2\log(N))$ for all the views. As for the $\boldsymbol{\mathcal{G}}$ subproblem, according to [22], it will spend $\mathcal{O}(2N^2V\log(N) + N^2V^2) \approx \mathcal{O}(2N^2V\log(N))$ in each iteration, in which the former is the cost for FFT and IFFT operations, the latter represents the cost of SVD for N $V \times N$ matrices. Since we keep the nearest neighbors to 5 for constructing the hypergraph, solving the sparsified Laplacian linear system for $\mathbf{Q}$ will take $\mathcal{O}(m\log^c(m))$ for a constant c [58], where m denotes the number of non-zero entries in Laplacian matrix. By considering the number of iterations and the cost of spectral clustering, for Algorithm 2, we have the following computational complexity:

$$\mathcal{O}(N^3) + \mathcal{O}(K(2N^2V\log(N))), \tag{38}$$

where K is the number of iterations.

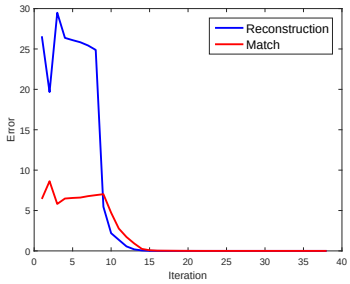


Figure 7: Convergence plot on Scene-15 dataset.

Practically, the derived optimization method converges fast. As it is shown in Fig. 7, the two error terms, namely the reconstruction error (RE) and the match error (ME), are defined according to the convergence conditions (see steps (11) in Algorithm 2):

$$\text{RE} \doteq \frac{1}{V}\sum_{v=1}^{V}||\mathbf{I} - \mathbf{Z}^{(v)} - \mathbf{P}^{(v)}||_\infty \tag{39}$$

$$\text{ME} \doteq \frac{1}{V}\sum_{v=1}^{V}||\mathbf{Z}^{(v)} - \mathbf{G}^{(v)}||_{\infty} \tag{40}$$

Usually, the number of optimization iteration is between 20 and 40.

VII. Conclusions

In this paper, a hyper-Laplacian regularized multilinear self-representation model is derived to conduct clustering and semi-supervised classification by using multi-view heterogeneous features. In the proposed model, all the subspace coefficients will be alternatively optimized both in unified tensor space and view-specific feature spaces. On the one hand, in unified tensor space, the global consensus could be captured through low-rank approximation of the rotated subspace coefficient tensor by using t-SVD based tensor multi-rank minimization. On the other hand, in view-specific feature spaces, the local high-order geometrical structure will be discovered by imposing the hyper-Laplacian regularization on view-specific self-representation coefficient matrix. These two aspects can be regarded as the *global* and *local* constraints to ensure the consensus principle among multiple views and preserve geometrical structure in each view, respectively. Furthermore, the proposed model can be extended to semi-supervised classification without introducing any additional parameter. Extensive evaluation is conducted on several challenging datasets, in which a remarkable advance over state-of-the-art multi-view clustering and multi-view semi-supervised classification approaches is obtained.

VIII. Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61772524, Grant 61772525, Grant 61876161, Grant 61701235, Grant 61373077, Grant 61602482; by the Beijing Municipal Natural Science Foundation under Grant 4182067; by the Fundamental Research Funds for the Central Universities under Grant 30917011323; by the Australian Research Council Projects FT-130101457, DP-120103730, LP-150100671.

References

[1] S. Xiao, M. Tan, D. Xu, and Z. Dong, "Robust kernel low-rank representation," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 27, no. 11, pp. 2268-2281, 2016.

[2] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using laplacianfaces," *IEEE Trans. on Pattern Recognition and Machine Intelligence*, vol. 27, no. 3, pp. 328-340, 2005.

[3] V. Patel, and R. Vidal, "Kernel sparse subspace clustering," *Proc. International Conference on Image Processing*, 2014.

[4] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Liu, and D. Cai, "Graph regularized sparse coding for image representation," *IEEE Trans. on Image Processing*, vol. 20, no. 5, pp. 1327-1336, 2011.

[5] S. Gao, I. Tsang, and L. Chia, "Laplacian sparse coding, hypergraph laplacian sparse coding, and applications," *IEEE Trans. on Pattern Recognition and Machine Intelligence*, vol. 35, no. 1, pp. 92-104, 2013.

[6] S. T. Roweis, and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323-2326, 2000.

[7] X. He, D. Cai, S. Yan, and H. Zhang, "Neighborhood preserving embedding," *Proc. International Conference on Computer Vision*, 2005.

[8] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *Journal of Machine Learning Research*, vol. 7, pp. 2399-2434, 2006.

[9] M. Yin, J. Gao, and Z. Lin, "Laplacian regularized low-rank representation and its applications," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 38, no. 3, pp. 504-517, 2016.

[10] R. Xia, Y. Pan, L. Du, and J. Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," In *AAAI*, 2014.

[11] K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan, "Multi-view clustering via canonical correlation analysis," *Proc. International Conference on Machine Learning*, pp. 129-136, 2009.

[12] M. B. Blaschko, and C. H. Lampert, "Correlational spectral clustering," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1-8, 2008.

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

[14] W. Wang, R. Arora, K. Livescu, and J. Bilmes, "On deep multi-view representation learning," *Proc. International Conference on Machine Learning*, 2015.

[15] M. White, X. Zhang, D. Schuurmans, and Y. I. Yu, "Convex multi-view subspace learning," In NIPS, 2012.

[16] M. Kilmer, K. Braman, N. Hao, and R. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148-172, 2013.

[17] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-SVD," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2014.

[18] O. Semerci, Ning Hao, M. Kilmer, and E. Miller, "Tensor-based formulation and nuclear norm regularization for multienergy computed tomography," *IEEE Trans. Image Processing*, vol. 23, no. 4, pp. 1678-1693, 2014.

[19] E. Elhamifar, and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and application," In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765-2781, 2013.

[20] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171-184, 2013.

[21] C. Zhang, H. Fu, S. Liu, G. Liu, and X. Cao. "Low-rank tensor constrained multiview subspace clustering," *Proc. International Conference on Computer Vision*, pp. 2439-2446, 2015.

[22] Y. Xie, D. Tao, W. Zhang, L. Zhang, Y. Liu, and Y. Qu, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *International Journal of Computer Vision*, preprint, 2018.

[23] Z. Lin, M. Chen, Y. Ma. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. *Technical Report UILU-ENG-09-2215*, UIUC, 2009.

[24] D. Zhou, J. Huang, and B. Schölkopf. "Learning with hypergraphs: Clustering, classification, and embedding," *In Proc. Advances in Neural Information Processing Systems*, 2006.

[25] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," *Proc. International Conference on Machine Learning*, 2003.

[26] L. Zhuang, H. Yao, Z. Lin, Y. Ma, X. Zhang, and N. Yu, "Non-negative low rank and sparse graph for semi-supervised learning," *Proc. Computer Vision and Pattern Recognition*, 2012.

[27] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455-500, 2009.

[28] J. Eckstein and D. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Programming*, vol. 55, pp. 293-318, 1992.

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

[30] M. Christopher D., P. Raghavan, and H. Schtze, *Introduction to Information Retrieval*, vol. 1, Cambridge University Press, Cambridge, 2008.

[31] C. Lu, S. Yan, and Z. Lin, "Convex sparse spectral clustering: single-view to multi-view," *IEEE Trans. on Image Processing*, vol. 25, no. 6, pp. 2833-2843, 2016.

[32] M. Karasuyama, and H. Mamitsuka, "Multiple graph label propagation by sparse integration," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 24, no. 12, pp. 1999-2012, 2013.

[33] X. Cai, F. Nie, W. Cai, and H. Huang, "Heterogeneous image features integration via multi-modal semi-supervised learning model," *Proc. Computer Vision and Pattern Recognition*, 2013.

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

[35] F. Nie, J. Li, and X. Li, "Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification," *International Joint Conference on Artifical Intelligence*, 2016.
[36] F. Nie, G. Cai, and X. Li, "Multi-view clustering and semi-supervised classification with adaptive neighbours," *In Proc. AAAI Conference on Artificial Intelligence*, 2017.
[37] L. Gui, and L. P. Morency, "Learning and transferring deep ConvNet representations with group-sparse factorization," *Proc. IEEE International Conference on Computer Vision*, 2015.
[38] Y. Zhang, C. Xu, H. Lu, and Y. M. Huang, "Character identification in feature-length films using global face-name matching," *IEEE Trans. on Multimedia*, vol. 11, no. 7, pp. 1276-1288, 2009.
[39] A. Oliva, and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, pp. 145-175, 2001.
[40] L. Fei-Fei, and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 524-531, 2005.
[41] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2169-2178, 2006.
[42] J. Wu, and J. M. Rehg. "Centrist: A visual descriptor for scene categorization," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1489-1501, 2011.
[43] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forest and ferns," In *Proc. IEEE International Conference on Computer Vision*, 2007.
[44] X. Qi, R. Xiao, C. Li, Y. Qiao, J. Guo, and X. Tang, "Pairwise rotation invariant co-occurrence local binary pattern," *IEEE Trans. on Pattern Recognition and Machine Intelligence*, vol. 36, no. 11, 2014.
[45] A. Vedaldi, and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," http://www.vlfeat.org/, 2008.
[46] X. Cao, C. Zhang, C. Zhou, H. Fu, and H. Foroosh, "Constrained multi-view video face clustering," *IEEE Trans. on Image Processing*, vol. 24, no. 11, pp. 4381-4393, 2015.
[47] X. Cao, C. Zhang, H. Fu, S. Liu, and H. Zhang, "Diversity-induced multi-view subspace clustering," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015.
[48] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. on Pattern Recognition and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002.
[49] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. on Computer*, vol. 42, no. 3, pp. 300-311, 1993.
[50] A. Quattoni, and A. Torralba, "Recognizing indoor scenes," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 413-420, 2009.
[51] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *International Conference on Learning Representations*, 2014.
[52] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "ImageNet: a large-scale hierarchical image database," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009.
[53] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. on Pattern Recognition and Machine Intelligence*, vol. 33, no. 8, pp. 1548-1560, 2011.
[54] H. Gao, F. Nie, X. Li, and H. Huang, "Multi-view subspace clustering," *Proc. IEEE International Conference on Computer Vision*, 2015.
[55] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 59-70, 2007.
[56] G. Griffin, A. Holub, and P. Perona, "The Caltech 256," *Caltech Technical Report*, 2007.
[57] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception architecture for computer vision," http://arxiv.org/abs/1512.00567v1.
[58] D. Spielman, and S. Teng, "Nearly-linear time algorithm for graph partitioning, graph sparsification, and solving linear systems," *Proc. ACM Symposium on Theory of Computing*, 2004.
[59] X. Wang, X. Guo, Z. Lei, C. Zhang, and S. Z. Li, "Exclusivity-consistency regularized multi-view subspace clustering," *Proc. IEEE Conference Computer Vision and Pattern Recognition*, 2017.

Yuan Xie (M'12) received the Ph.D. degree in Pattern Recognition and Intelligent Systems from the Institute of Automation, Chinese Academy of Sciences (CAS), in 2013.

He is currently an associated professor with the Research Center of Precision Sensing and Control, Institute of Automation, CAS. His research interests include image processing, computer vision, machine learning and pattern recognition. He has published around 30 papers in major international journals including the IJCV, IEEE TIP, TNNLS, TCYB, TCSVT, TGRS, TMM, etc. He also has served as a reviewer for more than 15 journals and conferences. Dr. Xie received the Hong Kong Scholar Award from the Society of Hong Kong Scholars and the China National Postdoctoral Council in 2014.

Wensheng Zhang received the Ph.D. degree in Pattern Recognition and Intelligent Systems from the Institute of Automation, Chinese Academy of Sciences (CAS), in 2000. He is a Professor of Machine Learning and Data Mining and the Director of Research and Development Department, Institute of Automation, CAS. His research interests include computer vision, pattern recognition, artificial intelligence and computer human interaction. Email: wensheng.zhang@ia.ac.cn.

Yanyun Qu (M'12) received the B.S. and the M.S. degrees in Computational Mathematics from Xiamen University and Fudan University, China, in 1995 and 1998, respectively, and received the Ph.D. degrees in Automatic Control from Xian Jiaotong University, China, in 2006. She joined the faculty of Department of Computer Science in Xiamen University since 1998. She is currently a professor in School of Information Science and Engineer of Xiamen University. She is a member of IEEE and a member of ACM. Her current research interests include pattern recognition, computer vision, and image processing etc. Email: yyqu@xmu.edu.cn.

Longquan Dai is currently an Assistant Professor in Intelligent Media Analysis Group (IMAG) at School of Computer Science and Engineering, Nanjing University of Science and Technology (NJUST). He received his B.A., M.S. and Ph.D. degrees from Henan University of Technology (HAUT) in 2006, Shantou Univeristy (STU) in 2010 and National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA) in 2016, respectively. His current research interests lie in computer graphics, computer vision, and optimization-based techniques for image analysis and synthesis.

Dacheng Tao (F'15) is Professor of Computer Science and ARC Laureate Fellow in the School of Information Technologies and the Faculty of Engineering and Information Technologies, and the Inaugural Director of the UBTECH Sydney Artificial Intelligence Centre, at the University of Sydney. He mainly applies statistics and mathematics to Artificial Intelligence and Data Science. His research results have expounded in one monograph and 200+ publications at prestigious journals and prominent conferences, such as IEEE T-PAMI, T-IP, T-NNLS, IJCV, JMLR, NIPS, ICML, CVPR, ICCV, ECCV, ICDM; and ACM SIGKDD, with several best paper awards, such as the best theory/algorithm paper runner up award in IEEE ICDM07, the best student paper award in IEEE ICDM13, the distinguished paper award in the 2018 IJCAI, the 2014 ICDM 10-year highest-impact paper award, and the 2017 IEEE Signal Processing Society Best Paper Award. He is a Fellow of the Australian Academy of Science, AAAS, IEEE, IAPR, OSA and SPIE.