Northumbria Research Link

Multiview Subspace Clustering Using Low-rank Representation

Jie Chen, Shengxiang Yang, *Senior Member, IEEE,* Hua Mao, *Member, IEEE*, Conor Fahy

Abstract—Multiview subspace clustering is one of the most widely used methods for exploiting the internal structures of multiview data. Most previous studies have performed the task of learning multiview representations by individually constructing an affinity matrix for each view without simultaneously exploiting the intrinsic characteristics of multiview data. In this paper, we propose a multiview low-rank representation (MLRR) method to comprehensively discover the correlation of multiview data for multiview subspace clustering. MLRR considers symmetric low-rank representations (LRRs) to be an approximately linear spatial transformation under the new base, i.e., the multiview data themselves, to fully exploit the angular information of the principal directions of LRRs, which is adopted to construct an affinity matrix for multiview subspace clustering, under a symmetric condition. MLRR takes full advantage of LRR techniques and a diversity regularization term to exploit the diversity and consistency of multiple views, respectively, and this method simultaneously imposes a symmetry constraint on LRRs. Hence, the angular information of the principal directions of rows is consistent with that of columns in symmetric LRRs. The MLRR model can be efficiently calculated by solving a convex optimization problem. Moreover, we present an intuitive fusion strategy for symmetric LRRs from the perspective of spectral clustering to obtain a compact representation, which can be shared by multiple views and comprehensively represents the intrinsic features of multiview data. Finally, the experimental results based on benchmark datasets demonstrate the effectiveness and robustness of MLRR compared with several state-of-the-art multiview subspace clustering algorithms.

Index Terms—Multiview data, low-rank representation, sub-space clustering, spectral clustering

I. INTRODUCTION

Subspace clustering is an effective tool to explore the underlying structures of high-dimensional data [3], [13], [28], [37]. It shows significant compactness in developing applications involving high-dimensional data, such as image representation [50], motion segmentation [7] and multimedia understanding [44]. It is well known that high-dimensional data often have a

J. Chen is with the College of Computer Science, Sichuan University, Chengdu 610065, China (E-mail: chenjie2010@scu.edu.cn).

S. Yang is with the School of Computer Science and Informatics, De Montfort University, Leicester LE1 9BH, U.K. (e-mail: syang@dmu.ac.uk).

H. Mao is with the Department of Computer and Information Sciences, Northumbria University, Newcastle, NE1 8ST, U.K. (e-mail: hua.mao@northumbria.ac.uk).

C. Fahy is with the School of Computer Science and Informatics, De Montfort University, Leicester LE1 9BH, U.K. (e-mail: conor.fahy@dmu.ac.uk).

smaller intrinsic dimension [1], [6], [56]. We assume that high-dimensional data are approximately drawn from multiple low-dimensional subspaces. Therefore, the objective of subspace clustering addresses the problem of clustering data into their respective subspaces.

A number of subspace clustering methods aim to identify groups and achieve impressive performance for single-view data [22]. For example, spectral clustering-based methods and their variants are widely used clustering methods [13], [18], [28], [33]. However, many machine learning problems deal with data involving several different representations or views from multiple sources of features, where each representation or view represents a type of feature set [26], [61]. Each of these individual views may be insufficient to perform a clustering task. Multiple views providing complementary information can improve the performance of clustering tasks. For example, the features of a handwritten image can be simultaneously described by multiple styles, such as with different rotations, translations, and thicknesses. These different features can provide more differentiable information from different styles of handwritten images and can help improve the clustering performance in image information retrieval.

The goal of multiview subspace clustering is to simultaneously cluster data into respective clusters using multiview feature sets, where each cluster corresponds to a low-dimensional subspace [45], [53], [57], [60], [63].Multiview data consist of different views and can be represented by multiple distinct feature sets. The consistency of different feature sets indicates that common information is shared among different views. Different features can provide useful diversity information extracted from different views. Hence, one of the main challenges in multiview subspace clustering is to seek a consensus feature representation that is able to effectively exploit the consistency and diversity among different views. Additionally, graph construction is an effective way to evaluate the memberships between features in multiview subspace clustering [43], [59]. For instance, Tang *et al.* presented a cross-view graph diffusion method that was parameter free to effectively learn a unified graph for multiview data clustering [43]. Zhan *et al.* presented a graph learning-based multiview clustering algorithm to learn a global graph into which the initial graphs are integrated with the rank constraint on the Laplacian matrix [59]. In particular, the mutual relationship between pairwise data samples should be equivalent in graph construction [23], [5], [19]. Jia *et al.* used symmetric nonnegative matrix factorization to learn a similarity graph without additional postprocessing for subspace clustering [19].

A variety of multiview representation learning approaches

have been presented to exactly solve the problem of learning a shared representation of multiple views [25], [42], [44], [55], [64]. The existing representative multiview representation learning approaches, such as multimodal probabilistic-based models [9] and neural network-based models [15], [34], effectively integrate the distinct features learned from multiple different views into a compact representation. In general, the multimodal probabilistic-based models, such as multiview topic learning and multiview latent space Markov networks [9], consider the problem of multiview feature learning as seeking a compact set of latent random variables that estimate a distribution for observed multiview data. Specifically, these models construct probabilistic models of feature representation using the probabilistic distributions of multiview data. In addition, neural network-based models, such as multimodal autoencoders [34] and multiview convolutional neural networks [34], usually construct a successive network layer structure to learn the respective features for individual views and then fuse these features into a shared multiview representation. However, these methods lack the ability to exploit the intrinsic low-dimensional structures of high-dimensional data. In fact, high-dimensional data of multiple views often imply low-dimensional subspaces. Therefore, the multiple subspace structure of high-dimensional data should be taken into account when exploiting the complementary knowledge contained in multiview data.

Inspired by the success of l_1-norm and nuclear norm techniques, two major categories of data representation learning theory have recently been introduced to learn feature representations of high-dimensional data, i.e., sparse representation and low-rank representation (LRR) [3], [38], [40]. Sparse subspace clustering (SSC) and LRR are two representative methods of data representation that uncover the underlying structure of high-dimensional data for subspace clustering [13], [28]. SSC uses the sparsest representation of data points by solving the l_1-norm minimization problem. This method shows its ability to capture the local structure of high-dimensional data. LRR imposes a low-rank constraint on data points to find subspace structures, which can be solved by the convex optimization problem of nuclear norm minimization. This method is able to effectively capture the global structure of high-dimensional data. These methods show impressive clustering performance in subspace clustering but mainly focus on learning feature representations from a single view rather than multiple views. Several impressive extensions of LRR techniques have been proposed to discover the manifold structure embedded in high-dimensional data [10], [30], [31]. These extensions provide a discriminative representation that preserves projection learning for image recognition and obtains competitive classification results.

In recent years, data representation learning theories have been widely extended to learn the joint feature representations of multiview data for multiview subspace clustering [2], [11], [29], [32], [47], [58]. By introducing structured sparsity constraints on a weight matrix, the work in [2] proposed a multiview learning model to integrate all features and learn the joint weight matrix for an individual feature representation with respect to each cluster. Although the $l_{2,1}$-norm encourages the columns of a joint weight matrix to be zero, the fusion mechanism of multiview features that adopts the characteristic of structured sparsity for the matrix is ambiguous. In [35], a multiview learning model, named multiview learning with adaptive neighbors (MLAN), was proposed to find complex structures of high-dimensional data and learn relationships among multiple views. MLAN has the ability to perform multiview clustering and local manifold structure learning simultaneously. However, this model is insufficient for capturing the global structure of high-dimensional data. In addition, Wang proposed an iterative low-rank-based structured optimization (ILRSO) method to capture the local manifold structure of high-dimensional data and obtain agreement among multiple views for multiview spectral clustering [52]. Moreover, Wang further presented a structured low-rank matrix factorization (SLRMF) method by factorizing LRR as a clustered symmetric matrix [51]. Hence, significant characteristics of LRR have been developed for extracting essential features of high-dimensional data, especially in evaluating the membership of data samples, and should be further researched.

In this paper, we present a multiview low-rank representation (MLRR) model to comprehensively discover the correlation of multiview data for multiview subspace clustering. Generally, different views contain complementary information. MLRR adopts LRR techniques to explore the diversity of multiple views, i.e., the different intrinsic low-rank structures of multiview data. Such differences in low-rank structures enhance the diversity of multiple views to capture the inherent differences among individual views. In addition, MLRR introduces a diversity regularization term to guarantee the consistency of multiple views. We simultaneously consider a symmetry constraint imposed on the MLRR model to improve its learning ability of complementary knowledge. Moreover, each multiview feature representation shows distinct degrees of importance in the fusion of feature representations. Consequently, MLRR presents an intuitive fusion strategy of symmetric LRRs to obtain a compact representation from the perspective of spectral clustering. The compact representation is shared by multiple views to comprehensively represent the intrinsic features of multiview data. As multiview data are often contaminated by noise and outlying entries, an affinity matrix can be finally constructed by calculating the angular information of the principal directions of symmetric LRRs.

The main contributions of this paper are summarized below:
1) The MLRR model effectively exploits the intrinsic low-dimensional structures of multiple views by an LRR and a diversity regularization term under a symmetric constraint.
2) The MLRR model considers symmetric LRRs to be an approximately linear spatial transformation under the new base, i.e., the multiview data themselves, to fully exploit the angular information of the principal directions of LRRs under a symmetric condition.
3) The fusion mechanism of symmetric LRRs is presented to pursue a compact representation of multiple views from the perspective of spectral clustering.
4) The MLRR model can be efficiently calculated by solving

a convex optimization problem, and a rigorous convergence proof for MLRR is provided.

5) Extensive experimental results obtained from benchmark datasets demonstrate the effectiveness and robustness of MLRR for multiview subspace clustering.

The rest of this paper is organized as follows. Related work on LRR and multi-view subspace clustering are summarized in Section II. Section III presents our proposed approach in detail. Extensive experimental results and discussions are shown in Section IV. Finally, conclusions are drawn in Section V.

II. RELATED WORK

This section provides a brief review of LRR techniques and some related work on multiview subspace clustering.

A. Low-rank Representation

Let $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2..., \mathbf{y}_n] \in \mathbb{R}^{d \times n}$ be a set of d-dimensional data vectors. We assume that each data vector is drawn from the union of k linear subspaces $\{\mathbf{S}_i\}_{i=1}^{k}$ of unknown dimensions. To find an LRR $\mathbf{Z} \in \mathbb{R}^{n \times n}$, LRR solves the following rank minimization problem:

$$\min_{\mathbf{Z}} rank(\mathbf{Z}) \qquad s.t. \qquad \mathbf{Y} = \mathbf{D}\mathbf{Z}, \tag{1}$$

where $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_n] \in \mathbb{R}^{d \times n}$ is a given dictionary used to linearly represent the data vectors [28].

Since Problem (1) is nonconvex and NP-hard, LRR uses the nuclear norm as a common surrogate for the rank function to attain a convex solution as follows:

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_* \qquad s.t. \qquad \mathbf{Y} = \mathbf{D}\mathbf{Z}, \tag{2}$$

where $\|\mathbf{Z}\|_*$ denotes the nuclear norm of matrix $\mathbf{Z}$. The above optimization problem can be efficiently solved by singular value decomposition (SVD). The solution $\mathbf{Z}$ of Problem (2) is considered the set of LRR features of $\mathbf{Y}$. In a single view, the affinity matrix constructed as $|\mathbf{Z}| + |\mathbf{Z}|^T$ can be used for spectral clustering [41] to obtain the final clustering result of $\mathbf{Y}$.

B. Multiview Subspace Clustering

Suppose that each data vector $\mathbf{x}_i$ has n_v views in $\mathbf{X}$. Without loss of generality, we have $\mathbf{X} = [\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, ..., \mathbf{X}^{(n_v)}]$, where $\mathbf{X}^v \in \mathbb{R}^{d \times n}$ consists of the vectors of $\mathbf{x}_i^v$ in the v-th view. The purpose of multiview representation learning is to learn comprehensive representations of multiview data, and this approach is closely related to the expressive ability of data representation models. The fused representation of multiple views is critical for multiview subspace clustering, and ILRSO methods are typical examples of multiview subspace clustering involving LRR techniques from feature representation [51], [52], [54].

To exploit the complementary information among multiple views, RMSC adopts a low-rank and sparse decomposition strategy for the feature fusion of multiple views [54]. This algorithm first constructs affinity matrices $\mathbf{S}_i(i = 1, 2, ..., n_v)$ for the multiple views and then calculates the corresponding

transition probability matrix $\mathbf{P}_i = (\mathbf{D}_i)^{-1}\mathbf{S}_i$ for each single view, where $\mathbf{D_i}$ is a diagonal matrix with $d_k = \sum_{j=1}^{n} \mathbf{S}_{kj}$. The algorithm further imposes a low-rank constraint on the transition probability matrix $\mathbf{P}$ to implement feature fusion by minimizing the following objective function:

$$\min_{\mathbf{P},\{\mathbf{E}_i\}_{i=1}^{n_v}} \|\mathbf{P}\|_* + \lambda \sum_{i=1}^{n_v} (\|\mathbf{E}_i\|_1)$$
$$s.t. \quad \mathbf{P}_i = \mathbf{P} + \mathbf{E}_i, \mathbf{P} \geq 0, \mathbf{P}\mathbf{1} = \mathbf{1}, i = 1, 2, ..., n_v, \tag{3}$$

where $\lambda > 0$ is a parameter. Finally, RMSC uses the shared low-rank transition probability matrix as an input for spectral clustering via Markov chains [62].

By combining low-rank and sparse decomposition, ILRSO was proposed to explicitly learn a shared representation in multiview spectral clustering [52]. By utilizing the differences in low-rank and sparse representations from different views, the objective function of the structured LRR is formulated as:

$$\min_{\{\mathbf{Z}_i\}_{i=1}^{n_v},\{\mathbf{E}_i\}_{i=1}^{n_v}} \sum_{i \in \mathbf{V}} f(\mathbf{Z}_i, \mathbf{E}_i)$$
$$s.t \quad \mathbf{X}_i = \mathbf{X}_i\mathbf{Z}_i + \mathbf{E}_i, \quad \mathbf{Z}_i \geq 0, \quad i = 1, ..., n_v, \tag{4}$$

where λ_1, λ_2, λ_3 and β are parameters and

$$f(\mathbf{Z}_i, \mathbf{E}_i) = \|\mathbf{Z}_i\|_* + \lambda_1\|\mathbf{E}_i\|_1 + \lambda_2\|\mathbf{Z}_i\|_1$$
$$+ \lambda_3 Tr(\mathbf{Z}_i^T \mathbf{L}_i \mathbf{Z}_i) + \frac{\beta}{2} \sum_{j \in \mathbf{V}, j \neq i} \|\mathbf{Z}_i - \mathbf{Z}_j\|_2^2. \tag{5}$$

The above optimization problem can be solved by an alternating direction method with adaptive penalty. The shared representation of multiple views can be obtained by fusing individual LRRs of multiple views as follows:

$$\mathbf{W} = \frac{1}{n_v} \sum_{i=1}^{n_v} \mathbf{W}_i, \tag{6}$$

where $\mathbf{W}_i(j, k) = \frac{1}{2}(\mathbf{Z}_i(j, k) + \mathbf{Z}_i(k, j))$. As a variant of ILRSO, SLRMF employs clustered symmetric matrix factorization instead of low-rank minimization, i.e.,

$$\|\mathbf{Z}_i\|_* = \min_{\mathbf{U}_i, \mathbf{Z}_i = \mathbf{U}_i} \mathbf{U}_i^T \frac{1}{2} \|\mathbf{U}_i\|_F^2, \tag{7}$$

where $\mathbf{U}_i$ represents the factorized low-rank data-cluster representation of $\mathbf{X}_i$ for the i-th view [51]. To avoid the computational complexity of $O(n^3)$ caused by computing the inverse matrix, an efficient updating strategy of updating a low-rank data-cluster representation of high-dimensional data is introduced in SLRMF.

III. MULTIVIEW LOW-RANK REPRESENTATION FOR SUBSPACE CLUSTERING

This section presents an MLRR algorithm for multiview subspace clustering. We first propose an MLRR model for learning comprehensive representations that enhances diversity and explores consistency. Then, we introduce a symmetric constraint into LRR to perform an approximately linear spatial transformation of multiview data. Third, we present a fusion strategy of LRRs to obtain a compact representation, which

exhibits a great ability to capture complementary information contained in multiple views. The angular information of the principal directions of symmetric LRRs is further exploited to construct a final affinity matrix, which enriches the relationship among the data samples. Finally, we discuss the convergence properties and computational complexity of MLRR.

A. The Diversity and Consistency of Different Views in MLRR

Consider the given multiview data $\mathbf{X} = [\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, ..., \mathbf{X}^{(n_v)}]$ with n observations obtained from n_v views, each view of which consists of vectors $\{\mathbf{x}_i^v\}_{i=1}^n$. We assume that the majority of vectors are approximately drawn from a union of multiple low-dimensional subspaces. This assumption is reasonable because the intrinsic structures of high-dimensional data often have a smaller dimension. Multiple views usually describe the different types of features of multiview data in various aspects, where each view contains particular information. The diversity of different views indicates that these views contain complementary information. This implies that multiple views have their own approximate low-dimensional structures of multiple views with noise and corrupted observations. In other words, the different low-dimensional structures of multiple views can be used to characterize the diversity of the multiple views. Consequently, learning separate LRRs from multiview subspaces indicates that the separate information from different views should be contained in a self-representation with low-rank constraints. Specifically, a general optimization problem is considered as follows:

$$\min_{\{\mathbf{Z}_i,\mathbf{E}_i\}_{i=1}^{n_v}} \sum_{i=1}^{n_v} f(\mathbf{X}_i, \mathbf{Z}_i) + \lambda \sum_{i=1}^{n_v} g(\mathbf{X}_i, \mathbf{E}_i), \qquad (8)$$

where $f(\cdot)$ represents a strategy of exploiting the separate low-rank structures of the original views, $g(\cdot)$ characterizes a certain type of corruption, and the parameter $\lambda > 0$ balances the effects of the LRR term and the corruption regularization term. Hence, $f(\cdot)$ with low-rank constraints is adopted to enhance the diversity of multiple views, i.e., the intrinsic low-dimensional structures of the multiple views.

In addition, each view usually shares some common information that characterizes feature correlations among the multiple views. Each LRR characterizes the same multidata from different views. If the multidata are a collection of data vectors that are strictly drawn from a union of multiple subspaces, the fact that each LRR matrix $\mathbf{Z}_i$ is block-diagonal can be proven rigorously [28]. As a result, it is reasonable that each LRR shares the consistency of multiple views.

Suppose we have multiview data consisting of two views, i.e., $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$, which have two corresponding LRRs $\mathbf{Z}^{(i)}$ and $\mathbf{Z}^{(j)}$, respectively. We propose the following cost function as a measure of the consistency between the two views:

$$c(\mathbf{X}^{(i)}, \mathbf{X}^{(j)}) = tr\left(\left(\mathbf{Z}^{(i)} - \mathbf{Z}^{(j)}\right)\left(\mathbf{Z}^{(i)} - \mathbf{Z}^{(j)}\right)^T\right), \quad (9)$$

where tr denotes the matrix trace. A smaller value of the affinity $c(\mathbf{X}^{(i)}, \mathbf{X}^{(j)})$ indicates that the i-th view and the j-th view are more similar. Clearly, the degree of diversity

decreases to zero when the two views are consistent with each other. However, usually, there are no two identical views in multiview data. Hence, we encourage the pairwise similarities of each pair of multiview LRRs to be closer.

Considering the diversity of different views, we add a diversity regularization term to Problem (8) as follows:

$$\min_{\{\mathbf{Z}_i,\mathbf{E}_i\}_{i=1}^{n_v}} \sum_{i=1}^{n_v} f(\mathbf{X}_i, \mathbf{Z}_i) + \lambda \sum_{i=1}^{n_v} g(\mathbf{X}_i, \mathbf{E}_i) + \eta \sum_{i,j=1,i\neq j}^{n_v} c\left(\mathbf{X}^{(i)}, \mathbf{X}^{(j)}\right), \qquad (10)$$

where η is a scalar constant. Hence, $f(\cdot)$ and $c(\cdot)$ characterize the diversity and consistency of different views, respectively.

B. Symmetry Constraint

Generally, an LRR can be considered a linear combination of vectors in a single view. The LRR effectively measures the relationship of the weights among the vectors from the perspective of matrix factorization. In many previous multiview subspace clustering methods, $\mathbf{Z}_{ij}^{(v)}$ and $\mathbf{Z}_{ji}^{(v)}$ represent the similarity measures of vectors x_i^v and x_j^v in the v-th view after a symmetrization postprocessing step. Actually, the mutual relationship between x_i and x_j should be consistent from the perspective of graph theory. However, the contribution of x_i^v and x_j^v may not be equal to that of x_i^v and x_j^v in practice. These works directly adopt the LRR as the affinity matrix for spectral clustering [4], [28], [49]. They ignore the intrinsic characteristics preserved in LRRs, i.e., the angular information of the principal directions of LRRs, since $\mathbf{Z}^{(v)}$ is low-rank. Consequently, $\mathbf{Z}^{(v)}$ is considered to be an approximate result of linear spatial transformation under the new base $\mathbf{X}^{(v)}$. In addition, the angular information of the principal directions of rows should be consistent with that of columns in LRRs. To address this issue, we employ a symmetry scheme where the strategy is to impose a symmetry constraint on multiview LRRs. This scheme ensures the final weight consistency for each pair of data samples. By integrating the symmetry constraint into Problem (10), we formulate our final MLRR model as follows:

$$\min_{\{\mathbf{Z}^{(v)},\mathbf{E}^{(v)}\}_{v=1}^{n_v}} \sum_{v=1}^{n_v} \left\|\mathbf{Z}^{(v)}\right\|_* + \lambda \sum_{v=1}^{n_v} \left\|\mathbf{E}^{(v)}\right\|_l + \frac{\eta}{2} \sum_{w,v=1,v\neq w}^{n_v} \left\|\mathbf{Z}^{(v)} - \mathbf{Z}^{(w)}\right\|_F^2 \qquad (11)$$

$$s.t. \quad \mathbf{X}^{(v)} = \mathbf{X}^{(v)}\mathbf{Z}^{(v)} + \mathbf{E}^{(v)}, \mathbf{Z}^{(v)} = \left(\mathbf{Z}^{(v)}\right)^T,$$

$$v = 1, 2..., n_v.$$

On the one hand, the symmetry of $\mathbf{Z}^{(v)}$ enables each pairwise element, i.e., $\mathbf{Z}_{ij}^{(v)}$ and $\mathbf{Z}_{ji}^{(v)}$, to become more appropriate when this mathematical optimization problem is solved. On the other hand, the introduction of a symmetry constraint is conducive to improving the ability of the MLRR model to learn the feature correlations among multiple views. Hence, the MLRR model makes full use of symmetric LRRs for exploiting the low-dimensional structures of multiple views.

C. Optimization Strategy

The MLRR model in Problem (11) learns symmetric LRRs from different views. Problem (11) can be solved iteratively by the inexact augmented Lagrange multipliers (ALM) algorithm [27]. In the iterative procedure, the specific parameters updated in the previous stage can be considered constants in the current stage. For instance, the variables $\mathbf{Z}^{(v)}$ $(1 \le v \le n_v)$ can be updated alternately, while the other variables are fixed. Hence, we further consider the following problem:

$$
\min_{\mathbf{Z}^{(v)}, \mathbf{E}^{(v)}} \left\| \mathbf{Z}^{(v)} \right\|_* + \lambda \left\| \mathbf{E}^{(v)} \right\|_l
$$
$$
+ \frac{\eta}{2} \sum_{w=1, w \neq v}^{n_v} \left\| \mathbf{Z}^{(v)} - \mathbf{Z}^{(w)} \right\|_F^2 \tag{12}
$$
$$
s.t. \quad \mathbf{X}^{(v)} = \mathbf{X}^{(v)} \mathbf{Z}^{(v)} + \mathbf{E}^{(v)}, \mathbf{Z}^{(v)} = \left(\mathbf{Z}^{(v)} \right)^T.
$$

The objective function in Problem (12) is nonseparable. By introducing an auxiliary variable $\mathbf{J}$, the objective function in Problem (12) can be converted to the following equivalent problem:

$$
\min_{\mathbf{Z}^{(v)}, \mathbf{J}^{(v)}, \mathbf{E}^{(v)}} \left\| \mathbf{J}^{(v)} \right\|_* + \lambda \left\| \mathbf{E}^{(v)} \right\|_l
$$
$$
+ \frac{\eta}{2} \sum_{w=1, w \neq v}^{n_v} \left\| \mathbf{Z}^{(v)} - \mathbf{Z}^{(w)} \right\|_F^2 \tag{13}
$$
$$
s.t. \quad \mathbf{X}^{(v)} = \mathbf{X}^{(v)} \mathbf{Z}^{(v)} + \mathbf{E}^{(v)},
$$
$$
\mathbf{Z}^{(v)} = \mathbf{J}^{(v)}, \mathbf{J}^{(v)} = \left(\mathbf{J}^{(v)} \right)^T.
$$

The augmented Lagrangian function in Problem (13) is

$$
\min_{\mathbf{Z}^{(v)}, \mathbf{J}^{(v)}, \mathbf{E}^{(v)}, \mathbf{J}^{(v)} = \left(\mathbf{J}^{(v)}\right)^T, \mathbf{Y}_1, \mathbf{Y}_2} \left\| \mathbf{J}^{(v)} \right\|_* + \lambda \left\| \mathbf{E}^{(v)} \right\|_l
$$
$$
+ \frac{\eta}{2} \sum_{w=1, w \neq v}^{n_v} \left\| \mathbf{Z}^{(v)} - \mathbf{Z}^{(w)} \right\|_F^2
$$
$$
+ tr[\mathbf{Y}_1^T \left(\mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} - \mathbf{E}^{(v)} \right)]
$$
$$
+ tr[\mathbf{Y}_2^T \left(\mathbf{Z}^{(v)} - \mathbf{J}^{(v)} \right)]
$$
$$
+ \frac{\mu}{2} \left(\left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} - \mathbf{E}^{(v)} \right\|_F^2 + \left\| \mathbf{Z}^{(v)} - \mathbf{J}^{(v)} \right\|_F^2 \right), \tag{14}
$$

where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are Lagrange multipliers, and $\mu > 0$ is an adaptive penalty parameter. The variables $\mathbf{J}^{(v)}$, $\mathbf{Z}^{(v)}$ and $\mathbf{E}^{(v)}$ can be updated alternately at each step, while the other two variables are fixed. The updating rules at iteration $t + 1$ are formulated as follows:

$$
\mathbf{J}_{t+1}^{(v)} = \underset{\mathbf{J}_{t+1}^{(v)} = \left(\mathbf{J}_{t+1}^{(v)} \right)^T}{\arg \min} \frac{1}{\mu} \left\| \mathbf{J}_{t+1}^{(v)} \right\|_*
$$
$$
+ \frac{1}{2} \left\| \mathbf{J}_{t+1}^{(v)} - \left(\mathbf{Z}_t^{(v)} + \frac{\mathbf{Y}_2^{(v)}}{\mu} \right) \right\|_F^2, \tag{15}
$$

$$
\mathbf{Z}_{t+1}^{(v)} = \left(\left(\mathbf{X}^{(v)} \right)^T \mathbf{X}^{(v)} + \left(\frac{\eta}{\mu} (n_v - 1) + 1 \right) \cdot \mathbf{I} \right)^{-1} \mathbf{H}, \tag{16}
$$

Algorithm 1 Solving Problem (11) by the inexact ALM algorithm

Input: Data matrices $\mathbf{X} = \left\{ \mathbf{X}^{(v)} \right\}_{v=1}^{n_v}$, parameters λ and η

Initialize:
$\left\{ \mathbf{J}^{(v)} = \mathbf{Z}^{(v)} = \mathbf{0} \right\}_{v=1}^{n_v}, \left\{ \mathbf{Y}_1^{(v)} = \mathbf{Y}_2^{(v)} = \mathbf{0} \right\}_{v=1}^{n_v}, \rho = 1.1,$
$\mu_1 = \mu_2 = 10^{-2}, \mu_{\max} = 10^{10}, \varepsilon = 10^{-6}$

1: **while** not converged **do**
2:    **for** $v = 1$ to n_v **do**
3:       Fix others and update $\mathbf{J}^{(v)}$ by solving Problem (15)
4:       Fix others and update $\mathbf{Z}^{(v)}$ by solving Problem (16)
5:       Fix others and update $\mathbf{E}^{(v)}$ by solving Problem (17)
6:       Update the multipliers:
         $\mathbf{Y}_1^{(v)} = \mathbf{Y}_1^{(v)} + \mu_1 \left(\mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} - \mathbf{E}^{(v)} \right)$
         $\mathbf{Y}_2^{(v)} = \mathbf{Y}_2^{(v)} + \mu_2 \left(\mathbf{Z}^{(v)} - \mathbf{J}^{(v)} \right)$
7:    **end for**
8:    Update the parameters μ_1 and μ_2:
      $\mu_1 = \min(\rho \mu_1, \mu_{\max})$;
      $\mu_2 = \min(\rho \mu_2, \mu_{\max})$;
9:    Check the convergence conditions:
10:   **for** $v = 1$ to n_v **do**
11:      **if** $\left\| \mathbf{Z}^{(v)} - \mathbf{J}^{(v)} \right\|_{\max} \le \varepsilon$
         and $\left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}^{(v)} - \mathbf{E}^{(v)} \right\|_{\max} \le \varepsilon$ **then**
12:         break;
13:      **end if**
14:   **end for**
15: **end while**
Output: $\left\{ \mathbf{Z}^{(v)} \right\}_{v=1}^{n_v}$.

$$
\mathbf{E}_{t+1}^{(v)} = \arg \min \lambda \left\| \mathbf{E}_{t+1}^{(v)} \right\|_l
$$
$$
+ \frac{\mu}{2} \left\| \mathbf{E}_{t+1}^{(v)} - \left(\mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{Z}_{t+1}^{(v)} + \frac{\mathbf{Y}_1^{(v)}}{\mu} \right) \right\|_F^2, \tag{17}
$$

where

$$
\mathbf{H} = \left(\mathbf{X}^{(v)} \right)^T \mathbf{X}^{(v)} - \left(\mathbf{X}^{(v)} \right)^T \mathbf{E}_t^{(v)} + \mathbf{J}_{t+1}^{(v)}
$$
$$
+ \frac{1}{\mu} \left(\left(\mathbf{X}^{(v)} \right)^T \mathbf{Y}_1^{(v)} - \mathbf{Y}_2^{(v)} + \eta \sum_{w=1, w \neq v}^{n_v} \mathbf{Z}_t^{(w)} \right). \tag{18}
$$

The first equation in Problem (15) has a closed form solution, which can be solved by Lemma 1 [5]. We assume that some data vectors are corrupted in multiple views. Hence, we adopt the $l_{2,1}$-norm to characterize a certain type of corruption, which encourages the columns of $\mathbf{E}^{(v)}$ to be zero. The last equation in Problem (17) is a convex problem, which can be solved using the $l_{2,1}$-norm minimization operator [28]. The whole procedure for solving Problem (11) is outlined in Algorithm 1.

Lemma 1 *([5]) For any given square matrix* $\mathbf{Z} \in \mathbb{R}^{n \times n}$, *the minimizer of the following optimization problem is unique,*

$$
\mathbf{J}^* = \arg \min_{\mathbf{J}} \frac{1}{\mu} \|\mathbf{J}\|_* + \frac{1}{2} \|\mathbf{J} - \mathbf{Z}\|_F^2, \mathbf{J} = \mathbf{J}^T, \tag{19}
$$

where μ is a parameter. The unique solution of the above problem has the following closed form

$$\mathbf{J}^* = \mathbf{U}_r(\boldsymbol{\Sigma}_\mathbf{r} - \frac{1}{\mu} \cdot \mathbf{I}_r)\mathbf{V}_r^T, \qquad (20)$$

where $\widetilde{\mathbf{Z}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ is an SVD of the symmetric matrix $\widetilde{\mathbf{Z}} = (\mathbf{Z} + \mathbf{Z}^T)/2$, $\boldsymbol{\Sigma}_r = diag(\delta_1, \delta_2, ..., \delta_r)$ where $\{r : \delta_r > \frac{1}{\mu}\}$ are strictly positive singular values, $\mathbf{U}_r$ and $\mathbf{V}_r$ are the corresponding singular vectors of matrix $\widetilde{\mathbf{Z}}$, and $\mathbf{I}_r$ is an $r \times r$ identity matrix.

D. LRR Fusion for Multiview Clustering

Multiple views usually share some common information, which can be represented by a compact representation. To fully characterize the feature correlations of different views, we present the fusion mechanism of LRRs learned from different views from the spectral clustering perspective [41].

To efficiently integrate the features from multiple views, we first consider the following general spectral clustering model for multiple views:

$$\min_{\mathbf{F}} \sum_{v=1}^{n_v} Tr\left(\mathbf{F}^T\mathbf{L}^{(v)}\mathbf{F}\right) \qquad s.t. \qquad \mathbf{F}^T\mathbf{F} = \mathbf{I}, \qquad (21)$$

where $\mathbf{F}$ is the clustering indicator matrix, $\mathbf{L}^{(v)} = \left(\mathbf{D}^{(v)}\right)^{-\frac{1}{2}}\left(\mathbf{D}^{(v)} - \mathbf{Z}^{(v)}\right)\left(\mathbf{D}^{(v)}\right)^{-\frac{1}{2}}$ is the Laplacian matrix, and $\mathbf{D}^{(v)}$ is a diagonal matrix with diagonal elements $\mathbf{D}_{ii}^{(v)} = \sum_j \mathbf{Z}_{ij}^{(v)}$. Theorem 1 shows the fusion mechanism of the Laplacian matrices $\mathbf{L}^{(v)}$ $(v = 1, 2, ..., n_v)$.

Theorem 1 *Let $\mathbf{F}$ be an optimal solution of Problem* (21). *There exists a final fusion result of $\mathbf{L}^{(v)}$ as follows:*

$$\mathbf{L} = n_v \cdot \mathbf{I} - \sum_{v=1}^{n_v}\left(\mathbf{D}^{(v)}\right)^{-\frac{1}{2}}\mathbf{Z}^{(v)}\left(\mathbf{D}^{(v)}\right)^{-\frac{1}{2}},$$

where $\mathbf{Z}^{(v)}$ represents the single representation of the v-th view $(v = 1, 2, ..., n_v)$.

Proof *By linear algebra techniques, Problem* (21) *can be converted to the following problem:*

$$\min_{\mathbf{F}} Tr\left(\mathbf{F}^T\sum_{v=1}^{n_v}\left(\mathbf{L}^{(v)}\right)\mathbf{F}\right) \qquad s.t. \qquad \mathbf{F}^T\mathbf{F} = \mathbf{I}.$$

The above problem can be equivalently rewritten as the standard spectral clustering model as follows:

$$\min_{\mathbf{F}} Tr\left(\mathbf{F}^T\mathbf{L}\mathbf{F}\right) \qquad s.t. \qquad \mathbf{F}^T\mathbf{F} = \mathbf{I}, \qquad (22)$$

where $\mathbf{L} = \sum_{v=1}^{n_v}\left(\mathbf{L}^{(v)}\right)$ is a Laplacian matrix.

Algorithm 2 The MLRR algorithm

Input: Data matrix $\mathbf{X} = \left\{\mathbf{X}^{(v)}\right\}_{v=1}^{n_v}$, the number of subspaces k, parameters λ, η and $\alpha > 0$

1: Solve Problem (11) by Algorithm 1, and obtain the optimal solution $\left\{\mathbf{Z}^{(v)}\right\}_{v=1}^{n_v}$.

2: Calculate the sum of $\left\{\mathbf{Z}^{(v)}\right\}_{v=1}^{n_v}$: $\mathbf{Z}^* = \sum_{v=1}^{n_v}\mathbf{Z}^{(v)}$

3: Compute the skinny SVD $\mathbf{Z}^* = \mathbf{U}^*\sum^*(\mathbf{V}^*)^T$.

4: Calculate $\mathbf{M} = \mathbf{U}^*(\sum^*)^{1/2}$.

5: Construct the affinity matrix $\mathbf{W}$, i.e.,

$$[\mathbf{W}]_{ij} = \left(\frac{\mathbf{m}_i^T\mathbf{m}_j}{\|\mathbf{m}_i\|_2\|\mathbf{m}_j\|_2}\right)^{2\alpha}.$$

6: Compute the eigenvectors $\{\mathbf{v}_i\}_{i=1}^k \in \mathbb{R}^n$ corresponding to the smallest k eigenvalues of the Laplacian matrix $\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}\left(\mathbf{D} - \mathbf{W}\right)\mathbf{D}^{-\frac{1}{2}}$.

7: Let each point $\mathbf{y}_i \in \mathbb{R}^k$ $(i = 1...n)$ be the vector corresponding to the i−th row of $\mathbf{V} = \{v_i\}_{i=1}^k \in \mathbb{R}^{n \times k}$

8: Cluster the points $\{\mathbf{y}_i\}_{i=1}^n$ by the k−means algorithm into clusters $\{\mathbf{C}_i\}_{i=1}^k$.

Output: The clustering results.

Furthermore, we obtain a final fusion result of $\mathbf{L}^{(v)}$ as follows:

$$\begin{aligned}\mathbf{L} &= \sum_{v=1}^{n_v}\left(\mathbf{D}^{(v)}\right)^{-\frac{1}{2}}\left(\mathbf{D}^{(v)} - \mathbf{W}^{(v)}\right)\left(\mathbf{D}^{(v)}\right)^{-\frac{1}{2}} \\ &= \sum_{v=1}^{n_v}\left(\mathbf{I} - \left(\mathbf{D}^{(v)}\right)^{-\frac{1}{2}}\mathbf{W}^{(v)}\left(\mathbf{D}^{(v)}\right)^{-\frac{1}{2}}\right) \\ &= n_v \cdot \mathbf{I} - \sum_{v=1}^{n_v}\left(\mathbf{D}^{(v)}\right)^{-\frac{1}{2}}\mathbf{W}^{(v)}\left(\mathbf{D}^{(v)}\right)^{-\frac{1}{2}}.\end{aligned}$$

$\square$

According to Theorem 1, we can construct the new Laplacian matrix $\mathbf{L}^{'}$ as follows:

$$\begin{aligned}\mathbf{L} &> \mathbf{I} - (\mathbf{D})^{-\frac{1}{2}}\left(\sum_{v=1}^{n_v}\mathbf{Z}^{(v)}\right)(\mathbf{D})^{-\frac{1}{2}} \\ &= \mathbf{D}^{-\frac{1}{2}}\left(\mathbf{D} - \mathbf{Z}^*\right)\mathbf{D}^{-\frac{1}{2}} \\ &= \mathbf{L}^{'}.\end{aligned} \qquad (23)$$

This approach further motivates an intuitive fusion strategy of symmetric LRRs of multiple views. Specifically, we calculate the sum of LRRs of multiple views, i.e., $\mathbf{Z}^* = \sum_{v=1}^{n_v}\mathbf{Z}^{(v)}$, which represents the fusion representation of multiple views.

We then construct an affinity matrix, which measures the relationship among the samples in multiple views. Considering a given matrix of LRR $\mathbf{Z}^*$, the angular information of the principal directions of any two low-rank vectors approximately drawn from the same subspace has a higher value than that of those drawn from different subspaces. Specifically, we consider $\mathbf{Z}^*$ with the skinny SVD $\mathbf{U}^*\sum^*(\mathbf{V}^*)^T$. We further use the angular information from all of the row vectors of

matrix $\mathbf{M}$, i.e., $\mathbf{M} = \mathbf{U}^*(\sum^*)^{1/2}$, to define an affinity matrix $\mathbf{W}$ as follows:

$$[\mathbf{W}]_{ij} = \left(\frac{\mathbf{m}_i^T \mathbf{m}_j}{\|\mathbf{m}_i\|_2 \|\mathbf{m}_j\|_2} \right)^{2\alpha}, \qquad (24)$$

where $\mathbf{m}_i$ and $\mathbf{m}_j$ denote the i-th and j-th rows of matrix $\mathbf{M}$ and α ensures that each value of the affinity matrix $\mathbf{W}$ is positive for subspace clustering [5]. As $\mathbf{Z}^*$ is symmetric, we can obtain the same affinity matrix $\mathbf{W}$ using the angular information of the column vectors of matrix $\mathbf{M}$. Consequently, the introduction of the symmetry constraint imposed on LRRs of multiple views allows us to effectively evaluate the membership of data samples in multiview data. Algorithm 2 summarizes the complete multiview subspace clustering algorithm of MLRR. MLRR integrates symmetric LRRs into a compact representation and makes full use of spectral clustering techniques for multiview subspace clustering.

E. Analysis of MLRR

1) Convergence Analysis: The inexact variation of the ALM algorithm has been extensively studied [27]. Although the convergence property of a general ALM algorithm is still an open issue, we provide a rigorous convergence proof for MLRR. Algorithm 1 includes three critical steps, i.e., iteratively updating $\mathbf{J}$, $\mathbf{Z}$ and $\mathbf{E}$, in each iterative computation. Each step of Algorithm 1 has a closed solution. There are two convergence conditions in Algorithm 1, i.e., $\left\| \mathbf{Z}^{(v)} - \mathbf{J}^{(v)} \right\|_{\max} \leq \varepsilon$ and $\left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{Z}^{(v)} - \mathbf{E}^{(v)} \right\|_{\max} \leq \varepsilon$, where $\|\mathbf{A}\|_{\max}$ represents the maximum absolute value of all the entries in a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, i.e., $\|\mathbf{A}\|_{\max} = \max\limits_{1 \leq i \leq m, 1 \leq j \leq n} |a_{ij}|$. To analyze the convergence of Algorithm 1, we focus on these two convergence conditions in the $(t+1)$-th iteration. Theorem 2 shows the convergence of Algorithm 1.

Theorem 2 *Let* $\mathbf{Q} = \mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{Z}_{t+1}^{(v)} - \mathbf{E}_t^{(v)}$ *and* $q_{\max} = \|\mathbf{Q}\|_{\max}$, *where each entry* $q \geq 0$ *in* $\mathbf{Q}$. *Then, let* $q_{\max}$ *be an entry in* $\mathbf{Q}$ *and* $x_{ij}^{(v)}$ *be any of the corresponding entries in* $\mathbf{X}^{(v)}$ *during* $\left(\mathbf{X}^{(v)}\right)^T \mathbf{Q}$. *In Algorithm 1, two convergence conditions, i.e.,* $\left\| \mathbf{Z}^{(v)} - \mathbf{J}^{(v)} \right\|_{\max} \leq \varepsilon$ *and* $\left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{Z}^{(v)} - \mathbf{E}^{(v)} \right\|_{\max} \leq \varepsilon$, *will be satisfied eventually as* t *increases if it satisfies the following conditions, i.e.,*

$$x_{ij}^{(v)} > 0 \text{ and } \forall p \in [1, d], \forall q \in [1, n], p \neq i, q \neq j, x_{pq}^{(v)} \geq 0$$

in $\mathbf{X}^{(v)}$, *where* t *represents the number of iterations and* ε *is a small constant, e.g.,* $\varepsilon = 10^{-6}$.

Proof *We first consider the first convergence condition, i.e.,* $\left\| \mathbf{Z}_{t+1}^{(v)} - \mathbf{J}_{t+1}^{(v)} \right\|_{\max} \leq \varepsilon$. *According to Step 3 of Algorithm 1 and Lemma 1, we have*

$$\mathbf{J}_{t+1}^{(v)} = \mathbf{U}_r^{(v)}(\mathbf{\Sigma}_r - \frac{1}{\mu} \cdot \mathbf{I}_r)\left(\mathbf{V}_r^{(v)}\right)^T,$$

where $\widetilde{\mathbf{Z}_t^{(v)}} = \mathbf{U}_r^{(v)}\mathbf{\Sigma}\left(\mathbf{V_r^{(v)}}\right)^\mathbf{T}$ *is the skinny SVD of the symmetric matrix* $\widetilde{\mathbf{Z}_t^{(v)}} = \left(\mathbf{Z}_t^{(v)} + \left(\mathbf{Z}_t^{(v)}\right)^T\right)\Big/2$. *Clearly,* μ *dramatically rises as* t *gradually increases. This means that* $\mathbf{J}_{t+1}^{(v)}$

become closer to the skinny SVD of the symmetric matrix $\widetilde{\mathbf{Z}_t^{(v)}}$ *as* μ *increases. This implies that* $\mathbf{J}_{t+1}^{(v)} = \mathbf{U}_r^{(v)}\mathbf{\Sigma}_r\left(\mathbf{V}_r^{(v)}\right)^T$ *when* $\mu \to +\infty$. *Hence, the value of the first condition will significantly decrease until it satisfies a certain condition as* t *increases during iterative computations.*

Then, we consider the other convergence condition, i.e., $\left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{Z}^{(v)} - \mathbf{E}^{(v)} \right\|_\infty \leq \varepsilon$. *According to Step 4 of Algorithm 1, we have*

$$\mathbf{Z}_{t+1}^{(v)} = \mathbf{P}\left(\left(\mathbf{X}^{(v)}\right)^T \mathbf{X}^{(v)} - \left(\mathbf{X}^{(v)}\right)^T \mathbf{E}_t^{(v)} + \mathbf{J}_{t+1}^{(v)}\right),$$

where $\mathbf{P} = \left(\left(\mathbf{X}^{(v)}\right)^T \mathbf{X}^{(v)} + \mathbf{I}\right)^{-1}$ *and* $\mu \to +\infty$. *Furthermore, we obtain*

$$\left(\mathbf{X}^{(v)}\right)^T \left(\mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{Z}_{t+1}^{(v)} + \mathbf{E}_t^{(v)}\right) = \mathbf{Z}_{t+1}^{(v)} - \mathbf{J}_{t+1}^{(v)}.$$

Suppose $x_{ij}^{(v)} > 0$ *and* $\forall p \in [1, d], \forall q \in [1, n], p \neq i, q \neq j, x_{pq}^{(v)} \geq 0$ *in* $\mathbf{X}^{(v)}$, *and we get* $\left\| \mathbf{Z}_{t+1}^{(v)} - \mathbf{J}_{t+1}^{(v)} \right\|_{\max} \leq x_{ij}^{(v)}\varepsilon$ *when* $\mu \to +\infty$. *Then,*

$$x_{ij}^{(v)} q_{\max} \leq \left\| \left(\mathbf{X}^{(v)}\right)^T \mathbf{Q} \right\|_{\max} \leq x_{ij}^{(v)}\varepsilon.$$

Hence,

$$q_{\max} \leq \varepsilon.$$

This indicates that the second convergence condition will hold

$$\left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{Z}_{t+1}^{(v)} - \mathbf{E}_t^{(v)} \right\|_{\max} \leq \varepsilon$$

as the first convergence condition converges to $x_{ij}^{(v)}\varepsilon$, *i.e.,*

$$\left\| \mathbf{Z}_{t+1}^{(v)} - \mathbf{J}_{t+1}^{(v)} \right\|_{\max} \leq x_{ij}^{(v)}\varepsilon,$$

if it satisfies the following conditions, i.e.,

$$x_{ij}^{(v)} > 0 \text{ and } \forall p \in [1, d], \forall q \in [1, n], p \neq i, q \neq j, x_{pq}^{(v)} \geq 0$$

in $\mathbf{X}^{(v)}$. $\qquad\square$

In fact, Algorithm 1 generally converges well in practical applications. The number of iterations was always less than 150 in our experiments.

2) Computational Complexity: We assume that the size of $\mathbf{X}$ is $d \times n \times m$, where $\mathbf{X}$ has m views, each view has n data vectors and each vector has d dimensions. The first step of Algorithm 1 involves computing the SVD of an $n \times n$ matrix, and the computational complexity is $O(n^3)$. The computational complexity of the second step of Algorithm 1 is also $O(n^3)$ because it requires finding the inverse of an $n \times n$ matrix. The computational complexity of the last step is $O(dn^2)$ when solving $\|\cdot\|_2$. Hence, the overall computational complexity of Algorithm 1 is $O(m(2n^3 + dn^2))$ in each iterative computation. In practice, we always have $n > d$. The final computational complexity of Algorithm 1 can be considered to be $O(t \times m \times n^3)$, where t is the number of iterations. In Algorithm 2, the computational complexity of the last three steps is $O(n^3)$. Therefore, the final overall complexity of Algorithm 2 is $O(t \times m \times n^3)$.

3) The Arrangement of Data Vectors and Multi-views: Here, we analyze whether the arrangement of data vectors and multiple views may lead to any potential effect. The solution of Algorithm 1 calculated by MLRR is globally optimal. We assume that data vectors are arranged by sorting clustering labels without noise. The matrix $\mathbf{Z}^*$ is a block. If the arrangement assumption is violated, we always find a matrix transpose $\mathbf{T}$ to obtain a new block $\mathbf{Z}'$, i.e., $\mathbf{Z}' = \mathbf{T} \times \mathbf{Z}^*$. In addition, we further consider any potential effect caused by multiple views. We can obtain the same values for the variables, i.e., $\mathbf{Z}^{(v)}$, $\mathbf{J}^{(v)}$, and $\mathbf{E}^{(v)}$, for each view after the first-loop computation in Algorithm 1 when considering any arrangement of multiple views. This indicates that the values of $\mathbf{Z}^{(v)}$, $\mathbf{J}^{(v)}$, and $\mathbf{E}^{(v)}$ cannot be affected by the arrangement of multiple views in successive iterative steps. Consequently, whether data vectors and multiple views are randomly shuffled is irrelevant to the optimal solution of Algorithm 1. This enables MLRR to be feasible in practical applications.

4) Connections to LRR-based Clustering Techniques: Like this work, most LRR-based techniques claim that the goal of finding the LRR is to obtain low-dimensional structures of high-dimensional data. However, the analysis of the low-dimensional matrix in this work differs from previous analyses in two key ways. First, the original meaning of intrinsic low-dimensional structures was ignored in previous works since these LRR-based techniques employed the low-rank matrix to directly measure the similarity among data samples. In this work, the low-dimensional structures are still regarded as intrinsic low-rank features of high-dimensional data, i.e., symmetric LRRs. Specifically, the symmetric LRR results of MLRR are considered to be an approximately linear spatial transformation under the new base, i.e., the multiview data themselves. Consequently, it is reasonable to exploit the angular information of the principal directions of symmetric LRRs, and a fused representation strategy is feasible. For example, a few elements in $\mathbf{Z}$ deviate strongly from the real results, and the influence of this deviation can be eliminated by exploiting the angular information of the principal directions of symmetric LRRs. This further demonstrates that the assumption of high-dimensional data is reasonable; i.e., high-dimensional data are collections of data samples approximately drawn from a union of multiple low-dimensional subspaces. Second, the introduction of the symmetric constraint makes pursuing low-rank features of high-dimensional data more reasonable, and the angular information of the principal directions of rows is consistent with that of columns in LRRs, eliminating the need for a symmetrization postprocessing step. The purpose of MLRR is to obtain a more reasonable affinity matrix produced from the angular information of the principal directions of symmetric LRRs from a different point of view.

IV. EXPERIMENTAL STUDY

In this section, we conduct extensive experiments on multiple benchmark datasets to evaluate the performance of the proposed MLRR algorithm [1] in comparison with some state-of-the-art baselines.

TABLE I
STATISTICS OF THE MULTIVIEW DATASETS.

Datasets	Clusters	Views	Data samples
BBC	5	4	685
3-sources	6	3	169
Reuters	6	5	600
Caltech-101	7	6	1474
Wiki	10	2	693
UCI Digit	10	3	2000
Flower17	17	7	1360
ProteinFold	27	12	694

A. Experimental Settings

1) Datasets: Eight publicly available datasets are used in our experiments. The statistics of these datasets are summarized in Table I. In addition, a brief description of each dataset is summarized below.

- **BBC Dataset** [17]: It includes 4 views, each of which contains 685 documents. The documents are naturally categorized into five classes: business, politics, sports, entertainment and technology.
- **3-sources Dataset** [16]: It is collected from three online news sources: BBC, Reuters, and The Guardian. This dataset consists of 948 news articles, which are manually categorized into six classes, covering 416 distinct news stories. There are 169 stories in all three sources, each of which can be considered a single view of a story. Each story is manually annotated with labels of topics in the dataset.
- **Reuters Dataset** [46]: It contains documents written in five different languages and their translations over a common set of six categories. We use documents in English as one view and their translations to the other four languages, e.g., French, German, Spanish and Italian, as the other four views. We randomly select 600 documents for this dataset, with each class containing 100 documents.
- **Caltech-101 Dataset** [14]: It contains 8,677 images of objects that belong to 101 classes. We choose a subset that has 1474 images with seven classes, i.e., Face, Motorbikes, Dolla-Bill, Garfield, Snoopy, Stop-Sign and Windsor-Chair.
- **Wiki Dataset** [39]: This dataset is a document corpus with paired texts and images. It consists of 693 Wikipedia documents, which are divided into ten classes.
- **UCI Digit Dataset** [12]: It consists of 2,000 examples of handwritten digits (0-9). There are 200 examples available for each digit, with examples of each digit represented with six types of features. We construct three views: 76 Fourier coefficients of the character shapes, 216 profile correlations and 64 Karhunen-Love coefficients.
- **Flower17 Dataset** [36] It consists of 17 different flower categories with 80 images for each class. There are a total of 7 views in this dataset.
- **ProteinFold Dataset** [2] This dataset includes 12 views, each of which contains 694 protein domains that belong to 27 fold classes.

2) Compared Methods: We compared the performance of MLRR against the following methods:

---

[1]https://github.com/chenjie20/MLRR

[2]http://mkl.ucsd.edu/dataset/protein-fold-prediction

TABLE II
CLUSTERING PERFORMANCE COMPARISON OF DIFFERENT METHODS ON EIGHT MULTIVIEW DATASETS. THE BEST AND SECOND BEST RESULTS ARE SHOWN IN BOLD AND UNDERLINED, RESPECTIVELY.

Datasets	LRR$_{BestSV}$	LRR$_{Concat}$	MLAN	LRRSC	GMC	RMSL	MCLES	LMVSC	MLRR
ACC									
BBC	79.27 (0)	87.15 (0)	86.57 (0)	87.16 (0)	77.52 (0)	90.91 (0.81)	85.8 (1.05)	76.79 (0)	**91.39** (0)
3-sources	60.95 (0)	66.27 (0)	81.07 (0)	78.7 (0)	85.21 (0)	78.46 (1.58)	83.02 (1.37)	76.92 (0)	**89.29** (0.52)
Reuters	50.43 (0.2)	49.33 (0)	52.17 (0)	52 (0)	33.5 (0)	48.55 (1.8)	49.47 (1.57)	48.17 (0)	**54** (0)
Caltech-101	54.75 (0)	51.7 (0)	81.95 (0)	57.21 (0)	80.94 (0)	80.05 (0.5)	82.94 (1.7)	71.25 (0.23)	**84.06** (0.02)
Wiki	62.34 (0.07)	60.26 (0.28)	47.76 (1.83)	59.56 (0.17)	61.47 (0)	59.55 (1.19)	60.76 (1.36)	56.12 (0.32)	**64.52** (0.05)
UCI digit	77.3 (0)	92.15 (0)	97.7 (0)	91.9 (0)	87.15 (0)	90.24 (1.81)	91.37 (1.34)	86.38 (0.24)	**98.1** (0)
Flower17	42.21 (0.41)	55.57 (1.58)	51.43 (2.51)	58.63 (1.83)	48.93 (2.47)	55.42 (1.17)	49.92 (0.47)	52.9 (1.08)	**64.79** (0.2)
ProteinFold	25.78 (1.4)	29.55 (1.82)	29.68 (1.98)	30.48 (0.79)	31.86 (1.93)	30.48 (1.58)	37.12 (0.91)	30.3 (0.56)	**38.79** (0.48)
NMI									
BBC	54.87 (0)	72.3 (0)	73.04 (0)	72.82 (0)	69.1 (0)	77.46 (1.63)	71.02 (0.17)	60.96 (0)	**78.73** (0)
3-sources	51.96 (0)	63.11 (0)	75.34 (0)	69.7 (0)	73.97 (0)	73.14 (1.47)	74.6 (1.04)	65.38 (0)	**80.36** (0.9)
Reuters	32.75 (0.16)	36.65 (0)	**40.68** (0)	34.63 (0.42)	28.32 (0)	33.77 (1.87)	35.79 (0.57)	37.28 (0)	38.12 (0)
Caltech-101	42.8 (0)	57.39 (0)	62.37 (0)	61.53 (0)	61.18 (6.46)	66.94 (1.63)	56 (0.15)	55.3 (0.08)	**69.56** (0)
Wiki	54.95 (0.08)	52.5 (0.24)	46.55 (0.78)	54.36 (0.26)	55.86 (0)	50.78 (1.75)	52.57 (1.58)	52.02 (0.19)	**59.46** (0.01)
UCI digit	76.19 (0)	85.26 (0)	94.78 (0)	86.34 (0)	90.83 (0)	89.08 (1.27)	83.8 (1.58)	79.94 (0.02)	**95.49** (0)
Flower17	46.36 (0.26)	56.25 (0.91)	54.41 (1.56)	56.58 (0.95)	52.15 (1.06)	55.47 (0.76)	51.35 (0.24)	51.28 (0.91)	**61.73** (0.25)
ProteinFold	32.62 (0.83)	40.88 (1.2)	37.45 (1.01)	40.24 (0.49)	40.42 (1.19)	42.1 (1.13)	46.46 (0.52)	40.37 (0.71)	**48.66** (0.54)
F-measure									
BBC	79.14 (0)	87.2 (0)	85.74 (0)	87.35 (0)	79.97 (0)	90.89 (1.8)	84.19 (1.38)	77.55 (0)	**91.25** (0)
3-sources	66.12 (0)	71.2 (0)	82.46 (0)	79.3 (0)	83.27 (0)	80.47 (1.46)	84.21 (1.69)	79.24 (0)	**89.67** (0.51)
Reuters	50.26 (0.16)	51.25 (0)	50.26 (0)	51.67 (0.03)	41.75 (0)	50.89 (1.78)	47.04 (1.27)	48.89 (0)	**56.98** (0)
Caltech-101	58.04 (0	60 (0)	79.25 (0)	71.64 (0)	79.5 (0)	82.43 (0.75)	80.34 (1.65)	76 (0.4)	**85.64** (0)
Wiki	64.21 (0.1)	63.07 (0.21)	54.9 (1.83)	62.98 (0.12)	64.02 (0)	60.44 (1.13)	59.41 (1.39)	62.25 (0.28)	**65.91** (0.05)
UCI digit	82.48 (0)	92.16 (0)	97.72 (0)	97.05 (0)	90.74 (0)	92.36 (1.54)	90.5 (1.25)	86.57 (0.02)	**98.11** (0)
Flower17	48.53 (0.49)	57.78 (1.27)	54.14 (1.96)	60.5 (1.5)	53.84 (2.44)	60.09 (0.99)	55.23 (0.08)	56.33 (1.23)	**66.01** (0.22)
ProteinFold	27.13 (1.31)	32.51 (1.75)	32.02 (1.66)	31.5 (0.81)	33.83 (2.23)	33.97 (1.39)	**40.75** (0.99)	34.44 (0.82)	39.03 (0.89)
Adj-RI									
BBC	57.81 (0)	75.08 (0)	72.74 (0)	74.03 (0)	68.66 (0)	80.51 (1.75)	71.92 (0.28)	51.01 (0)	**81.71** (0)
3-sources	44.48 (0)	53.46 (0)	66.64 (0)	63.65 (0)	73.19 (0)	70.38 (1.42)	78.28 (1.34)	60.57 (0)	**85.22** (0.55)
Reuters	23.75 (0.14)	24.62 (0)	20.96 (0)	24.56 (0.26)	4.94 (0)	22.36 (1.6)	21.07 (0.18)	21.73 (0)	**25.33** (0)
Caltech-101	28.95 (0)	38.48 (0)	60.77 (0)	42.57 (0)	58.88 (0)	63.52 (1.31)	71.12 (0.91)	40.13 (0.24)	**80.63** (0))
Wiki	45.44 (0.11)	44.76 (0.18)	26.23 (2.1)	42.1 (0.28)	37.07 (0)	44.09 (1.27)	48.44 (1.45)	52.23 (1.31)	**52.6** (0.08)
UCI digit	69.35 (0)	83.67 (0)	94.9 (0)	83.63 (0)	83.38 (0)	88.27 (1.52)	86.96 (1.45)	71.85 (0.05)	**95.81** (0)
Flower17	21.28 (0.32)	36.13 (1.43)	30.97 (2.66)	38.41 (1.23)	29.65 (0.44)	33.62 (2.62)	29.39 (0.14)	27.24 (0.71)	**47.03** (0.26)
ProteinFold	10.58 (1.37)	14.67 (1.75)	11.32 (1.54)	15.73 (0.6)	12.33 (2.4)	12.37 (1.25)	18.64 (0.53)	10.87 (0.81)	**20.99** (0.5)

- **LRR$_{BestSV}$**[3] [28]. We perform a standard LRR on an individual view and report the best subspace clustering result with a single view of data.
- **LRR$_{Concat}$**[28]. We first obtain an LRR by concatenating the aligned features from all views and then directly perform standard spectral clustering on the LRR.
- **MLAN**[35]. It presents a multiview learning mode to learn an optimal graph that can be partitioned into clusters directly. We apply principal component analysis (PCA) to reduce the same number of dimensions for all views of multidata.
- **LRSSC**[4] [2] . It learns a joint subspace representation across all views, i.e., an affinity matrix, by combining sparsity with a low-rank property.
- **GMC**[5] [48]. It proposes a general graph-based multiview clustering (GMC) model to generate a unified graph matrix for multiview clustering by fusing the data graph matrices of all views.
- **RMSL**[6] [24]. The reciprocal multi-layer subspace learning (RMSL) algorithm consists of two main components, i.e., the hierarchical self-representative layers and backward encoding networks.
- **MCLES**[7] [8]. It presents a unified optimization framework to implement multi-view clustering in latent embedding space (MCLES).
- **LMVSC**[8] [21]. The large-scale multi-view subspace clustering (LMVSC) algorithm implements multi-view clustering with linear order complexity.

For the competing algorithms, we adopted the source codes provided by their authors. We applied a PCA algorithm to preprocess the samples if the number of samples was less than the dimension of any single view in the multiview data [20]. All competing methods were implemented in MATLAB, and all experiments were performed on a Windows platform with

---

[3]https://sites.google.com/site/guangcanliu/
[4]https://github.com/mbrbic/Multi-view-LRSSC
[5]https://github.com/cshaowang/gmc

[6]https://github.com/lslrh/RMSL
[7]https://github.com/Ttuo123/MCLES
[8]https://github.com/sckangz/LMVSC

an Intel i7-9700K CPU and 32 GB RAM.

3) Evaluation Metrics: We employed four metrics to evaluate the clustering performance of all competing algorithms, i.e., clustering accuracy (ACC), normalized mutual information (NMI), F-measure and adjusted rand index (Adj-RI). The last three metrics have specific definitions. In our experiments, the clustering accuracy is defined as follows:

$$ACC = \frac{1}{n} \sum_{i=1}^{n} \delta(c_i, map(x_i)), \quad (25)$$

where c_i denotes the class label of x_i, $\delta(x, y)$ measures the equivalence of x and y, and $map(x_i)$ is the permutation mapping function that maps each cluster label x_i to the class label contained in one permutation of all class labels. A larger value indicates better clustering performance for these metrics.

4) Parameter Settings: For a fair comparison, we manually tuned the parameters of the competing methods following the advice of their respective authors and reported the best results. The final step of the competing methods was run with the k-means algorithm. We adopted the standard implementation of k-means provided by MATLAB, and these methods shared the same parameters of k-means in the experiments. The performance of MLRR is independent of the arrangement of data vectors and multiple views. Hence, we performed all experiments only once and reported the metrics as the final results, i.e., ACC, NMI, F-measure and Adj-RI.

There are three parameters, i.e., λ, η and α, in the MLRR. The initial parameter λ is chosen from $\{1e^{-2}, 5e^{-2}, 0.1, 0.5, 1, 10, 1e^2, 5e^2\}$ while the other initial parameter η is chosen from $\{5e^{-3}, 1e^{-2}, 5e^{-2}, 0.1, 0.5, 1\}$. In fact, parameter λ is closely related to the prior of the corruption level of multidata. Parameter η is usually dependent on the differences among multiple views. Parameter α usually ranges from 1 to 5. The specific values of the parameters are given in the experiments.

B. Evaluation of Clustering Quality

1) Performance Comparison: We evaluated the proposed algorithm on eight multiview datasets. Table II shows the clustering results of the competing methods on the datasets. The eight groups of MLRR parameters were: (1) $\lambda = 5e^{-2}$, $\eta = 0.1$, $\alpha = 1$; (2) $\lambda = 0.5$, $\eta = 0.1$, $\alpha = 5$; (3) $\lambda = 0.6$, $\eta = 5e^{-3}$, $\alpha = 3$; (4) $\lambda = 0.5$, $\eta = 0.1$, $\alpha = 5$; (5) $\lambda = 0.4$, $\eta = 0.1$, $\alpha = 5$; (6) $\lambda = 9e^{-2}$, $\eta = 9e^{-3}$, $\alpha = 5$; (7) $\lambda = 600$, $\eta = 0.1$, $\alpha = 1$ and (8) $\lambda = 50$, $\eta = 0.1$, $\alpha = 2$ for these experiments, respectively.

Table II clearly shows that our approach generally performs better than the other baselines regarding the four test metrics. For example, MLRR achieved high clustering ACCs of 91.39%, 89.29%, 54%, 84.06%, 64.52%, 98.1%, 59.27% and 38.79% for the BBC, 3-sources, Reuters, Caltech-101, Wiki, UCI digit, Flower17 and ProteinFold datasets, respectively. These clustering ACCs were improved by at least 0.48%, 4.08%, 1.83%, 1.12%, 2.18%, 0.4%, 0.64% and 1.67% when compared with the second-best methods. We also observed significant advantages for the F-measures in all the datasets. In addition, MLRR obtained competitive clustering results for

almost all NMI and Adj-RI and significantly outperformed the other methods, except on the Reuters dataset. Specifically, MLRR yielded NMI improvements of at least 2.27%, 5.02%, 2.62%, 3.6%, 0.71%, 0.85% and 2.2% compared with the other methods for the BBC, 3-sources, Caltech-101, Wiki, UCI digit, Flower17 and ProteinFold, datasets, respectively. These results confirm that our proposed method is very effective with respect to different numbers of views and clusters in multiview subspace clustering. As expected, the clustering performance of the multiview methods is much better than that of the best single-view algorithms. This is because different views contain complementary information in multidata.

Considering LRRs in multiple views, we obtained the following observations for LRR$_{BestSV}$ and LRR$_{Concat}$ in Table II. We reported the best clustering results of individual views in Table II, i.e., LRR$_{BestSV}$. We also simultaneously reported the clustering results in Table II by concatenating aligned features from all views, i.e., LRR$_{Concat}$. Table II shows that LRR$_{Concat}$ consistently achieved higher clustering ACCs than LRR$_{BestSV}$ for the BBC, 3-sources, UCI digit, Flower17 and ProteinFold datasets. However, LRR$_{BestSV}$ slightly outperformed LRR$_{Concat}$ for the other datasets. Hence, the clustering performance of LRR on the datasets cannot remain relatively stable when simply concatenating LRRs of all views in multidata. Compared with the clustering results of MLRR, this finding also confirms the effectiveness and robustness of the fusion strategy of MLRR for performing multiview subspace clustering tasks.

2) Computational Time: To evaluate the efficiency of the proposed MLRR algorithm, we performed experiments on all the datasets. For the sake of fairness, we did not report the computational cost of LRR$_{BestSV}$ because this method only perform a clustering task on an individual view. Instead of LRR$_{BestSV}$, LRR$_{Concat}$ is the appropriate representative of multiview subspace clustering to evaluate the computational cost of a standard LRR.

We report the computational costs of all the methods in Table III. MLAN achieved a lower computational cost by iteratively modifying the similarity matrix. In addition, Table III shows that MLRR and LMVSC achieve similar computational costs overall, which are lower than those of RMSL and MCLES. Clearly, MLRR's computational cost is much lower than LRR$_{Concat}$'s. This confirms that it is worth developing a symmetric LRR model rather than simply concatenating separate LRRs into a compact representation, which significantly reduces the computational cost of multiview subspace clustering from the perspective of LRR.

C. The Effect of the Symmetry Constraint

The symmetry constraint of LRR is a critical factor that guarantees weight consistency for each pair of data vectors in multiple views. To show the effect of the symmetry factor in MLRR, we performed two groups of experiments to evaluate NMI and ACC on different datasets. We first considered two special variants of MLRR, i.e., MLRR$_a$ and MLRR$_b$. Symmetrization is only performed on $\mathbf{Z}^{(v)}$ after each iteration of Algorithm 1, i.e., $\mathbf{Z}_{t+1}^{(v)} = \frac{1}{2}\left(\mathbf{Z}^{(v)} + \left(\mathbf{Z}^{(\mathbf{v})}\right)^T\right)$ while the

TABLE III
COMPUTATIONAL TIME (IN SECONDS) OF THE COMPETING METHODS ON DIFFERENT DATASETS.

| Datasets	LRR$_{Concat}$	MLAN	LRRSC	GMC	RMSL	MCLES	LMVSC	MLRR
BBC	88.49	**1.01**	7.65	<u>1.5</u>	60.29	315.18	20.48	54.5
3-sources	4.56	**0.13**	0.9	<u>0.68</u>	3.3	10.05	1.87	2.26
Reuters	81.15	**0.97**	<u>2.01</u>	3.33	89.29	186.15	57.9	72.58
Caltech-101	1179.4	<u>5.72</u>	24.79	**4.81**	489.04	11535	250.59	234.3
Wiki	55.58	**0.58**	<u>2.37</u>	5.91	22.22	283.22	55.25	18.85
UCI digit	622.51	**8.5**	168.51	<u>49.21</u>	704.25	8235.6	58.91	362.32
Flower17	918.03	<u>47.13</u>	123.39	**20.9**	1373.83	8598.3	1373.4	246.27
ProteinFold	314.57	<u>24.7</u>	35.44	**13.2**	688.29	1998.9	36.5	76.61
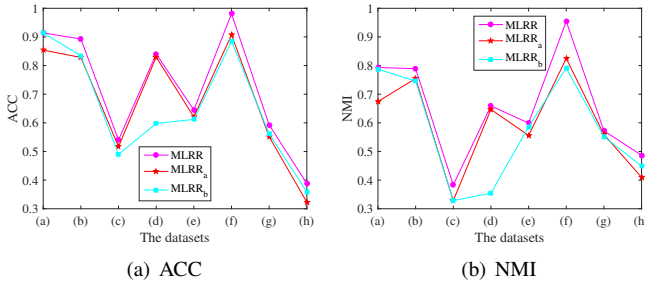
Fig. 1. ACC and NMI of MLRR, MLRR$_a$ and MLRR$_b$ on different datasets: (a) BBC, (b) 3-sources, (c) Reuters, (d) Caltech-101, (e) Wiki, (f) UCI Digit, (g) Flower17 and (h) ProteinFold.
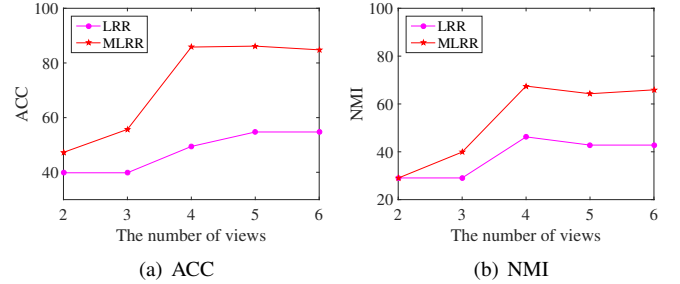

Fig. 3. ACC and NMI of MLRR and LRR$_{BestSV}$ using different numbers of views for the Caltech-101 dataset.
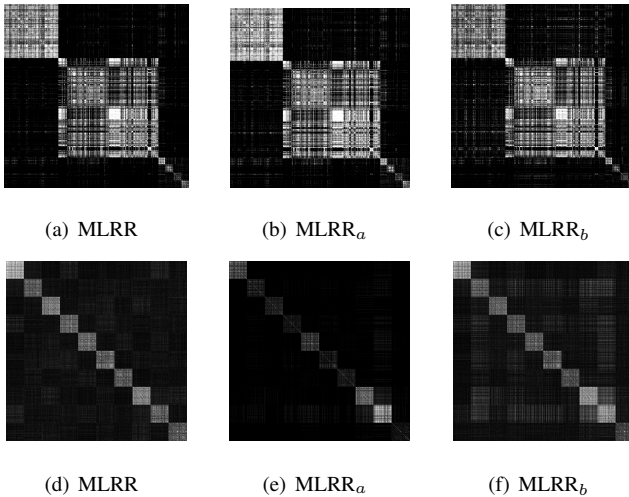

Fig. 2. Comparisons of the affinity graph matrices produced by MLRR, MLRR$_a$ and MLRR$_b$ on two datasets: (a)-(c) Caltech-101, (d)-(f) UCI Digit.

symmetry constraint is ignored in MLRR$_a$. In addition, the symmetrization postprocessing step is applied in MLRR$_b$, where the symmetry constraint is ignored. Fig. 1(a) shows a comparison of the experimental results of MLRR, MLRR$_a$ and MLRR$_b$. The experimental results of MLRR and MLRR$_b$ were obtained under the same parameter settings, whereas the optimal experimental results of MLRR$_b$ were obtained by tuning the parameters. We can see that MLRR consistently outperformed MLRR$_a$ and MLRR$_b$ on different datasets. Moreover, MLRR$_a$ achieved better results than MLRR$_b$ in most cases. These findings indicate that performing symmetrization during iterative computations has advantages over the symmetrization

postprocessing step. We also show intuitive examples of the affinity graphs produced by MLRR, MLRR$_a$ and MLRR$_b$ on the Caltech-101 and UCI Digit datasets. The results are illustrated in Fig. 2. The affinity graphs for MLRR reveal a distinct block-diagonal structure, which is attributed to the final good clustering performance. However, there are a small number of noisy values in the affinity graphs of the variant. These examples demonstrate that the symmetry constraint plays an important role in the MLRR model.

D. Evaluation of Increasing Number of Views

Since the clustering performance of MLRR is closely related to the number of views, we conducted another experiment to illustrate the effect of various numbers of views. Specifically, we designed five groups of views of the Caltech-101 dataset, where each group included 2 to 6 views of each sample. We also reported the results of LRR$_{BestSV}$, which provided some useful statistical information for the Caltech-101 dataset. Figs. 3(a) and 3(b) show the ACC and NMI values of MLRR and LRR$_{BestSV}$ when the number of views varies from 2 to 6. The ACC of MLRR rapidly increases when the number of views varies from 2 to 4. Then, the ACC of MLRR remains relatively stable when the number of views is between 4 and 6. We also observed that the ACC of MLRR is similar to that of LRR$_{BestSV}$ under different numbers of views. The clustering results of LRR$_{BestSV}$ show that multiple views contain more information when the number of views increases. This confirms that an increasing number of views can improve the clustering performance of MLRR. Moreover, MLRR can effectively capture important information as the number of views varies.
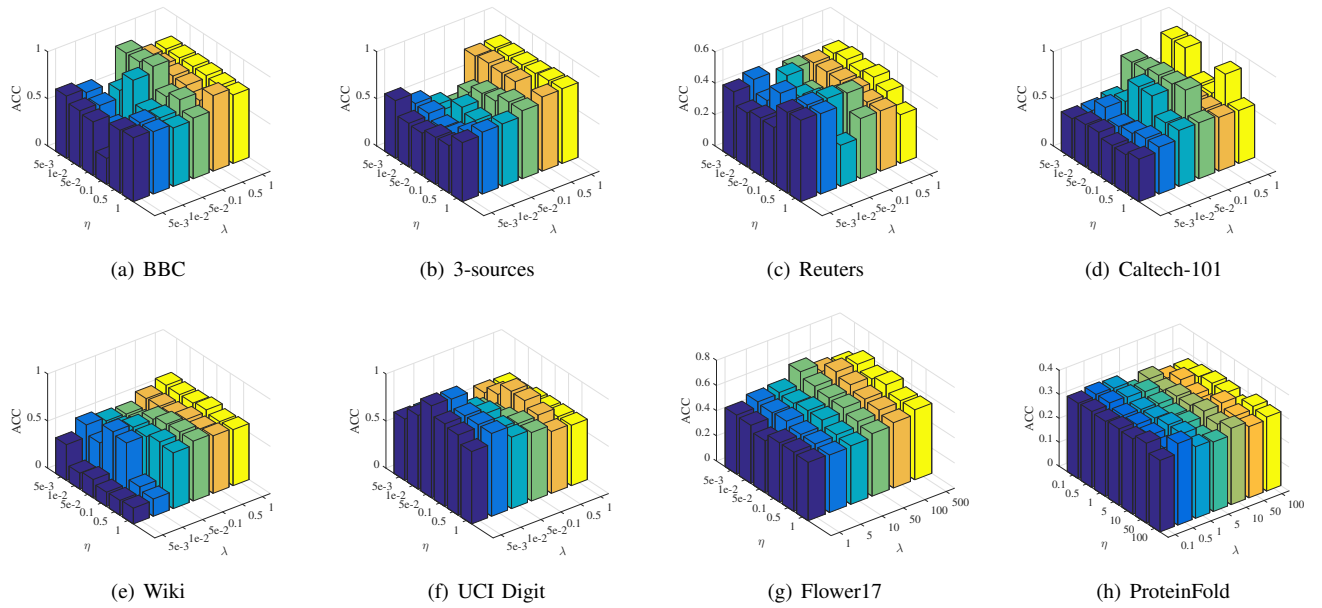
Fig. 4. ACC with different λ and η combinations on different datasets: (a) $\alpha = 1$, (b) $\alpha = 1$, (c) $\alpha = 1$, (d) $\alpha = 5$, (e) $\alpha = 5$, (f) $\alpha = 5$, (g) $\alpha = 1$ and (h) $\alpha = 2$.
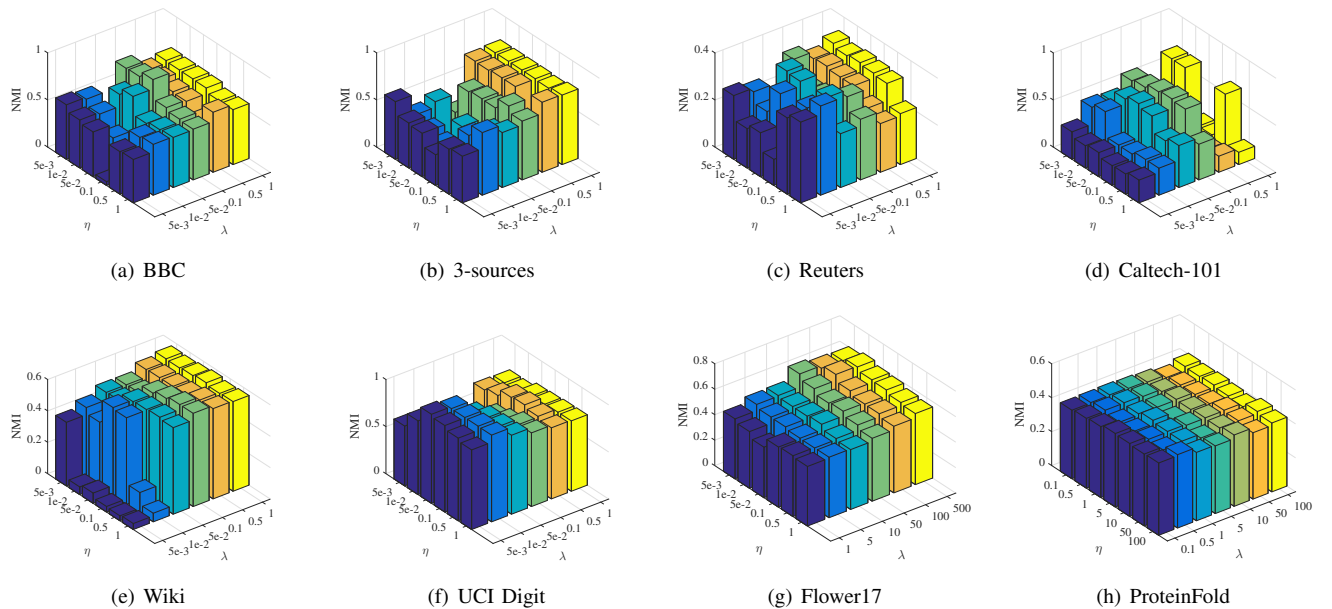

Fig. 5. NMI with different λ and η combinations on different datasets: (a) $\alpha = 1$, (b) $\alpha = 1$, (c) $\alpha = 1$, (d) $\alpha = 5$, (e) $\alpha = 5$, (f) $\alpha = 5$, (g) $\alpha = 1$ and (h) $\alpha = 2$.

E. Parameter Sensitivity Analysis and Convergence Study

1) Parameter Sensitivity Analysis: For the sake of studying the sensitivity of parameters in the proposed method, we considered different values of α, i.e., (a) $\alpha = 1$, (b) $\alpha = 1$, (c) $\alpha = 1$, (d) $\alpha = 5$, (e) $\alpha = 5$, (f) $\alpha = 5$, (g) $\alpha = 1$ and (h) $\alpha = 2$, where MLRR achieved the best performance in the previous experiments. We conducted experiments to study the influences of the parameters λ and η of MLRR with a fixed α. We let λ and η range from $5e^{-3}$ to 1 or 0.1 to 500 in steps of half-orders of magnitude. Figs. 4 and 5 show the influences of the parameters λ and η in terms of ACC and

NMI, respectively. We can observe that the ACC and NMI values of MLRR remain very stable under a fixed range of values of λ and a wide range of values of η.

2) Convergence Study: The convergence of Algorithm 1 was analyzed theoretically in Section III. We further conducted an experimental analysis to validate its convergence speed. For comparison, we also analyzed the speed of LRR_{BestSV}, and the best clustering results were chosen. Fig. 6 shows the convergence curves of the proposed Algorithm 1 and LRR_{BestSV} on eight datasets. The numbers of iterations of Algorithm 1 are less than 150 for the datasets. For example, the
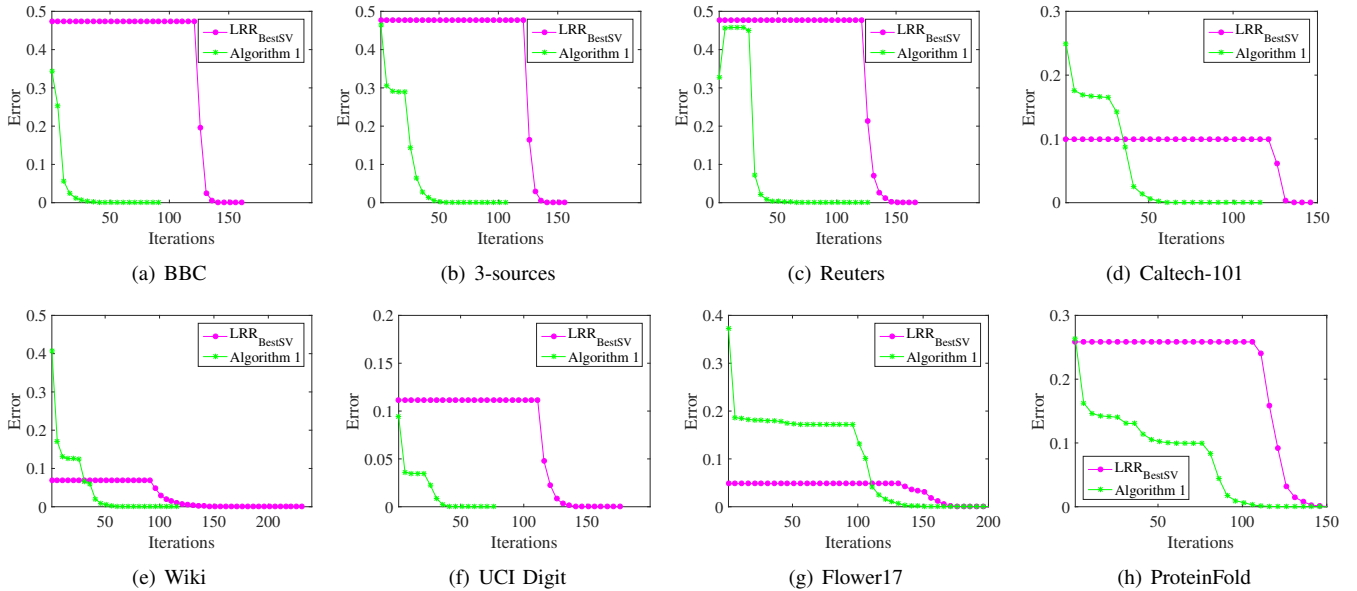
Fig. 6. Convergence curves of the proposed Algorithm 1 and LRR$_{BestSV}$ for different datasets.

minimum and maximum iterations are 78 and 118 for the UCI Digit and Reuters datasets, respectively. Hence, Algorithm 1 always converges well for the datasets.

In addition, the number of iterations of Algorithm 1 is much lower than that of LRR$_{BestSV}$. Compared with LRR$_{BestSV}$, the objective value of Algorithm 1 decreases dramatically as the number of iterations increases. This is because the LRRs of multidata can share the diversity and consistency among different views of multidata. Consequently, this approach effectively improves the convergence speed of learning the LRRs of multidata.

F. Discussion

To learn the intrinsic structure of multiple views, various criteria are employed to characterize the data representation, e.g., LRR and sparse representation. For example, ILRSO simultaneously considers low-rank and sparse representation as well as the local manifold structure in each view. It considers the sum of all similarity matrices as an affinity matrix, which effectively reveals the membership of data vectors in multiple views. Moreover, the SLRMF technique is used to improve the efficiency of the ILRSO method. Although it is similar to SLRMF and ILRSO in adopting LRR techniques, MLRR does not intuitively pursue an affinity matrix. It recovers the multiple subspaces of multiple views by exploiting the diversity and consistency among different views. MLRR focus on learning the optimal affinity matrix preserved in an approximately linear spatial transformation under the new base, which is completely different from several of the representative works mentioned above.

The LRR can be applied in single-view subspace clustering. However, each view of the data may contain distinct knowledge in multidata. Hence, MLRR is employed to exploit the complementarity of multiple views by pursing the symmetric LRRs of multiple views, where different views share complementary information with each other. The complementary

information underlying multiple views can be exploited by MLRR to improve the learning performance of data representation. As expected, MLRR improves subspace clustering performance when the number of views gradually increases. In theory, if there is only one view, MLRR is regarded as an LRR-based model for single-view subspace clustering.

V. CONCLUSIONS

In this paper, we proposed an MLRR model to solve the multiview subspace clustering problem. MLRR effectively exploits the diversity of multiple views, i.e., different low-dimensional structures of multiple views, and simultaneously measures the consistency among different views. In contrast with most multiview subspace clustering methods, MLRR presents a new construction of the affinity matrix in multiview subspace clustering from another point of view, i.e., by fully exploiting the angular information of the principal directions of LRRs under a symmetric condition. Specifically, it focuses on two important aspects, i.e., feature learning and feature fusion. It successfully learns the symmetric LRRs that imply the complementary knowledge contained in multiple views. Moreover, a fusion strategy of symmetric LRRs is presented to obtain a compact representation, which can be shared by multiple views and comprehensively represents the intrinsic features of multiview data. The angular information of the principal directions of the compact representation is employed to evaluate the membership of multiview data. Hence, the intrinsic characteristics preserved in symmetric LRRs are adopted to effectively characterize multiple subspaces. In addition, MLRR naturally avoids the symmetrization postprocessing step. Compared with standard LRR algorithms, MLRR converges much faster due to sharing the common information of multiple views, as shown in the experiments. The experimental results on benchmark datasets showed that MLRR produces very competitive results compared with several state-of-the-art multiview subspace clustering algorithms.

REFERENCES

[1] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear sub-spaces," *IEEE Trans. Pattern Analysis Machine Intell.*, vol. 25, no. 2, pp. 218–233, Feb. 2003.

[2] M. Brbić and I. Kopriva, "Multi-view low-rank sparse subspace clustering," *Pattern Recognition*, vol. 73, pp. 247–258, Jan. 2018.

[3] M. Brbić and I. Kopriva, "l_0 motivated low-rank sparse subspace clustering," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1711–1725, Apr. 2020.

[4] X. Cao, C. Zhang, H. Fu, S. Liu, and H. Zhang, "Diversity-induced multi-view subspace clustering," in *IEEE Computer Society Conf. Computer Vision Pattern Recognition*, Boston, USA, Jun. 2015, pp. 586–594.

[5] J. Chen, H. Mao, Y. Sang, and Z. Yi, "Subspace clustering using a symmetric low-rank representation," *Knowledge-Based Systems*, vol. 127, no. 1, pp. 46–57, july 2017.

[6] J. Chen, H. Mao, H. Zhang, and Z. Yi, "Symmetric low-rank preserving projections for subspace learning," *Neurocomputing*, vol. 315, no. 13, pp. 381–393, Nov. 2018.

[7] J. Chen, H. Zhang, H. Mao, Y. Sang, and Z. Yi, "Symmetric low-rank representation for subspace clustering," *Neurocomputing*, vol. 173, no. 15, pp. 1192–1202, january 2016.

[8] M. Chen, L. Huang, C. Wang, and D. Huang, "Multi-view clustering in latent embedding space," in *Proc. 34th AAAI Conf. Artif. Intell.*, New York, USA, Feb. 2020, pp. 1–8.

[9] N. Chen, J. Zhu, and E. P. Xing, "Predictive subspace learning for multi-view data: a large margin approach," in *Advances in Neural Information Processing Systems*, Vancouver, Canada, december 2010, pp. 361–369.

[10] Y. Chen, Z. Lai, W. K. Wong, L. Shen, and Q. Hu, "Low-rank linear embedding for image recognition," *IEEE Trans. Multimedia*, vol. 20, no. 12, pp. 3212–3222, Dec. 2018.

[11] M. Cheng, L. Jing, and M. K. Ng, "Tensor-based low-dimensional representation learning for multi-view clustering," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2399–2414, Oct. 2018.

[12] D. Dua and C. Graff. (2017) UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences. [Online]. Available: http://archive.ics.uci.edu/ml

[13] E. Elhamifar and R. Vidal, "Sparse subspace clustering algorithm, theory, and applications," *IEEE Trans. Pattern Analysis Machine Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

[14] 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," in *IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Washington, DC, USA, june 2004, pp. 178–178.

[15] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," in *IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Las Vegas, Nevada, USA, june 2016, pp. 1933–1941.

[16] D. Greene. (2009, april) The reuters dataset. [Online]. Available: http://mlg.ucd.ie/datasets/3sources.html

[17] D. Greene and P. Cunningham, "Producing accurate interpretable clusters from high-dimensional data," in *Knowledge Discovery in Databases: PKDD 2005*, Porto, Portugal, oOct. 2005, pp. 486–494.

[18] Z. Huang, H. Zhu, J. T. Zhou, and X. Peng, "Multiple marginal fisher analysis," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9798–9807, Dec. 2018.

[19] Y. Jia, H. Liu, J. Hou, and S. Kwong, "Semisupervised adaptive symmetric non-negative matrix factorization," *IEEE Trans. Cybern.*, in press, 2020.

[20] I. Jolliffe, *Principal component analysis*. Heidelberg, Berlin: Springer, 2011.

[21] Z. Kang, W. Zhou, Z. Zhao, J. Shao, M. Han, and Z. Xu, "Large-scale multi-view subspace clustering in linear time," in *Proc. 34th AAAI Conf. Artif. Intell.*, New York, USA, Feb. 2020, pp. 1–8.

[22] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. S., and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Trans. Pattern Analysis Machine Intell.*, vol. 24, no. 7, pp. 881–892, Jul. 2002.

[23] D. Kuang, C. Ding, and H. Park, "Symmetric nonnegative matrix factorization for graph clustering," in *Proc. 2012 SIAM Int. Conf. Data Mining*, Brussels, Belgium, Dec. 2012, pp. 106–117.

[24] R. Li, C. Zhang, H. Fu, X. Peng, T. Zhou, and Q. Hu, "Reciprocal multi-layer subspace learning for multi-view clustering," in *Proc. IEEE Int. Conf. Computer Vision*, Seoul, Korea, Oct. 2019, pp. 8172–8180.

[25] Y. Li, F. Nie, H. Huang, and J. Huang, "Large-scale multi-view spectral clustering via bipartite graph," in *Proc. 29th AAAI Conf. Artif. Intell.*, Austin, Texas, USA, Jan. 2015, pp. 2750–2756.

[26] 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.

[27] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Advances in Neural Information Processing Systems*, Vancouver, Canada, Dec. 2011, pp. 612–620.

[28] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[29] C. Lu, S. Yan, and Z. Lin, "Convex sparse spectral clustering: Single-view to multi-view," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2833–2843, Jun. 2016.

[30] Y. Lu, Z. Lai, W. K. Wong, and X. Li, "Low-rank discriminative regression learning for image classification," *Neural Networks*, vol. 125, pp. 245–257, june 2020.

[31] Y. Lu, C. Yuan, X. Li, Z. Lai, D. Zhang, and L. Shen, "Structurally incoherent low-rank 2dlpp for image classification," *IEEE Trans. Circuits Syst. for Video Tech.*, vol. 29, no. 6, pp. 1701–1714, Jun. 2018.

[32] S. Luo, C. Zhang, W. Zhang, and X. Cao, "Consistent and specific multi-view subspace clustering," in *Proc. 32nd AAAI Conf. Artif. Intell.*, New Orleans, Louisiana, USA, Feb. 2018, pp. 3730–3737.

[33] U. V. Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, Dec. 2019.

[34] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proc. 28th Int. Conf. Machine Learning*, Bellevue, Washington, USA, Jun. 2011, pp. 689–696.

[35] F. Nie, G. Cai, and X. Li, "Multi-view clustering and semi-supervised classification with adaptive neighbours," in *Proc. 31st AAAI Conf. Artif. Intell.*, San Francisco, California, USA, Feb. 2017, pp. 2408–2414.

[36] M. E. Nilsback and A. Zisserman, "A visual vocabulary for flower classification," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, USA, june 2006, pp. 1447–1454.

[37] X. Peng, J. Feng, S. Xiao, W. Y. Yau, J. T. Zhou, and S. Yang, "Structured autoencoders for subspace clustering," *IEEE Trans. Image Processing*, vol. 27, no. 10, pp. 5076–5086, Aug. 2018.

[38] X. Peng, C. Lu, Z. Yi, and H. Tang, "Connections between nuclear norm and frobenius norm based representation," *IEEE Trans. Neural Networks Learning Syst.*, vol. 29, no. 1, pp. 218–224, Jan. 2018.

[39] N. Rasiwasia, J. C. Pereira, E. Coviello, G. Doyle, G. R. Lanckriet, R. Levy, and N. Vasconcelos, "A new approach to cross-modal multimedia retrieval," in *Proc. 18th ACM Int. Conf. Multimedia*, Firenze, Italy, October 2010, pp. 251–260.

[40] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, Aug. 2010.

[41] J. Shi, J. Malik, and S.Sastry, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 181–214, may 2000.

[42] S. Sun, "A survey of multi-view machine learning," *Neural computing and applications*, vol. 23, no. 7, pp. 2031–2038, december 2013.

[43] C. Tang, X. Liu, X. Zhu, E. Zhu, Z. Luo, L. Wang, and W. Gao, "CGD: Multi-view clustering via cross-view graph diffusion," in *Proc. AAAI Conf. Artif. Intell.*, Boston, USA, june 2020, pp. 5924–5931.

[44] C. Tang, X. Zhu, X. Liu, P. Wang, C. Zhang, and L. Wang, "Learning a joint affinity graph for multiview subspace clustering," *IEEE Trans. Multimedia*, vol. 21, no. 7, pp. 1724–1736, Jul. 2019.

[45] H. Tao, C. Hou, D. Yi, J. Zhu, and D. Hu, "Joint embedding learning and low-rank approximation: A framework for incomplete multiview learning," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1690–1703, Mar. 2021.

[46] M. Thoma. (2017, july) The reuters dataset. [Online]. Available: https://martin-thoma.com/nlp-reuters

[47] H. Wang, F. Nie, and H. Huang, "Multi-view clustering and feature learning via structured sparsity," in *Proc. 30th Int. Conf. Machine Learning*, Atlanta, GA, USA, June 2013, pp. 352–360.

[48] H. Wang, Y. Yang, and B. Liu, "GMC: graph-based multi-view clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 6, pp. 1116–1129, Jun. 2020.

[49] X. Wang, X. Guo, Z. Lei, C. Zhang, and S. Z. Li, "Exclusivity-consistency regularized multi-view subspace clustering," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Boston, USA, june 2017, pp. 923–931.

[50] Y. Wang, X. Lin, L. Wu, W. Zhang, Q. Zhang, and X. Huang, "Robust subspace clustering for multi-view data by exploiting correlation con-

sensus," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 3939–3949, Nov. 2015.

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

[52] Y. Wang, W. Zhang, L. Wu, X. Lin, M. Fang, and S. Pan, "Iterative views agreement: an iterative low-rank based structured optimization method to multi-view spectral clustering," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, New York, USA, Jul. 2016, pp. 2153–2159.

[53] J. Wen, Y. Xu, and H. Liu, "Incomplete multiview spectral clustering with adaptive graph learning," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1418–1429, Apr. 2020.

[54] R. Xia, Y. Pan, L. Du, and J. Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Proc. 28th AAAI Conf. Artif. Intell.*, Quebec, Canada, Jul. 2014, pp. 2149–2155.

[55] D. Xie, X. Zhang, Q. Gao, J. Han, S. Xiao, and X. Gao, "Multiview clustering by joint latent representation and similarity learning," *IEEE Transs. Cybern.*, vol. 50, no. 11, pp. 4848–4854, Nov. 2020.

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

[57] L. Xing, B. Chen, S. Du, Y. Gu, and N. Zheng, "Correntropy-based multiview subspace clustering," *IEEE Trans. Cybern.*, in press, 2019.

[58] Q. Yin, S. Wu, R. He, and L. Wang, "Multi-view clustering via pairwise sparse subspace representation," *Neurocomputing*, vol. 156, no. 25, pp. 12–21, May 2015.

[59] K. Zhan, C. Zhang, J. Guan, and J. Wang, "Graph learning for multiview clustering," *IEEE Trans. Cybern.*, vol. 48, no. 10, pp. 2887–2895, Oct. 2018.

[60] C. Zhang, H. Fu, Q. Hu, X. Cao, Y. Xie, D. Tao, and D. Xu, "Generalized latent multi-view subspace clustering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 42, no. 1, pp. 86–99, Jan. 2020.

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

[62] D. Zhou, J. Huang, and B. Schölkopf, "Learning from labeled and unlabeled data on a directed graph," in *Proc. 22th Int.. Conf. Machine Learning*, Bonn, Germany, August 2005, pp. 1036–1043.

[63] T. Zhou, C. Zhang, X. Peng, H. Bhaskar, and J. Yang, "Dual shared-specific multiview subspace clustering," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp.. 3517 - 3530, Aug. 2020.

[64] X. Zhu, S. Zhang, W. He, R. Hu, C. Lei, and P. Zhu, "One-step multi-view spectral clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 10, pp. 2022–2034, Oct. 2018.

Shengxiang Yang (M'00–SM'14) received the Ph.D. degree from Northeastern University, Shenyang, China in 1999. He is currently a Professor in Computational Intelligence and Director of the Centre for Computational Intelligence, School of Computer Science and Informatics, De Montfort University, Leicester, U.K. He has over 330 publications with an H-index of 57 according to Google Scholar. His current research interests include evolutionary computation, swarm intelligence, artificial neural networks, data mining and data stream mining, and relevant real-world applications. He serves as an Associate Editor/Editorial Board Member of a number of international journals, such as the *IEEE Transactions on Evolutionary Computation*, *IEEE Transactions on Cybernetics*, *Information Sciences*, *Enterprise Information Systems*, and *CAAI Transactions on Intelligence Technology*.

Hua Mao received the B.S. degree and M.S. degree in Computer Science from University of Electronic Science and Technology of China (UESTC) in 2006 and 2009, respectively. She received her Ph.D. degree in Computer Science and Engineering from Aalborg University, Denmark in 2013. She is currently a Senior Lecturer in Department of Computer and Information Sciences, Northumbria University, U.K. Her current research interests include Deep Neural Networks and Big Data.

Jie Chen received the B.Sc. degree in Software Engineering, M.Sc. and Ph.D. degrees in Computer Science from Sichuan University, Chengdu, China, in 2005, 2008 and 2014, respectively. He is currently an Associate Professor in the College of Computer Science, Sichuan University, China. His current research interests include machine learning, big data analysis and deep neural networks.

Conor Fahy received the BSc degree in Computer Science from Dublin City University, Ireland in 2004. He received the MSc degree in Intelligent Systems and PhD degree in Computer Science from De Montfort University, Leicester, UK in 2016 and 2019, respectively. Since June 2019, he has been appointed as a VC2020 Lecturer in the School of Computer Science and Informatics, De Montfort University, Leicester, UK. His research interests include swarm intelligence and ensemble methods for unsupervised and semi-supervised learning in dynamic environments.