

Simultaneously learning feature-wise weights and local structures for multi-view subspace clustering

Shi-Xun Lin ^{a,1}, Guo Zhong ^{b,*}, Ting Shu ^c

^a School of Mathematics and Statistics, Zhaotong University, Zhaotong 657000, China

^b Department of Computer and Information Science, University of Macau, Macau, China

^c Guangdong-Hongkong-Macao Greater Bay Area Weather Research Center for Monitoring Warning and Forecasting, Shenzhen 518000, China

ARTICLE INFO

Article history:

Received 2 April 2020

Received in revised form 1 June 2020

Accepted 15 July 2020

Available online 24 July 2020

Keywords:

Local adaptive learning

Multi-view clustering

Subspace clustering

ABSTRACT

Multi-view clustering integrates multiple feature sets, which usually have a complementary relationship and can reveal distinct insights of data from different angles, to improve clustering performance. It remains challenging to productively utilize complementary information across multiple views since there is always noise in real data, and their features are highly redundant. Moreover, most existing multi-view clustering approaches only aimed at exploring the consistency of all views, but overlooked the local structure of each view. However, it is necessary to take the local structure of each view into consideration, because individual views generally present different geometric structures while admitting the same cluster structure. To ease the above issues, in this paper, a novel multi-view subspace clustering method is established by concurrently assigning weights for different features and capturing local information of data in view-specific self-representation feature spaces. In particular, a common clustering assignment regularization is adopted to explore the consistency among multiple views. An alternating iteration algorithm based on the augmented Lagrangian multiplier is also developed for optimizing the associated objective. Experiments conducted on diverse multi-view datasets manifest that the proposed method achieves state-of-the-art performance. We provide the Matlab code on <https://github.com/Ekin102003/JFLMSC>.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Clustering aims to organize reasonably unlabeled data to discover meaningful patterns [1–3]. Clustering has an important role in many fields, including machine learning [4,5], data mining [6,7], and pattern recognition [8,9]. Clustering methods can be mainly categorized into two groups [10]: partitioning clustering and hierarchical clustering. In particular, spectral clustering [11] is a graph-based algorithm for partitioning arbitrarily shaped data structure into disjoint clusters. Numerous spectral clustering methods and their variants have been proposed, such as Ratio Cut [12], K-way Ratio Cut [13], Min Cut [14], Normalized Cut (NCut) [15], and Spectral Embedded Clustering [16]. The clustering performance of all of these methods is largely dependent on the quality of the so-called similarity graph, which is learned according to the similarities between the corresponding data points.

Recent works [17–19] on spectral clustering-based subspace clustering have attracted considerable attention due to the

promising performance in data clustering. Subspace clustering works on the assumption that data are drawn from a union of low-dimensional subspaces, i.e., every subspace is equivalent to a cluster. According to the self-expression of data [17,20] and by imposing a properly chosen constraint on the representation coefficients, a representation matrix uncovering the intrinsic subspace structure of data can be obtained. For example, Elhamifar and Vidal [17] proposed sparse subspace clustering (SSC), which learns a graph by adaptively and flexibly selecting data points. Liu et al. [21] imposed a low-rank constraint on the representation matrix to capture the global structure of data. Dornaiika and Weng [22] incorporated a manifold regularization term into SSC to capture the manifold structure. More recently, Peng et al. [23] proposed a new deep model—Structured autoEncoder by preserving the global and local structure of subspace for clustering. Nevertheless, these approaches are mostly applied to group single-view data.

In the era of Big Data, various data sources are represented by multiple distinct feature sets [24]. For instance, one can describe an image via LBP, SIFT, HOG, CENTRIST, and GIST features. Each kind of feature is a particular view that describes different aspects of data. Traditional clustering methods can be used to group multi-view data by simply concatenating all views into a monolithic one. However, compatible and complementary information

* Corresponding author.

E-mail address: yb77410@um.edu.mo (G. Zhong).

¹ These authors contributed equally to this work.

across all views can be typically under-utilized. Recently, numerous multi-view clustering approaches [25–32] have become available. Bickel and Scheffer [33] generalized the conventional K -means and expectation–maximization clustering methods to the multi-view case. However, these methods are directly conducted in the original feature space, and then they fail to discover the geometric structures existing in each feature space. Canonical correlation analysis [34] is a prominent statistical approach for learning from multiple views, but is restricted to linear transformation for each view. Liu et al. [35] proposed a multi-view non-negative matrix factorization algorithm for clustering that learns a common representation of all views and then feeds it into K -means clustering. Such a common coefficient matrix does not consider the flexible structures of different feature spaces because of the heterogeneity of different views. Assuming that each sample should share the same cluster in all views, co-regularized spectral clustering (CRMSC) [36] proposed two co-regularization terms to force the consistency of clustering labels criss-crossing different views. The Laplacian matrix used in CRMSC is learned from the original data, and then the obtained low-dimensional representation of the data is used to perform clustering. The quality of the Laplacian matrix may not be good enough because of the influence of noise. As a result, the clustering performance may be seriously affected. To construct a reliable similarity graph from different views, a low-rank and sparse decomposition technique is adopted to improve the robustness of multi-view spectral clustering [37]. However, the above methods ignore the importance of different views, i.e., treating all of the features equally.

Recently, several multiple-graph learning-based clustering methods [26,29,38–41] have been established to identify clustering ability of different views. Auto-weighted multiple-graph learning [29] simultaneously learns an optimal weight for each Laplacian matrix and conducts spectral clustering. Multi-view learning with adaptive neighbors [26] can automatically learn the weights of all views during each iteration. Simultaneously, the local manifold structure of multi-view data is captured. The above methods achieve better performances compared to those methods that do not consider the weights of views. To learn a better graph from multiple views, Liang et al. [42] proposed to simultaneously explore the consistency and inconsistency across different views. However, the similarity graph is usually learned in the original feature space. Such a graph might be severely damaged, and the true similarities among samples cannot be guaranteed due to the influence of noise and redundancy.

The above-mentioned work assumes that data lie in a single subspace. This assumption contradicts the observation, i.e., in many problems, data in a cluster often lie in a low-dimensional subspace of the high-dimensional ambient space [43]. Recent studies [27,30,44–47] have focused on developing multi-view subspace clustering methods. For example, Gao et al. [44] unified classic spectral clustering and self-representation learning into a framework. Brbić and Kopriva [47] introduced simultaneous sparsity and low-rankness constraints on the coefficient matrix to extract underlying structure in multi-view data. Luo et al. [27] jointly exploited consistency and specificity for subspace representation learning by using a set of specific representations and a shared consistent representation. The above methods mainly treat all of the features of each view as a whole to learn a single representation or common representation of all views. However, features of real data may exhibit a high degree of redundancy, and hence unrelated information can be introduced to degrade clustering performance. Although various feature-selection methods have been proposed, conducting feature selection and clustering in two separate steps may not obtain the optimal clustering results [48].

Despite the recent progress within multi-view clustering, most existing methods ignore the feature-level relationship or the local

structure of multi-view data. Taking Fig. 1 as an example, five independent subspaces are constructed and then 20 vectors from each subspace are sampled. Fig. 1(a) shows an affinity matrix in which all of the vectors of the same subspace are connected, while those vectors belonging to other subspaces are not connected. Fig. 1(b) and (c) show the affinity matrices learned by LSR [49,50] and SSC [17], respectively. Compared to LSR and SSC, which capture the global and local structure of data, respectively, and ignore the feature-level relationship of data, the affinity matrix in Fig. 1(d) provided by the proposed method much better approximates the ideal affinity matrix. To ease the above limitations, a novel multi-view subspace clustering method is proposed based on data self-representation, which simultaneously weights features and learns the local structure of each view. By weighting the original features, the effective and robust features can be extracted, thus alleviating the influence of redundancy. Moreover, data self-representation and local structure learning are integrated into a unified framework such that the inherent difference in each view can be captured via simultaneously exploiting the global and local information of each view. In particular, a common cluster structure regularization is used to capture consistency across different views.

The significant contributions in this paper are the following.

1. A new multi-view subspace clustering method is proposed that captures view-specific information by simultaneously exploiting the underlying global representation information and local distance information from each view, while exploring the consistency of all views by spectral embedding. In particular, an adaptive distance regularization is integrated into the self-representation framework to explore the intrinsic structure of each view instead of using a predefined similarity matrix of each view as in previous methods.
2. View-specific weight vectors for each type of feature are learned so that the influence of redundancy and noise can be alleviated, and the similarities between samples can be accurately obtained to form a high-quality similarity graph.
3. A new augmented Lagrangian alternating direction method for solving the associated optimization problem is developed. Numerous experimental results on diverse multi-view datasets show the effectiveness of the proposed method.

The rest of this paper is organized as follows. First, a brief review of relevant work is given in Section 2. Next, in Section 3, the proposed method is described in detail. In Section 4, an optimization algorithm for solving the minimization problem is presented. In Section 5, the proposed method is verified through extensive experiments. The paper is concluded in Section 6.

2. Related work

For convenience, X is used to denote the sample matrix, i.e., each column of X denotes a sample, and the normal italic upper-case letters denote matrices. Symbols are summarized in Table 1. In this section, brief reviews of the related work, i.e., spectral clustering and spectral-clustering-based multi-view subspace clustering methods, are given.

2.1. Spectral clustering

Assuming that $X = [x_1, \dots, x_n] \in \mathbb{R}^{d \times n}$ is a given data matrix consisting of n samples, one can represent the relationships between samples by forming a similarity graph $G = (V, E)$, where E is a set of edges and V a set of vertices. Each vertex of G represents a sample and each edge has a weight reflecting the

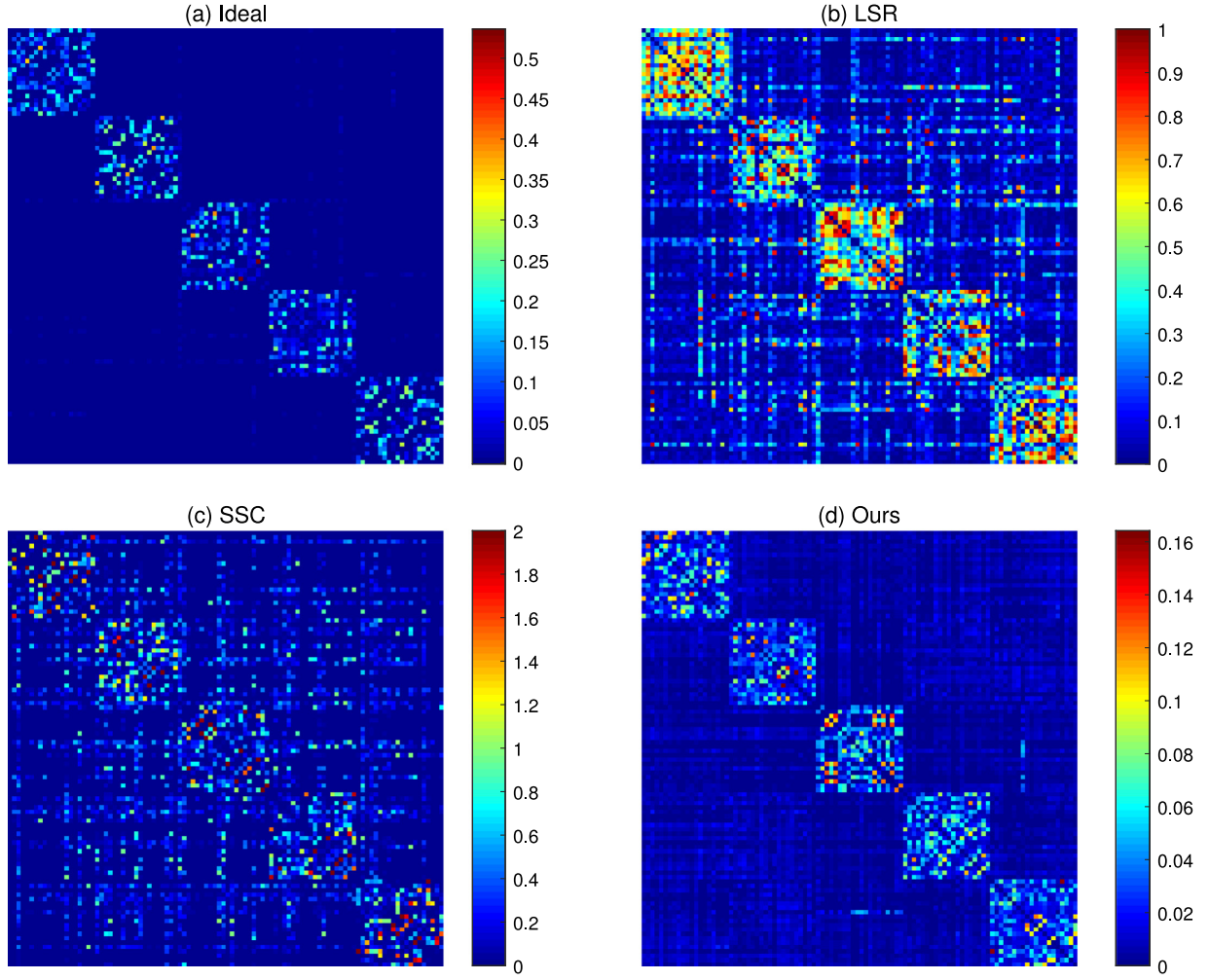


Fig. 1. Data consisting of five independent subspaces. Some data vectors are randomly chosen to corrupt by adding Gaussian noise with variance 0.1.

Table 1

Symbols used in this paper.

Symbols	Description
n	Number of samples
c	Number of clusters
n_v	Number of views
d_v	Dimension of views v
I	Unit matrix
$X \geq 0$	Non-negative matrix
x_i	i th column of a matrix X
x^j	j th row of X
x_{ij}	ij th entry of X
$\mathbf{1}$	Column vector of ones
$\mathbf{0}$	Column vector of zeros
$\ X\ _F$	$\ X\ _F = \sqrt{\sum_{i,j} x_{ij}^2}$
$X^{(v)}$	Data matrix of v th view
$\ X\ _1$	$\ X\ _1 = \sum_{i,j} x_{ij} $
$\ x\ _2$	2-norm of a vector x
X^T	Transpose of X
$Tr(X)$	Trace of X
$diag(u)$	Square diagonal matrix with elements of vector u on main diagonal
$diag(X)$	Column vector of main diagonal elements of X

similarity between two connected samples. Spectral clustering aims at partitioning the similarity graph such that the edges between different groups have very low weights and those within a group have high weights [11]. For this purpose, the objective of

the Ratio Cut spectral clustering method is stated as:

$$\min_{Q^T Q = I, Q \in \mathbb{R}^{n \times c}} Tr(Q^T L Q), \quad (1)$$

where the optimal solution Q^* of this problem consists of the first c eigenvectors of L . $L = D - S$ is the graph Laplacian, where $S \in \mathbb{R}^{n \times n}$ is the similarity graph and the diagonal elements of D can be computed by $d_{ii} = \sum_{j=1}^n s_{ij}$, i.e., $D = diag(d_{11}, \dots, d_{nn})$. The spectral embedding matrix Q^* is regarded as a clustering assignment.

To more effectively group multi-view data, a number of multi-view spectral clustering methods have been advanced [29,36,40,51,52]. Most of them learn a shared Laplacian matrix by unifying different views, while neglecting to learn high-quality similarity graphs to improve clustering performance further.

2.2. Multi-view subspace clustering

Assuming that X is a sample matrix, in which each sample is drawn from a union of subspaces, subspace clustering groups $x_1, \dots, x_n$ according to their subspaces.

Spectral-clustering-based subspace clustering methods first learn a representation matrix from the self-representation learning framework, and then utilize the learned representation matrix to form a Laplacian matrix for spectral clustering. In general, its model can be formulated as:

$$\min_{Z \in \mathbb{R}^{n \times n}} \mathcal{R}(X, XZ) + \lambda \Psi(Z), \quad (2)$$

where $Z = [z_1, z_2, \dots, z_n] \in \mathbb{R}^{n \times n}$ is a self-representation matrix. Every column z_i of Z is considered as a new representation of the sample x_i in terms of other samples in X . Ideally, those nonzero elements z_{ij} in Z mean that samples x_i and x_j are from the same cluster. $\lambda > 0$ is used to balance the regularization term $\Psi(\cdot)$ and the reconstruction error term $\mathcal{R}(\cdot, \cdot)$. By solving problem (2) to obtain the representation matrix Z , one easily builds the similarity graph $S = \frac{|Z| + |Z^T|}{2}$. Finally, S is fed into a spectral clustering algorithm, such as [13],² to obtain the clustering result.

Recently, multi-view subspace clustering methods have been designed to fuse complementary and compatible information from multiple views. Cao et al. [53] proposed utilizing the Hilbert Schmidt Independence Criterion as a diversity term with which to explore the complementarity of multi-view representations. Wang et al. [54] explored the complementary information between different representations by introducing a novel position-aware exclusivity term. State-of-the-art consistent and specific multi-view subspace clustering [27] simultaneously learns a view-consistent representation and a set of view-specific representations for multi-view subspace clustering. The general formulation of multi-view subspace clustering can be written as the following form:

$$\min_{Z^{(v)} \in \mathbb{R}^{n \times n}} \sum_{v=1}^{n_v} [\mathcal{R}(X^{(v)}, X^{(v)}Z^{(v)}) + \lambda \Psi(Z^{(v)})]. \quad (3)$$

By solving problem (3) to obtain $Z^{(v)}$, $v = 1, \dots, n_v$, the final clustering results can be obtained by performing spectral clustering [13] on the similarity graph

$$S = \frac{1}{n_v} \sum_{v=1}^{n_v} \frac{|Z^{(v)}| + |Z^{(v)T}|}{2}. \quad (4)$$

3. Methodology

In this section, to explore better representation capacity of data and more flexible local manifold structures of different views, the proposed method, which unifies the featurewise weight learning and adaptive local structure learning into a framework, is derived.

Previous works [55–57] demonstrate the importance of local structure learning in subspace clustering. Most of them mainly utilized a pre-computed graph regularizer to capture local manifold structure of data, i.e.,

$$\min_Z \frac{1}{2} \sum_{i,j} \|z_i - z_j\|_2^2 b_{ij} = \text{Tr}(Z^T L_B Z), \quad (5)$$

where L_B is a Laplacian matrix and B a pre-computed similarity matrix. By integrating Eq. (5) into the self-representation learning framework Eq. (2), the learned representations can preserve the local geometry structure embedded in a high-dimensional ambient space. However, b_{ij} must be pre-computed; that is, to capture the local structure in self-representation learning framework, a two-stage process is adopted, which can result in information loss. We will integrate self-representation learning and local structure learning into a unified framework. However, the non-negativity is more consistent with the biological modeling of visual data and often leads to better performance for data representation [58]. Moreover, each view should share a consensus clustering assignment. As a result, by Eq. (1), the following objective is proposed first:

$$\min_{Z^{(v)}, E^{(v)}, Q} \sum_{v=1}^{n_v} \sum_{i=1}^n \sum_{j=1}^n \|x_i^{(v)} - x_j^{(v)}\|_2^2 z_{ij}^{(v)} \quad (6)$$

$$+ \sum_{v=1}^{n_v} [\lambda_2 \|E^{(v)}\|_1 + \lambda_1 \text{Tr}(Q^T L_{Z^{(v)}} Q)],$$

$$\text{s.t. } X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)}, Z^{(v)} \geq 0, Z^{(v)}\mathbf{1} = \mathbf{1},$$

$$\text{diag}(Z^{(v)}) = \mathbf{0}, Q^T Q = I,$$

where $Q \in \mathbb{R}^{n \times c}$ is the consensus clustering assignment across all views. $L_{Z^{(v)}} = D^{(v)} - \frac{Z + Z^T}{2}$ is the Laplacian matrix, where $D^{(v)} = \text{diag}(\frac{Z^{(v)} + Z^{(v)T}}{2} \mathbf{1})$. λ is a positive penalty parameter. Owing to the non-negative constraint, each entry z_{ij} of $Z^{(v)}$ can directly expose the similarity between $x_i^{(v)}$ and $x_j^{(v)}$. Concretely, the first term of Eq. (6) guides the representation coefficients to capture the local structure. In practical applications, data may be contaminated by noise. Thus, a sparse error term $\|E^{(v)}\|_1$ is introduced. Close samples should possess a high similar degree and distant samples should possess a low or even zero similar degree. Specifically, $Z^{(v)}\mathbf{1} = \mathbf{1}$ is further constrained from the point of view of probability, i.e., $0 \leq z_{ij} < 1$.

To capture the global structure, following [59], a 2-norm constraint is added onto the representation matrix instead of a low-rank constraint since one may obtain bad representations when the dimension of a certain view is far smaller than the number of samples [44]. Thus, one has the following minimization problem:

$$\min_{Z^{(v)}, Q, E^{(v)}} \sum_{v=1}^{n_v} \sum_{i,j} \|x_i^{(v)} - x_j^{(v)}\|_2^2 z_{ij}^{(v)} + \lambda_1 \text{Tr}(Q^T L_{Z^{(v)}} Q)$$

$$+ \sum_{v=1}^{n_v} [\lambda_2 \|Z^{(v)}\|_2 + \lambda_3 \|E^{(v)}\|_1], \quad (7)$$

$$\text{s.t. } X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)}, Z^{(v)} \geq 0,$$

$$\text{diag}(Z^{(v)}) = \mathbf{0}, Q^T Q = I, Z^{(v)}\mathbf{1} = \mathbf{1},$$

where $\|Z^{(v)}\|_2$ is the maximum singular value of $Z^{(v)}$. λ_1, λ_2 , and λ_3 are positive penalty parameters.

Note that the first term of Eq. (7) uses Euclidean distance computed on the original data to guide the local structure learning. Doing this may not reflect the true closeness between x_i and x_j due to the sensitivity of Euclidean distance to noise and redundancy. To alleviate this problem, the importance of different features is learned. Consequently, the final multi-view subspace clustering framework is stated as follows:

$$\min_{Z^{(v)}, Q, E^{(v)}, w^{(v)}} \sum_{v=1}^{n_v} \sum_{i=1}^n \sum_{j=1}^n \|W^{(v)} x_i^{(v)} - W^{(v)} x_j^{(v)}\|_2^2 z_{ij}^{(v)} \quad (8)$$

$$+ \sum_{v=1}^{n_v} [\lambda_2 \|Z^{(v)}\|_2 + \lambda_3 \|E^{(v)}\|_1 + 2\lambda_1 \text{Tr}(Q^T L_{Z^{(v)}} Q)],$$

$$\text{s.t. } X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)}, Z^{(v)} \geq 0,$$

$$\text{diag}(Z^{(v)}) = \mathbf{0}, Q^T Q = I, Z^{(v)}\mathbf{1} = \mathbf{1},$$

$$W^{(v)} = \text{diag}(w^{(v)}), w^{(v)} \geq 0, w^{(v)}\mathbf{1} = \mathbf{1},$$

where $w^{(v)} \in \mathbb{R}^{1 \times d_v}$ is the self-weighted vector for feature of the v th view. $w^{(v)}$ adaptively learns the importance of different features in the v th view.

One can see that the objective function in Eq. (8) simultaneously takes into consideration the global and local structures of each view. Specifically, the local structure is learned through more discriminative features instead of pre-computing a locality constraint on the original feature space. Furthermore, the consistency interweaving in different views according to the clustering assignment matrix Q shared by all views is captured, since each view should have a different representation matrix resulting from the heterogeneity of features from different views, but they have the same clustering structure. The proposed method is referred to as Joint Featurewise Weighting and Local Structure Learning for Multi-view Subspace Clustering (JFLMSC).

² The code is available at <http://www.cis.upenn.edu/~jshi/software/>.

4. Optimization algorithm

The augmented Lagrange multiplier (ALM) method is used to solve the challenging problem (8). Specifically, the problem (8) can be optimized alternatively. Auxiliary variables, i.e., $Z^{(v)} = A^{(v)}$ and $Z^{(v)} = U^{(v)}$, $v \in \{1, \dots, n_v\}$, are introduced. Problem (8) is converted to the following optimization problem:

$$\begin{aligned} \min_{\substack{Z^{(v)}, E^{(v)}, A^{(v)}, \\ U^{(v)}, Q, w^{(v)}}} & \sum_{v=1}^{n_v} \sum_{i=1}^n \sum_{j=1}^n \|W^{(v)}x_i^{(v)} - W^{(v)}x_j^{(v)}\|_2^2 a_{ij}^{(v)} \\ & + \sum_{v=1}^{n_v} [\lambda_2 \|U^{(v)}\|_2 + \lambda_3 \|E^{(v)}\|_1 + 2\lambda_1 \text{Tr}(Q^T L_{A^{(v)}} Q)], \\ \text{s.t. } & X^{(v)} = X^{(v)} Z^{(v)} + E^{(v)}, Z^{(v)} = U^{(v)}, Z^{(v)} = A^{(v)}, \\ & A^{(v)} \geq 0, \text{diag}(A^{(v)}) = \mathbf{0}, A^{(v)} \mathbf{1} = \mathbf{1}, Q^T Q = I, \\ & W^{(v)} = \text{diag}(w^{(v)}), w^{(v)} \geq 0, w^{(v)} \mathbf{1} = 1. \end{aligned} \quad (9)$$

The augmented Lagrangian function of Eq. (9) is

$$\begin{aligned} \mathcal{L}(Z^{(v)}, E^{(v)}, Q, A^{(v)}, U^{(v)}, w^{(v)}) & \\ = & \sum_{v=1}^{n_v} \sum_{i=1}^n \sum_{j=1}^n \|W^{(v)}x_i^{(v)} - W^{(v)}x_j^{(v)}\|_2^2 a_{ij}^{(v)} \\ & + \sum_{v=1}^{n_v} [\lambda_2 \|U^{(v)}\|_2 + \lambda_3 \|E^{(v)}\|_1 + 2\lambda_1 \text{Tr}(Q^T L_{A^{(v)}} Q)] \\ & + \sum_{v=1}^{n_v} \left\langle \Lambda_1^{(v)}, X^{(v)} - X^{(v)} Z^{(v)} - E^{(v)} \right\rangle \\ & + \sum_{v=1}^{n_v} \left\langle \Lambda_2^{(v)}, Z^{(v)} - U^{(v)} \right\rangle + \left\langle \Lambda_3^{(v)}, Z^{(v)} - A^{(v)} \right\rangle \\ & + \sum_{v=1}^{n_v} \frac{\mu}{2} \|X^{(v)} - X^{(v)} Z^{(v)} - E^{(v)}\|_F^2 \\ & + \sum_{v=1}^{n_v} \frac{\mu}{2} (\|Z^{(v)} - U^{(v)}\|_F^2 + \|Z^{(v)} - A^{(v)}\|_F^2), \end{aligned} \quad (10)$$

where $\Lambda_1^{(v)}$, $\Lambda_2^{(v)}$, and $\Lambda_3^{(v)}$ are Lagrangian multipliers, μ is a positive penalty scalar, and $\langle \cdot, \cdot \rangle$ denotes the matrix inner product. Now, Eq. (10) can be solved by alternately optimizing each variable with the other variables fixed. It can be seen that Eq. (10) is reduced to six sub-problems. The optimization for each sub-problem is as follows:

4.0.1. $Z^{(v)}$ -subproblem

When fixing other variables except $Z^{(v)}$, Eq. (10) becomes

$$\begin{aligned} \mathcal{L}(Z^{(v)}) &= \|X^{(v)} - X^{(v)} Z^{(v)} - E^{(v)}\|_F^2 + \frac{\Lambda_1^{(v)}}{\mu} \|X^{(v)} - X^{(v)} Z^{(v)} - E^{(v)}\|_F^2 \\ &+ \|Z^{(v)} - U^{(v)}\|_F^2 + \frac{\Lambda_2^{(v)}}{\mu} \|Z^{(v)} - U^{(v)}\|_F^2 \\ &+ \|Z^{(v)} - A^{(v)}\|_F^2 + \frac{\Lambda_3^{(v)}}{\mu} \|Z^{(v)} - A^{(v)}\|_F^2. \end{aligned} \quad (11)$$

It suffices to take the derivative of $\mathcal{L}(Z^{(v)})$ w.r.t. $Z^{(v)}$ and set it to zero. The closed-form solution of Eq. (11) is:

$$Z^{(v)} = (X^{(v)T} X^{(v)} + 2I)^{-1} (X^{(v)T} V_1 + V_2 + V_3), \quad (12)$$

where $V_1 = X^{(v)} - E^{(v)} + \frac{\Lambda_1^{(v)}}{\mu}$, $V_2 = U^{(v)} - \frac{\Lambda_2^{(v)}}{\mu}$, and $V_3 = A^{(v)} - \frac{\Lambda_3^{(v)}}{\mu}$.

4.0.2. $A^{(v)}$ -subproblem

By ignoring the irrelevant terms of Eq. (10), one can obtain $A^{(v)}$ by minimizing

$$\begin{aligned} \min_{A^{(v)}} & \sum_{i=1}^n \sum_{j=1}^n \|W^{(v)}x_i^{(v)} - W^{(v)}x_j^{(v)}\|_2^2 a_{ij}^{(v)} \\ & + 2\lambda_1 \text{Tr}(Q^T L_{A^{(v)}} Q) + \frac{\mu}{2} \|Z^{(v)} - A^{(v)} + \frac{\Lambda_3^{(v)}}{\mu}\|_F^2, \\ \text{s.t. } & A^{(v)} \geq 0, \text{diag}(A^{(v)}) = \mathbf{0}, A^{(v)} \mathbf{1} = \mathbf{1}. \end{aligned} \quad (13)$$

Note that

$$\frac{1}{2} \sum_{i,j=1}^n \|q^i - q^j\|_2^2 a_{ij}^{(v)} = \text{Tr}(Q^T L_{A^{(v)}} Q). \quad (14)$$

To simplify the notations, the view index is tentatively ignored. Eq. (13) can be immediately decoupled into rows, and then the following problem is solved to obtain the i th row of A :

$$\begin{aligned} \min_{a^i} & \sum_{j=1}^n (\|Wx_i - Wx_j\|_2^2 a_{ij} + \lambda_1 \|q^i - q^j\|_2^2 a_{ij}) \\ & + \sum_{j=1}^n (\frac{\mu}{2} a_{ij}^2 - \mu a_{ij} h_{ij}), \\ \text{s.t. } & a^i \geq 0, a_{ii} = 0, a^i \mathbf{1} = 1, \end{aligned} \quad (15)$$

where h_{ij} is the (i, j) th element of $H = Z + \frac{\Lambda_3}{\mu}$. Letting

$$d_{ij} = \|Wx_i - Wx_j\|_2^2 + \lambda_1 \|q^i - q^j\|_2^2 - \mu h_{ij} \quad (16)$$

be the j th element of $d_i \in \mathbb{R}^{1 \times n}$, it is easy to verify that Eq. (15) can be transformed into:

$$\min_{a^i \geq 0, a_{ii}=0, a^i \mathbf{1}=1} \|a^i + \frac{d_i}{\mu}\|_2^2. \quad (17)$$

The Lagrangian function of Eq. (17) is:

$$\mathcal{L}(a^i, \eta, \beta) = \|a^i + \frac{d_i}{\mu}\|_2^2 + \eta(1 - a^i \mathbf{1}) + \beta^T (-a^i), \quad (18)$$

where η and $\beta \geq \mathbf{0}$ are the Lagrangian multipliers. It follows from the Karush-Kuhn-Tucker (KKT) conditions [60] that the optimal solution a^i is:

$$a^i = (-\frac{d_i}{\mu} + \eta \mathbf{1}^T)_+, \quad (19)$$

where $(\cdot)_+ = \max(\cdot, 0)$.

4.0.3. Q -subproblem

Treating other variables as constants, Eq. (10) transforms into the following problem:

$$\min_{Q \in \mathbb{R}^{n \times c}, Q^T Q = I} \text{Tr}(Q^T (\sum_{v=1}^{n_v} L_{A^{(v)}}) Q). \quad (20)$$

To solve problem (20), it suffices to calculate the eigen-decomposition of $\sum_{v=1}^{n_v} L_{A^{(v)}}$. The optimal solution Q^* consists of the c eigenvectors of $\sum_{v=1}^{n_v} L_{A^{(v)}}$ w.r.t. the c smallest eigenvalues.

4.0.4. $U^{(v)}$ -subproblem

Fixing all variables except $U^{(v)}$, the following problem is solved:

$$\min_{U^{(v)}} \lambda_2 \|U^{(v)}\|_2 + \frac{\mu}{2} \|Z^{(v)} - U^{(v)} + \frac{\Lambda_2^{(v)}}{\mu}\|_F^2, \quad (21)$$

which can be solved by Proposition 4 [59].

4.0.5. $E^{(v)}$ -subproblem

All variables except $E^{(v)}$ are fixed to solve the following problem:

$$\min_{E^{(v)}} \lambda_3 \|E^{(v)}\|_1 + \frac{\mu}{2} \|X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)} + \frac{\Lambda_1^{(v)}}{\mu}\|_F^2. \quad (22)$$

Following [61], one can obtain:

$$E^{(v)} = \Omega_{\frac{\lambda_3}{\mu}}(X^{(v)} - X^{(v)}Z^{(v)} + \frac{\Lambda_1^{(v)}}{\mu}), \quad (23)$$

where Ω denotes the shrinkage operator.

4.0.6. $w^{(v)}$ -subproblem

When fixing the other variables, the problem (10) is equivalent to the following problem:

$$\begin{aligned} \min_{w^{(v)}} \sum_{i,j} \|W^{(v)}x_i^{(v)} - W^{(v)}x_j^{(v)}\|_2^2 a_{ij}^{(v)} \\ \text{s.t. } W^{(v)} = \text{diag}(w^{(v)}), w^{(v)} \geq 0, w^{(v)}\mathbf{1} = 1. \end{aligned} \quad (24)$$

According to Eq. (14), problem (24) is rewritten as follows:

$$\begin{aligned} \min_{w^{(v)}} \text{Tr}(W^{(v)}X^{(v)}L_{A^{(v)}}X^{(v)T}W^{(v)}) \\ \text{s.t. } W^{(v)} = \text{diag}(w^{(v)}), w^{(v)} \geq 0, w^{(v)}\mathbf{1} = 1. \end{aligned} \quad (25)$$

The problem (25) can be further transformed into:

$$\begin{aligned} \min_{w^{(v)}} \text{Tr}(W^{(v)}YW^{(v)}) \\ \text{s.t. } W^{(v)} = \text{diag}(w^{(v)}), w^{(v)} \geq 0, w^{(v)}\mathbf{1} = 1, \end{aligned} \quad (26)$$

where $Y = X^{(v)}L_{A^{(v)}}X^{(v)T}$. Eq. (26) is simplified as:

$$\begin{aligned} \min_{w^{(v)}} w^{(v)}\mathbf{y}w^{(v)T} \\ \text{s.t. } w^{(v)} \geq 0, w^{(v)}\mathbf{1} = 1, \end{aligned} \quad (27)$$

The Lagrangian function of Eq. (27) is:

$$\mathcal{L}(w^{(v)}, \eta) = w^{(v)}\mathbf{y}w^{(v)T} - \eta(w^{(v)}\mathbf{1} - 1), \quad (28)$$

where $\mathbf{y} = \text{diag}(y_{11}, \dots, y_{d_v d_v})$ and η is the Lagrangian multiplier.

By taking the derivative of Eq. (28) w.r.t. $w^{(v)}$ and setting it to zero, the closed-form solution of Eq. (24) is obtained as follows:

$$w^{(v)} = \left(\frac{1}{y_{11} \sum_{j=1}^{d_v} \frac{1}{y_{jj}}}, \dots, \frac{1}{y_{d_v d_v} \sum_{j=1}^{d_v} \frac{1}{y_{jj}}} \right), \quad (29)$$

where the constraint $w^{(v)}\mathbf{1} = 1$ is used.

4.0.7. Multiplier:

μ , $\Lambda_1^{(v)}$, $\Lambda_2^{(v)}$, and $\Lambda_3^{(v)}$ are updated as follows:

$$\begin{aligned} \mu &= \min(\rho\mu, \mu_{\max}), \\ \Lambda_1^{(v)} &= \Lambda_1^{(v)} + \mu(X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}), \\ \Lambda_2^{(v)} &= \Lambda_2^{(v)} + \mu(Z^{(v)} - U^{(v)}), \\ \Lambda_3^{(v)} &= \Lambda_3^{(v)} + \mu(Z^{(v)} - S^{(v)}), \end{aligned} \quad (30)$$

where ρ controls the rate of convergence, $\mu_{\max}$ and ρ are pre-set constants.

4.1. Computation complexity

An augmented Lagrange multiplier with an alternating direction minimizing (ALM-ADM) algorithm has been derived for the proposed JFLMSC. For clarity, the optimization algorithm is summarized in Algorithm 1. The computational complexity is comparable with many state-of-the-art multi-view subspace clustering methods [27,30,47,62]. In particular, the highest computational

Algorithm 1 Solving Problem (8) via ALM-ADM Algorithm

Input: Multi-view data: $X^{(1)} \in \mathbb{R}^{d_1 \times n}, \dots, X^{(n_v)} \in \mathbb{R}^{d_{n_v} \times n}$, and regularization parameters λ_1, λ_2 , and λ_3 .

Output: $Q, Z^{(v)}, U^{(v)}, E^{(v)}, A^{(v)}, w^{(v)}, \forall v = 1, \dots, n_v$.

```

1: initialization:  $\mu > 0, \rho > 1, \Lambda_1^{(v)} = \mathbf{0}, \Lambda_2^{(v)} = \Lambda_3^{(v)} = \mathbf{0}, Z^{(v)} = A^{(v)}, U^{(v)} = Z^{(v)}, E^{(v)} = \mathbf{0}$  and  $Z^{(v)}$  is an affinity matrix based on  $k$ th nearest-neighbor graph.
2: repeat
3:   while  $v \leq n_v$  do
4:     Update  $Z^{(v)}$  using Eq. (12).
5:     Update  $A^{(v)}$  line by line using Eq. (19).
6:     Update  $U^{(v)}$  by solving problem (21).
7:     Update  $E^{(v)}$  using Eq. (23).
8:     Update  $w^{(v)}$  using Eq. (29).
9:     Update  $\Lambda_1^{(v)}, \Lambda_2^{(v)}$ , and  $\Lambda_3^{(v)}$  using Eq. (30).
10:  end while
11:  Update  $Q$  by solving problem (20).
12:  Update  $\mu$  using Eq. (30).
13: until convergence

```

costs in 1 are incurred in Steps 1, 3, and 4. The complexity for updating $Z^{(v)}$ in Step 1 is $O(n^3)$. The Woodbury formula [63] is utilized to compute the inversion in Eq. (12). Then, the complexity in Step 1 is reduced to $O(d_v n^2)$. One only needs to compute the inversion once since it is not updated during the iterations. The complexity for updating F is $O(nd_v c)$ by using the partial singular value decomposition (SVD). Generally, Step 4 takes $O(n^3)$ due to the full SVD. One can utilize the partial SVD to improve the efficiency of Step 4. The complexity for updating $U^{(v)}$ is reduced to $O(r_v n^2)$, where r_v is the rank of $Z^{(v)} - U^{(v)} + \frac{\Lambda_2^{(v)}}{\mu}$, and the rank prediction strategy [64] is used. Therefore, the total time complexity of Algorithm 1 is $O(n_v(d_v n^2 + tr_v n^2))$, where t is the iteration number. Moreover, one may exploit a more efficient approach motivated by the recent studies [65,66] on large-scale clustering. Theoretically proving the convergence of the ALM-ADM algorithm is difficult [27,30,45–47]. As in previous work [27,46,47], the convergence in the experiment is empirically verified. However, under mild conditions, one can show that any limit point of the iteration sequence generated by Algorithm 1 is a stationary point that satisfies the KKT conditions. Detailed convergence analysis can be found in [67].

5. Experiments

In this section, experiments with five public datasets are performed to evaluate the clustering performance of the proposed JFLMSC. Benchmarking clustering is generally difficult [27]. Following previous work [40], five evaluation metrics, i.e., ACC (Clustering Accuracy), NMI (Normalized Mutual Information), Precision, F-score, and ARI (Adjusted Rand Index), are adopted.

5.1. Datasets

In the experiments, a diversity of five datasets, i.e., 100leaves, MSRC-v1, ORL, BBCSport, and Outdoor Scene, was collected to confirm that the proposed JFLMSC is effective.

1. MSRC-v1 dataset [68] is from Microsoft Research in Cambridge and includes 210 images and nine object classes. Following [69], GIST is extracted with dimension 512, HOG with dimension 100, LBP with dimension 256, SIFT with dimension 210, color moment with dimension 48, and CENTRIST with dimension 1302 visual features from each image.

Table 2
Descriptions of datasets.

Datasets	100leaves	MSRC-v1	ORL	Outdoor Scene	BBCSport
Number of samples	1600	210	400	2688	544
Number of clusters	100	7	40	8	2
#d ₁	64	254	4096	432	3183
#d ₂	64	24	3304	256	3203
#d ₃	64	512	6750	512	–
#d ₄	–	576	–	48	–
#d ₅	–	256	–	–	–

2. ORL dataset [70] is a database of faces from AT&T Laboratories Cambridge, which contains 10 different images of each of 40 distinct subjects. For ORL, three types of features are extracted, i.e., the intensity with dimension 4096, LBP with dimension 3304, and Gabor with dimension 6750.
3. Outdoor Scene [40] is commonly used for evaluating the performance of full-scene segmentation. For each image, GIST is extracted with dimension 512, color moment with dimension 432, HOG with dimension 256, and LBP with dimension 48.
4. BBCSport [42] consists of 544 documents from the BBC sports website, including sports news articles in five hot areas from 2004 to 2005.
5. One-hundred plant species leaves dataset (100leaves) [71] consists of 1600 samples from each of 100 plant species. For each sample, texture histogram, fine-scale margin, and shape descriptor are given.

The datasets used are summarized in Table 2, in which d_v denotes the dimension of features in view v .

5.2. Compared methods

The proposed JFLMSC method was compared with the following baseline methods, including several closely related state-of-the-art approaches.

1. Normalized Cut (NCut) [15] is a traditional spectral clustering approach for clustering single-view data. The features of each view of multi-view data are concatenated in a column-wise manner to feed into NCut.
2. Co-regularized Multi-view Spectral Clustering (CRMSC) [36]³ is a spectral clustering framework that achieves exploiting information from multiple views by co-regularizing the consistency across the views.
3. Multi-View Clustering with Adaptive Neighbors (MCAN) [26]⁴ is a novel multi-view learning model that performs clustering and local structure learning simultaneously. Moreover, MCAN can allocate an ideal weight for each view automatically.
4. Auto-Weighted Multiple Graph Learning (AMGL) [29]⁵ is a novel framework obtained by the reformulation of the standard spectral learning model, which can learn an optimal weight for each graph automatically.
5. Self-weighted Multi-view Clustering with Multiple Graphs (SWMC) [39]⁶ explores a Laplacian rank-constrained graph, which can be approximated as the centroid of the built graph for each view with different confidences.

³ The code is available at <http://users.umi.acs.umd.edu/~abhishek/papers.html>.

⁴ The code is available at <http://www.esience.cn/people/fpnie/index.html>.

⁵ The code is available at <http://www.esience.cn/people/fpnie/index.html>.

⁶ The code is available at <http://www.esience.cn/people/fpnie/index.html>.

Table 3
Clustering results on MSRC-v1 dataset.

Methods	Metrics				
	ACC (%)	NMI (%)	ARI (%)	Precision (%)	F-score (%)
NCut [15]	63.81	51.00	40.81	47.16	49.14
CRMSC [36]	78.10	66.77	60.59	65.12	66.15
MCAN [26]	88.10	77.74	74.25	80.36	77.86
AMGL [29]	80.00	73.94	65.44	75.04	70.15
LMSC [30]	78.57	69.69	61.82	66.18	67.20
SWMC [39]	78.10	73.75	67.23	78.29	72.03
CSMSC [27]	84.76	75.02	69.96	73.39	74.17
MLRSSC [72]	82.71	74.22	68.49	74.83	72.98
NESE [40]	82.38	73.28	68.52	71.47	72.65
JFLMSC	91.43	82.30	81.06	83.54	83.69

6. Latent Multi-view Subspace Clustering (LMSC) [30]⁷ clusters data points with latent representation and simultaneously explores underlying complementary information from multiple views.
7. Consistent and Specific Multi-View Subspace Clustering (CSMSC) [27]⁸ is a novel multi-view subspace clustering method, where consistency and specificity are jointly exploited for subspace representation learning.
8. Multi-view Low-rank Sparse Subspace Clustering (MLRSSC) [72]⁹ learns a joint subspace representation by constructing affinity matrix shared among all views.
9. Multi-view Spectral Clustering via Integrating Non-negative Embedding and Spectral Embedding (NESE) [40]¹⁰ inherits the advantages of both graph-based and matrix factorization methods.

5.3. Experimental settings

In this work, the cluster number was set to just the true number of classes, which can be estimated by existing algorithms. All views of multi-view datasets were concatenated to form a new single-view dataset for NCut. NCut, CRMSC, AMGL, and NESE constructed the affinity matrix by the Gaussian kernel, in which the constant 1 was used as the Gauss kernel parameter. The parameter λ in CRMSC was chosen from 0.01 to 0.05 with a step 0.01. The k th-nearest-neighbor graph was constructed for MCAN and SWMC. k was selected from 3 to 15. For the state-of-the-art multi-view subspace clustering methods MLRSSC, LMSC, CSMSC, and the proposed JFLMSC, the parameter values were chosen from the set {0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10, 100}. To guarantee convergence of algorithms, the maximum number of iterations was set to 200. Each algorithm was run 20 times, and then their best performances are recorded.

5.4. Clustering results

The proposed JFLMSC was compared with the competing methods in terms of ACC, NMI, ARI, F-score, and Precision. Tables 3–7 present the clustering results on the five public datasets, and the best results are highlighted in bold and the second-best are underlined for each metric. A larger value implies better clustering performance for all five metrics. In the majority of cases, the proposed JFLMSC method achieves the best performance compared to the other state-of-the-art multi-view clustering methods. Additional important findings follow.

⁷ The code is available at <http://cic.tju.edu.cn/faculty/zhongchangqing/code.html>.

⁸ The code is available at <https://github.com/XIAOCHUN-CAS/Consistent-and-Specific-Multi-View-Subspace-Clustering>.

⁹ The code is available at <https://github.com/mbrbic/Multi-view-LRSSC>.

¹⁰ The code is available at <https://github.com/sudalvxin/SMSC.git>.

Table 4
Clustering results on 100leaves dataset.

Methods	Metrics				
	ACC (%)	NMI (%)	ARI (%)	Precision (%)	F-score (%)
NCut [15]	67.19	83.76	56.48	51.61	56.93
CRMSC [36]	80.06	92.12	75.55	71.01	76.56
MCAN [26]	89.95	94.78	83.30	87.89	83.46
AMGL [29]	82.06	90.64	60.34	83.93	62.42
LMSC [30]	77.44	88.33	65.15	65.70	68.60
SWMC [39]	91.25	95.35	78.44	92.16	78.67
CSMSC [27]	79.19	89.93	72.41	69.22	72.69
MLRSSC [72]	70.94	85.30	60.45	64.25	60.84
NESE [40]	90.69	94.25	85.15	84.45	84.80
JFLMSC	95.44	97.83	93.46	92.22	93.53

Table 5
Clustering results on Outdoor Scene dataset.

Methods	Metrics				
	ACC (%)	NMI (%)	ARI (%)	Precision (%)	F-score (%)
NCut [15]	40.33	26.78	19.86	26.59	32.17
CRMSC [36]	60.27	46.07	38.61	46.52	46.38
MCAN [26]	62.17	52.06	41.19	33.00	50.07
AMGL [29]	60.49	54.97	42.52	44.28	51.09
LMSC [30]	69.12	51.16	44.77	50.62	51.97
SWMC [39]	60.83	53.48	42.65	50.32	51.67
CSMSC [27]	73.48	57.96	52.19	56.51	58.47
MLRSSC [72]	69.20	54.52	47.54	56.35	54.42
NESE [40]	71.47	54.95	45.58	52.32	54.52
JFLMSC	72.58	61.08	54.95	57.13	61.47

Table 6
Clustering results on ORL dataset.

Methods	Metrics				
	ACC (%)	NMI (%)	ARI (%)	Precision (%)	F-score (%)
NCut [15]	65.75	79.07	50.71	50.06	51.90
CRMSC [36]	73.78	86.93	65.89	62.41	66.65
MCAN [26]	72.75	83.84	49.46	75.89	50.97
AMGL [29]	74.00	85.20	56.95	48.86	58.15
LMSC [30]	83.75	92.93	79.94	76.13	80.42
SWMC [39]	77.25	88.22	66.68	80.78	67.52
CSMSC [27]	83.00	90.05	75.02	71.25	75.62
MLRSSC [72]	78.75	87.89	68.73	74.61	69.47
NESE [40]	77.00	88.14	70.11	64.19	70.85
JFLMSC	84.75	93.31	81.39	77.38	81.83

Table 7
Clustering results on BBCSport dataset.

Methods	Metrics				
	ACC (%)	NMI (%)	ARI (%)	Precision (%)	F-score (%)
NCut [15]	60.85	33.57	31.08	40.25	52.54
CRMSC [36]	65.74	54.39	42.90	55.97	62.59
MCAN [26]	72.43	70.98	60.65	96.27	72.37
AMGL [29]	71.51	67.42	57.45	55.64	70.26
LMSC [30]	88.97	71.40	72.22	80.14	78.74
SWMC [39]	73.16	65.68	55.23	91.81	68.61
CSMSC [27]	96.26	88.09	90.18	92.79	93.15
MLRSSC [72]	96.69	89.38	91.59	92.43	93.56
NESE [40]	92.83	83.57	84.04	83.69	88.04
JFLMSC	96.88	89.92	92.18	94.05	94.03

1. Compared to the single view spectral clustering method NCut, all of the multi-view clustering methods gain better results. For example, CRMSC outperforms NCut by nearly 20% in terms of ARI on the MSRC-v1 dataset. In particular, the improvement of the proposed JFLMSC method with respect to NCut is approximately 28%, 31%, 41%, 36%, and 34% on ACC, NMI, ARI, Precision, and F-score, respectively. This phenomenon demonstrates that multi-view clustering methods can better mine the latent clustering structure of

Table 8
Clustering results on contaminated data.

Methods	Metrics				
	ACC (%)	NMI (%)	ARI (%)	Precision (%)	F-score (%)
CRMSC [36]	28.48	16.07	8.17	19.13	24.17
MCAN [26]	72.86	65.33	56.39	68.08	62.78
AMGL [29]	70.95	65.70	54.77	55.73	61.72
LMSC [30]	55.24	45.65	34.31	43.09	43.68
SWMC [39]	81.43	70.56	62.76	71.36	68.16
CSMSC [27]	64.29	53.11	48.17	48.87	56.48
MLRSSC [72]	54.76	42.43	30.40	43.32	40.16
NESE [40]	75.71	67.32	58.93	63.18	64.80
JFLMSC	83.33	69.73	65.41	69.64	70.25

data, while concatenating all views of multi-view datasets cannot effectively leverage information from multiple data sources.

2. Compared to multi-view spectral clustering methods, i.e., CRMSC, NESE, and AMGL, the proposed JFLMSC method also shows the best results on five databases. For example, on the MSRC-v1 dataset, it outperforms the state-of-the-art for NESE by a more than 10% improvement in terms of ACC, ARI, Precision, and F-score. This validates the effectiveness of the proposed method, which can learn a higher-quality similarity graph by exploring the global and local structures of data simultaneously.
3. The state-of-the-art multi-view subspace clustering methods MLRSSC, LMSC, and CSMSC show better performance on the ORL, BBCSport, and Outdoor Scene datasets compared to non-subspace clustering methods. For example, CSMSC shows the best result in terms of ACC on the Outdoor Scene dataset. However, LMSC shows lower performance on the MSRC-v1 dataset compared to the other multi-view clustering methods. CSMSC lags behind JFLMSC on the 100leaves dataset. This is because LMSC and CSMSC completely ignore the local structure of data. MLRSSC also shows competitive performance compared to CRMSC, AMGL, and SWMC. This is because MLRSSC integrates low-rank representation and sparse representation into a framework while ignoring the manifold structure of the data. JFLMSC outperforms MLRSSC on all datasets in all metrics due to adaptive distance regularization.
4. The proposed JFLMSC method consistently outperforms all competitors in terms of NMI, ARI, and F-score on all five datasets. This fact indicates that simultaneously unifying the global representation learning and local structure learning enables it to learn the optimal graph for clustering. Specifically, the JFLMSC method learns weights for different features, which can further improve clustering performance.

5.5. Robustness study

Following [73], the robustness of the proposed JFLMSC method was examined by adding commonly seen Gaussian noise. Specifically, views 3, 4, and 5 of the MSRC-v1 dataset were chosen to form a new multi-view dataset. Several samples of each view of the dataset were randomly corrupted with Gaussian noise with different variances, i.e., 0.01, 0.03, and 0.05. As shown in Table 8, the proposed method exhibits the best performance compared to other compared multi-view clustering methods in terms of ACC, ARI, and F-score.

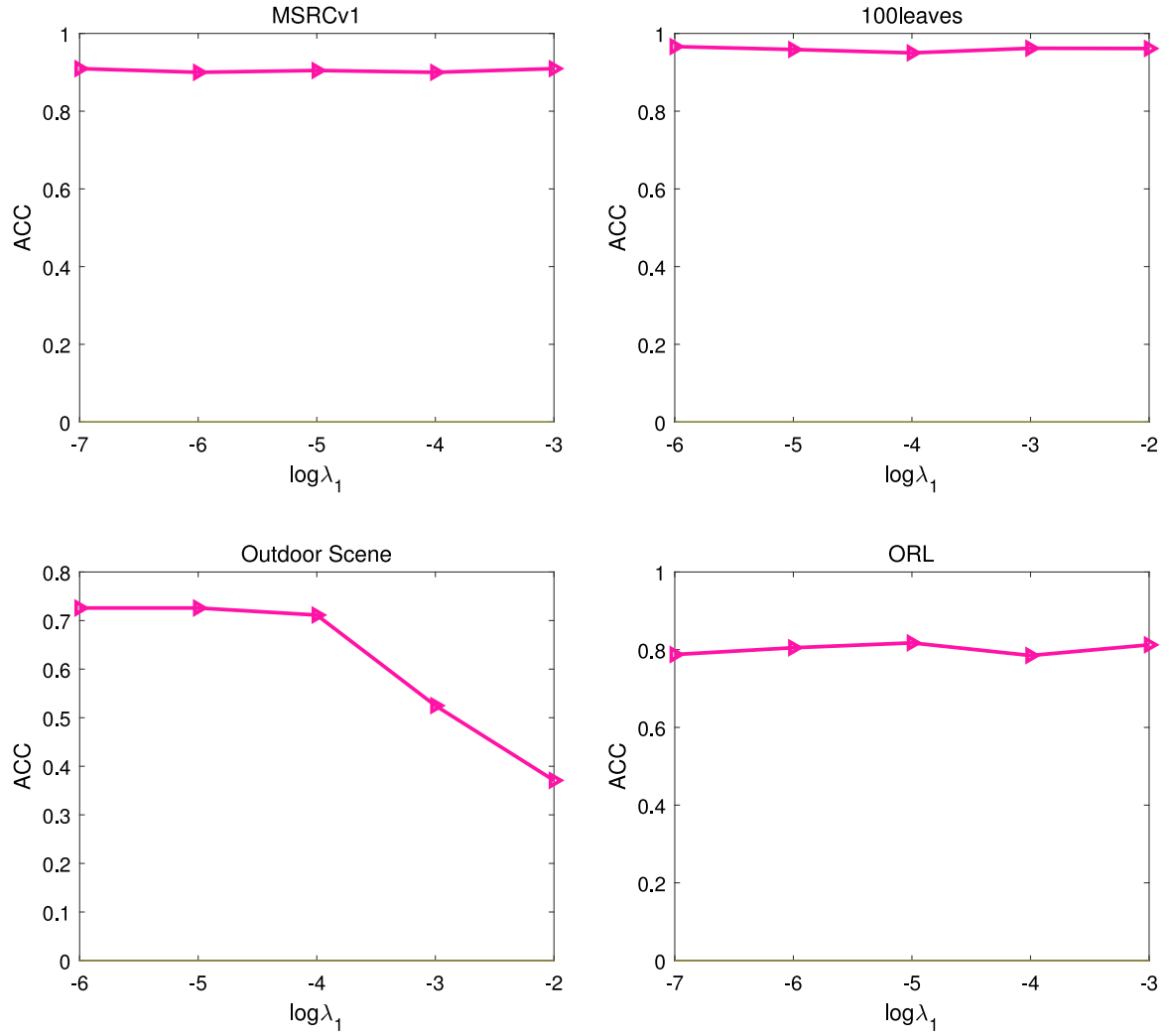


Fig. 2. ACC of JFLMSC method vs. different values of parameter λ_1 .

Table 9

Ablation study of proposed method.

Methods	Metrics				
	ACC (%)	NMI (%)	ARI (%)	Precision (%)	F-score (%)
MSRC-v1					
Eq. (6)	80.95	73.67	69.08	70.29	73.64
Eq. (7)	83.26	76.91	70.76	70.64	75.18
JFLMSC	91.43	82.30	81.06	83.54	83.69
100leaves					
Eq. (6)	85.88	92.87	80.37	76.47	80.57
Eq. (7)	94.00	97.33	92.04	90.56	92.11
JFLMSC	95.44	97.83	93.46	92.22	93.53
Outdoor Scene					
Eq. (6)	69.53	56.77	48.11	54.22	54.84
Eq. (7)	72.32	61.07	54.71	56.55	61.27
JFLMSC	72.58	61.08	54.95	57.13	61.47
ORL					
Eq. (6)	79.75	89.02	72.73	68.77	73.39
Eq. (7)	84.50	91.79	78.20	74.89	78.72
JFLMSC	84.75	93.31	81.39	77.38	81.83

5.6. Ablation study

An ablation study of the proposed method was conducted; that is, results only capturing local structure and results without

considering featurewise weight learning were analyzed separately. Table 9 reports the clustering results on the four datasets.

Eq. (6) was used to adaptively learn the local structure of each view, while Eq. (7) simultaneously learns the global and local structure of each view. Table 9 demonstrates that the improvement of Eq. (7) is significant compared to that of Eq. (6). The clustering ACC of Eq. (7) on the 100leaves dataset achieves an improvement of more than 9% in comparison with Eq. (6). Similarly, Eq. (7) outperforms Eq. (6) by nearly 7% on the Outdoor Scene dataset in terms of F-score.

Furthermore, one can see that the proposed JFLMSC method achieves the best clustering results. In particular, on the MSRC-v1 dataset JFLMSC achieves more than 10% of ARI and Precision compared to the second-best method. In conclusion, the proposed JFLMSC method effectively captures the global and local structure of multi-view data, and if the original data are noisy JFLMSC can adaptively assign small weights for noisy features and large weights for discriminative features. Thus JFLMSC outperforms Eqs. (6) and (7).

5.7. Computational speed, parameter, and convergence analysis

The clustering times of the proposed JFLMSC method on five multi-view datasets, as well as a comparison between it and CRMSC, MCAN, AMGL, LMSC, SWMC, CSMSC, and NESE, are reported in Table 10. All of the methods are implemented on a

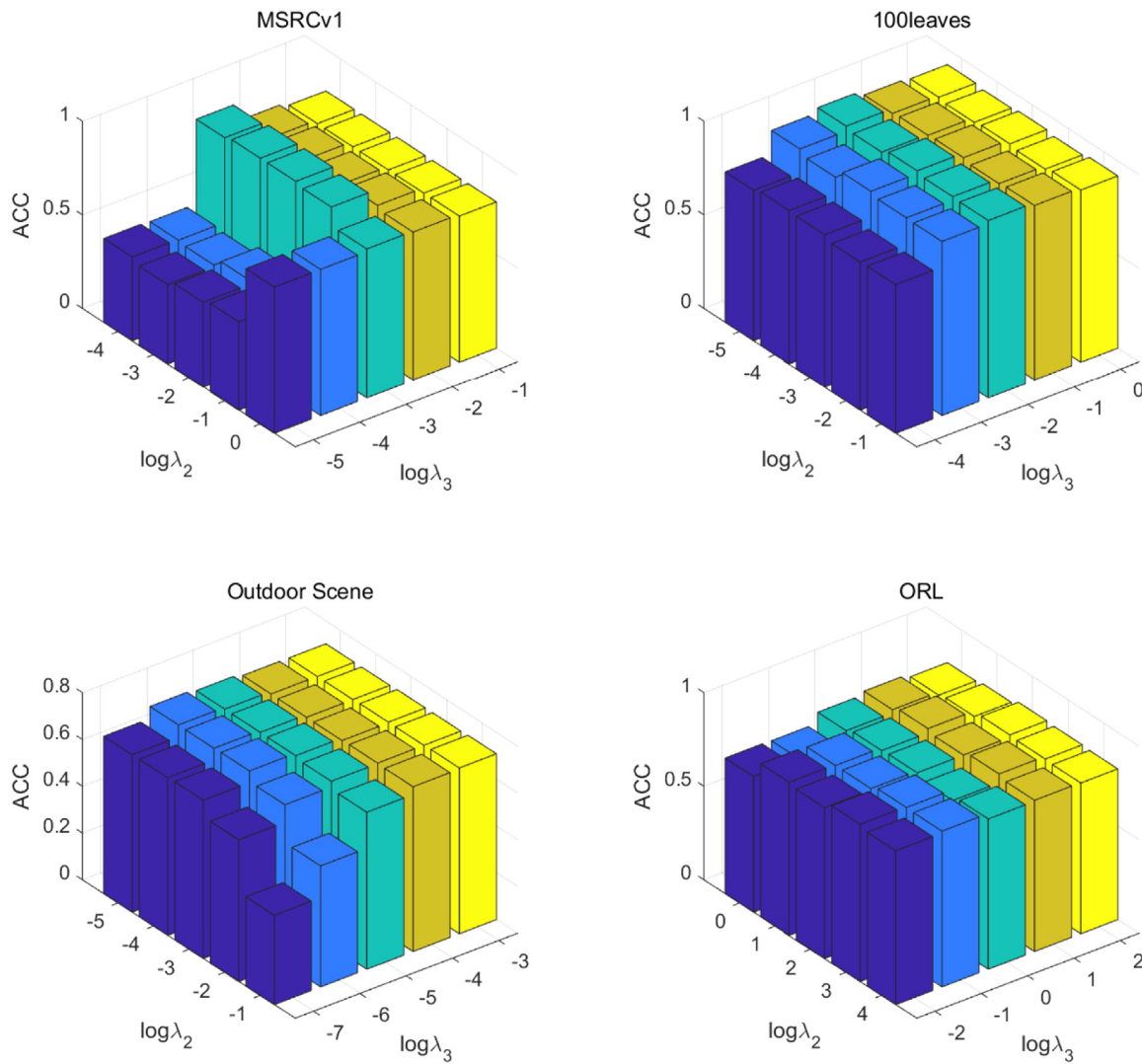


Fig. 3. ACC of JFLMSC method vs. different values of λ_2 and λ_3 .

Table 10

Runtime (in s) comparison on five multi-view datasets.

Datasets	Methods								
	CRMSC	MCAN	AMGL	LMSC	SWMC	CSMSC	NESE	MLRSSC	JFLMSC
100leaves	22.73	3.45	12.98	159.70	8.47	7.96	7.36	62.96	105.21
MSRC-v1	2.37	0.39	0.60	2.49	0.86	1.48	0.37	6.14	2.42
ORL	6.42	0.96	1.04	23.80	0.90	6.41	1.15	5.64	16.46
Outdoor Scene	47.43	21.74	64.59	655.71	47.45	157.75	13.81	599.94	326.72
BBCSport	3.74	1.16	1.02	18.56	1.84	5.11	36.91	4.28	14.18

computer with a 3.3-GHz Intel I9-9940X CPU and 64 GB of RAM running MatLab version R2018a (MathWorks, USA). The methods were run 10 times and the average running times recorded. As shown in Table 10, one can see that although the proposed JFLMSC is not the most efficient, its runtime is at the same level as the state-of-the-art LMSC and MLRSSC methods.

For the proposed JFLMSC method, three parameters in Eq. (8) balance the importance of the 2-norm constraint term, error term, and consistency term of all views. How the clustering ACC varies with the change of these three parameters was investigated. To deeply study the influence of parameters on clustering performance, parameter values from $\{1e-7, 1e-6, 1e-5, \dots, 1e-1, 1\}$ were selected.

First, the parameters λ_2 and λ_3 were fixed, and then the proposed JFLMSC method performed with different values of λ_1

to observe the influence of λ_1 on the clustering ACC. From Fig. 2, it can be seen that ACC is insensitive to λ_1 when $\lambda_1 \leq 1e-4$. Notably, when $\lambda_1 > 1e-4$, ACC drops sharply on the Outdoor Scene dataset. This is because the consistency term may destroy the diversity learning of each view if λ_1 is too large.

Fig. 3 shows ACC versus different values of λ_2 and λ_3 with fixed λ_1 . It is observed that ACC of the proposed JFLMSC method is sensitive to λ_2 and λ_3 . Different datasets need different parameter combinations to achieve the best performance. In reality, one can first fix parameter λ_1 since it is not sensitive to the clustering performance, and then use cross-validation to find the optimal λ_2 and λ_3 .

Fig. 4 exhibits the trend of the objective value computed by the proposed Algorithm 1 regarding the number of iterations

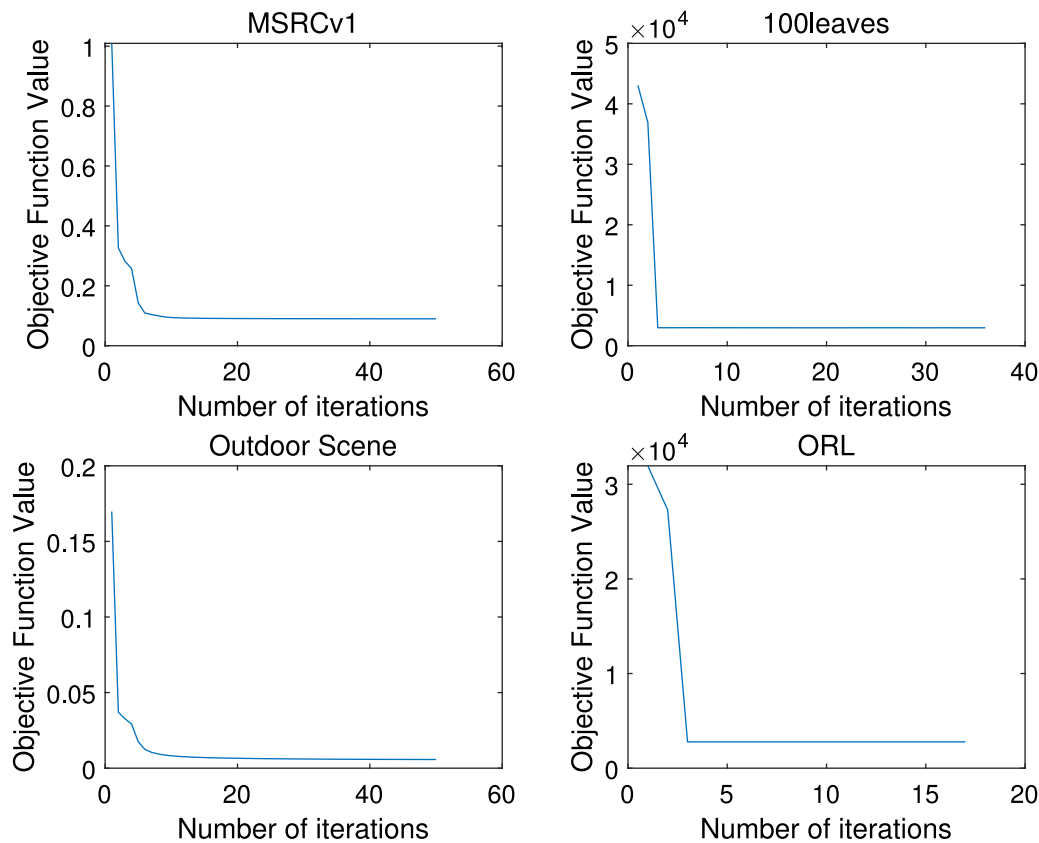


Fig. 4. Convergence analysis of proposed method.

on four datasets separately. The result empirically confirms the convergence behavior of Algorithm 1.

6. Conclusions

In this paper, a novel multi-view subspace clustering method named Joint Featurewise Weighting and Lobal Structure Learning for Multi-view Subspace Clustering (JFLMSC) is presented. Specifically, adaptive local learning, self-representation learning, and weight learning for features are seamlessly integrated into a unified framework to exploit the global and local structures of each view for multi-view subspace clustering. In the proposed method, discriminative features are assigned high weights, while noisy features have low weights. As a result, one can obtain a high-quality similarity graph for each view, which also extracts the consistency information across different views using the consensus clustering assignment matrix. Extensive experimental analysis was carried out on five benchmark multi-view datasets in terms of performance comparison, robustness study, ablation study, sensitivity analysis, and convergence analysis, the results of which demonstrate the superiority of the proposed method. The complexity of the proposed method is largely determined by the size of samples, which may restrict its usage to small- to middle-sized datasets. In planned future research, the landmark-based strategy [74] may be utilized to improve the scalability of JFLMSC such that it can apply to large-scale multi-view subspace clustering.

CRedit authorship contribution statement

Shi-Xun Lin: Conceptualization, Methodology, Validation.
Guo Zhong: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing - original draft. **Ting Shu:** Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Y. Wang, Y.Y. Tang, L. Li, Minimum error entropy based sparse representation for robust subspace clustering, *IEEE Trans. Signal Process.* 63 (15) (2015) 4010–4021, <http://dx.doi.org/10.1109/TSP.2015.2425803>.
- [2] G. Zhong, C.-M. Pun, Nonnegative self-representation with a fixed rank constraint for subspace clustering, *Inform. Sci.* 518 (2020) 127–141.
- [3] Y. Wang, Y.Y. Tang, C. Zou, L. Li, H. Chen, Modal regression based greedy algorithm for robust sparse signal recovery, clustering and classification, *Neurocomputing* 372 (2020) 73–83.
- [4] C. Yang, Z. Ren, Q. Sun, M. Wu, M. Yin, Y. Sun, Joint correntropy metric weighting and block diagonal regularizer for robust multiple kernel subspace clustering, *Inform. Sci.* 500 (2019) 48–66.
- [5] M. Liao, X. Gu, Subspace clustering based on alignment and graph embedding, *Knowl.-Based Syst.* 188 (2020) 105029.
- [6] J. Nguyen, A. Armisen, G. Sánchez-Hernández, M. Casabayó, N. Agell, An owa-based hierarchical clustering approach to understanding users' lifestyles, *Knowl.-Based Syst.* 190 (2020) 105308.
- [7] S. McCloskey, B. Jeffries, I. Koprinska, C.B. Miller, R.R. Grunstein, Data-driven cluster analysis of insomnia disorder with physiology-based qeeg variables, *Knowl.-Based Syst.* 183 (2019) 104863.
- [8] S. Hu, X. Yan, Y. Ye, Joint specific and correlated information exploration for multi-view action clustering, *Inform. Sci.* (2020).
- [9] Y. Song, Y. Wang, X. Ye, D. Wang, Y. Yin, Y. Wang, Multi-view ensemble learning based on distance-to-model and adaptive clustering for imbalanced credit risk assessment in p2p lending, *Inform. Sci.* (2020).
- [10] A.K. Jain, Data clustering: 50 years beyond k-means, *Pattern Recognit. Lett.* 31 (8) (2010) 651–666, award winning papers from the 19th International Conference on Pattern Recognition (ICPR).
- [11] U. Von Luxburg, A tutorial on spectral clustering, *Statist. Comput.* 17 (4) (2007) 395–416.

- [12] L. Hagen, A.B. Kahng, New spectral methods for ratio cut partitioning and clustering, *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* 11 (9) (1992) 1074–1085, <http://dx.doi.org/10.1109/43.159993>.
- [13] Yu, Shi, Multiclass spectral clustering, in: *Proceedings Ninth IEEE International Conference on Computer Vision*, Vol. 1, 2003, pp. 313–319, <http://dx.doi.org/10.1109/ICCV.2003.1238361>.
- [14] Z. Wu, R. Leahy, An optimal graph theoretic approach to data clustering: theory and its application to image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 15 (11) (1993) 1101–1113, <http://dx.doi.org/10.1109/34.244673>.
- [15] Jianbo Shi, J. Malik, Normalized cuts and image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (8) (2000) 888–905, <http://dx.doi.org/10.1109/34.868688>.
- [16] F. Nie, Z. Zeng, I.W. Tsang, D. Xu, C. Zhang, Spectral embedded clustering: A framework for in-sample and out-of-sample spectral clustering, *IEEE Trans. Neural Netw.* 22 (11) (2011) 1796–1808, <http://dx.doi.org/10.1109/TNN.2011.2162000>.
- [17] E. Elhamifar, R. Vidal, Sparse subspace clustering: Algorithm, theory, and applications, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (11) (2013) 2765–2781, <http://dx.doi.org/10.1109/TPAMI.2013.57>.
- [18] C. Li, C. You, R. Vidal, Structured sparse subspace clustering: A joint affinity learning and subspace clustering framework, *IEEE Trans. Image Process.* 26 (6) (2017) 2988–3001, <http://dx.doi.org/10.1109/TIP.2017.2691557>.
- [19] C. Lu, J. Feng, Z. Lin, T. Mei, S. Yan, Subspace clustering by block diagonal representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 41 (2) (2019) 487–501, <http://dx.doi.org/10.1109/TPAMI.2018.2794348>.
- [20] Z. Zhang, Y. Xu, L. Shao, J. Yang, Discriminative block-diagonal representation learning for image recognition, *IEEE Trans. Neural Netw. Learn. Syst.* 29 (7) (2018) 3111–3125.
- [21] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (1) (2013) 171–184, <http://dx.doi.org/10.1109/TPAMI.2012.88>.
- [22] F. Dornaika, L. Weng, Sparse graphs with smoothness constraints: Application to dimensionality reduction and semi-supervised classification, *Pattern Recognit.* 95 (2019) 285–295.
- [23] X. Peng, J. Feng, S. Xiao, W. Yau, J.T. Zhou, S. Yang, Structured autoencoders for subspace clustering, *IEEE Trans. Image Process.* 27 (10) (2018) 5076–5086.
- [24] J. Zhao, X. Xie, X. Xu, S. Sun, Multi-view learning overview: Recent progress and new challenges, *Inf. Fusion* 38 (2017) 43–54.
- [25] M. Li, Z. Li, C. Tang, X. Liu, L. Wang, Robust multi-view clustering with a unified weight learning paradigm, *IEEE Access* 7 (2019) 177965–177973, <http://dx.doi.org/10.1109/ACCESS.2019.2958493>.
- [26] F. Nie, G. Cai, X. Li, Multi-view clustering and semi-supervised classification with adaptive neighbours, in: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, in: AAAI'17, AAAI Press, 2017, pp. 2408–2414.
- [27] S. Luo, C. Zhang, W. Zhang, X. Cao, Consistent and specific multi-view subspace clustering, in: *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [28] T. Shu, B. Zhang, Y.Y. Tang, Multi-view classification via a fast and effective multi-view nearest-subspace classifier, *IEEE Access* 7 (2019) 49669–49679.
- [29] F. Nie, J. Li, X. Li, Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification, in: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, in: IJCAI'16, AAAI Press, 2016, pp. 1881–1887.
- [30] C. Zhang, H. Fu, Q. Hu, X. Cao, Y. Xie, D. Tao, D. Xu, Generalized latent multi-view subspace clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (1) (2020) 86–99, <http://dx.doi.org/10.1109/TPAMI.2018.2877660>.
- [31] X. Peng, H. Zhu, J. Feng, C. Shen, H. Zhang, J.T. Zhou, Deep clustering with sample-assignment invariance prior, *IEEE Trans. Neural Netw. Learn. Syst.* (2019).
- [32] X. Peng, Z. Huang, J. Lv, H. Zhu, J.T. Zhou, Comic: Multi-view clustering without parameter selection, in: *International Conference on Machine Learning*, 2019, pp. 5092–5101.
- [33] S. Bickel, T. Scheffer, Multi-view clustering, in: *Fourth IEEE International Conference on Data Mining (ICDM'04)*, 2004, pp. 19–26, <http://dx.doi.org/10.1109/ICDM.2004.10095>.
- [34] K. Chaudhuri, S.M. Kakade, K. Livescu, K. Sridharan, Multi-view clustering via canonical correlation analysis, in: *Proceedings of the 26th Annual International Conference on Machine Learning*, in: ICML '09, Association for Computing Machinery, New York, NY, USA, 2009, pp. 129–136.
- [35] J. Liu, C. Wang, J. Gao, J. Han, Multi-view clustering via joint nonnegative matrix factorization, in: *Proceedings of the 2013 SIAM International Conference on Data Mining*, SIAM, 2013, pp. 252–260.
- [36] A. Kumar, P. Rai, H. Daumé, Co-regularized multi-view spectral clustering, in: *Proceedings of the 24th International Conference on Neural Information Processing Systems*, in: NIPS'11, Curran Associates Inc., Red Hook, NY, USA, 2011, pp. 1413–1421.
- [37] R. Xia, Y. Pan, L. Du, J. Yin, Robust multi-view spectral clustering via low-rank and sparse decomposition, in: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, in: AAAI'14, AAAI Press, 2014, pp. 2149–2155.
- [38] Z. Zhang, L. Liu, F. Shen, H.T. Shen, L. Shao, Binary multi-view clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* 41 (7) (2019) 1774–1782.
- [39] F. Nie, L. Jing, X. Li, Self-weighted multiview clustering with multiple graphs, in: *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, in: IJCAI'17, AAAI Press, 2017, pp. 2564–2570.
- [40] Z. Hu, F. Nie, R. Wang, X. Li, Multi-view spectral clustering via integrating nonnegative embedding and spectral embedding, *Inf. Fusion* 55 (2020) 251–259.
- [41] G. Zhong, C. Pun, A unified framework for multi-view spectral clustering, in: *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, 2020, pp. 1854–1857.
- [42] Y. Liang, D. Huang, C. Wang, Consistency meets inconsistency: A unified graph learning framework for multi-view clustering, in: *2019 IEEE International Conference on Data Mining (ICDM)*, 2019, pp. 1204–1209.
- [43] R. Vidal, Subspace clustering, *IEEE Signal Process. Mag.* 28 (2) (2011) 52–68, <http://dx.doi.org/10.1109/MSP.2010.939739>.
- [44] H. Gao, F. Nie, X. Li, H. Huang, Multi-view subspace clustering, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4238–4246.
- [45] Q. Zheng, J. Zhu, Z. Li, S. Pang, J. Wang, Y. Li, Feature concatenation multi-view subspace clustering, *Neurocomputing* 379 (2020) 89–102.
- [46] W. Weng, W. Zhou, J. Chen, H. Peng, H. Cai, Enhancing multi-view clustering through common subspace integration by considering both global similarities and local structures, *Neurocomputing* 378 (2020) 375–386.
- [47] M. Brbić, I. Kopriva, Multi-view low-rank sparse subspace clustering, *Pattern Recognit.* 73 (2018) 247–258.
- [48] G. Zhong, C.-M. Pun, Subspace clustering by simultaneously feature selection and similarity learning, *Knowl.-Based Syst.* 193 (2020) 105512.
- [49] X. Peng, C. Lu, Z. Yi, H. Tang, Connections between nuclear-norm and frobenius-norm-based representations, *IEEE Trans. Neural Netw. Learn. Syst.* 29 (1) (2018) 218–224.
- [50] C.-Y. Lu, H. Min, Z.-Q. Zhao, L. Zhu, D.-S. Huang, S. Yan, Robust and efficient subspace segmentation via least squares regression, in: *European Conference on Computer Vision*, Springer, 2012, pp. 347–360.
- [51] X. Cai, F. Nie, W. Cai, H. Huang, Heterogeneous image features integration via multi-modal semi-supervised learning model, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1737–1744.
- [52] A. Kumar, H.D. III, A co-training approach for multi-view spectral clustering, in: *Proceedings of the 28th International Conference on International Conference on Machine Learning*, in: ICML'11, Omnipress, Madison, WI, USA, 2011, pp. 393–400.
- [53] X. Cao, C. Zhang, H. Fu, Si Liu, Hua Zhang, Diversity-induced multi-view subspace clustering, in: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 586–594.
- [54] X. Wang, X. Guo, Z. Lei, C. Zhang, S.Z. Li, Exclusivity-consistency regularized multi-view subspace clustering, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 923–931.
- [55] X. Lu, Y. Wang, Y. Yuan, Graph-regularized low-rank representation for destriping of hyperspectral images, *IEEE Trans. Geosci. Remote Sens.* 51 (7) (2013) 4009–4018.
- [56] Z. Liu, J. Wang, G. Liu, J. Pu, Sparse low-rank preserving projection for dimensionality reduction, *IEEE Access* 7 (2019) 22941–22951.
- [57] M. Yin, J. Gao, Z. Lin, Laplacian regularized low-rank representation and its applications, *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (3) (2016) 504–517.
- [58] Z. Lin, H. Zhang, *Low-Rank Models in Visual Analysis: Theories, Algorithms, and Applications*, Academic Press, 2017.
- [59] K. Tang, D.B. Dunson, Z. Su, R. Liu, J. Zhang, J. Dong, Subspace segmentation by dense block and sparse representation, *Neural Netw.* 75 (2016) 66–76.
- [60] S. Boyd, S.P. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge university press, 2004.
- [61] E.J. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis?, *J. ACM* 58 (3) (2011).
- [62] G.-Y. Zhang, Y.-R. Zhou, X.-Y. He, C.-D. Wang, D. Huang, One-step kernel multi-view subspace clustering, *Knowl.-Based Syst.* 189 (2020) 105126.
- [63] K.S. Riedel, A sherman-morrison-woodbury identity for rank augmenting matrices with application to centering, *SIAM J. Matrix Anal. Appl.* 13 (2) (1992) 659–662, <http://dx.doi.org/10.1137/0613040>.
- [64] Z. Shen, C. chuan Toh, S. Yun, An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems, *Pac. J. Optim.* (2010) 615–640.
- [65] D. Huang, C. Wang, J. Wu, J. Lai, C. Kwok, Ultra-scalable spectral clustering and ensemble clustering, *IEEE Trans. Knowl. Data Eng.* 32 (6) (2020) 1212–1226.
- [66] X. Peng, H. Tang, L. Zhang, Z. Yi, S. Xiao, A unified framework for representation-based subspace clustering of out-of-sample and large-scale data, *IEEE Trans. Neural Netw. Learn. Syst.* 27 (12) (2016) 2499–2512.

- [67] T. Zhou, C. Zhang, X. Peng, H. Bhaskar, J. Yang, Dual shared-specific multiview subspace clustering, *IEEE Trans. Cybern.* (2019) 1–14.
- [68] J. Winn, A. Criminisi, T. Minka, Object categorization by learned universal visual dictionary, in: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, Vol. 2, IEEE, 2005, pp. 1800–1807.
- [69] X. Cai, F. Nie, H. Huang, Multi-view k-means clustering on big data, in: Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, in: IJCAI '13, AAAI Press, 2013, pp. 2598–2604.
- [70] F.S. Samaria, A.C. Harter, Parameterisation of a stochastic model for human face identification, in: Proceedings of 1994 IEEE Workshop on Applications of Computer Vision, IEEE, 1994, pp. 138–142.
- [71] H. Wang, Y. Yang, B. Liu, Gmc: Graph-based multi-view clustering, *IEEE Trans. Knowl. Data Eng.* 32 (6) (2020) 1116–1129.
- [72] M. Brbic, I. Kopriva, Multi-view low-rank sparse subspace clustering, *Pattern Recognit.* 73 (2018) 247–258.
- [73] Z. Kang, W. Zhou, Z. Zhao, J. Shao, M. Han, Z. Xu, Large-scale multi-view subspace clustering in linear time, in: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, in: AAAI'20, AAAI Press, 2020, pp. 2149–2155.
- [74] D. Cai, X. Chen, Large scale spectral clustering via landmark-based sparse representation, *IEEE Trans. Cybern.* 45 (8) (2015) 1669–1680.