Feature Concatenation Multi-view Subspace Clustering

Qinghai Zheng[a], Jihua Zhu[a,*], Zhongyu Li[a], Shanmin Pang[a], Jun Wang[b] and Yaochen Li[a]

[a]*School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China*
[b]*School of Communication and Information Engineering, Shanghai University, Shanghai 200444, People's Republic of China*

ABSTRACT

Multi-view clustering is a learning paradigm based on multi-view data. Since statistic properties of different views are diverse, even incompatible, few approaches implement multi-view clustering based on the concatenated features straightforward. However, feature concatenation is a natural way to combine multi-view data. To this end, this paper proposes a novel multi-view subspace clustering approach dubbed Feature Concatenation Multi-view Subspace Clustering (FCMSC), which boosts the clustering performance by exploring the consensus information of multi-view data. Specifically, multi-view data are concatenated into a joint representation firstly, then, $l_{2,1}$-norm is integrated into the objective function to deal with the sample-specific and cluster-specific corruptions of multiple views. Moreover, a graph regularized FCMSC is also proposed in this paper to explore both the consensus information and complementary information of multi-view data for clustering. It is noteworthy that the obtained coefficient matrix is not derived by simply applying the Low-Rank Representation (LRR) to concatenated features directly. Finally, an effective algorithm based on the Augmented Lagrangian Multiplier (ALM) is designed to optimize the objective functions. Comprehensive experiments on six real-world datasets illustrate the superiority of the proposed methods over several state-of-the-art approaches for multi-view clustering.

1. Introduction

Multi-view data, which are collected from different measurements or fields to give a comprehensive description of objects, are popular in many real-world applications [46, 51, 19, 22, 52, 10]. For example, in computer vision fields, an image can be presented by multiple views (GIST [34], SIFT [31], LBP [33], etc.); the words presented on a webpage and the words presented in URL are two distinct views of the webpage; video signals and audio signals are two common representations and can be applied to multimedia content understanding. Compared with single-view data, multi-view data contain both the consensus and complementary information among multiple views. And the goal of multi-view learning, which has achieved success in many applications [46, 65, 38, 69, 55, 6], is to improve the generalization performance by leveraging multiple views.

As a fundamental task in unsupervised learning, clustering, which is often used to mine underlying information of data, can be a stand-alone exploratory tool or a preprocessing step to assist other learning tasks in machine learning as well [71]. Many clustering approaches have been proposed, and subspace clustering, which assumes that high dimensional data lie in a union of low-dimensional subspaces and tries to group data points into clusters and find the corresponding subspace simultaneously, attracts lots of researches owing to its promising performance and good interpretability. In recent years, many clustering algorithms based on the subspace clustering with different constraints have been proposed [20, 41, 14, 35, 36, 11, 67, 12, 37, 21]. Low-Rank Subspace Clustering (LRSC) [41] finds a low-rank linear representation of data in a dictionary of themselves and then employs the spectral clustering on an adjacent matrix,

which is derived from the low-rank representation [29], to obtain clustering results. Besides, Sparse Subspace Clustering (SSC) [14], which tries to find a sparse representation based on the l_1-norm, is a powerful subspace clustering algorithms as well. Additionally Low-Rank Sparse Subspace Clustering (LRSSC) [35] applies low-rank and sparse constraints simultaneously based on the trace norm and l_1-norm according to the fact that the coefficient matrix is often sparse and low-rank at the same time. By combining the labels and the affinity, Discriminative and Coherent Subspace Clustering (DCSC) [11] tries to enhance the labels' discrimination for data in variant clusters and the affinity for data in the same cluster. Although these algorithms mentioned above can get promising clustering results in practice, they are designed for single-view data rather than multi-view data.

Based on the subspace clustering, many multi-view subspace clustering approaches have been proposed [8, 57, 5, 47, 15, 66]. Most of them process multiple views separately and obtain clustering results by finding a common shared coefficient matrix or fusing clustering results of different views directly. Although good performance has been achieved in practice, the underlying information of multi-view data is insufficiently explored in these methods. To this end, in this paper, we propose a novel multi-view subspace clustering named Feature Concatenation Multi-view Subspace Clustering (FCMSC), which performs clustering on all views simultaneously and takes advantage of the consensus information of multi-view data to improve clustering results.

For multi-view clustering, a naive idea is concatenating features of all views and then running a clustering algorithm to get clustering results. Concatenated features have the following merits: 1) original information of multi-view data can be maximum preserved by concatenating features of all views into a joint view representation; 2) all views of multi-

view data can be processed simultaneously during clustering. However, it is ineffective in practice and even gets worse clustering results [46, 65, 24, 25, 44] by simply performing a single-view clustering algorithm on the concatenated features straightforward to obtain clustering results, since each view contains its own statistical properties. It is noteworthy that our proposed FCMSC can achieve the promising clustering performance on the joint view representation. To be specific, by introducing the concept of cluster-specific corruptions, our FCMSC decomposes the original coefficient matrix, which is derived from concatenated features by employing low-rank representation straightforward, to obtain a new low-rank coefficient matrix, which enjoys the consensus property of multi-view data. Moreover, a graph regularized FCMSC (gr-FCMSC) is also proposed, which can explore both the consensus information and complementary information simultaneously during clustering. Finally, an effective optimization algorithm based on the Augmented Lagrangian Multiplier (ALM) [50, 28] is designed for the objective functions of the proposed FCMSC and gr-FCMSC. Extensive experiments on six benchmark datasets compared with several state-of-the-arts illustrate the effectiveness and competitiveness of the proposed methods.

The main contributions of this paper can be summarized as follows:

1) An effective feature concatenation multi-view subspace clustering is proposed in this paper. By introducing the cluster-specific corruptions brought by different views, the proposed method can perform clustering on multiple views simultaneously and explore the consensus information of multi-view data based on the joint view representation directly.

2) A graph regularized feature concatenation multi-view subspace clustering (gr-FCMSC) is also proposed. By employing the graph Laplacians, both the consensus information and the complementary information of multi-view data can be fully explored during clustering.

3) Comprehensive experiments are conducted on public available datasets, and experimental results show the effectiveness and superiority of the proposed methods compared with several state-of-the-arts.

The rest of this paper is organized as follows. The next section reviews related works briefly. Section 3 introduces our methods, including FCMSC and gr-FCMSC, in detail. And Section 4 presents the related optimizations. Comprehensive experimental results and discussions are provided in Section 5. Finally, Section 6 provides the conclusions.

2. Related Work

A lot of approaches have been proposed recently to solve the multi-view clustering problem [70, 58, 57, 5, 15, 66, 24, 25, 44, 56, 39, 8, 49, 40, 42, 64, 30, 3]. Most existing multi-view clustering methods can be grouped into two main categories roughly: generative methods and discriminative methods [8]. The idea of generative methods is to construct generative models for variant clusters respectively. For example, multi-view convex mixture models [40] assign different weights for multiple views automatically and consider the diversity of different views. Although most generative algorithms are robust to the missing entries and have global optimization, they are accompanied with a series of hypotheses and parameters, which make the optimization more difficult and time consuming.

Discriminative methods, the goal of which is to minimize both the intrinsic similarity of data points between different clusters and dissimilarity of data points within the same cluster through all multiple views simultaneously, have achieved good clustering results in many applications and attract the most attention of researchers in this research field [8]. Taking examples of multi-view subspace clustering, Latent Multi-view Subspace Clustering (LMSC) [58] and generalized Latent Multi-view Subspace Clustering (gLMSC) [56] introduce a latent representation to explore the relationships of data points among all views, obtain the underlying complementary information and seek the latent representation as well; And Multi-view Low-rank Sparse Subspace Clustering (MLRSSC) [3] obtains multi-view clustering results by constructing an affinity matrix with the low-rank and sparsity constraints; Multi-view subspace clustering by learning a joint affinity graph [39] pursuits a low-rank subspace representation with diversity regularization and a rank constraint for multi-view clustering. Besides, many spectral clustering based methods are also proposed in recent years. The co-training approach for multi-view spectral clustering [24] and the co-regularized multi-view spectral clustering [25] try to get clustering results that can maximize the similarity graph agreement among different views; Robust Multi-view Spectral Clustering (RMSC) [44] recovers a common transition probability matrix via low-rank and sparse decomposition and employs the Markov chain approach to obtain clustering results. In addition, some multi-view clustering methods based on the matrix factorization method [27] are proposed by exploring the consensus information among views [64, 30]. For most of discriminative multi-view clustering methods, the essential difference is the style they use to explore the underlying information of multiple views.

Inspire by the success of deep learning [26, 53, 59, 54], some clustering methods based on deep learning are proposed [23, 68]. By introducing a self-expressive layer, Deep Subspace Clustering Networks can map data to a latent space non-linearly and learns the affinity matrix straightforward. Deep Canonical Correlation Analysis (DCCA) [1] and Deep Canonically Correlated Autoencoders (DCCAE) [43] are two deep learning based methods which can be employed for multi-view clustering. Although good clustering results can be achieved, most existing multi-view clustering methods deal with different views separately, and that is an ineffective way since the relationships among multiple views are ignored. A natural way is to combine all views before clustering, and some related approaches have been proposed
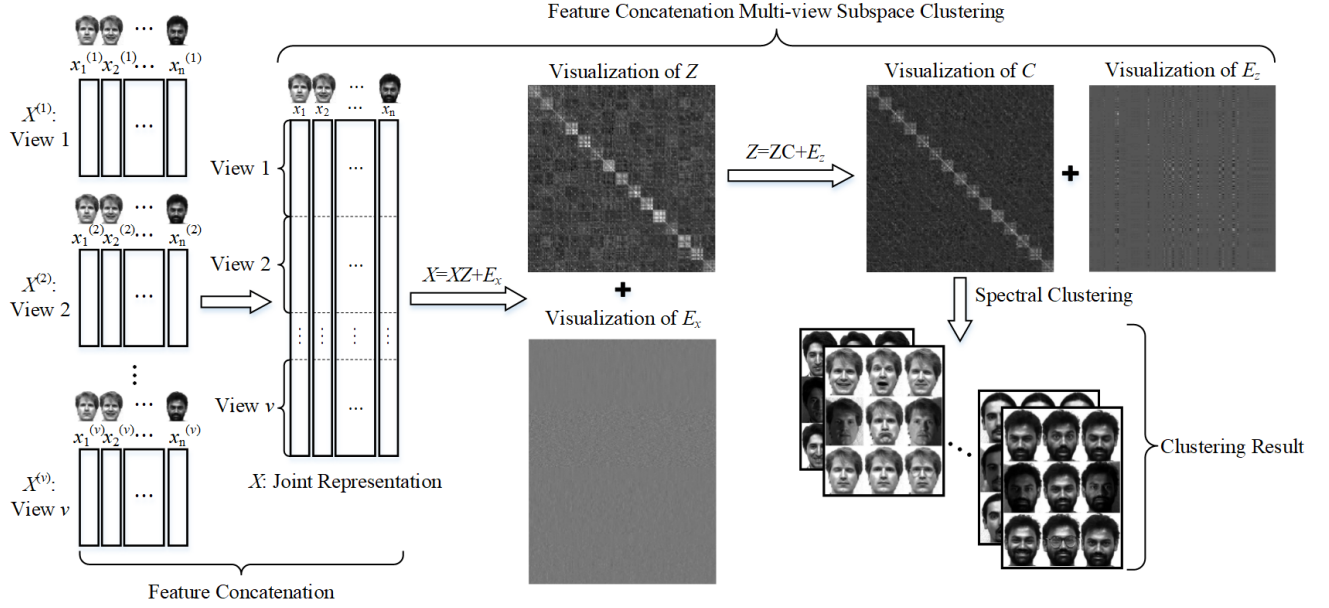
Figure 1: Illustration of the proposed method, in which multi-view subspace clustering is implemented on the joint view representation. Multiple views are concatenated firstly, then the FCMSC is employed to obtain a desired coefficient matrix, i.e. C, and the last step is to infer the clustering results of data points by leveraging the spectral clustering approach with the adjacency matrix $(\text{abs}(C) + \text{abs}(C^T))/2$.

[9, 7, 58, 60, 18]. However, these methods may corrupt either the consensus information or the complementary information among views during combination to varying degrees. Taking the joint view representation into consideration, multi-view clustering results achieved by employing a single-view clustering algorithm to the joint view representation directly are uncompetitive [57, 5, 25, 44, 8], and few works focus on this kind of combination styles. However, it is obvious that original information contained among multiple views can get maximum preservation by concatenating features of all views straightforward. It is notable that the proposed FCMSC and gr-FCMSC can get promising and competitive clustering results by utilizing the concatenated features of multiple views straightforward.

3. Feature Concatenation Multi-view Subspace Clustering

In this section, we propose the Feature Concatenation Multi-view Subspace Clustering (FCMSC) method by exploring the consensus information of multi-view data. Moreover, a graph regularized FCMSC method, termed as gr-FCMSC, is also proposed, and it can explore both the consensus information and complementary information of multiple views for multi-view clustering.

3.1. FCMSC

For convenience, Table 1 lists main symbols leveraged throughout this paper. Given a multi-view dataset with v views and n samples, i.e. $\{x_1^{(i)}, x_2^{(i)}, \cdots, x_n^{(i)}\}_{i=1}^{v}$, data points

of which are drawn from m multiple subspaces. In order to obtain a matrix that each column has the same magnitude, data of each view are normalized within the range of $[0, 1]$, and then multiple views are concatenated into a joint view representation matrix X, which is defined as follows:

$$
X = \begin{bmatrix}
x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\
x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\
\vdots & \vdots & \ddots & \vdots \\
x_1^{(v)} & x_2^{(v)} & \cdots & x_n^{(v)}
\end{bmatrix},
\tag{1}
$$

where $x_i^{(k)}$ denotes the features of the i-th sample from the k-th view, and the i-th column of X contains features of all views of the i-th sample. Based on the concatenated features, Fig. 1 displays the framework of the proposed FCMSC.

Since statistic properties of different views are diverse, even incompatible among views, it is difficult to explore the mutual information of multiple views effectively and fully. In order to get a preliminary exploration of multi-view data, we consider the following objective function in the beginning:

$$
\begin{aligned}
&\min_{Z, E_x} \left\| E_x \right\|_{2,1} + \lambda \|Z\|_* \\
&\text{s.t.} \quad X = XZ + E_x,
\end{aligned}
\tag{2}
$$

where Z indicates an original coefficient matrix of X, E_x denotes the sample-specific corruptions of data points, and λ is a trade-off parameter. The $l_{2,1}$-norm of E_x enforces E_x to be sparse in columns and columns of E_x to be zero. Equation (2) is a standard low-rank representation [29] of the concatenated features. However, experimental results presented in

Table 1
Main Symbols

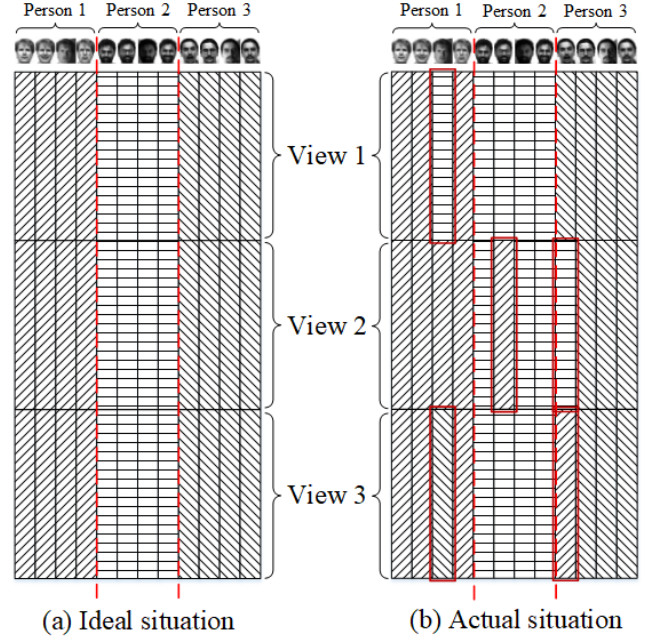| Symbol	Meaning
n	The number of samples.
v	The number of views.
m	The number of clusters.
d_i	The dimension of features in i-th view.
d	The dimension of the concatenated features.
$x_k^{(i)} \in R^{d_i}$	The features of k-th sample from i-th view.
$X \in R^{d \times n}$	The joint view representation matrix.
$Z \in R^{n \times n}$	The original coefficient matrix.
$E_x \in R^{d \times n}$	The sample-specific corruptions.
$C \in R^{n \times n}$	The desired coefficient matrix.
$E_{cs} \in R^{d \times n}$	The cluster-specific corruptions.
$E_z \in R^{n \times n}$	The term derived from E_{cs}.
$L_i \in R^{n \times n}$	The Laplacian matrix of i-th view.
$\|A\|_*$	The trace-norm of matrix A.
$\|A\|_{2,1}$	The $l_{2,1}$-norm of matrix A.

Figure 2: Illustrating the cluster-specific corruptions brought by multiple views. Taking 12 images of 3 individuals for examples, and they are described by 3 views. With the joint view representation, (a) is an ideal situation that clustering results of all views are coincident. (b) is the actual situation since clustering results of different views are different to some degree. The columns of joint view representation matrix, containing red rectangles, indicate the cluster-specific corruptions, which are brought by multiple views obviously.

[57, 5, 8] and later section of this paper show that the clustering performance is uncompetitive if a spectral clustering algorithm is performed based on the coefficient matrix Z. This is because each view has specific statistical properties, which may be contradictory among views, and it is unreasonable to explore the joint views representation by directly employing single-view clustering algorithm.

In this paper, we introduce the cluster-specific corruptions, which are accompanied with multi-view data, as shown in Fig. 2. Without considering the cluster-specific corruptions, it is expected that running a single-view clustering algorithm on the concatenated features is hard to get satisfied clustering results. And the original coefficient matrix Z, obtained in (2), is far from good enough for multi-view clustering. In order to handle the concatenated features better and get a desired coefficient matrix, it is suggested to consider the following formulation:

$$X = DC + E_{cs} + E_x, \tag{3}$$

where D indicates a dictionary matrix, C denotes the desired coefficient matrix, and E_{cs} represents the cluster-specific corruptions among multiple views. Equation (3) considers both the cluster-specific and sample-specific corruptions.

Obviously, the choices of D and E_{cs} are vital for the final multi-view clustering performance. Since matrix D is free of the sample-specific corruptions, it is reasonable to employ the reconstructed features, obtained from (2), as the dictionary matrix, i.e. $D = XZ$. For E_{cs}, most existing norms are not suitable for it. As shown in Fig. 2, under the assumption that the true underlying clustering would assign corresponding points in each view to the same cluster, the number of columns with cluster-specific corruptions in matrix X should be small, and the major part of columns achieves

the same clustering results. However, it is difficult to process the cluster-specific corruptions directly. In this paper, we decomposed E_{cs} as $E_{cs} = XE_z$. By assuming that E_z is sparse in columns, XE_z can capture the cluster-specific corruption approximately. So it is reasonable to impose the $l_{2,1}$-norm minimization constraint on E_z, and $E_{cs} = XE_z$ can characterize the cluster-specific corruptions of multi-view data properly. Accordingly, (3) is rewritten as follows:

$$X = XZC + XE_z + E_x. \tag{4}$$

For simplicity, we can reformulate the above equation as follows:

$$X = X(ZC + E_z) + E_x. \tag{5}$$

As a consequence, it is straightforward to design the following objective function for multi-view clustering based on the joint view representation X:

$$\min_{Z,C,E_x,E_z} \|E_x\|_{2,1} + \lambda_1 \|E_z\|_{2,1} + \lambda_2 \|C\|_* \tag{6}$$
$$\text{s.t.} \quad X = XZ + E_x, \ Z = ZC + E_z,$$

where λ_1 and λ_2 are trade-off parameters. Although E_x and E_z are both imposed with the $l_{2,1}$-norm constraint, they are totally different in essence. More specifically, E_x illustrates

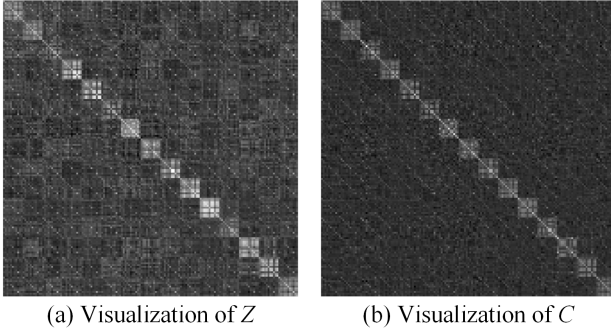(a) Visualization of Z (b) Visualization of C

Figure 3: Visualization of coefficient matrices obtained from the Yale Face dataset. (a) is derived from (2), and (b) is calculated from (6). where Z and C are both displayed in the form of $(\text{abs}(Z^T)+\text{abs}(Z))/2$ and $(\text{abs}(C^T)+\text{abs}(C))/2$, and obviously, C characters the underlying clustering structures of data much better than Z.

the sample-specific corruptions, and $E_{cs} = XE_z$ is used to process the cluster-specific corruptions caused by multiple-views. Theoretically, compared with the coefficient matrix obtained in (2), the coefficient matrix C is much better for multi-view clustering. To view the difference in a more intuitive way, Fig. 3 displays a visualization of Z and C conducted on the Yale Face dateset[1]. As shown in Fig. 3, it is clear that the matrix C has more suitable structures than Z for clustering.

3.2. gr-FCMSC

In FCMSC, only the consensus information of multi-view data is employed for clustering. In order to leverage the complementary information as well, a graph regularized Feature Concatenated Multi-view Subspace Clustering (gr-FCMSC) is also proposed in this paper. Graph regularization can preserve local manifold structures [4, 45, 16], and inspired by [45] we impose the following graph Laplacian regularizer of multiple views on the FCMSC to explore the complementary of multi-view data:

$$\sum_{i=1}^{v} Tr(C^T L_i C), \qquad (7)$$

where C^T denotes the transpose of C, L_i represents the graph Laplacian matrix of i-th view, $L_i = D_i - W_i$, and D_i is the degree matrix of the i-th view, W_i is the adjacency matrix of the i-th view [32]. And, the objective function of gr-FCMSC can be formulated as follows:

$$\min_{Z,C,E_x,E_z} \|E_x\|_{2,1} + \lambda_1 \|E_z\|_{2,1} + \lambda_2 \|C\|_*$$
$$+ \lambda_3 \sum_{i=1}^{v} Tr(C^T L_i C) \qquad (8)$$
$$\text{s.t.} \ \ X = XZ + E_x, \ \ Z = ZC + E_z$$

---

[1] The Yale Face database contains 165 grayscale images in GIF format of 15 individuals. More details will be presented in the section of experiment.

Algorithm 1 Optimization of the proposed FCMSC

INPUT:
  Multi-view data $\{x_1^{(i)}, x_2^{(i)}, \cdots, x_n^{(i)}\}_{i=1}^{v}$;
  $E_x = 0$, $E_z = 0$, $C = 0$, $J = 0$,
  $Y_1 = 0$, $Y_2 = 0$, $Y_3 = 0$,
  $\mu = 10^{-4}$, $\mu_{\max} = 10^6$, $\varepsilon = 10^{-6}$,
  Initialize Z with random values;
OUTPUT:
  C, Z, E_z, E_x;
REPEAT
  Update E_x according to the problem (12);
  Update E_z according to the problem (14);
  Update J according to the problem (15);
  Update C according to the problem (18);
  Update Z according to the problem (20);
  Update Y_1, Y_2, Y_3 and μ according to the problem (23);
UNTIL
  $\|X - XZ - E_x\|_\infty < \varepsilon$,
  $\|Z - ZC - E_z\|_\infty < \varepsilon$,
  and $\|C - J\|_\infty < \varepsilon$.

where λ_1, λ_2, and λ_3 denote trade-off parameters. Obviously, the desired coefficient matrix C, derived from (8), takes specific manifold structures of different views into consideration, in other words, the complementary information of multi-view data are also leveraged for clustering.

Once the desired coefficient matrix C is learned, we construct an adjacency matrix for spectral clustering to get multi-view clustering results as follows:

$$W = \frac{\text{abs}(C) + \text{abs}(C^T)}{2}, \qquad (9)$$

where $\text{abs}(\cdot)$ denotes the absolution function, which can deal with a matrix and return the absolute value of each element in the matrix.

4. Optimization

In this section, the optimization algorithms of the objective functions, including FCMSC and gr-FCMSC, are introduced in detail, then the computational complexity and convergence are analyzed as well.

4.1. Optimization for FCMSC

Although the optimization problem of (6) is not convex with respect to the variables, i.e. Z, C, E_x, and E_z, jointly, subproblems with respect to each of them are convex. So we apply the Alternating Direction Minimization strategy based on the Augmented Lagrangian Multiplier (ALM) [28] method to solve the objective function (6) effectively. Additionally, an auxiliary variable is introduced here so as to make the objective function separable and convenient for optimization. Accordingly, (6) can be reformulated equivalently as follows:

$$\min_{Z,C,E_z,E_x,J} \|E_x\|_{2,1} + \lambda_1 \|E_z\|_{2,1} + \lambda_2 \|J\|_*$$
$$\text{s.t.} \ \ X = XZ + E_x, \ \ Z = ZC + E_z, \ \ C = J. \qquad (10)$$

where J denotes the auxiliary variable, λ_1 and λ_2 are trade-off parameters.

The corresponding ALM problem of (10), which should be minimized in this section, can be shown as follows:

$$
\begin{aligned}
&\mathcal{L}(E_x, E_z, J, C, Y_1, Y_2, Y_3, \mu) \\
&= \|E_x\|_{2,1} + \lambda_1 \|E_z\|_{2,1} + \lambda_2 \|J\|_* \\
&+ \langle Y_1, X - XZ - E_x \rangle + \frac{\mu}{2} \|X - XZ - E_x\|_F^2 \\
&+ \langle Y_2, Z - ZC - E_z \rangle + \frac{\mu}{2} \|Z - ZC - E_z\|_F^2 \\
&+ \langle Y_3, C - J \rangle + \frac{\mu}{2} \|C - J\|_F^2,
\end{aligned}
\tag{11}
$$

where Y_1, Y_2, and Y_3 are Laplacian multipliers, μ indicates a positive adaptive penalty parameter, $\langle A, B \rangle$ denotes the trace of $A^T B$.

Since the Alternating Direction Minimization strategy is employed to minimize the above ALM problem, the whole problem is decomposed into several subproblems, which are convex and can be optimized effectively.

1) Updating E_x: To update E_x with other variables fixed, the following minimization problem should be optimized:

$$
\min_{E_x} \|E_x\|_{2,1} + \frac{\mu}{2} \left\| E_x - (X - XZ + \frac{Y_1}{\mu}) \right\|_F^2,
\tag{12}
$$

which has a closed-form solution.

Specifically, the solution of the above subproblem is denoted as E_x^*, and we can get the following closed-form solution [48]:

$$
[E_x^*]_{:,j} = \begin{cases} \frac{\|[T_E]_{:,j}\|_2 - \frac{1}{\mu}}{\|[T_E]_{:,j}\|_2} [T_E]_{:,j}, & \text{if } \left\| [T_E]_{:,j} \right\|_2 > \frac{1}{\mu} \\ 0 & , \text{ otherwise} \end{cases}
\tag{13}
$$

where $[A]_{:,j}$ represents the j-th column of the matrix A, and $T_E = X - XZ + \frac{Y_1}{\mu}$.

2) Updating E_z: The subproblem of updating E_z, in which other variables are all fixed, can be written as follows:

$$
\min_{E_z} \lambda_1 \|E_z\|_{2,1} + \frac{\mu}{2} \left\| E_z - (Z - ZC + \frac{Y_2}{\mu}) \right\|_F^2.
\tag{14}
$$

This subproblem is similar to the subproblem of updating E_x, and can be optimized effectively in the same way.

3) Updating J: With other variables fixed, we solve the following problem to update variable J:

$$
\min_J \lambda_2 \|J\|_* + \frac{\mu}{2} \left\| J - (C + \frac{Y_3}{\mu}) \right\|_F^2,
\tag{15}
$$

which can be optimized by leveraging the singular value threshold method [28]. Specifically, by setting $T_J = C + Y_3/\mu$ and performing singular value decomposition (SVD) on T_J, i.e. $T_J = U\Sigma V^T$, we achieve the optimization as follows:

$$
J = U S_{\lambda_2/\mu}(\Sigma) V^T,
\tag{16}
$$

where S_ε denotes a soft-thresholding operator as following and can be extended to matrices by applying it element-wise.

$$
S_\varepsilon(x) = \begin{cases} x - \varepsilon, & \text{if } x - \varepsilon > 0 \\ x + \varepsilon, & \text{if } x - \varepsilon < 0 \\ 0, & \text{otherwise.} \end{cases}
\tag{17}
$$

4) Updating C: When other variables are fixed, the subproblem with respect to C can be written as follows:

$$
\min_C \langle Y_2, Z - ZC - E_z \rangle + \frac{\mu}{2} \|Z - ZC - E_z\|_F^2 + \langle Y_3, C - J \rangle + \frac{\mu}{2} \|C - J\|_F^2.
\tag{18}
$$

In order to get an optimization, we take the derivative of the above function with respect to variable C and let the derivative to be zero, then obtain the following solution:

$$
\begin{aligned}
C &= T_{CA}^{-1} T_{CB}, \\
T_{CA} &= \mu(I + Z^T Z), \\
T_{CB} &= \mu J - Y_3 + Z^T Y_2 + \mu(Z^T Z - Z^T E_z),
\end{aligned}
\tag{19}
$$

where I is an identity matrix with the proper size.

5) Updating Z: With other variables being fixed, the subproblem of updating Z can be written as follows:

$$
\min_Z \langle Y_1, X - XZ - E_x \rangle + \frac{\mu}{2} \|X - XZ - E_x\|_F^2 + \langle Y_2, Z - ZC - E_z \rangle + \frac{\mu}{2} \|Z - ZC - E_z\|_F^2.
\tag{20}
$$

Differentiating (20) with respect to Z and letting it to be zero, the following equivalent equation can be achieved, solution of which is the optimization of this subproblem:

$$
T_{ZA}Z + Z T_{ZB} = T_{ZC},
\tag{21}
$$

where T_{ZA}, T_{ZB}, and T_{ZC} can be written as follows:

$$
\begin{aligned}
T_{ZA} &= X^T X + I, \\
T_{ZB} &= C C^T - C - C^T, \\
T_{ZC} &= X^T X - X^T E_x + E_z - E_z C^T \\
&+ \frac{1}{\mu} X^T Y_1 + \frac{1}{\mu}(Y_2 C^T - Y_2).
\end{aligned}
\tag{22}
$$

Equation (21) is a Sylvester equation and can be optimized effectively referring to [2].

6) Updating Lagrange multipliers and μ: According to [28], we update the Lagrange multipliers and the parameter μ as following:

$$
\begin{aligned}
Y_1 &= Y_1 + \mu(X - XZ - E_x), \\
Y_2 &= Y_2 + \mu(Z - ZC - E_z), \\
Y_3 &= Y_3 + \mu(C - J), \\
\mu &= \min(\rho\mu, \mu_{\max}),
\end{aligned}
\tag{23}
$$

where $\rho > 1$ and the parameter μ is monotonically increased by ρ until reaching the maximum, $\mu_{\max}$.

Algorithm 1 outlines the whole procedure of optimization for FCMSC. It is should be noticed that we random initialize Z in practice to avoid all zeros solutions.

4.2. Optimization for gr-FCMSC

Algorithm 1 can be generalized to optimize the problem of (8) in this section, and following ALM problem is constructed:

$$
\begin{aligned}
&\mathcal{L}(E_x, E_z, J, C, Y_1, Y_2, Y_3) \\
&= \|E_x\|_{2,1} + \lambda_1 \|E_z\|_{2,1} \\
&\quad + \lambda_2 \|J\|_* + \lambda_3 \sum_{i=1}^{v} Tr(C^T L_i C) \\
&\quad + \langle Y_1, X - XZ - E_x \rangle + \frac{\mu}{2} \|X - XZ - E_x\|_F^2 \\
&\quad + \langle Y_2, Z - ZC - E_z \rangle + \frac{\mu}{2} \|Z - ZVC - E_z\|_F^2 \\
&\quad + \langle Y_3, C - J \rangle + \frac{\mu}{2} \|C - J\|_F^2 .
\end{aligned}
\tag{24}
$$

Clearly, the subproblem with respect to C, which can be formulated as follows, is different from (18):

$$
\begin{aligned}
\min_C \ &\langle Y_2, Z - ZC - E_z \rangle + \frac{\mu}{2} \|Z - ZC - E_z\|_F^2 \\
&+ \langle Y_3, C - J \rangle + \frac{\mu}{2} \|C - J\|_F^2 + \lambda_3 \sum_{i=1}^{v} Tr(C^T L_i C)
\end{aligned}
\tag{25}
$$

And the optimization of the above problem is

$$
\begin{aligned}
C &= T_{CA}^{-1} T_{CB}, \\
T_{CA} &= \lambda_3 \sum_{i=1}^{v} (L_i^T + L_i) + \mu(I + Z^T Z), \\
T_{CB} &= \mu J - Y_3 + Z^T Y_2 + \mu(Z^T Z - Z^T E_z)
\end{aligned}
\tag{26}
$$

As for other subproblems, we optimize they according to Algorithm 1 straightforward. And we skip they over for the compactness of this paper.

4.3. Computational Complexity and Convergence

As shown in Algorithm 1, the main computational burden is composed of five parts, i.e. the five corresponding subproblems. The complexity of updating E_x is $O(dn^2 + n^3)$, and the complexity of updating E_z is $O(n^3)$, both of which are matrix multiplication. As for the subproblem of updating J, the complexity is $O(n^3)$. In the subproblem of updating C, the complexity is $O(n^3)$, since matrix inversion is included during optimization process. For updating Z, the Sylvester equation is optimized, and the complexity of this subproblem is $O(d^3 + n^3)$. To sum up, the computational complexity of each iteration is $O(dn^2 + d^3 + n^3)$.

For the convergence analysis, unfortunately, we find that it is difficulty to give any solid proof on the convergence of the proposed algorithm, since more than two subproblems are involved during the optimization. Inspired by [50, 63, 61], convergence discussion will be presented in the experiments section, extensive experimental results on the real-world datasets show that the proposed algorithm can converge effectively with all-zero initialization except for variable Z, which is initialized with random values.

5. Experiments

In this section, extensive experiments are conducted on six benchmark datasets. Accordingly, experimental results

are presented with the corresponding analyses. Both validation experiments and comparison experiments are provided, and the convergence properties and parameters sensitivity are analyzed as well. All codes are implemented in Matlab on a desktop with a four-core 3.6GHz processor and 8GB of memory.

5.1. Experimental Settings

To evaluate the performance of the proposed FCMSC, we employ six real-world datasets in experiments, including BBCSport[1] [24, 17], Movies 617[2], MSRCV1[3], Olympics[4], ORL[5], and Yale Face[6]. To be specific, BBCSport is collected from the BBC Sport website corresponding to sports news in 5 topical areas, and it consists of 544 documents, each which is divided into two sub-parts as two different views, and the standard TF-IDF normalization is utilized to obtain the corresponding features. Movie 617 is a movie dataset containing 617 movies of 17 genres, and it consists of two views, including keywords-mapping (view 1) and actors-mapping (view 2). MSRCV1 used in this paper consists of 210 images of 7 object classes, including building, cow, car airplane, tree, face, and bicycle, and 6 types of features are utilized, including: CENT (view1), CMT (view2), GIST (view3), HOG (view4), LBP (view5), and SIFT (view6). Olympics consists of 464 London 2012 Summer Olympics players' information active on Twitter of 28 different sports, and 9 different views are provide, including followedby-dictionary (view 1), follows-dictionary (view 2), mentionedby-dictionary (view 3), mentions-dictionary (view 4), retweets-dictionary (view 5), retweetedby-dictionary (view 6), listmerged500-dictionary (view 7), lists500-dictionary (view 8), tweets500-dictionary (view 9). ORL, which contains 400 images from 40 individuals, and Yale Face, which consists of 165 images from 15 individuals, are both face image datasets, three types of features, i.e., intensity (view1), LBP (view2) and Gabor (view3), are employed in datasets.

Meanwhile, three metrics are employed in this section to evaluate the clustering performance, including NMI (Normalized Mutual Information), ACC (accuracy), and F-score, which are commonly used in multi-view clustering [56, 3]. To make more precise statements, the definition of the ACC utilized in this paper can be written as follows:

$$
\text{ACC} = \frac{\sum_{i=1}^{n} \sigma(\tau_i, \text{map}(\omega_i))}{n},
\tag{27}
$$

where x_i denotes the i-th sample, ω_i is the clustering label of the i-th sample, and τ_i illustrates the corresponding ground-truth label. The function of $\text{map}(\omega_i)$ is the permutation map function, in which the Kuhn-Munkres algorithm is employed. And $\sigma(\cdot, \cdot)$ denotes the Dirac delta function. It

---

[1] http://mlg.ucd.ie/datasets/
[2] http://lig-membres.imag.fr/grimal/data/movies617.tar.gz
[3] http://research.microsoft.com/en-us/projects/objectclassrecognition/
[4] http://mlg.ucd.ie/aggregation/
[5] https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
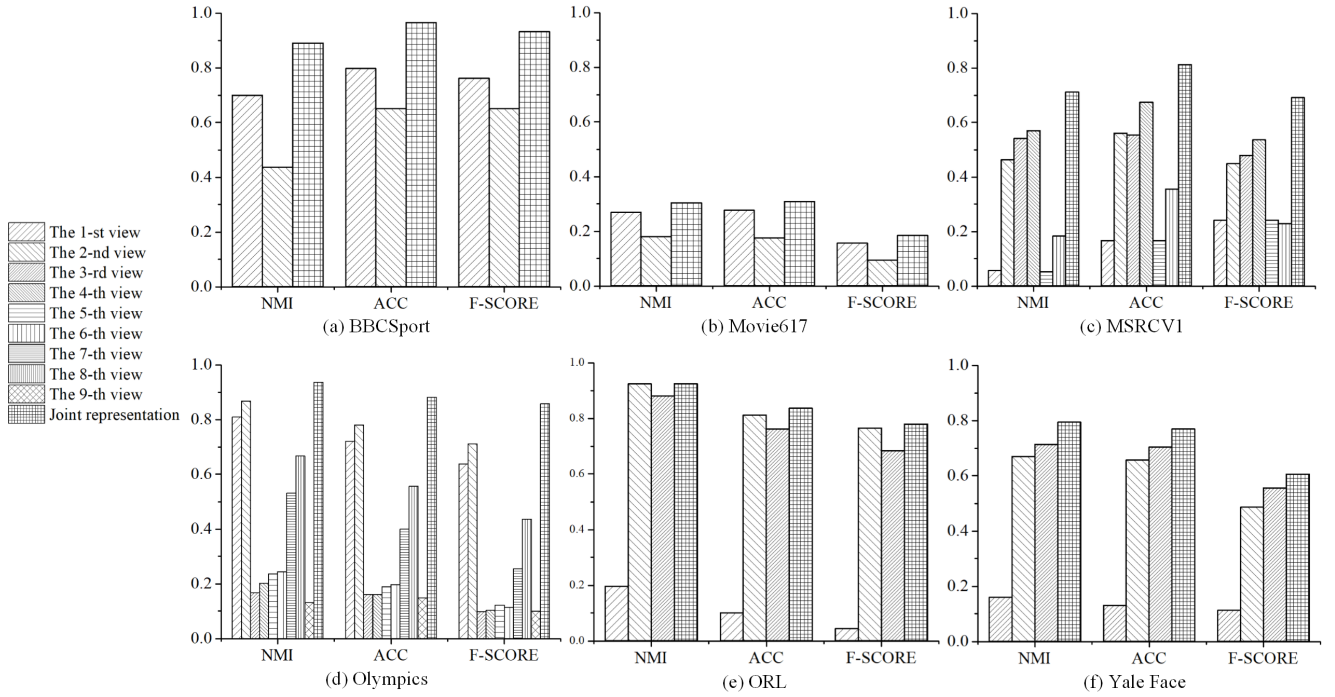[6] http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html

Figure 4: Clustering results comparison between the proposed FCMSC conducted on the joint view representation and LRR performed on each single view. Six benchmark datasets are employed and clustering results are presented in the metric of NMI, ACC, and F-Score.

should be noted that the higher value of all metrics corresponds the better clustering performance. All parameters of the competed methods are fine-tuned. To eliminate the randomness, 30 Monte Carlo (MC) trials are conducted with respect to each benchmark dataset. Experimental results are reported in form of the mean value and the standard deviation, and the best and the second best clustering results are present in bold font.

5.2. Validation Experiments

To validate our method, we compare the clustering results, achieved by the proposed FCMSC, with the clustering results, obtained by performing LRR on each single view. Specifically, validation experiments are conducted on all six benchmark datasets, and we shown the clustering performance of our methods and LRR on each single view with respect to NMI, ACC and F-score.

As shown in Fig. 4, the clustering performance of our FCMSC based on the joint view representation is much better than those of all single view. Taking BBCSport as example, NMI and ACC obtained by LRR [29] with the best single view are 69.96% and 79.70% respectively. As for the proposed FCMSC based on the concatenated features, NMI and ACC are 89.04% and 96.51% respectively. In other words, our FCMSC achieves a relative increase of 27.27% and 21.09% with respect to NMI and ACC. Since FCMSC performs clustering on multiple view simultaneously and handles the cluster-specific corruptions properly, it can take advantage of consensus information to improve

clustering results. Therefore, the proposed FCMSC is valid and can achieve promising clustering performance for multi-view data.

5.3. Comparison Experiments

To demonstrate the competitiveness of our FCMSC and gr-FCMSC, eleven approaches are employed for comparison and listed as follows:

1) SC_{BSV} [32]: Spectral Clustering of the Best Single View. Spectral clustering algorithm is employed on each single view, and the best clustering results of these views are presented.

2) SC_{FC}: Spectral Clustering based on the Concatenated Features. Features of multiple views are concatenated, and then spectral clustering algorithm is applied to the joint view representation.

3) LRR_{BSV} [29]: Low-Rank Representation of the Best Single View. Similar to the SC_{BSV}, low-rank representation algorithm is conducted on each view, and the results of the view with the best clustering performance are reported.

4) LRR_{FC}: Low-Rank Representation based on the Concatenated Features. We apply the low-rank representation algorithm to the joint view representation to get multi-view clustering results.

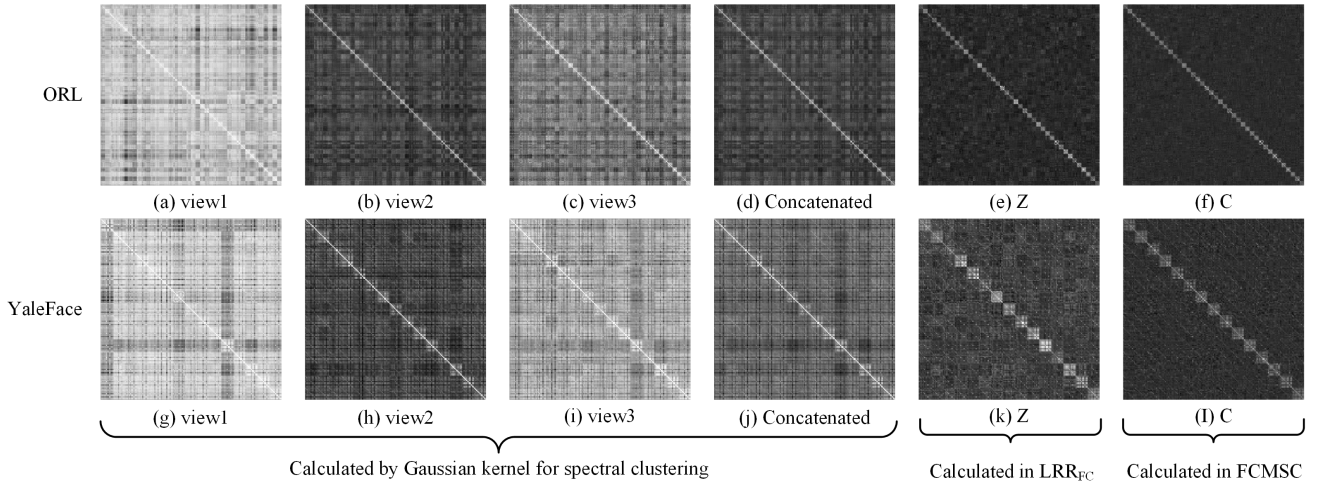5) Kernel Addition [13]: This approach combines information of multi-view data by averaging the sum of ker-

Figure 5: Visualization of adjacency matrices (a, b, c, d, g, h, i, and j) and coefficient matrices (e, f, k, and l), where Z and C are both displayed in the form of $(\text{abs}(Z^T)+\text{abs}(Z))/2$ and $(\text{abs}(C^T)+\text{abs}(C))/2$.

nel matrices of all views, then gets clustering results based on spectral clustering.

6) Co-reg [25]: Co-regularized multi-view spectral clustering. This approach clusters multi-view data by pursuing graph similarity agreement of multiple views.

7) RMSC [44]: Robust Multi-View Spectral Clustering via Low-Rank and Sparse Decomposition. RMSC recovers a common low-rank transition probability matrix via low-rank and sparse decomposition, and then obtain clustering results based on the standard Markov chains.

8) LMSC [58]: Latent Multi-view Subspace Clustering. It learns a latent multi-view representation and attains the corresponding subspace coefficient matrix simultaneously.

9) MLRSSC [3]: Multi-view low-rank sparse subspace clustering. It learns the affinity matrix of multi-view data with the low-rank and sparsity constraints. Linear kernel MLRSSC algorithm is employed here for comparison.

10) DCCA [1]: Deep Canonical Correlation Analysis. It is a learning method which employs the neural networks to extend Canonical Correlation Analysis. We use k-means to obtain cluster results.

11) DCCAE [43]: Deep Canonically Correlated Autoencoders. Two autoencoders are employed and DCCAE tries to maximize the canonical correlation between two views. K-means is utilized for clustering.

As shown in Table 2 and 3, experimental results are reported in form of the mean score, as well as the standard deviation. Overall, the proposed FCMSC and gr-FCMSC can achieve better clustering results on six benchmark datasets than other competed multi-view clustering approaches with respect to all the clustering metrics. For example, on MSRCV1 dataset with six views, our FCMSC gains a relative increase of 6.01%, 4.48% and 5.92% with respect to NMI, ACC, and F-Score, respectively, comparing with the corresponding best competed method. Moreover, the proposed gr-FCMSC get 13.74%, 10.69%, and 16.03% relative improvement in metrics of NMI, ACC, and F-Score as well. Furthermore, take experiments on ORL for examples, some statistical experimental results on ORL dataset are presented in Fig. 6, it can be observed clearly that the proposed FCMSC and gr-FCMSC can achieve significantly improvement compared with competitors. Compared with deep learning based methods, i.e., DCCA and DCCAE, our methods can also achieve the better clustering performance, since both the sample-specific and the cluster-specific corruptions are taken into account for clustering.

Compared with LRR_{FC}, the proposed FCMSC takes the cluster-specific corruptions, which are brought by multiple views, into consideration, and the clustering results indicate that it is essential to handle the clustering-specific corruptions during multi-view clustering. Since each view has its own specific properties that may be contrary to other views, it is difficult to explore and utilize the consensus information of multi-view data by performing some existing single-view clustering approaches on the concatenated features. To get an intuitive analysis, taking experiments on ORL and Yale Face datasets for example, Fig. 5 presents the visualization of adjacency matrices, which are calculated from each view and concatenated features by different methods. Clearly, the adjacency matrix C achieved by our proposed FCMSC has more suitable underlying structures for clustering.

5.4. Comparisons Between FCMSC and gr-FCMSC

Both the proposed FCMSC and gr-FCMSC perform clustering on multiple views simultaneously, and get promising clustering results. The consensus information of multi-

Table 2
Comparison results of different methods on the benchmark datasets

| Dataset	Method	NMI	ACC	F-SCORE
BBCSport	SC_{BSV}	0.7182 (0.0054)	0.8456 (0.0099)	0.7671 (0.0067)
	SC_{FC}	0.8019 (0.0095)	0.8505 (0.0262)	0.8452 (0.0206)
	LRR_{BSV}	0.6996 (0.0001)	0.7970 (0.0015)	0.7612 (0.0001)
	LRR_{FC}	0.5580 (0.0110)	0.6684 (0.0088)	0.6064 (0.0055)
	Kernel Addition	0.6170 (0.0085)	0.7347 (0.0099)	0.6684 (0.0059)
	Co-reg	0.7185 (0.0031)	0.8465 (0.0050)	0.7674 (0.0041)
	RMSC	0.8124 (0.0074)	0.8562 (0.0198)	0.8514 (0.0132)
	LMSC	0.8393 (0.0043)	0.9180 (0.0031)	0.8996 (0.0033)
	MLRSSC	0.8855 (0.0000)	**0.9651 (0.0000)**	0.9296 (0.0000)
	DCCA	0.2779 (0.0041)	0.5490 (0.0042)	0.4116 (0.0010)
	DCCAE	0.3298 (0.0046)	0.5438 (0.0089)	0.4088 (0.0080)
	FCMSC	**0.8904 (0.0000)**	**0.9651 (0.0000)**	**0.9317 (0.0000)**
	gr-FCMSC	**0.8973 (0.0000)**	**0.9670 (0.0000)**	**0.9348 (0.0000)**
Movies 617	SC_{BSV}	0.2606 (0.0020)	0.2579 (0.0035)	0.1481 (0.0025)
	SC_{FC}	0.2668 (0.0017)	0.2604 (0.0033)	0.1542 (0.0019)
	LRR_{BSV}	0.2667 (0.0059)	0.2747 (0.0071)	0.1545 (0.0047)
	LRR_{FC}	0.2839 (0.0075)	0.2824 (0.0135)	0.1813 (0.0063)
	Kernel Addition	0.2917 (0.0026)	0.2901 (0.0049)	0.1764 (0.0033)
	Co-reg	0.2454 (0.0018)	0.2396 (0.0017)	0.1381 (0.0016)
	RMSC	0.2957 (0.0032)	0.2971 (0.0040)	0.1810 (0.0028)
	LMSC	0.2813 (0.0098)	0.2747 (0.0094)	0.1606 (0.0068)
	MLRSSC	0.2975 (0.0061)	0.2887 (0.0111)	0.1766 (0.0068)
	DCCA	0.1764 (0.0002)	0.1948 (0.0018)	0.1141 (0.0014)
	DCCAE	0.1759 (0.0061)	0.2009 (0.0118)	0.1214 (0.0056)
	FCMSC	**0.3043 (0.0052)**	**0.3090 (0.0063)**	**0.1852 (0.0034)**
	gr-FCMSC	**0.3169 (0.0059)**	**0.3051 (0.0049)**	**0.1930 (0.0035)**
MSRCV1	SC_{BSV}	0.6047 (0.0112)	0.6826 (0.0171)	0.5724 (0.0122)
	SC_{FC}	0.4398 (0.0021)	0.5073 (0.0077)	0.3978 (0.0032)
	LRR_{BSV}	0.5704 (0.0054)	0.6732 (0.0091)	0.5368 (0.0076)
	LRR_{FC}	0.6257 (0.0105)	0.6871 (0.0105)	0.5913 (0.0142)
	Kernel Addition	0.6176 (0.0087)	0.7102 (0.0130)	0.5973 (0.0097)
	Co-reg	0.6583 (0.0106)	0.7674 (0.0169)	0.6459 (0.0128)
	RMSC	0.6696 (0.0064)	0.7819 (0.0125)	0.6614 (0.0093)
	LMSC	0.6162 (0.0676)	0.6992 (0.0700)	0.5936 (0.0763)
	MLRSSC	0.6709 (0.0352)	0.7774 (0.0497)	0.6524 (0.0470)
	DCCA	0.6606 (0.0000)	0.7429 (0.0000)	0.6270 (0.0000)
	DCCAE	0.6782 (0.0040)	0.7662 (0.0015)	0.6462 (0.0025)
	FCMSC	**0.7112 (0.0031)**	**0.8122 (0.0030)**	**0.6910 (0.0046)**
	gr-FCMSC	**0.7631 (0.0036)**	**0.8605 (0.0022)**	**0.7570 (0.0036)**
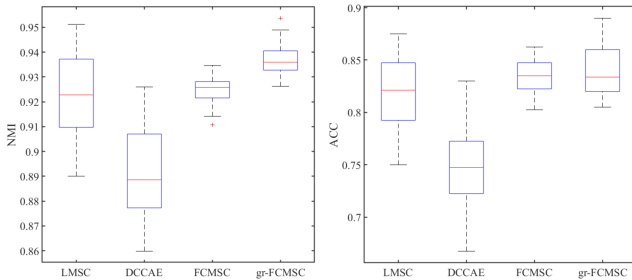
Figure 6: Clustering results on ORL in metrics of NMI (left column) and ACC (right column).

view data is both leveraged by FCMSC and gr-FCMSC. Additionally, gr-FCMSC also takes advantage of the complementary of multiple views by means of graph Laplacian regularizers for multi-view clustering. As shown in Table 2, gr-FCMSC improves FCMSC significantly by exploring the complementary information. For example, Compared with FCMSC, gr-FCMSC gains 7.30% and 5.95% relative improvement on MSRCV1 in metrics of NMI and ACC. And on Movie 617 dataset, gr-FCMSC achieves 3.42% and 4.04% relative improvement with respect to the metrics of NMI and F-Score.

5.5. Parameters Sensitivity and Convergence Analysis

Besides, Convergence analysis, shown in Fig. 7, and parameters influence of the proposed methods, shown in Fig.

Table 3
Comparison results of different methods on the benchmark datasets

Dataset	Method	NMI	ACC	F-SCORE
Olympics	SC_{BSV}	0.7617 (0.0046)	0.6288 (0.0112)	0.5178 (0.0134)
	SC_{FC}	0.5625 (0.0038)	0.4610 (0.0078)	0.3194 (0.0085)
	LRR_{BSV}	0.8674 (0.0038)	0.7830 (0.0093)	0.7112 (0.0088)
	LRR_{FC}	0.8910 (0.0054)	0.7746 (0.0190)	0.7532 (0.0190)
	Kernel Addition	0.7245 (0.0038)	0.6093 (0.0073)	0.5189 (0.0071)
	Co-reg	0.8308 (0.0027)	0.7341 (0.0071)	0.6707 (0.0079)
	RMSC	0.7573 (0.0063)	0.6372 (0.0108)	0.5687 (0.0117)
	LMSC	0.8902 (0.0065)	0.8043 (0.0140)	0.7814 (0.0154)
	MLRSSC	0.9122 (0.0067)	0.8454 (0.0208)	0.8236 (0.0285)
	DCCA	0.7782 (0.0015)	0.6474 (0.0047)	0.4024 (0.0072)
	DCCAE	0.7686 (0.0042)	0.6515 (0.0115)	0.4096 (0.0172)
	FCMSC	**0.9357 (0.0062)**	**0.8815 (0.0199)**	**0.8576 (0.0264)**
	gr-FCMSC	**0.9389 (0.0037)**	**0.8890 (0.0137)**	**0.8649 (0.0176)**
ORL	SC_{BSV}	0.8868 (0.0069)	0.7459 (0.0121)	0.6805 (0.0159)
	SC_{FC}	0.8084 (0.0027)	0.6323 (0.0061)	0.5236 (0.0069)
	LRR_{BSV}	0.9240 (0.0054)	0.8122 (0.0203)	0.7650 (0.0166)
	LRR_{FC}	0.8497 (0.0085)	0.7178 (0.0190)	0.6119 (0.0218)
	Kernel Addition	0.8028 (0.0033)	0.6349 (0.0074)	0.5224 (0.0061)
	Co-reg	0.8277 (0.0040)	0.6653 (0.0080)	0.5672 (0.0092)
	RMSC	0.8885 (0.0056)	0.7482 (0.0128)	0.6866 (0.0139)
	LMSC	0.9215 (0.0168)	0.8193 (0.0360)	0.7623 (0.0419)
	MLRSSC	0.9102 (0.0113)	0.8042 (0.0234)	0.7459 (0.0281)
	DCCA	0.8589 (0.0025)	0.7093 (0.0041)	0.5716 (0.0001)
	DCCAE	0.8906 (0.0171)	0.7476 (0.0356)	0.6741 (0.0399)
	FCMSC	**0.9249 (0.0055)**	**0.8359 (0.0165)**	**0.7792 (0.0180)**
	gr-FCMSC	**0.9370 (0.0065)**	**0.8382 (0.0229)**	**0.7991 (0.0227)**
Yale Face	SC_{BSV}	0.6229 (0.0354)	0.5715 (0.0497)	0.4319 (0.0472)
	SC_{FC}	0.5761 (0.0335)	0.5145 (0.0460)	0.3653 (0.0420)
	LRR_{BSV}	0.7134 (0.0098)	0.7034 (0.0125)	0.5561 (0.0159)
	LRR_{FC}	0.6917 (0.0190)	0.6667 (0.0236)	0.4941 (0.0303)
	Kernel Addition	0.5872 (0.0320)	0.5352 (0.0397)	0.3823 (0.0390)
	Co-reg	0.6146 (0.0084)	0.5638 (0.0108)	0.4208 (0.0110)
	RMSC	0.6590 (0.0108)	0.6091 (0.0161)	0.4773 (0.0133)
	LMSC	0.7073 (0.0105)	0.6758 (0.0116)	0.5138 (0.0172)
	MLRSSC	0.7005 (0.0311)	0.6733 (0.0384)	0.5399 (0.0377)
	DCCA	0.7642 (0.0004)	0.7392 (0.0011)	0.6159 (0.0012)
	DCCAE	0.6888 (0.0226)	0.6442 (0.0291)	0.5152 (0.0307)
	FCMSC	**0.7939 (0.0206)**	**0.7691 (0.0267)**	**0.6058 (0.0306)**
	gr-FCMSC	**0.7979 (0.0202)**	**0.7717 (0.0222)**	**0.6095 (0.0335)**

8, Fig. 9, and Fig. 10, are discussed in the this section as well.

In the proposed methods, as shown in the objective functions (6) and (8), there are two trade-off parameters required to be fine-tuned, i.e. λ_1 and λ_2, for FCMSC, and an extra trade-off parameter λ_3 for gr-FCMSC. Values of λ_1 and λ_2 are selected from $\{1, 10, 100, 1000, 10000\}$. As shown in Fig. 8 and Fig. 9, it can be observed that promising clustering results can be attained when λ_1 is relative large and λ_2 equals to 100. To be specific, curves drawn in Fig. 8 demonstrate the influence of λ_1 and λ_2, respectively. For λ_1, it can be observed that promising clustering results can be achieved with a relatively large value. As for λ_2, generally speaking, the prior knowledge of the dataset error level determines the choice of λ_2 mainly, and the promis-

ing clustering performance can be achieved with $\lambda_2 = 100$ on MSRCV1 dataset. Meanwhile, the influence of λ_3 for gr-FCMSC is also discussed as shown in Fig. 9, it can be observed that the best clustering performance can be attained with $\lambda_3 = 0.01$, and the clustering performance degenerates when the value of λ_3 is larger than 0.01.

Meanwhile, we explore the convergence properties of the proposed FCMSC and gr-FCMSC. Fig. 7 displays the convergence of our approaches conducted on three datasets, including BBCSport, MSRCV1, and Yale Face. It can be observed that both FCMSC and gr-FCMSC can achieve the quick convergence within 40 iterations. Although it is difficult for us to give an solid proof on the convergence, experimental results demonstrate the effectiveness and convergence of our methods empirically.
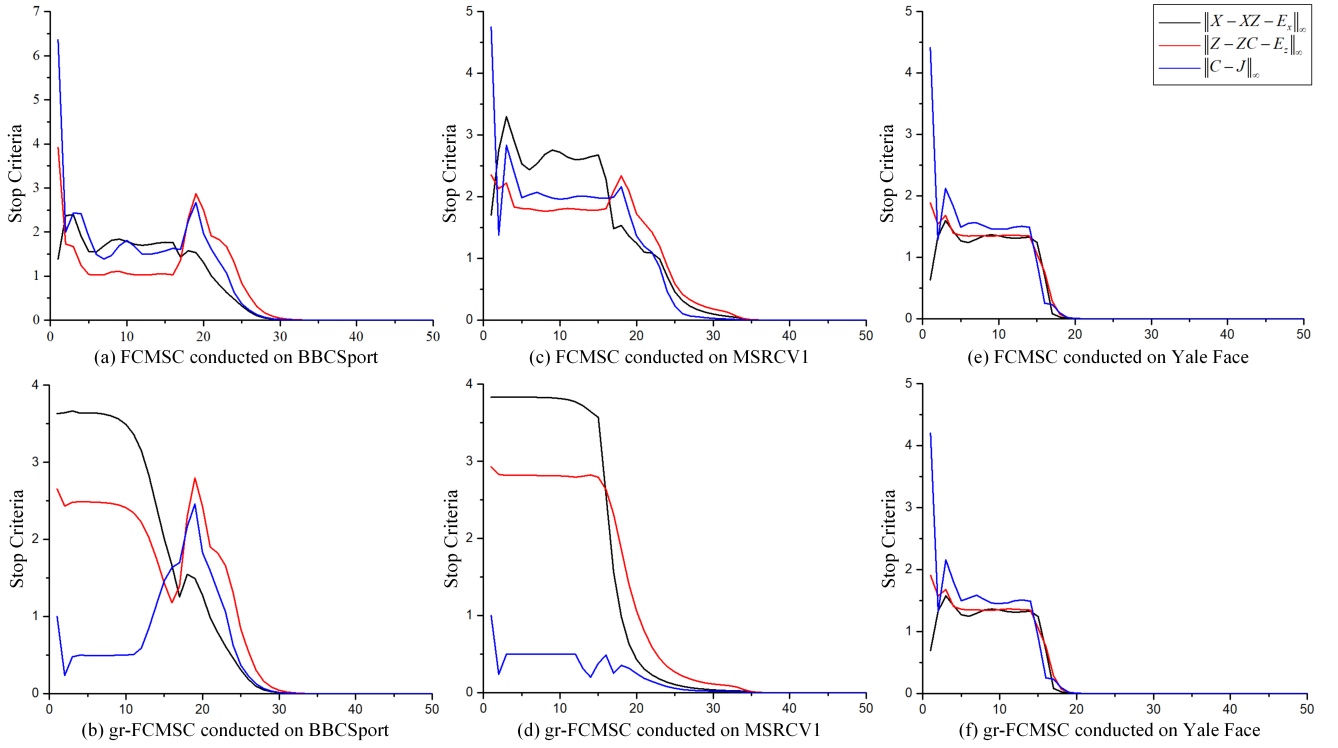
Figure 7: Convergence of the proposed FCMSC and gr-FCMSC. Convergence curves about the stop criteria of reconstruction errors versus the iteration numbers on three datasets, including (a)-(b) BBCSport, (c)-(d) MSRCV1, and (e)-(f) Yale Face, are displayed in this section.
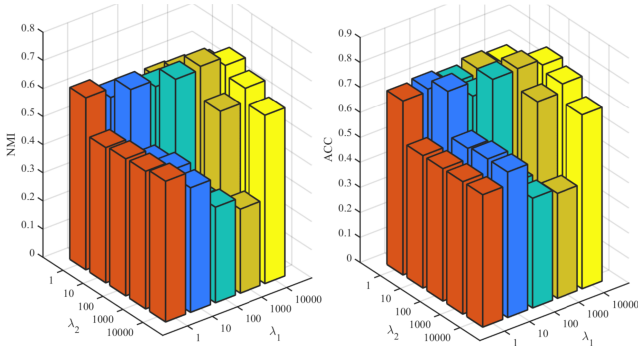

Figure 8: Clustering results of proposed FCMSC with different λ_1 and λ_2 on MSRCV1 dataset.

Figure 9: Influence of λ_1 and λ_2 on MSRCV1 dataset. (a) illustrates the influence of λ_1 with $\lambda_2 = 100$, (b) illustrates the influence of λ_2 with $\lambda_1 = 1000$.

6. Conclusion

This paper proposes a feature concatenation multi-view subspace clustering approach, termed FCMSC, and a graph regularized FCMSC (gr-FCMSC) as well. Different from most of existing approaches, the proposed methods can perform clustering on all views simultaneously by exploring the consensus information and complementary information of multi-view data based on the concatenated features. By taking the cluster-specific corruptions into consideration, the proposed methods can obtain a desired coefficient matrix and achieve promising clustering results. Extensive exper-
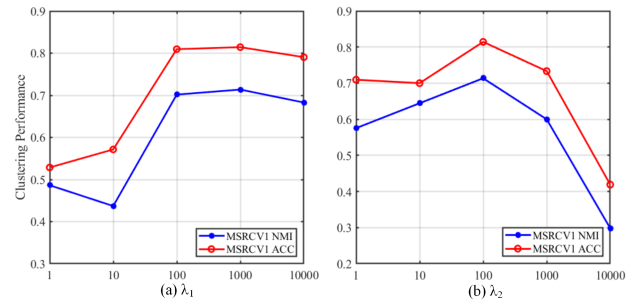
iments on six benchmark datasets demonstrate the superiority of our approach over some state-of-the-arts.

Despite effectiveness of the proposed methods, they are time consuming due to the operation of matrix inversion and SVD decomposition involved in the optimization, especially when the number of data is large. Further work will focus on the improvement of proposed methods for large-scale data, by employing the dimensionality reduction and the binary representation [62] strategies. And for the gr-FCMSC, the incompatible graph information among multiple views is ignored, further work will also focus on this problem.
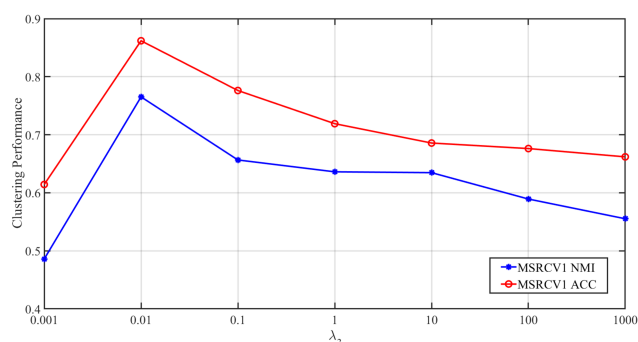
Figure 10: Clustering results of proposed gr-FCMSC with different λ_3 on MSRCV1 dataset.

Acknowledgements

References

[1] Andrew, G., Arora, R., Bilmes, J., Livescu, K., 2013. Deep canonical correlation analysis, in: International conference on machine learning, pp. 1247–1255.

[2] Bartels, R.H., Stewart, G.W., 1972. Solution of the matrix equation ax+xb=c [f4] (algorithm 432). Communications of the Acm 15, 820–826.

[3] Brbić, M., Kopriva, I., 2018. Multi-view low-rank sparse subspace clustering. Pattern Recognition 73, 247–258.

[4] Cai, D., He, X., Han, J., Huang, T.S., 2010. Graph regularized nonnegative matrix factorization for data representation. IEEE transactions on pattern analysis and machine intelligence 33, 1548–1560.

[5] Cao, X., Zhang, C., Fu, H., Liu, S., Zhang, H., 2015. Diversity-induced multi-view subspace clustering, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 586–594.

[6] Chao, G., Sun, S., 2016a. Consensus and complementarity based maximum entropy discrimination for multi-view classification. Information Sciences 367, 296–310.

[7] Chao, G., Sun, S., 2016b. Multi-kernel maximum entropy discrimination for multi-view learning. Intelligent Data Analysis 20, 481–493.

[8] Chao, G., Sun, S., Bi, J., 2017. A survey on multi-view clustering. arXiv preprint arXiv:1712.06246 .

[9] Chaudhuri, K., Kakade, S.M., Livescu, K., Sridharan, K., 2009. Multi-view clustering via canonical correlation analysis, in: Proceedings of the 26th annual international conference on machine learning, ACM. pp. 129–136.

[10] Chen, C., Qian, H., Chen, W., Zheng, Z., Zhu, H., 2019. Auto-weighted multi-view constrained spectral clustering. Neurocomputing .

[11] Chen, H., Wang, W., Feng, X., He, R., 2018a. Discriminative and coherent subspace clustering. Neurocomputing 284, 177–186.

[12] Chen, M., Wang, Q., Li, X., 2018b. Adaptive projected matrix factorization method for data clustering. Neurocomputing 306, 182–188.

[13] Cortes, C., Mohri, M., Rostamizadeh, A., 2009. Learning non-linear combinations of kernels, in: Advances in neural information processing systems, pp. 396–404.

[14] Elhamifar, E., Vidal, R., 2013. Sparse subspace clustering: Algorithm, theory, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 2765–2781.

[15] Gao, H., Nie, F., Li, X., Huang, H., 2015. Multi-view subspace clustering, in: Proceedings of the IEEE international conference on computer vision, pp. 4238–4246.

[16] Gao, S., Tsang, I.W.H., Chia, L.T., 2012. Laplacian sparse coding, hypergraph laplacian sparse coding, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 92–104.

[17] Greene, D., Cunningham, P., 2006. Practical solutions to the problem of diagonal dominance in kernel document clustering, in: Proceedings of the 23rd international conference on Machine learning, ACM. pp. 377–384.

[18] Guo, D., Zhang, J., Liu, X., Cui, Y., Zhao, C., 2014. Multiple kernel learning based multi-view spectral clustering, in: International Conference on Pattern Recognition, pp. 3774–3779.

[19] Hong, C., Yu, J., Wan, J., Tao, D., Wang, M., 2015. Multimodal deep autoencoder for human pose recovery. IEEE Transactions on Image Processing 24, 5659–5670.

[20] Hu, H., Lin, Z., Feng, J., Zhou, J., 2014. Smooth representation clustering, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3834–3841.

[21] Hu, L., Xing, Y., Gong, Y., Zhao, K., Wang, F., 2019. Nonnegative matrix tri-factorization with user similarity for clustering in point-of-interest. Neurocomputing 363, 58–65.

[22] Huang, S., Ren, Y., Xu, Z., 2018. Robust multi-view data clustering with multi-view capped-norm k-means. Neurocomputing 311, 197–208.

[23] Ji, P., Zhang, T., Li, H., Salzmann, M., Reid, I., 2017. Deep subspace clustering networks, in: Advances in Neural Information Processing Systems, pp. 24–33.

[24] Kumar, A., Daumé, H., 2011. A co-training approach for multi-view spectral clustering, in: Proceedings of the 28th International Conference on Machine Learning (ICML-11), pp. 393–400.

[25] Kumar, A., Rai, P., Daume, H., 2011. Co-regularized multi-view spectral clustering, in: Advances in neural information processing systems, pp. 1413–1421.

[26] LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. nature 521, 436.

[27] Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401, 788.

[28] Lin, Z., Liu, R., Su, Z., 2011. Linearized alternating direction method with adaptive penalty for low-rank representation, in: Advances in neural information processing systems, pp. 612–620.

[29] Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., Ma, Y., 2013a. Robust recovery of subspace structures by low-rank representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 171–184.

[30] Liu, J., Wang, C., Gao, J., Han, J., 2013b. Multi-view clustering via joint nonnegative matrix factorization, in: Proceedings of the 2013 SIAM International Conference on Data Mining, SIAM. pp. 252–260.

[31] Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International journal of computer vision 60, 91–110.

[32] Luxburg, U.V., 2007. A tutorial on spectral clustering. Statistics and Computing 17, 395–416.

[33] Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution grayscale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence 24, 971–987.

[34] Oliva, A., Torralba, A., 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of computer vision 42, 145–175.

[35] Patel, V.M., Van Nguyen, H., Vidal, R., 2015. Latent space sparse and low-rank subspace clustering. IEEE Journal of Selected Topics in Signal Processing 9, 691–701.

[36] Song, J., Yoon, G., Hahn, K., Yoon, S.M., 2019a. Subspace clustering via structure-enforced dictionary learning. Neurocomputing 362, 1–10.

[37] Song, J., Yoon, G., Hahn, K., Yoon, S.M., 2019b. Subspace clustering via structure-enforced dictionary learning. Neurocomputing 362, 1–10.

[38] Sun, S., Chao, G., 2013. Multi-view maximum entropy discrimination., in: IJCAI, pp. 1706–1712.

[39] Tang, C., Zhu, X., Liu, X., Li, M., Wang, P., Zhang, C., Wang, L., 2018. Learning joint affinity graph for multi-view subspace cluster-

ing. IEEE Transactions on Multimedia .

[40] Tzortzis, G.F., Likas, A.C., 2010. Multiple view clustering using a weighted combination of exemplar-based mixture models. IEEE Transactions on neural networks 21, 1925–1938.

[41] Vidal, R., Favaro, P., 2014. Low rank subspace clustering (lrsc). Pattern Recognition Letters 43, 47–61.

[42] Wang, Q., Dou, Y., Liu, X., Lv, Q., Li, S., 2016. Multi-view clustering with extreme learning machine. Neurocomputing 214, 483–494.

[43] Wang, W., Arora, R., Livescu, K., Bilmes, J., 2015. On deep multi-view representation learning, in: International Conference on Machine Learning, pp. 1083–1092.

[44] Xia, R., Pan, Y., Du, L., Yin, J., 2014. Robust multi-view spectral clustering via low-rank and sparse decomposition., in: AAAI, pp. 2149–2155.

[45] Xia, T., Tao, D., Mei, T., Zhang, Y., 2010. Multiview spectral embedding. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 40, 1438–1446.

[46] Xu, C., Tao, D., Xu, C., 2013. A survey on multi-view learning. arXiv preprint arXiv:1304.5634 .

[47] Xue, Z., Du, J., Du, D., Lyu, S., 2019. Deep low-rank subspace ensemble for multi-view clustering. Information Sciences 482, 210–227.

[48] Yang, J., Yin, W., Zhang, Y., Wang, Y., 2009. A fast algorithm for edge-preserving variational multichannel image restoration. SIAM Journal on Imaging Sciences 2, 569–592.

[49] Yi, X., Xu, Y., Zhang, C., 2005. Multi-view em algorithm for finite mixture models, in: International Conference on Pattern Recognition and Image Analysis, Springer. pp. 420–425.

[50] Yin, M., Gao, J., Lin, Z., Shi, Q., Guo, Y., 2015a. Dual graph regularized latent low-rank representation for subspace clustering. IEEE Transactions on Image Processing 24, 4918–4933.

[51] Yin, Q., Wu, S., He, R., Wang, L., 2015b. Multi-view clustering via pairwise sparse subspace representation. Neurocomputing 156, 12–21.

[52] Yu, J., Rui, Y., Tao, D., 2014. Click prediction for web image reranking using multimodal sparse coding. IEEE Transactions on Image Processing 23, 2019–2032.

[53] Yu, J., Tan, M., Zhang, H., Tao, D., Rui, Y., 2019a. Hierarchical deep click feature prediction for fine-grained image recognition. IEEE transactions on pattern analysis and machine intelligence .

[54] Yu, J., Zhu, C., Zhang, J., Huang, Q., Tao, D., 2019b. Spatial pyramid-enhanced netvlad with weighted triplet loss for place recognition. IEEE transactions on neural networks and learning systems .

[55] Zhai, L., Zhu, J., Zheng, Q., Pang, S., Li, Z., Wang, J., 2019. Multi-view spectral clustering via partial sum minimisation of singular values. Electronics Letters 55, 314–316.

[56] Zhang, C., Fu, H., Hu, Q., Cao, X., Xie, Y., Tao, D., Xu, D., 2018a. Generalized latent multi-view subspace clustering. IEEE transactions on pattern analysis and machine intelligence .

[57] Zhang, C., Fu, H., Liu, S., Liu, G., Cao, X., 2015. Low-rank tensor constrained multiview subspace clustering, in: Proceedings of the IEEE international conference on computer vision, pp. 1582–1590.

[58] Zhang, C., Hu, Q., Fu, H., Zhu, P., Cao, X., 2017a. Latent multi-view subspace clustering, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE. pp. 4333–4341.

[59] Zhang, J., Yu, J., Tao, D., 2018b. Local deep-feature alignment for unsupervised dimension reduction. IEEE transactions on image processing 27, 2420–2432.

[60] Zhang, T., Popescul, A., Dom, B., 2006. Linear prediction models with graph regularization for web-page categorization, in: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM. pp. 821–826.

[61] Zhang, Z., Lai, Z., Xu, Y., Shao, L., Wu, J., Xie, G.S., 2017b. Discriminative elastic-net regularized linear regression. IEEE Transactions on Image Processing 26, 1466–1481.

[62] Zhang, Z., Liu, L., Shen, F., Shen, H.T., Shao, L., 2018c. Binary multi-view clustering. IEEE Transactions on Pattern Analysis and

Machine Intelligence PP, 1–1.

[63] Zhang, Z., Xu, Y., Shao, L., Yang, J., 2017c. Discriminative block-diagonal representation learning for image recognition. IEEE transactions on neural networks and learning systems 29, 3111–3125.

[64] Zhao, H., Ding, Z., Fu, Y., 2017a. Multi-view clustering via deep matrix factorization., in: AAAI, pp. 2921–2927.

[65] Zhao, J., Xie, X., Xu, X., Sun, S., 2017b. Multi-view learning overview: Recent progress and new challenges. Information Fusion 38, 43–54.

[66] Zhao, L., Chen, Z., Yang, Y., Wang, Z.J., Leung, V.C., 2018. Incomplete multi-view clustering via deep semantic mapping. Neurocomputing 275, 1053–1062.

[67] Zheng, W., Yan, H., Yang, J., 2019. Robust unsupervised feature selection by nonnegative sparse subspace learning. Neurocomputing 334, 156–171.

[68] Zhou, P., Hou, Y., Feng, J., 2018a. Deep adversarial subspace clustering, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1596–1604.

[69] Zhou, T., Zhang, C., Gong, C., Bhaskar, H., Yang, J., 2018b. Multiview latent space learning with feature redundancy minimization. IEEE Transactions on Cybernetics .

[70] Zhou, T., Zhang, C., Peng, X., Bhaskar, H., Yang, J., 2019. Dual shared-specific multiview subspace clustering. IEEE transactions on cybernetics .

[71] Zhou, Z.H., 2012. Ensemble methods: foundations and algorithms. Chapman and Hall/CRC.