

# Robust Rank-Constrained Sparse Learning: A Graph-Based Framework for Single View and Multiview Clustering

Qi Wang<sup>1</sup>, Senior Member, IEEE, Ran Liu, Mulin Chen<sup>1</sup>, and Xuelong Li<sup>1</sup>, Fellow, IEEE

**Abstract**—Graph-based clustering aims to partition the data according to a similarity graph, which has shown impressive performance on various kinds of tasks. The quality of similarity graph largely determines the clustering results, but it is difficult to produce a high-quality one, especially when data contain noises and outliers. To solve this problem, we propose a robust rank constrained sparse learning (RRCSL) method in this article. The  $L_{2,1}$ -norm is adopted into the objective function of sparse representation to learn the optimal graph with robustness. To preserve the data structure, we construct an initial graph and search the graph within its neighborhood. By incorporating a rank constraint, the learned graph can be directly used as the cluster indicator, and the final results are obtained without additional postprocessing. In addition, the proposed method cannot only be applied to single-view clustering but also extended to multiview clustering. Plenty of experiments on synthetic and real-world datasets have demonstrated the superiority and robustness of the proposed framework.

**Index Terms**—Graph clustering, graph learning, machine learning, multiview clustering, sparse representation.

## I. INTRODUCTION

CLUSTERING is a fundamental technique for processing unlabeled data, which aims to partition the samples into clusters. In this age of information, a large number of unlabeled data show exponential growth [1]. Therefore, clustering has become an active research area, and has been widely used in various applications, such as gene expression analysis, motion segmentation, document clustering, social media analysis, and image segmentation [2], [3]. In the past few decades, plenty of methods have been proposed toward this topic [4], such as  $K$ -means, hierarchical clustering, subspace

clustering, spectral clustering, and graph-based clustering [5]–[8]. Some clustering methods based on deep learning also obtain extensive attention [9], [10].

Among numerous clustering techniques, graph-based clustering methods focus on the internal data structure, and have shown better performance [11], [12]. Most of them construct the similarity graph as the first step, and fix it during the clustering stage. Therefore, the clustering performance relies highly on the graph construction procedure. However, it is difficult to guarantee the graph quality. To alleviate this problem, some methods update the graph during the clustering procedure, such as clustering with adaptive neighbors (CANs) [13], constrained Laplacian rank (CLR) [14], and simplex sparse representation (SSR) [15]. However, they are susceptible to noises and outliers. Moreover, most of the existing works cannot obtain the clustering indicator intuitively, so they use  $K$ -means or spectral clustering as the postprocessing, which leads to the suboptimal result [16].

In addition, multiview clustering is a necessary and important issue, since data are usually represented by different views [17]–[19]. A variety of methods is proposed to fuse the complementary information hidden in the views [20]–[25]. An intuitive way to make use of the information from each view is combining all the feature vectors together. This strategy ignores the diversity of different views and is often affected severely when some views are confusing. Recently, some methods try to learn the weight of each view automatically [26]–[28]. But they treat the optimal graph as a linear combination of graphs with different views, which restricts the flexibility of the desired graph. Meanwhile, postprocessing is necessary in most methods [29], [30], which degrades clustering performance.

In this article, a robust rank constrained sparse learning (RRCSL) method is proposed to solve the above-mentioned problems. We combine the sparse representation with the  $L_{2,1}$ -norm to learn the desired graph, which also reduces the impact of data noises and outliers. Furthermore, in order to preserve the data relationship accurately, the desired graph is searched within the initial graph's neighborhood. To avoid the shortcomings caused by two separate steps in most existing works, we introduce the Laplacian rank constraint, which makes the learned graph have a clear structure. In this way, data can be classified directly and, therefore, the quality and efficiency of clustering have been improved. In addition, we extend the RRCSL method to multiview clustering and make

Manuscript received August 12, 2020; revised November 25, 2020 and February 19, 2021; accepted March 13, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1107403 and in part by the National Natural Science Foundation of China under Grant U1864204, Grant 61773316, Grant U1801262, and Grant 61871470. This article was recommended by Associate Editor N. Zhang. (Corresponding authors: Xuelong Li; Mulin Chen).

The authors are with the School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China, and also with the School of Artificial Intelligence, Optics and Electronics, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: crabwq@gmail.com; liuran\_npu@foxmail.com; chenmulin@mail.nwpu.edu.cn; xuelong\_li@nwpu.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCYB.2021.3067137>.

Digital Object Identifier 10.1109/TCYB.2021.3067137

the weights of views adjusted adaptively. Instead of using the linear combination of the graphs directly, the unified graph is searched within its neighborhood such that the searching space is extended.

In summary, the major contributions of this article are as follows.

- 1) The sparse representation is combined with  $L_{2,1}$ -norm, which upgrades the construction quality of the graph while reducing the impact of data noises and outliers.
- 2) The optimal similarity graph is searched within the neighborhood of a predefined graph, which ensures that the similarity graph reflects the accurate relationship between the data.
- 3) The rank constraint is added into the proposed method and makes the learned graph have a clear structure, which avoids additional postprocessing.
- 4) The proposed RRCSL method is extended to multi-view clustering and the weights of views are adjusted adaptively, while the flexibility of the optimal graph is improved.

## II. SINGLE-VIEW CLUSTERING BY RRCSL METHOD

In this section, the sparse representation method is first introduced as the preliminary knowledge. Then, the RRCSL method and the corresponding optimization algorithm are proposed.

### A. Sparse Representation Revisited

There are many ways to construct a similarity graph by given data, such as the k-nearest neighbor method (kNN) and the Gaussian kernel function method. But the hyperparameter in these methods is a deficiency because of its difficulty to tune. The sparse representation can be applied to build a data similarity graph without specifying parameters and improve the quality of the graph. Suppose  $X = [x_1, \dots, x_n] \in \mathbb{R}^{d \times n}$  is the data matrix with  $d$  features and  $n$  data points.  $X$  is dense with lots of irrelevant and redundant data. In most cases, it is necessary to convert  $X$  into an appropriate sparse matrix. Suppose the new representation of  $X$  is  $y \in \mathbb{R}^{d \times 1}$ , and the sparse representation method computes a representation vector  $\alpha \in \mathbb{R}^{n \times 1}$  to satisfy  $y \approx X\alpha$ . To seek a sparse solution, it is natural to solve the following problem:

$$\min_{\alpha} \|y - X\alpha\|_2^2 + \omega_0 \|\alpha\|_1. \quad (1)$$

Based on (1), Huang *et al.* [15] proposed the SSR method. According to SSR,  $\alpha_i \in \mathbb{R}^{(n-1) \times 1}$  is the similarity between the  $i$ th sample and other samples, which can be represented by the following objective function:

$$\min_{\alpha_i \geq 0} \|x_i - X_{-i}\alpha_i\|_2^2 + \omega_0 \|\alpha_i\|_1 \quad (2)$$

where  $X_{-i} = [x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n] \in \mathbb{R}^{d \times (n-1)}$  is the data matrix that removes column  $i$ . To obtain the shift invariant similarities, it is necessary to constrain  $\alpha_i^T \mathbf{1} = 1$ . In this way,  $\|\alpha_i\|_1$  becomes constant, so (2) can be rewritten as

$$\min_{\alpha_i \geq 0, \alpha_i^T \mathbf{1} = 1} \|x_i - X_{-i}\alpha_i\|_2^2. \quad (3)$$

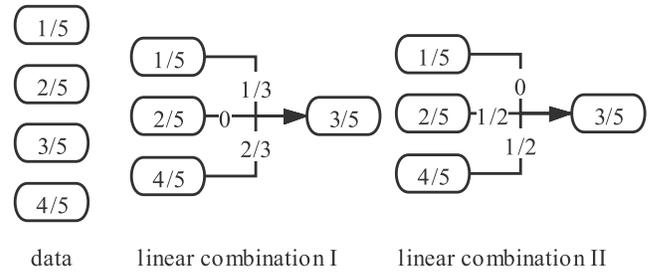


Fig. 1. Example of multiple linear combinations.

Equation (3) is the objective function of the SSR method, and it assumes that a data point can be approximated by the linear combination of the other points such that a sparse data similarity graph can be learned without prior analysis or parameter. However, since (3) squares the residue errors, SSR is sensitive to outliers and noise. Furthermore, there may be multiple linear combinations of  $\alpha_i$  for the same set of  $x_i$  and  $X_{-i}$ . To demonstrate this statement, a simple example is shown in Fig. 1. When  $(3/5)$  is reconstructed from  $(1/5)$ ,  $(2/5)$ , and  $(4/5)$ , there are two completely different linear combinations of the coefficients. For the same sample, minimizing (3) leads to multiple solutions. Therefore, the similarity graph may deviate from the relationship of samples, which further produces incorrect clustering results.

### B. Robust Rank-Constrained Sparse Learning Method

In order to facilitate subsequent optimization, we rewrite (3) into the matrix form

$$\begin{aligned} \min_S \|X - XS\|_F^2 \\ \text{s.t. } s_{ij} \geq 0, \sum_j s_{ij} = 1, \text{diag}(S) = 0 \end{aligned} \quad (4)$$

where  $S \in \mathbb{R}^{n \times n}$  is the desired graph. To solve the problem of sensitivity to outliers and noises, we introduce the  $L_{2,1}$ -norm to replace the square of  $F$ -norm. The  $L_{2,1}$ -norm shows good performance in fields of feature selection and outlier detection [31]–[33]. The objective function becomes

$$\begin{aligned} \min_S \|X - XS\|_{2,1} \\ \text{s.t. } s_{ij} \geq 0, \sum_j s_{ij} = 1, \text{diag}(S) = 0. \end{aligned} \quad (5)$$

The  $L_{2,1}$ -norm of the matrix  $M$  is defined as

$$\|M\|_{2,1} = \sum_i \sqrt{\sum_j |m_{ij}|^2} = \sum_i \|m_j\|_2. \quad (6)$$

Rather than pursuing the flat sparsity, the  $L_{2,1}$ -norm ensures the structural sparsity within the learned graph. In addition, for any rotation matrix  $R$ ,  $\|MR\|_{2,1} = \|M\|_{2,1}$  always holds, which makes the model insensitive to feature rotation.

However, in (5), there are still multiple linear combinations of  $S$  for the same  $X$ . To make the learned graph  $S$  reflect the relationship of  $X$ , we propose to construct an initial graph

TABLE I  
ALGORITHM OF ALM METHOD TO SOLVE  $\min_{h(X)=0} f(X)$

ALGORITHM 1: The Algorithm of ALM method.
Set $1 < \rho < 2$ . Initialize $\mu > 0, \Lambda$
<b>While</b> not converge <b>do</b>
1. Update $X$ by $\min_X f(X) + \frac{\mu}{2} \left\  h(X) + \frac{1}{\mu} \Lambda \right\ _F^2$ ;
2. Update $\Lambda$ by $\Lambda = \Lambda + \mu h(X)$ ;
3. Update $\mu$ by $\mu = \rho \mu$ ;
<b>End While</b>

and find the desired graph within its neighborhood. Thus, the objective function turns out to be

$$\begin{aligned} \min_S \quad & \|X - XS\|_{2,1} + \alpha \|S - B\|_F^2 \\ \text{s.t.} \quad & s_{ij} \geq 0, \sum_j s_{ij} = 1, \text{diag}(S) = 0 \end{aligned} \quad (7)$$

where  $B$  is the initial similarity graph, which can be generated by Nie *et al.* [14]. Although this simple initial graph has low quality and poor clustering results will be obtained if it is directly used for clustering. But to some extent, it can reflect accurate relationships between the data. Equation (7) can make the learned optimal similarity graph close to the initial similarity graph. In this way, the similarity graph will only be the one that fits the true relationship of the data in multiple linear combinations, which ensures it well represented. It also makes up for the shortcomings of sparse representation learning that cannot grasp the local relationship.

In order to obtain the cluster indicator directly, we propose to impose an additional constraint into the objective function, which requires the following theorem.

*Theorem 1* [34]: The multiplicity  $k$  of the eigenvalue zero of the Laplacian matrix  $L_s$  is equal to the number of connected components in the graph associated with  $S$ .

The Laplacian matrix is  $L_s = D_s - (S^T + S)/2$ , where  $D_s$  is the diagonal matrix and the  $i$ th diagonal element is  $\sum_j (s_{ij} + s_{ji})/2$ . Given the cluster number  $k$ , we can constrain the rank of  $L_s$  to be  $n - k$ , so that the learned graph contains exact  $k$  connected components. In this way, the clustering result can be directly obtained according to the connection relationship of the data points in the similarity graph [14]. Then, (7) becomes

$$\begin{aligned} \min_S \quad & \|X - XS\|_{2,1} + \alpha \|S - B\|_F^2 \\ \text{s.t.} \quad & s_{ij} \geq 0, \sum_j s_{ij} = 1, \text{diag}(S) = 0, \text{rank}(L_s) = n - k. \end{aligned} \quad (8)$$

Besides, the constraint  $\text{rank}(L_s) = n - k$  is a complex non-linear constraint. Equation (8) is difficult to solve. But it can be equivalently converted into an easy-to-solve form by an effective method.  $L_s$  is known to be semidefinite, assuming  $\sigma_i(L_s)$  is its  $i$ th smallest eigenvalue, and then  $\sigma_i(L_s)$  is non-negative. When  $\lambda$  is large enough, (8) is equivalent to

$$\begin{aligned} \min_S \quad & \|X - XS\|_{2,1} + \alpha \|S - B\|_F^2 + 2\lambda \sum_{i=1}^k \sigma_i(L_s) \\ \text{s.t.} \quad & s_{ij} \geq 0, \sum_j s_{ij} = 1, \text{diag}(S) = 0. \end{aligned} \quad (9)$$

With the method, the rank constraint is incorporated into (9) as a regularizer. According to Fan [35]

$$\begin{aligned} \sum_{i=1}^k \sigma_i(L_s) = \min_F \text{Tr}(F^T L_s F) \\ \text{s.t.} \quad F \in \mathbb{R}^{n \times k}, F^T F = I \end{aligned} \quad (10)$$

where  $F \in \mathbb{R}^{n \times k}$  is an indicator matrix. Therefore, (9) is equivalent to

$$\begin{aligned} \min_{S,F} \quad & \|X - XS\|_{2,1} + \alpha \|S - B\|_F^2 + \lambda \text{Tr}(F^T L_s F) \\ \text{s.t.} \quad & s_{ij} \geq 0, \sum_j s_{ij} = 1, \text{diag}(S) = 0, F^T F = I. \end{aligned} \quad (11)$$

The above formula is the objective equation of the RRCSL method for single-view clustering. By solving (11), we obtain a robust and sparse data similarity graph with the clear cluster structure. In the following part, we will present an efficient optimization algorithm to solve (11).

### C. Optimization Algorithm

Since  $\|X - XS\|_{2,1}$ ,  $\|S - B\|_F^2$ , and  $L_s$  all depend on  $S$ , (8) is a complex optimization problem. An efficient approach, augmented Lagrangian multiplier (ALM) [36], can be used for tackling (11). Consider the constrained optimization problem  $\min_{h(X)=0} f(X)$ , the algorithm using the ALM method to solve this problem is described in Table I. It has been demonstrated that under some rather general conditions, Algorithm 1 converges linearly to the optimal solution [37]. This property makes the ALM method very attractive.

Let  $E = X - XZ$ ,  $Z = S$ , and (11) is transformed into the following ALM problem:

$$\begin{aligned} \min_S \quad & \|E\|_{2,1} + \alpha \|Z - B\|_F^2 + \lambda \text{Tr}(F^T L_s F) \\ & + \frac{\mu}{2} \left\| E - X + XZ + \frac{\Lambda_1}{\mu} \right\|_F^2 + \frac{\mu}{2} \left\| Z - S + \frac{\Lambda_2}{\mu} \right\|_F^2 \\ \text{s.t.} \quad & s_{ij} \geq 0, \sum_j s_{ij} = 1, \text{diag}(S) = 0, F^T F = I \end{aligned} \quad (12)$$

where  $\mu \in \mathbb{R}^{1 \times 1}$  is a regularity coefficient, and  $\Lambda_1 \in \mathbb{R}^{d \times n}$  and  $\Lambda_2 \in \mathbb{R}^{d \times n}$  are penalty parameters.

*Update E:* When updating  $E$ , we fix  $Z$ ,  $S$ , and  $F$ . Denoting  $C = X - XZ - (\Lambda_1/\mu)$ , thus (12) becomes

$$\min_E \quad \|E\|_{2,1} + \frac{\mu}{2} \|E - C\|_F^2. \quad (13)$$

The following Lemma is useful for solving this problem [38].

*Lemma 1* [39]: Given a matrix  $W = [W_1, \dots, W_n] \in \mathbb{R}^{m \times n}$  and a positive scalar  $\delta$ , then  $X^*$  is the optimal solution of

$$\min \delta \|X\|_{2,1} + \frac{1}{2} \|X - W\|_F^2 \quad (14)$$

and the  $i$ th column of  $X^*$  is

$$X^*(:, i) = \begin{cases} \frac{\|W_i\|_2 - \delta}{\|W_i\|} W_i, & \text{if } \|W_i\|_2 > \delta \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

Let  $\delta = (1/\mu)$ , and the solution of  $E$  is

$$E^*(:, i) = \begin{cases} \left(1 - \frac{1}{\mu \|C(:, i)\|}\right) C(:, i), & \text{if } \|C(:, i)\|_2 > \frac{1}{\mu} \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

where  $E(:, i)$  and  $C(:, i)$  are the  $i$ th columns of  $E$  and  $C$ , respectively.

*Update Z:* When updating  $Z$ , we fix  $E$ ,  $S$ , and  $F$ . Thus, (12) becomes

$$\min_Z \alpha \|Z - B\|_F^2 + \frac{\mu}{2} \left\| E - X + XZ + \frac{\Lambda_1}{\mu} \right\|_F^2 + \frac{\mu}{2} \left\| Z - S + \frac{\Lambda_2}{\mu} \right\|_F^2. \quad (17)$$

Taking the derivation of (17) with respect to  $Z$

$$\min_Z 2\alpha \text{Tr}(Z - B) + \mu \text{Tr} \left( Z - S + \frac{\Lambda_2}{\mu} \right) + \frac{\mu}{2} \text{Tr} \left[ 2X^T XZ + 2X^T \left( E - X + \frac{\Lambda_1}{\mu} \right) \right]. \quad (18)$$

Let (18) be zero and we obtain the optimal solution for  $Z$  as follows:

$$Z = \frac{\left[ \frac{2\alpha}{\mu} B - 2X^T \left( E - X + \frac{\Lambda_1}{\mu} \right) + S - \frac{\Lambda_2}{\mu} \right]}{\left( \frac{2\alpha}{\mu} I + X^T X + I \right)}. \quad (19)$$

*Update S:* When updating  $S$ , we fix  $E$ ,  $Z$ , and  $F$ . Thus, (12) becomes

$$\min_S \lambda \text{Tr}(F^T L_S F) + \frac{\mu}{2} \left\| Z - S + \frac{\Lambda_2}{\mu} \right\|_F^2. \quad (20)$$

Let  $s_{ij}$  denote the element in  $S$ . According to the properties of the Laplace matrix, we have the following equation for any  $F$ :

$$\text{Tr}(F^T L_S F) = \frac{1}{2} \sum_{ij} \left( \|F(i, :) - F(j, :)\|_2^2 s_{ij} \right). \quad (21)$$

Denoting  $H = Z + (\Lambda_2/\mu)$  and  $d_{ij} = \sum_{ij} \|F(i, :) - F(j, :)\|_2^2$ , (20) becomes

$$\min_S \frac{\lambda}{2} \sum_{ij} d_{ij} s_{ij} + \frac{\mu}{2} \|H - S\|_F^2. \quad (22)$$

Let  $h_{ij}$  denote the element in  $H$ . Equation (22) becomes

$$\min_S \frac{\lambda}{2} \sum_{ij} d_{ij} s_{ij} + \frac{\mu}{2} \sum_{ij} (h_{ij} - s_{ij})^2. \quad (23)$$

Note that the solution to (23) is independent for each row, so we can optimize each row separately

$$\min_S \frac{\lambda}{2} \sum_j d_{ij} s_{ij} + \frac{\mu}{2} \sum_j (h_{ij} - s_{ij})^2. \quad (24)$$

After expanding (24), we obtain

$$\min_S \sum_j \left( s_{ij}^2 - 2 \left( h_{ij} - \frac{\lambda}{2\mu} d_{ij} \right) s_{ij} + \frac{\mu}{2} h_{ij}^2 \right). \quad (25)$$

For each row, (25) is equivalent to the following form further:

$$\min_S \left\| s_j - \left( h_j - \frac{\lambda}{2\mu} d_j \right) \right\|_2^2 \quad (26)$$

which can be solved with an efficient iterative algorithm [15].

TABLE II

ALGORITHM OF SINGLE-VIEW CLUSTERING BY RRCSL METHOD

**ALGORITHM 2:** Single-view Clustering by RRCSL Method**Input:**  $X \in \mathbb{R}^{d \times n}$ , the number of clusters  $k$ , parameter  $\alpha > 0$ **Output:** the similarity graph  $S$  with  $k$  connected components**While** not converge **do**1. update  $E$  by Eq.(16);2. update  $Z$  by Eq.(19);3. update  $S$  by solving Eq.(26);4. update  $F$  by calculating  $k$  eigenvectors corresponding to the first  $k$  minimum eigenvalues of  $L_S$ ;

5. update the ALM parameters.

**End While**

*Update F:* When updating  $F$ , we fix  $E$ ,  $Z$ , and  $S$ . Thus, (12) becomes

$$\min_F \text{Tr}(F^T L_S F) \quad \text{s.t. } F \in \mathbb{R}^{n \times k}, F^T F = I. \quad (27)$$

The number of clusters is  $k$ . Calculating the first  $k$  minimum eigenvalues of  $L_S$  and the optimal solution of  $F$  consists of  $k$  eigenvectors corresponding to these eigenvalues.

*Update  $\Lambda_1$ ,  $\Lambda_2$ , and  $\mu$ :* In each iteration, the ALM parameters are updated as follows [38]:

$$\begin{aligned} \Lambda_1 &= \Lambda_1 + \mu(E - X + XZ) \\ \Lambda_2 &= \Lambda_2 + \mu(Z - S) \\ \mu &= \rho \mu \end{aligned} \quad (28)$$

where  $\rho > 1$  is the update rate, and a larger  $\rho$  brings faster convergence speed, but accompanied with poorer results. The algorithm of single-view clustering by RRCSL method is shown in Table II.

### III. MULTIVIEW CLUSTERING BY RRCSL METHOD

In this section, the RRCSL method is extended to multiview clustering. We will show the differences in model and the optimization algorithm separately from the previous section.

#### A. Multiview Clustering Formulation

Supposing the data contain  $n_v$  views, the data corresponding to the  $n_v$  views can be written as  $X^{(1)}, X^{(2)}, \dots, X^{(n_v)} \in \mathbb{R}^{d \times n}$ , where each column corresponds to one sample, and each row corresponds to one feature in each view. Denoting that scalar variable  $w^{(1)}, w^{(2)}, \dots, w^{(n_v)} \geq 0$  are the weights of  $X^{(1)}, X^{(2)}, \dots, X^{(n_v)}$ , respectively, and the optimization problem becomes the following formula on the basis of (11):

$$\begin{aligned} \min_S \left\| \sum_{v=1}^{n_v} w^{(v)} X^{(v)} - \sum_{v=1}^{n_v} w^{(v)} X^{(v)} S \right\|_{2,1}^2 \\ + \alpha \left\| S - \sum_{v=1}^{n_v} w^{(v)} B^{(v)} \right\|_F^2 + \lambda \text{Tr}(F^T L_S F) \\ \text{s.t. } s_{ij} \geq 0, \sum_j s_{ij} = 1, \text{diag}(S) = 0 \\ F^T F = I, \sum_v w^{(v)} = 1, w > 0 \end{aligned} \quad (29)$$

where  $B^{(v)}$  is the initial graph of the  $v$ th view.

Compared with (11), (29) deals with the multiview data, and it learns the weight of each view automatically. Some methods directly average the weights of the various views [28]. This strategy is straightforward, but it neglects the different roles of the views. In order to solve this problem, a method is proposed to assign the weight of the view automatically [40]. This way is embedded in the optimization process of the RRCSL method. Based on the algorithm of single-view clustering in the previous section, we realized the optimal distribution of weights of various views by adding an iterative link, which is shown in Table III.

### B. Optimization Algorithm

Let  $E = \sum_{v=1}^{n_v} w^{(v)} X^{(v)} - \sum_{v=1}^{n_v} w^{(v)} X^{(v)} Z$ ,  $Z = S$ , then the ALM method is applied to construct the following objective equation based on (29):

$$\begin{aligned} \min_S \quad & \|E\|_{2,1} + \alpha \left\| Z - \sum_{v=1}^{n_v} w^{(v)} B^{(v)} \right\|_F^2 + \lambda \text{Tr}(F^T L_s F) \\ & + \frac{\mu}{2} \left\| E - \sum_{v=1}^{n_v} w^{(v)} X^{(v)} + \sum_{v=1}^{n_v} w^{(v)} X^{(v)} Z + \frac{\Lambda_1}{\mu} \right\|_F^2 \\ & + \frac{\mu}{2} \left\| Z - S + \frac{\Lambda_2}{\mu} \right\|_F^2 \\ \text{s.t.} \quad & s_{ij} \geq 0, \sum_j s_{ij} = 1, \text{Tr}(S) = 0 \\ & F^T F = I, \sum_v w^{(v)} = 1, w > 0. \end{aligned} \quad (30)$$

*Update E, Z, S, and F:* When updating  $E$ ,  $Z$ ,  $S$ , or  $F$ ,  $w^{(v)}$  is fixed. Equation (30) is equivalent to the optimization problem for single-view clustering, just as (13)–(27).

*Update  $w^{(v)}$ :* When updating  $w^{(v)}$ , we fix  $E$ ,  $Z$ ,  $F$ , and  $S$ ; thus, (30) becomes

$$\begin{aligned} \min_{w^{(v)}} \quad & \alpha \left\| Z - \sum_{v=1}^{n_v} w^{(v)} B^{(v)} \right\|_F^2 \\ & + \frac{\mu}{2} \left\| E - \sum_{v=1}^{n_v} w^{(v)} X^{(v)} + \sum_{v=1}^{n_v} w^{(v)} X^{(v)} Z + \frac{\Lambda_1}{\mu} \right\|_F^2 \\ \text{s.t.} \quad & \sum_v w^{(v)} = 1, w > 0. \end{aligned} \quad (31)$$

To solve this difficult optimization problem, we convert it from matrix form into the vector form [40]. We convert  $E \in \mathbb{R}^{d \times n}$  and  $Z \in \mathbb{R}^{n \times n}$  into two column vectors  $\hat{e} \in \mathbb{R}^{dn \times 1}$  and  $\hat{z} \in \mathbb{R}^{n^2 \times 1}$ , respectively. Meanwhile, we combine  $[w^{(1)}, w^{(2)}, \dots, w^{(n_v)}]$  into a column vector  $\hat{w} \in \mathbb{R}^{n_v \times 1}$ .  $X^{(1)}, X^{(2)}, \dots, X^{(n_v)} \in \mathbb{R}^{d \times n}$  and  $B^{(1)}, B^{(2)}, \dots, B^{(n_v)} \in \mathbb{R}^{n \times n}$  are also converted into two matrixes  $\hat{X} \in \mathbb{R}^{dn \times n_v}$  and  $\hat{B} \in \mathbb{R}^{n^2 \times n_v}$ . Then, (31) becomes

$$\min_{\hat{w}} \alpha \left\| \hat{z} - \hat{B} \hat{w} \right\|_2^2 + \frac{\mu}{2} \left\| \hat{e} - \hat{X} \hat{w} + \hat{X} \hat{z} \hat{w} + \frac{\Lambda_1}{\mu} \right\|_2^2. \quad (32)$$

TABLE III  
ALGORITHM OF MULTIVIEW CLUSTERING BY RRCSL METHOD

ALGORITHM 3: Multi-view Clustering by RRCSL Method	
<b>Input:</b>	$X^{(n_v)} \in \mathbb{R}^{d \times n}$ , the number of clusters $k$ , the controlling parameter $\alpha > 0$
<b>Output:</b>	the similarity matrix $S$ with $k$ connected components
<b>While</b> not converge <b>do</b>	
	1. update $E$ by Eq.(16);
	2. update $Z$ by Eq.(19);
	3. update $S$ by solving Eq.(26);
	4. update $F$ by calculating $k$ eigenvectors corresponding to the first $k$ minimum eigenvalues of $L_s$ ;
	5. update $w^{(v)}$ by solving Eq.(37);
	6. update the parameters of the ALM method.
<b>End While</b>	

In order to be more concise in form, we let  $\hat{\mathbf{I}} = \hat{\mathbf{e}} + (\Lambda_1/\mu)$  and  $\hat{R} = \hat{X} - \hat{X}\hat{z}$ . Then, (32) becomes

$$\min_{\hat{w}} \alpha \left\| \hat{z} - \hat{B} \hat{w} \right\|_2^2 + \frac{\mu}{2} \left\| \hat{\mathbf{I}} - \hat{R} \hat{w} \right\|_2^2. \quad (33)$$

Spreading the first term, we have

$$\begin{aligned} \min_{\hat{w}} \quad & \alpha \left\| \hat{z} - \hat{B} \hat{w} \right\|_2^2 \\ = \min_{\hat{w}} \quad & \alpha \left( \hat{z}^T \hat{z} - \hat{w}^T \hat{B}^T \hat{z} - \hat{z}^T \hat{B} \hat{w} + \hat{w}^T \hat{B}^T \hat{B} \hat{w} \right) \end{aligned} \quad (34)$$

which is transformed into

$$\min_{\hat{w}} \alpha \left( \hat{w}^T \hat{B}^T \hat{B} \hat{w} - 2 \hat{z}^T \hat{B} \hat{w} \right). \quad (35)$$

Similarly, the second term becomes

$$\min_{\hat{w}} \frac{\mu}{2} \left( \hat{w}^T \hat{R}^T \hat{R} \hat{w} - 2 \hat{\mathbf{I}}^T \hat{R} \hat{w} \right). \quad (36)$$

Therefore, (33) finally becomes

$$\begin{aligned} \min_{\hat{w}} \quad & \hat{w}^T \left( \alpha \hat{B}^T \hat{B} + \frac{\mu}{2} \hat{R}^T \hat{R} \right) \hat{w} - \left( 2 \alpha \hat{z}^T \hat{B} + \mu \hat{\mathbf{I}}^T \hat{R} \right) \hat{w} \\ \text{s.t.} \quad & \sum_v w^{(v)} = 1, w > 0. \end{aligned} \quad (37)$$

So far, we have converted (31) into the standard quadratic programming (QP) problem. There are various existing algorithms that can solve this problem effectively [15], such as the Lagrangian method. In this way, the weight of each view is learned adaptively.

## IV. EXPERIMENTS OF SINGLE-VIEW CLUSTERING

The RRCSL method is evaluated on nine commonly used real-world benchmark datasets. The clustering experimental results will be compared with seven state-of-the-art clustering methods. In addition, the robustness of the RRCSL method is also demonstrated.

### A. Dataset Descriptions

Nine commonly used real-world benchmark datasets are employed to demonstrate the effectiveness of the proposed RRCSL, including SRBCT [41], iris [42], Yale [43], wine [44], glass [45], umist [46], coil20 [47], yeast [48], and Semeion [49].

SRBCT contains 83 sets of data, with 2308 features in each set. This is a bioinformatics dataset consisting of four classes,

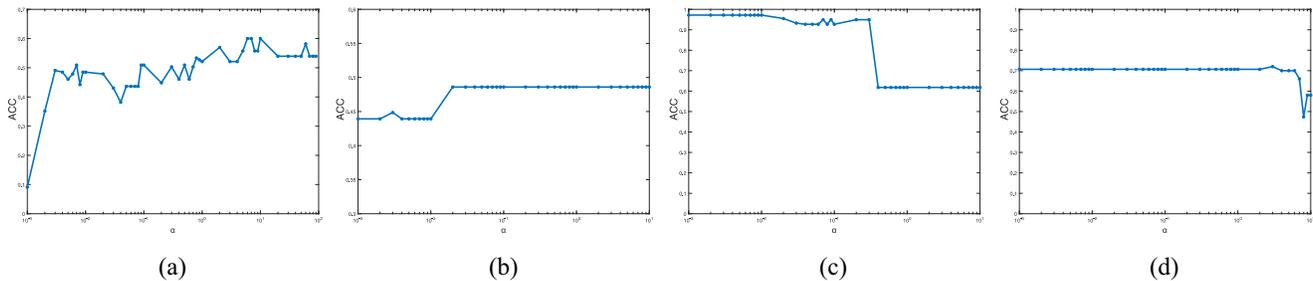


Fig. 2. Effect of parameter on the datasets. (a) YALE. (b) GLASS. (c) WINE. (d) IRIS.

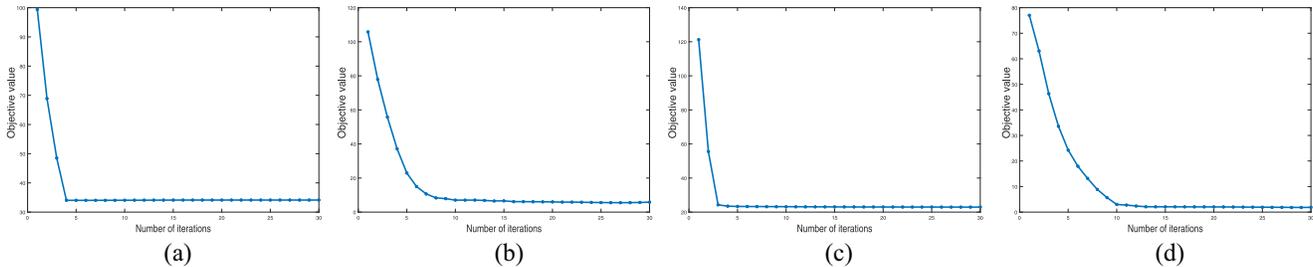


Fig. 3. Convergence analysis on datasets. (a) YALE. (b) GLASS. (c) WINE. (d) IRIS.

TABLE IV  
SINGLE-VIEW CLUSTERING RESULT IN TERMS OF ACCURACY

	Method	SRBCT	Iris	Yale	Wine	Glass	UMIST	Coil20	Yeast	Semeion
ACC	K-means	0.424	0.797	0.472	0.960	0.482	0.431	0.550	0.401	0.572
	NMF	0.435	0.636	0.339	0.722	0.444	0.340	0.475	0.136	0.375
	NCut	0.430	0.586	0.511	0.959	0.365	0.439	0.493	0.352	0.513
	CAN	0.506	0.573	0.521	0.926	0.457	0.671	0.834	0.409	0.601
	CLR	0.397	0.846	0.509	0.915	0.411	0.610	0.821	0.498	0.414
	SSR	0.421	0.686	0.593	0.955	0.383	0.563	0.692	0.418	0.615
	KMM	0.425	0.817	0.442	0.541	0.463	0.382	0.433	0.427	0.589
	RRCSL	<b>0.590</b>	<b>0.873</b>	<b>0.600</b>	<b>0.972</b>	<b>0.486</b>	<b>0.807</b>	<b>0.878</b>	<b>0.499</b>	<b>0.633</b>

corresponding to four subtypes of the small round blue cell tumor.

*Iris* contains 150 sets of data from three classes, and four features (sepal length, sepal length, petal length, and petal width) are used for clustering.

The *Yale* dataset contains 165 face images from 15 classes. The resolution of each image is 16 by 16.

The *Wine* dataset contains 178 sets of data from 3 classes, and 13 features are used for clustering.

The *Glass Identification* dataset contains 214 sets of data from seven classes, and nine features are used for clustering.

*UMIST* is another face dataset from 20 classes, containing 575 face images.

The *Coil20* dataset contains 1440 images from 20 classes. These 20 classes correspond to 20 objects, and the resolution of each image is 32 by 32.

The *Yeast* dataset contains 1484 sets of data from ten classes, and 1470 features are used for clustering.

The *Semeion* handwritten digit dataset contains 1593 images from ten classes. The resolution of each image is 16 by 16.

## B. Experiments Setup

We compared our clustering methods with seven clustering methods, including *K*-means, NMF [50] methods, normalized

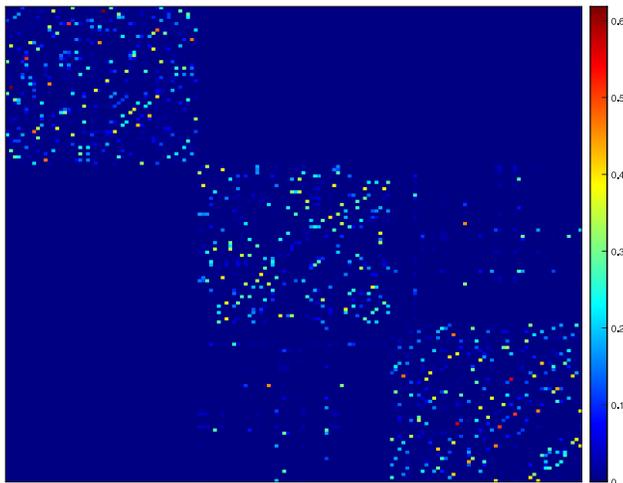
cut (NCut) [51], CAN [13], CLR [14], SSR [15], and *K*-multiple-means (KMMs) [52]. The comparison methods are all set by default. As for our clustering method, the initial similarity graph is generated by Nie *et al.* [14]. A heuristic way is employed to determine the value of  $\lambda$ : at the beginning, let  $\lambda$  be a small value like  $1e-4$ ; then, we calculate whether the number of zero eigenvalues of  $L_S$  is equal to  $k$  in each iteration. The controlling parameter  $\alpha$  is set as 0.01. Because the learned graph  $S$  has a clear structure, clustering results can be obtained by directly dividing  $S$  into  $k$  connected components and each of these components corresponds to a class.

## C. Performance

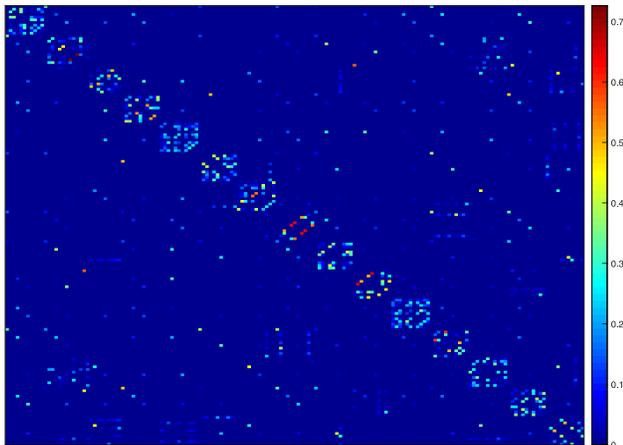
After the same normalization processing of the data from the nine datasets, the RRCSL method and seven comparison algorithms are applied, respectively, to carry out clustering experiments. Two widely used clustering performance measures are adopted to evaluate the clustering results, namely, accuracy (ACC) [53] and normalized mutual information (NMI) [15], [54]. In experiments, the greater the value of ACC and NMI, the better the clustering result. The performances are shown in Tables IV and V (bold indicates the best). *K*-means, NMF, and NCut are widely used clustering methods, and KMM is a novel clustering algorithm improved from *K*-means. Compared with them, our method has significant

TABLE V  
SINGLE-VIEW CLUSTERING RESULT IN TERMS OF NMI

	Method	SRBCT	Iris	Yale	Wine	Glass	UMIST	Coil20	Yeast	Semeion
NMI	K-means	0.140	0.637	0.540	0.862	0.337	0.637	0.722	0.269	0.522
	NMF	0.171	0.404	0.412	0.447	0.306	0.489	0.602	0.017	0.320
	NCut	0.159	0.572	0.561	<b>0.881</b>	0.276	0.625	0.661	0.236	0.483
	CAN	0.271	0.427	0.549	0.780	0.289	0.816	0.916	0.158	0.613
	CLR	0.180	0.723	0.582	0.732	0.284	0.792	0.905	0.296	0.407
	SSR	0.165	0.580	0.584	0.848	0.306	0.714	0.805	0.238	<b>0.617</b>
	KMM	0.243	0.682	0.506	0.361	0.324	0.594	0.641	0.267	0.533
	RRCSL	<b>0.408</b>	<b>0.754</b>	<b>0.631</b>	<b>0.881</b>	<b>0.344</b>	<b>0.901</b>	<b>0.932</b>	<b>0.308</b>	<b>0.617</b>



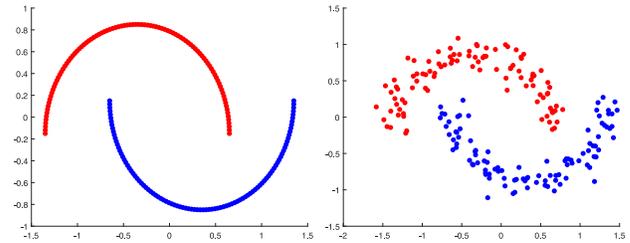
(a)



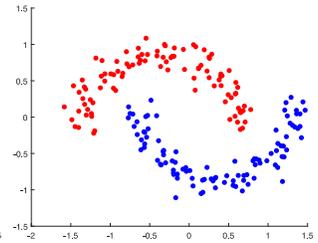
(b)

Fig. 4. Visualization results of (a)  $S$  for Iris dataset and (b)  $S$  for Yale dataset.

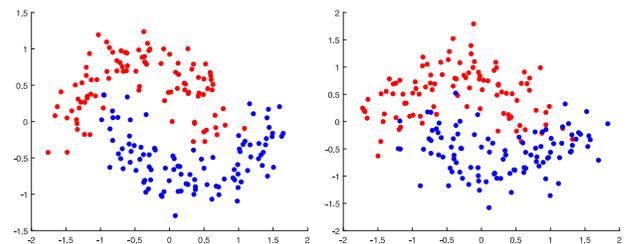
performance advantages. As representative graph-based clustering methods, CAN, CLR, and SSR achieved good results in clustering experiments. But we can see that RRCSL outperforms these methods in all the experiments. This is because the RRCSL method has stronger ability on data representation compared with CAN and CLR, and it avoids the problem of SSR that the similarity graph is not unique. While improving the quality of similarity graph, this method is more robust than others. The results of the experiments demonstrate the superiority of the proposed approach. Using the Iris dataset and



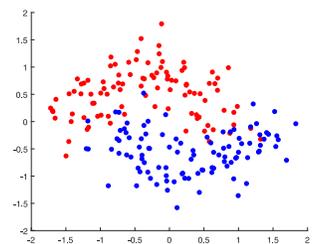
(a)



(b)



(c)



(d)

Fig. 5. Two-Moon synthetic datasets. (a) Original two-moon. (b) Noise percentage is 0.1. (c) Noise percentage is 0.2. (d) Noise percentage is 0.3.

Yale dataset as examples, we show the visualization results of data similarity graph  $S$ , as shown in Fig. 4.

We also investigate the parameter sensitivity of the proposed method. Taking the four datasets of Yale, Glass, Yeast, and Wine as examples, we let the variable parameter  $\alpha$  fluctuate within a large range of 0.001–10 and calculate the ACC as results. It can be observed that the performance of the RRCSL method is very stable as shown in Fig. 2. Meanwhile, the convergence curves of the proposed algorithm are shown in Fig. 3.

#### D. Robustness to Noises and Outliers

The robustness of RRCSL is verified by experiments on “Two-Moon” synthetic dataset and Yale face dataset. For Two-Moon dataset, there are two clusters of data distributed in the moon shape. Each cluster has a volume of 100 samples and the noise percentage is set to be from 0.1 to 0.3 as shown in Fig. 5. In this figure, we set the color of the two clusters to be red and blue, respectively. Table VI shows the clustering results, and it can be seen that the RRCSL method is less affected by noise. It is particularly noteworthy that our method achieves 100% clustering accuracy when the noise percentage is 0.1. Meanwhile, we also conduct experiments on the occluded Yale face dataset [55]. On each 16-by-16-pixel image, a part is randomly selected for occlusion processing

TABLE VI  
ACC/NMI OF THE ROBUSTNESS EXPERIMENTS ON TWO-MOON DATASET

Noise	10%	20%	30%
K-means	0.816/0.312	0.869/0.441	0.822/0.326
NMF	0.845/0.398	0.740/0.187	0.700/0.126
NCut	0.964/0.812	0.880/0.471	0.818/0.316
CAN	1.000/1.000	0.890/0.502	0.810/0.388
CLR	1.000/1.000	0.725/0.234	0.810/0.388
SSR	0.680/0.096	0.845/0.378	0.820/0.320
KMM	0.910/0.574	0.855/0.407	0.840/0.366
RRCSL	<b>1.000/1.000</b>	<b>0.920/0.627</b>	<b>0.875/0.487</b>

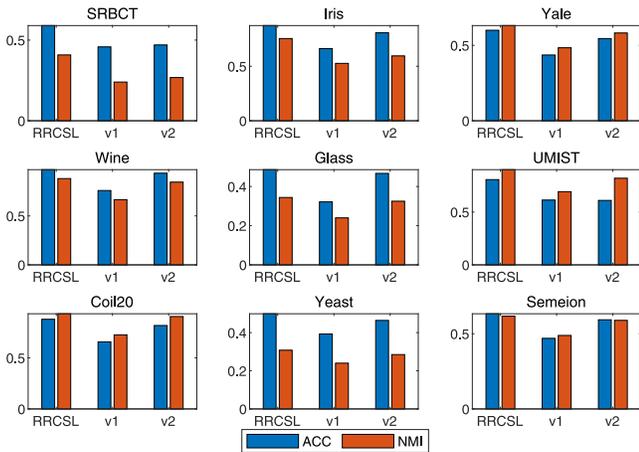


Fig. 6. Ablation experiments for single-view clustering.

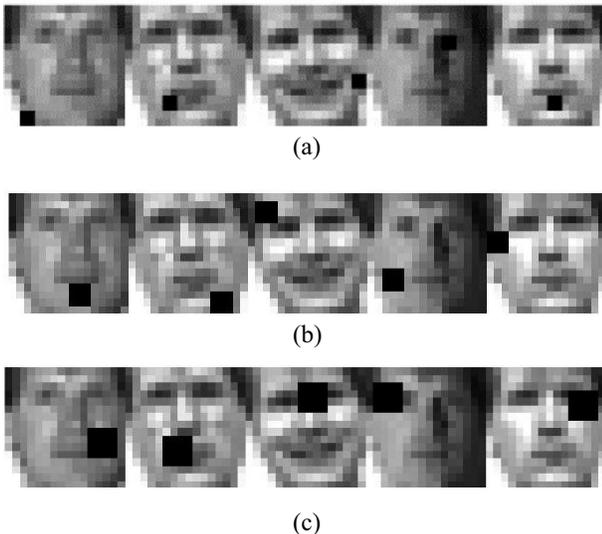


Fig. 7. Occluded Yale datasets. (a) Randomly add  $2 \times 2$  occlusion. (b) Randomly add  $3 \times 3$  occlusion. (c) Randomly add  $4 \times 4$  occlusion.

(as shown in Fig. 7). Table VII summarizes the results for all methods on the occluded Yale dataset. Obviously, the RRCSL method has better robustness than others.

### E. Ablation Experiments

In order to analyze the influence of components in the proposed method on clustering results, ablation experiments were conducted. The ablation experiments consist of two parts. First, to investigate the impact of the initial similarity graph,

TABLE VII  
ACC/NMI OF THE ROBUSTNESS EXPERIMENTS ON YALE DATASET

Occlusion	$2 \times 2$	$3 \times 3$	$4 \times 4$
K-means	0.460/0.527	0.446/0.510	0.420/0.483
NMF	0.343/0.421	0.327/0.399	0.319/0.393
NCut	0.517/0.565	0.508/0.558	0.484/0.534
CAN	0.509/0.564	0.436/0.454	0.430/0.485
CLR	0.466/0.559	0.509/0.547	0.430/0.464
SSR	0.551/0.557	0.533/0.534	0.478/0.509
KMM	0.364/0.411	0.343/0.387	0.332/0.403
RRCSL	<b>0.593/0.627</b>	<b>0.569/0.603</b>	<b>0.497/0.578</b>

we remove  $\|S - B\|_F^2$  and replace it with  $\|S\|_F^2$ , which is used to prevent  $S$  from becoming the identity matrix. Therefore, (11) becomes

$$\begin{aligned} \min_{S, F} \quad & \|X - XS\|_{2,1} + \alpha \|S\|_F^2 + \lambda \text{Tr}(F^T L_s F) \\ \text{s.t.} \quad & s_{ij} \geq 0, \sum_j s_{ij} = 1, F^T F = I \end{aligned} \quad (38)$$

and we name it RRCSL-Ablation-v1.

Besides the initial similarity graph, how much influence the Laplacian rank constraint has on the proposed RRCSL method is also an issue worth exploring. So we remove the Laplacian rank constraint and use  $K$ -means as the postprocessing to obtain the clustering result from the similarity graph, just like the traditional spectral clustering. Therefore, (11) becomes

$$\begin{aligned} \min_S \quad & \|X - XS\|_{2,1} + \alpha \|S - B\|_F^2 \\ \text{s.t.} \quad & s_{ij} \geq 0, \sum_j s_{ij} = 1 \end{aligned} \quad (39)$$

which is called RRCSL-Ablation-v2.

The optimizations and algorithms of RRCSL-Ablation-v1 and RRCSL-Ablation-v2 are similar to those in Section II, which are listed in Appendices A and B. As shown in Fig. 6, if the similarity graph is not limited to the neighborhood of the initial similarity graph  $B$ , the accuracy of clustering will decrease. The reason is that there are multiple linear combinations of  $S$  for the same  $X$ . By comparison, using  $K$ -means as postprocessing instead of the rank constraint has some minor impact on clustering. But considering that  $K$ -means is unstable, it is effective and necessary to add the Laplacian rank constraint into the proposed method and make the learned graph have a clear structure.

## V. EXPERIMENTS OF MULTIVIEW CLUSTERING

In this section, we evaluate the effectiveness of RRCSL on multiview clustering.

### A. Dataset Descriptions

The *MSRC-v1* [56] dataset contains 240 images from eight classes, and we extract four visual features (CMT, LBP, GIST, and CENT) for clustering.

The *Caltech101-7* [57] dataset contains 101 categories of object recognition images. Following Wang *et al.* [40], we select the widely used seven classes (Dolla-Bill, Face, Garfield, Motorbikes, Snoopy, Stop-Sign, and WindsorChair), and obtain 1474 images.

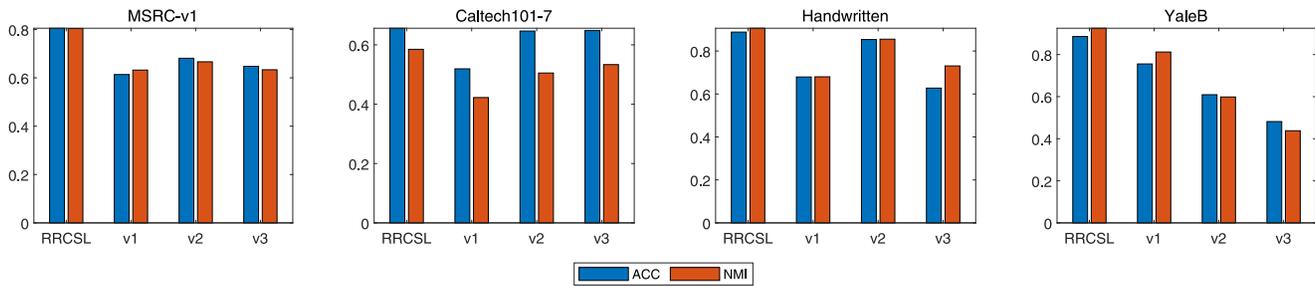


Fig. 8. Ablation experiments for multiview clustering.

TABLE VIII  
ACC OF THE ROBUSTNESS MULTIVIEW CLUSTERING EXPERIMENTS

Method	MSRC-v1	Caltech101-7	Handwritten	YaleB
SC(best)	0.762	0.663	0.819	0.132
Co-train	0.438	0.406	0.633	0.231
Co-reg	0.723	0.384	0.853	0.200
AMGL	0.619	0.636	0.877	0.434
ECMSC	0.652	0.540	0.807	0.463
MLAN	0.738	0.794	0.865	0.386
SwMC	0.781	0.391	0.670	0.392
AWP	0.767	0.564	0.831	0.459
RRCSL	<b>0.805</b>	<b>0.828</b>	<b>0.888</b>	<b>0.886</b>

TABLE IX  
NMI OF THE ROBUSTNESS MULTIVIEW CLUSTERING EXPERIMENTS

Method	MSRC-v1	Caltech101-7	Handwritten	YaleB
SC(best)	0.625	0.506	0.860	0.406
Co-train	0.498	0.372	0.603	0.160
Co-reg	0.630	0.298	0.769	0.146
AMGL	0.603	0.512	0.886	0.452
ECMSC	0.652	0.421	0.845	0.441
MLAN	0.724	0.450	0.868	0.356
SwMC	0.718	0.139	0.728	0.412
AWP	0.703	0.503	0.874	0.463
RRCSL	<b>0.805</b>	<b>0.552</b>	<b>0.888</b>	<b>0.925</b>

TABLE X  
ACC/NMI OF THE ROBUSTNESS EXPERIMENTS ON YALEB DATASET

Methods	1%	5%	10%
SC(best)	0.432/0.406	0.482/0.440	0.434/0.404
Co-train	0.191/0.111	0.205/0.114	0.260/0.186
Co-reg	0.225/0.167	0.199/0.128	0.294/0.251
AMGL	0.480/0.463	0.422/0.436	0.471/0.457
ECMSC	0.471/0.443	0.489/0.479	0.491/0.467
MLAN	0.420/0.349	0.548/0.479	0.520/0.453
SwMC	0.469/0.416	0.492/0.448	0.472/0.465
AWP	0.500/0.482	0.526/0.512	0.454/0.443
RRCSL	<b>0.886/0.925</b>	<b>0.726/0.862</b>	<b>0.539/0.522</b>

The *Handwritten* [58] dataset contains 2000 images from ten classes, and six kinds of features (FOU, FAC, KAR, PIX MOR, and ZER) are employed in the experiment.

*YaleB* [43] is a challenging dataset due to the large variation of luminance, containing three kinds of features. As done in [59], we used the first ten classes.

### B. Experimental Setup

Eight state-of-the-art multiview clustering methods are taken as competitors, including spectral clustering [60] (we run SC on all single views and the best results are set as the baseline), co-trained spectral clustering (Co-train) [27], co-regularized spectral clustering (Co-reg) [28], autoweighted multiple graph learning (AMGL) [61], exclusivity-consistency regularized multiview subspace clustering (ECMSC) [62], multiview learning with adaptive neighbors (MLANs) [63], self-weighted multiview clustering (SwMC) [64], and adaptively weighted procrustes (AWPs) [26]. The parameter setting is the same as that in Section IV. The weight  $w^{(v)}$  is initialized as  $w^{(1)}, w^{(2)}, \dots, w^{(n_v)} = (1/n_v)$ . Our algorithm requires the weighted sum of each view matrix, but the dimension of each view is usually not the same. In order to solve this problem, we fill the smaller view with 0 in the preprocessing stage.

### C. Performance

Tables VIII and IX show the ACC and NMI of comparison experiments, respectively, (bold indicates the best). SC, Co-train, and Co-reg neglect the different roles of the views. In contrast, our strategy of adaptively distributing weights of various views makes the RRCSL method perform better. Meanwhile, compared with self-weighted multiview clustering methods such as SwMC, our method can construct data similarity graph with higher quality and has better robustness

against noise and outliers. All these ensure that the RRCSL method achieves the best performance on all datasets. Our proposed method is applicable to both single-view clustering and multiview clustering, which fully illustrates its superiority and extendibility.

### D. Robustness to Noises and Outliers

To demonstrate the robustness of our RRCSL method for multiview clustering, we conduct experiments on the YaleB dataset. For all views, different proportions of data (1%, 5%, and 10%) were randomly selected to be 0. Table X shows the clustering results, and it can be seen that RRCSL outperforms these methods in all the experiments. Our method maintains excellent robustness in multiview clustering.

### E. Ablation Experiments

For multiview clustering, we also conducted ablation experiments to analyze the influence of components in the proposed RRCSL method. Using the same settings as in Section IV, we analyze the influence of initial similarity graph and Laplacian rank constraint on multiview clustering. As with single-view clustering, the clustering results show that these components are necessary. In addition, to verify that our strategy of adaptively distributing weights of various views is effective, we

TABLE XI

ALGORITHM OF SINGLE-VIEW CLUSTERING BY RRCSL-ABLATION-V1

---

**ALGORITHM 4:**Single-view Clustering by RRCSL-Ablation-v1

---

**Input:**  $X \in \mathbb{R}^{d \times n}$ , the number of clusters  $k$ , parameter  $\alpha > 0$

**Output:**the similarity graph  $S$  with  $k$  connected components

---

**While** not converge **do**

1. update  $E$  by Eq.(16);
2. update  $Z$  by Eq.(43);
3. update  $S$  by solving Eq.(26);
4. update  $F$  by calculating  $k$  eigenvectors corresponding to the first  $k$  minimum eigenvalues of  $L_s$ ;
5. update the ALM parameters.

**End While**

---

TABLE XII

ALGORITHM OF SINGLE-VIEW CLUSTERING BY RRCSL-ABLATION-V2

---

**ALGORITHM 4:**Single-view Clustering by RRCSL-Ablation-v2

---

**Input:**  $X \in \mathbb{R}^{d \times n}$ , the number of clusters  $k$ , parameter  $\alpha > 0$

**Output:**the similarity graph  $S$  with  $k$  connected components

---

**repeat**

1. update  $E$  by Eq.(16);
2. update  $Z$  by Eq.(19);
3. update  $S$  by solving Eq.(26);
5. update the ALM parameters.

**until** converge

Perform k-means with the  $S$ .

---

remove the step of updating the weights. This clustering method with averaging the weights is called RRCSL-Ablation-v3. As shown in Fig. 8, the complete version of the RRCSL method performs the best. The experiment demonstrates that if different views are treated equally, the clustering will be affected by the low-performing views and the accuracy of the results will be reduced. Our adaptive distribution strategy of weights solves this problem effectively and each component is reasonable.

## VI. CONCLUSION

In this article, we proposed a new graph-based RRCSL framework for both single-view and multiview clustering. In our method, the sparse representation with a rank constraint is adopted to learn the desired similarity graph, and a predefined graph is used to guide the graph learning procedure. Therefore, the obtained graph has more capability on data representation and relationship preservation. Besides,  $L_{2,1}$ -norm is combined to reduce the impact of data noises and outliers. For multiview clustering, our adaptive distribution strategy of weights solves the problem that clustering is affected by the low-performing views. Extensive experiments on real-world datasets have demonstrated the superiority and robustness of the proposed framework.

## APPENDIX A

### OPTIMIZATIONS AND ALGORITHMS OF RRCSL-ABLATION-V1

Let  $E = X - XZ$ ,  $Z = S$ , and (38) is transformed into the following ALM problem:

$$\min_S \|E\|_{2,1} + \alpha \|Z\|_F^2 + \lambda \text{Tr}(F^T L_s F) + \frac{\mu}{2} \left\| E - X + XZ + \frac{\Lambda_1}{\mu} \right\|_F^2 + \frac{\mu}{2} \left\| Z - S + \frac{\Lambda_2}{\mu} \right\|_F^2$$

$$\text{s.t. } s_{ij} \geq 0, \sum_j s_{ij} = 1, F^T F = I \quad (40)$$

where  $\mu \in \mathbb{R}^{1 \times 1}$  is a regularity coefficient, and  $\Lambda_1 \in \mathbb{R}^{d \times n}$  and  $\Lambda_2 \in \mathbb{R}^{d \times n}$  are penalty parameters.

**Update  $E$ ,  $S$ , and  $F$ :** When updating  $E$ ,  $S$ , or  $F$ , (40) is equivalent to the optimization problem in Section II, just as (13)–(16) and (20)–(27).

**Update  $Z$ :** When updating  $Z$ , we fix  $E$ ,  $S$ , and  $F$ . Thus, (40) becomes

$$\min_Z \alpha \|Z\|_F^2 + \frac{\mu}{2} \left\| E - X + XZ + \frac{\Lambda_1}{\mu} \right\|_F^2 + \frac{\mu}{2} \left\| Z - S + \frac{\Lambda_2}{\mu} \right\|_F^2. \quad (41)$$

Taking the derivation of (17) with respect to  $Z$

$$\min_Z 2\alpha \text{Tr}Z + \mu \text{Tr} \left( Z - S + \frac{\Lambda_2}{\mu} \right) + \frac{\mu}{2} \text{Tr} \left[ 2X^T XZ + 2X^T \left( E - X + \frac{\Lambda_1}{\mu} \right) \right]. \quad (42)$$

Let (42) be zero and we obtain the optimal solution for  $Z$  as follows:

$$Z = \frac{\left[ -2X^T \left( E - X + \frac{\Lambda_1}{\mu} \right) + S - \frac{\Lambda_2}{\mu} \right]}{\left( \frac{2\alpha}{\mu} I + X^T X + I \right)}. \quad (43)$$

See Table XI.

## APPENDIX B

### OPTIMIZATIONS AND ALGORITHMS OF RRCSL-ABLATION-V2

Let  $E = X - XZ$ ,  $Z = S$ , and (39) is transformed into the following ALM problem:

$$\min_S \|E\|_{2,1} + \alpha \|Z - B\|_F^2 + \frac{\mu}{2} \left\| Z - S + \frac{\Lambda_2}{\mu} \right\|_F^2 + \frac{\mu}{2} \left\| E - X + XZ + \frac{\Lambda_1}{\mu} \right\|_F^2$$

$$\text{s.t. } s_{ij} \geq 0, \sum_j s_{ij} = 1 \quad (44)$$

where  $\mu \in \mathbb{R}^{1 \times 1}$  is a regularity coefficient, and  $\Lambda_1 \in \mathbb{R}^{d \times n}$  and  $\Lambda_2 \in \mathbb{R}^{d \times n}$  are penalty parameters.

**When updating  $E$ ,  $Z$ , or  $S$ ,** (44) is equivalent in Section II, just as (13)–(26). The objective value of (44) decreases in each iteration. When (44) converges, we perform  $k$ -means with  $S$  to obtain the final clustering result.

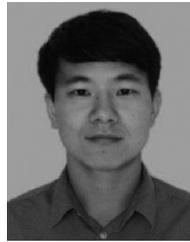
See Table XII.

## REFERENCES

- [1] L. He, N. Ray, Y. Guan, and H. Zhang, "Fast large-scale spectral clustering via explicit feature mapping," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1058–1071, Mar. 2019.
- [2] A. Segatori, A. Bechini, P. Ducange, and F. Marcelloni, "A distributed fuzzy associative classifier for big data," *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2656–2669, Sep. 2018.
- [3] X. Li, M. Chen, F. Nie, and W. Qi, "Locality adaptive discriminant analysis," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017, pp. 2201–2207.

- [4] Y. Zhao, K. Xu, E. Zhu, X. Liu, X. Zhu, and J. Yin, "Triangle lasso for simultaneous clustering and optimization in graph datasets," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 8, pp. 1610–1623, Aug. 2019.
- [5] J. Xu, M. Yu, L. Shao, W. Zuo, and D. Zhang, "Scaled simplex representation for subspace clustering," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1493–1505, Mar. 2021.
- [6] Y. Pang, J. Xie, F. Nie, and X. Li, "Spectral clustering by joint spectral embedding and spectral rotation," *IEEE Trans. Cybern.*, vol. 50, no. 1, pp. 247–258, Jan. 2020.
- [7] M. Brbić and I. Kopriva, "L0-motivated low-rank sparse subspace clustering," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1711–1725, Apr. 2020.
- [8] Y. Zhao, A. K. Shrivastava, and K. L. Tsui, "Regularized Gaussian mixture model for high-dimensional clustering," *IEEE Trans. Cybern.*, vol. 49, no. 10, pp. 3677–3688, Oct. 2019.
- [9] J. Wen *et al.*, "DIMC-Net: Deep incomplete multi-view clustering network," in *Proc. ACM Int. Conf. Multimedia*, 2020, pp. 3753–3761.
- [10] H. Fan, L. Zheng, C. Yan, and Y. Yang, "Unsupervised person re-identification: Clustering and fine-tuning," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 14, no. 4, pp. 1–18, 2018.
- [11] K. Zhan, F. Nie, J. Wang, and Y. Yang, "Multiview consensus graph clustering," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1261–1270, Oct. 2019.
- [12] K. Zhan, C. Niu, C. Chen, F. Nie, C. Zhang, and Y. Yang, "Graph structure fusion for multiview clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 10, pp. 1984–1993, Mar. 2019.
- [13] F. Nie, X. Wang, and H. Huang, "Clustering and projected clustering with adaptive neighbors," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Disc. Data Mining*, 2014, pp. 977–986.
- [14] F. Nie, X. Wang, M. I. Jordan, and H. Huang, "The constrained Laplacian rank algorithm for graph-based clustering," in *Proc. AAAI*, 2016, pp. 1969–1976.
- [15] J. Huang, F. Nie, and H. Huang, "A new simplex sparse learning model to measure data similarity for clustering," in *Proc. IJCAI*, 2015, pp. 3569–3575.
- [16] X. Li, M. Chen, F. Nie, and Q. Wang, "A multiview-based parameter free framework for group detection," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 4147–4153.
- [17] X. Liu *et al.*, "Late fusion incomplete multi-view clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 3, pp. 2410–2423, Oct. 2018.
- [18] Y. Wang, Z. Wenjie, L. Wu, X. Lin, M. Fang, and S. Pan, "Iterative views agreement: An iterative low-rank based structured optimization method to multi-view spectral clustering," in *Proc. Int. Joint Conf. Artif. Intell.*, 2016, pp. 14–28.
- [19] J. Wen, Z. Zhang, Z. Zhang, L. Fei, and M. Wang, "Generalized incomplete multiview clustering with flexible locality structure diffusion," *IEEE Trans. Cybern.*, vol. 51, no. 1, pp. 101–114, Jan. 2021.
- [20] P. Zhou, L. Du, and X. Li, "Self-paced consensus clustering with bipartite graph," in *Proc. Int. Joint Conf. Artif. Intell.*, 2020, pp. 2133–2139.
- [21] K. Zhan, C. Zhang, J. Guan, and J. Wang, "Graph learning for multiview clustering," *IEEE Trans. Cybern.*, vol. 48, no. 10, pp. 2887–2895, Oct. 2018.
- [22] D. Xie, Q. Gao, Q. Wang, X. Zhang, and X. Gao, "Adaptive latent similarity learning for multi-view clustering," *Neural Netw.*, vol. 121, pp. 409–418, Jan. 2020.
- [23] Y. Wang, L. Wu, X. Lin, and J. Gao, "Multiview spectral clustering via structured low-rank matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4833–4843, Oct. 2018.
- [24] X. Lan, S. Zhang, P. C. Yuen, and R. Chellappa, "Learning common and feature-specific patterns: A novel multiple-sparse-representation-based tracker," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2022–2037, Apr. 2018.
- [25] Q. Gao, P. Zhang, W. Xia, X. Deyan, X. Gao, and D. Tao, "Enhanced tensor rpca and its application," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Aug. 18, 2020, doi: [10.1109/TPAMI.2020.3017672](https://doi.org/10.1109/TPAMI.2020.3017672).
- [26] F. Nie, L. Tian, and X. Li, "Multiview clustering via adaptively weighted procrustes," in *Proc. Int. Conf. Knowl. Disc. Data Min.*, 2018, pp. 2022–2030.
- [27] A. Kumar and H. Daumé, "A co-training approach for multi-view spectral clustering," in *Proc. ICML*, 2011, pp. 393–400.
- [28] A. Kumar, P. Rai, and H. Daume, "Co-regularized multi-view spectral clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 1413–1421.
- [29] Y. Fan, J. Liang, R. He, B.-G. Hu, and S. Lyu, "Robust localized multi-view subspace clustering," 2017. [Online]. Available: [arXiv:1705.07777](https://arxiv.org/abs/1705.07777).
- [30] Y. Cai, Z. Zhang, Z. Cai, X. Liu, X. Jiang, and Q. Yan, "Graph convolutional subspace clustering: A robust subspace clustering framework for hyperspectral image," 2020. [Online]. Available: [arXiv:2004.10476](https://arxiv.org/abs/2004.10476).
- [31] Y. Yang, H. T. Shen, Z. Ma, Z. Huang, and X. Zhou, "L2, 1-NORM regularized discriminative feature selection for unsupervised," in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, 2011, pp. 1589–1594.
- [32] S. Li, M. Shao, and Y. Fu, "Multi-view low-rank analysis for outlier detection," in *Proc. SIAM Int. Conf. Data Mining*, 2015, pp. 748–756.
- [33] Q. Gao, S. Xu, F. Chen, C. Ding, X. Gao, and Y. Li, "R1-2-DPCA and face recognition," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1212–1223, Apr. 2019.
- [34] B. Mohar, Y. Alavi, G. Chartrand, and O. Oellermann, "The Laplacian spectrum of graphs," *Graph Theory Comb. Appl.*, vol. 2, nos. 871–898, p. 12, 1991.
- [35] K. Fan, "On a theorem of weyl concerning eigenvalues of linear transformations I," *Proc. Nat. Acad. Sci. USA*, vol. 35, no. 11, pp. 652–655, 1949.
- [36] M. J. Powell, "A method for nonlinear constraints in minimization problems," in *Optimization*. New York, NY, USA: Academic, 1969, pp. 283–298.
- [37] D. P. Bertsekas, "Constrained optimization and lagrange multiplier methods," in *Computer Science and Applied Mathematics*. Boston, MA, USA: Academic, 1982.
- [38] J. Huang, F. Nie, H. Huang, and C. Ding, "Robust manifold nonnegative matrix factorization," *ACM Trans. Knowl. Disc. Data*, vol. 8, no. 3, p. 21, 2014.
- [39] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Stat. Soc. B Stat. Methodol.*, vol. 68, no. 1, pp. 49–67, 2006.
- [40] Q. Wang, M. Chen, F. Nie, and X. Li, "Detecting coherent groups in crowd scenes by multiview clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 46–58, Jan. 2020.
- [41] J. Khan *et al.*, "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," *Nat. Med.*, vol. 7, no. 6, p. 673, 2001.
- [42] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.*, vol. 7, no. 2, pp. 179–188, 1936.
- [43] A. S. Georghiadis, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.
- [44] S. Aeberhard, D. Coomans, and O. D. Vel, "Improvements to the classification performance of RDA," *J. Chemometrics*, vol. 7, no. 2, pp. 99–115, 1993.
- [45] P. Zhong and M. Fukushima, "Regularized nonsmooth Newton method for multi-class support vector machines," *Optim. Meth. Softw.*, vol. 22, no. 1, pp. 225–236, 2007.
- [46] D. B. Graham and N. M. Allinson, "Characterising virtual eigensignatures for general purpose face recognition," in *Face Recognition*. Heidelberg, Germany: Springer, 1998, pp. 446–456.
- [47] S. A. Nene *et al.*, *Columbia Object Image Library (COIL-20)*. Princeton, NJ, USA: Citeseer, 1996.
- [48] P. Horton and K. Nakai, "A probabilistic classification system for predicting the cellular localization sites of proteins," in *Proc. ISMB*, vol. 4, 1996, pp. 109–115.
- [49] M. Buscema, "MetaNet\*: The theory of independent judges," *Substance Use Misuse*, vol. 33, no. 2, pp. 439–461, 1998.
- [50] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, p. 788, 1999.
- [51] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [52] F. Nie, C.-L. Wang, and X. Li, "K-multiple-means: A multiple-means clustering method with specified k clusters," in *Proc. ACM SIGKDD Int. Conf. Knowl. Disc. Data Mining*, 2019, pp. 959–967.
- [53] Q. Wang, Z. Qin, F. Nie, and X. Li, "Spectral embedded adaptive neighbors clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 4, pp. 1265–1271, Apr. 2019.
- [54] Q. Wang, X. He, X. Jiang, and X. Li, "Robust bi-stochastic graph regularized matrix factorization for data clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Jul. 7, 2020, doi: [10.1109/TPAMI.2020.3007673](https://doi.org/10.1109/TPAMI.2020.3007673).
- [55] R. Liu, M. Chen, Q. Wang, and X. Li, "Robust rank constrained sparse learning: A graph-based method for clustering," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2020, pp. 4217–4221.

- [56] J. M. Winn and N. Jojic, "LOCUS: Learning object classes with unsupervised segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, pp. 756–763.
- [57] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," *Comput. Vis. Image Understand.*, vol. 106, no. 1, pp. 59–70, 2007.
- [58] M. V. Breukelen, R. P. W. Duin, D. M. J. Tax, and J. E. Den Hartog, "Handwritten digit recognition by combined classifiers," *Kybernetika*, vol. 34, no. 4, pp. 381–386, 1998.
- [59] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang, and Y. Qu, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *Int. J. Comput. Vis.*, vol. 126, no. 11, pp. 1157–1179, 2018.
- [60] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, pp. 849–856.
- [61] F. Nie, J. Li, and X. Li, "Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification," in *Proc. Int. Joint Conf. Artif. Intell.*, 2016, pp. 1881–1887.
- [62] X. Wang, X. Guo, Z. Lei, C. Zhang, and S. Z. Li, "Exclusivity-consistency regularized multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 923–931.
- [63] F. Nie, G. Cai, and X. Li, "Multi-view clustering and semi-supervised classification with adaptive neighbours," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 2408–2414.
- [64] F. Nie, J. Li, and X. Li, "Self-weighted multiview clustering with multiple graphs," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017, pp. 362–377.



**Ran Liu** received the B.E. degree in transportation equipment and control engineering from Northwestern Polytechnical University, Xi'an, China, in 2018, where he is currently pursuing the Ph.D. degree in computer science with the School of Computer Science and the School of Artificial Intelligence, Optics and Electronics.

His research interests include machine learning and computer vision.



**Mulin Chen** received the B.E. degree in software engineering and the Ph.D. degree in computer application technology from Northwestern Polytechnical University, Xi'an, China, in 2014 and 2019, respectively.

He is currently a Researcher with the School of Artificial Intelligence, Optics and Electronics, Northwestern Polytechnical University. His current research interests include computer vision and machine learning.



**Qi Wang** (Senior Member, IEEE) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively.

He is currently a Professor with the School of Artificial Intelligence, Optics and Electronics, Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision, pattern recognition, and remote sensing.

**Xuelong Li** (Fellow, IEEE) is currently a Full Professor with the School of Artificial Intelligence, Optics and Electronics, Northwestern Polytechnical University, Xi'an, China.