

Research Article

Robust Spectral Clustering via Low-Rank Sample Representation

Hao Liang,¹ Hai-Tang Guan,² Stanley Ebhohimhen Abhadiomhen ,^{3,4} and Li Yan³

¹Graduate School of Jiangsu University, Zhenjiang 212013, Jiangsu, China

²Haian Experimental High School, Nantong, Jiangsu, China

³School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, Jiangsu, China

⁴Department of Computer Science, University of Nigeria, Nsukka, Nigeria

Correspondence should be addressed to Stanley Ebhohimhen Abhadiomhen; stanley.abhadiomhen@unn.edu.ng

Received 31 January 2022; Accepted 13 April 2022; Published 29 April 2022

Academic Editor: Babak Daneshvar Rouyendegh (B. Erdebilli)

Copyright © 2022 Hao Liang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Traditional clustering methods neglect the data quality and perform clustering directly on the original data. Therefore, their performance can easily deteriorate since real-world data would usually contain noisy data samples in high-dimensional space. In order to resolve the previously mentioned problem, a new method is proposed, which builds on the approach of low-rank representation. The proposed approach first learns a low-rank coefficient matrix from data by exploiting the data's self-expressiveness property. Then, a regularization term is introduced to ensure that the representation coefficient of two samples, which are similar in original high-dimensional space, is close to maintaining the samples' neighborhood structure in the low-dimensional space. As a result, the proposed method obtains a clustering structure directly through the low-rank coefficient matrix to guarantee optimal clustering performance. A wide range of experiments shows that the proposed method is superior to compared state-of-the-art methods.

1. Introduction

Clustering is a powerful technique in unsupervised machine learning, which requires a measure of similarity to group data samples into k classes. Therefore, traditional clustering algorithms such as k -means [1] perform clustering on original data samples directly by assuming that alike samples reside around a centroid. This assumption is noted in [2] as too restrictive in a real-world setting where two samples can be far away from each other yet belong to the same cluster, and vice versa. Furthermore, as real-world data is often high dimensional, the k -means approach is also computationally inefficient. Thus, efforts were made in literature [3] to enhance its efficiency. However, the robustness needed for accurate clustering under the aforementioned real-world circumstances was largely ignored, as they do not consider data quality.

Therefore, considering that corresponding low-dimensional subspaces exist for high-dimensional data, Zheng et al. [4] and Cai et al. [5] proposed graph regularized sparse coding (GraphSC) and graph regularized nonnegative

matrix (GNMF), respectively, to tackle the previously mentioned lapses. The object of both methods is such that discriminability and computational efficiency can be enhanced simultaneously to improve clustering performance. Nonetheless, as they do not fully utilize the self-expressiveness property of data, a good representation of data cannot be guaranteed in all cases. As a result, subsequent spectral-based approaches were built on the self-expressiveness property of data where a data sample can be represented by a combination of the bases in a whole dataset [6]. Specifically, low-rank representation (LRR) [7] and sparse subspace clustering (SSC) [8] are the two classical methods, and the newer ones are formed using either of their principles. An example is Laplacian regularized low-rank representation (LapLRR) [9] proposed to improve LRR's global data structure initiative by further capturing data's intrinsic nonlinear geometric information. Similarly, Zhang et al. [10] proposed spectral-spatial sparse subspace clustering (S^4C) using the SSC's $L1$ -norm regularization strategy. Besides, considering the local data structure that LRR previously ignored, low-rank representation with adaptive graph

regularization (LRR_AGR) [11] was proposed to learn optimal clustering. However, because real-world data often contain noise, these spectral methods cannot guarantee that two similar samples in the original high-dimensional space will have a close representation coefficient in the low-dimensional space.

To address the previously mentioned concern, a new method that adopts LRR's approach to first learn data's low-rank coefficient matrix is proposed. A regularization term is then introduced to ensure that the samples' neighborhood structure is maintained in the low-dimensional space. Thus, the proposed method learns optimal clustering structure directly from the low-rank coefficient matrix without spectral postprocessing. It is achieved by imposing a constraint on the low-rank coefficient matrix to promote a robust affinity matrix, a rank constraint is further utilized to make the affinity matrix express the clustering structure.

- (1) We propose a novel method that uses the LRR strategy first to obtain data's low-rank coefficient matrix, of which a regularization term is also introduced to ensure that samples' neighborhood structure is maintained in the low-dimensional space. This approach is different from the existing ones, which simply assume such a structure without a robust strategy to tackle the influence of noise.
- (2) We constrain the low-rank coefficient matrix to guide a robust affinity matrix to obtain optimal clustering directly while avoiding spectral post-processing of the affinity matrix, unlike most existing methods.
- (3) Several experiments are performed to evaluate the effectiveness of the proposed method using accuracy (ACC), normalized mutual information (NMI), and purity evaluation metrics. The results demonstrate the superiority of the proposed method over similar state-of-the-art ones.

2. Related Work

To resolve the inefficiency of traditional clustering methods, several spectral type methods based on the self-expressiveness property of data have been proposed over the years. Therefore, one can categorize these methods into two groups according to the strategy used to obtain a given data's low-dimensional coefficient matrix. The first group has methods like those mentioned in [10, 12–14], which use the L_1 -norm regularization [8] to acquire the coefficient matrix. For example, Zhang et al. [10] proposed a spectral-spatial sparse subspace clustering algorithm for hyperspectral remote sensing images, which obtains a final clustering by applying a spectral clustering algorithm on an adjacent matrix. Also, using the L_1 -norm approach, Li et al. [12] proposed structured sparse subspace clustering (S^3C), which learns an affinity matrix and data segmentation jointly to improve clustering accuracy. The L_1 -norm's regularization main advantage lies in the fact that it can obtain a sparse representation of data samples. However, it ignores the data's global structure; hence, it can be vulnerable to noise.

In contrast, the nuclear norm regularization approach was proposed in [7] to capture a global data structure, thus providing some robustness against noise and possible outliers. Thus, the second category contains methods like those in [6, 9, 11, 15–24], which apply the aforementioned nuclear norm technique to learn the coefficient matrix. Illustratively, the LapLRR method was proposed in [9] based on the nuclear norm. LapLRR mainly focuses on capturing data's nonlinear geometric structures to improve clustering performance. The compound rank- k projection (CRP) [23] algorithm was proposed for bilinear analysis. Specifically, CRP uses multiple rank- k projection models to enhance discriminant ability. Meanwhile, the constrained low-rank representation (CLRR) [15] emphasizes on increasing discriminating ability by incorporating supervision information as hard constraints. Since the LRR methods are based on affinity graph construction, Luo et al. [24] argue that the affinity measurement in the original feature space would usually suffer from the curse of dimensionality. Thus, a method was proposed, which assumes similarity between instances only if they have a larger probability of being neighbours. However, because a fixed graph often does not guarantee optimal performance, Wen et al. [11] then introduced flexibility into graph learning and proposed the LRR-AGR method. Besides, a technique founded on a finite mixture of exponential power (MoEP) was proposed in [19] to handle complex noise contamination in data. Furthermore, the recent work of [25] introduced an adaptive kernel into LRR to boost the accuracy of clustering. Besides, a coupled low-rank representation (CLR) strategy was presented in [6] to learn accurate clustering from data using the k block diagonal regularizer [26]. Aside from that, Yan et al. [27] proposed a novel self-weighted robust linear discriminant analysis (LDA) for multiclass classification, especially with edge classes.

This study also adopts the nuclear norm regularization approach. However, unlike the previously mentioned techniques, the proposed method further injects a regularization term to guarantee that two similar samples in the original high-dimensional space would have a close representation coefficient in the low-dimensional space to handle noise distortion in data more holistically.

3. The Proposed Method

In this section, the proposed method is formulated first. Then, an optimization method is proposed to solve the model.

3.1. Model Formulation. LRR's nuclear norm regularization has been shown in many studies, not limited to those previously cited, as an effective technique for capturing a robust data coefficient matrix due to its global orientation. Thus, LRR is integrated into our model to capture the data's low-rank coefficient matrix as follows:

$$\begin{aligned} & \|Z\|_* + \lambda_2 \|E\|_{2,1}, \\ \text{s.t. } & X = XZ + E, \end{aligned} \quad (1)$$

where $Z \in \mathbb{R}^{n \times n}$ is the low-rank coefficient matrix. $E \in \mathbb{R}^{d \times n}$ is the error matrix, assuming that part of the given data $X \in \mathbb{R}^{d \times n}$ is corrupt. However, because the level of corruption in real-world data is unknown in advance, one cannot guarantee that Z will capture an accurate similarity of the data samples. To address this concern, a regularization term $\sum_{i,j} \|\hat{x}_i - \hat{x}_j\|^2 w_{i,j}$ is introduced into (1) to ensure that when samples x_i and x_j are similar in the original high-dimensional space, z_i and z_j should be similar in low-dimensional space, so that $Z_{ij} = 1$, otherwise 0. Thus, we have

$$\begin{aligned} & \min_{\hat{X}, Z, E} \sum_{i,j} \|\hat{x}_i - \hat{x}_j\|^2 w_{i,j} + \lambda_1 \|Z\|_* + \lambda_2 \|E\|_{2,1} \\ & \text{s.t. } X = XZ + E, \hat{X} = XZ, \end{aligned} \quad (2)$$

where $\sum_{i,j} \|\hat{x}_i - \hat{x}_j\|^2 w_{i,j} = 2tr(\hat{X}L_W\hat{X}^T)$. $W = (W + W^T)/2$, $L_W = D - W$ and $D = \text{diag}(\sum_i W_{ij})$. Once Z is obtained, existing methods directly learn an affinity matrix U such that $U = (|Z| + |Z^T|)/2$. Most of these methods then apply a spectral clustering algorithm on U to obtain the clustering structure. Differently, we utilize a constraint $Z = U$ to ensure that the coefficient of Z_{ij} and U_{ij} is the same, in which some nonnegative constraints, $\text{diag}(U) = 0, U \geq 0, U^T \mathbf{1} = 1$, imposed on U ensures that its entries are nonnegative. As a result, a spectral postprocessing of U is then dodged by imposing a rank constraint $\text{rank}(L_U) = n - c$ on its Laplacian matrix to allow it to express our clustering structure using Theorem 1. Therefore, the proposed model is formulated as

$$\begin{aligned} & \min_{\hat{X}, Z, S, E} 2tr(\hat{X}L_W\hat{X}^T) + \lambda_1 \|S\|_* + \lambda_2 \|E\|_{2,1} \\ & \text{s.t. } X = XZ + E, Z = U, \hat{X} = XZ, \text{diag}(U) = 0, \\ & U \geq 0, U^T \mathbf{1} = 1, \text{rank}(L_U) = n - c, \end{aligned} \quad (3)$$

where $L_U = D_U - U$ is the Laplacian matrix of U and $D_U \in \mathbb{R}^{n \times n}$ is the diagonal matrix with i th entry $\sum_j U_{ij} = 1$. λ_1 and λ_2 are parameters to balance the terms, respectively.

Theorem 1 (see [6]). *If U is nonnegative, the multiplicity c of the zero eigenvalue of the graph Laplacian L_U corresponds to the number of connected components in the graph associated with U .*

3.2. Optimization. To solve (3), the augmented Lagrange multiplier (ALM) [28, 29] approach is adopted, but an auxiliary term $Z = S$ is introduced first to make it easily solvable. Besides, because the rank constraint is not linear, a similar strategy employed in [6] is followed to express the i th smallest eigenvalue of L_U as $\theta_i(L_U)$. Thus, given sufficient λ_3 , (3) is equal to

$$\begin{aligned} & \min_{\hat{X}, S, U, E} 2tr(\hat{X}L_W\hat{X}^T) + \lambda_1 \|S\|_* + \lambda_2 \|E\|_{2,1} + 2\lambda_3 \sum_{i=1}^c \theta_i(L_U) \\ & \text{s.t. } Z = S, X = XZ + E, Z = U, \hat{X} = XZ, \text{diag}(U) = 0, \\ & U \geq 0, U^T \mathbf{1} = 1. \end{aligned} \quad (4)$$

According to Ky Fan's theorem [30], $\sum_{i=1}^c \theta_i(L_U)$ is equivalent to minimizing $Tr(F^T L_U F)$ subject to $F^T F = I$. Hence, we have

$$\begin{aligned} & \min_{\hat{X}, Z, S, E} 2tr(\hat{X}L_W\hat{X}^T) + \lambda_1 \|S\|_* + \lambda_2 \|E\|_{2,1} + 2\lambda_3 Tr(F^T L_U F) \\ & \text{s.t. } Z = S, X = XZ + E, Z = U, \hat{X} = XZ, F^T F = I, \text{diag}(U) \\ & = 0, U \geq 0, U^T \mathbf{1} = 1. \end{aligned} \quad (5)$$

Following conventional practice, the augmented Lagrangian function of (5) is obtained as follows:

$$\begin{aligned} & \min_{\hat{X}, Z, S, E, U, F} 2tr(\hat{X}L_W\hat{X}^T) + \lambda_1 \|S\|_* + \lambda_2 \|E\|_{2,1} + 2\lambda_3 Tr(F^T L_U F) \\ & + \langle Y_1, Z - S \rangle + \langle Y_2, X - XZ - E \rangle + \langle Y_3, \hat{X} - XZ \rangle \\ & + \langle Y_4, Z - U \rangle + \frac{\mu}{2} \left(\|Z - S\|_F^2 + \|X - XZ - E\|_F^2 + \|\hat{X} - XZ\|_F^2 + \|Z - U\|_F^2 \right) \\ & \text{s.t. } F^T F = I, \text{diag}(U) = 0, U \geq 0, U^T \mathbf{1} = 1. \end{aligned} \quad (6)$$

At this point, the terms not connected in (6) are separated, where Y_1, Y_2, Y_3 , and Y_4 are Lagrangian multipliers. Hence, each term's optimal value is obtained in the following order.

3.2.1. \hat{X} Problem. When other variables are fixed, \hat{X} can be obtained by minimizing the following formula:

$$L(\hat{X}) = 2tr(\hat{X}L_W\hat{X}^T) + \langle Y_3, \hat{X} - XZ \rangle + \frac{\mu}{2} \|\hat{X} - XZ\|_F^2. \quad (7)$$

By setting the derivative $\partial L(\hat{X})/\partial \hat{X} = 0$, variable \hat{X} is obtained as follows:

$$\begin{aligned} & 4\hat{X}L + \mu(\hat{X} - XZ) + Y_3 = 0, \\ & \hat{X} = (\mu XZ - Y_3)(4L + \mu I)^{-1}. \end{aligned} \quad (8)$$

3.2.2. *Z Problem.* When other variables are fixed, Z can be obtained by minimizing the following formula:

$$\begin{aligned} & \min_Z \langle Y_1, Z - S \rangle + \langle Y_2, X - XZ - E \rangle + \langle Y_3, \hat{X} - XZ \rangle + \langle Y_4, Z - U \rangle \\ & + \frac{\mu}{2} (\|Z - S\|_F^2 + \|X - XZ - E\|_F^2 + \|\hat{X} - XZ\| + \|Z - U\|_F^2). \end{aligned} \quad (9)$$

$$Z = (I + X^T X)^{-1} \frac{((1/\mu)X^T(Y_2 + Y_3) - ((Y_1 + Y_4)/\mu) + S + U + X^T X + X^T \hat{X} - X^T E)}{2}. \quad (10)$$

3.2.3. *S Problem.* S can be achieved by solving the following problem with other variables fixed

$$S = \arg \min_S \frac{\lambda_1}{\mu} \|S\|_* + \frac{1}{2} \left\| S - \left(Z + \frac{Y_1}{\mu} \right) \right\|_F^2. \quad (11)$$

Then, S is obtained as follows by using the singular value thresholding (SVT) operator [31]:

$$S = S_{\lambda_1/\mu} \left[Z + \frac{Y_1}{\mu} \right]. \quad (12)$$

3.2.4. *E Problem.* E is obtained by solving the following minimization problem:

$$E = \arg \min_E \lambda_2 \|E\|_{2,1} + \frac{\mu}{2} \left\| E - \left(X - XZ + \frac{Y_2}{\mu} \right) \right\|_F^2. \quad (13)$$

Denote $T = (X - XZ + (1/\mu)Y_2)$, $\tau = \lambda_2/\mu$.

The i -th column of E is

$$E(:, i) = \begin{cases} \frac{\|t_i\| - \tau}{\|t_i\|} t_i, & \text{if } \tau < \|t_i\|, \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

3.2.5. *U Problem.* U is obtained by solving the following minimization problem:

$$\min_U 2\lambda_3 \text{Tr}(F^T L_U F) + \langle Y_4, Z - U \rangle + \frac{\mu}{2} \|Z - U\|_F^2. \quad (15)$$

For simplicity, $\sum_{i,j} \|F_i - F_j\|^2 U_{i,j}$ is equivalent to $\text{Tr}(G^T U)$, where $G_{ij} = \|F_i - F_j\|_2^2$; then, we rewrite (15) as

$$\min_U \text{Tr}(G^T U) + \frac{\mu}{2} \left\| U - \left(Z + \frac{Y_4}{\mu} \right) \right\|_F^2. \quad (16)$$

Denoting $N = Z + (Y_4/\mu)$, we get

$$\begin{aligned} & \min_{U_i} \|U_i - (N_i - G_i/\mu)\|_2^2 \\ & \text{s.t. } \text{diag}(U) = 0, U \geq 0, U^T \mathbf{1} = 1. \end{aligned} \quad (17)$$

By setting the derivative $\partial L(Z)/\partial Z = 0$, variable Z is obtained as follows:

Considering the previously mentioned constraints, we have the following:

$$\min_{U_i} \sum_j \left\| U_i - \frac{(N_i - G_i/\mu)}{2\lambda_4} \right\|_2^2 - \eta(1^T U_i - 1) - \zeta^T U_i \quad (18)$$

s.t. $U_{ii} = 0$.

Then, taking the derivative with respect to U_i and setting it to zero, that is,

$$U_i - \frac{1(N_i - G_i/\mu)}{2\lambda_4} - \eta \mathbf{1} - \zeta = 0. \quad (19)$$

The j entry of U_i is shown as follows:

$$U_{ij} - \frac{N_{ij} - G_{ij}/\mu}{2\lambda_4} - \eta - \zeta_i = 0. \quad (20)$$

According to KKT conditions,

$$U_{ij} = \left(\frac{N_{ij} - G_{ij}/\mu}{2\lambda_3} + \eta \right)_+. \quad (21)$$

3.2.6. *F Problem.* When other variables are fixed, F can be obtained by solving the following minimization problem:

$$L(F) = \min_F \text{Tr}(F^T L_U F) \quad (22)$$

s.t. $F^T F = I$,

where L_U is the Laplacian matrix of U . This problem can be simply solved via eigenvalue decomposition, and its solutions are the set of c eigenvectors corresponding to the first c smallest eigenvalues of L_U .

The complete solution is given in Algorithm 1.

3.3. *Computational Complexity.* The inverse of a matrix, eigen decomposition, and SVT's nuclear norm minimization problem are three key aspects that define the computing cost of our proposed algorithm. Specifically, the eigen decomposition costs $O(n^2c)$, where c is the number of clusters. The inverse of a matrix and nuclear norm minimization cost is $O(n^3)$ each. As a result, the total computational complexity of each iteration is

Input: Data X , clusters size c , parameter $\lambda_1, \lambda_2, \lambda_3$
 Initialize: $\hat{X} = Z = S = E = U = F = 0, Y_1 = 0, Y_2 = 0, Y_3 = 0, Y_4 = 0; \rho = 1.01; \mu = 0.01; \mu_{\max} = 10^8; \epsilon = 10^{-5};$
While not converged **do**
 (1) Update \hat{X} by equation (8);
 (2) Update Z by equation (10);
 (3) Update S by equation (12);
 (4) Update E by equation (14);
 (5) Update U by equation (21);
 (6) Update F by equation (22);
 $Y_1 = Y_1 + \mu(Z - S)$
 $Y_2 = Y_2 + \mu(X - XZ - E)$
 $Y_3 = Y_3 + \mu(\hat{X} - XZ)$
 $Y_4 = Y_4 + \mu(Z - U)$
 (11) Update μ by $\mu = \min(\rho\mu, \max(\mu))$
 (12) Check the convergence conditions:
 $\|Z - S\|_{\infty} < \epsilon, \|X - XZ - E\|_{\infty} < \epsilon, \|\hat{X} - XZ\|_{\infty} < \epsilon$ and $\|Z - U\|_{\infty} < \epsilon$
Output: U

ALGORITHM 1: The proposed algorithm.

TABLE 1: Summary of the datasets.

Dataset	Class	Dimension	Number of samples
COIL20	20	1024	1440
UCI	10	240	2000
ORL	40	4096	400
FERET	200	1024	1400
BBC	5	4659	685



FIGURE 1: Example images of (a) ORL, (b) FERET, and (c) UCI datasets.

$O(n^3 + 2n^2c)$ because the cost of computing multipliers and basic matrix operations like addition, subtraction, and division are negligible compared with the cost of other operations.

4. Experiments

4.1. *Experimental Settings.* In order to demonstrate the effectiveness of the proposed method, several experiments

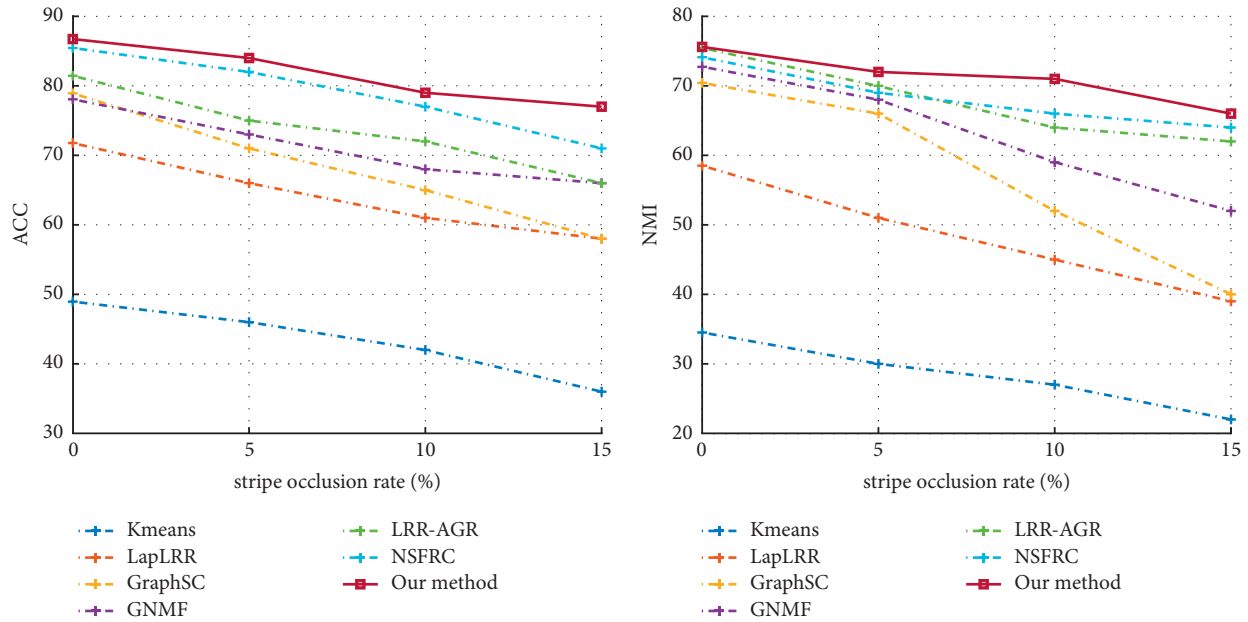


FIGURE 2: Clustering performances of different algorithms on corrupted ORL dataset.

TABLE 2: Clustering results of different algorithms on Coil 20 dataset.

	ACC	NMI	PUR
Kmeans	67.78	83.47	78.91
LapLRR	66.06	79.14	82.23
GraphSC	82.12	92.05	79.51
GNMF	81.63	90.31	82.03
LRR_AGR	85.12	95.33	85.14
NSFRC	88.56	93.12	87.67
Proposed method	88.70	94.39	88.47

TABLE 5: Clustering results of different algorithms on FERET dataset.

	ACC	NMI	PUR
Kmeans	43.23	22.71	30.38
LapLRR	55.13	42.94	52.73
GraphSC	49.34	44.15	50.66
GNMF	55.30	49.28	53.31
LRR_AGR	60.66	53.23	52.34
NSFRC	61.96	52.87	52.15
Proposed method	63.28	53.77	53.52

TABLE 3: Clustering results of different algorithms on UCI dataset.

	ACC	NMI	PUR
Kmeans	75.58	72.71	58.44
LapLRR	81.73	78.92	72.11
GraphSC	80.12	81.56	81.58
GNMF	82.12	79.96	82.72
LRR_AGR	82.77	83.52	86.47
NSFRC	83.21	88.56	88.71
Proposed method	85.53	90.36	87.05

TABLE 6: Clustering results of different algorithms on BBC dataset.

	ACC	NMI	PUR
Kmeans	38.12	10.91	15.66
LapLRR	55.12	31.55	37.83
GraphSC	60.27	44.62	38.91
GNMF	59.65	43.92	45.08
LRR_AGR	61.23	47.12	36.12
NSFRC	53.67	45.23	41.97
Proposed method	62.13	48.23	42.39

TABLE 4: Clustering results of different algorithms on ORL dataset.

	ACC	NMI	PUR
Kmeans	48.95	34.51	13.63
LapLRR	71.77	58.51	53.13
GraphSC	78.92	70.43	65.56
GNMF	78.06	72.73	69.31
LRR_AGR	81.45	75.45	71.22
NSFRC	85.45	74.12	72.34
Proposed method	86.73	75.61	73.87

were performed on COIL20 (<https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php>), UCI (<https://archive.ics.uci.edu/ml/datasets/>

Optical+Recognition+of+Handwritten+Digits), ORL (<http://cam-orl.co.uk/facedatabase.html>), FERET (<https://www.nist.gov/itl/products-and-services/color-FERET-database>), and BBC (<http://mlg.ucd.ie/datasets/segment.html>) datasets (See Table 1 for a summary of each dataset and Figure 1 for images of some datasets). Thus, using accuracy (ACC), normalized mutual information (NMI), and purity (PUR) metrics, the performance of the proposed

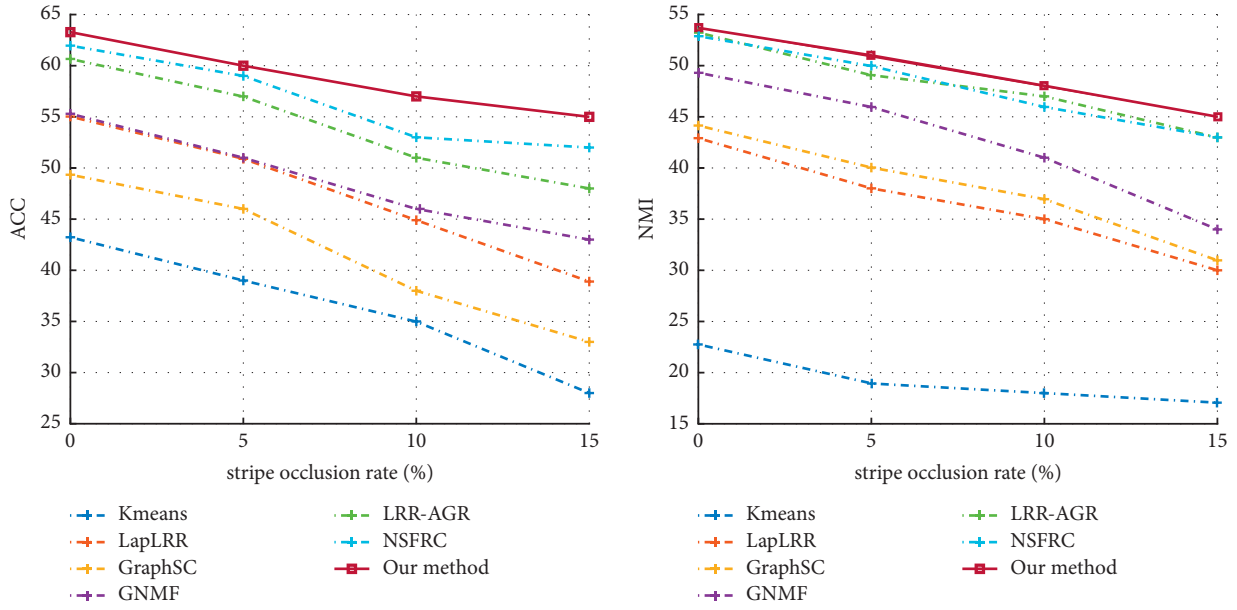


FIGURE 3: Clustering performances of different algorithm on corrupted FERET dataset.

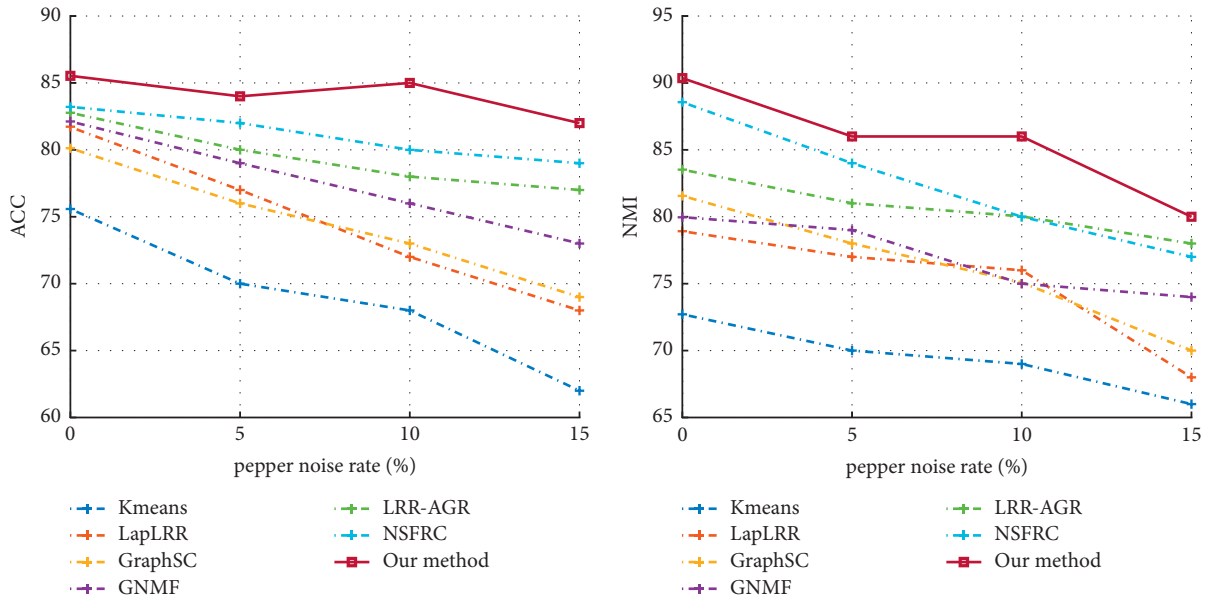


FIGURE 4: Clustering performances of different algorithm on corrupted UCI dataset.

method was evaluated in comparison with Kmeans, LapLRR [9], GraphSC [4], GNMf [5], LRR_AGR [11], and non-negative self-representation (NSFRC) [32] state-of-the-art methods. For each method, the parameter settings in the corresponding literature were adopted. Therefore, each of these methods are described as follows.

LapLRR [9]: it uses LRR’s nuclear norm strategy to capture data’s intrinsic nonlinear geometric information

GraphSC [4]: it considers the local manifold structure of the data to learn a sparse representation

GNMF [5]: it constructs an affinity graph to encode the geometrical information of data

LRR-AGR [11]: it learns optimal clustering by considering the local data structure that LRR previously ignored

NSFRC [32]: it uncovers data’s intrinsic structure by joint nonnegative self-representation and adaptive distance regularization

4.2. Experimental Results. This section analyzes the results of various experiments conducted to evaluate the proposed method’s effectiveness. Note that the best results are highlighted in bold fonts in the tables

4.2.1. Clustering Performance on Original Data. Tables 2–6 display the clustering results concerning ACC, NMI, and PUR of different algorithms on COIL20, UCI,

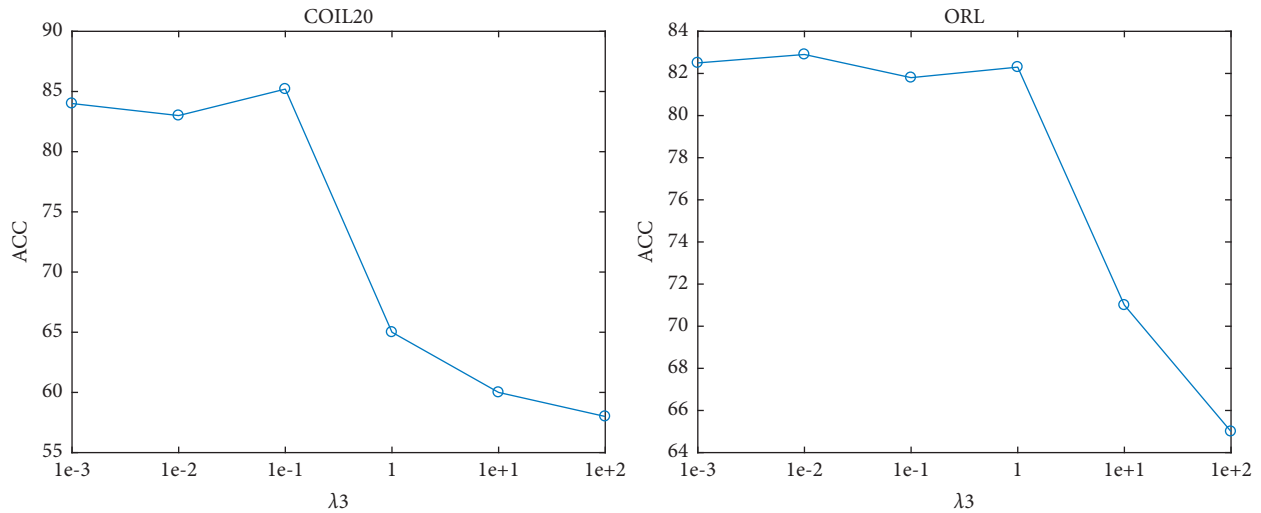


FIGURE 5: The parameter study of λ_3 on COIL20 and ORL.

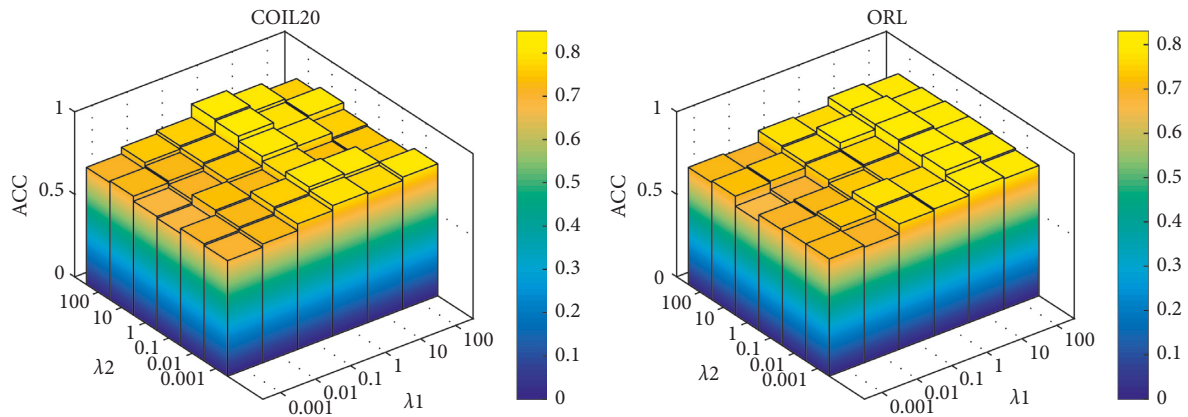


FIGURE 6: The parameter study of λ_2 and λ_3 on COIL20 and ORL.

ORL, FERET, and BBC benchmark datasets, respectively. On COIL20 dataset, the proposed method has the best result in all evaluation metrics, followed closely by NSFRC in ACC and LRR_AGR in NMI. This performance is not surprising because LRR_AGR considers the local data structure of the data to improve its performance, while NSFRC, on the other hand, uses an adaptive affinity matrix learning approach to uncover the intrinsic structure of data to boost clustering performance. Besides, it is also not shocking that Kmeans has the worst performance on COIL20 dataset because it performs clustering directly on the original data, promoting noise interference on the clustering structure. Furthermore, one may observe that Kmeans has its best performance on the UCI dataset compared to other datasets. However, its performance is far below that of others methods, especially the proposed method and NSFRC. It can also be observed that NSFRC and LRR_AGR performances are consistently close to the proposed method on ORL, FERET, and BBC datasets. Specifically, on ORL datasets, both NSFRC and LRR_AGR have a performance of 85.45% and 81.45%, respectively, in ACC, which is lesser than that of the proposed method with 1.28% and 5.28%, respectively. This

performance is also maintained in the relatively difficult datasets of FERET and BBC. Overall, the proposed method has more excellent performance to demonstrate the effectiveness of the regularization term in our model.

4.2.2. Clustering Performance on Data Corrupted with Pepper Noise and Stripe Occlusion. In this section, several experiments are performed to evaluate the robustness of each algorithm against noise and occlusion. To perform this experiment, two settings were adopted. First, stripe occlusion of various degrees (0, 5%, 10%, and 15%) were randomly applied to ORL and FERET datasets. Second, pepper noise of various degrees (the same as previously mentioned) were randomly applied to the UCI dataset. Therefore, Figures 2–4 present each algorithm’s performance on corrupted ORL, FERET, and UCI datasets, respectively. Although it can be seen clearly from these figures that all methods have performance degradation as the corruption level increases, our proposed method shows more robustness than other methods on the three datasets. More specifically, the proposed method’s performance maintained a

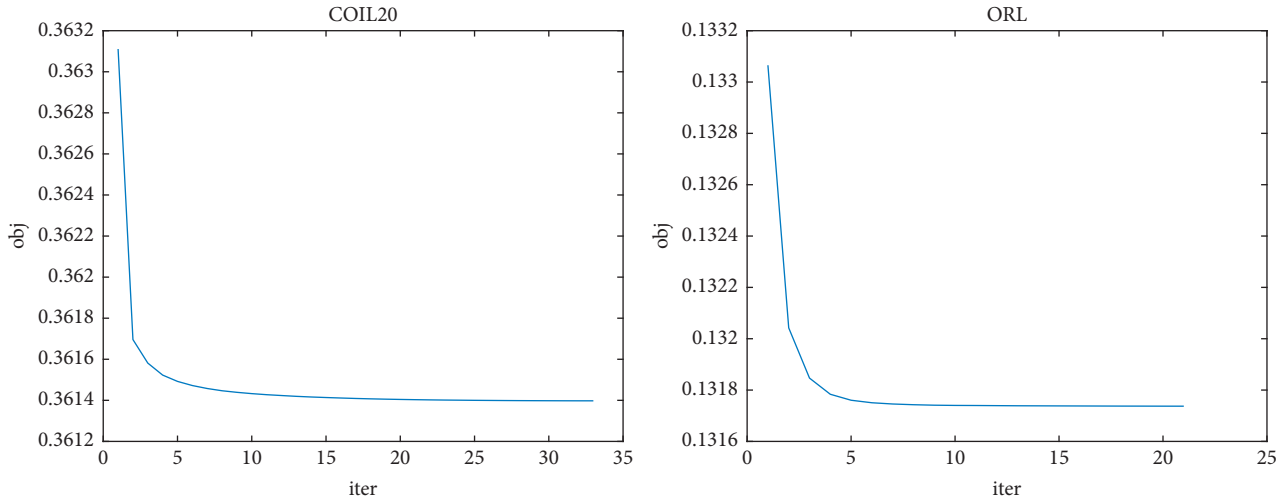


FIGURE 7: The convergence curve of the proposed method on COIL20 and ORL.

TABLE 7: Computation time (in seconds) of each method on three benchmark datasets. The experiments are conducted on an eight-core windows with each core's CPU at 4.2 GHz; the total memory is 16 G.

	COIL20	UCI	BBC
Kmeans	7.18	3.72	6.41
LapLRR	92.13	63.57	75.93
GraphSC	53.54	32.18	42.42
GNMF	78.49	58.81	68.35
LRR_AGR	77.21	65.51	72.61
NSFRC	120.15	87.52	81.64
Proposed method	88.93	67.59	73.46

steadier drop on the ORL dataset than a sharp one obtained by the NSFRC method in the ACC evaluation metric. It can also be observed that its performance in NMI is better with a 10% degree of stripe occlusion than in the 5% case, which confirms that the proposed method can guarantee superior robustness to a large extent. Besides, its performance on the UCI dataset under the random pepper noise corruption further validate above with a clear improvement over the other compared methods.

4.3. Parameter Sensitivity. This section uses some experiments to show the sensitivity of parameters concerning clustering ACC of the proposed method. We can find that there are three regularization parameters, that is, λ_1 , λ_2 , and λ_3 , needed to be set in advance. These parameters, respectively, balance the importance of low-rank constraint term, error term, and rank constraint. Generally, the larger the parameter value is, the more important or impact the corresponding term is. To demonstrate the effects of these three parameters for data clustering, a candidate parameter range set of $\{10 - 3, 10 - 2, 10 - 1, 1, 10 + 1, 10 + 2\}$ is first defined for the three parameters, and then, the proposed method is performed with different combinations of the parameters for data clustering. We first fixed parameters λ_1 and λ_2 and then executed the proposed method with

different parameter values of λ_3 to show the influence on the clustering ACC.

From Figure 5, it is obvious to see that the clustering ACC is insensitive to parameter λ_3 when $\lambda_3 \leq 0.01$. This is mainly because if parameter λ_3 is too large, the corresponding rank constraint term will play the dominant role in the graph learning while ignoring the local and global structure preservation. In this case, although the obtained graph still has corresponding connected components, it cannot reveal the intrinsic structure of data. In the experiments, we can select a small value in the range of $\{10 - 3, 10 - 2, 10 - 1\}$ for parameter λ_3 . Figure 6 shows the clustering ACC versus different values of parameters λ_1 and λ_2 when parameter λ_3 is fixed. As shown in this figure, the clustering ACC is sensitive to parameter λ_2 to some extent, and the best clustering result can be obtained when parameters λ_1 and λ_2 are in a feasible range. This is mainly because a very large or very small parameter λ_2 leads to a small error or large error that cannot compensate well for the sparse data noise. In this case, the model cannot learn the intrinsic similarity graph for data clustering. Thus, in the experiments, we can select the two parameters in the candidate range of $\{10 - 1, 1, 10, 10 + 2\}$ according to the degree of noise corruptions of data.

As far as we know, it is still an open problem to adaptively select these optimal parameters for different datasets. In the experiments, we first fix parameter λ_3 since this parameter is insensitive to the clustering ACC and then perform the method to find the optimal λ_1 and λ_2 in a candidate domain where the optimal parameters may exist. Then, by a similar strategy, we fix parameters λ_1 and λ_2 to find the optimal value of parameter λ_3 in a candidate domain. Finally, the optimal combination of these parameters can be obtained in the 3D candidate space, composed of three candidate domains of parameters.

4.4. Convergence Study. The convergence of the ADMM-style algorithm with two blocks has been generally proven in [29]. There are six blocks (including \tilde{X}, Z, S, E, U, F) in

Algorithm 1, and the objective function is not smooth; it would not be easy to prove the convergence in theory. For \hat{X} , Z , and U , their subproblems have clear analytical solutions. With continuous iteration, they can be expected to reach the optimal solution. The subproblems of F can be solved by solving the characteristic equations. For the S subproblem, the SVT [33, 34] method is used to solve it. For the E subproblem, the effectiveness and convergence of the solution have also been confirmed [35]. Therefore, we can expect the algorithm to reach a local optimal solution. In Figure 7, we further verify the algorithm's convergence on the actual dataset.

Furthermore, Table 7 lists each method's average computation runtime on three benchmark datasets. According to the record, it can be observed that although the computational time of the proposed method is not the most efficient, it is at the same level as most methods.

5. Conclusion

A novel method is proposed in this study that injects a regularization term into LRR to ensure that the neighborhood structure of data samples in the original high-dimensional space is also maintained in the low-dimensional space. With this strategy, the proposed method guarantees robust clustering performance that resolves the limitation of most existing methods. This fact is demonstrated experimentally, with several results showing that the proposed method substantially outperformed similar state-of-the-art methods in ACC, NMI, and PUR evaluation metrics. Thus, our approach will be extended to multiview learning in future work based on several works [36], which shows that it can improve single view learning models.

Data Availability

All datasets used in this paper are open source, meaning they are freely available for research purposes. They can be accessed using the following links. COIL-20: <https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php>; UCI: <https://archive.ics.uci.edu/ml/datasets/Optical+Recognition+of+Handwritten+Digits/>; ORL: <http://cam-orl.co.uk/facedatabase.html>; FERET: <https://www.nist.gov/itl/products-and-services/color-FERET-database>; and BBC: <http://mlg.ucd.ie/datasets/segment.html>.

Conflicts of Interest

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

Acknowledgments

This work was funded in part by the National Natural Science Foundation of China (no.61572240).

References

- [1] K. Alsabti, S. Ranka, and V. Singh, "An efficient k-means clustering algorithm," *Electrical Engineering and Computer Science*, vol. 43, 1997.
- [2] M. C. Naldi, R. J. Campello, E. R. Hruschka, and A. C. P. L. F. Carvalho, "Efficiency issues of evolutionary k-means," *Applied Soft Computing*, vol. 11, no. 2, pp. 1938–1952, 2011.
- [3] R. Kalam and K. Manikandan, "Enhancing k-means algorithm for image segmentation," in *Proceedings of the 2011 International Conference on Process Automation, Control and Computing*, pp. 1–4, IEEE, Coimbatore, India, July 2011.
- [4] M. Zheng, J. Bu, C. Chen et al., "Graph regularized sparse coding for image representation," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 20, no. 5, pp. 1327–1336, 2011.
- [5] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548–1560, 2011.
- [6] S. E. Abhadiomhen, Z. Wang, and X. Shen, "Coupled low rank representation and subspace clustering," *Applied Intelligence*, vol. 52, pp. 1–17, 2021.
- [7] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.
- [8] E. Elhamifar and R. Vidal, "Sparse subspace clustering: algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.
- [9] M. Yin, J. Gao, and Z. Lin, "Laplacian regularized low-rank representation and its applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 3, pp. 504–517, 2016.
- [10] H. Zhang, H. Zhai, L. Zhang, and P. Li, "Spectral-spatial sparse subspace clustering for hyperspectral remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3672–3684, 2016.
- [11] J. Wen, X. Fang, Y. Xu, C. Tian, and L. Fei, "Low-rank representation with adaptive graph regularization," *Neural Networks*, vol. 108, pp. 83–96, 2018.
- [12] C. G. Li, C. You, and R. Vidal, "Structured sparse subspace clustering: a joint affinity learning and subspace clustering framework," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2988–3001, 2017.
- [13] R. Zheng, Z. Liang, X. Chen, Y. Tian, C. Cao, and M. Li, "An adaptive sparse subspace clustering for cell type identification," *Frontiers in Genetics*, vol. 11, p. 407, 2020.
- [14] C. G. Li, C. You, and R. Vidal, "On geometric analysis of affine sparse subspace clustering," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1520–1533, 2018.
- [15] J. Wang, X. Wang, F. Tian, C. H. Liu, and H. Yu, "Constrained low-rank representation for robust subspace clustering," *IEEE Transactions on Cybernetics*, vol. 47, no. 12, pp. 4534–4546, 2017.
- [16] J. Wang, D. Shi, D. Cheng, Y. Zhang, and J. Gao, "LRSR: low-rank-sparse representation for subspace clustering," *Neurocomputing*, vol. 214, pp. 1026–1037, 2016.
- [17] J. Chen, H. Zhang, H. Mao, Y. Sang, and Z. Yi, "Symmetric low-rank representation for subspace clustering," *Neurocomputing*, vol. 173, pp. 1192–1202, 2016.

- [18] H. Zhai, H. Zhang, L. Zhang, and P. Li, "Laplacian-regularized low-rank subspace clustering for hyperspectral image band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1723–1740, 2019.
- [19] X. Guo, X. Xie, G. Liu, M. Wei, and J. Wang, "Robust low-rank subspace segmentation with finite mixture noise," *Pattern Recognition*, vol. 93, pp. 55–67, 2019.
- [20] W. Sun, J. Peng, G. Yang, and Q. Du, "Fast and latent low-rank subspace clustering for hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 3906–3915, 2020.
- [21] K. F. Hui, X. J. Shen, S. E. Abhadiomhen, and Y. Z. Zhan, "Robust low-rank representation via residual projection for image classification," *Knowledge-Based Systems*, vol. 241, Article ID 108230, 2022.
- [22] W. Gao, X. Li, S. Dai, X. Yin, and S. E. Abhadiomhen, "Recursive sample scaling low-rank representation," *Journal of Mathematics*, vol. 202114 pages, Article ID 2999001, 2021.
- [23] X. Chang, F. Nie, S. Wang, Y. Yang, X. Zhou, and C. Zhang, "Compound rank- k projections for bilinear analysis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 7, pp. 1502–1513, 2016.
- [24] M. Luo, X. Chang, L. Nie, Y. Yang, A. G. Hauptmann, and Q. Zheng, "An adaptive semisupervised feature analysis for video semantic recognition," *IEEE Transactions on Cybernetics*, vol. 48, no. 2, pp. 648–660, 2018.
- [25] M. Liu, Y. Wang, J. Sun, and Z. Ji, "Adaptive low-rank kernel block diagonal representation subspace clustering," *Applied Intelligence*, vol. 52, no. 2, pp. 2301–2316, 2021.
- [26] C. Lu, J. Feng, Z. Lin, T. Mei, and S. Yan, "Subspace clustering by block diagonal representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 487–501, 2019.
- [27] C. Yan, X. Chang, M. Luo et al., "Self-weighted robust LDA for multiclass classification with edge classes," *ACM Transactions on Intelligent Systems and Technology*, vol. 12, no. 1, pp. 1–19, 2021.
- [28] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2010, <https://arxiv.org/abs/1009.5055>.
- [29] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," *Journal of Structural Biology*, vol. 181, pp. 116–127, 2013.
- [30] K. Fan, "On a theorem of weyl concerning eigenvalues of linear transformations," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 36, no. 1, pp. 31–35, 1950.
- [31] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1615–1622, IEEE, Barcelona, Spain, November 2011.
- [32] G. Zhong and C. M. Pun, "Nonnegative self-representation with a fixed rank constraint for subspace clustering," *Information Sciences*, vol. 518, pp. 127–141, 2020.
- [33] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, pp. 1956–1982, 2010.
- [34] S. E. Abhadiomhen, R. C. Nzeh, E. D. Ganaa, H. C. Nwagwu, G. E. Okereke, and S. Routray, "Supervised shallow multi-task learning: analysis of methods," *Neural Processing Letters*, vol. 2022, pp. 1–18, 2022.
- [35] J. Yang, W. Yin, Y. Zhang, and Y. Wang, "A fast algorithm for edge-preserving variational multichannel image restoration," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 569–592, 2009.
- [36] K. Zhan, F. Nie, J. Wang, and Y. Yang, "Multiview consensus graph clustering," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1261–1270, 2019.