Robust subspace clustering via two-way manifold representation

Nnamdi Johnson Ezeora[1] · Gregory Emeka Anichebe[1] · Royransom Chiemela Nzeh[1,2] · Izuchukwu Uchenna Uzo[1]

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Subspace clustering has shown great potential in discovering the hidden low-dimensional subspace structures in high-dimensional data. However, most existing methods still face the problem of noise distortion and overlapping subspaces. To tackle this problem and ensure that each sample is only assigned to a single subspace, a new method is proposed in this paper. Specifically, a two-way learning technique is introduced by inducing data manifold via two representative structures. The first is a low-rank structure learned directly from original data. The second structure is an affinity matrix obtained via a k symmetric nearest neighbor graph. By further introducing a dual regularization term, both structures are allowed to guild themselves adaptively to find robust clustering directly without spectral post-processing. In order to evaluate the effectiveness of the proposed method, several experiments are conducted on multiple benchmark datasets. The experimental results, evaluated using six standard metrics, clearly demonstrate that the proposed method outperforms state-of-the-art methods.

1 Introduction

Due to the so-called "curse of dimensionality," many traditional clustering algorithms often do not perform well on high-dimensional data [1]. Thus, subspace clustering (SC) was conceived to handle such a problem. SC operates under the assumption that high-dimensional spaces can be represented as a union of multiple lower-dimensional subspaces, with each subspace corresponding to a different underlying factor or pattern in the data [2].

✉ Gregory Emeka Anichebe
   gregory.anichebe@unn.edu.ng

✉ Izuchukwu Uchenna Uzo
   izuchukwu.uzo@unn.edu.ng

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

2   School of Computer Science and Communication Engineering, JiangSu University, JiangSu 212013, China

The obtained subspace structure is further used as data prior in constructing an affinity matrix for clustering, usually through spectral post-processing. Notably, sparse subspace clustering (SSC) [3] and low-rank representation (LRR) [4, 5] are two popular methods for constructing the affinity matrix. Unlike SSC, which assumes each data point can be represented as a linear combination of a limited number of other data points in the same subspace, LRR believes that the data can be represented as a low-rank matrix. The choice between the two methods often depends on the specific characteristics of the dataset. For example, LRR is exceptional when it comes to dealing with the influence of noise and outliers [6, 7]. Moreover, many researchers have also proposed hybrid methods that combine the sparsity effect of SSC and the global nature of LRR to improve the performance of SC.

Notably, Wang et al. [8] proposed a low-rank subspace sparse representation (LRSR) framework for subspace segmentation, which can recover and segment embedding subspaces simultaneously. Zhu et al. [9] introduced the low-rank sparse subspace (LSS) clustering method. LSS addresses the problems associated with the two-stage approach of traditional graph clustering methods by avoiding spectral post processing. LSS method dynamically learns the affinity matrix from the low-dimensional space of the original data by using a transformation matrix to project the original data to their low-dimensional space, conducting feature selection and subspace learning, and utilizing the rank constraint to obtain the clustering results directly. Sui, Wang and Zhang [10] presented a subspace clustering method that formulates the problem as structured representation learning. Specifically, they discovered that propagating a low-rank structure promotes sparsity in the data representation, resulting in a more robust clustering description. Therefore, their proposed method leverages two cascade self-expressions to implement the propagation. Xia et al. [11] proposed a nonconvex low-rank learning framework, which exploits the low-rank property of nonlinear data and induces a high-dimensional Hilbert space that approaches the true feature space. Recently, Teng et al. [12] introduced a new method, called kernel-based sparse representation learning with global and local low-rank label (KSR-GL^3). KSR-GL^3 uses the global and local low-rank label constraint to ensure the semantic invariance, low-rankness, and discrimination of features during learning. Despite the successes of these existing SC methods, most of them still face the problem of noise distortion and overlapping subspaces, which can result in samples being wrongly assigned to more than one subspace. The ideal is for each sample to only belong to one subspace.

To address the above problem, this paper proposes a new method that induces data manifold via two representative structures. Motivated by the robustness of LRR, we begin by learning a low-rank matrix to characterize the original data and capture its global structure. In order to facilitate our approach, a two-way manifold learning strategy is incorporated into our model such that an affinity matrix is learned simultaneously through a k-symmetric nearest neighbor graph of data. Thus, we further introduced a dual regularization term to enable the LRR and the affinity graph to guide each other adaptively to better capture both global and neighborhood structures of the data together, which leads to more accurate clustering results. By imposing several non-negative constraints and a rank constraint on the affinity matrix, we also avoid spectral post processing procedure, allowing us to obtain the final clustering results directly to prevent extra computational cost. Our main contributions are as follows.

1) We propose a new method that uses two-way representative structures to effectively capture the data manifold and improve clustering performance, unlike, traditional subspace clustering methods, which obtain clustering through one data structure.

2) We avoid spectral post-processing procedures by constraining the learned affinity matrix with a rank constraint and several nonnegative constraints in order to directly derive the clustering matrix.

3) Extensive experiments are performed to evaluate the effectiveness of the proposed method. The results shows that our method performs better than state-of-the-art methods (SOTA) in six evaluation metrics.

The rest of this paper is structured as follows. Section 2 presents the related works. The proposed method formulation is described in Section 3. Results and analysis are provided in Section 4 while Section 5 concludes the paper.

2 Related work

Over the years, several SC approaches have been proposed to tackle the problem of subspace segmentation. These methods can be categorised based on their approach to learning an affinity matrix. In the following paragraphs, we describe those LRR based methods that are closely related to our proposed approach. Firstly, given an unlabelled data matrix X, the traditional LRR model can be formulated as.

$$\min_{A,E} \|A\|_* + \lambda_2 \|E\|_{2,1}, s.t. \ \mathrm{X} = \mathrm{XA} + E, \tag{1}$$

where $X \in \mathbb{R}^{d \times n}$ is used as the self-dictionary to learn a low-rank matrix $A \in \mathbb{R}^{n \times n}$ to characterize the data such that A_{ij}, is only non-zero if samples x_i and x_j are members of the same subspace. $E \in \mathbb{R}^{d \times n}$ is an error matrix to capture the error components since real-world data are not always clean. Thus, considering that data corruption is usually sample specific, the L_{21}-norm is imposed to encourage the columns of E to be zero. The goal is to recover clean data through the product of the original data and a low-rank matrix. That is, using XA while eliminating corrupt samples through the constraint $E = X - XA$. Specifically, in order to reveal the true membership of X, LRR expects that A would have a k block diagonal structure. However, this is difficult to obtain in complex applications, so LRR also expects that when the block diagonal fails to hold, spectral clustering [13] (via a spectral post-processing of A) may ensure the robustness of the segmentation, which is also very uncertain. As a result, various methods have been proposed to strengthen the robustness of LRR.

Structure-constrained LRR (SC-LRR) [14] uses a predefined weight matrix to analyse the structure of multiple disjoint subspaces and reveal their relationship for clustering. Laplacian regularized low-rank representation (LRLRR) [15] incorporates hypergraph Laplacian regularization to capture intrinsic non-linear geometric information in data. LRLRR was motivated by LRR's inability to consider non-linear geometric structures within data, which may result in the missing locality and similarity information during the learning process. Low-rank representation with adaptive graph regularization (LRR_AGR) [16] integrates a distance regularization term and a non-negative constraint to enable simultaneous exploitation of global and local information for graph learning. A novel rank constraint is also introduced to encourage the learned graph to have clear clustering structures. Adaptive Structure-constrained Low-Rank Coding (AS-LRC) [17] combines an adaptive weighting-based block-diagonal structure-constrained low-rank representation and the group sparse salient feature extraction into a unified framework. First, AS-LRC performs latent decomposition of given data into three matrices: a low-rank reconstruction by a block-diagonal codes matrix, a group sparse locality-adaptive salient feature part, and

a sparse error part. The auto-weighting matrix is computed based on the locality-adaptive features, encouraging the codes to be block-diagonal, and avoids the problem of choosing optimal neighbourhood size or kernel width for weight assignment. Furthermore, considering that it is difficult for SC-LRR to determine the best value for the corresponding parameter of the constraint term, an Improved structured low-rank representation (ISLRR) [18] was introduced as an extension of SC-LRR. ISLRR includes the structure information of datasets into the equality constraint term of LRR, avoiding the need to adjust the extra parameter. Similarly, SinNLRR [19] uses an adaptive penalty selection method to avoid sensitivity to the parameters. Also, a non-negative and low rank structure is imposed on the affinity matrix to ensure more robust and accurate clustering results.

Recently, some adaptive affinity graphs methods have emerged. For example, Adaptive Low-rank Representation (ALRR) [20] adaptively learns a dictionary from the data to make the obtained model more resistant to the negative impact of noise. Adaptive Data Correction-based Graph Clustering (ADCGC) [21] works to improve clustering performance by removing errors and noise from the original data. Adaptive Graph Regularization from Clean Data (RLRR-AGR) [22] unifies graph construction and subsequent optimization in a framework to capture the data's underlying non-linear geometric information. Despite the effectiveness of these adaptive methods over the fixed graph ones, one drawback is that the data's manifold structure is pre-captured using a non-flexible low-rank matrix, which is a disadvantage. Especially in cases where the data is severely corrupted, resulting in a significant distance between similar samples and causing unrelated samples to remain together. In such cases, the constructed graph may not accurately represent the actual geometric structure of the data. Hence, our proposed method induces the geometric structure of the data via a two-way representative technique to ensure that only similar samples are connected while samples from different classes do not stay together.

3 The proposed method

This section formulates the proposed method and provides an optimization method to solve it.

3.1 Model formulation

To achieve excellent performance in subspace clustering, an accurate k block diagonal subspace structure, with each sample belonging to one subspace is required [23]. But most SC methods are not robust enough to guarantee that because; (1) they focus either on local or the global data structure and simply assume that similar samples habitually reside close to one another (and the corrupt ones always stay far away from others, including their likes) in reconstructing each sample with similar ones using the self-expressiveness property of data. In actuality, two unrelated samples could stay close and be regarded as similar due to corruption causing the data to be arbitrarily distributed. Thus, the subspace structure learned in this way would have negative correlations embedded in it. (2) they construct an affinity matrix from such a subspace structure in advance for subspace clustering through a spectral post-processing procedure. In other words, an error in the first step may impact the final step. To address this, our core idea in this paper is to allow two representative subspace structures to guide each other to simultaneously capture the neighbourhood and global data structures through a dual

regularization term without spectral post-processing procedure. Thus, by adopting the LRR (A) of Eq. (1) as our first structure, we further learn the second structure $B \in \mathbb{R}^{n \times n}$. For this, we use the k-nearest neighbour graph approach to generate a matrix $P \in \mathbb{R}^{n \times n}$ from X. This involves identifying k symmetric nearest neighbours for each sample in X using Euclidean distance as the distance metric (Reference [24] contains a detailed explanation of k-nearest neighbor graph). The spectral clustering algorithm is then applied directly to obtain the affinity matrix B through the low-dimensional embedding matrix of P as follows.

$$\min Tr\left(Q^T P Q\right) + \|B - QQ^T\|_F^2, s.t., Q^T Q = I, B1 = 1, B \geq 0, \tag{2}$$

where constraints $B1 = 1$ and $B \geq 0$ is to ensure that the entries of B are non-negative. The low embedding matrix Q is used to obtain the affinity matrix B because, according to Zhan et al. [25], such an embedding matrix can guarantee a block diagonal structure. Ideally, for such validation, B can be used directly for clustering. However, to further ensure a robust solution, since A and B are obtained independently, we want to align both structures to capture the global and neighbourhood structures of data simultaneously. Thus, we introduce a dual regularization term $\|B - A\|_F^2$, with the Frobenius norm applied to further enforce a block-diagonal solution. Therefore, combining Eq. (1), Eq. (2) and the dual regularization term, we arrive at the following model.

$$\min \|A\|_* + \|B - QQ^T\|_F^2 + Tr\left(Q^T P Q\right) + \|B - A\|_F^2 + \|E\|_{2,1}, \\ s.t., \mathrm{X} = \mathrm{X}A + E, Q^T Q = I, B1 = 1, B \geq 0, \tag{3}$$

As depicted by Eq. (3), A and B matrices are simultaneously learned by allowing them to approximate one another to find more better solution adaptively. Therefore, unlike most existing approach, we avoid the spectral post processing procedure by imposing a rank constraint $rank\left(L_B\right) = n - c$ on the Laplacian matrix of B to allow it to denote our clustering structure using Theorem 1. So, our proposed model described in Fig. 1 is:

$$\min \|A\|_* + \|B - QQ^T\|_F^2 + Tr\left(Q^T P Q\right) + \|B - A\|_F^2 + \lambda_4 \|E\|_{2,1}, \\ s.t., \mathrm{X} = \mathrm{X}A + E, Q^T Q = I, \\ B1 = 1, B \geq 0, rank\left(L_B\right) = n - c, \tag{4}$$

where $L_B = D_B - B$ such that $D_B \in \mathbb{R}^{n \times n}$ denotes the diagonal matrix with ith entry $\sum_j B_{ij} = 1$. And parameters λ_1 and $\lambda_2, \lambda_3, \lambda_4$ are used balance the different terms. To relax Eq. (4), and make it more solvable, we introduce an auxiliary term $S = A$. Additionally, to address the non-linearity of the rank constraint, the method follows a similar strategy as in a previous study [7] to express the smallest eigenvalue of L_B as $\theta_i(L_B)$. Hence, with sufficient λ_5, Eq. (4) can be rewritten as:
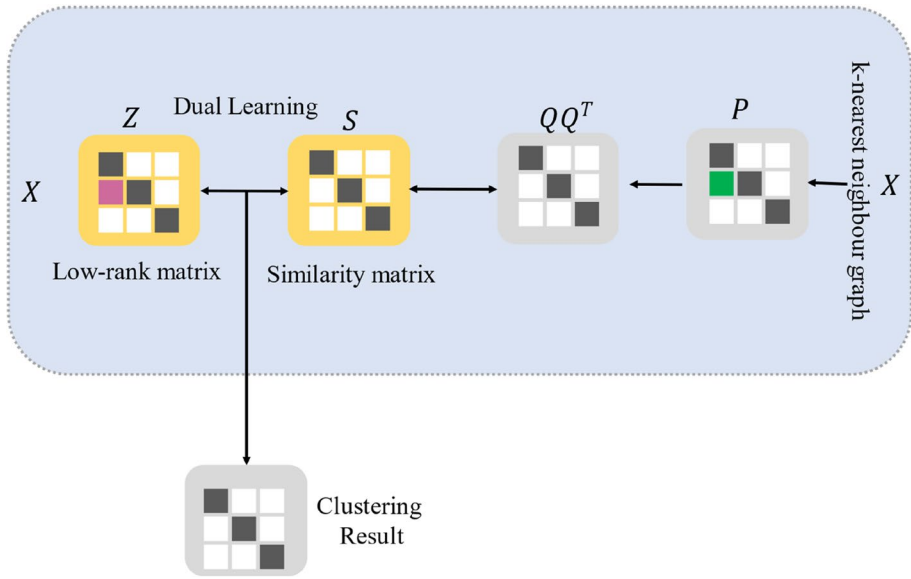
$$\min \|S\|_* + \lambda_1 \|B - QQ^T\|_F^2 + \lambda_2 Tr\left(Q^T P Q\right) + \lambda_3 \|B - A\|_F^2 + \lambda_4 \|E\|_{2,1} + \lambda_5 \sum_{i=1}^{c} \theta_i(L_B), \\ s.t., \mathrm{X} = \mathrm{X}A + E, Q^T Q = I, \\ B1 = 1, B \geq 0, S = A \tag{5}$$

Theorem 1[7]: if B is non-negative, the multiplicity c of the zero eigenvalue of the graph Laplacian L_B corresponds to the number of connected components in the graph associated with B.

Refer to preposition 2 of reference [26] for a detailed proof of the Theorem 1.

Fig. 1 The framework of the proposed method. Firstly, by adopting the LRR A (denoted as Z in the Figure) of Eq. (1) as our first structure, we further learn a second structure B, a similarity matrix (denoted as S in the Figure). For this, we use the *k*-nearest neighbour graph approach to first generate a matrix P, through which we learn the low embedding matrix Q used to obtain B directly

3.2 Optimization

Based on its efficiency, we adopt the Augmented Lagrange Multiplier (ALM) [22, 23] method to solve the objective function of Eq. (5). Firstly, by following Ky Fan's theorem [28], $\sum_{i=1}^{c} \theta_i(L_B)$ is the same as minimizing $Tr(F^T L_B F)$ subject to $F^T F = I$. Thus, we have:

$$
\begin{aligned}
\min \|S\|_* & \\
+\lambda_1 \|B - QQ^T\|_F^2 + \lambda_2 Tr(Q^T PQ) + \lambda_3 \|B - A\|_F^2 + \lambda_4 \|E\|_{2,1} + \lambda_5 Tr(F^T L_B F), & \\
s.t., X = XA + E, Q^T Q = I, & \\
B1 = 1, B \geq 0, S = A, F^T F = I, &
\end{aligned}
\tag{6}
$$

Consequently, we obtain the LaGrange function of Eq. (6) as follows:

$$
\begin{aligned}
\min \|S\|_* + \lambda_1 \|B - QQ^T\|_F^2 + \lambda_2 Tr(Q^T PQ) + \lambda_3 \|B - A\|_F^2 + \lambda_4 \|E\|_{2,1} + \lambda_5 Tr(F^T L_B F), & \\
+ Tr[Y_1^T(X - XA - E)] + Tr[Y_{2,}^T(S - A)] + \frac{\mu}{2}(\|X - XA - E\|_F^2 + \|S - A\|_F^2), &
\end{aligned}
\tag{7}
$$

where Y_1 and Y_2 are Lagrange multipliers, which are necessary for solving constrained problems. Therefore, separating the unconnected terms in Eq. (7), the minimization problem and the ideal solution of each variable are given below in no particular order.

S subproblem:

$$
\underset{S}{\arg\min} \frac{1}{\mu} \|S\|_* + \frac{1}{2} \left\| S - \left(A + \frac{Y_2}{\mu} \right) \right\|_F^2
\tag{8}
$$

The optimal solution to Eq. (8) can be found by first representing the singular value thresholding [27] of $Z_\mu[Y]$ as $U\vartheta_\mu[\sum]V^T$ to obtain the following

$$S = Z_{\frac{1}{\mu}}\left[A + \frac{Y_2}{\mu}\right], \tag{9}$$

where $\vartheta_\mu[\Sigma] = min(0, \Sigma + \mu) + max(0, \Sigma - \mu)$.

E subproblem:

$$E = \underset{E}{\operatorname{argmin}} \frac{\lambda_4}{\mu}\|E\|_{2,1} + \frac{1}{2}\|E - \left(X - XA + \frac{Y_1}{\mu}\right)\|_F^2 \tag{10}$$

Q subproblem

$$\underset{H}{\operatorname{argmin}} \lambda_1\|B - QQ^T\|_F^2 + \lambda_2 Tr(Q^T PQ), \\ s.t., Q^T Q = I. \tag{11}$$

Considering the fact that Eq. (11) is different for each i, we can rewrite it as follows:

$$\underset{Q}{\operatorname{argmax}} Tr(Q^T(P + 2\lambda_1 B)Q), \\ s.t., Q^T Q = I. \tag{12}$$

The optimal value of Q can be found by deriving it from the c eigenvectors of the corresponding topmost c greatest eigenvalues of $(P + 2\lambda_1 B)$.

A subproblem:

$$\underset{A}{\operatorname{argmin}} \lambda_3\|B - A\|_F^2 + Tr[Y_1^T(X - XA - E)] + Tr[Y_2^T(A - S)] + \frac{\mu}{2}\left(\|X - XZ - E\|_F^2 + \|S - A\|_F^2\right). \tag{13}$$

Setting the derivative $\frac{\partial}{\partial A} = 0$, we can obtain A using the following formula.

$$A = \left[\left(\frac{2}{\mu} + 1\right)I + X^T X\right]^{-1}\left((X^T X - X\prime * E + S) + \frac{2\lambda_3 B + X^T Y_1 - Y_2}{\mu}\right). \tag{14}$$

B subproblem

$$\underset{B}{\operatorname{argmin}} \lambda_1\|B - QQ^T\|_F^2 + \lambda_3\|B - A\|_F^2 + \lambda_5 Tr(F^T L_B F), \\ s.t., B1 = 1, B \geq 0, F^T F = I. \tag{15}$$

For simplicity, we rewrite Eq. (15) as:

$$\underset{B}{\min}\|B\|_F^2 - 2\langle\lambda_1 QQ^T + \lambda_3 A, B\rangle + \lambda_5 Tr(F^T L_B F), \\ s.t., B1 = 1, B \geq 0, F^T F = I. \tag{16}$$

By denoting $2(\lambda_1 QQ^T + \lambda_3 A)$ by W, we have:

$$\underset{B}{\operatorname{argmin}}\|B\|_F^2 - Tr(WB^T) + \lambda_5 Tr(F^T L_B F), \\ s.t., B1 = 1, B \geq 0, F^T F = I. \tag{17}$$

Like Eq. (12), the optimal value of F can be found by deriving it from the c eigenvectors of the corresponding topmost c greatest eigenvalues of L_B. Thus, Eq. (18) can be recast as:

$$\min_{b_j}\left(\lambda_5\|f_i - f_j\|_2^2 - w_{ij}\right)b_{ij} + b_j^T b_j,$$
$$s.t., b_j \geq 0, 1^T b_j = 1. \tag{18}$$

Then, denoting $\lambda_5\|f_i - f_j\|_2^2 - w_{ij}$ as q_{ij}, Eq. (19) is simplified as:

$$\min_{b_j}\frac{1}{2}\|b_j + \frac{q_j}{2}\|_2^2,$$
$$s.t., b_j \geq 0, 1^T b_j = 1. \tag{19}$$

Same as Eq. (30) of [28], b_j can be obtained in the following way:

$$b_j = \left(-\frac{w_j}{2} + \eta 1\right)_+. \tag{20}$$

Algorithm 1 outlines the entire solution for the proposed algorithm.

|	
Input: Data X, clusters size c, parameter $\lambda_1, \lambda_2, \lambda_3, \lambda_4$	
Initialize：, $Y_1 = \mathbf{0}, Y_2 = \mathbf{0}$; $\epsilon = 10^{-5}$; $\rho = 1.01$; $\mu_{max} = 10^8$; $\mu = 0.01$	
While not converged **do**	
1	Update S by Eq. (9);
2	Update E by Eq. (10);
3	Update Q by Eq. (12);
4	Update A by Eq. (14);
5	Update B by Eq. (21);
6	Update the multipliers.
	$\begin{aligned} Y_1 &= Y_1 + \mu(X - XA - E) \\ Y_2 &= Y_2 + \mu(S - A) \end{aligned}$
7	Update μ by $\mu = min(\rho\mu, max(\mu))$
8	Check the convergence conditions: $\|X - XA - E\|_\infty < \epsilon$ and $\|S - A\|_\infty < \epsilon$
Output: B	

Algorithm 1 Our Proposed Algorithm

4 Experiments

In this section, the effectiveness of the proposed method is evaluated through several experiments. The experimental settings, including the datasets used, and the compared methods are described in Section 4.1. The results of the experiments and their analysis are presented in Section 4.2.

Table 1 Summary of the datasets

Dataset	No. of Sample	Dimension	Cluster size
COIL20	1440	1024	20
UCI Digits	2000	240	10
ORL	400	1024	40
WebKB	203	1703	4

(a)ORL (b) COIL20

(c) UCI digits
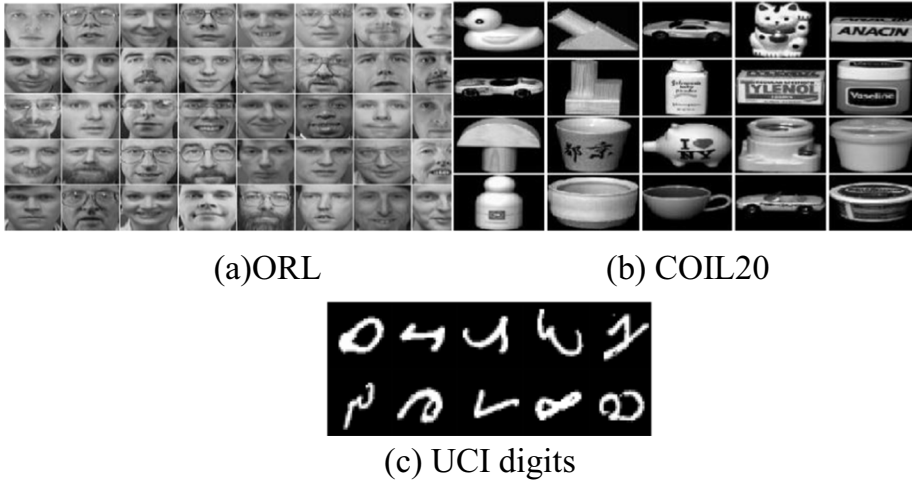
Fig. 2 Illustration images of (a) ORL, (b) COIL20 and (c) USPS datasets

4.1 Experimental settings

To demonstrate the effectiveness of our proposed method, we utilized several datasets: COIL20[1], USPS[2], ORL[3], and WebKB[4] datasets (See Table 1 for a summary of each dataset and Fig. 2 for images of some datasets) to perform various experiments. We compare the performance of our methods with seven state-of-the-art methods: LRR [5], LRR + SSC [29], SinNLRR [19], Implicit Block Diagonal Low-Rank Representation (IBDLR) [30], Nonlinear Orthogonal Non-Negative Matrix Factorization (NONMF) [31], Adaptive Low-Rank (ALR) [32], and Coupled Low-Rank Representation (CLRR) [33] using six standard evaluation metrics: accuracy (ACC), normalized mutual information (NMI), F-score, Recall, Precision and adjusted rand index (AR). For the parameter settings of each method, we adopt the optimal settings in the corresponding literature. We describe the datasets and compared methods below.

---

[1] https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php
[2] https://archive.ics.uci.edu/ml/datasets/Optical+Recognition+of+Handwritten+Digits
[3] http://cam-orl.co.uk/facedatabase.html
[4] https://linqs.soe.ucsc.edu/datal

- COIL20: It consists of a collection of 20 different objects, each captured from various angles under controlled conditions. The dataset contains a total of 1440 samples, where each sample represents an image of one of the objects. We extracted 1024 features for each image.
- UCI Digits: It holds a collection of handwritten digit images, where each image represents one of the ten digits (0–9). The dataset contains a total of 2000 samples, with each sample represented by a feature vector of 240 dimensions.
- ORL: It is a collection of grayscale facial images captured from 40 distinct subjects, with each subject having 10 different images taken under varying conditions, resulting in 400 samples. Each image in the ORL dataset has a dimension of 1024, representing a 32×32 pixel image.
- WebKB: It is a widely used benchmark in the field of text classification and information retrieval. It consists of a collection of web pages from various university websites, with each web page representing a document. The dataset contains a total of 203 samples. Each document in the WebKB dataset is represented by a feature vector with a dimension of 1703.

The following is a description of each comparison method.

- LRR: It aims to find the representation of data samples as linear combinations of bases in a given dictionary with the lowest possible rank.
- LRR + SSC: It considers data correlation and performs automatic data selection while grouping together correlated data.
- SinNLRR: It imposes a nonnegative and low rank structure on the similarity matrix to accurately identify and classify cell types.
- IBDLR: It incorporates implicit feature representation and a block diagonal prior to improve the accuracy of the LRR model.
- NONMF: It extends the nonlinear orthogonal Nonnegative Matrix Factorization (NMF) framework by incorporating a graph regularization.
- ALR: It uses an adaptive solution of a low-rank kernel matrix such that the mapped data in the feature space possesses both low-rank characteristics and self-expression properties.
- CLRR: It uses a k block diagonal regularization term to ensure a block diagonal clustering structure.

4.2 Experimental results

Table 2 shows the result concerning ACC, NMI, F-score, Recall, Precision, and AR obtained by different methods on the COIL20 dataset. As can be seen in the table, the proposed method outperforms the other methods. Specifically, the performance of 88.78% (ACC), 95.19%(NMI), 86.72%(F-score), 76.80%(Precision), and 85.87%(AR) acquired by the proposed method is better than that of the closest method, CLRR, by over 1%, and much better than the other methods. The much closer performance of CLRR is due to its utilization of the k block diagonal regularization, which is shown to be very effective in several works such as [34]. Nonetheless, when the computational runtime is compared, one can easily see that by avoiding the k block diagonal regularization and inducing the dual regularization strategy through the k nearest neighbour graph and avoiding the complexity of manifold recovery structure, the proposed approach is more efficient than CLRR.

Table 2 Clustering results of different algorithms on COIL20 Dataset

	ACC	NMI	F-score	Recall	Precision	AR	Runtime(s)
LRR	65.21	74.41	52.06	68.80	41.87	48.93	638.21
LRR+SSC	60.17	70.66	53.05	57.37	49.34	50.42	1.27e+04
SinNLRR	68.75	79.81	63.55	70.67	57.73	61.45	62.69
IBDLR	84.58	91.84	80.86	88.96	73.10	76.56	401.64
NONMF	66.41	75.71	60.57	65.30	56.53	58.37	236.79
ALR	85.15	89.54	82.42	92.15	72.61	78.92	88.14
CLRR	87.22	94.07	84.97	**96.86**	75.68	84.09	26.45
Ours	**88.78**	**95.19**	**86.72**	96.78	**76.80**	**85.87**	**18.13**

Table 3 Clustering results of different algorithms on UCI Digits

	ACC	NMI	F-score	Recall	Precision	AR	Runtime(s)
LRR	55.10	51.48	52.80	68.67	41.55	40.80	50.90
LRR+SSC	72.14	66.70	63.20	66.22	60.46	58.92	2.24e+03
SinNLRR	87.00	79.40	77.80	78.22	77.38	75.33	2.24e+04
IBDLR	77.85	76.59	73.63	76.67	70.82	70.58	511.80
NONMF	88.30	86.23	80.72	80.34	81.10	76.99	624.34
ALR	90.23	89.80	92.49	91.62	89.40	88.60	211.46
CLRR	**97.25**	**93.86**	94.52	**94.64**	94.40	93.91	61.65
Ours	97.20	93.78	**94.57**	94.60	**94.55**	**93.93**	**31.35**

Table 3 displays the result concerning ACC, NMI, F-score, Recall, Precision, and AR obtained by different methods on the UCI dataset. It can be observed that the proposed method maintained its superiority in handwritten recognition. Although both CLRR and the proposed method obtained competitive performance, with the proposed method outperforming CLRR in only three of the six metrics: F-score, Precision, and AR, the computational runtime of the proposed method is almost half of that of CLRR. Following the performance of CLRR and our method is NONMF and ALR methods. ALR introduces a novel kernel subspace clustering method capable of handling non-linear models, making it a more effective approach for non-linear subspace clustering. NONMF, on the other hand, allows for a factorization that preserves the local geometric structure of the data after nonlinear mapping. In other words, NONMF considers the nonlinearity of the underlying manifold in subspace clustering tasks to enhance its performance.

Similar results are obtained on ORL dataset where CLRR and our method also obtained competitive performance, as shown in Table 4. However, as can be seen in Table 5, the proposed method demonstrates strong performance when compared to its closest competitors, CLRR and IBDLR on the WebKB dataset. In terms of accuracy, the proposed method achieves an impressive 81.77%, surpassing CLRR's 79.80% and outperforming IBDLR's 71.67%. Specifically, the difference between the proposed method and CLRR, which is the closest method is about 2% in ACC, NMI, Precision and AR, respectively, confirming that our dual manifold learning strategy is very effective. In addition, Fig. 3 further demonstrate the effectiveness of the proposed method by showing the scatter plot of the optimized clustering results of different methods on the COIL20 dataset.
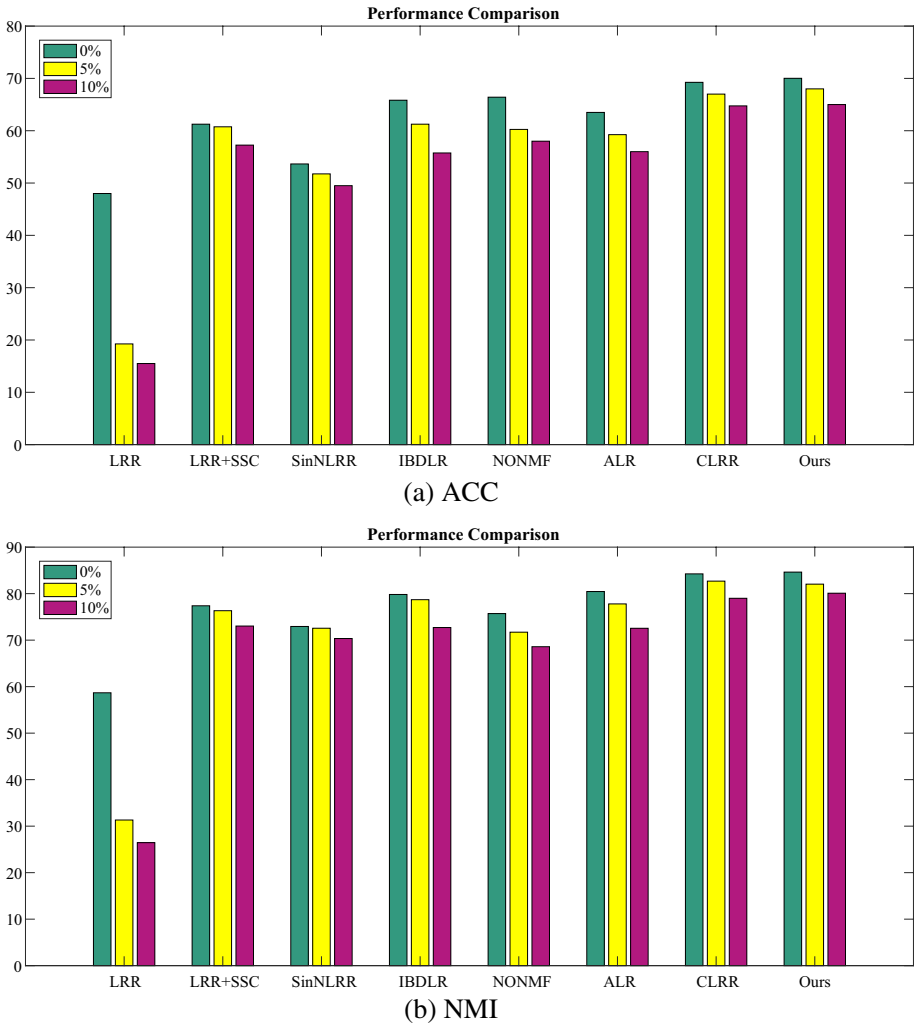
Table 4 Clustering results of different algorithms on ORL Dataset

	ACC	NMI	F-score	Recall	Precision	AR	Runtime(s)
LRR	48.00	58.68	33.99	38.28	30.89	31.14	19.68
LRR + SSC	61.25	77.39	48.70	52.06	45.75	47.43	282.71
SinNLRR	53.65	72.94	40.01	42.37	37.90	38.54	3.89
IBDLR	65.83	79.82	52.40	55.37	49.74	51.24	27.26
NONMF	66.41	75.71	60.57	65.30	56.53	58.37	**0.32**
ALR	63.50	80.45	53.25	59.33	48.30	52.06	3.37
CLRR	69.25	84.25	**59.16**	62.17	**56.43**	**58.17**	1.75
Ours	**70.02**	**84.63**	58.90	**65.39**	54.81	57.95	1.46

Table 5 Clustering results of different algorithms on WebKB Dataset

	ACC	NMI	F-score	Recall	Precision	AR	Runtime(s)
LRR	71.42	79.80	66.98	68.13	65.86	44.97	14.00
LRR + SSC	65.52	79.31	66.71	72.19	62.01	42.30	134.18
SinNLRR	63.05	79.80	64.40	65.19	63.62	40.86	9.64
IBDLR	71.67	78.32	67.70	79.55	58.93	41.09	6.10
NONMF	52.71	55.17	44.59	45.26	43.94	20.80	**0.07**
ALR	52.71	75.86	51.50	42.06	66.40	30.22	0.80
CLRR	79.80	81.28	78.19	89.73	66.65	58.77	1.19
Ours	**81.77**	**83.25**	**78.41**	**91.16**	**68.78**	**60.87**	0.09
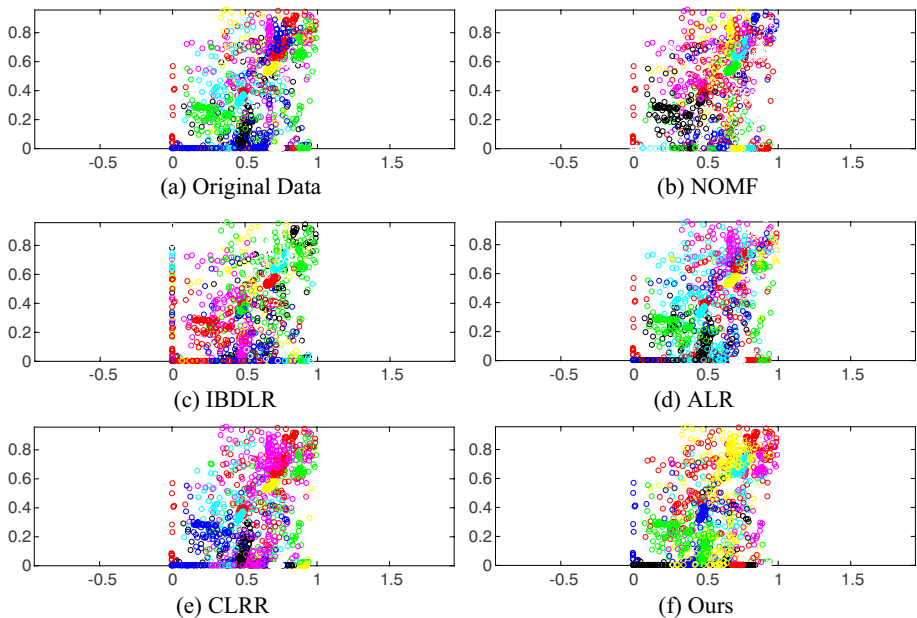
Fig. 3 A scatter plot of the optimized clustering results of different methods on the COIL20 dataset

(a) Original Data
(b) NOMF
(c) IBDLR
(d) ALR
(e) CLRR
(f) Ours

Fig. 4 Clustering performances of different algorithms on corrupted ORL dataset

Furthermore, considering that the degree of corruption in real-world data is not usually known in advance [35], the robustness of various algorithms to corruption is further investigated by adding different degrees (5% and 10%) of occlusion into the arbitrarily selected ORL dataset. Precisely, we randomly introduced some black blocks into the images of ORL. According the results shown Fig. 4, all methods have performance degradation as the corruption level increases. The reason is straightforward, because, as the sizes of occlusion grows, it tends to obscure the discriminative details present in the data, making it challenging for various clustering methods to accurately identify the underlying structure. The presence of occlusion can lead to misinterpretation and incorrect clustering assignments. However, the proposed method exhibits a relatively higher level of robustness compared to the other methods that were compared. This is because the proposed method is able to
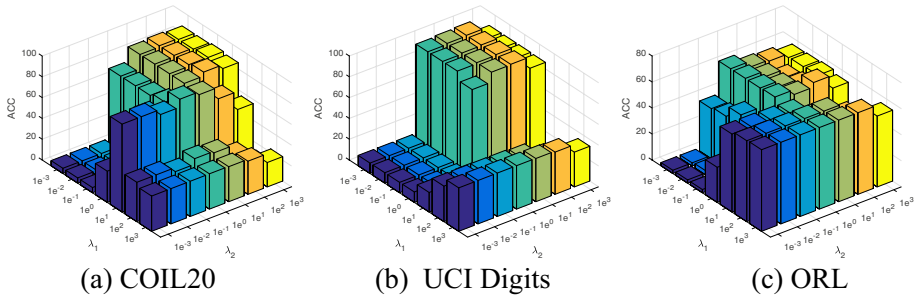
Fig. 5 Visualization of the clustering performance of our method w.r.t. ACC by fixing other parameters and varying-parameters λ_1 and λ_2 on COIL20, UCI Digits, and ORL datasets
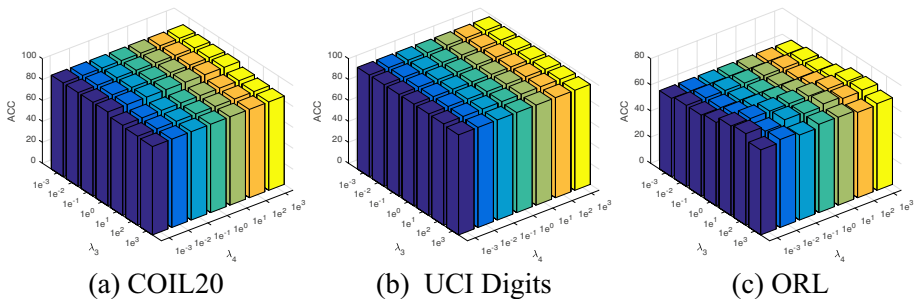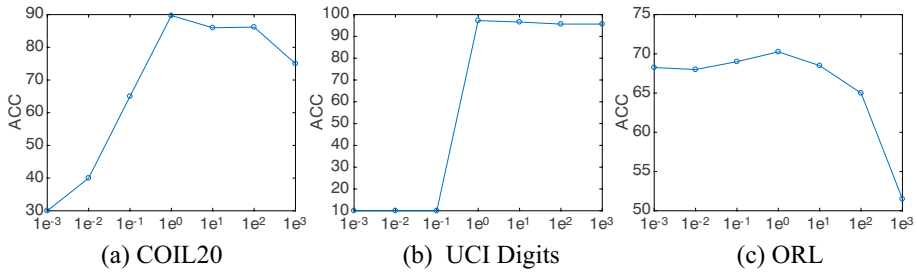

Fig. 6 Visualization of the clustering performance of our method w.r.t. ACC by fixing other parameters and varying-parameters λ_3 and λ_4 on COIL20, UCI Digits, and ORL datasets

mitigate the detrimental effects of data corruption to a greater extent via the dual regularization term, allowing it to generate more accurate and reliable clustering results.

4.3 Parameter sensitivity

This section presents experimental results to examine the sensitivity of parameters related to ACC of the proposed clustering method using COIL20, UCI Digits and ORL datasets. The objective function of the method includes five regularization parameters: λ_1, λ_2, λ_3, λ_4 and λ_5 which needs to be set beforehand. These parameters determine the impact of the various terms in our objective function. In order to investigate their effects on data clustering, the study defines a range of candidate values: {0.001, 0.01, 0.1, 1, 10, 100, 1000}, for each parameter by applying the proposed method with various combinations. Firstly, we fix other parameters to 1 while varying λ_1 and λ_2 to observe its influence on the clustering accuracy (ACC). Then, by fixing λ_1 and λ_2 to their optimal value from the previous stage, we vary λ_3 and λ_4 to understand how the different combinations affect the accuracy of the clustering results. Finally, we vary λ_5 by fixing all the other parameters to obtain the final optimal sets. Although this process seems expensive, it is reasonable in practice. Moreover, to the best of our knowledge, how to select optimal parameters for different dataset is still an open problem in the literature.

Fig. 7 Visualization of the clustering performance of our method w.r.t. ACC by fixing other parameters and varying-parameters λ_5 on COIL20, UCI Digits, and ORL datasets

Fig. 8 The convergence curve of the proposed method on the four datasets
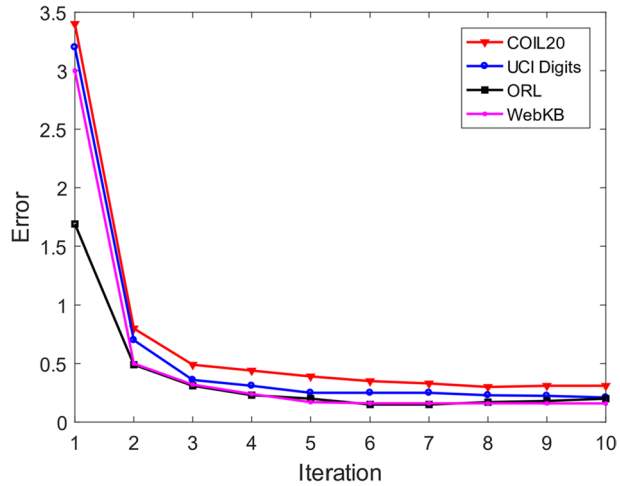
From Fig. 5, 6, it is obvious to see that the clustering ACC is relatively insensitive to parameter λ_3 and λ_4, especially on COIL20 and UCI Digits datasets. Figure 7 indicates that the clustering accuracy (ACC) is not significantly affected by the parameter λ_5. This insensitivity is primarily attributed to the behaviour of the rank constraint term associated with λ_5 in the graph learning process. Because when the value of λ_5 is too large, the rank constraint term becomes dominant in the graph learning, overshadowing the preservation of the local and global structure. As a result, even though the generated graph may still have connected components, it fails to reveal the actual underlying intrinsic structure of the data.

4.4 Convergence study

The convergence proof of the ALM algorithm typically involves establishing mathematical guarantees that the algorithm will converge to an optimal solution under certain conditions. Notably, the convergence of the ALM algorithm with two sub-blocks has been generally proven [4]. However, it is still difficult to prove that with more sub-blocks. Ours has five sub-blocks, making it even more challenging to prove. As a result, we study the convergence behaviour of the proposed method by calculating the relative error of the dual regularization term. As shown in Fig. 8, the proposed method has strong convergence

properties in simultaneously learning two manifold structures to achieve a more optimal solution. Thus, the relative error practically lowers after every iteration, causing it to converge quickly between 10 iterations.

5 Conclusion

This study proposes a novel method, which uses a two-way learning technique to induce the data manifold via two representative structures. The first structure is a low-rank matrix obtained directly from the original data. The second structure called the similarity matrix is learned via a k-symmetric nearest neighbor graph. In order to guarantee better clustering results, a dual regularization term is further introduced to allow both structures to guide themselves adaptively to find a more optimal solution without spectral post-processing procedure. Several experiments were performed to evaluate the effectiveness of the proposed method using four popular benchmark datasets. The results obtained concerning ACC, NMI, F-score, Recall, Precision, and AR show that the proposed method has some advantages over compared methods. Despite its effectiveness, our proposed method has several limitations. Notably, its computation efficiency may be limited in very large-scale problems [36] since it has many sub-blocks to update in each iteration. Secondly, finding the parameters' good combination is tricky for several datasets [37]. Thus, this paper can be improved with further studies exploring large-scale techniques to enhance computational efficiency, and self-parameter tuning strategy to improve performance. Additionally, we will exploit multiview learning [38, 39] and the deep learning paradigm to further enhance the performance of our method.

Data availability The datasets used in this study are available on the following websites.
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
WebKB: https://linqs.soe.ucsc.edu/datal

Declarations

Conflict 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.

References

1. Parsons L, Haque E, Liu H (2004) Subspace clustering for high dimensional data: a review. ACM SIGKDD Explor Newsl 6(1):90–105
2. Vidal R (2011) Subspace clustering. IEEE Signal Process Mag 28(2):52–68
3. Elhamifar E, Vidal R (2013) Sparse subspace clustering: Algorithm, theory, and applications. IEEE Trans Pattern Anal Mach Intell 35(11):2765–2781
4. Liu G, Lin Z, Yu Y (2010) Robust subspace segmentation by low-rank representation. Proc 27th Int Conf Mach Learn (ICML-10) 663–670. https://dl.acm.org/doi/10.5555/3104322.3104407
5. Liu G, Lin Z, Yan S, Sun J, Yu Y, Ma Y (2012) Robust recovery of subspace structures by low-rank representation. IEEE Trans Pattern Anal Mach Intell 35(1):171–184
6. Zhuang L, Wang J, Lin Z, Yang AY, Ma Y, Yu N (2016) Locality-preserving low-rank representation for graph construction from nonlinear manifolds. Neurocomputing 175:715–722

7. Gao W, Li X, Dai S, Yin X, Abhadiomhen SE (2021) Recursive sample scaling low-rank representation. J Math 2021:1–14

8. Wang J, Shi D, Cheng D, Zhang Y, Gao J (2016) LRSR: low-rank-sparse representation for subspace clustering. Neurocomputing 214:1026–1037

9. Zhu X, Zhang S, Li Y, Zhang J, Yang L, Fang Y (2018) Low-rank sparse subspace for spectral clustering. IEEE Trans Knowl Data Eng 31(8):1532–1543

10. Sui Y, Wang G, Zhang L (2019) Sparse subspace clustering via low-rank structure propagation. Pattern Recogn 95:261–271

11. Xia G, Chen B, Sun H, Liu Q (2020) Nonconvex low-rank kernel sparse subspace learning for keyframe extraction and motion segmentation. IEEE Trans Neural Netw Learn Syst 32(4):1612–1626

12. Teng L, Tang F, Zheng Z, Kang P, Teng S (2022) Kernel-Based Sparse Representation Learning With Global and Local Low-Rank Label Constraint. IEEE Trans Comput Soc Syst 11(1):488–502. https://doi.org/10.1109/TCSS.2022.3227406

13. Ng A, Jordan M, Weiss Y (2001) On spectral clustering: Analysis and an algorithm. Adv Neural Inf Process Systs 14

14. Tang K, Liu R, Su Z, Zhang J (2014) Structure-constrained low-rank representation. IEEE Trans Neural Netw Learn Syst 25(12):2167–2179

15. Yin M, Gao J, Lin Z (2016) Laplacian regularized low-rank representation and its applications. IEEE Trans Pattern Anal Mach Intell 38(3):504–517

16. Wen J, Fang X, Xu Y, Tian C, Fei L (2018) Low-rank representation with adaptive graph regularization. Neural Netw 108:83–96

17. Zhang Z, Wang L, Li S, Wang Y, Zhang Z, Zha Z, Wang M (2019) Adaptive structure-constrained robust latent low-rank coding for image recovery. 2019 IEEE Int Conf Data Min (ICDM) 846–855. (IEEE). https://www.computer.org/csdl/proceedings-article/icdm/2019/460400a846/1h5XOYozIe4

18. Wei L, Zhang Y, Yin J, Zhou R, Zhu C, Zhang X (2019) An improved structured low-rank representation for disjoint subspace segmentation. Neural Process Lett 50:1035–1050

19. Zheng R, Li M, Liang Z, Wu FX, Pan Y, Wang J (2019) SinNLRR: a robust subspace clustering method for cell type detection by non-negative and low-rank representation. Bioinformatics 35(19):3642–3650

20. Chen J, Mao H, Wang Z, Zhang X (2021) Low-rank representation with adaptive dictionary learning for subspace clustering. Knowl-Based Syst 223:107053

21. Guo L, Zhang X, Zhang R et al (2023) Robust graph representation clustering based on adaptive data correction. Appl Intell 53:17074–17092. https://doi.org/10.1007/s10489-022-04268-8

22. Lu GF, Wang Y, Tang G (2022) Robust low-rank representation with adaptive graph regularization from clean data. Appl Intell 52(5):5830–5840

23. Feng J, Lin Z, Xu H, Yan S (2014) Robust subspace segmentation with block-diagonal prior. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA, pp 3818–3825. https://doi.org/10.1109/CVPR.2014.482

24. Lucińska M, Wierzchoń ST (2012) Spectral clustering based on k-Nearest neighbor graph. In: Cortesi A, Chaki N, Saeed K, Wierzchoń S (eds) Computer Information Systems and Industrial Management. CISIM 2012. Lecture Notes in Computer Science, vol 7564. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33260-9_22

25. Zhan K, Nie F, Wang J, Yang Y (2019) Multiview consensus graph clustering. IEEE Trans Image Process 28(3):1261–1270

26. Von Luxburg U (2007) A tutorial on spectral clustering. Stat Comput 17:395–416

27. Cai JF, Candès EJ, Shen Z (2010) A singular value thresholding algorithm for matrix completion. SIAM J Optim 20(4):1956–1982

28. Nie F, Wang X, Jordan M, Huang H (2016) The constrained laplacian rank algorithm for graph-based clustering. Proc AAAI Conf Artif Intell 30(1)

29. Lu C, Feng J, Lin Z, Yan S (2013) Correlation Adaptive Subspace Segmentation by Trace Lasso. 2013 IEEE International Conference on Computer Vision. Sydney, NSW, Australia, pp 1345–1352. https://doi.org/10.1109/ICCV.2013.170

30. Xie X, Guo X, Liu G, Wang J (2017) Implicit block diagonal low-rank representation. IEEE Trans Image Process 27(1):477–489

31. Tolić D, Antulov-Fantulin N, Kopriva I (2018) A nonlinear orthogonal non-negative matrix factorization approach to subspace clustering. Pattern Recogn 82:40–55

32. Ji P, Reid I, Garg R, Li H, Salzmann M (2017) Adaptive low-rank kernel subspace clustering. arXiv preprint arXiv:1707.04974

33. Abhadiomhen SE, Wang Z, Shen X (2022) Coupled low rank representation and subspace clustering. Appl Intell 52(1):530–546

34. Lu C, Feng J, Lin Z, Mei T, Yan S (2018) Subspace clustering by block diagonal representation. IEEE Trans Pattern Anal Mach Intell 41(2):487–501
35. Tang H, Yuan C, Li Z, Tang J (2022) Learning attention-guided pyramidal features for few-shot fine-grained recognition. Pattern Recogn 130:108792
36. Liu Z, Tang C, Abhadiomhen SE, Shen X-J, Li Y (2023) Robust label and feature space co-learning for multi-label classification. IEEE Transactions on Knowledge and Data Engineering 35(11):11846–11859. https://doi.org/10.1109/TKDE.2022.3232114
37. Abhadiomhen SE, Nzeh RC, Ganaa ED, Nwagwu HC, Okereke GE, Routray S (2022) Supervised shallow multi-task learning: analysis of methods. Neural Process Lett 54(3):2491–2508
38. Dong N, Yan S, Tang H, Tang J, Zhang L (2024) Multi-view information integration and propagation for occluded person re-identification. Inf Fusion 104:102201
39. Tang H, Liu J, Yan S, Yan R, Li Z, Tang J (2023) M3net: multi-view encoding, matching, and fusion for few-shot fine-grained action recognition. Proc 31st ACM Int Conf Multimedia 1719–1728. https://dl.acm.org/doi/abs/10.1145/3581783.3612221