ORIGINAL ARTICLE

Discriminative low-rank projection for robust subspace learning

Zhihui Lai[1,2] · Jiaqi Bao[1,2] · Heng Kong[3] · Minghua Wan[4] · Guowei Yang[4]

Abstract

The robustness to outliers, noises, and corruptions has been paid more attention recently to increase the performance in linear feature extraction and image classification. As one of the most effective subspace learning methods, low-rank representation (LRR) can improve the robustness of an algorithm by exploring the global representative structure information among the samples. However, the traditional LRR cannot project the training samples into low-dimensional subspace with supervised information. Thus, in this paper, we integrate the properties of LRR with supervised dimensionality reduction techniques to obtain optimal low-rank subspace and discriminative projection at the same time. To achieve this goal, we proposed a novel model named Discriminative Low-Rank Projection (DLRP). Furthermore, DLRP can break the limitation of the small class problem which means the number of projections is bound by the number of classes. Our model can be solved by alternatively linearized alternating direction method with adaptive penalty and the singular value decomposition. Besides, the analyses of differences between DLRP and previous related models are shown. Extensive experiments conducted on various contaminated databases have confirmed the superiority of the proposed method.

Keywords Feature selection · Low-rank representation · Image classification · Small-class problem · Subspace learning

1 Introduction

Dimensionality reduction is one of the most effective and simple techniques in machine learning and classification problem [1, 2]. It is proposed to solve the critical problem

✉ Heng Kong
  generaldoc@126.com

  Zhihui Lai
  lai_zhi_hui@163.com

  Jiaqi Bao
  837292365@qq.com

  Minghua Wan
  wmh36@sina.com

  Guowei Yang
  38090605@qq.com

1 College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China

2 Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen 518060, China

3 BaoAn Central Hospital of Shenzhen, Shenzhen 518060, China

4 School of Information Engineering, Nanjing Audit University, Nanjing 211815, Jiangsu, P.R. China

called "curse of dimensionality" [3] which means models cannot perform well when processing high-dimensional data, even cause the singular problem. Recently, many classification models based on image recognition have been proposed, such as multiple-instance learning (MIL) [4] and subspace learning. The most classical subspace learning methods include principal component analysis (PCA) [5–7] and Linear discriminant analysis (LDA) [8–10]. PCA aims to project the high dimensionality data into the low dimensionality subspace along the direction of the maximum variance and LDA needs to find the optimal discriminative subspace so that the ratio of between-class scatter to within-class scatter is maximized. However, both PCA and LDA cannot learn the local structure information for dimensionality reduction task. To address this problem, many manifold learning techniques based on locality information were proposed. The representative methods include locality preserving projections (LPP) [11, 12], locally linear embedding (LLE) [13, 14], neighborhood preserving projection (NPP) [15], neighborhood preserving embedding (NPE) [16] and the variant of NPE named discriminant sparse neighborhood preserving embedding (DSNPE) [17]. Recently, a structured optimal graph based sparse feature extraction (SOGSFE) [18] method was proposed to obtain

the local manifold sparse structure and semi-supervised feature extraction simultaneously.

Previous studies present that a least square regression (LSR) formation [19, 20] characterizes the classical subspace learning methods. LSR is usually mathematically tractable and it has a lot of versions, such as weight LSR [21], partial LSR [22] and nonnegative least squares (NNLS) [23]. However, all above regression methods have two disadvantages. Firstly, for most LSR-based methods, they are only devoted to keeping the local geometry information which is sensitive to the illumination, corruptions and outliers. Thus, these methods may fail to achieve good results in the recognition task. Secondly, LSR-based techniques possess small-class problem which means the number of classes limits the number of projected vectors when learning a low-dimensional subspace.

To address the first problem mentioned above, researchers studied the $l_{2,1}$-norm-based least square regression methods and low-rank learning methods. There were a lot of works [24–27] proved that the robustness of least square regression methods by imposing the $l_{2,1}$-norm as the basic metric for pattern recognition. Therefore, more and more $l_{2,1}$-norm regression models were used in subspace learning [28–32]. For example, $l_{2,1}$-norm regularized discriminative feature selection (UDFS) [33] was designed for robust unsupervised subspace learning. However, UDFS is incapable of modeling the global structure of image points. Fortunately, the low-rank learning technique can also be used to improve robustness on noise/corrupted data and has been greatly noticed recently in the fields of data mining, machine learning and computer vision. A large number of low-rank representation (LRR) methods emerged in the past decade [34–39], which aims at corrupted data recovery and noise removal. LRR and its extended versions can discover potential relations among the data vectors by imposing a low-rank constraint. As a result, LRR-based methods can catch the global information of data and weaken the disturbance of noise. For example, robust PCA (RPCA) [40] learns a clean matrix and a sparse error matrix from the input matrix to obtain robust subspace segmentation. However, RPCA presumes that all data points are distributed in a single subspace, which is inconsistent with the fact that data vectors usually locate in multiple subspaces. To solve the above problem, Liu et al. [41, 42] proposed LRR to better acquire the underlying data structure with a low-rank matrix. Moreover, by taking full advantage of the graph regularizer, Yin et al. [39] developed a laplacian regularized LRR-based method that exploited both the global and local geometric information in data for graph learning and subspace projection. Based on [39], Li et al. [43] adopted a rank constraint on the proposed model to further learn an optimal graph with clear clustering structure and obtained encouraging performance. What's more, Zhuang et al. [44] introduced a non-negative low-rank and sparse graph (NNLRS) for classification and discriminative analysis task. NNLRS-graph not only captured the local linear relationships and global structures between image points but also preserved sparse to ensure the robustness of the algorithm. To integrate dimensionality reduction with LRR, Liu and Yan [45] put forward latent LRR which recovered the low-rank data representation and projected the clean data points into a low-dimensional subspace simultaneously. Since two low-rank matrices must be learned separately, the effectiveness of latent LRR cannot be guaranteed. Thus, approximate low-rank projection (ALRP) [46] treated the two matrices in LatLRR as a whole so that they can be boosted mutually in the learning process. By combining with least square regression, ALRP was extended to the supervised version called SALRP which can do feature extraction and classification task at the same time. Nevertheless, SALRP is limited by the size of the class label matrix to obtain c projections at most, where c represents the number of classes. Especially, when c is small, the dimension of the learned features is not enough for achieving good performance, which is undesirable in many applications.

To solve the problem mentioned above, we develop a novel model called discriminative low-rank projection (DLRP), which combines the properties of LRR and LSR to settle down the underlying problem existing in LRR and its variants. Our purpose is to obtain enough projections in a supervised manner and learn a robust discriminative subspace with low-rank representation. To ensure the effectiveness, we employ some adaptive constraints and $l_{2,1}$-norm as the basic measurement. The major contributions of this paper can be concluded as follows:

(1) Different from previous LLR-based techniques, our method alleviates the small-class problem that the number of features is bound by the number of classes. This problem may degrade the recognition rate for image feature extraction and classified task. DLRP can not only learn the projected vectors/matrix with arbitrary dimensions during the process of dimensionality reduction but also equip a robust low-rank subspace to gain better results.

(2) DLRP shares some advantages of both LSR and LRR. Therefore, it can not only utilize clean data recovered by LRR to obtain more promising image classification performances but also employ the projecting subspace to achieve better regression results for feature extraction.

(3) An efficient algorithm, which integrates the linearized alternating direction method with adaptive penalty (LADMAP) and singular value decomposition (SVD), is designed for solving the proposed model. The theoretical studies about the convergence and computation complexity of our algorithm are also shown. The algo-

rithm has validated its superiority on no matter corrupted datasets or deep learning dataset.

The remainder of this paper could be structured as below. In Sect. 2, some related methods will be reviewed for ease of understanding. The proposed DLRP and its optimization algorithm to solve the problem iteratively will be introduced in Sect. 3. Theoretical studies of our model will be shown in Sect. 4. Experimental analysis and results will be shown in Sect. 5 and conclusion in Sect. 6.

2 Related works

In this section, we present the notations and definitions related to our model firstly, and then review some related LRR-based models, such as subspace learning model RPCA, low-rank representation method LRR and its extended version latent LRR.

2.1 Notations and definitions

In this paper, given a collection of data points $X = [x_1, x_2, \ldots, x_n]$, n is the number of training samples, and Y is label information matrix which includes c classes. The matrix Y is a binary label matrix and when x_i belongs to jth class, the $Y_{ij} = 1$; if not, $Y_{ij} = 0$. $\|X\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij}^2}$ is Frobenius norm (F-norm) of X. $\|X\|_*$ denotes the nuclear norm that computes the sum of all the singular values. $diag(X)$ means the diagonal matrix X. Table 1 lists the detail notations of this paper.

2.2 Robust principal component analysis

As a dimensionality reduction technique, RPCA [40] can decompose the contaminated image information into a clean low-rank term and a sparse error term simultaneously. Because PCA assumes that the noises of data is Gaussian, its effectiveness will be influenced by heavy noises or unexcepted outliers. However, RPCA has not this assumption (RPCA assumes

Table 1 The notations used in this paper

Notation	Size	Description
X	$m \times n$	The data matrix
Y	$c \times n$	The label matrix
P	$c \times d$	The coefficient matrix
Q	$m \times d$	The transformation matrix
Z	$n \times n$	The low-rank representation matrix
E	$m \times n$	Error matrix
J	$n \times n$	Auxiliary matrix

that the noises are sparse). Thus, RPCA is a robust extended version of PCA. This advantage of RPCA is widely used in a variety of applications, such as video surveillance [47], face recognition [48] and singing-voice separation from monaural records [49].

RPCA is a low-rank matrix recovery technique, and its goal is to decompose data X to $A + E$, where A is the clean low-rank data on the low-dimensional subspace and E records sparse error. Such decomposition can be obtained through solving the following principal component pursuit problem [40]:

$$\min_{A,E} rank(A) + \eta \|E\|_0, \ s.t. \ X = A + E \tag{1}$$

where $rank(A)$ means the rank of A, η is a positive parameter to trade off the low rankness and sparsity.

2.3 Low-rank representation-based clustering methods

Firstly, we introduce the LRR [41, 42] method. Then, we review the LatLRR [45].

As a subspace learning method, LRR aims to find the lowest-rank representation for data clustering. The general formulation of LRR can be written as:

$$\min_{Z} rank(Z), \ s.t. \ X = DZ \tag{2}$$

where D is treated as the dictionary matrix while Z reveals the potential subspace structure. In real applications, we select X itself as dictionary matrix, so (2) can be reformulated as:

$$\min_{Z} rank(Z), \ s.t. \ X = XZ \tag{3}$$

Due to problem (3) is a NP-hard, we transform (3) into the base minimization problem:

$$\min_{Z} \|Z\|_*, \ s.t. \ X = XZ \tag{4}$$

Similar to RPCA, LRR can also be rewritten to the following formulation with error component:

$$\min_{Z,E} \|Z\|_* + \lambda \|E\|_p, \ s.t. \ X = XZ + E \tag{5}$$

where λ is a positive penalty parameter to balance the low-rank part and reconstruction error as η of RPCA. $\|\cdot\|_p$ denotes certain norm, such as F-norm, l_1-norm, $l_{2,1}$-norm, etc. However, when training samples are high-dimensional and insufficient, LRR fails to obtain good performance. LatLRR is proposed to exploit more unobserved samples. The function of LatLRR is formulated as:

$$\min_{Z,E} \|Z\|_* + \lambda \|E\|_p, \ s.t. \ X = [X, X_H]Z + E \tag{6}$$

where X, X_H represent the observed data and unobserved data separately. With Bayesian inference [50], X can equal to $XZ + LX$ so that LatLRR can use L to reduce the dimension of original datasets. (6) is rewritten as:

$$\min_{Z,E} \|Z\|_* + \|L\|_* + \lambda\|E\|_p, \, s.t. \, X = XZ + LX + E \tag{7}$$

Although LatLRR is a good solution to the problem that LRR-based methods cannot directly learn the low-dimensional subspace, but the discriminant information is not used. Therefore, LatLRR is not suitable for feature extraction and classification problems in such case.

In addition, it is expected that label information of the data can be well integrated with LRR-based methods, which will derive the clean data with discriminating power. Moreover, if directly used the terms $\|Y - XZ\|_F^2$ or $\|Y - LX\|_F^2$ in the above low-rank learning models, we had to encounter the small-class problem. i.e. the upper bound of the learned subspace of L or Z will be the maximum of c. Hence, how to recover the discriminant representative coefficient matrix Z for robust regression task is a critical problem.

3 Objective function and optimization solution

In this subsection, we first the introduce motivations of our model. The object function and update rules are also presented later. At last, an efficient algorithm is designed to iteratively solve the optimization problem.

3.1 Motivations of proposed method

The classical unsupervised dimensionality reduction techniques use Euclidean distances as the metric to measure the local geometric structure and construct the graph by local patches. In real applications, the data are often noisy and even grossly corrupted. Local structural information of data is easily damaged in the contaminated circumstances and then affect the recognition results in previous method. As a result, it is desirable to develop a novel technique that focuses on global information and keeps the robustness to some unexpected situations, such as disguise, occlusion, and specular reflections or pixel corruptions.

To this end, LRR is proposed as its ability to obtain the global structure of the data information. It also has great robustness for subspace learning and even recovers the true data [42]. Therefore, it is expected to integrate the low-rank properties of LRR into the dimensionality reduction process. Furthermore, to make the algorithm obtain more projections and inherit discriminative ability of LSR, we propose a novel supervised subspace learning method called DLRP. The key idea is to use the clean recovered data to fit the label so that data samples with the same labels can keep close to each other on the low-dimensional subspace. In this way, our method can not only use the discriminant information to obtain higher recognition rates but also guarantee the robustness in low-dimensional subspace. Different from the previous LRR-based methods, we impose some adaptive constraints on representation matrix Z for better performance. In conclusion, our method can jointly generate a robust low-dimensional subspace, where the dimension of features is greater than the number of classes, and obtain a discriminate low-rank representation.

3.2 Objective function of DLRP

Unlike the previous LRR-based methods which cannot directly introduce the label matrix for regression task, we aim to embed the label information into LRR and use the recovered data for supervised subspace learning. In order to make our model perform more robust, we utilize the $l_{2,1}$-norm as the basic metric to record the reconstructive errors. Then we have

$$\min_{Z,P,Q} \|Z\|_* + \lambda\|X - XZ\|_{2,1} + \alpha\left\|Y - PQ^TXZ\right\|_F^2 \tag{8}$$

For evading the trivial solution of the problem (8) and ease of measuring the reconstructive error, we introduce the orthogonal constraint on matrix P and error matrix E, so we can reformulate (8) as:

$$\min_{Z,E,P,Q} \|Z\|_* + \lambda\|E\|_{2,1} + \alpha\left\|Y - PQ^TXZ\right\|_F^2$$
$$s.t. \, X = XZ + E, P^TP = I \tag{9}$$

where P is an orthogonal matrix and Q is the projected matrix which maps the recovered data representation XZ onto the low-dimensional space, α is a positive weighted parameter to control the fitness of least squares term. To address the small-class problem, which may degrade the performance in real applications, it is expected to utilize the two particular matrices $P \in R^{c \times d}$ and $Q \in R^{m \times d}$ to replace the single projector $L \in R^{c \times m}$ in LatLRR (PQ^T is the same size as L obviously).

In conventional LSR, the single projection matrix can only learn at most c projections. However, the proposed method can obtain d ($d \geq c$) projections at least for extracting salient features. In this way, the small-class problem in LSR or LatLRR can be released. In conclusion, in the term $\|Y - PQ^TXZ\|_F^2$, the matrix Q^T first projects the low-rank representation XZ to robust low-dimensional subspace, then the orthogonal matrix P is used to obtain more projective information and regress the projected features to the label matrix Y with an orthogonal rotation. This regularization

term can make the low-dimensional representation PQ^TXZ own the similar structure as that of Y.

ALGORITHM 1 DLRP iterative algorithm

Input: Image matrix $X = [x_1, x_2, ..., x_n]$, the label matrix Y, the number of iteration T, the parameters λ, α.

Initialization: set all entries of $Z, E, J, Q, C_1, C_2, C_3$ to zero; $Z = J$; set $\mu = 0.1$, $\mu_{max} = 10^6$, $\rho = 1.01$.

Procedure:
while not converge do
Step 1: Update Z by using (19).
Step 2: Update E by using (21).
Step 3: Update J by using (23).
Step 4: Update C_1, C_2, C_3, μ by using (24), (25), (26), (27), respectively.
Step 5: Update P by using (32).
Step 6: Update Q by using (33).
Step 7: Update $T = T + 1$.
Step 8: Obtain the optimal solution (Z, E, J, P, Q).
end while
Output: The projection vector Q.

In (9), the first part is the minimization of the rank of Z, which can keep the global structure of original data, while the second part is for recording errors derived from noise and corruptions. For balance the representation coefficients, we further introduce a new constraint on the low-rank representation Z. Thus, the final objective function of DLRP is defined as follows:

$$\min_{Z,E,P,Q} \|Z\|_* + \lambda\|E\|_{2,1} + \alpha\left\|Y - PQ^TXZ\right\|_F^2 \quad (10)$$
$$s.t. \ X = XZ + E, P^TP = I, 1_n^TZ = 1_n^T$$

where $1 \in R^{n \times 1}$ represents the vector with all elements equal to one. To avoid all row elements of Z equal to zero, it is sensible to enforce low-rank representation Z by constraint $1_n^TZ = 1_n^T$, which means the sum of each row of low-rank matrix Z must be one.

(10) is our final objective function which can be iteratively optimized. Therefore, in next subsection, we design an iterative algorithm to compute the optimal solution of DLRP.

3.3 Iterative algorithms

In this part, we introduce the solution of optimization problem. We first use the LADMAP [51] to optimal Z and E in (11) and (42). Then we optimal P and Q in sequence. To make the object function easily separated and solved, we introduce an auxiliary matrix J to replace Z. So, we have

$$\min_{Z,E,P,Q,J} \|J\|_* + \lambda\|E\|_{2,1} + \alpha\left\|Y - PQ^TXZ\right\|_F^2 \quad (11)$$
$$s.t. \ X = XZ + E, P^TP = I, 1_n^TZ = 1_n^T, Z = J$$

The augmented Lagrangian function of problem (11) is:

$$L(Z, E, J, C_1, C_2, C_3, \mu)$$
$$= \|J\|_* + \lambda\|E\|_{2,1} + \alpha\left\|Y - PQ^TXZ\right\|_F^2$$
$$+ \langle C_1, X - XZ - E\rangle + \langle C_2, Z - J\rangle + \langle C_3, 1_n^TZ - 1_n^T\rangle$$
$$+ \frac{\mu}{2}\left(\|X - XZ - E\|_F^2 + \|Z - J\|_F^2 + \left\|1_n^TZ - 1_n^T\right\|_F^2\right)$$
$$= \|J\|_* + \lambda\|E\|_{2,1} + \alpha\left\|Y - PQ^TXZ\right\|_F^2$$
$$+ \frac{\mu}{2}\left(\left\|X - XZ - E + \frac{C_1}{\mu}\right\|_F^2 + \left\|Z - J + \frac{C_2}{\mu}\right\|_F^2 + \left\|1_n^TZ - 1_n^T + \frac{C_3}{\mu}\right\|_F^2\right)$$
$$- \frac{1}{2\mu}\left(\|C_1\|_F^2 + \|C_2\|_F^2 + \|C_3\|_F^2\right) \quad (12)$$

where C_1, C_2, C_3 are Lagrange multipliers, $\mu > 0$ is a penalty parameter. With other variables fixed, we can solve the $Z, E, J, C_1, C_2, C_3, \mu$ by minimizing the augmented Lagrangian function and P, Q respectively.

Z-step Given the other variables, the terms in (12) and (42) that depend on Z are

$$\alpha\left\|Y - PQ^TXZ\right\|_F^2$$
$$+ \frac{\mu}{2}\left(\left\|X - XZ - E + \frac{C_1}{\mu}\right\|_F^2 + \left\|Z - J + \frac{C_2}{\mu}\right\|_F^2 + \left\|1_n^TZ - 1_n^T + \frac{C_3}{\mu}\right\|_F^2\right) \quad (13)$$

Then, (13) can be transformed to the minimization problem as follow:

$$\min_Z \alpha\left\|Y - PQ^TXZ\right\|_F^2$$
$$+ \frac{\mu}{2}\left(\|M - XZ\|_F^2 + \|Z - N\|_F^2 + \left\|1_n^TZ - L\right\|_F^2\right) \quad (14)$$

where $M = X - E + \frac{C_1}{\mu}$, $N = J - \frac{C_2}{\mu}$ and $L = 1_n^T - \frac{C_3}{\mu}$. Then (14) can be solved by setting its derivative with respect to Z be 0, we have

$$W_aZ = W_b + W_c \quad (15)$$

where

$$W_a = 2\alpha X^TQQ^TX + \mu\left(X^TX + I_n + 1_n1_n^T\right) \quad (16)$$

$$W_b = 2\alpha X^TQP^TY \quad (17)$$

$$W_c = \mu\left(X^TM + N + 1_nL\right) \quad (18)$$

Finally, we can have the solution as follow:

$$Z = W_a^{-1}(W_b + W_c) \tag{19}$$

E-step E is obtained by solving the following function:

$$E = \arg\min_E \lambda\|E\|_{2,1} + \frac{\mu}{2}\left\|X - XZ - E + \frac{C_1}{\mu}\right\|_F^2 \tag{20}$$

According to [52], E can be presented as the following closed solution:

$$E = \Omega_{\lambda/\mu}\left(X - XZ + \frac{C_1}{\mu}\right) \tag{21}$$

where Ω is the shrinkage operator.

J-step For solving J, (12) is rewritten to the following minimization problem:

$$J = \arg\min_J \|J\|_* + \frac{\mu}{2}\left\|Z - J + \frac{C_2}{\mu}\right\|_F^2 \tag{22}$$

then J can be achieved as follow by using the singular value thresholding (SVT) operator [53]:

$$J = \Theta_{1/\mu}\left(Z + \frac{C_2}{\mu}\right) \tag{23}$$

where Θ is SVT operator.

C_1, C_2, C_3, μ**-step** Lagrange multipliers C_1, C_2, C_3 and penalty parameter μ update as follow:

$$C_1 \leftarrow C_1 + \mu(X - XZ - E) \tag{24}$$

$$C_2 \leftarrow C_2 + \mu(Z - J) \tag{25}$$

$$C_3 \leftarrow C_3 + \mu\left(1_n^T Z - 1_n^T\right) \tag{26}$$

$$\mu \leftarrow \min\left(\rho\mu, \mu_{\max}\right) \tag{27}$$

where $\rho > 0$ and μ are constants

P-step Given the other variables, P can be solved by minimizing the following problem:

$$P = \arg\min_P \alpha\left\|Y - PQ^T XZ\right\|_F^2 s.t. P^T P = I \tag{28}$$

(28) can be expanded to:

$$\begin{aligned} P &= \arg\min_P \alpha\left\|Y - PQ^T XZ\right\|_F^2 \\ &= \arg\min_P \alpha tr\left[\left(Y - PQ^T XZ\right)^T\left(Y - PQ^T XZ\right)\right] \\ &= \arg\min_P \alpha tr\left(Y^T Y - 2Y^T PQ^T XZ + Z^T X^T QQ^T XZ\right) \\ &\quad s.t.\ P^T P = I \end{aligned} \tag{29}$$

When other variables are fixed, (29) is equivalent to optimize the follow problem:

$$\begin{aligned} &\min_P tr\left(-2Q^T XZY^T P\right) s.t. P^T P = I \\ &\Leftrightarrow \max_P tr\left(Q^T XZY^T P\right) s.t. P^T P = I \end{aligned} \tag{30}$$

Then P can be optimal by computing singular value decomposition (SVD) [40] of matrix $Q^T XZY^T$ as follow:

$$Q^T XZY^T = \tilde{U}\tilde{D}\tilde{V}^T \tag{31}$$

Finally, the optimal P can be presented as

$$P = \tilde{V}\tilde{U}^T \tag{32}$$

Q-step For (29), P and Z are given already, we take partial derivative of (29) with respect to Q and set it to 0, it is simple to get

$$Q = \left(XZZ^T X^T + I\right)^{-1} XZY^T P \tag{33}$$

By updating the iterative algorithm mentioned above, we finally obtained the optimal solution to DLRP. Algorithm 1 presents the steps that how to employ the algorithm iteratively. Theoretical studies of our model, such as convergence and computational complexity, can be found in the next section.

4 Algorithm analysis and comparisons

In this section, we first introduce the convergence of our algorithm. And then, we analyze its computational complexity. Finally, we show the details about comparisons between our algorithm and other most related algorithms.

4.1 Convergence discussion

As shown in previous section, we optimized the proposed model with an iterative algorithm. So it is necessary to prove the convergence of our model.

Let

$$U(Z, E, J, P, Q) = \|J\|_* + \lambda\|E\|_{2,1} + \alpha\left\|Y - PQ^T XZ\right\|_F^2 \tag{34}$$

We can obtain the following theorem.

Theorem 1 *The iterative scheme of algorithm in Algorithm 1 monotonically decreases the objective function value of (10) in each iteration.*

Proof Suppose the objective function in the tth iteration can be presented as

$$U(Z_t, E_t, J_t, P_t, Q_t) = \|J_t\|_* + \lambda \|E_t\|_{2,1} + \alpha \left\| Y - P_t Q_t^T X Z_t \right\|_F^2 \tag{35}$$

where Z_t, E_t, J_t, P_t and Q_t mean matrix Z, E, J, P and Q in tth iteration, respectively. When E_t, J_t, P_t and Q_t are given. We know that inverse operation in (19) to compute Z_t will finally decrease the objective function, thus we have

$$U(Z_{t+1}, E_t, J_t, P_t, Q_t) \leq U(Z_t, E_t, J_t, P_t, Q_t) \tag{36}$$

When Z_t, J_t, P_t and Q_t are given. The convergence in (21) has proven in [52], thus we have

$$U(Z_{t+1}, E_{t+1}, J_t, P_t, Q_t) \leq U(Z_{t+1}, E_t, J_t, P_t, Q_t) \tag{37}$$

When Z_t, E_t, P_t and Q_t are given. J_t can be achieved by using the SVT operator which further decreases the objective function value and has been proven in [53]. Then, we have

$$U(Z_{t+1}, E_{t+1}, J_{t+1}, P_t, Q_t) \leq U(Z_{t+1}, E_t, J_t, P_t, Q_t) \tag{38}$$

When Z_t, E_t, J_t and Q_t are given. The optimal value of P_{t+1} derives from the SVD decomposition on $Q_t^T X Z_t Y^T$, which continuously reduces the objective function value. Therefore, we have

$$U(Z_{t+1}, E_{t+1}, J_{t+1}, P_{t+1}, Q_t) \leq U(Z_{t+1}, E_t, J_t, P_t, Q_t) \tag{39}$$

When Z_t, E_t, J_t and P_t are given, the solution derived by partial deviation provides the optimal Q_{t+1}. Finally we have

$$U(Z_{t+1}, E_{t+1}, J_{t+1}, P_{t+1}, Q_{t+1}) \leq U(Z_t, E_t, J_t, P_t, Q_t) \tag{40}$$

which completes the proof. □

We will present some experimental results to show the convergence property of DLRP in the following sections. In fact, the proposed algorithm can converge fast even within 5 iterations.

4.2 Computation complexity

The major time-consuming computations are

(1) the matrix inverse operation to solve Z and Q in step 1 and 6;
(2) the SVT to solve the values of J in step 3;
(3) the SVD to solve the value of P in step 5;

Thus, we just give the complexity about these steps in algorithm of DLRP. Firstly, the computational complexity of variable Q is about $O(m^3)$ while computing Z needs $O(n^2 m + nmd)$; Secondly, To compute the variable J costing $O(n^3)$; At last, optimizing the matrix P may cost $O(d^3)$;

computational complexity at most. To sum up, the overall cost for DLRP is up to $O[T(m^3 + n^2 m + nmd + n^3 + d^3)]$, where T means the number of iterations. Thus, in this case, PCA should be used to reduce the data dimension as preprocessing, which could greatly degree the computational cost.

4.3 Difference between our method and some similar methods

In this section, a further explanation of the essence of DLRP will be presented. We compare the proposed model with the most related methods on their performance in feature extraction and classification. The first model is (41) proposed in [38] and the second is (42) proposed in [46].

$$\min_{Z,P} \|Z\|_* + \lambda \left\| Q^T X - Q^T X Z \right\|_{2,1} \\ s.t.\ Q^T Q = I \tag{41}$$

$$\min_{P,Q,Z,E} \frac{1}{2} \left\| Y - Q^T X \right\|_F^2 + \frac{1}{2}\lambda_1 \|Q\|_F^2 + \lambda_2 \|E\|_1 + \lambda_3 \|Z\|_* \\ s.t.\ X = P Q^T X Z + E, P^T P = I \tag{42}$$

It is easy to find that all of DLRP, (41) and (42) aim to seek a projection matrix Q to reduce the dimension of original data X in low-rank subspace and use the low rankness to improve robustness to the noise for feature extraction simultaneously. However, our model (10) is essentially different from the models (41) and (42). Although model (41) can reduce the negative impact from the occlusions and corruptions by integrating $l_{2,1}$-norm loss function and low-rank property, the subspace it obtained does not capture the expected properties, e.g., utilizing supervised information to improve the discriminating power of recovered data. Furthermore, (41) cannot solve the small-class problem obviously. Model (42) is very similar to our model, both two models aim to obtain the discriminant low-rankness representation Z as well as the robust projecting subspace Q. From (42), we can find that $\left\| Y - Q^T X \right\|_F^2$ aims to fit the class label matrix Y with the low-dimensional representation $Q^T X$. While the term $\left\| Y - P Q^T X Z \right\|_F^2$ in (10) can not only keep closely correlated to the label indicator Y with low-rank and low-dimensional representation $P Q^T X Z$ in discriminant subspace but also release the small-class problem with $P Q^T$. In general, it can be seen that (42) cannot completely solve the small-class problem in essence for the regression task. In addition, in order to strengthen the recognition power and representation ability of the data, we append some additional constraints on (10), which is different from (41) and (42).

Figure 1 presents the recognition performance of LPP, RPCA, UDFS, LatLRR, DLRP, model (41) (LRE) and (42)

(SALRP) on the PIE face database with 5×5 block noise. From Fig. 1, it is can be found that DLRP and model (41) behave significantly superior to the model (42) on corrupted datasets, which indicates that integrating low-rank property and $l_{2,1}$-norm as the distance metric can greatly improve the robustness to the contaminated image samples.

5 Experiments

In this section, to systematically estimate the performance of our algorithm, we test the proposed algorithm on the five databases with the random pixel corruptions and different level uninterrupted occlusions. The first four databases are used for testing the robustness of our model in different cases, the last one the LFW database is used to test the performance of DLRP based on deep learning features. Except for the proposed DLRP, the conventional low-rank representation models RPCA, LRR and its extension LatLRR, LRE [38] and SALRP [46] are involved for comparisons. Besides, both $l_{2,1}$-norm regularized unsupervised discriminative feature selection model UDFS [33] and locality preserving method LPP are also compared.

5.1 The description of datasets

The AR face dataset [54] includes 2400 face pictures collected from 120 different individuals. The FERET face dataset [55] contains 200 people's face images and 1400 images in total. The Yale face database [56] includes 2414 people face pictures collected from 38 individuals. The CMU PIE face dataset has 41 368 face images which were shot from 68 people. The LFW face dataset [57] is only selected a subset

Fig. 1 Recognition rate (%) of different models on PIE face database with 5×5 block noise

which includes 4324 images of 158 subjects. Similar to [58], the deep conventional neural network (CNN) extracted about 1024 the deep features of original data before we use LFW for the recognition task.

All pictures of databases are cropped to 50×40, 40×40, 32×32, 32×32 and 112×96 size for the AR, FERET, Yale, CMU PIE and LFW face databases, respectively. Figure 2a–e display samples of the people's face pictures on AR, FERET, Yale, CMU PIE and LFW face databases.

5.2 Experimental settings

In our experiments, we compare the proposed model against the RPCA, LPP, UDFS, LatLRR, LRE, and SALRP. We use the Yale, FERET, CMU PIE and AR face datasets to prove the robustness of DLRP when data contaminated by random pixel corruptions and different degree contiguous occlusion. According to the number of samples in different datasets, we randomly selected T_w images of each individual on Yale, AR, FERET, CMU PIE and LFW datasets as training sets and the remainders as the test set, where $T_w = 4, 5, 6$, $T_w = 4, 5, 6$, $T_w = 2, 3, 4$, $T_w = 4, 5, 6$ and $T_w = 5, 6, 7$ respectively. The best parameters were determined by the validation set, which was also used to choose the optimal parameters for other compared models. The regularization parameters λ and α is selected from 10^{-6} to 10^6. To achieve believable results from experiments, all algorithms on each dataset conducted ten times independently and average recognition accuracy were shown. The optimal dimensions of low-dimensional subspace selected in the range $[50, 200]$ at the intervals of 10. After the feature extraction process, the nearest neighbor (1NN) classifier based on Euclidean distance as the metric was utilized for accuracy of image classification. The average recognition rate and corresponding the number of dimensions of Yale, FERET, PIE databases are presented in the Tables 2, 3 and 4 when $T_w = 6$. The average recognition rates versus the number of dimensions on LFW and AR databases are also reported in Fig. 5a, b.

5.3 Robustness evaluation with random pixel noises

To evaluate the robustness of the DLRP, RPCA, LPP, UDFS, LatLRR, LRE and SALRP algorithms in the case that there are random pixel corruptions in face pictures, we add the Gaussian noise in images of Yale, FERET and PIE datasets. Figure 3 shows original data and corrupted data images with 0.1 densities and 0.15 densities respectively. Tables 2, 3 and 4 present the recognition rate and the number of dimensions of RPCA, LPP, UDFS, LatLRR, LRE and SALRP on random pixel noises with different densities of Gaussian noise. The robustness of DLRP to random pixel noises can be proven from the experimental results.
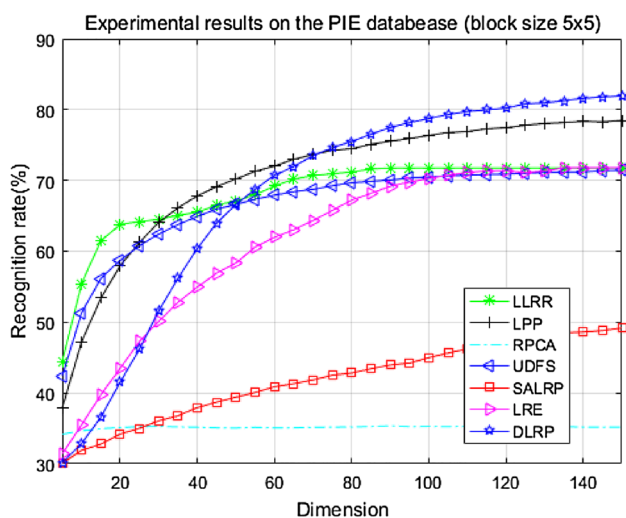
Fig. 2 Face samples of different datasets. **a** AR dataset, **b** FERET dataset, **c** Yale dataset, **d** PIE dataset, **e** LFW dataset

(a)

(b)
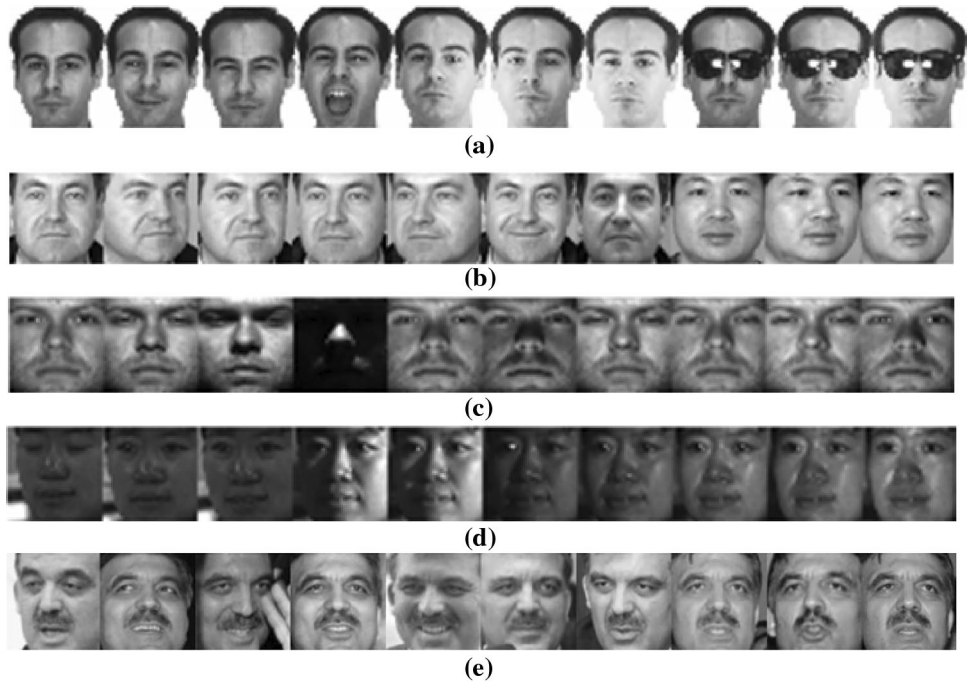
(c)

(d)

(e)

Table 2 Best recognition rate and corresponding dimensions of different algorithms on contaminated Yale datasets ($T_w = 6$)

	LatLRR	RPCA	LPP	UDFS	SALRP	LRE	DLRP
Den = 0.1	60.1202 (150)	26.4663 (150)	56.2139 (150)	22.8125 (120)	44.2427 (150)	65.9014 (200)	**67.5000** (150)
Den = 0.15	59.8798 (150)	26.6466 (150)	56.0817 (150)	58.0409 (110)	44.8197 (150)	67.9808 (200)	**72.3558** (150)
5 × 5	57.0673 (150)	26.9315 (150)	60.3606 (150)	55.1322 (120)	51.6707 (150)	61.8630 (200)	**72.6803** (150)
7 × 7	50.4087 (150)	26.7548 (150)	55.2404 (150)	49.3269 (100)	48.2692 (150)	53.5216 (200)	**70.1082** (150)

The bold values represent the highest recognition rate of all the methods in this datasets

Table 3 Best recognition rate and corresponding dimensions of different algorithms on contaminated FERET datasets ($T_w = 4$)

	LatLRR	RPCA	LPP	UDFS	SALRP	LRE	DLRP
Den = 0.1	74.7833 (150)	65.9833 (148)	76.5333 (150)	75.3512 (140)	66.2667 (150)	**82.0167** (150)	77.7833 (150)
Den = 0.15	75.3167 (150)	66.0000 (148)	75.3167 (150)	77.5612 (140)	65.8167 (150)	76.9833 (150)	**77.8833** (150)
5 × 5	70.5667 (150)	66.7167 (145)	70.5667 (150)	**77.3500** (145)	65.2000 (150)	72.8833 (150)	74.6500 (150)
7 × 7	67.7167 (150)	65.9667 (148)	69.6833 (150)	71.8667 (145)	65.1167 (150)	71.8667 (150)	**72.4667** (150)

The bold values represent the highest recognition rate of all the methods in this datasets

Table 4 Best recognition rate and corresponding dimensions of different algorithms on contaminated CMU PIE datasets ($T_w = 6$)

	LatLRR	RPCA	LPP	UDFS	SALRP	LRE	DLRP
Den = 0.1	76.4951 (150)	34.2801 (140)	68.0392 (145)	73.0229 (200)	63.3660 (200)	78.2761 (145)	**78.5212** (200)
Den = 0.15	76.3889 (150)	34.5670 (140)	69.4853 (145)	73.5376 (200)	62.1243 (200)	76.2827 (145)	**79.0114** (200)
5 × 5	71.7565 (150)	35.3595 (140)	78.4804 (145)	71.4297 (200)	49.1830 (200)	75.5964 (145)	**84.9183** (200)
7 × 7	66.9199 (150)	34.5180 (140)	69.3137 (145)	60.8578 (200)	47.0507 (200)	67.8023 (145)	**83.3333** (200)

The bold values represent the highest recognition rate of all the methods in this datasets

Fig. 3 Some examples of the original and noise images with different densities (Den) of Gaussian noise on Yale, FERET and PIE databases

5.4 Robustness evaluation with random block corruptions

To estimate the robustness of the DLRP, RPCA, LPP, UDFS, LatLRR, LRE and SALRP algorithms in the case that there are various level contiguous noises and corruptions in face pictures, we add some blocks with random positions in images. Figure 4 shows original data and samples of Yale, FERET and PIE datasets with block size 5×5 and 7×7. Tables 2, 3 and 4 display the best recognition rates of RPCA, LPP, UDFS, LatLRR, LRE and SALRP. As

shown in Tables 2, 3 and 4, DLRP achieves the best results among all compared algorithms in most cases. Thus, the robustness of our model to contiguous occlusions has been proved.

5.5 Experimental analysis on AR face dataset

To investigate the performance of our model and its competing models in the case that there are some variations in expressions of face, conditions of light and different degrees of occlusion, such as sunglasses and scarf. For all we know,

Fig. 4 Some examples of the original and noise images with different sizes of block noise on Yale, FERET and PIE datasets

the AR face database is a challenging dataset for recognition tasks for most of the regression methods. However, our proposed DLRP can outperform its competitions in most situations just as Fig. 5a presents.

5.6 Experimental analysis on LFW database in deep learning conditions

The LFW database is utilized to test the recognition results of our model and the compared models on the deep learning circulation. CNN was utilized as a first step to extract the deep features, then the subspace learning techniques were used to extract the features (i.e. RPCA, SPCA, LPP, UDFS, LatLRR, LRE and ALRP). The experimental results on recognizing deep features are shown in Fig. 5b. It shows that DLRP is still the best choice among all the methods.

5.7 Parameter selection and convergence study

Figure 6 draws the curves of iteration-based convergence on Yale dataset with 0.1 Gaussian noise, AR database and LFW dataset. As vividly depicted in figures, we can observe that the algorithm of DLRP could converge within 5 iterations on three databases. Similar properties on the convergence of the algorithm can also be discovered on other datasets.

There are two parameters in the proposed model, namely λ and α. Figure 7 shows the classification rates versus λ with α fixed, and α with λ fixed on the AR, LFW and Yale, PIE and FERET dataset with 0.1 Gaussian noise when λ and α ranged from $10^{-6}, 10^{-5} \ldots 10^5, 10^{-6}$. As can be seen, DLRP is very robust to the values of α and λ except on LFW database. Therefore, all of these Experiment results prove that DLRP can achieve better average recognition rates when parameter λ lies in 10^0 to 10^4 and α lies in 10^{-6} to 10^1 respectively. On the other side, when α ranges from 10^1 to 10^6 and
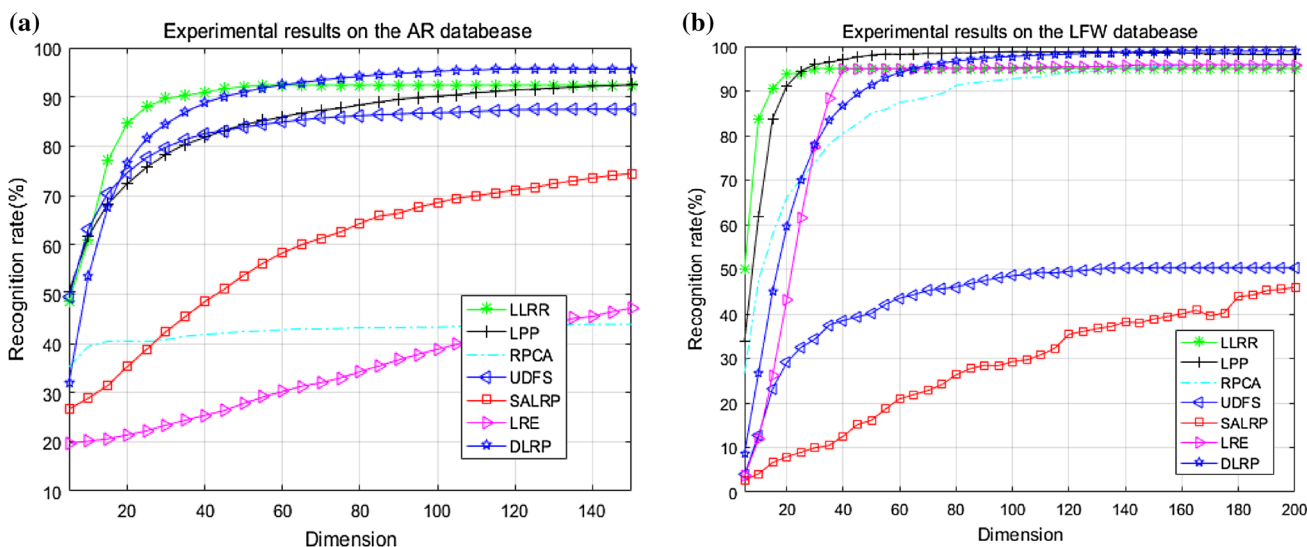

Fig. 5 Recognition rate (%) of algorithms on **a** AR database ($T_w = 6$) and **b** LFW database ($T_w = 7$)
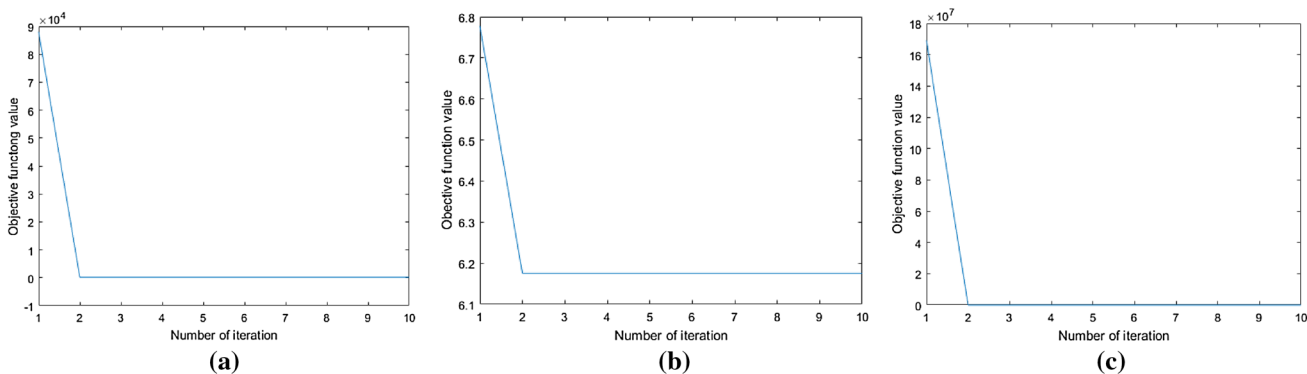

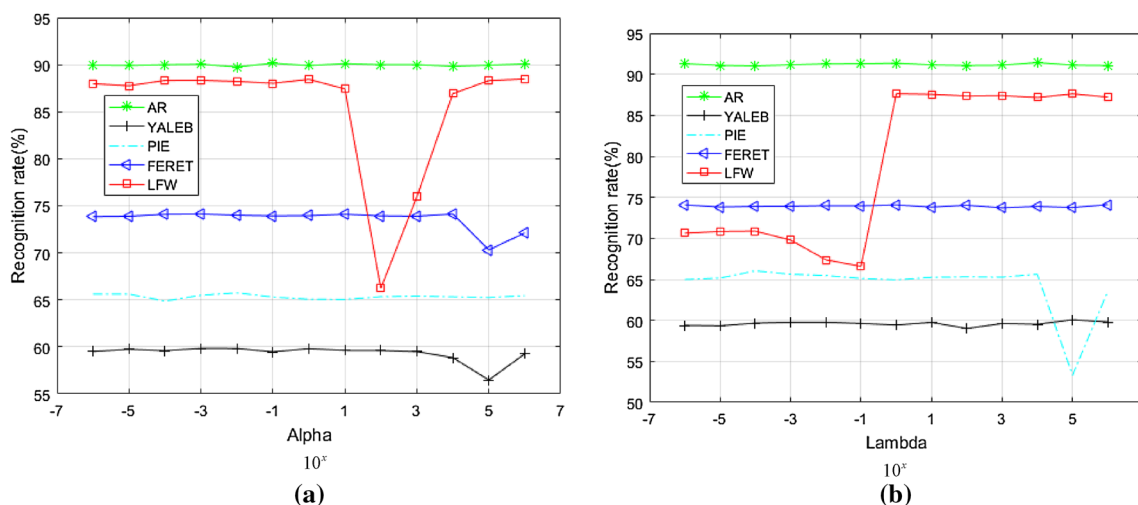**Fig. 6** Convergence curves on **a** AR database and **b** Yale database with 0.1 Gaussian noise **c** LFW database

Fig. 7 Different properties of DLRP. Testing the variation of recognition rate with **a** α regularization term coefficients **b** λ regularization term coefficients

λ ranges from 10^{-6} to 10^0, these two regularization terms are important for our models.

5.8 Experimental conclusions

Based on the experimental results on the various datasets, the following observations and conclusions are reached.

(1) Compared with the traditional dimension reduction methods, such as LPP, RPCA and UDFS, the proposed model can perform better when there are data specific corruptions on the pictures. DLRP combining the LRR with dimensionality reduction function is the key reason for high classification accuracy.

(2) From the Tables 2, 3 and 4, we can find that when the degree of noises and occlusions of the data samples are increased, the effect of most methods are affected, while our method is not affected too much or even performs better. This indicates our model is the best choice to perform on the contaminated circumstance.

(3) DLRP is superior to LRE for robust face image feature extraction and classification in most cases, which is due to the fact that our method not only fully uses the properties of LRR but also integrates the supervised information to improve discriminant ability. In addition, releasing the small-class problem and imposing adaptive constraints also have a positive effect on the recognition rates of the face image.

(4) DLRP outperforms SALRP in most cases, especially increasing about 30% to 40% on contaminated Yale database (see Table 2). The possible reason for the high recognition rate is that DLRP directly projects the clean data XZ into the label subspace while SALRP projects the original data X which includes more noises and corruptions. That means DLRP can use the low-rank representation to weaken the disturbance of noise in projected subspace, such that the learned low-dimensional subspace can be more robust than SALRP.

(5) From Figs. 1 and 5a, b, we can find that DLRP cannot achieve the consistent winner on low-dimensional subspace, especially under 80 dimensions. In other words, DLRP is unable to classify some data with low dimensionality. The major reason is that when dimensionality is so low that DLRP cannot maintain the manifold structure of samples, which may occur the problem of overfitting in this situation. However, DLRP can perform well on data with high dimensionality, which means it is favorable to select DLRP for classifying the high dimensional data.

Based on the above experimental observations and analysis, we can find that the DLPR can be superior to the classical models (i.e., LLRR, LPP, RPCA and UDFS) and most related models (i.e., SALRP and LRE). This indicates that integrating the low-rank properties with supervised information and releasing small-class problem are significant to obtain the high recognition rates. Furthermore, DLRP essentially inherits the advantage from the low rank representation to improve its robustness for corrupted datasets. Thus, the proposed model can learn an optimal robust subspace to achieve good performance in our experiments.

6 Conclusion

In this paper, a novel model named DLRP, which combines the low-rank representation with subspace learning, is proposed for image classification and feature extraction. Different from the previous LRR-based methods, DLRP can not only obtain clean data under the low-rank constraint but also learn a robust projecting subspace. In addition, DLRP figures out the small-class problem that exists in the typical least square regression or its derivatives. Thus, the proposed model can capture enough projections for feature extraction. Moreover, we impose some adaptive constraints on low-rank matrix to improve the performances which the most low-rank learning methods lack. To demonstrate the effectiveness and robustness of DLRP, we conducted mass comprehensive experiments on five benchmark datasets for comparing with related techniques. The limitation of our algorithm is that it cannot perform well on some low dimensional data, which we try to improve in the future.

References

1. Wang X, Xing H, Li Y, Hua Q, Dong C, Pedrycz W (2015) A study on relationship between generalization abilities and fuzziness of base classifiers in ensemble learning. IEEE Trans Fuzzy Syst 23(5):1638–1654
2. Wang X, Wang R, Xu C (2018) Discovering the relationship between generalization and uncertainty by incorporating complexity of classification. IEEE Trans Cybern 48(2):703–715
3. Donoho D (2000) High-dimensional data analysis: the curses and blessings of dimensionality. In: AMS conference on math challenges of the 21st century, pp 1–33
4. Wang R, Wang X, Kwong S, Xu C (2017) Incorporating diversity and informativeness in multiple-instance active learning. IEEE Trans Fuzzy Syst 25(6):1460–1475
5. Sirovich L, Kirby M (1987) Low-dimensional procedure for the characterization of human faces. J Opt Soc Am A 4(3):519–524
6. Turk M, Pentland A (1991) A eigenfaces for face detection/recognition. J Cognit Neurosci 3(1):71–86
7. Yi S, Lai Z, He Z, Cheung Y, Liu Y (2017) Joint sparse principal component analysis. Pattern Recognit 61:524–536
8. Ben X, Gong C, Zhang P, Yan R, Wu Q, Meng W (2019) Coupled bilinear discriminant projection for cross-view gait recognition. IEEE Trans Circuits Syst Video Technol. https://doi.org/10.1109/TCSVT.2019.2893736
9. Ye Q, Yang J, Liu F, Zhao C, Ye N, Yin T (2016) L1-norm distance linear discriminant analysis based on an effective iterative algorithm. IEEE Trans Circuits Syst Video Technol 28(1):114–129
10. Sun W, Xie S, Han N (2019) Robust discriminant analysis with adaptive locality preserving. Int J Mach Learn Cybern. https://doi.org/10.1007/s13042-018-00903-4
11. Zhang W, Kang P, Fang X, Teng L, Han N (2019) Joint sparse representation and locality preserving projection for feature extraction. Int J Mach Learn Cybern 10(7):1731–1745
12. Lu G-F, Wang Y, Zou J, Wang Z (2018) Matrix exponential based discriminant locality preserving projections for feature extraction. Neural Netw 97:127–136
13. Roweis S, Saul L (2000) Nonlinear dimensionality reduction by locally linear embedding. Sicence 290:2323–2326
14. Sun B, Zhang X, Li J, Mao X (2010) Feature fusion using locally linear embedding for classification. IEEE Trans Neural Networks Learn Syst 21(1):163–168
15. Pang Y, Zhang L, Liu Z (2005) Neighborhood preserving projections (NPP): a novel linear dimension reduction method. In: International conference on intelligent computing, pp 117–125
16. He X, Cai D, Yan S, Zhang H (2005) Neighborhood preserving embedding. In: IEEE international conference on computer vision, pp 1208–1213
17. Gui J, Sun Z, Jia W, Hu R, Lei Y, Ji S (2012) Discriminant sparse neighborhood preserving embedding for face recognition. Pattern Recognit 45(8):2884–2893
18. Liu Z, Lai Z, Ou W, Zhang K, Zheng R (2020) Structured optimal graph based sparse feature extraction for semi-supervised learning. Signal Process. https://doi.org/10.1016/j.sigpro.2020.107456
19. Xiao C, Ding C, Nie F, Huang H (2013) On the equivalent of low-rank linear regressions and linear discriminant analysis based regressions In: International conference on knowledge discovery and data mining, pp1124–1132
20. Torre F (2012) A least-squares framework for component analysis. IEEE Trans Pattern Anal Mach Intell 34(6):1041–1055
21. Strutz T (2010) Data fitting and uncertainty: a practical introduction to weighted least squares and beyond. Springer Vieweg and Teubner, Wiesbaden, Germany
22. Liu H, Ma Z, Han J, Chen Z, Zheng Z (2018) Regularized partial least squares for multi-label learning. Int J Mach Learn Cybern 9(2):335–346
23. Wold S, Ruhe H, Wold H, Dunn W (1984) The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. SIAM J Sci Stat Comput 5(3):735–743
24. Darmstadt T (2017) A compact formulation for the l21 mixed-norm minimization problem. In: IEEE international conference on acoustics, speech and signal processing, pp 4730–4734
25. Lai Z, Mo D, Wong W, Xu Y, Miao D, Zhang D (2018) Robust discriminant regression for feature extraction. IEEE Trans Cybern 48(8):2472–2484
26. Wang S, Xie D, Chen F, Gao Q (2018) Dimensionality reduction by LPP-L21. IET Comput Vis 12(5):659–665
27. Zhao L, Sun Y, Hu Y, Yin B (2014) Robust face recognition based l21-norm sparse representation. In: International conference on digital home, pp 25–29
28. Gu Q, Li Z, Han J (2011) Joint feature selection and subspace learning. In: International joint conference on artificial intelligence, pp 1294–1299
29. Ji S, Tang L, Yu S, Ye J (2010) A shared-subspace learning framework for multi-label classification. ACM Trans Knowl Discov Data 4(2):1–29
30. Lai Z, Xu Y, Yang J, Shen L, Zhang D (2017) Rotational invariant dimensionality reduction algorithms. IEEE Trans Cybern 47(11):3733–3746
31. Ma Z, Nie F, Yang Y, Uijlings J, Sebe N, Hauptmann A (2012) Discriminating joint feature analysis for multimedia data understanding for multimedia data understanding. IEEE Trans Multimed 14(6):1662–1672
32. Ma Z, Yang Y, Sebe N, Zheng K, Hauptmann A (2013) Multimedia event detection using a classifier-specific intermediate representation. IEEE Trans Multimed 15(7):1628–1637

33. Yang Y, Shen H, Ma Z, Huang Z, Zhou X (2010) L2,1-norm regularized discriminative feature selection for unsupervised learning. In: International joint conference on artificial intelligence, pp 1589–1594

34. Cheng L, Yang M (2018) Graph regularized weighted low-rank representation for image clustering. In: Chinese control conference, pp 9051–9055

35. Li S, Fu Y (2016) Learning robust and discriminative subspace with low-rank constraints. IEEE Trans Neural Networks Learn Syst 27(11):2160–2173

36. Liu Z, Wang X, Pu J, Wang L, Zhang L (2017) Nonnegative low-rank representation based manifold embedding for semi-supervised learning. Knowl Based Syst 136:121–129

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

38. Wong W, Lai Z, Wen J, Fang X, Lu Y (2017) Low rank embedding for robust feature extraction. IEEE Trans Image Process 26(6):2905–2917

39. 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

40. Candes E, Li X, Ma Y, Wright J (2011) Robust principal component analysis? J ACM 58(3):1–39

41. Liu G, Lin Z, Yan S, Sun J, Yu Y, Ma Y (2013) Robust recovery of subspace structures by low-rank representation. IEEE Trans Pattern Anal Mach Intell 35(1):171–184

42. Chen J, Yang J (2014) Robust subspace segmentation via low-rank representation. IEEE Trans Cybern 44(8):1432–1445

43. Li P, Yu J, Wang M, Zhang L, Cai D, Li X (2017) Constrained low-rank learning using least squares-based regularization. IEEE Trans Cybern 47(12):4250–4262

44. Zhuang L, Gao S, Tang J, Wang J, Lin Z, Ma Y (2015) Constructing a non-negative low rank and sparse graph with data-adaptive features. IEEE Trans Image Process 24(11):3717–3728

45. Liu G, Yan S (2011) Latent low-rank representation for subspace segmentation and feature extraction. In: International conference on computer vision, pp 1615–1622

46. Fang X, Han N, Wu J, Xu Y, Yang J, Wong W, Li X (2018) Approximate low-rank projection learning for feature extraction. IEEE Trans Neural Netw Learn Syst 29(11):5228–5241

47. Liu Y, Pados D (2016) Compressed-sensed-domain l1-PCA video surveillance. IEEE Trans Multimed 18(3):351–363

48. Wang Z, Xie X (2010) An efficient face recognition algorithm based on robust principal component analysis. In: International conference on internet multimedia computing and service, pp 99–102

49. Umap P, Chaudhari K (2015) Singing voice separation from polyphonic music accompanient using compositional model. Int J Adv Res Electr Instrum Eng 4:541–546

50. Zhou P, Lin Z, Zhang C (2015) Integrated low-rank-based discriminative feature learning for recognition. IEEE Trans Neural Networks Learn Syst 27(5):1080–1093

51. Lin Z, Liu R, Su Z (2011) Linearized alternating direction method with adaptive penalty for low-rank representation. In: Proceedings of 25th annual conference on neural information processing systems, pp 1–9

52. Yang J, Yin W, Zhang Y, Wang Y (2009) A fast algorithm for edge-preserving variational multichannel image restoration. SIAM J Sci Stat Comput 2(2):569–592

53. Cai J, Candes E, Shen Z (2010) A singular value thresholding algorithm for matrix completion. SIAM J Optim 20(4):1956–1982

54. Martinez A (1998) The AR face database. Cvc technical report, 24

55. Phillips P, Moon H, Rizvi S, Rauss P (2000) The FERET evaluation methodology for face-recognition algorithms. IEEE Trans Pattern Anal Mach Intell 22(10):1090–1104

56. Georghiades S, Belhumeur P, Kriegman D (2001) From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans Pattern Anal Mach Intell 23(6):643–660

57. Learned-miller E, Huang G, Mattar M, Berg T (2007) Labeled faces in the wild : a database for studying face recognition in unconstrained environments. Technical Report

58. Wen Y, Zhang K, Li Z, Qiao Y (2016) A discriminative feature learning approach for deep face recognition. In: European conference on computer vision, pp 499–515