

# Double Layer Robust 2DPCA: Joint Flexible Cut Norm and Adaptive Weighted Learning for Small Sample Size Image Recognition

Pengfei Bi, *Member, IEEE*, Mei Chen, Xue Du, *Member, IEEE*, Fahad Sohrab, *Member, IEEE*, and Moncef Gabbouj, *Fellow, IEEE*

**Abstract**—Recently, many 2D principal component analysis (2DPCA) methods with different distance metric mechanisms have been successfully proposed and applied to improve the robustness of information perception based on image data. However, these methods ignore the possibility of minimizing reconstruction errors in a real sense. To overcome this shortcoming, we propose a method called double layer robust 2DPCA (DLR-2DPCA), which is a novel formulation of 2DPCA with robustness. DLR-2DPCA uses the cut  $l_{2,p}$ -norm as the similarity measure criterion between data in the first layer, which not only effectively utilizes a flexible distance metric mechanism with rotational invariance, but also fully suppresses the amplification of reconstruction errors caused by  $p$ -value selection, greatly improving the robustness of the method. Moreover, we introduce an adaptive weighted learning strategy in the second layer, which can adaptively assign different weights to the reconstruction errors obtained from each sample in the first layer, clearly considering the difference in the contribution of each sample to the reconstruction error. Finally, we design a novel iterative algorithm with nongreedy properties to seek the DLR-2DPCA expected solution. Numerous experimental results have effectively validated the theoretical analysis and fully demonstrated the excellent performance of our method.

**Index Terms**—2-D principal component analysis (2DPCA), cut  $l_{2,p}$ -norm, adaptive weighted learning, image recognition.

## I. INTRODUCTION

MANY computer vision and pattern recognition methods have been successfully applied in consumer electronics products, such as face classification, traffic sign detection and handwritten digit recognition, the actual available image information not only has high-dimensional attributes, but is also mostly rendered in small sample data, meaning that the number of training samples cannot match the dimensions of the observation data [1]. Moreover, these image data are inevitably affected by various environmental factors during the collection

This work was supported by the National Natural Science Foundation of China under Grant 5217110332, in part by the Natural Science Basic Research Plan in Jiangsu Province of China under Grant BK20220452 (*Corresponding author: Pengfei Bi*).

Pengfei Bi and Mei Chen are with the School of Artificial Intelligence/School of Future Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China. 210044, China (e-mails: pfcx@nuist.edu.cn; chenmei0718@outlook.com;).

Xue. Du is with the College of Intelligent Systems Science and Engineering, Harbin Engineering University, Heilongjiang 150001, China (e-mail: duxue@hrbeu.edu.cn).

Fahad Sohrab and Moncef Gabbouj are with the Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, 33720, Finland (e-mails: moncef.gabbouj@tuni.fi; fahad.sohrab@tuni.fi)

process, resulting in some samples accompanied by the phenomena appear of occlusion, gaussian noise, salt-and-pepper noise, and even complex mixed noise [2]. Therefore, based on the above analysis, effectively improving the ability to depict important features of image data in the presence of noise interference is particularly important.

With the rapid development and breakthroughs of subspace learning techniques in recent years, it has gradually occupied a place in the research field of small sample image recognition, among which principal component analysis (PCA) [3], linear discriminant analysis (LDA) [4] and locality preserving projection (LPP) [5] are considered the three classic methods. PCA seeks to minimize the reconstruction error of projection data under the Euclid space metric. LDA aims to explore encoding discriminative features that can satisfy the minimum within-class variance and the maximum between-class variance simultaneously. LPP focuses on effectively protecting the manifold structure of sample data.

When applying the above techniques to perform the image feature extraction task, we usually consider converting the two-dimensional (2D) matrix transmitted by each image into one-dimensional (1D) vector in advance, which virtually destroys the inherent spatial structure attribute of the original data. To solve the negative impact of this abuse, Yang *et al.* [6] developed the 2DPCA method, which directly utilizes the strategy of replacing vector transformation with the input image to evaluate the covariance matrix. Inspired by 2DPCA method, a series of feature extraction methods oriented to matrix information, such as 2DLDA [7], 2DLPP [8], tensor PCA [9] and multilinear PCA (MPCA) [10], have been developed. Although the motivations of the objective functions established by the above methods are different from each other, they all utilize the square of the Euclidean norm as a metric between the original data and the reconstructed samples, which leads to the adverse impact of noise presenting exponential amplification. Therefore, these methods generally lack robustness in noisy environments.

To improve the shortcomings of the large distance metric, the idea of using the  $l_1$ -norm instead of the squared Euclidean norm strategy has emerged and is prioritized for PCA architecture applications [11]. For example, Kwak [12] presented PCA-L1, which maximizes variance by selecting the  $l_1$ -norm as the distance metric mechanism. Nie *et al.* [13] developed nongreedy algorithm for finding the optimal solution of PCA-L1, which improved the computational efficiency of the model. Enlightened by the above methods, many subspace methods  $l_1$ -

norm-based have been proposed. Among these, LDA-L1 [14] and LPP-L1 [15] are the two representative methods for image feature description. Guided by this idea and considering the crucial importance of the original spatial structure information for image recognition results, a series of 2DPCA methods using the  $l_1$ -norm have been continuously proposed. For instance, Li *et al.* [16] and Wang *et al.* [17] explored the invented 2DPCA-L1 solving process, where the former selects a solution strategy with greedy properties while the latter considers a nongreedy algorithm. Pang *et al.* [18] based on the  $l_1$ -norm to describe the reconstruction error of PCA in tensor space and named it TPCA-L1. Ju *et al.* [19] developed 2DPPCA-L1, which can learn the key parameters in the probabilistic method to improve the ability of image outlier detection and feature extraction. Wang *et al.* [20] proposed a generalized 2DPCA method  $l_p$ -norm-based metric by reasonably extending the application interval of the  $l_1$ -norm.

Although using the  $l_1$ -norm enhances the robustness of 2D methods to a certain extent, it also loses the inherent rotational invariance that exists in Euclidean space [21], and the actual solution obtained is also irrelevant to the covariance matrix with the ability to protect the geometric structure of the data [22]. To address these shortcomings effectively, Li *et al.* [23] proposed the  $F$ -norm 2DPCA, which uses the  $F$ -norm as a distance metric to maximize the objective function structure. On this basis, Gao *et al.* [24] presented the  $R_1$ -2DPCA method, which selects a robust metric mode with better scalability when performing image feature extraction operations. However, the above methods usually consider the use of fixed means for data centralization, invisibly ignoring the impact of noise on the mean data. To compensate for such defects, Wang *et al.* [25] introduced the optimal mean mechanism in the  $F$ -2DPCA and proposed the OMF-2DPCA. Inspired by 2DPCA, Zhou *et al.* [26] applied a centralized weight matrix to projection data and developed GC-2DPCA, further protecting the global structural information of the data. The focus of these methods is mainly on the robust selection of data similarity metric criteria, but they ignore the consideration of the inherent correlation between reconstruction error and variance in the structural relationship. Thus, building a reasonable system architecture will improve method robustness.

Naturally, many methods that establish potential correlation between reconstruction error and variance have gradually been incorporated into the research scope. For example, Gao *et al.* [27] proposed Angle-2DPCA via its tangent angle value. Wang *et al.* [28] proposed Cos-2DPCA by utilizing adjustable parameters, both of which essentially found the optimal solution of the objective function based on the right triangle relationship framework. Considering that the inherent system of the above methods indirectly restricts the match ability of robust metric strategies, the ability to accurately describe important features in noisy environments weakens. To address this dilemma, Zhang *et al.* [29] proposed 2DPCA-2- $L_p$ , which uses a joint 2-norm and  $l_p$ -norm to increase the flexibility of distance metric selection. However, it has the following defects in the method building process. First, the flexible variation in  $p$ -values may to some extent cause imbalanced evaluations

based on the importance of certain data points, and lack of rotational invariance when  $p = 2$ . Second, the dominant idea is to treat each sample data equally, ignoring the differences in the contributions of reconstruction errors between outliers and normal samples. From the perspective of further enhancing method robustness, we present a novel robust matrix-based subspace learning framework for image recognition, namely DLR-2DPCA. It employs the cut  $l_{2,p}$ -norm metric criterion combined with an adaptive weighted strategy to form a double layer robust structure, which effectively improves the method ability to withstand noise interference while successfully preserving rotational invariance. Compared with many existing robust 2DPCA methods, we emphasize the contribution of the proposed method in the following aspects.

(1) DLR-2DPCA selects to establish a novel cut  $l_{2,p}$ -norm as a distance metric strategy in the first layer robust structure, which not only fully enhances the flexibility of the Euclidean norm application but also reasonably improves the ability to suppress outliers under different  $p$ -values. More importantly, the use of the cut  $l_{2,p}$ -norm ensures a balance between data structure sparsity and spatial geometric information.

(2) DLR-2DPCA considers introducing an adaptive learning mechanism in the second layer robust structure to match the corresponding weights for each input sample data, which can automatically evaluate the effects of samples with different reconstruction errors without the need for manual intervention. Therefore, our proposed method truly achieves the core demand of minimizing reconstruction errors.

(3) DLR-2DPCA retains the valuable properties inherent in classical 2DPCA methods, specifically the rotational invariance possessed by Euclidean norm metrics, and the solution is closely related to the weighted covariance matrix with protect the global geometric structure of the data, which potentially improves the recognition performance.

(4) We develop a fast iterative algorithm with nongreedy properties for the proposed objective function, which can obtain a closed form solution in each iteration and ultimately converge to a local optimal solution. The relevant theoretical analysis provided in Section III C fully proves that the algorithm has good convergence.

The rest of this paper is organized as follows. Section II reviews the main related work, including 2DPCA-L1 and  $R_1$ -2DPCA. The double layer robust 2DPCA method and its important attribute analysis are presented in Section III. In Section IV, we report the experimental results in several image databases. Finally, conclusions are drawn in Section V.

## II. RELATED WORKS

### A. 2DPCA-L1

Assume that we have a set of  $N$  training sample data  $\mathbf{A}_i \in \mathcal{R}^{m \times n}$  ( $i = 1, 2, \dots, N$ ),  $\mathbf{A}_i$  indicates the  $i$ th image sample. In general, we need to complete centralized processing of input sample, i.e.,  $\sum_{i=1}^N (\mathbf{A}_i - \mathbf{M}) = \mathbf{0}$ ,  $\mathbf{M} = 1/N \sum_{i=1}^N \mathbf{A}_i$  is the mean data of all training samples. 2DPCA-L1 [16] seeks the projection matrix  $\mathbf{V} \in \mathcal{R}^{n \times q}$  ( $q \ll n$ ) by using the  $l_1$ -norm to maximize the sum of variance for each projection data.

Specifically, its objective function is defined as follows

$$\begin{cases} \operatorname{argmax} \sum_{i=1}^N \|\mathbf{A}_i \mathbf{V}\|_{l_1} = \operatorname{argmax} \sum_{i=1}^N \sum_{j=1}^m \|\mathbf{A}_i(j,:)\mathbf{V}\| \\ \text{s.t. } \mathbf{V}^T \mathbf{V} = \mathbf{I}_q \end{cases} \quad (1)$$

where  $\mathbf{I}_q \in \mathbb{R}^{q \times q}$  is an identity matrix,  $\|\cdot\|_1$  indicates the 1-norm of a vector,  $\|\cdot\|_{l_1}$  indicates the  $l_1$ -norm operation of a matrix, and  $\mathbf{A}_i(j,:)$  indicates the  $j$ th row of matrix  $\mathbf{A}_i$ .

Unlike the classical 2DPCA method, 2DPCA-L1 is more robust in the selection of the distance metric strategy. However, it also has corresponding limitations. First, the adoption of the  $l_1$ -norm will lead to the loss of rotational invariance and the solving process is complex. Moreover, the solution obtained by 2DPCA-L1 is not closely related to the covariance matrix with the global geometric structure attribute of the protected data. To address these deficiencies, the  $R_1$ -norm-based 2DPCA method is proposed, which is introduced in the following subsection.

### B. $R_1$ -2DPCA

The essential demand of  $R_1$ -2DPCA [24] is to solve the projection matrix  $\mathbf{V}$  by establishing the objective function

$$\begin{cases} \operatorname{argmax} \sum_{i=1}^N \|\mathbf{A}_i \mathbf{V}\|_{R_1} = \operatorname{argmax} \sum_{i=1}^N \sum_{j=1}^m \|\mathbf{A}_i(j,:)\mathbf{V}\|_2 \\ \text{s.t. } \mathbf{V}^T \mathbf{V} = \mathbf{I}_q \end{cases} \quad (2)$$

where  $\|\cdot\|_{R_1}$  indicates the  $R_1$ -norm operation of a matrix and  $\|\cdot\|_2$  indicates the  $l_2$ -norm operation of a vector.

As seen in objective function (2), the use of the  $R_1$ -norm can retain more valuable characteristics than that of  $l_1$ -norm, such as rotational invariance, but it still does not improve the expressiveness of robust 2DPCA technology to the greatest extent. First, the  $R_1$ -norm, as a distance measurement criterion, has defects in flexibility considerations and is not the optimal choice. Second,  $R_1$ -2DPCA cannot adaptively match the contribution of each sample in the optimization process. Finally,  $R_1$ -2DPCA failed to truly implement the core concept of minimizing data reconstruction error. To further improve the robustness of the 2DPCA-based method and enhance the effective suppression of noise interference, we present a novel framework for 2DPCA in Section III.

## III. DOUBLE LAYER ROBUST 2DPCA

### A. Motivation and Method Formulation

As analyzed in Section II, although the distance metric mechanism adopted by objective function (2) demonstrates greater robustness than objective function (1), its ability to adjust the optional range under the Euclidean norm metric criterion is insufficient, resulting in the inability to achieve the essential demands of minimizing reconstruction errors in a true sense. To attack these problems, it is necessary to introduce a flexible robust norm as a distance metric criterion. Among the numerous norms, the  $l_{2,p}$ -norm as a generalized representation of the  $R_1$ -norm, not only possesses good adjustability but also preserves structural sparsity, which has a positive promoting effect on the essential requirement for minimizing projected

data reconstruction error. However, selecting the appropriate  $p$ -value for the  $l_{2,p}$ -norm in a real application is not easy, and inappropriate matching results inevitably overemphasize or underemphasize the importance degree of certain data points, seriously damaging the subspace data structure. Inspired by the successful idea of the cut norm, we consider constructing a cut  $l_{2,p}$ -norm to achieve effective avoidance of the above problems, which is defined as follows.

*Definition* [30]: Given an arbitrary matrix  $\mathbf{M} \in \mathbb{R}^{m \times n}$ , its cut  $l_{2,p}$ -norm can be written as

$$\begin{aligned} \|\mathbf{M}\|_{c,l_{2,p}} &= \sum_{i=1}^m \min \left( c, \left( \sum_{j=1}^n m_{ij}^2 \right)^{\frac{p}{2}} \right) \\ &= \sum_{i=1}^m \min \left( c, \|\mathbf{M} i,: \|_2^p \right) \end{aligned} \quad (3)$$

where  $c > 0$  is a parameter,  $\mathbf{M} i,:$  indicates the  $i$ th row of matrix  $\mathbf{M}$  and  $0 < p < 2$ .

Compared with the existing 2DPCA method that adjusts the distance metric strategy, using cut  $l_{2,p}$ -norm to measure the similarity between data can effectively improve the robustness of the method. There are three main reasons for this. Firstly, the adoption of cut  $l_{2,p}$ -norm inherits the inherent advantages of traditional  $l_{2,p}$ -norm, which can successfully suppress noise interference by flexibly selecting the  $p$ -value. Secondly, the cut  $l_{2,p}$ -norm belongs to the Euclidean distance metric category, which ensures that the low-dimensional description remains unchanged under the action of orthogonal matrices, enhancing the stability of the sample space structure. Most importantly, the parameter  $c$  in the cut  $l_{2,p}$ -norm has excellent control ability, which can reasonably improve the imbalance of sample data importance evaluation caused by flexible changes in  $p$ -values. Therefore, we have decided to employ the cut  $l_{2,p}$ -norm with excellent robustness as the distance metric criterion under the 2DPCA minimization reconstruction error mechanism, which is defined as the first layer robust structure.

On this basis and after careful consideration, it can be found that there may be deviations in the selection of the parameter  $c$  when using the cut  $l_{2,p}$ -norm, and the contribution of each training sample data to the reconstruction error is also different, which to some extent restricts the overall performance of the first layer robust structure. From the perspective of further improving method ability, we consider introducing an adaptive weighted learning framework with the support of a robust distance metric. The core goal is to adaptively find suitable weights for each sample data without the need for human intervention, achieving positive effects on samples with smaller reconstruction errors and weakening the adverse effects with larger reconstruction error samples. This allows us to truly minimize the weighted sum of reconstruction errors for image data in practical application. Therefore, the introduction of an adaptive weighting framework can be seen as a second layer robust structure that compensates for the previous layer.

Finally, we hope that the proposed method can enhance robustness while still retaining the valuable qualities of traditional 2DPCA, which means that the obtained optimal

solution is closely related to the covariance matrix of the input data. This not only fully protects the global geometric structure of the sample space but also potentially contributes to the improvement of recognition accuracy.

Based on the analysis mentioned above, we introduce a novel double layer robust learning architecture for 2DPCA. The objective function for determining the optimal projection direction  $\mathbf{V}$  is

$$\begin{cases} \operatorname{argmin} \sum_{i=1}^N \sum_{j=1}^m \frac{1}{s_{ij}} \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V} \mathbf{V}^T \right\|_2^p \right) \\ \text{s.t. } \mathbf{s} \geq \mathbf{0}, \mathbf{s}^T \mathbf{1} = 1, \mathbf{V}^T \mathbf{V} = \mathbf{I}_q \end{cases} \quad (4)$$

where  $1/s_{ij} \in (1, +\infty)$  is the weighted coefficient assigned to the reconstruction error of the  $ij$  th input sample data, and satisfies  $\sum_{i=1}^N \sum_{j=1}^m s_{ij} = 1$ ,  $0 \leq s_{ij} \leq 1$ ,  $\mathbf{s} = (s_{11}, s_{12}, \dots, s_{Nm})^T$  indicates the weighted vector.

To demonstrate the robustness advantages of the proposed method, we consider using MATLAB function to create some manually sample points as instance validation data, including clean data (green circles) and noisy points (red pentagrams), as shown in Fig. 1. On this basis, we sequentially obtain the optimal projection directions of 2DPCA, 2DPCA-L1, F-2DPCA, R1-2DPCA, GC-2DPCA, Angle-2DPCA, Cos-2DPCA, 2DPCA-2-Lp and DLR-2DPCA, where  $\mathbf{V}_{2DPCA}$  indicates the expected projection direction based on the 2DPCA method under noise free interference conditions,  $\mathbf{V}_{2DPCA-L1}^{\text{noise}}$ ,  $\mathbf{V}_{F-2DPCA}^{\text{noise}}$ ,  $\mathbf{V}_{R1-2DPCA}^{\text{noise}}$ ,  $\mathbf{V}_{GC-2DPCA}^{\text{noise}}$ ,  $\mathbf{V}_{\text{Angle-2DPCA}}^{\text{noise}}$ ,  $\mathbf{V}_{\text{Cos-2DPCA}}^{\text{noise}}$ ,  $\mathbf{V}_{2DPCA-2-Lp}^{\text{noise}}$  and  $\mathbf{V}_{\text{DLR-2DPCA}}^{\text{noise}}$  represent the actual projection directions of each comparison method when the training data contains noise. From the experimental results in Fig. 1, it can be seen that compared with other comparison methods, DLR-2DPCA is closest to the expected projection direction, which effectively verifies the excellent suppression ability of our proposed method against noise interference.

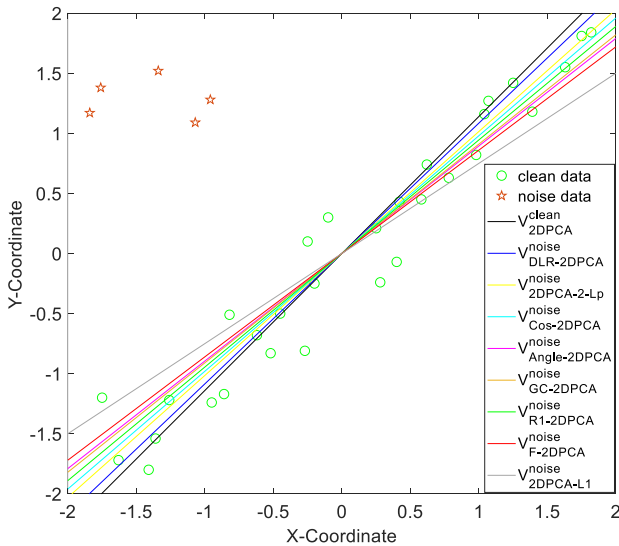


Fig. 1 Projection vectors of 2DPCA and all comparison methods on manually dataset with/without noise.

## B. Algorithm

According to basic algebraic knowledge, objective function (4) can be transformed to

$$\begin{aligned} & \sum_{i=1}^N \sum_{j=1}^m \frac{1}{s_{ij}} \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V} \mathbf{V}^T \right\|_2^p \right) \\ & = \sum_{i=1}^N \sum_{j=1}^m \frac{2}{p} \left( \frac{1}{s_{ij}} u_{ij} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V} \mathbf{V}^T \right\|_2^p \right) + c \end{aligned} \quad (5)$$

where  $u_{ij}$  can be defined as

$$u_{ij} = \begin{cases} \frac{p}{2} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V} \mathbf{V}^T \right\|_2^{p-2} & \text{if } \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V} \mathbf{V}^T \right\|_2^p < c \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

By utilizing the property of the matrix trace, Eq. (5) becomes

$$\frac{2}{p} \sum_{i=1}^N \sum_{j=1}^m \frac{1}{s_{ij}} \left( \operatorname{tr} \left( \mathbf{A}_i(j,:)^T \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V} \mathbf{V}^T \mathbf{A}_i(j,:)^T \right) u_{ij} + c \right) \quad (7)$$

where  $\operatorname{tr}(\cdot)$  denotes the trace operator of a matrix.

Since  $2/p (0 < p < 2) > 0$  and  $c > 0$  are constants, objective function (4) is equivalent to

$$\begin{cases} \operatorname{argmin} \sum_{i=1}^N \sum_{j=1}^m \frac{u_{ij}}{s_{ij}} \left( \operatorname{tr} \left( \mathbf{A}_i(j,:)^T \mathbf{A}_i(j,:) \right) - \operatorname{tr} \left( \mathbf{V}^T \mathbf{A}_i(j,:)^T \mathbf{A}_i(j,:) \mathbf{V} \right) \right) \\ \text{s.t. } \mathbf{s} \geq \mathbf{0}, \mathbf{s}^T \mathbf{1} = 1, \mathbf{V}^T \mathbf{V} = \mathbf{I}_q \end{cases} \quad (8)$$

The objective function (8) contains three unknown joint variables  $u_{ij}$ ,  $s_{ij}$  and  $\mathbf{V}$ , where  $u_{ij}$  and  $s_{ij}$  are closely related to  $\mathbf{V}$ . Therefore, it is very difficult to directly obtain the optimal projection matrix  $\mathbf{V}$  with a closed form based on objective function (8). To overcome this obstacle, we consider seeking the optimal projection direction  $\mathbf{V}$  via the optimization strategy of alternating solutions and block coordinate descent. Specifically, the solution process can be divided into the following three steps.

**Step 1:** Fix  $\mathbf{V}$ , perform the update operation of  $u_{ij}$ .

According to Eq. (6), we can capture the solution of  $u_{ij}$ .

Notably when  $\left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V} \mathbf{V}^T \right\|_2^p < c$ ,  $u_{ij}$  can be defined as follows

$$\begin{aligned} u_{ij} & = \frac{p}{2} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V} \mathbf{V}^T \right\|_2^{p-2} \\ & = \frac{p}{2} \frac{1}{\left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V} \mathbf{V}^T \right\|_2^{2-p}} \end{aligned} \quad (9)$$

In Eq. (9), the denominator  $\left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V} \mathbf{V}^T \right\|_2^{2-p}$  may have a state equal to zero, which results in the inability to provide an effective definition of  $u_{ij}$ . Based on the above analysis, we consider introducing a constant strategy to avoid the occurrence of this phenomenon. Therefore, Eq. (6) can be rewritten as

$$u_{ij} = \begin{cases} \frac{p}{2} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V} \mathbf{V}^T \right\|_2^{p-2} + z & \text{if } \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V} \mathbf{V}^T \right\|_2^p < c \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

where  $z$  is a small non-zero constant.

**Step 2:** Fix  $u_{ij}$  and  $\mathbf{V}$ , perform the update operation of  $s_{ij}$ . Because  $u_{ij}$  and  $\mathbf{V}$  are treated as fixed terms, we can make

$k_{ij} = \min\left(c, \|\mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V} \mathbf{V}^T\|_2^p\right)$ . Thus, the objective function function (7) can be rewritten as

$$\begin{cases} \operatorname{argmin} \sum_{i=1}^N \sum_{j=1}^m \frac{k_{ij}}{s_{ij}} \\ \text{s.t. } \sum_{i=1}^N \sum_{j=1}^m s_{ij} = 1, 0 \leq s_{ij} \leq 1 \end{cases} \quad (11)$$

Combined with the Lagrange multiplier method, the Lagrange function of objective function (11) can be defined as

$$L = \sum_{i=1}^N \sum_{j=1}^m \frac{k_{ij}}{s_{ij}} + \lambda \left(1 - \sum_{i=1}^N \sum_{j=1}^m s_{ij}\right) + \sum_{i=1}^N \sum_{j=1}^m \mu_{ij} (-s_{ij}) \quad (12)$$

where  $\lambda$  and  $\mu_{ij}$  are Lagrange multipliers. The KKT conditions of Eq. (12) are displayed as follows

$$\begin{cases} -\frac{k_{ij}}{(s_{ij})^2} - \lambda - \mu_{ij} = 0 \\ \mu_{ij} s_{ij} = 0 \\ \mu_{ij} \geq 0 \\ \sum_{i=1}^N \sum_{j=1}^m s_{ij} = 1 \end{cases} \quad (13)$$

By integrating Eq. (13) and performing algebraic operations, we can obtain the optimal solution of  $s_{ij}$ .

$$s_{ij} = \frac{\sqrt{k_{ij}}}{\sum_{i=1}^N \sum_{j=1}^m \sqrt{k_{ij}}} \quad (14)$$

**Step 3:** Fix  $u_{ij}$  and  $s_{ij}$ , perform the update operation of  $\mathbf{V}$ .

When the solved  $u_{ij}$  (using step 1) and  $s_{ij}$  (using step 2) are fixed, the first term in objective function (8) can be considered a constant. Therefore, Eq. (8) further becomes

$$\begin{cases} \operatorname{argmax} \operatorname{tr}(\mathbf{V}^T \tilde{\mathbf{A}}^T \mathbf{G} \tilde{\mathbf{A}} \mathbf{V}) \\ \text{s.t. } \mathbf{s} \geq \mathbf{0}, \mathbf{s}^T \mathbf{1} = 1, \mathbf{V}^T \mathbf{V} = \mathbf{I}_q \end{cases} \quad (15)$$

where  $\tilde{\mathbf{A}} \in \mathfrak{R}^{Nm \times n}$  is the new sample enhancement matrix constructed by sequentially arranging each row of the original input image data  $\mathbf{A}_i(j,:)$ ,  $\mathbf{G} \in \mathfrak{R}^{Nm \times Nm}$  is a diagonal matrix composed of diagonal elements  $u_{ij}/s_{ij}$ . Combined with the matrix correlation theorem, the optimal projection matrix  $\mathbf{V}$  can be determined based on the eigenvector of the weighted covariance matrix  $\tilde{\mathbf{A}}^T \mathbf{G} \tilde{\mathbf{A}}$  corresponding to the first  $q$  largest eigenvalues. Finally, according to Eq. (9), Eq. (13) and Eq. (14), we summarize an iterative algorithm with nongreedy properties to capture the optimal solution of objective function (7). The pseudo code of the detailed solution process is presented in *Algorithm 1*.

### C. Convergence Analysis

We first introduce the following lemma to help verify the convergence of *Algorithm 1* [31].

### Algorithm 1 DLR-2DPCA

**Input:**  $\mathbf{A}_i \in \mathfrak{R}^{m \times n}$ ,  $p$ ,  $q$ ,  $c$  and  $h$ , where  $\mathbf{A}_i$  is centralized.

**Initialize:**  $\mathbf{V}^{t-1} \in \mathfrak{R}^{n \times q}$  which satisfies  $(\mathbf{V}^{t-1})^T \mathbf{V}^{t-1} = \mathbf{I}_q$ ,  $t = 1$ ,  $\delta = 0.01$ .

**while** not converge **do**

1. Compute  $u_{ij}^{t-1}$  by using Eq. (9).
2. Compute  $s_{ij}^{t-1}$  obtained by Eq. (13).
3. Compute diagonal matrix  $\mathbf{G}^{t-1}$  based on  $u_{ij}^{t-1}$  and  $s_{ij}^{t-1}$ , where the diagonal element of  $\mathbf{G}^{t-1}$  is  $u_{ij}^{t-1}/s_{ij}^{t-1}$ .
4. Build augmented matrix  $\tilde{\mathbf{A}} = (\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_I) \in \mathfrak{R}^{Nm \times n}$ .
5. Solve the optimal solution  $\mathbf{V}^t$  in objective function (15).
6. Determine convergence condition  $J(\mathbf{V}^t) - J(\mathbf{V}^{t-1}) \leq \delta$ , if satisfied, proceed step 8, else proceed step 7.
7. Capture  $\mathbf{V}^t$  that satisfies  $J(\mathbf{V}^t) \leq J(\mathbf{V}^{t-1})$  by using sub-gradient with Armijo line search [32]. If the solution exists proceed step 6, else end while.
8. Update  $u_{ij}^t$  and  $s_{ij}^t$  based on Eq. (9) and Eq. (14).
9.  $t \leftarrow t + 1$ .

**end while**

**Output:**  $\mathbf{V}^t \in \mathfrak{R}^{n \times q}$ .

**Lemma 1:** For any two nonzero vectors  $\mathbf{e}^t, \mathbf{e}^{t+1} \in \mathfrak{R}^m$ , when  $0 < p < 2$ , the inequality

$$\begin{aligned} & \left\| \mathbf{e}^{(t+1)} \right\|_2^p - \frac{p}{2} \left\| \mathbf{e}^{(t)} \right\|_2^{p-2} \left\| \mathbf{e}^{(t+1)} \right\|_2^2 \\ & \leq \left\| \mathbf{e}^{(t)} \right\|_2^p - \frac{p}{2} \left\| \mathbf{e}^{(t)} \right\|_2^{p-2} \left\| \mathbf{e}^{(t)} \right\|_2^2 \end{aligned} \quad (16)$$

can be obtained.

**Theorem 1:** In the  $(t+1)$  th iteration of *Algorithm 1*, for any indices  $i$  and  $j$ , we have the inequality

$$\begin{aligned} & \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \right) \\ & \quad - u_{ij}^{(t)} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^2 \\ & \leq \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p \right) \\ & \quad - u_{ij}^{(t)} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^2 \end{aligned} \quad (17)$$

satisfied.

**Proof:** Considering that the validation based on Eq. (17) is closely related to the selection of the minimum value between  $c$  and  $\left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p$ . Therefore, it is necessary for us to discuss the following two situations separately, which are described as follows.

**First case:** Assuming  $\left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p > c$  holds, then  $u_{ij}^{(t)} = 0$ . If  $\left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p > c$ , we have

$$\begin{aligned}
& \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \right) \\
& \quad - u_{ij}^{(t)} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^2 \\
& = \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p \right) \\
& \quad - u_{ij}^{(t)} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^2 = c
\end{aligned} \tag{18}$$

If  $\left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \leq c$ , we can obtain the following inequality

$$\begin{aligned}
& \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \right) \\
& \quad - u_{ij}^{(t)} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^2 \\
& \leq \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p \right) \\
& \quad - u_{ij}^{(t)} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^2 = c
\end{aligned} \tag{19}$$

Combining Eq. (18) and Eq. (19), it can be demonstrated that *Theorem 1* holds when the assumption in the first case, i.e.  $\left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p > c$ , is met.

**Second case:** Assuming  $\left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p < c$  holds, we have

$$\begin{aligned}
& \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p \right) \\
& = \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p
\end{aligned} \tag{20}$$

Assuming  $\mathbf{e}^{(t+1)} = \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p$  and  $\mathbf{e}^{(t)} = \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p$ , combined with *Lemma 1*, we have

$$\begin{aligned}
& \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \\
& \quad - u_{ij}^{(t)} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^2 \\
& \leq \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p \\
& \quad - u_{ij}^{(t)} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^2
\end{aligned} \tag{21}$$

According to the assumptions of the second case and the information conveyed by *Algorithm 1*, we can derive the following inequality

$$\begin{aligned}
& \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \\
& \leq \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p \leq c
\end{aligned} \tag{22}$$

According to Eq. (22), we can obtain

$$\begin{aligned}
& \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \right) \\
& = \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p
\end{aligned} \tag{23}$$

Substituting Eq. (20) and Eq. (23) into Eq. (21) yields

$$\begin{aligned}
& \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \right) \\
& \quad - u_{ij}^{(t)} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^2 \\
& \leq \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p \right) \\
& \quad - u_{ij}^{(t)} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^2
\end{aligned} \tag{24}$$

Eq. (24) indicates that *Theorem 1* still holds when the second case occurs, i.e.,  $\left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p < c$  is met.

**Theorem 2:** In the  $(t+1)$  th iteration of *Algorithm 1*, we have

$$\begin{aligned}
& \sum_{i=1}^N \sum_{j=1}^m \frac{1}{s_{ij}^{(t+1)}} \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \right) \\
& \leq \sum_{i=1}^N \sum_{j=1}^m \frac{1}{s_{ij}^{(t)}} \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p \right)
\end{aligned} \tag{25}$$

i.e., *Algorithm 1* monotonically decreases the objective function value of Eq. (4).

*Proof:* Based on the  $(t+1)$  th iteration of *Algorithm 1*, we have

$$\text{tr} \left( \left( \mathbf{V}^{(t+1)} \right)^T \tilde{\mathbf{A}}^T \mathbf{G}^{(t)} \tilde{\mathbf{A}} \mathbf{V}^{(t+1)} \right) \geq \text{tr} \left( \left( \mathbf{V}^{(t)} \right)^T \tilde{\mathbf{A}}^T \mathbf{G}^{(t)} \tilde{\mathbf{A}} \mathbf{V}^{(t)} \right) \tag{26}$$

Substituting  $\tilde{\mathbf{A}}^T \mathbf{G} \tilde{\mathbf{A}} = p/2 \sum_{i=1}^N \sum_{j=1}^m \left( \mathbf{A}_i(j,:) \right)^T \left( u_{ij} + 1/s_{ij} \right) \mathbf{A}_i(j,:)$  into Eq. (20), we have

$$\begin{aligned}
& \sum_{i=1}^N \sum_{j=1}^m \frac{u_{ij}^{(t)}}{s_{ij}^{(t)}} \text{tr} \left( \left( \mathbf{V}^{(t+1)} \right)^T \mathbf{A}_i(j,:) \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \right) \\
& \geq \sum_{i=1}^N \sum_{j=1}^m \frac{u_{ij}^{(t)}}{s_{ij}^{(t)}} \text{tr} \left( \left( \mathbf{V}^{(t)} \right)^T \mathbf{A}_i(j,:) \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \right)
\end{aligned} \tag{27}$$

Multiplying  $-1$  and adding  $\sum_{i=1}^N \sum_{j=1}^m \left( u_{ij}^{(t)} / s_{ij}^{(t)} \right) \text{tr} \left( \mathbf{A}_i(j,:) \mathbf{A}_i(j,:) \right)$  on both sides of Eq. (27), we can obtain

$$\begin{aligned}
& \sum_{i=1}^N \sum_{j=1}^m \frac{u_{ij}^{(t)}}{s_{ij}^{(t)}} \text{tr} \left( \mathbf{A}_i(j,:)^T \mathbf{A}_i(j,:) \right) \\
& \quad - \sum_{i=1}^N \sum_{j=1}^m \frac{u_{ij}^{(t)}}{s_{ij}^{(t)}} \text{tr} \left( \left( \mathbf{V}^{(t+1)} \right)^T \mathbf{A}_i(j,:)^T \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \right) \\
& \leq \sum_{i=1}^N \sum_{j=1}^m \frac{u_{ij}^{(t)}}{s_{ij}^{(t)}} \text{tr} \left( \mathbf{A}_i(j,:)^T \mathbf{A}_i(j,:) \right) \\
& \quad - \sum_{i=1}^N \sum_{j=1}^m \frac{u_{ij}^{(t)}}{s_{ij}^{(t)}} \text{tr} \left( \left( \mathbf{V}^{(t)} \right)^T \mathbf{A}_i(j,:)^T \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \right)
\end{aligned} \tag{28}$$

Owing to  $\text{tr} \left( \mathbf{A}_i(j,:)^T \mathbf{A}_i(j,:) \right) - \text{tr} \left( \mathbf{V}^T \mathbf{A}_i(j,:)^T \mathbf{A}_i(j,:) \mathbf{V} \right) = \left\| \mathbf{V}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V} \mathbf{V}^T \right\|_2^2$ , we can convert Eq. (28) as follows

$$\begin{aligned}
& \sum_{i=1}^N \sum_{j=1}^m \frac{u_{ij}^{(t)}}{s_{ij}^{(t)}} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^2 \\
& \leq \sum_{i=1}^N \sum_{j=1}^m \frac{u_{ij}^{(t)}}{s_{ij}^{(t)}} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^2
\end{aligned} \tag{29}$$

According to  $u_{ij}^{(t)} > 0$  and  $s_{ij}^{(t)} > 0$ , it is known that for each index  $i$  and  $j$ , Eq. (29) can be transformed as follows

$$\begin{aligned}
& \sum_{i=1}^N \sum_{j=1}^m u_{ij}^{(t)} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^2 \\
& \leq \sum_{i=1}^N \sum_{j=1}^m u_{ij}^{(t)} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^2
\end{aligned} \tag{30}$$

Based on the *Theorem 1*, we can deduce that the sum of all indexes satisfies the following inequality

$$\begin{aligned}
& \sum_{i=1}^N \sum_{j=1}^m \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \right) \\
& \quad - u_{ij}^{(t)} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^2 \\
& \leq \sum_{i=1}^N \sum_{j=1}^m \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p \right) \\
& \quad - u_{ij}^{(t)} \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^2
\end{aligned} \tag{31}$$

Combining Eq. (30) and Eq. (31), we have

$$\begin{aligned}
& \sum_{i=1}^N \sum_{j=1}^m \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \right) \\
& \leq \sum_{i=1}^N \sum_{j=1}^m \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p \right)
\end{aligned} \tag{32}$$

Multiplying  $1/s_{ij}^{(t)}$  on both sides of Eq. (32), then Eq. (32) can become

$$\begin{aligned}
& \sum_{i=1}^N \sum_{j=1}^m \frac{1}{s_{ij}^{(t)}} \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \right) \\
& \leq \sum_{i=1}^N \sum_{j=1}^m \frac{1}{s_{ij}^{(t)}} \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t)} \left( \mathbf{V}^{(t)} \right)^T \right\|_2^p \right)
\end{aligned} \tag{33}$$

According to the definitions of  $k_{ij}$  and  $s_{ij}$ , it is not difficult to know that  $\sum_{i=1}^N \sum_{j=1}^m \sqrt{k_{ij}^{(t)}} \geq \sum_{i=1}^N \sum_{j=1}^m \sqrt{k_{ij}^{(t+1)}} > 0$ ,  $\sum_{i=1}^N \sum_{j=1}^m s_{ij}^{(t)} = \sum_{i=1}^N \sum_{j=1}^m s_{ij}^{(t+1)} = 1$ . Therefore, we have

$$\begin{aligned}
& \sum_{i=1}^N \sum_{j=1}^m \frac{1}{s_{ij}^{(t)}} \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \right) \\
& \geq \sum_{i=1}^N \sum_{j=1}^m \frac{1}{s_{ij}^{(t+1)}} \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \right)
\end{aligned} \tag{34}$$

By combining Eq. (33) and Eq. (34), we obtain the following inequality

$$\begin{aligned}
& \sum_{i=1}^N \sum_{j=1}^m \frac{1}{s_{ij}^{(t+1)}} \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \right) \\
& \leq \sum_{i=1}^N \sum_{j=1}^m \frac{1}{s_{ij}^{(t)}} \min \left( c, \left\| \mathbf{A}_i(j,:) - \mathbf{A}_i(j,:) \mathbf{V}^{(t+1)} \left( \mathbf{V}^{(t+1)} \right)^T \right\|_2^p \right)
\end{aligned} \tag{35}$$

The results presented in Eq. (35) fully validate that *Algorithm 1* monotonically reduces the value of objective function (4) during each iteration process. Moreover, the objective function (4) clearly has a lower boundary and its value is 0. Therefore, by combining the effects of monotonicity and boundedness, it can be determined that the iterative process of *Algorithm 1* has good convergence.

#### IV. EXPERIMENTS

We consider evaluating the robust performance of DLR-2DPCA based on small sample size image databases from three different application fields (CMU PIE [33], GTSRB [34] and MNIST [35]). To fully demonstrate the advantages of DLR-2DPCA under the same experimental conditions, we choose to compare it with the existing robust 2DPCA methods 2DPCA-L1 [17],  $F$ -2DPCA [23],  $R_1$ -2DPCA [24], GC-2DPCA [26], Angle-2DPCA [27], Cos-2DPCA [28] and 2DPCA-2-L<sub>p</sub> [29]. Moreover, for experiments in the CMU PIE, GTSRB and MNIST databases, we mainly validated the recognition accuracy and reconstruction performance of the method by changing the feature dimension (range interval set between 0 and 25). The recognition accuracy can be evaluated via the results of the 1-nearest neighbor (1NN) classifier, and the reconstruction error  $e$  can be obtained via the following expression

$$e = \frac{1}{N} \frac{1}{m} \sum_{i=1}^N \sum_{j=1}^m \left\| \mathbf{A}_i(j,:)_{\text{clean}} - \mathbf{A}_i(j,:)_{\text{clean}} \mathbf{V} \mathbf{V}^T \right\|_2 \tag{36}$$

where  $\mathbf{A}_i(j,:)_{\text{clean}}$  denotes the  $ij$  th sample data without noise interference.

Finally, to facilitate quick observation of the method with the best performance, we consider bolding the optimal values of the above evaluation indicators among all comparison methods.

##### A. Database Introduction

The CMU PIE database [33] contains 2856 facial images of 68 individuals under different lighting conditions, with each

person containing 42 images. In the experiments, the size of each image sample was standardized to  $32 \times 32$  pixels. 13 images were randomly selected from the 42 images of each person and rectangular block occlusion was introduced. The occlusion was randomly distributed and its proportion is set between 5% and 25% of the interest area. Fig. 2 shows sample images of several facial based on the CMU PIE database. We randomly selected 25 images, which included noisy images, each person for training, and the rest for testing. We repeat the above process 10 times.

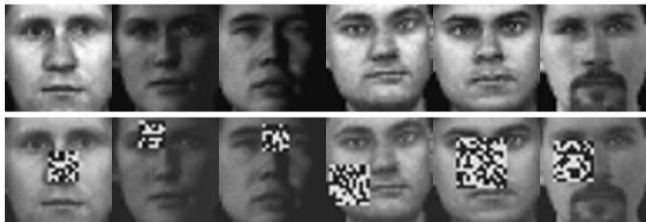


Fig. 2 Several image samples containing clean data and noise interference on the CMU PIE database

The GTSRB database [34] collects over 6500 sample images containing 42 types of traffic signs, each of which has multiple states, such as occlusion, fog, different lighting conditions and background changes. We consider 35 types of traffic signs with both clean and noisy images for data analysis, with 30 images for each type of traffic sign. The size of each image is uniformly processed to  $32 \times 32$  pixels and subjected to grayscale operation. Fig. 3 shows some sample images in the GTSRB database. In the experiment, we randomly selected 18 images of each traffic sign (including noisy images) as training samples, and the remaining images were used for testing. The above process was executed 10 times.



Fig. 3 Several image samples containing clean data and noise interference on the GTSRB database

The MNIST database [35] contains handwritten digit images from 250 different individuals, all of which are between 0 and 9. The entire image data was manually divided into a training set (60000 images) and a testing set (10000 images), with each image being resized to  $28 \times 28$  pixels. In practical applications, it is not always convenient to collect such a large dataset for handwritten digits other than 0 to 9, and images may not always have appropriate clarity. We chose 40 images from each digit in the original training and testing databases for experimental analysis, and arbitrarily selected several images from them to place mixed noise, which is composed of Gaussian noise and salt-and-pepper noise. The noise intensity is set to vary between 0.005 and 0.050 and is randomly distributed. Fig. 4 shows several clean images and corresponding noisy images in the MNIST database. We randomly select 24 images (including noisy images) of each handwritten digit for training, and the rest for testing. We performed the above operation 10 times.



Fig. 4 Several image samples containing clean data and noise interference on the MNIST database

## B. Parameter Selection

In the proposed DLR-2DPCA method, the cut parameter  $c$ , as an important indicator to measure its robustness, needs to be reasonably selected, which is closely related to the number of noisy sample data [30]. Starting from the strategy of reducing the search range, we usually consider setting the  $c$  value to be  $h \in [0, 1]$  times the training sample, which is also considered the content of noisy data. Thus, the cut parameter  $c$  can be written as follows

$$c = Nm \times h \quad (37)$$

where  $Nm$  indicates the number of training samples,  $h$  is a new parameter.

Now, the issue of selecting the cut parameter  $c$  has been transformed into the optimal determination of  $h$ . However, considering the unpredictable nature of data distributions in real-world application environments, accurately estimating the quantity of noisy data is difficult. To improve the universality for actual conditions and select a more suitable  $h$  value, we have decided to utilize a  $k$ -fold cross validation scheme to complete parameter acquisition [36]. For the CMU PIE, GTSRB and MNIST image databases, the  $k$  values are determined to be 10, 7 and 10, respectively, with corresponding training sample numbers of 1700, 630 and 240.

## C. Experimental Results

Tables I-III list the average recognition accuracy and average reconstruction error with the corresponding standard deviation (Std) of all methods on the CMU PIE, GTSRB and MNIST databases, respectively. Table IV lists the average time consumption with the corresponding Std of all experimental comparison methods on the three image databases. Figs. 6(a), 7(a) and 8(a) plot the average recognition rate curve versus the number of different feature dimensions on the CMU PIE, GTSRB and MNIST databases, respectively. Figs. 6(b), 7(b) and 8(b) show the reconstruction error histograms of each comparison method under ten sets of experiments on the three image databases, respectively. Fig. 9 present the convergence curve of the proposed method on the CMU PIE, GTSRB and MNIST databases.

Based on a reasonable analysis of the results obtained from the above experimental content, we can draw the following meaningful conclusions.

(1) According to the results presented in Tables I-III and Figs. 6-8, it is not difficult to concluded that compared with other robust 2DPCA methods, DLR-2DPCA has relatively excellent image feature extraction ability in complex noise environments. One of the main reasons is that DLR-2DPCA uses the cut  $l_{2,p}$ -norm as a distance metric criterion based on the first layer robust

TABLE I  
THE AVERAGE RECOGNITION ACCURACY (%) AND AVERAGE RECONSTRUCTION ERROR  
WITH THE CORRESPONDING STD OF EACH METHOD ON THE CMU PIE DATABASE

Methods	2DPCA-L1 [17]	F-2DPCA [23]	R1-2DPCA [24]	GC-2DPCA [26]	Angle-2DPCA [27]	Cos-2DPCA [28]	2DPCA-2-L <sub>p</sub> [29]	DLR-2DPCA		
								$p = 1.5$	$p = 1$	$p = 0.5$
Accuracy	69.97 ±0.82	76.94 ±0.78	80.06 ±0.76	78.27 ±0.85	77.65 ±0.80	81.32 ±0.73	81.82 ±0.77	83.85 ±0.73	82.24 ±0.71	82.71 ±0.75
Error	36.65 ±4.86	34.21 ±3.45	30.15 ±2.11	31.42 ±2.65	32.97 ±2.91	29.92 ±1.70	28.52 ±2.02	23.40 ±2.46	25.21 ±2.38	24.48 ±2.52

TABLE II  
THE AVERAGE RECOGNITION ACCURACY (%) AND AVERAGE RECONSTRUCTION ERROR  
WITH THE CORRESPONDING STD OF EACH METHOD ON THE GTSRB DATABASE

Methods	2DPCA-L1 [17]	F-2DPCA [23]	R1-2DPCA [24]	GC-2DPCA [26]	Angle-2DPCA [27]	Cos-2DPCA [28]	2DPCA-2-L <sub>p</sub> [29]	DLR-2DPCA		
								$p = 1.5$	$p = 1$	$p = 0.5$
Accuracy	64.93 ±0.89	69.21 ±0.83	71.43 ±1.06	71.79 ±0.84	71.05 ±0.85	72.31 ±0.74	72.78 ±0.88	73.27 ±0.79	73.64 ±0.83	<b>74.24</b> <b>±0.81</b>
Error	151.36 ±1.30	138.53 ±1.22	132.28 ±1.14	130.62 ±1.28	133.49 ±1.17	128.17 ±1.21	126.91 ±1.25	125.23 ±1.11	124.04 ±1.04	<b>122.75</b> <b>±1.07</b>

TABLE III  
THE AVERAGE RECOGNITION ACCURACY (%) AND AVERAGE RECONSTRUCTION ERROR  
WITH THE CORRESPONDING STD OF EACH METHOD ON THE MNIST DATABASE

Methods	2DPCA-L1 [17]	F-2DPCA [23]	R1-2DPCA [24]	GC-2DPCA [26]	Angle-2DPCA [27]	Cos-2DPCA [28]	2DPCA-2-L <sub>p</sub> [29]	DLR-2DPCA		
								$p = 1.5$	$p = 1$	$p = 0.5$
Accuracy	75.44 ±1.67	80.50 ±1.61	82.32 ±1.56	82.94 ±1.82	82.00 ±1.74	83.50 ±1.72	83.96 ±1.77	84.75 ±1.59	85.00 ±1.50	<b>85.57</b> <b>±1.57</b>
Error	100.49 ±2.53	93.87 ±2.46	91.64 ±2.37	90.93 ±2.68	92.05 ±2.72	89.87 ±2.60	88.95 ±2.33	87.81 ±2.09	87.33 ±2.16	<b>86.62</b> <b>±2.25</b>

TABLE IV  
THE AVERAGE TRAINING TIME AND THE CORRESPONDING STD OF EACH METHOD ON ALL EXPERIMENTAL DATABASES

Methods	2DPCA-L1 [17]	F-2DPCA [23]	R1-2DPCA [24]	GC-2DPCA [26]	Angle-2DPCA [27]	Cos-2DPCA [28]	2DPCA-2-L <sub>p</sub> [29]	DLR-2DPCA (optimal $p$ )
CMU PIE	4.37 ±0.35	<b>0.82</b> <b>±0.16</b>	1.04 ±0.28	1.77 ±0.40	2.64 ±0.42	6.29 ±0.38	1.06 ±0.33	1.13 ±0.21
GTSRB	5.08 ±0.32	<b>0.97</b> <b>±0.10</b>	1.15 ±0.08	2.32 ±0.24	2.70 ±0.19	8.41 ±0.16	1.20 ±0.18	1.32 ±0.13
MNIST	5.79 ±0.71	<b>1.12</b> <b>±0.19</b>	1.34 ±0.25	2.66 ±0.31	3.18 ±0.37	8.75 ±0.43	1.39 ±0.41	1.56 ±0.24

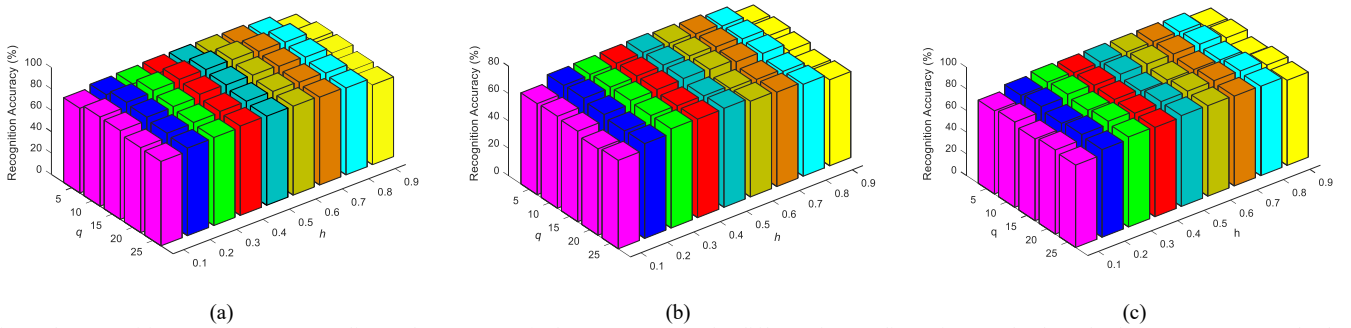
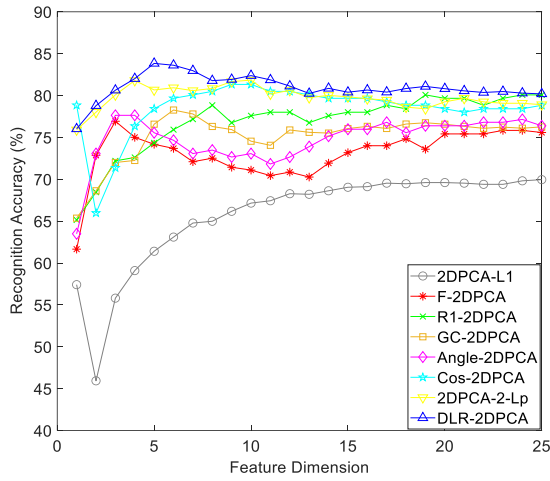


Fig. 5 The recognition accuracy corresponding to the parameter  $h$  of DLR-2DPCA under different feature dimensions on the three databases. (a) CMU PIE database. (b) GTSRB database. (c) MNIST database

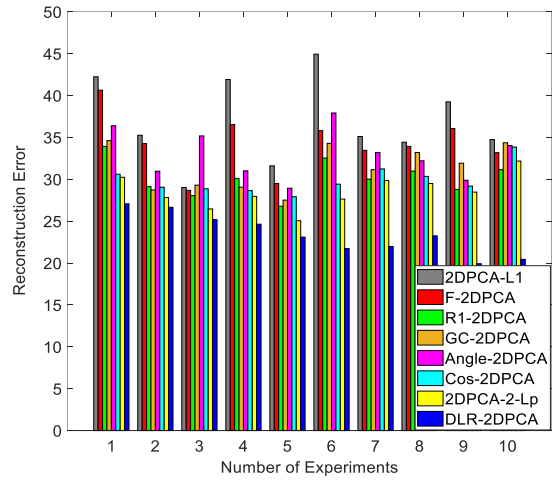
structure, which not only increases the flexibility of the robust metric strategy, but also avoids the phenomenon of some data point evaluation imbalances due to improper  $p$ -value selection, effectively weakening the adverse effects of noise in measuring the similarity between data. Another important reason is that DLR-2DPCA reasonably builds an adaptive weighted learning framework for the second layer robust structure, which can adaptively allocate weights based on the state of each sample data without human intervention, further improving the accurate capture ability of important information in noisy

environments. Moreover, DLR-2DPCA retains the rotational invariance of the traditional 2DPCA method, fully ensuring the stability of low-dimensional structures during feature extraction tasks. Finally, the objective function of DLR-2DPCA was described via the strategy of minimizing the reconstruction error, which is the true core requirement based on the 2DPCA method.

(2) Further observation of the experimental results reflected in Tables I-III and Figs. 6-8 reveals that although 2DPCA-L1 utilizes the  $l_1$ -norm with a robust attribute as a distance metric

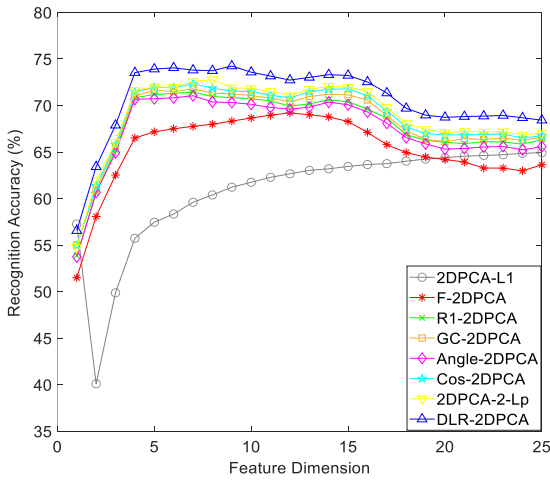


(a)

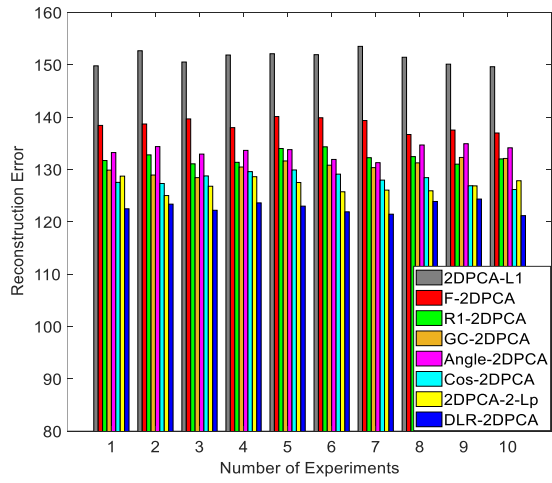


(b)

Fig. 6 The optimal recognition accuracy and minimum reconstruction error of each comparison method on the NF database. (a) Recognition accuracy versus the different feature dimension. (b) Reconstruction error versus the number of experiments

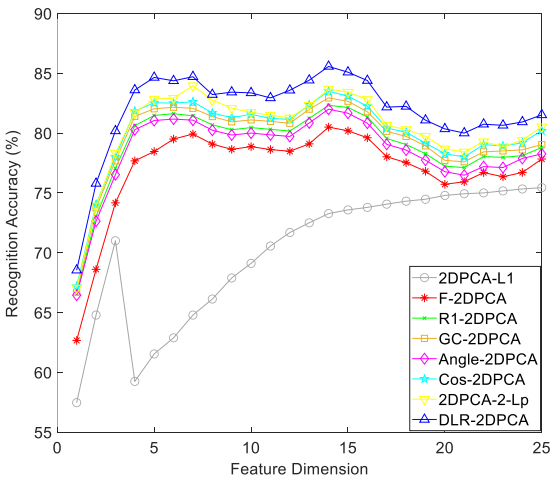


(a)

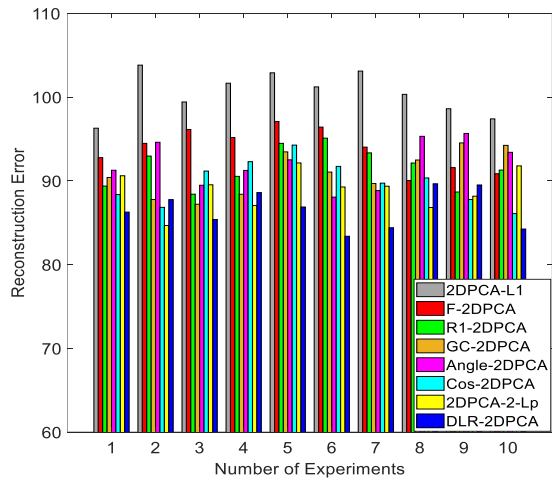


(b)

Fig. 7 The optimal recognition accuracy and minimum reconstruction error of each comparison method on the GTSRB database. (a) Recognition accuracy versus the different feature dimension. (b) Reconstruction error versus the number of experiments



(a)



(b)

Fig. 8 The optimal recognition accuracy and minimum reconstruction error of each comparison method on the MNIST database. (a) Recognition accuracy versus the different feature dimension. (b) Reconstruction error versus the number of experiments

mechanism, its overall performance is significantly inferior to that of the other comparison methods. The emergence of this result may be due to the following main reasons. First, 2DPCA-L1 adopts a metric strategy of maximizing variance based on the objective function, which is not the essential requirement of the 2DPCA method. Moreover, the optimization process of 2DPCA-L1 presents a relatively high computational cost, which invisibly increases the sensitivity of the method itself. Most importantly, the use of the  $l_1$ -norm metric mode achieved an improvement in robustness while losing valuable properties such as rotational invariance, which did not meet the expected goal.

(3) Based on recognition accuracy and reconstruction error experiments, the overall ability of  $F$ -2DPCA in the Euclidean norm distance metric method is relatively insufficient, mainly because it considers only the importance of robust measurement strategies for noise data, while neglecting the exploration of structural establishment and recognition patterns. For example,  $R_1$ -2DPCA utilizes a nuclear norm INN classifier to improve the impact of physical changes on recognition results. GC-2DPCA utilizes the center-weight-based strategy to reduce the adverse effects of noise information interference on feature description under the premise of robust measurement. Angle-2DPCA and Cos-2DPPCA successfully balances the respective demands of reconstruction error and variance in structural relationships, enabling low-dimensional features to have sufficient capability to capture important classification information. 2DPCA-2- $L_p$  selects the joint use of 2-norm and  $l_p$ -norm to obtain flexible distance metric criteria, which comprehensively suppresses noise interference.

(4) Via careful observation of Tables I-III, it can be observed that the recognition accuracy and reconstruction error obtained in the GTSRB database are clearly inferior to those of the other two databases. This is mainly because the GTSRB database contains images with significant differences in lighting and background, whereas low-dimensional PCA-based methods are sensitive to physical changes. However, this issue is not the key focus of the current research. Our core appeal is to explore how to improve robustness against noise interference and outliers.

(5) The experimental results presented in Table IV indicate that in algorithms with iterative characteristics, the average training time of the DLR-2DPCA is close to that of the  $R_1$ -2DPCA and 2DPCA-2- $L_p$ , mainly due to the solving algorithm designed for its method structure is more convenient. Our method is slightly faster than GC-2DPCA, Angle-2DPCA and Cos-2DPCA, mainly because GC-2DPCA inevitably needs to introduce more intermediate variables to obtain the weight matrix that protects the geometric structure of the method, whereas Angle-2DPCA and Cos-2DPPCA not only consider the relationship between the reconstruction error and variance, but also perform the singular value decomposition in each iteration task. Finally, the efficiency of the proposed method is significantly better than that of 2DPCA-L1, principally because of the complexity of the nongreedy solving algorithm designed for 2DPCA-L1. which directly increases the dual requirements of storage space and running time in the calculation process. Therefore, based on the recognition results of each method in

different application environments, it can be concluded that DLR-2DPCA has the best overall performance.

(6) The convergence speed of *Algorithm 1* directly affects the efficiency of DLR-2DPCA in executing feature extraction tasks. To show the results of each iteration more intuitively, we clearly depict in Fig. 9 the objective function values obtained by DLR-2DPCA during each iteration of executing *Algorithm 1*. The experimental results in Fig. 9 indicate that as the number of iterations gradually increases, the objective function values of DLR-2DPCA based on the three image databases also show a monotonically decreasing trend, and all tend to stabilize within 10 iterations. It effectively demonstrates that *Algorithm 1* can quickly find the local optimal solution for DLR-2DPCA, thereby verifying the outstanding ability of the proposed model from another perspective.

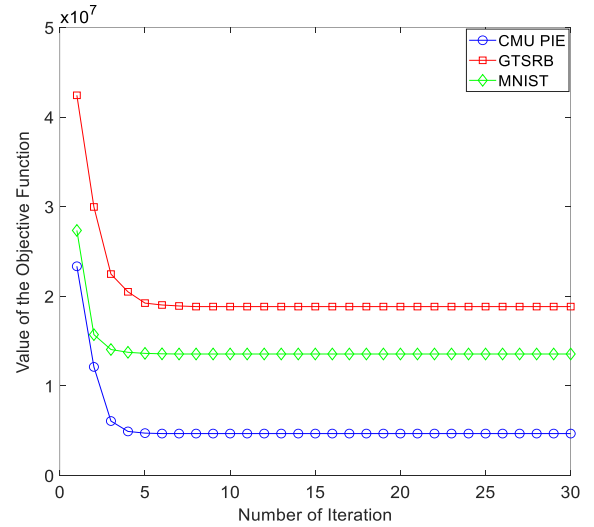


Fig. 9 The convergence curve of the proposed method on the CMU PIE, GTSRB and MNIST databases

## V. CONCLUSION

This paper presents a robust 2DPCA method for image recognition and expression, which is called DLR-2DPCA. Unlike most existing robust 2DPCA methods, DLR-2DPCA considers using cut  $l_2$ ,  $p$ -norm as distance metric in the first layer, which not only effectively increases the flexibility of the measurement mechanism, but also greatly weakens the risk of noise interference being amplified with  $p$ -value changes. Moreover, DLR-2DPCA introduces an adaptive weighted learning strategy in the second layer, which can reasonably assign different weights to the reconstruction errors obtained by each sample in the first layer, further reducing the negative impact of samples with large reconstruction errors on the model. Meanwhile, DLR-2DPCA has valuable properties such as rotational invariance and optimal solution related to covariance matrix, which fully protects the geometric structure of the data. Finally, we propose an iterative algorithm with fast nongreedy properties for the task of obtaining the optimal solution, which ensures a closed form solution during each iteration process and ultimately converges to a stationary point. The experimental results on several image databases convincingly illustrate the outstanding ability of the proposed method.

## ACKNOWLEDGMENTS

The authors thank the anonymous reviewers and associate editor for their valuable suggestions, which significantly improved the manuscript, and sincerely thank Q. Gao of Xidian University for providing comparison method codes.

## REFERENCES

- [1] Z. Li, C. Guo, and G. Han, "Small object detection based on lightweight feature pyramid," *IEEE Trans. Consum. Electron.*, early published, doi: 10.1109/TCE.2024.3412168.
- [2] Y.-H. Kim and J. Lee, "Image feature and noise detection based on statistical hypothesis tests and their applications in noise reduction," *IEEE Trans. Consum. Electron.*, vol. 51, no.4, pp. 1367–1378, Nov. 2005.
- [3] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 1991, pp. 586–591.
- [4] Y. Lu and Q. Tian, "Discriminant subspace analysis: An adaptive approach for image classification," *IEEE Trans. Multimedia*, vol. 11, no. 7, pp. 1289–1300, Nov. 2009.
- [5] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 153–160.
- [6] J. Yang, D. Zhang, X. Yong, and J.-Y. Yang, "Two-dimensional discriminant transform for face recognition," *Pattern Recognit.*, vol. 38, no. 7, pp. 1125–1129, 2005.
- [7] Y. Lu, C. Yuan, Z. Lai, X. Li, D. Zhang, and W. Wong, "Horizontal and vertical nuclear norm-based 2DLDA for image representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 4, pp. 941–955, Apr. 2019.
- [8] Y. Lu, C. Yuan, Z. Lai, X. Li, W. K. Wong, and D. Zhang, "Nuclear norm-based 2DLPP for image classification," *IEEE Trans. Multimedia*, vol. 19, no. 11, pp. 2391–2403, Nov. 2017.
- [9] K. Lee, and H. Park, "Probabilistic learning of similarity measures for tensor PCA," *Pattern Recognition Lett.*, vol. 33, no. 10, pp. 1364–1372, 2012.
- [10] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 18–39, Jan. 2008.
- [11] H. Kamrani, A. Z. Asli, P. P. Markopoulos, M. Langberg, D. A. Pados, and G. N. Karystinos, "Reduced-rank L1-norm principal-component analysis with performance guarantees," *IEEE Trans. Signal Process.*, vol. 69, pp. 240–255, Jan. 2021.
- [12] N. Kwak, "Principal component analysis based on L1-norm maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1672–1680, Sep. 2008.
- [13] F. Nie, H. Huang, C. Ding, D. Luo, and H. Wang, "Robust principal component analysis with non-greedy l-norm maximization," in *Proc. Int. Joint Conf. Artif. Intell.*, 2011, pp. 1433–1438.
- [14] Q. Ye, J. Yang, F. Liu, C. Zhao, N. Ye, and T. Yin, "L1-norm distance linear discriminant analysis based on an effective iterative algorithm," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 1, pp. 114–129, Jan. 2018.
- [15] W. Yu, R. Wang, F. Nie, F. Wang, Q. Yu, and X. Yang, "An improved locality preserving projection with L1-norm minimization for dimensionality reduction," *Neurocomputing*, vol. 316, pp. 322–331, 2018.
- [16] X. Li, Y. Pang, and Y. Yuan, "L1-norm-based 2DPCA," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 4, pp. 1170–1175, Aug. 2010.
- [17] R. Wang, F. Nie, X. Yang, F. Gao, and M. Yao, "Robust 2DPCA with non-greedy L1-norm maximization for image analysis," *IEEE Trans. Cybern.*, vol. 45, no. 5, pp. 1108–1112, May 2015.
- [18] Y. Pang, X. Li, and Y. Yuan, "Robust tensor analysis with L1-norm," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 2, pp. 172–178, Feb. 2010.
- [19] F. Ju, Y. Sun, J. Gao, Y. Hu, and B. Yin, "Image outlier detection and feature extraction via L1-norm-based 2D probabilistic PCA," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4834–4846, Dec. 2015.
- [20] J. Wang, "Generalized 2-D principal component analysis by Lp-norm for image analysis," *IEEE Trans. Cybern.*, vol. 46, no. 3, pp. 792–803, Mar. 2016.
- [21] J. Mi, Y. Zhang, Y. Li, and Y. Shu, "Generalized two-dimensional PCA based on 2,p-norm minimization," *Int J. Mach Learn Cyb.*, vol. 11, no. 7, pp. 1–18, 2020.
- [22] Y. Xue and L. Zhang, "Adaptive pair-weight vector projection classifier for semi-supervised classification," *IEEE Trans. Consum. Electron.*, vol. 70, no.1, pp. 3269–3278, Feb. 2024.
- [23] T. Li, M. Li, Q. Gao, and D. Xie, "F-norm distance metric based robust 2DPCA and face recognition," *Neural Netw.*, vol. 94, pp. 204–211, 2017.
- [24] Q. Gao, S. Xu, F. Chen, C. Ding, X. Gao, and Y. Li, "R1-2-DPCA and face recognition," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1212–1223, Apr. 2019.
- [25] Q. Wang, Q. Gao, X. Gao, and F. Nie, "Optimal mean two-dimensional principal component analysis with F-norm minimization," *Pattern Recognit.*, vol. 68, pp. 286–294, 2017.
- [26] G. Zhou, G. Xu, J. Hao, S. Chen, J. Xu, and X. Zheng, "Generalized centered 2-D principal component analysis," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1666–1677, Mar. 2021.
- [27] Q. Gao, L. Ma, Y. Liu, X. Gao, and F. Nie, "Angle 2DPCA: A new formulation for 2DPCA," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1672–1678, May. 2018.
- [28] X. Wang, L. Shi, J. Liu, and M. Zhang, "Cosine 2DPCA with weighted projection maximization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 12, pp. 9643–9656, Dec. 2023.
- [29] H. Zhang, H. Bi, X. Wang, and P. Zhang, "A joint-norm distance metric 2DPCA for robust dimensionality reduction," *Inf. Sci.*, vol. 640, 2023.
- [30] F. Nie, D. Wu, R. Wang, and X. Li, "Truncated robust principal component analysis with a general optimization framework," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 2, pp. 1081–1097, Feb. 2022.
- [31] Q. Wang, Q. Gao, X. Gao, and F. Nie, "L2p-norm based PCA for image recognition," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1336–1346, Mar. 2018.
- [32] P. Bi and X. Du, "Arbitrary triangle structure adaptive mean PCA and image recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 2, pp. 754–769, Feb. 2024.
- [33] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Washington, DC, USA, 2002, pp. 46–51.
- [34] W. Min, R. Liu, D. He, Q. Han, Q. Wei, and Q. Wang, "Traffic sign recognition based on semantic scene understanding and structural traffic sign location," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 15794–15807, Sep. 2022.
- [35] A. Agarwal, G. Goswami, M. Vatsa, R. Singh, and N. K. Ratha, "DAMAD: Database, attack, and model agnostic adversarial perturbation detector," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 3277–3289, Aug. 2022.
- [36] T. T. Wong and P. Y. Yeh, "Reliable accuracy estimates from k-fold cross validation," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 8, pp. 1586–1594, Aug. 2020.