Inductive robust projective Non-negative matrix factorization with adaptive graph embedding*

Ying Zhao[1[0009−0002−9913−2355]], Rigui Zhou[1[0000−0002−8894−8108]], Zizhu Fan[2[0000−0001−5354−4827]], and Minghua Wan[1,**[0000−0001−5725−6240]]

[1] Shanghai Maritime University, Shanghai, China
[2] Shanghai University of Electric Power, Shanghai, China
yingzhao0721@163.com, rgzhou@shmtu.edu.cn, zzfan3@163.com, wmh36@sina.com

Abstract. Although low-rank matrices have gained widespread popularity in recent years, traditional low-rank derivation algorithms, being conductive methods, cannot handle new samples. Inductive robust principal component analysis (IRPCA) can learn low-rank projections but cannot perform feature extraction. Therefore, this paper introduces a innovative robust feature extraction algorithm, named inductive robust projective Non-negative matrix factorization with adaptive graph embedding (IRPNMF-AGE). The algorithm is built in several steps. To begin with, we decompose the projection matrix into two components. The first component focuses on capturing the low-rank structure, while the other uses a projection Non-negative matrix factorization algorithm to train the low-rank recovery matrix and obtain the base matrix. Subsequently, the low-rank recovery matrix is employed to modify the graph structure. Ultimately, to create a projection matrix that preserves the low-rank structure, we merge the trained low-rank matrix with the base matrix. Experiments demonstrate that the IRPNMF-AGE algorithm achieves superior low-rank recovery performance on damaged test samples while enhancing robustness and image classification capabilities.

Keywords: Inductive robust PCA · Low-rank representation · Projective NMF · Feature extraction · Image classification.

1 Introduction

In recent years, advanced feature extraction algorithms have been continuously developed. However, all these algorithms have relied on clean datasets. When datasets are affected by noise and outliers, maintaining superiority becomes a

* Supported by the National Key R&D Plan under Grant No. 2021YFF0601200 and 2021YFF0601204, the Key Areas Special Project for Regular Higher Education Institutions in Guangdong Province under Grant 2025ZDZX1023, in part by Guangdong Key Discipline Research Capacity Building Project under Grant 2024ZDJS056, in part by Ganzhou City Unveiling and Leading Project under Grant 2023ULGX0004,in part by Guangzhou Municipal Education Bureau under Grant 2024312000.
** Corresponding author. Email: wmh36@sina.com

challenge for most algorithms. Due to their robustness against noise, a series of low-rank methods have gained significant attention recently [1]. Low-rank projection learning via graph embedding [2] balances considerations of local neighborhoods and robustness. Low-rank preserving projection via graph regularized reconstruction (LRPP-GRR) [3] introduces reconstruction error into the graph structure. This reduces model complexity. Low-rank and sparsity preserving embedding [4] aims to learn both the graph and projection together in a unified framework. Robust auto-weighted projective low-rank and sparse [5] integrates adaptive weighted graphs. It combines this with joint low-rank/sparse representation into a single, cohesive model. In low-rank local embedding representation [6], the local and global manifold structures of the dataset exhibit correlation. Low-rank adaptive graph embedding [7] concurrently performs engages in subspace learning and adaptive probability neighborhood graph embedding. Latent low-rank representation with weighted distance penalty (LLRRWD) [8] introduces weighted distance penalties into the LatLRR model. This enhances the discriminative power of affinity matrix learning. The dictionary learning-based unsupervised feature selection method [9] applies dictionary learning concepts to low-rank representation. It preserves sample similarity through spectral analysis.

The non-negative property of NMF imparts practical and interpretable meaning to it. Considering that NMF has the capability to employ the coefficient matrix in order to influence the base matrix, this aids in the reconstruction of the initial data, it yields a part-based data representation. Studies [10, 11] have indicated that this part-based representation aligns with the perceptual information processing in the human brain. NMF has found successful applications in various domains, including text analysis [12], image processing [13], and bioinformatics [14]. Recently, a range of robust NMF methods have been developed to bolster the generalization capability of NMF. Low-rank nonnegative factorization (LRNF) [15] incorporates RPCA into the foundation of NMF to achieve low-rank recovery of corrupted samples. Building upon this, exponential graph regularized Non-negative low-rank factorization [16] introduces matrix exponentiation and sparse constraints to further enhance its robustness. Low-rank NMF on the Stiefel manifold (LNMFS) [17] leverages the Stiefel manifold to impose orthogonal constraints and transform low-rank structures into the Frobenius norm. Many algorithms often consider either global or local geometric structures. However, low-rank sparse NMF (NMF-LRSR) [18] combines both aspects. Double manifold regularized NMF (DMR-NMF) [19] does not employ RPCA for low-rank recovery but instead employs LRR to learn adaptive global affinity matrices. Furthermore, there are robust methods that incorporate adaptive graph structures [20]. Although low-rank enhancement methods based on RPCA and LRR may achieve more robust feature extraction than the original approaches, they fail to effectively leverage low-rank properties.

To address the aforementioned issues, this paper proposes the inductive robust projective Non-negative matrix factorization with adaptive graph embedding (IRPNMF-AGE). By decomposing the projection matrix into a basis matrix and a low-rank recovery matrix, and optimizing the nearest-neighbor graph

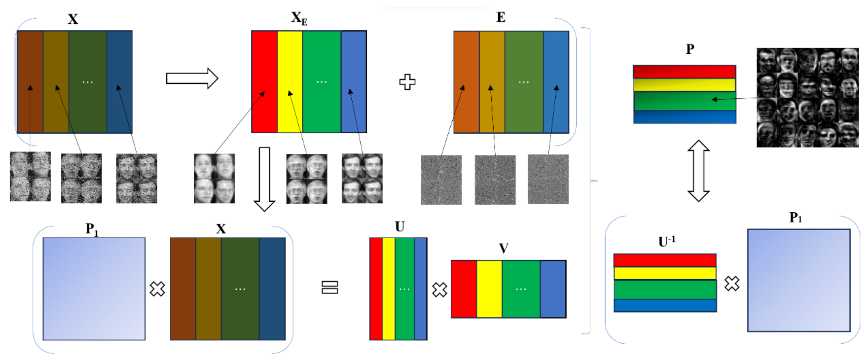


Fig. 1: Illustration of IRPNMF-AGE.

structure based on the low-rank recovery image, it obtains a feature extraction projection matrix capable of both processing new samples and restoring low-rank features of damaged images. This algorithm integrates non-negative matrix projection, low-rank structure, and nearest neighbor graph structure into a single framework. By combining the learned optimal base matrix with the low-rank matrix, we achieve a robust projection algorithm that can be used for training new corrupted samples. As shown in Fig.1, the obtained low-rank recovery image can be broken down into the low-rank projection matrix and original noise-corrupted matrix.

This paper makes contributions in the following areas:

1) Proposing a novel and scalable low-rank feature extraction model that effectively addresses the issues of low-rank algorithms not being able to handle new samples and perform feature extraction.

2) Pioneering the utilization of graph structures, optimized using low-rank matrices during the training process, to effectively rectify and improve the accuracy of the introduced manifold structure in the algorithm.

3) This adaptive graph correction mechanism fundamentally mitigates the sensitivity of traditional neighborhood graphs to noise and outliers. By utilizing low-rank matrices with denoising properties to optimize the graph structure, it ensures that the graph accurately reflects the intrinsic manifold of the data.

The structure of this paper is as follows: Section 2 presents the motivation, objective function, derivations, and algorithm complexity analysis for IRPNMF-AGE. Section 3 provides a comprehensive account of the experimental results. Finally, we draw the paper to its conclusion in Section 4.

2 IRPNMF-AGE

2.1 Motivation

Based on the introduction to related work, we already know that IRPCA can effectively preserve low-rank structures for projecting new samples. Our algorithm model is also based on this premise. Given a dataset $X \in R^{m\times n}$, we attempt to find a projection matrix $P \in R^{r\times m}$ such that:

$$Y = PX. \tag{1}$$

where $Y \in R^{r\times n}$ is the dimension-reduced data, and $r \ll m$.

Clearly, techniques like extracting noise in the dimension-reduced matrix, as in LRE, are too abstract and do not allow for the direct observation of its low-rank structure. Therefore, we decompose P into P_1 and P_2:

$$P = P_2 P_1. \tag{2}$$

where $P_1 \in R^{m\times m}$ is used to preserve low-rank structures, and $P_2 \in R^{r\times m}$ is used to project $P_1 X$ for dimension reduction. By combining these two methods, we obtain:

$$\begin{aligned} &\min_{P_1,E,U,X_E} \|P_1\|_* + \alpha\|E\|_1 + \beta\|X_E - UU^T X_E\|_F^2 \\ &\text{s.t. } X = X_E + E, X_E = P_1 X \end{aligned}. \tag{3}$$

where E serves as the noise matrix, X_E represents the low-rank recovered image, and $\alpha, \beta > 0$ is utilized as a balancing parameter.

It is worth noting that (3) contains fourth-order terms, which are theoretically challenging to analyze for convergence. To eliminate these higher-order terms, we introduce auxiliary variables to reformulate (3) as follows:

$$\begin{aligned} &\min_{P_1,E,U,V,X_E} \|P_1\|_* + \alpha\|E\|_1 + \beta\|X_E - UV\|_F^2 + \gamma\|V - U^T X_E\|_F^2 \\ &\text{s.t. } X = X_E + E, X_E = P_1 X \end{aligned}. \tag{4}$$

the parameter γ is introduced to mitigate the impact of orthogonality on the algorithm's convergence.

In real-world applications, data such as facial images is frequently obtained from nonlinear manifolds [21]. Using the low-rank recovery matrix during the iterative process to recover the disrupted graph structure can further boost the algorithm's recognition rate, as shown in (5):

$$Z_{ij} = \begin{cases} \exp\left(-\dfrac{\|X_{Ei} - X_{Ej}\|^2}{2\sigma^2}\right), & X_{Ej} \in N_{K_c}^{+}(X_{Ei}) \\ 0, & \text{otherwise} \end{cases}. \tag{5}$$

where $W_{ij} = \frac{1}{2}(Z_{ij} + Z_{ji})$, $N_{K_C}^{+}(X_{E_i})$ represents the index set for the K_C nearest neighbors of X_{E_i}.

So, in combination with (5), we formulate the objective function of the IRPNMF-AGE algorithm as:

$$\begin{aligned} &\min \|P_1\|_* + \alpha\|E\|_1 + \beta\|X_E - UV\|_F^2 + \gamma\|V - U^T X_E\|_F^2 + \frac{\omega}{2}\sum_{i,j}\|v_i - v_j\|^2 W_{ij} \\ &s.t. X = X_E + E, X_E = P_1 X \end{aligned}. \tag{6}$$

Thus, we have completed the motivation and principle introduction of the proposed IRPNMF-AGE algorithm model.

2.2 Objective Function of IRPNMF-AGE

Building on the model motivation in Section 2.1, we first simplify the graph regularization term to derive the final unified objective function, then reformulate it for efficient optimization via the ADMM algorithm. To simplify the derivation and solution, it is easy to obtain:

$$R = \frac{1}{2}\sum_{i,j}\|v_i - v_j\|^2 W_{ij} = \mathrm{Tr}(V^T DV) - \mathrm{Tr}(V^T WV) = \mathrm{Tr}(V^T LV). \quad (7)$$

where D is a diagonal matrix and $D_{ii} = \sum_i W_{ii}$, $L = D - W$.

So, we obtain the objective function:

$$\min \|P_1\|_* + \alpha\|E\|_1 + \beta\|X_E - UV\|_F^2 + \gamma\|V - U^T X_E\|_F^2 + \omega\mathrm{Tr}(VLV^T). \quad (8)$$

According to Theorem 1 in [22], the objective function for IRPNMF-AGE can alternatively be expressed as:

$$\min \|J\|_* + \alpha\|E\|_1 + \beta\|X_E - UV\|_F^2 + \gamma\|V - U^T X_E\|_F^2 + \omega\mathrm{Tr}(VLV^T). \quad (9)$$

where $P_1 = LQ^T$ and Q are obtained from the orthogonally derived columns of X, and $A = Q^T X$.

To make function (8) separable, we introduce an auxiliary matrix J and define $P_1 = J$, which transforms function (8) into the following representation:

$$\min \|J\|_* + \alpha\|E\|_1 + \beta\left\|X_E - UV\right\|_F^2 + \gamma\left\|V - U^T X_E\right\|_F^2 + \omega\mathrm{Tr}\left(VLV^T\right). \quad (10)$$

The augmented Lagrangian function of (10) is:

$$\begin{aligned} L &= \|J\|_* + \alpha\|E\|_1 + \beta\|X_E - UV\|_F^2 + \gamma\|V - U^T X_E\|_F^2 \\ &+ \langle Y_1, X - X_E - E\rangle + \langle Y_2, P_1 - J\rangle + \langle Y_3, X_E - P_1 X\rangle + \omega\mathrm{Tr}(VLV^T). \\ &+ \frac{\mu}{2}(\|X - X_E - E\|_F^2 + \|P_1 - J\|_F^2 + \|X_E - P_1 X\|_F^2) \end{aligned} \quad (11)$$

where $Y_1, Y_3 \in R^{m\times n}$ and $Y_2 \in R^{m\times m}$ are Lagrange multipliers, $\mu > 0$ is a parameter, $< A, B >$ represents the trace of the matrix $A^T B$.

2.3 Iterative Solution of IRPNMF-AGE

(1) Update J

By fixing variables other than J and updating J, we obtain:

$$J = \arg\min \frac{1}{\mu}\|J\|_* + \frac{1}{2}\left\|J - P_1 - \frac{Y_2}{\mu}\right\|_F^2. \quad (12)$$

This optimization problem can be addressed through the application of the Singular Value Thresholding operator:

$$J = \Gamma_{\frac{1}{\mu}}\left(P_1 + \frac{Y_2}{\mu}\right). \quad (13)$$

where $\Gamma_\tau(T) = Udiag(max(0, \sigma - \tau))V^T$, σ represents the singular value of T.

(2) Update P_1

By fixing variables other than P_1 and updating P_1, we obtain:

$$L(P_1) = \langle Y_2, P_1 - J\rangle + \langle Y_3, X_E - P_1 X\rangle + \frac{\mu}{2}\left(\|P_1 - J\|_F^2 + \|X_E - P_1 X\|_2^2\right). \quad (14)$$

Take the partial derivative of (14), then obtain the following equation:

$$\frac{\partial L}{\partial P_1} = Y_2 - Y_3 X^T + \mu\left(P_1(XX^T + I) - X_E X^T - J\right) = 0. \quad (15)$$

So, we obtain the update iteration formula for P_1:

$$P_1 = \left(X_E X^T + J + \frac{Y_3 X^T - Y_2}{\mu}\right)\left(XX^T + I\right)^{-1}. \quad (16)$$

(3) Update E

By keeping variables other than E fixed and updating E, we obtain:

$$E = \arg\min \frac{\alpha}{\mu}\|E\|_1 + \frac{1}{2}\left\|E - X + X_E - \frac{Y_1}{\mu}\right\|_F^2. \quad (17)$$

This optimization problem can be resolved using the soft-thresholding operator:

$$E = \vartheta_{\frac{\alpha}{\mu}}\left(X - X_E + \frac{Y_1}{\mu}\right). \quad (18)$$

where $\vartheta_\tau(t) = sign(t) \cdot max(|t| - \tau, 0)$.

(4) Update X_E

By fixing variables other than X_E and updating X_E, we obtain:

$$\begin{aligned} L(X_E) &= \beta\|X_E - UV\|_F^2 + \langle Y_1, (X - X_E - E)\rangle + \langle Y_3, (X_E - P_1 X)\rangle \\ &+ \gamma\|V - U^T X_E\|_F^2 + \frac{\mu}{2}\left(\|X - X_E - E\|_F^2 + \|X_E - P_1 X\|_F^2\right) \end{aligned}. \quad (19)$$

Take the partial derivative of (19), then obtain the following equation:

$$\frac{\partial L}{\partial X_E} = 2\beta(X_E - UV) + 2\gamma(UU^T X_E - UV) - Y_1 + Y_3 + \mu(2X_E - X - P_1 X + E) = 0. \quad (20)$$

So, we obtain the update iteration formula for X_E:

$$X_E = \left(2\frac{\beta}{\mu}I + 2I + 2\frac{\gamma}{\mu}UU^T\right)^{-1}\left(\frac{Y_1 - Y_3}{\mu} + 2\frac{\beta + \gamma}{\mu}UV + (X - E + P_1 X)\right). \quad (21)$$

(5) Update U and V

Updating U and V can be achieved by fixing variables other than U and V, and we obtain:

$$\begin{aligned} L(U, V) &= \mathrm{Tr}\left(-2X_E^T UV + V^T U^T UV\right) \\ &+ \frac{\gamma}{\beta}\mathrm{Tr}\left(V^T V - 2X_E^T UV + X_E^T UU^T X_E\right) + \frac{\omega}{\beta}\mathrm{Tr}\left(V(D - W)V^T\right) \end{aligned}. \quad (22)$$

Let ψ_{ik} and ϕ_{ik} be the Lagrange multipliers for constraints $u_{ik} \geq 0$ and $v_{ik} \geq 0$, let $\psi = [\psi_{ik}]$, $\Phi = [\phi_{ik}]$:

$$\begin{aligned} L(U,V) &= \operatorname{Tr}\left(-2X_E^T UV + V^T U^T UV\right) + \frac{\omega}{\beta}\operatorname{Tr}\left(V(D-W)V^T\right) \\ &+ \frac{\gamma}{\beta}\operatorname{Tr}\left(V^T V - 2X_E^T UV + X_E^T UU^T X_E\right) - \operatorname{Tr}\left(\Psi U^T\right) - \operatorname{Tr}\left(\Phi V^T\right) \end{aligned}. \tag{23}$$

Using the KKT condition $\psi_{ik}u_{ik} = 0$ and $\phi_{ik}v_{ik} = 0$, for u_{ik} and v_{ik}, we get the equations below:

$$\frac{\partial L}{\partial U} = 2UVV^T - 2X_E V^T + \frac{2\gamma}{\beta}(X_E X_E^T U - X_E V^T) = 0. \tag{24}$$

$$\frac{\partial L}{\partial V} = 2U^T UV - 2U^T X_E + \frac{2\gamma}{\beta}(V - U^T X_E) + \frac{2\omega}{\beta}V(D-W) = 0. \tag{25}$$

So, we have the update iterations for U and V:

$$U_{ij} \leftarrow U_{ij}\left(\left(1+\frac{\gamma}{\beta}\right)X_E V^T\right)_{ij} \Big/ \left(UVV^T + \frac{\gamma}{\beta}X_E X_E^T U\right)_{ij}. \tag{26}$$

$$V_{ij} \leftarrow V_{ij}\left(\left(1+\frac{\gamma}{\beta}\right)U^T X_E + \frac{\omega}{\beta}VW\right)_{ij} \Big/ \left(U^T UV + \frac{\gamma}{\beta}V + \frac{\omega}{\beta}VD\right)_{ij}. \tag{27}$$

The comprehensive optimization framework is detailed in Algorithm 1. Since the optimization procedure for objective function (11) follows a similar iterative scheme, we omit the redundant derivation for brevity.

Algorithm 1 Solving IRPNMF-AGE by ADMM.

Input($\mathbf{X}, i, \alpha, \beta, \gamma, \omega, \mu$)
  $Y_1 = Y_3 = 0, Y_2 = 0, X_E = X, E = 0$
  calculate W by (5)
  $U = abs(rand(m,r)), V = abs(rand(r,n))$
repeat
    Update J by (13);
    Update P_1 by (16);
    Update E by (18);
    Update X_E by (21);
    Update W by (5);
    Update U by (26);
    Update V by (27);
    $i = i + 1$
    $Y_1 = Y_1 + \mu(X - X_E - E)$
    $Y_2 = Y_2 + \mu(P_1 - J)$
    $Y_3 = Y_3 + \mu(X_E - P_1 X)$
    $\mu = \min(\rho\mu, \mu_{\max})$
until the maximum number of iterations is met
return $P = U^T P_1 U$

2.4 Algorithm Complexity Analysis

We consider this in the context of large-scale datasets, where $m < n$. The reduced dimension after dimensionality reduction is denoted as $r \ll min(m, n)$, and the number of neighbors is c. The computational scale of the associated matrices in a single iteration and the algorithm complexity are presented in the Table 1. Due to $m < n$, assuming the algorithm converges after k iterations, the time complexity of this algorithm is $O(km^2n)$.

Table 1: Computational Complexity Analysis

Matrix	Computational scale	Algorithm complexity
J	$3m^3 + m^2$	$O(m^3)$
P_1	$3m^2n + 2m^3 + m^2$	$O(m^2n)$
E	mn	$O(mn)$
X_E	$2m^2n + mnr + m^2r + 2mn + m^2 + 2m$	$O(m^2n)$
W	mnc	$O(mnc)$
U	$m^2n + mnr + 2m^2r + nr^2 + 4mr$	$O(m^2n)$
V	$mnr + mr^2 + nr^2 + nrc + 7nr$	$O(mnr)$

3 Experiments Analysis

3.1 Databases

The Extended YALE-B dataset comprises around 2,432 images distributed among 38 individuals, each having 64 images captured under various lighting conditions.

The CMU PIE database is an extensive collection, housing 11,560 facial images belonging to 68 different individuals. These images showcase diverse variations in facial expressions, head poses, and lighting conditions.

The AR dataset consists of 26 facial images for each of the 120 individuals, encompassing variations in lighting conditions, facial occlusions, and expressions.

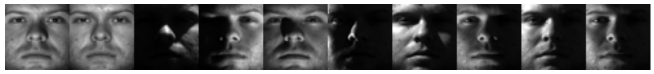
Meanwhile, the COIL-20 dataset is composed of 1,440 images, each representing one of 20 distinct objects. These objects are depicted from various angles within a 5-degree range, and each object is depicted in 70 images.

The MNIST dataset contains 10,000 images illustrating handwritten digits, with each digit contributed by 250 different individuals.

For a visual representation of these datasets, please refer to Fig.2.

3.2 Verification Experiments of Classification Accuracy

To assess the classification performance of IRPNMF-AGE, we conducted classification experiments on faces, objects, and handwritten digits. A nearest neighbor classifier was employed and compared against NMF, PNMF, LRNF, IRPCA, LNMFS, LRPP-GRR, and DMR-NMF. Specific settings are detailed in Table 2. Each method underwent 10 times, and the results were averaged. The experimental outcomes are presented in Table 3, demonstrating that IRPNMF-AGE achieved the highest recognition rate compared to other low-rank derivative algorithms.

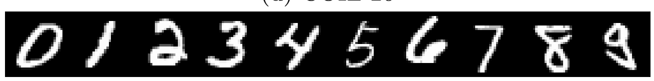


(a) Extended YALE-B

(b) CMU PIE

(c) AR

(d) COIL-20



(e) MNIST

Fig. 2: Five databases have a few images each.

Table 2: Experimental configurations.

Dataset	**Number of training samples**	**Total number of samples**
Extended YALE-B	30*(20, 30, 40)	30*60
CMU PIE	68*(60, 70, 80)	68*170
AR	120*(8, 10, 12)	120*26
COIL-20	20*(10, 20, 30)	20*70
MNIST	10*(100, 200, 300)	10*1000

Table 3: Mean classification accuracy (%) and standard deviation (%) for various models across five datasets.

datasets	**N**	**NMF**	**PNMF**	**LRNF**	**IRPCA**	**LNMFS**	**LRPP-GRR**	**DMR-NMF**	**IRPNMF-AGE**
Extended YALE-B	20	79.93±1.48	85.08±0.83	82.18±1.18	61.58±0.93	85.82±1.05	86.00±1.16	82.67±1.02	**87.40±0.88**
	30	82.27±0.68	88.69±1.27	87.09±0.82	69.11±2.32	89.60±0.41	88.67±1.23	86.33±0.98	**91.31±0.65**
	40	83.03±0.61	91.10±0.94	89.70±0.81	75.97±1.66	91.57±0.48	89.67±1.62	89.00±0.66	**91.90±0.71**
CMU PIE	60	82.75±0.46	81.95±0.54	85.55±0.44	78.65±0.50	87.20±0.80	85.11±0.72	89.60±0.55	**90.01±0.41**
	70	85.10±0.32	83.84±0.36	88.06±0.64	82.29±0.25	89.09±0.39	87.09±0.45	90.99±0.28	**91.63±0.18**
	80	86.80±0.50	88.19±0.49	89.52±0.47	85.22±0.79	90.42±0.33	87.29±0.91	92.27±0.48	**92.65±0.35**
AR	8	78.94±1.11	67.31±1.13	87.42±1.10	70.00±1.03	80.17±1.16	86.94±1.34	84.68±1.15	**87.71±1.07**
	10	82.68±0.91	69.95±0.95	91.02±0.76	75.23±1.06	83.60±1.20	90.73±0.96	88.70±0.86	**91.14±0.53**
	12	86.10±0.61	74.17±1.02	92.88±1.31	78.20±1.31	86.04±1.13	92.02±1.45	90.98±0.99	**93.58±0.72**
COIL-20	10	85.38±1.13	91.67±0.70	90.45±1.16	90.37±1.36	91.42±1.96	90.33±1.78	90.38±1.55	**92.32±1.29**
	20	90.56±1.11	96.44±1.45	95.74±0.90	95.56±1.02	96.02±1.11	96.01±0.95	96.32±0.97	**96.68±0.52**
	30	91.00±2.04	98.28±0.44	97.67±0.65	97.85±0.43	98.33±0.66	97.75±0.61	98.33±0.66	**98.90±0.48**
MNIST	100	76.64±1.07	86.70±0.61	84.40±0.68	87.12±0.35	86.70±0.74	80.26±1.21	86.25±0.54	**87.34±0.25**
	200	79.05±0.57	87.80±1.24	88.72±0.60	89.38±0.56	89.01±0.95	82.62±1.06	82.62±0.51	**89.58±0.28**
	300	80.30±1.55	90.10±0.92	90.28±1.02	90.49±0.25	90.23±1.15	83.29±1.11	90.13±0.45	**90.91±0.49**

3.3 Face Recognition with Synthetic Corruptions

Within the section, we performed experiments on the CMU PIE and Extended YALE-B datasets to evaluate the robustness of the IRPNMF-AGE algorithm. To validate its robustness in the context of face databases, we conducted experiments on data subjected to varying degrees of "salt & pepper" noise. Specifically, for the "salt & pepper" noise experiments on these two databases, we introduced noise with densities of 0.05, 0.1, and 0.15, respectively.

Fig.3 presents examples of corrupted images from the CMU PIE, Extended YALE-B, AR, COIL-20 and MNIST databases. Fig.4 illustrates the experimental results for these databases, including varying densities of "salt & pepper" noise. In each row, you can observe the classification accuracy achieved by different methods as the number of training samples varies. As depicted in Fig.4, it is evident that IRPNMF-AGE achieves the highest classification accuracy. This suggests that the introduction of IRPCA's low-rank projection method can better handle new sample data, effectively enhancing the robustness of IRPNMF-AGE.

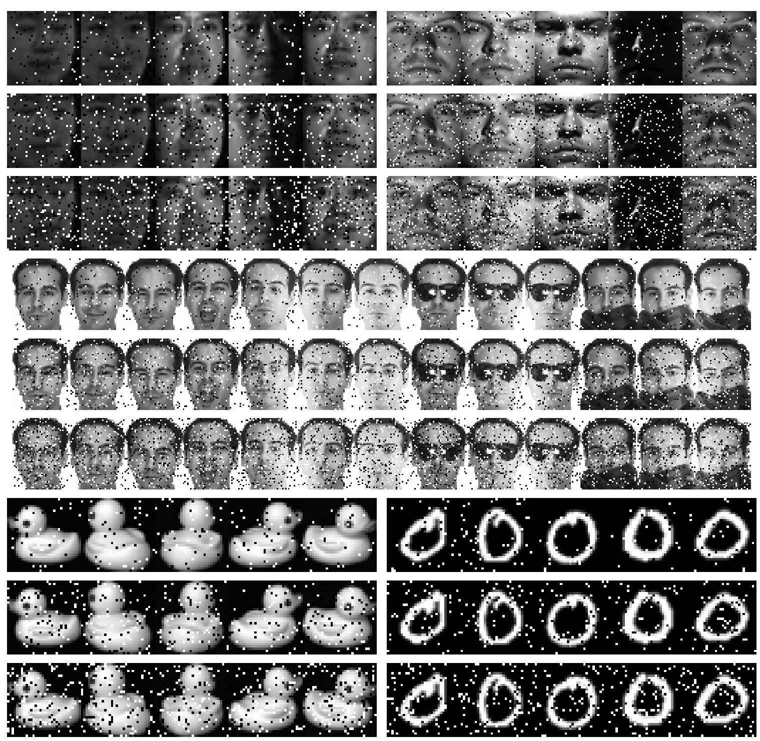
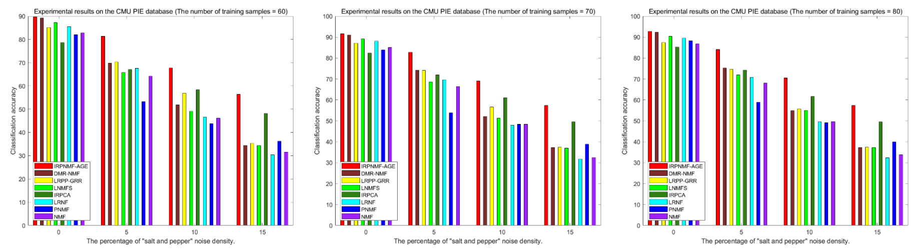
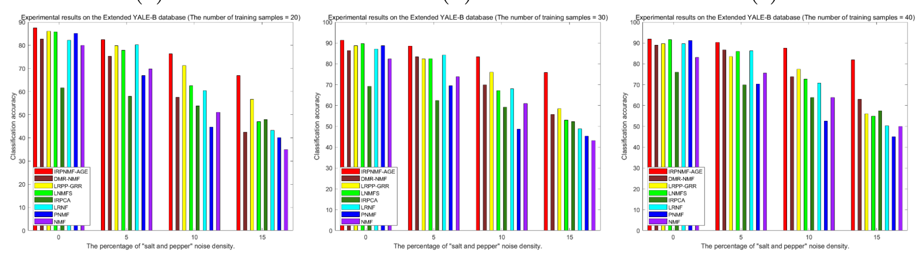


Fig. 3: Images within the CMU PIE, Extended YALE-B, AR, COIL-20 and MNIST databases affected by varying degrees of "salt & pepper" noise.

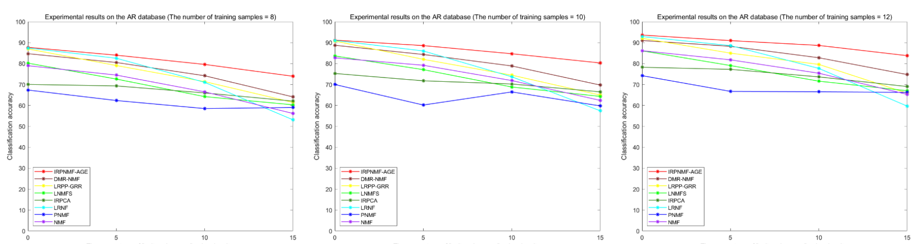


(a) (b) (c)



(d) (e) (f)

Fig. 4: The classification accuracies of eight algorithms on CMU PIE and Extended YALE-B databases with diverse "salt & pepper" noise.



(a) training number = 8 (b) training number = 10 (c) training number = 12

Fig. 5: Classification accuracies of eight algorithms on the AR database with "salt & pepper" noise.

3.4 Face Recognition with Real Occlusions and Noise breaking

Within the section, we conducted experiments on the AR database. As shown in Fig.3, "salt & pepper" noise was introduced to the images with densities of 0.05, 0.1, and 0.15. Fig.5 presents the experimental findings from the AR database. In this visualization, the x-axis represents the density of "salt & pepper" noise.

The results presented in Fig.5 clearly indicate that IRPNMF-AGE achieves the highest classification accuracy among all algorithms, regardless of noise presence, and its recognition rate is minimally affected by noise interference. This indicates that the IRPNMF-AGE algorithm demonstrates strong robustness in the task of facial recognition.

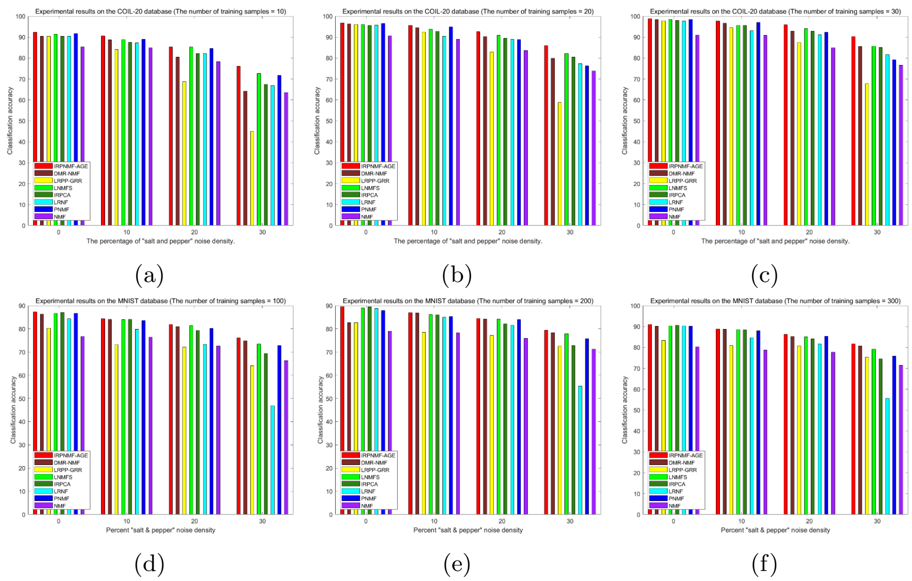


Fig. 6: The classification accuracies of eight algorithms on COIL-20 and MNIST databases with diverse “salt & pepper” noise.

3.5 Object Classification and Recognition of Handwritten Digits

In this subsection, we validate the robustness of IRPNMF-AGE in object and handwritten digit classification using the COIL-20 and MNIST databases. Similar to the method in Section 3.3, we introduced various levels of “salt & pepper” noise into the images. As shown in Fig.3, we added “salt & pepper” noise with densities spanning from 0.1 to 0.3. Fig.6 illustrates the experimental results for the COIL-20 and MNIST databases. It is evident that IRPNMF-AGE continues to achieve the highest recognition rates, indicating its superior robustness compared to other methods.

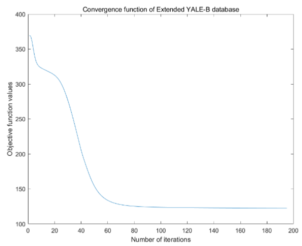
3.6 Convergence Analysis

Within the subsection, we further validate the convergence and investigate the convergence speed of IRPNMF-AGE through experiments on five datasets. Convergence curves are illustrated in Fig.7. In each graph, the x-axis signifies the number of iterations, with the y-axis denoting the objective function’s value.

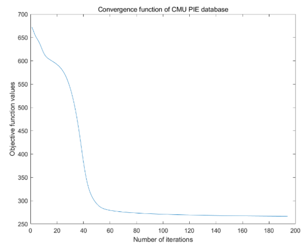
As shown in Fig.7, IRPNMF-AGE typically converges within 80 iterations across the five datasets examined in this study. This clearly demonstrates the effectiveness and rapid convergence of the IRPNMF-AGE.

3.7 Discussion on Parameter Sensitivity

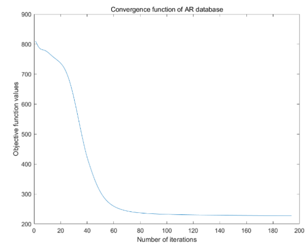
Within the section, we delve into the sensitivity of the model’s parameters. We conducted experiments with 10% “salt & pepper” noise on COIL-20. Given the



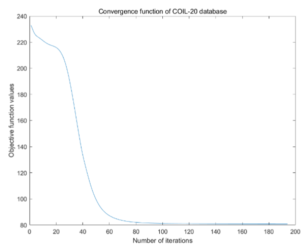
(a) YALE-B



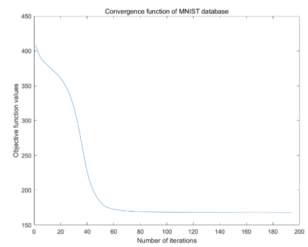
(b) CMU PIE



(c) AR

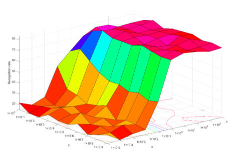


(d) COIL-20

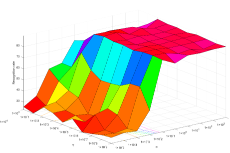


(e) MNIST

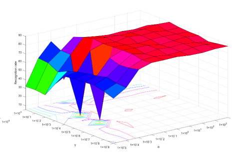
Fig. 7: The convergence function of IRPNMF-AGE in the five databases.



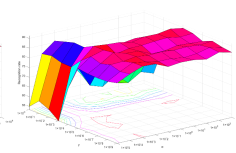
(a) $\beta = 1 \times 10^{0}$



(b) $\beta = 1 \times 10^{-1}$



(c) $\beta = 1 \times 10^{-2}$



(d) $\beta = 1 \times 10^{-3}$

Fig. 8: The impact of different combinations of values for the three parameters on the classification performance of IRPNMF-AGE.

relatively high number of parameters in our method, to simplify the experiments, we first assigned a very low value to the graph regularization term to eliminate its effect. Subsequently, we determined the values of α, β, and γ. Specifically, based on empirical knowledge, we selected α within a reasonable range $[10^{-5}, 10^{4}]$; β within another reasonable range $[10^{-3}, 10^{0}]$, and γ within yet another reasonable range $[10^{-9}, 10^{0}]$. Fig.8 presents a visual representation of how various combinations of values influence the classification performance of IRPNMF-AGE. It can be observed that except for extreme values, most of the reasonable value ranges result in classification accuracy fluctuating within a relatively high range.

4 Conclusions

This paper presents a novel approach, IRPNMF-AGE, which combines scalable low-rank learning with feature extraction. It effectively acquires a low-rank projection matrix from corrupted samples for recovery and performs efficient feature extraction from damaged new samples. Results from extensive experiments on five public datasets confirm the feasibility of the low-rank projection feature extraction method. IRPNMF-AGE exhibits superior robustness compared to other advanced low-rank algorithms. However, as the experiments unfolded, the limitations of IRPNMF-AGE became apparent. For instance, in experiments without noise interference, the classification accuracy of IRPNMF-AGE does not significantly outperform other methods, largely due to the limitations of the PNMF method. In addition, we plan to explore the integration of our low-rank learning mechanism into deep learning frameworks to further enhance representation ability and generalization performance in future research. In future work, we aim to overcome these limitations and further explore methods for extracting more efficient features.

References

1. G. Liu, Q. Liu, P. Li, Blessing of dimensionality: Recovering mixture data via dictionary pursuit, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (1) (2017) 47–60. doi:10.1109/TPAMI.2016.2539946.
2. Y. Liang, L. You, X. Lu, Z. He, H. Wang, Low-rank projection learning via graph embedding, Neurocomputing 348 (2019) 97–106, advances in Data Representation and Learning for Pattern Analysis. doi:https://doi.org/10.1016/j.neucom.2018.05.122.
URL https://www.sciencedirect.com/science/article/pii/S092523121831292X
3. J. Wen, N. Han, X. Fang, L. Fei, K. Yan, S. Zhan, Low-rank preserving projection via graph regularized reconstruction, IEEE Transactions on Cybernetics 49 (4) (2019) 1279–1291. doi:10.1109/TCYB.2018.2799862.
4. S. Zhan, J. Wu, N. Han, J. Wen, X. Fang, Unsupervised feature extraction by low-rank and sparsity preserving embedding, Neural Networks 109 (2019) 56–66. doi:https://doi.org/10.1016/j.neunet.2018.10.001.
URL https://www.sciencedirect.com/science/article/pii/S0893608018302867
5. L. Wang, B. Wang, Z. Zhang, Q. Ye, L. Fu, G. Liu, M. Wang, Robust auto-weighted projective low-rank and sparse recovery for visual representation, Neural Networks 117 (2019) 201–215. doi:https://doi.org/10.1016/j.neunet.2019.05.007.
URL https://www.sciencedirect.com/science/article/pii/S089360801930139X
6. T. Deng, D. Ye, R. Ma, H. Fujita, L. Xiong, Low-rank local tangent space embedding for subspace clustering, Information Sciences 508 (2020) 1–21. doi:https://doi.org/10.1016/j.ins.2019.08.060.
URL https://www.sciencedirect.com/science/article/pii/S0020025519308096
7. J. Lu, H. Wang, J. Zhou, Y. Chen, Z. Lai, Q. Hu, Low-rank adaptive graph embedding for unsupervised feature extraction, Pattern Recognition 113 (2021) 107758. doi:https://doi.org/10.1016/j.patcog.2020.107758.
URL https://www.sciencedirect.com/science/article/pii/S0031320320305616
8. Z. Fu, Y. Zhao, D. Chang, Y. Wang, J. Wen, Latent low-rank representation with weighted distance penalty for clustering, IEEE Transactions on Cybernetics 53 (11) (2023) 6870–6882. doi:10.1109/TCYB.2022.3166545.
9. M. G. Parsa, H. Zare, M. Ghatee, Low-rank dictionary learning for unsupervised feature selection, Expert Systems with Applications 202 (2022) 117149. doi:https://doi.org/10.1016/j.eswa.2022.117149.
URL https://www.sciencedirect.com/science/article/pii/S0957417422005437

10. Hierarchical structure in perceptual representation, Cognitive Psychology 9 (4) (1977) 441–474. doi:https://doi.org/10.1016/0010-0285(77)90016-0.
URL https://www.sciencedirect.com/science/article/pii/0010028577900160
11. E. Wachsmuth, M. W. Oram, D. I. Perrett, Recognition of objects and their component parts: Responses of single units in the temporal cortex of the macaque, Cerebral Cortex 4 (5) (1994) 509–522. arXiv:https://academic.oup.com/cercor/article-pdf/4/5/509/949457/4-5-509.pdf, doi:10.1093/cercor/4.5.509.
URL https://doi.org/10.1093/cercor/4.5.509
12. M. H. Aghdam, Automatic extractive and generic document summarization based on nmf, Journal of Artificial Intelligence and Soft Computing Research 13 (2022) 37 – 49.
URL https://api.semanticscholar.org/CorpusID:254019087
13. D. Fan, X. Zhang, W. Kang, H. Zhao, Y. Lv, Video watermarking algorithm based on nsct, pseudo 3d-dct and nmf, Sensors 22 (13) (2022).
URL https://www.mdpi.com/1424-8220/22/13/4752
14. X. Li, W. Yu, G. Xu, F. Liu, Msda-nmf: A multilayer complex system model integrating deep autoencoder and nmf, Mathematics 10 (15) (2022).
URL https://www.mdpi.com/2227-7390/10/15/2750
15. Y. Lu, Z. Lai, X. Li, D. Zhang, W. K. Wong, C. Yuan, Learning parts-based and global representation for image classification, IEEE Transactions on Circuits and Systems for Video Technology 28 (12) (2018) 3345–3360. doi:10.1109/TCSVT.2017.2749980.
16. G. Yang, L. Zhang, M. Wan, Exponential graph regularized non-negative low-rank factorization for robust latent representation, Mathematics 10 (22) (2022). doi:10.3390/math10224314.
URL https://www.mdpi.com/2227-7390/10/22/4314
17. P. He, X. Xu, J. Ding, B. Fan, Low-rank nonnegative matrix factorization on stiefel manifold, Information Sciences 514 (2020) 131–148. doi:https://doi.org/10.1016/j.ins.2019.12.004.
URL https://www.sciencedirect.com/science/article/pii/S0020025519311120
18. J. L. Shang Ronghua, Song Jiuzheng, L. Yangyang, Double feature selection algorithm based on low-rank sparse non-negative matrix factorization, International Journal of Machine Learning and Cybernetics 11 (2020) 1891–1908. doi:10.1007/s13042-020-01079-6.
URL https://doi.org/10.1007/s13042-020-01079-6
19. J. Guo, S. Yin, Y. Sun, Y. Hu, Double manifolds regularized non-negative matrix factorization for data representation, in: 2020 25th International Conference on Pattern Recognition (ICPR), 2021, pp. 901–906. doi:10.1109/ICPR48806.2021.9412606.
20. J. Tang, H. Feng, Robust local-coordinate non-negative matrix factorization with adaptive graph for robust clustering, Information Sciences 610 (2022) 1058–1077. doi:https://doi.org/10.1016/j.ins.2022.08.023.
URL https://www.sciencedirect.com/science/article/pii/S0020025522009070
21. J. Tang, L. Shao, X. Li, K. Lu, A local structural descriptor for image matching via normalized graph laplacian embedding, IEEE Transactions on Cybernetics 46 (2) (2016) 410–420. doi:10.1109/TCYB.2015.2402751.
22. G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (1) (2013) 171–184. doi:10.1109/TPAMI.2012.88.