Edge Structure Learning via Low Rank Residuals for Robust Image Classification

Xiang-Jun Shen[1], Stanley Ebhohimhen Abhadiomhen[1,2], Yang Yang[1], Zhifeng Liu[1*], Sirui Tian[3]

[1]School of Computer Science and Communication Engineering, JiangSu University, JiangSu, 212013, China
[2]Department of Computer Science, University of Nigeria, Nsukka, Nigeria
[3]Department of Electronic Engineering, School of Electronic and Optical Engineering,
Nanjing University of Science and Technology, Nanjing, 210094, China
xjshen@ujs.edu.cn, stanley.abhadiomhen@unn.edu.ng, yyoung@ujs.edu.cn,
liuzf@ujs.edu.cn, Tiansirui@njust.edu.cn

Abstract

Traditional low-rank methods overlook residuals as corruptions, but we discovered that low-rank residuals actually keep image edges together with corrupt components. Therefore, filtering out such structural information could hamper the discriminative details in images, especially in heavy corruptions. In order to address this limitation, this paper proposes a novel method named ESL-LRR, which preserves image edges by finding image projections from low-rank residuals. Specifically, our approach is built in a manifold learning framework where residuals are regarded as another view of image data. Edge preserved image projections are then pursued using a dynamic affinity graph regularization to capture the more accurate similarity between residuals while suppressing the influence of corrupt ones. With this adaptive approach, the proposed method can also find image intrinsic low-rank representation, and much discriminative edge preserved projections. As a result, a new classification strategy is introduced, aligning both modalities to enhance accuracy. Experiments are conducted on several benchmark image datasets, including MNIST, LFW, and COIL100. The results show that the proposed method has clear advantages over compared state-of-the-art (SOTA) methods, such as Low-Rank Embedding (LRE), Low-Rank Preserving Projection via Graph Regularized Reconstruction (LRPP_GRR), and Feature Selective Projection (FSP) with more than 2% improvement, particularly in corrupted cases.

1 Introduction

As real-world images often contain redundant and corrupt features, such as noise and occlusion, their high dimensionality nature not only puts a huge computational burden on machine learning algorithms but also makes interpretation very difficult. Hence several efforts have been made over the years to overcome these limitations. Recent study (Parsons, Haque, and Liu 2004) indicates that multiple low dimensional subspaces can be leveraged to characterize high dimensional data. Consequently, Low-Rank Representation (LRR) (Liu et al. 2010, 2013) was proposed to learn low dimensional subspaces corresponding to a given high dimensional data. LRR excels in capturing the global structure of data and can also remedy the influence of corruption (Zhang et al. 2019; Abhadiomhen et al. 2021). However, LRR lacks feature extraction functionality. As a result, it cannot be used directly for linear dimensionality reduction (LDR). Although some LDR approaches exist such as classical PCA (Belhumeur, Hespanha, and Kriegman 1997) and Linear Discriminant Analysis (LDA)(Turk and Pentland 1991), they are not robust and therefore unable to preserve the geometrical structure of data in the projected subspace.

Figure 1: The left image in each row is the original data, the middle is the low rank part while the right one is the residual part. Our goal is to preserve those discriminative edges by finding robust image projections from low-rank residuals.

Motivated by the above challenges, many methods have been proposed recently incorporating LDR into LRR (LRR and LDR), mainly for learning image subspace representation. Notably, Low-Rank Embedding (LRE) (Wong et al. 2017) tries to uncover the inherent relationship in data for robust feature extraction. Low-Rank Sparse Preserving Projections (LSPP) (Xie et al. 2018) focuses on enhancing accuracy in corrupted data. Sparse Low-Rank Preserving Projection (SLRPP) (Liu et al. 2019) combines LRE and PCA models; an L_1-norm is further imposed on the low-rank representation to capture the local structure of data simultaneously. Similarly, to improve interpretability, Low-Rank Preserving Projection via Graph Regularized Reconstruction (LRPP_GRR) (Wen et al. 2019) applies a graph constraint and a sparse constraint to data reconstruction error and data projection, respectively. Feature Selective Projection (FSP) (Tang et al. 2020) was subsequently introduced to ensure a selective feature projection by merging LRR, feature extraction, and feature selection in a unified model. The more current methods, Double Low-Rank Representation (DLRRPD) (Fu et al. 2021) and Robust Clustering with Low-Rank Linear Embedding (RCLR) (Zhou et al. 2022) exploits global and local geometrical structures of data together using projection distance penalty and local property preservation, respectively. Despite the promising performances of

existing LRR and LDR methods, there are still some limitations. Specifically, they strive to keep the overall data structure which often neglects image edges that help enhance discriminability between images. This drawback is inherited from the traditional LRR model, which treats image edges as high-frequency signals that must be filtered out together with corrupt elements to preserve a low-rank structure, as shown in Figure. 1.

To tackle the above problem, this paper proposes a new LDR and LRR method named ESL-LRR, which can effectively preserve image edges by regarding low-rank residuals as another view of image data to find image projections. Firstly, through a coupled regularization term, we further integrate manifold learning into the LRR framework to learn the structure of its residuals. A dynamic affinity graph regularization is also introduced simultaneously to capture their similarity by using similar points to reconstruct one another. In other words, if two different residuals are more similar (with minimal distance), the corresponding weight value in the affinity graph is larger; and these points are perceived as harboring image edges as their principal component. Edge-preserved image projections are then learned adaptively from such residuals, with the influence of significantly distant pairs (corrupt or unrelated) being suppressed by a much smaller weight in the affinity graph. Furthermore, as a more precise affinity graph would result in better preservation of the geometric structure of data, the proposed approach can also find an intrinsic low-rank representation concurrently. Therefore, fusing both optimal LRR and the edge preserved projections, ESL-LRR further achieves robust linear dimensionality reduction to enhance accuracy. Section. 3 contains more detailed explanation and implementation specifics of the proposed method.

Our major contributions are summarized as follows:

1) Different from the conventional LRR approach, which consider residuals as corruptions, we discovered that low-rank residuals also keep image edges together with corrupt components. As a result, a new method is proposed named ESL-LRR, which views low-rank residuals as a combination of both structures.
2) The proposed ESL-LRR preserves the image edges by pursuing robust image projections from low-rank residuals through a manifold learning framework. In this way, the influence of corrupt residuals are adaptively subdued using a dynamic affinity graph regularization.
3) With our adaptive approach, ESL-LRR can find image intrinsic low-rank representation and much discriminative edge preserved projections simultaneously, unlike existing LRR and LDR methods which focus only on extracting the low-rank part. Thus, our strategy helps achieve more robustness to corruption.
4) Several experiments are performed in classification tasks. Extensive evaluation of ESL-LRR and SOTA methods reveal that the proposed method has superior advantages, especially in corrupted cases.

The remainder of this paper is structured as follows. Section. 2 presents a review of related works. Section. 3 formulates the proposed method. Experiments and analysis are given in Section. 4. And then, Section. 5 concludes the paper and highlights future works.

2 Related Works

This section describes LRR and LDR methods that are most closely related to our method. First, the baseline method LRR can be formulated as follows.

$$\begin{aligned} &\min_{\mathbf{Z}} \parallel \mathbf{Z}\parallel_* + \lambda \parallel \mathbf{E}\parallel_{2,1}, \\ &s.t., \ \mathbf{X} = \mathbf{XZ} + \mathbf{E}, \end{aligned} \tag{1}$$

where $\mathbf{X}$ and $\mathbf{Z}$ denotes a given data (simultaneously used as self-dictionary) and its low-rank matrix, respectively, $\mathbf{E}$ is the residual matrix introduced to capture the corrupt components. Thus, LRR can recover clean data with $\mathbf{XZ}$ (Wang et al. 2021).

Considering this perspective, various LDR methods were proposed based on LRR. These methods mainly focus on resolving the limitations of previous LDR approaches, which can be categorized into classical and graph-based methods. The classical approaches, including PCA, LDA, and their variants, such as block PCA (Benyong 2005) and Regularized LDA (Guo, Hastie, and Tibshirani 2007; Xu, Zhang, and Yang 2010), have the drawback of being unable to keep the geometric structure of the data in the projected subspace. Whereas, the graph-based approaches like Locality Preserving Projections (LPP) (He and Niyogi 2004) and Neighborhood Preserving Embedding (NPE) (He et al. 2005), are limited because they do not fully exploit the data's self-expression property in generating a similarity graph to ensure that unrelated samples are not connected. As a result, noise and outliers can impact their accuracy negatively. To tackle these disadvantages jointly, the LRR and LDR methods such as LRE and FSP, which are closely related to our proposed method, provide interesting techniques. For instance, by incorporating LDR directly into LRR, LRE can simultaneously capture data's global structure while tackling corruption relatively through the following model.

$$\begin{aligned} &\min_{\mathbf{Z,E,P}} \parallel \mathbf{Z}\parallel_* + \lambda \parallel \mathbf{E}\parallel_{2,1}, \\ &s.t., \ \mathbf{P^T X} = \mathbf{P^T XZ} + \mathbf{E}. \end{aligned} \tag{2}$$

As can be seen in the above formulation, LRE mainly presents a new data representation with $\mathbf{P^T X}$, which is integrated into Eq. (1) to learn a projection $\mathbf{P}$ to transform $\mathbf{X}$ from d original high-dimensional space to a projected r low-dimensional subspace where $r \ll d$. FSP then combines LRR, feature extraction, and feature selection in a unified model to facilitate selective feature projections using a sparsity term $\parallel \mathbf{P^T}\parallel_{2,1}$. This term allows for row sparsity of $\mathbf{P}$. However, when it is removed, FSP can easily reduce to the LRE model.

3 Proposed Method

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

3.1 Model Formulation

Specifically, when viewed from the frequency domain, it can be seen that LRR actually learns low-frequency signals, which are presumed to be the overall data structure,

and they delete the high ones, which are often a combination of corrupt elements and edge signals. It thus limits the LRR and LDR methods since edges usually retain enough discriminative features. Therefore, our approach aims to preserve image edges by finding robust projections $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_r)$ from low-rank residuals. On the other hand, we want to use only a subset of the residuals to achieve that since some will hold more corrupt components than edge features. As a result, manifold learning is further incorporated into Eq. (1) to learn the structure of low-rank residuals and subdue the influence of corrupt ones in obtaining $\mathbf{P}$ using the following model.

$$\begin{aligned} &\min_{\mathbf{P},\mathbf{E},\mathbf{W}} \frac{1}{2} \sum_{i,j} \parallel \mathbf{P}^{\mathbf{T}}\mathbf{e}_i - \mathbf{P}^{\mathbf{T}}\mathbf{e}_j\parallel_2^2 \mathbf{W}_{ij} + \lambda_3 \parallel \mathbf{W}\parallel_F^2, \\ &s.t., \ \mathbf{P}^{\mathbf{T}}\mathbf{P} = \mathbf{I}, \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}, \mathbf{W}1 = 1, \mathbf{W}_{ii} = 0, \mathbf{W} \geq 0, \end{aligned} \tag{3}$$

where $\frac{1}{2}\sum_{i,j} \parallel \mathbf{P}^{\mathbf{T}}\mathbf{e}_i - \mathbf{P}^{\mathbf{T}}\mathbf{e}_j\parallel_2^2 \mathbf{W}_{ij}$ is a coupling term and is an integral part of our model since we do not know beforehand which part of the data is corrupt. Therefore, we suppose that $\mathbf{E}$ could easily be exploited to reveal such prior knowledge. The reason is simple because, for two similar samples, $\mathbf{x}_i$ and $\mathbf{x}_j$, much smaller values of $\mathbf{e}_i$ and $\mathbf{e}_j$ would mean a good fit between them and their reconstruction $\mathbf{X}\mathbf{z}_i$ and $\mathbf{X}\mathbf{z}_j$ in low dimensional space. As a result, the Euclidean distance between $\mathbf{e}_i$ and $\mathbf{e}_j$ will be very small. And it is easy to deduce that these points will have image edges as their main components. In contrast, the distance between $\mathbf{e}_i$ and $\mathbf{e}_j$ would increase if one or both predominantly holds corrupt elements. In principle, these points are perceived unrelated even though they may belong to the same class. The latter is also true if, ideally, they are not similar samples' residuals. Motivated by these, the affinity graph $\mathbf{W}$ is obtained by using similar points in $\mathbf{E}$ to reconstruct each other. In this way, a smaller distance $\parallel \mathbf{P}^{\mathbf{T}}\mathbf{e}_i - \mathbf{P}^{\mathbf{T}}\mathbf{e}_j\parallel$ between data points $\mathbf{P}^{\mathbf{T}}\mathbf{e}_i$ and $\mathbf{P}^{\mathbf{T}}\mathbf{e}_j$ corresponds to a larger weight $\mathbf{W}_{ij}$, and zero otherwise to suppress the influence of such a combination in obtaining $\mathbf{P}$. In facilitating our approach, the non-negative constraints $\mathbf{W}_{ii} = 0$ and $\mathbf{W} \geq 0$ prevents the entries of $\mathbf{W}$ from being negative. $\mathbf{W}1 = 1$ is imposed to guarantee a normalized affinity graph where the sum of each row is equal to 1. In addition, the Frobenius norm is applied on the graph to further enforce it to be block-diagonal.

Therefore, combining Eq. (1) and Eq. (3), the ESL-LRR model described in Figure. 2 is formulated as follows:

$$\begin{aligned} &\min_{\mathbf{P},\mathbf{Z},\mathbf{W},\mathbf{E}} \frac{1}{2} \sum_{i,j} \parallel \mathbf{P}^{\mathbf{T}}\mathbf{e}_i - \mathbf{P}^{\mathbf{T}}\mathbf{e}_j\parallel_2^2 \mathbf{W}_{ij} + \lambda_1 \parallel \mathbf{Z}\parallel_* \\ &\qquad +\lambda_2 \parallel \mathbf{E}\parallel_{2,1} + \lambda_3 \parallel \mathbf{W}\parallel_F^2, \\ &s.t., \ \mathbf{P}^{\mathbf{T}}\mathbf{P} = \mathbf{I}, \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}, \mathbf{W}1 = 1, \mathbf{W}_{ii} = 0, \mathbf{W} \geq 0, \end{aligned} \tag{4}$$

As depicted by Eq. (4), ESL-LRR adaptively learns $\mathbf{P}$ with the orthogonal constraint $\mathbf{P}^{\mathbf{T}}\mathbf{P} = \mathbf{I}$ to find more discriminative projections that would cause the discrepancy between similar residuals to be significantly minimized. In such a way, ESL-LRR can more effectively preserve the structural information in residuals and eliminate the actual corruptions. This helps to prevent negative correlations to a large extent since corruption could initially cause two unrelated residuals to stay close and be perceived as similar samples.

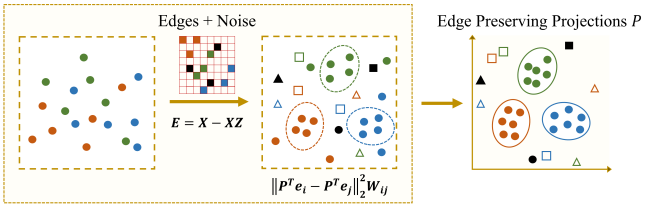


Figure 2: The framework of ESL-LRR. As illustrated using the above corrupted scenario, image edges are preserved by adaptively finding robust image projections P from low-rank residuals. $P^T X$ and $P^T XZ$ are then used for classification tasks where Z denotes the intrinsic low-rank representation of images and X is the actual data matrix.

Besides, it also induces a dynamic structure on $\mathbf{W}$ to capture the better similarity between residuals concurrently. Thus, a more intrinsic low-rank representation $\mathbf{Z}$ is learned alongside $\mathbf{P}$. Therefore, while the existing techniques like LRE and FSP can obtain low-rank projections, they do not provide such robustness to preserve the image edges to enhance accuracy. Hence, it is the main difference between them and the ESL-LRR model.

To solve Eq. (4), we propose an efficient optimization algorithm based on Augmented LaGrange Multiplier (ALM) strategy (Lin, Chen, and Ma 2010). First, the coupling term $\sum_{i,j} \parallel \mathbf{P}^{\mathbf{T}}\mathbf{e}_i - \mathbf{P}^{\mathbf{T}}\mathbf{e}_j\parallel^2 \mathbf{W}_{ij}$ is reformulated as $2tr(\mathbf{P}^{\mathbf{T}}\mathbf{E}\mathbf{L}_{\mathbf{W}}\mathbf{E}^{\mathbf{T}}\mathbf{P})$, where $\mathbf{L}_{\mathbf{W}} = \mathbf{D} - \mathbf{W}$ is the graph Laplacian of $\mathbf{W}$ in which $\mathbf{D} = diag(\sum_i \mathbf{W}_{ij})$ is a diagonal matrix. Then, we introduce the following auxiliary terms $\mathbf{S} = \mathbf{Z}$ and $\mathbf{E} = \hat{\mathbf{E}}$ to make Eq. (4) easier to solve (Abhadiomhen, Wang, and Shen 2021). As a result, Eq. (4) is rewritten as:

$$\begin{aligned} &\min_{\mathbf{P},\mathbf{Z},\mathbf{W},\mathbf{E},\hat{\mathbf{E}}} 2tr(\mathbf{P}^{\mathbf{T}}\hat{\mathbf{E}}\mathbf{L}_{\mathbf{W}}\hat{\mathbf{E}}^{\mathbf{T}}\mathbf{P}) + \lambda_1 \parallel \mathbf{S}\parallel_* \\ &\qquad +\lambda_2 \parallel \mathbf{E}\parallel_{2,1} + \lambda_3 \parallel \mathbf{W}\parallel_F^2, \\ &s.t., \ \mathbf{P}^{\mathbf{T}}\mathbf{P} = \mathbf{I}, \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}, \mathbf{E} = \hat{\mathbf{E}}, \mathbf{S} = \mathbf{Z}, \mathbf{W}1 = 1, \\ &\qquad \mathbf{W}_{ii} = 0, \mathbf{W} \geq 0. \end{aligned} \tag{5}$$

Besides, we also introduced another variable $\mathbf{F}$ to avoid Sylvester's equation (Feng, Wu, and Jing 2021) appearing in solving $\hat{\mathbf{E}}$. Thus, we have:

$$\begin{aligned} &\min_{\mathbf{P},\mathbf{Z},\mathbf{W},\mathbf{E},\hat{\mathbf{E}},\mathbf{F}} 2tr(\mathbf{P}^{\mathbf{T}}(\mathbf{F}\mathbf{L}_{\mathbf{W}}\hat{\mathbf{E}}^{\mathbf{T}})\mathbf{P}) + \lambda_1 \parallel \mathbf{S}\parallel_* \\ &\qquad +\lambda_2 \parallel \mathbf{E}\parallel_{2,1} + \lambda_3 \parallel \mathbf{W}\parallel_F^2, \\ &s.t., \ \mathbf{P}^{\mathbf{T}}\mathbf{P} = \mathbf{I}, \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}, \mathbf{E} = \hat{\mathbf{E}}, \mathbf{S} = \mathbf{Z}, \mathbf{W}1 = 1, \\ &\qquad \mathbf{W}_{ii} = 0, \mathbf{W} \geq 0. \end{aligned} \tag{6}$$

3.2 Optimization of ESL-LRR

Before the optimal solution of each variable in Eq. (6) can be described, the Augmented LaGrange function of Eq. (6) needs to be obtained first (Gong, Chen, and Chen 2020) as

Algorithm 1: The Algorithm of ESL-LRR.

Input: Data $\mathbf{X} \in R^{d\times n}$, parameter $\lambda_1, \lambda_2, \lambda_3$
Initialize: $\mathbf{P}$ with orthogonal column vector, $\mathbf{Z} = \mathbf{S} = \mathbf{E} = \mathbf{W} = \hat{\mathbf{E}} = \mathbf{F} = 0, Y_1 = Y_2 = Y_3 = Y_4 = 0, \rho = 1.1; \mu = 10^{-4}; \mu_{\max} = 10^6; \varepsilon = 10^{-5}$;
while not converged **do**
  Fix others and update $\mathbf{P}$ by Eq. (9).
  Fix others and update $\mathbf{Z}$ by Eq. (11).
  Fix others and update $\mathbf{S}$ by Eq. (13).
  Fix others and update $\mathbf{E}$ by Eq. (15).
  Fix others and update $\mathbf{W}$ by Eq. (22).
  Fix others and update $\hat{\mathbf{E}}$ by Eq. (24).
  Fix others and update $\mathbf{F}$ by Eq. (26).
  Update the multipliers.

$$\begin{aligned} Y_1 &= Y_1 + \mu(\mathbf{X} - \mathbf{XZ} - \mathbf{E}) \\ Y_2 &= Y_2 + \mu(\mathbf{S} - \mathbf{Z}) \\ Y_3 &= Y_3 + \mu\left(\hat{\mathbf{E}} - \mathbf{E}\right) \\ Y_4 &= Y_4 + \mu\left(\mathbf{F} - \hat{\mathbf{E}}\right) \end{aligned}$$

  Update μ by $\mu = \min(\rho\mu, \mu_{max})$.
  Check the convergence conditions: $\| \mathbf{X} - \mathbf{XZ} - \mathbf{E}\|_\infty < \varepsilon, \| \mathbf{S} - \mathbf{Z}\|_\infty < \varepsilon, \| \hat{\mathbf{E}} - \mathbf{E}\|_\infty < \varepsilon \, and \, \| \mathbf{F} - \hat{\mathbf{E}}\|_\infty < \varepsilon$
end while
Output: $\mathbf{P} \in R^{d\times r}$ and $\mathbf{Z} \in R^{n\times n}$.

follows.

$$\begin{gathered} \min_{\mathbf{P,Z,W,E},\hat{\mathbf{E}},\mathbf{F}} 2tr(\mathbf{P^T}(\mathbf{FL_W}\hat{\mathbf{E}}^\mathbf{T})\mathbf{P}) + \lambda_1 \| \mathbf{S}\|_* \\ +\lambda_2 \| \mathbf{E}\|_{2,1} + \lambda_3 \| \mathbf{W}\|_F^2 + tr(Y_1^T(\mathbf{X} - \mathbf{XZ} - \mathbf{E})) \\ +tr(Y_2^T(\mathbf{S} - \mathbf{Z})) + tr(Y_3^T(\hat{\mathbf{E}} - \mathbf{E})) + tr(Y_4^T(\mathbf{F} - \hat{\mathbf{E}})) \\ +\frac{\mu}{2}(\| \mathbf{X} - \mathbf{XZ} - \mathbf{E}\|_F^2 + \| \mathbf{S} - \mathbf{Z}\|_F^2 \\ + \| \hat{\mathbf{E}} - \mathbf{E}\|_F^2 + \| \mathbf{F} - \hat{\mathbf{E}}\|_F^2), \\ s.t., \ \mathbf{P^T P} = \mathbf{I}, \mathbf{W}1 = 1, \mathbf{W}_{ii} = 0, \mathbf{W} \geq 0, \end{gathered} \tag{7}$$

where Y_1, Y_2, Y_3, and Y_4 are LaGrange multipliers required for solving constrained problems. Therefore, separating the disconnected terms in Eq. (7), the minimization problem and each variable's ideal solution are given below.

P subproblem:

$$\begin{gathered} \min_{\mathbf{P}} tr(\mathbf{P^T}(\mathbf{FL_W}\hat{\mathbf{E}}^\mathbf{T})\mathbf{P}), \\ s.t., \ \mathbf{P^T P} = \mathbf{I}. \end{gathered} \tag{8}$$

Denoting $\mathbf{FL_W}\hat{\mathbf{E}}^\mathbf{T} = \mathbf{Q}$, $\mathbf{P}$ is obtained using the following eigenvalue decomposition problem whose optimal solution involves the set of k eigenvectors of the topmost k smallest eigenvalues of $\mathbf{Q}$.

$$\begin{gathered} \min_{\mathbf{P}} tr(\mathbf{P^T Q P}), \\ s.t., \ \mathbf{P^T P} = \mathbf{I}. \end{gathered} \tag{9}$$

Z subproblem:

$$\min_{\mathbf{Z}} \| \mathbf{X} - \mathbf{XZ} - \mathbf{E} + \tfrac{Y_1}{\mu}\|_F^2 + \| \mathbf{S} - \mathbf{Z} + \tfrac{Y_2}{\mu}\|_F^2. \tag{10}$$

Setting the derivative $\frac{\partial}{\partial \mathbf{Z}}$, the optimal $\mathbf{Z}$ is obtained through the following.

$$\mathbf{Z} = (\mathbf{X^T X} + \mathbf{I})^{-1}(\mathbf{X^T X} - \mathbf{X^T E} + \mathbf{S} + \tfrac{1}{\mu}\mathbf{X^T} Y_1 + \tfrac{Y_2}{\mu}). \tag{11}$$

S subproblem:

$$\min_{\mathbf{S}} \lambda_1 \| \mathbf{S}\|_* + \tfrac{\mu}{2} \| \mathbf{S} - \mathbf{Z} + \tfrac{Y_2}{\mu}\|_F^2. \tag{12}$$

Using the singular value thresholding (SVT)(Cai, Candès, and Shen 2010), the optimal $\mathbf{S}$ is obtained as:

$$\mathbf{S} = \Phi_{\frac{\lambda_1}{\mu}}(\mathbf{Z} - \tfrac{Y_2}{\mu}), \tag{13}$$

where Φ is a soft-thresholding operator.

E subproblem:

$$\begin{gathered} \min_{\mathbf{E}} \lambda_2 \| \mathbf{E}\|_{2,1} + \tfrac{\mu}{2} \| \mathbf{X} - \mathbf{XZ} - \mathbf{E} + \tfrac{Y_1}{\mu}\|_F^2 \\ +\tfrac{\mu}{2} \| \hat{\mathbf{E}} - \mathbf{E} + \tfrac{Y_3}{\mu}\|_F^2. \end{gathered} \tag{14}$$

Denoting $\mathbf{T} = \frac{1}{2}(\mathbf{X} - \mathbf{XZ} + \hat{\mathbf{E}} + \frac{1}{\mu}Y_1 + \frac{1}{\mu}Y_3)$, and $\tau = \frac{\lambda_2}{\mu}$, the i-th column of $\mathbf{E}$ is updated by the following formula.

$$\mathbf{E}(:,i) = \begin{cases} \frac{\|\mathbf{t}_i\| - \tau}{\|\mathbf{t}_i\|}\mathbf{t}_i, & \text{if } \tau < \|\mathbf{t}_i\| \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

W subproblem:

$$\begin{gathered} \min_{\mathbf{W}} \sum_{i,j} \| \mathbf{P^T}\hat{\mathbf{E}}_i - \mathbf{P^T}\hat{\mathbf{E}}_j\|_2^2 \mathbf{W}_{ij} + \tfrac{\lambda_3}{2} \| \mathbf{W}\|_F^2, \\ s.t., \ \mathbf{W}1 = 1, \mathbf{W}_{ii} = 0, \mathbf{W} \geq 0. \end{gathered} \tag{16}$$

Since Eq. (16) is independent for each i, $\mathbf{W}_i$ is solved separately as follows.

$$\begin{gathered} \min_{\mathbf{W}_i} \sum_i (\| \mathbf{P^T}\hat{\mathbf{E}}_i - \mathbf{P^T}\hat{\mathbf{E}}_j\|_2^2)\mathbf{W}_{ij} + \tfrac{\lambda_3}{2} \| \mathbf{W}_i\|_F^2, \\ s.t., \ \mathbf{W}1 = 1, \mathbf{W}_{ii} = 0, \mathbf{W} \geq 0. \end{gathered} \tag{17}$$

For simplicity, we use $\mathbf{A}_{ij}$ to symbolize $\mathbf{P^T}\hat{\mathbf{E}}_i - \mathbf{P^T}\hat{\mathbf{E}}_j$ such that $\mathbf{A}_i$ is a vector whose j-th element is $\mathbf{A}_{ij}$. As a result, we have:

$$\begin{gathered} \min_{\mathbf{W}_i} \sum_j \| \mathbf{W}_i + \tfrac{\mathbf{A}_i}{\lambda_3}\|_2^2, \\ s.t., \ \mathbf{W}1 = 1, \mathbf{W}_{ii} = 0, \mathbf{W} \geq 0. \end{gathered} \tag{18}$$

Furthermore, Eq. (18) is unconstrained as:

$$\min_{\mathbf{W}_i} \sum_j \| \mathbf{W}_i + \tfrac{\mathbf{A}_i}{\lambda_3}\|_2^2 - \eta(1^T\mathbf{W}_i - 1) - \zeta^T\mathbf{W}_i. \tag{19}$$

Therefore, setting the derivative $\frac{\partial}{\partial \mathbf{W}_i} = 0$, we have:

$$\mathbf{W}_i + \tfrac{1\mathbf{A}_i}{\lambda_3} - \eta 1 - \zeta = 0. \tag{20}$$

$\mathbf{W}_i$'s j entry is given as follows.

$$\mathbf{W}_{ij} + \tfrac{\mathbf{A}_{ij}}{\lambda_3} - \eta - \zeta_i = 0. \tag{21}$$

According to KKT conditions(Gordon and Tibshirani 2012), the optimal $\mathbf{W}_{ij}$ is obtained through the following formula.

$$\mathbf{W}_{ij} = (\tfrac{-\mathbf{A}_{ij}}{\lambda_3} + \eta)_+. \tag{22}$$

To arrive at the above, we suppose that $\mathbf{A}_{i1}$, $\mathbf{A}_{i2}$, ..., $\mathbf{A}_{in}$ are ordered from small to large. In order to impose $\mathbf{W}_{ii} = 0$, we always let $\mathbf{A}_{ii}$ place this value last despite having $\mathbf{A}_{ii} = 0$. Refer to (Nie et al. 2016) for a detailed derivation and proof.

$\hat{\mathbf{E}}$ subproblem:

$$\min_{\hat{\mathbf{E}}} tr(\mathbf{P^T F L_W \hat{E}^T P}) + \mu \parallel \hat{\mathbf{E}} - \mathbf{E} + \frac{Y_3}{\mu} \parallel_F^2. \quad (23)$$

Setting the derivative $\frac{\partial}{\partial \hat{\mathbf{E}}} = 0$, $\hat{\mathbf{E}}$ is obtained in the following way.

$$\hat{\mathbf{E}} = \frac{(2\mu \mathbf{E} - \mathbf{P P^T F L_W} - 2Y_3)}{2\mu}. \quad (24)$$

F subproblem:

$$\min_{\mathbf{F}} tr(\mathbf{P^T F L_W \hat{E}^T P}) + \mu \parallel \mathbf{F} - \hat{\mathbf{E}} + \frac{Y_4}{\mu} \parallel_F^2. \quad (25)$$

Setting the derivative $\frac{\partial}{\partial \mathbf{F}} = 0$, $\mathbf{F}$ is obtained as follows:

$$\mathbf{F} = \frac{(2\mu \hat{\mathbf{E}} - \mathbf{P P^T \hat{E} L_W}^T - 2Y_4)}{2\mu}. \quad (26)$$

The complete solution of the ESL-LRR model is summarized in Algorithm 1.

3.3 Computational Complexity

ESL-LRR's computational complexity mainly depends on multiplication operations, which cost $(t+1)O(n^3)$ (where t is the iteration count), nuclear norm minimization and inverse of a matrix operation, which cost $O(n^3)$ each, and eigen decomposition, which cost $O(d^2 r)$. As a result, the overall computational complexity of ESL-LRR in each of the iteration is $O(3n^3 + (r+1)n^3 + d^2 r)$.

3.4 Linear Multi-Classifier for Classification

Following the multivariate ridge regression (Zhang, Jiang, and Davis 2013), a new linear classifier is introduced for the proposed model as follows:

$$\min_{\mathbf{M}} \parallel Y - \mathbf{M P^T X}\ \mathbf{Z} \parallel_F^2 + \lambda \parallel \mathbf{M} \parallel_2^2. \quad (27)$$

where Y is the label matrix of the training data $\mathbf{X}$. Thus, by setting the derivative $\frac{\partial}{\partial \mathbf{M}} = 0$, $(\mathbf{M P^T X Z} - Y)\mathbf{Z^T X^T P} + \lambda \mathbf{M} = 0$, the weight $\mathbf{M}$ can be obtained using the following formula:

$$\mathbf{M} = Y \mathbf{Z^T X^T P} (\mathbf{P^T X Z Z^T X^T P} + \lambda \mathbf{I})^{-1} \quad (28)$$

Therefore, for a given test data $\mathbf{X}_{test}$, $\mathbf{Z}_{test}$, is calculated using Eq.(11), where $\mathbf{z}_i$ denotes the representation of i test sample. As a result, the label for i can be obtained as follows:

$$L = arg \max_L (\mathbf{M P^T X} z_i) \quad (29)$$

4 Experiments

In this section, several experiments are performed to assess the effectiveness of ESL-LRR in image classification. Firstly, we describe the experimental settings, including datasets, compared methods, and parameter selection, followed by the results and analysis in Section. 4.2.

Dataset	No of samples	Dimension	No of classes
MNIST	70000	784	10
USPS	1854	256	10
Yale	165	4096	15
LFW	2484	4096	38
COIL100	7200	4096	100

Table 1: Summary of the datasets

4.1 Experiment Settings

Dataset: To conduct various experiments, five popular benchmark image datasets are utilized: MNIST[1] and USPS [2] for handwritten image experiments, LFW[3] and Yale[4] for facial image experiments, and COIL100[5] for object image experiments. Section. 4.1 briefly describe each dataset.

Compared Methods: The performance of ESL-LRR is compared with that of seven SOTA methods, namely, Sparse Representation-based Classification (SRC) (Wright et al. 2008), Latent Low-rank Representation (LatLRR) (Liu and Yan 2011), LRE (Wong et al. 2017), LRPP_GRR (Wen et al. 2019), FSP (Tang et al. 2020), Generalized Least-squares Approach Regularized with Graph Embedding (GLSRGE) (Shen et al. 2020), and DLRRPD (Fu et al. 2021). In these experiments, each algorithm's recognition accuracy is first studied on original dataset using uniform selected projections of $r = 80$ dimensions, similar to the approach of FSP. Next, the datasets are unsystematically divided into two groups to study the robustness of each method to data corruption. Thus, pepper noise of 20% and 40% densities are randomly added to the images of MNIST, Yale, and COIL100 by replacing original pixel values. In the second category, block occlusions (some black blocks) of 12×12 and 18×18 sizes are randomly introduced into the images of LFW while USPS is corrupted with 3×3 and 9×9 sizes in concurrence with its original dimensions.

Parameter Selection: For parameters in each algorithm, we set the values according to optimal settings found in their respective literature. In the case of ESL-LRR, parameters λ_1, λ_2, and λ_3 need tuning to achieve a robust solution based on the objective function of Eq. (6). However, because there are three parameters and several datasets involved, finding a good combination of the parameters is tricky. As far as we know, it is still an open problem in the literature. Hence, a reasonable strategy is adopted from reference (Wen, Xu, and Liu 2020) to find optimum values of the parameters using $[0.001, 0.01, 0.1, \ldots, 1000]$ as the candidate set. First, a good combination of λ_1 and λ_2 is pursued by varying them while keeping λ_3 fixed to a random value from above. Then, after setting λ_1 and λ_2, to their best values from the previous round, λ_3 is tuned to arrive at the final optimal set.

For uniformity, the k-nearest neighbor classifier is applied on matrix $\mathbf{P^T X}$ of each algorithm with the k-nearest neigh-

[1] http://yann.lecun.com/exdb/mnist/

[2] https://www.kaggle.com/bistaumanga/usps-dataset

[3] http://vis-www.cs.umass.edu/lfw/

[4] http://vision.ucsd.edu/content/yale-face-database

[5] https://www.kaggle.com/jessicali9530/coil100

Noise density	0%	20%	40%
SRC	81.02±2.19(76)	77.87±1.12(77)	72.31±1.64(58)
LatLRR	83.27±2.82(74)	79.32±2.15(78)	74.43±2.20(63)
LRE	87.67±2.36(74)	83.94±2.04(79)	80.36±2.27(60)
LRPP_GRR	89.95±1.40(43)	88.87±1.55(55)	83.50±2.03(68)
FSP	93.08±1.57(75)	87.73±1.46(60)	85.38±1.61(64)
GLSRGE	95.11±1.18(70)	90.91±1.33(55)	88.22±1.21(53)
DLRRPD	96.70±0.76(73)	92.10±1.49(72)	87.65±1.76(64)
ESL-LRR	**97.86±0.58(62)**	**93.52±1.01(72)**	**91.41±1.38(78)**

Table 2: Mean Classification Accuracy ± standard deviation(%) of different algorithms on MNIST dataset corrupted with various degrees of random pepper noise. The best dimension is reported, and it is denoted in the bracket

Datasets	Yale			COIL100		
Noise density	0%	20%	40%	0%	20%	40%
SRC	35.48±5.16(34)	29.22±5.01(22)	28.62±4.68(45)	79.83±3.21(34)	74.60±2.30(77)	72.82±1.83(38)
LatLRR	71.17±4.90(43)	68.26±5.44(68)	62.73±6.12(28)	81.08±2.30(32)	78.62±2.12(25)	76.24±1.68(50)
LRE	51.13±5.13(31)	47.52±4.53(78)	45.93±4.05(26)	86.20±1.25(22)	81.71±1.54(18)	78.30±1.72(54)
LRPP_GRR	56.34±6.82(14)	51.46±5.60(70)	46.44±6.91(47)	95.69±1.10(33)	94.82±1.02(22)	89.68±1.24(68)
FSP	72.79±4.31(37)	69.70±5.22(73)	62.14±5.45(33)	97.15±1.23(58)	88.22±2.25(74)	86.74±1.44(35)
GLSRGE	79.62±5.10(61)	75.38±4.84(65)	72.33±6.42(46)	97.16±1.09(28)	96.00±0.77(26)	92.76±1.28(28)
DLRRPD	79.61±4.64(58)	74.77±5.75(73)	71.97±4.53(38)	97.77±0.24(36)	97.48±0.34(39)	94.47±1.16(27)
ESL-LRR	**86.19±4.82(59)**	**80.38±4.31(64)**	**78.02±5.24(56)**	**100(22)**	**99.69±0.02(45)**	**97.54±0.74(25)**

Table 3: Mean Classification Accuracy ± standard deviation(%) of different algorithms on MNIST, Yale, and COIL100 datasets corrupted with various degrees of random pepper noise. The best dimension is reported, and it is denoted in the bracket

bor equally set to 1. Therefore, to prevent generalization loss, all experiments are performed ten times, and the mean classification accuracy and standard deviation are reported. Note that 70% of samples are randomly selected as training set and 30% as testing set in experiments based on findings in the literature, which reveals that the best results can be obtained from such split. Also, ESL-LRR's classification accuracy is evaluated using the introduced classifier in Section. 3.4. And all experiments were performed using Matlab 2020a installed on an AMD Ryzen 9-5950X system with 64GB RAM.

4.2 Result and Analysis

Here, we analyze the results presented in Tables. 2 to 5 and Figure. 3.

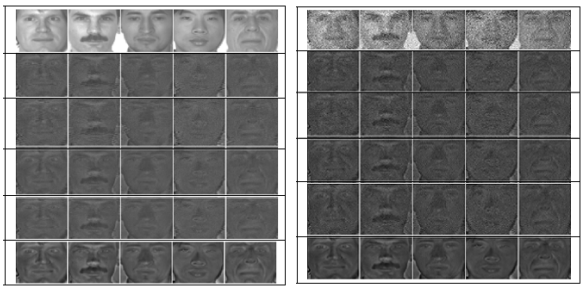
Tables. 2 and 3 displays the recognition accuracy of different algorithms on MNIST, Yale, and COIL100 datasets under various degrees of random pepper noise. According to the results, ESL-LRR achieves the best performance than the compared methods. However, one may notice that the performance gap between the proposed method and the newer LRR and LDR methods is not significant on clean data. The reason is that these methods can extract the discriminative low-rank features in such cases. Thus, the effectiveness of ESL-LRR is more evident in corrupted subjects. In particular, while its accuracy is better than DLRRPD and FSP by 1.16%, and 4.78% on MNIST, it becomes much better in noisy settings by 1.42%, 5.79% with 20% noise and 3.76%, 6.03% with 40% noise. ESL-LRR also obtains more superior results on Yale and COIL100, notably with 40% noise contamination. It validates the effectiveness of our edge-preserving approach. Besides, Figure. 3 shows the images recovered by newer algorithms with five top basis vectors of their respective projections acquired on Yale. As can be seen in the figure, ESL-LRR better preserves image edges than other methods.

Table. 4 presents the recognition accuracy of different algorithms on USPS and LFW datasets with various sizes of random continuous pixel block occlusions. As can be observed in the table, ESL-LRR maintains the best performance with only 3.57% degradation on USPS. GLSRGE and DLRRPD which uses graph embedding and projection distance penalty, respectively, to retain global and local geometrical data structures, have the closest performance to our method. However, DLRRPD displays slightly less resistance to occlusion in handwritten recognition than other algorithms, with a performance decrease of 6.98%. In addition, ESL-LRR is more robust in facial recognition (LFW dataset), with over 4% better accuracy than DLRRPD in the case of 18×18 occlusion block size.

Table. 5 shows the recognition rate of ESL-LRR using the new classification strategy presented in Section. 3.4. Compared with the results obtained on the KNN classifier, it can be observed that ESL-LRR achieves a higher classification accuracy using the proposed classifier due to the impact of added low-rank representation with $\mathbf{P^TXZ}$. Nonetheless, the results are fairly close, confirming that our image projec-

Datasets	USPS			LFW		
Occlusion size	Original data	3×3	9×9	Original data	12×12	18×18
SRC	92.78±3.81(57)	89.56±2.86(45)	85.13±2.95(37)	70.38±2.32(53)	69.26±2.31(75)	61.68±2.48(53)
LatLRR	97.91±0.89(37)	94.42±1.12(44)	90.52±1.47(46)	72.08±3.40(56)	65.98±3.78(74)	63.64±2.31(57)
LRE	88.26±3.81(22)	86.65±4.64(76)	83.85±5.38(20)	74.29±3.16(80)	66.49±1.82(80)	64.24±1.47(55)
LRPP_GRR	84.19±5.18(25)	82.82±4.27(34)	80.30±3.54(19)	76.23±2.77(76)	69.21±2.71(73)	68.06±1.85(65)
FSP	91.97±1.92(56)	89.55±3.51(35)	85.61±3.22(15)	74.80±1.54(79)	72.11±2.22(72)	67.58±2.12(60)
GLSRGE	95.99±1.69(22)	95.48±1.98(26)	91.88±2.17(37)	76.67±2.10(77)	71.17±1.88(62)	69.78±1.83(71)
DLRRPD	98.08±0.50(16)	97.75±0.69(38)	91.10±2.38(34)	77.57±2.41(70)	72.05±2.40(66)	70.30±1.83(57)
ESL-LRR	**99.80±0.02(13)**	**98.74±0.01(22)**	**96.23±0.02(27)**	**80.66±1.40(64)**	**75.85±2.01(53)**	**75.08±1.50(58)**

Table 4: Mean Classification Accuracy ± standard deviation(%) of different algorithms on USPS and LFW datasets corrupted with various degrees of random continuous pixel block occlusions. The best dimension is reported, and it is denoted in the bracket



(a) (b)

Figure 3: Images recovered by different algorithms from five top basis vectors of respective data projections obtained on the Yale dataset. (a) Original data, (b) 20% random pepper noise. The first row in each figure contains the original data. The second, third, fourth, fifth and sixth rows are the images recovered by LRE, LRPP_GRR, FSP, DLRRPD and ESL-LRR, respectively.

tions have a strong discriminability which is appropriate for classification, regardless of the classifier.

4.3 Parameter Sensitivity and Convergence Analysis

Parameter Sensitivity: ESL-LRR's sensitivity to the tunable parameters is investigated using MNIST and COIL100 (with 20% random pepper noise added) datasets by adopting the settings described in Section. 4.1. As can be seen in Figure. 4, the accuracy of ESL-LRR is relatively stable, with low or high values for λ_1. In contrast, more superior results are obtained on both datasets with high values from 10 up to 1000 for λ_2.

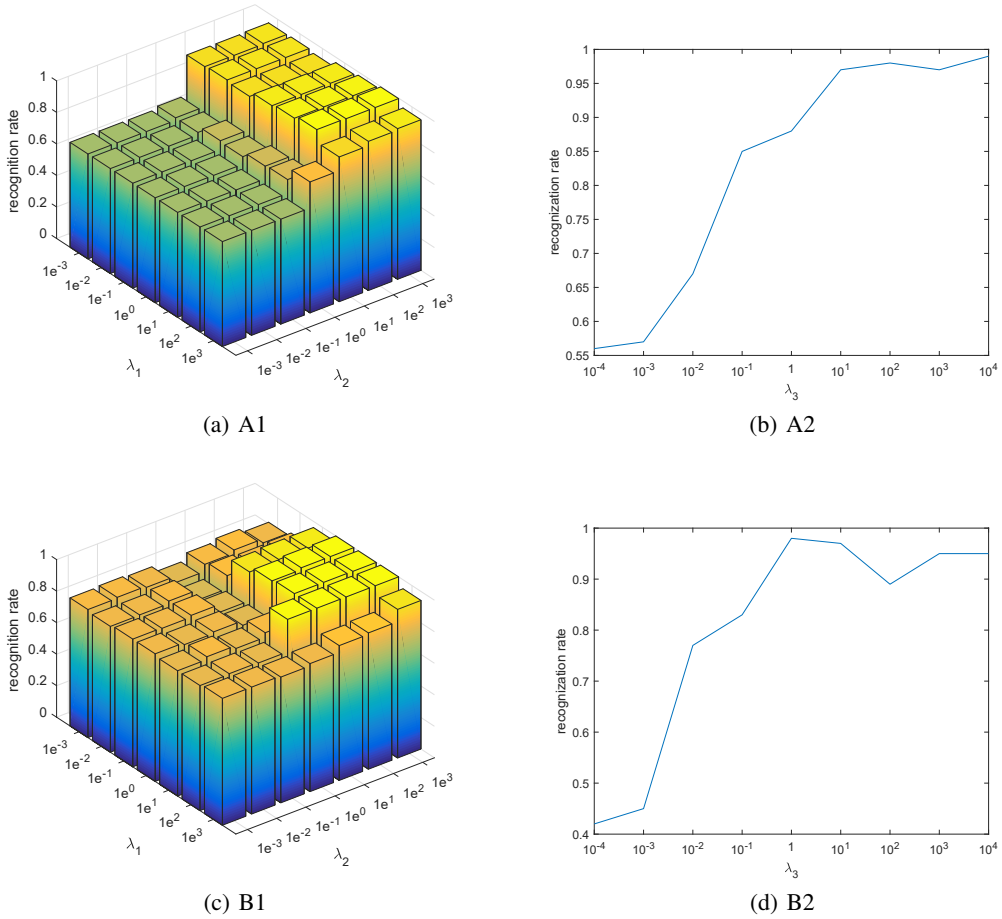
Convergence Analysis: The convergence of the inexact ALM method with a maximum of two sub-problems has been theoretically verified (Lin, Chen, and Ma 2010). However, proving the convergence of the inexact ALM method with three or more sub-problems is still difficult (Luo et al. 2018; Li et al. 2019). Moreover, ESL-LRR has seven sub-blocks, making it even more challenging to prove. Fortunately, as experimentally verified on actual data using Figure. 5(b), our algorithm has strong convergence property. Furthermore, as shown in Figure. 5(a), the runtime of ESL-LRR is comparable to that of most algorithms, affirming that it is efficient.

Data type	Original data	Corrupted data
MNIST	98.40±0.23(47)	96.31±1.02(59)
USPS	100(15)	100(17)
LFW	82.07±1.28(57)	78.13±1.49(59)
Yale	87.80±2.75(43)	82.76±2.35(43)
COIL100	100(18)	100(40)

Table 5: Mean Classification Accuracy ± standard deviation(%) of ESL-LRR using the introduced classifier on clean and corrupted data: MNIST, YALE, and COIL100 (20% random pepper noise). LFW (12×12 random continuous pixel block occlusions) and USPS (3×3 type)

5 Conclusion

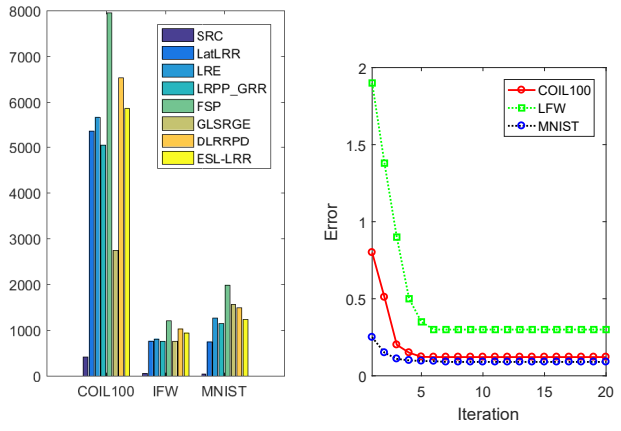
A new method named ESL-LRR is proposed in this paper for image classification. Unlike existing LRR and LDR methods which focus only on extracting the low-rank features of images, ESL-LRR preserves image edges by finding robust image projections from low-rank residuals through a manifold learning framework. ESL-LRR can also learn image intrinsic low-rank representation together to enhance accuracy. Several experiments were performed on five benchmark image datasets to evaluate the effectiveness of ESL-LRR on classification tasks. The experimental results clearly show that it significantly outperforms SOTA methods, especially in corrupted scenarios. Also, the computational time experiments reveal that ESL-LRR's runtime is at the same level as most compared methods. However, because the proposed algorithm has seven sub-blocks to update in each iteration, we will explore large-scale techniques to improve its efficiency in future work. Additionally, we intend to explore the deep paradigm to enhance accuracy further.



(a) A1 (b) A2

(c) B1 (d) B2

Figure 4: The parameter sensitivity study with respect to recognition rate on, (a) MNIST, and (b) COIL100 datasets where A1, B1 fixes λ_3 (A1= 10, B1=1) to adjust λ_1 and λ_2, and A2, B2 then fixes λ_1 and λ_2 to adjust λ_3.



(a) (b)

Figure 5: (a) Runtime (in seconds) of different algorithms and (b) ESL-LRR's convergence behavior on the norms of $(\hat{\mathrm{E}} - E)$ (a convergence condition) on three datasets.

Acknowledgments

This research was funded in part by the Key Program of National Natural Science Foundations of China under Grant No. 41930110.

References

Abhadiomhen, S. E.; Wang, Z.; and Shen, X. 2021. Coupled low rank representation and subspace clustering. *Applied Intelligence*, 1–17.

Abhadiomhen, S. E.; Wang, Z.; Shen, X.; and Fan, J. 2021. Multiview Common Subspace Clustering via Coupled Low Rank Representation. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 12(4): 1–25.

Belhumeur, P. N.; Hespanha, J. P.; and Kriegman, D. J. 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7): 711–720.

Benyong, S. X. L. B. L. 2005. Face Recognition Based on Block-PCA [J]. *Computer Engineering and Applications*, 27.

Cai, J.-F.; Candès, E. J.; and Shen, Z. 2010. A singular value thresholding algorithm for matrix completion. *SIAM Journal on optimization*, 20(4): 1956–1982.

Feng, Y.-Y.; Wu, Q.-B.; and Jing, X.-N. 2021. The MGHSS for Solving Continuous Sylvester Equation. *Complexity*, 2021.

Fu, Z.; Zhao, Y.; Chang, D.; Zhang, X.; and Wang, Y. 2021. Double Low-Rank Representation With Projection Distance Penalty for Clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5320–5329.

Gong, X.; Chen, W.; and Chen, J. 2020. A low-rank tensor dictionary learning method for hyperspectral image denoising. *IEEE Transactions on Signal Processing*, 68: 1168–1180.

Gordon, G.; and Tibshirani, R. 2012. Karush-kuhn-tucker conditions. *Optimization*, 10(725/36): 725.

Guo, Y.; Hastie, T.; and Tibshirani, R. 2007. Regularized linear discriminant analysis and its application in microarrays. *Biostatistics*, 8(1): 86–100.

He, X.; Cai, D.; Yan, S.; and Zhang, H.-J. 2005. Neighborhood preserving embedding. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, 1208–1213. IEEE.

He, X.; and Niyogi, P. 2004. Locality preserving projections. *Advances in neural information processing systems*, 16(16): 153–160.

Li, R.; Zhang, C.; Hu, Q.; Zhu, P.; and Wang, Z. 2019. Flexible Multi-View Representation Learning for Subspace Clustering. In *IJCAI*, 2916–2922.

Lin, Z.; Chen, M.; and Ma, Y. 2010. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*.

Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2013. Robust recovery of subspace structures by low-rank representation. *IEEE transactions on pattern analysis and machine intelligence*, 35(1): 171–184.

Liu, G.; Lin, Z.; Yu, Y.; et al. 2010. Robust subspace segmentation by low-rank representation. In *Icml*, volume 1, 8. Citeseer.

Liu, G.; and Yan, S. 2011. Latent low-rank representation for subspace segmentation and feature extraction. In *2011 international conference on computer vision*, 1615–1622. IEEE.

Liu, Z.; Wang, J.; Liu, G.; and Pu, J. 2019. Sparse low-rank preserving projection for dimensionality reduction. *IEEE Access*, 7: 22941–22951.

Luo, S.; Zhang, C.; Zhang, W.; and Cao, X. 2018. Consistent and specific multi-view subspace clustering. In *Thirty-second AAAI conference on artificial intelligence*.

Nie, F.; Wang, X.; Jordan, M.; and Huang, H. 2016. The constrained laplacian rank algorithm for graph-based clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30.

Parsons, L.; Haque, E.; and Liu, H. 2004. Subspace clustering for high dimensional data: a review. *Acm sigkdd explorations newsletter*, 6(1): 90–105.

Shen, X.-J.; Liu, S.-X.; Bao, B.-K.; Pan, C.-H.; Zha, Z.-J.; and Fan, J. 2020. A generalized least-squares approach regularized with graph embedding for dimensionality reduction. *Pattern Recognition*, 98: 107023.

Tang, C.; Liu, X.; Zhu, X.; Xiong, J.; Li, M.; Xia, J.; Wang, X.; and Wang, L. 2020. Feature selective projection with low-rank embedding and dual Laplacian regularization. *IEEE Transactions on Knowledge and Data Engineering*, 32(9): 1747–1760.

Turk, M.; and Pentland, A. 1991. Eigenfaces for Recognition: Journal of Cognitive Neurosicence.

Wang, Z.-y.; Abhadiomhen, S. E.; Liu, Z.-f.; Shen, X.-j.; Gao, W.-y.; and Li, S.-y. 2021. Multi-view intrinsic low-rank representation for robust face recognition and clustering. *IET Image Processing*, 15(14): 3573–3584.

Wen, J.; Han, N.; Fang, X.; Fei, L.; Yan, K.; and Zhan, S. 2019. Low-rank preserving projection via graph regularized reconstruction. *IEEE Transactions on Cybernetics*, 49(4): 1279–1291.

Wen, J.; Xu, Y.; and Liu, H. 2020. Incomplete multiview spectral clustering with adaptive graph learning. *IEEE transactions on cybernetics*, 50(4): 1418–1429.

Wong, W. K.; Lai, Z.; Wen, J.; Fang, X.; and Lu, Y. 2017. Low-rank embedding for robust image feature extraction. *IEEE Transactions on Image Processing*, 26(6): 2905–2917.

Wright, J.; Yang, A. Y.; Ganesh, A.; Sastry, S. S.; and Ma, Y. 2008. Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, 31(2): 210–227.

Xie, L.; Yin, M.; Yin, X.; Liu, Y.; and Yin, G. 2018. Low-rank sparse preserving projections for dimensionality reduction. *IEEE Transactions on Image Processing*, 27(11): 5261–5274.

Xu, Y.; Zhang, D.; and Yang, J.-Y. 2010. A feature extraction method for use with bimodal biometrics. *Pattern recognition*, 43(3): 1106–1115.

Zhang, Y.; Jiang, Z.; and Davis, L. S. 2013. Learning structured low-rank representations for image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 676–683.

Zhang, Z.; Ren, J.; Li, S.; Hong, R.; Zha, Z.; and Wang, M. 2019. Robust subspace discovery by block-diagonal adaptive locality-constrained representation. In *Proceedings of the 27th ACM international conference on multimedia*, 1569–1577.

Zhou, J.; Pedrycz, W.; Wan, J.; Gao, C.; Lai, Z.-H.; and Yue, X. 2022. Low-Rank Linear Embedding for Robust Clustering. *IEEE Transactions on Knowledge and Data Engineering*.