Robust Discriminative t-Linear Subspace Learning for Image Feature Extraction

Kangdao Liu, Xiaolin Xiao, *Member, IEEE*, Jinkun You, *Graduate Student Member, IEEE*, and Yicong Zhou, *Senior Member, IEEE*

Abstract—Subspace learning has been widely applied for joint feature extraction and dimensionality reduction, demonstrating significant efficacy. Numerous subspace learning methods with diverse assumptions regarding the criteria for the target subspaces have been developed to obtain compact and interpretable data representations. However, when applied to image data, existing methods fail to fully exploit the inherent correlations within the image set. This paper proposes a Robust Discriminative t-Linear Subspace Learning model (RDtSL) to tackle this issue using t-product. The model mainly has four strengths: 1) Taking advantage of t-product, RDtSL learns the projection basis directly from the image set while fully exploiting its internal correlations; 2) Based on its energy preservation module, RDtSL retains the primary energy of samples in the learned subspace, maintaining satisfactory performance even with low subspace dimensions; 3) Class-distinctive features are effectively preserved in the learned representations due to the incorporation of the classification module; 4) Relying on its graph embedding module, RDtSL learns an affinity graph of samples adaptively to enrich the data representations with locality and similarity information. The harmonious balance maintained between the three proposed modules helps RDtSL learn discriminative and informative data representations. We also develop an iterative algorithm to solve RDtSL. Extensive experiments on benchmark databases demonstrate the superiority of the proposed model.

Index Terms—Dimensionality reduction, image feature extraction, subspace learning, t-product.

I. INTRODUCTION

WITH the rapid advancement in information collection and storage capabilities, real-world data continue to grow in size. Although high-dimensional data contain a wealth of valuable information, an increase in dimensionality can substantially compromise both the effectiveness and efficiency of data processing algorithms. This phenomenon is commonly referred to as the "curse of dimensionality". Consequently, learning compact representations of high-dimensional data becomes pivotal for subsequent analyses; this can be achieved by extracting the most representative features from the samples.

In the past few decades, a wide range of feature extraction methods have been proposed for various types of data. For instance, in recent works by Wang et al. [1] and Qin et al. [2], unsupervised class-wise clustering and supervised representation learning are blended in a synergy. Significant local patterns are automatically identified to facilitate classification decisions, and the supervisory signal from classification boosts meaningful clustering and discriminative feature learning. Moreover, recent studies conducted by Wen et al. [3] and Hu et al. [4] focus on learning compact data representations through adaptively learning and embedding the affinity graph of samples. They also ensure inter-class separability through the joint learning of a classification model. In another work by Xiao et al. [5], a feature extraction method specifically designed for color images based on quaternion algebra is proposed. This approach considers the correlation between image channels, enabling a more comprehensive representation of color image features. Among the feature extraction methods, subspace learning methods learn explicit projection bases with specific physical interpretations and accomplish joint feature extraction and dimensionality reduction by mapping high-dimensional data samples into low-dimensional subspaces. When compared to non-linear methods, particularly those based on deep neural networks, subspace learning methods offer significant advantages in terms of robust interpretability, faster training and validation speed, and commendable performance even with limited training data. To date, a variety of subspace learning methods [6], [7], [8], [9] have been developed, with diverse assumptions regarding the criteria for the target subspace. Principal Component Analysis (PCA) [10] aims to maximize the post-projection variance to preserve the primary energy of the samples; Locality Preserving Projection (LPP) [11], Neighborhood Preserving Embedding (NPE) [12] and Low-Rank Preserving Projections [13] assume that the affinity of samples in the subspace is dependent on their similarity in the original data space; Linear Discriminant Analysis (LDA) [14] focuses on label information by maximizing inter-class scatter and minimizing intra-class one; Low-Rank Embedding (LRE) [15] aims to learn the projection basis by exploring the underlying low-rank structure of samples.

In practical applications, classical subspace learning methods treat data samples as vectors and require that the samples

Kangdao Liu, Jinkun You, and Yicong Zhou are with the Department of Computer and Information Science, University of Macau, Macau, China (e-mail: kangdaoliu@gmail.com; youjinkun09@gmail.com; yicongzhou@um.edu.mo).

Xiaolin Xiao is with the School of Computer Science, South China Normal University, Guangzhou 510631, China (e-mail: shellyxiaolin@gmail.com).

be flattened into long vectors in advance when applied to image data. However, image sets naturally possess a strong spatial and cross-sample structure (a third-order structure); simply stitching rows or columns of images into long vectors and then forming them into matrices breaks down the spatial coherency. As a result, the inherent correlations within the image set are overlooked in the subsequent learning phase, leading to deficiencies in data representations. Additionally, massive computation of high-dimensional image vectors poses various challenges, including the small sample size problem [16]. In response, Two-Dimensional PCA (2DPCA) [17] was introduced to mitigate these negative impacts by processing image columns or rows independently. Specifically, either columns or rows are projected separately into the target subspace, allowing for partial preservation of the spatial structures. Furthermore, 2DPCA has the benefit of avoiding the computation of high-dimensional vectors. Following the same rationale, 2DLDA [18] and 2DLPP [19] were developed. These 2D methods have proven more effective for processing image data as they take the valued spatial structures into consideration. However, since the 2D methods focus solely on either row or column space, they are unable to fully explore the third-order structure. To address this limitation, multi-linear methods [20], [21] were introduced, solving it with the aid of high-order tensors. When applied to image data, multi-linear methods can take the correlations among rows, columns, and samples into account, as they learn the projection basis from the unfoldings of the high-order tensor along all its directions. However, since the unfoldings are processed individually, correlations across different modes cannot be considered simultaneously, leading multi-linear methods to typically converge to a local optimum.

Apart from the deficiency in processing images, the aforementioned subspace learning methods mainly rely on the squared l_2-norm for measurement, which exacerbates the impact of outliers on the subspace learning process. As a result, these methods are highly susceptible to data corruption, such as noise and occlusions. To enhance robustness, RLSL [22] employs a l_1-norm-based reconstruction loss term to fit noise, allowing the uncontaminated samples to be adaptively recovered during the learning phase. Similarly, RSLDA [23] and LLRSE [24] incorporate a $l_{2,1}$-norm-based reconstruction loss term into the LDA model [14] and the low-rank embedding model [15], respectively, with the same purpose. In addition to enhancing the robustness, by minimizing the reconstruction loss, the primary energy of data samples can be effectively preserved in the learned subspace, even if the dimensionality of the target subspace is low. Experimental results reported in [23] show that RSLDA maintains satisfactory performance as the subspace dimension decreases, demonstrating the beneficial effects of the energy preservation module. In a similar vein, various subspace learning methods [25], [26], [27], [28] employ robust estimators to measure and minimize reconstruction loss, aiming to mitigate the impact of data contamination while preserving the primary energy of samples. Nevertheless, despite these advancements, the aforementioned robust methods encounter three limitations: 1) The similarity and locality information

of samples are ignored and are not retained in the learned representations, which are crucial for subsequent processing and analysis. 2) None of them addresses the deficiency of the classical subspace learning methods in processing images. The calculation of reconstruction loss also requires the vectorization operation when applied to images. Therefore, they always produce sub-optimal representations due to the irreversible loss of spatial information and the extensive computations of high-dimensional vectors. 3) Many of them incorporate a regression-type classification module to further enhance discriminative ability. However, due to this module, part of them cannot be flexibly applied to dimensionality reduction since they can only project the high-dimensional samples into the C-dimensional label space, where C denotes the number of classes. Besides, the classification module also requires the samples to be flattened beforehand when applied to images, thus neglecting the internal correlations.

To fully and simultaneously leverage the correlations within the image set, a tensor-tensor product known as the t-product was proposed [29], [30], providing a comprehensive framework for executing any tensor linear transformations and various tensor factorizations. By employing third-order tensor representation in conjunction with the t-product, image sets can be effectively represented and transformed without breaking down their spatial structures, thereby enabling the simultaneous exploitation of correlations within the image set and eliminating the need for computations of high-dimensional vectors. The t-product-based operations have been effectively applied in various fields, such as image restoration [31], [32] and clustering [33], [34], demonstrating superior performance compared to matrix-based and multi-linear-based approaches. To address the aforementioned limitations of existing subspace learning methods, this paper introduces the **R**obust **D**iscriminative **t**-Linear **S**ubspace **L**earning model (RDtSL), a novel t-linear subspace learning method for joint image feature extraction and dimensionality reduction. The utilization of third-order tensor representation and t-product sets RDtSL apart from existing feature extraction methods [4], [6], [23], [25]. Additionally, this approach simultaneously considers data energy preservation, class-distinctive feature extraction, and local pattern capture for discriminative feature learning. The key contributions of this paper are highlighted as follows:

1) We propose the Robust Discriminative t-Linear Subspace Learning model (RDtSL) for joint image feature extraction and dimensionality reduction. By utilizing third-order tensor representation in conjunction with the t-product, RDtSL can fully exploit the inherent correlations within the image set to enhance subspace learning while eliminating the need for high-dimensional vector computations.

2) RDtSL effectively preserves the primary energy of image samples within the reduced-dimensionality space by incorporating the proposed energy preservation module. We provide both theoretical and experimental evidence to demonstrate the feasibility of this module. Leveraging it, RDtSL maintains high levels of performance even with low subspace dimensions.

3) The discriminative ability of RDtSL is enhanced by incorporating the proposed classification module and graph

embedding module. The graph embedding module further enriches the representations with locality and similarity information, allowing for the automatic mining of local patterns.

4) We develop an iterative algorithm to solve RDtSL and theoretically analyze its computational complexity. Extensive experiments demonstrate that RDtSL outperforms the state-of-the-arts regarding discriminative ability and robustness.

The remainder of this paper is structured as follows. Section II reviews closely-related works and provides the preliminaries. Afterward, Section III introduces the proposed model with the optimization algorithm. Then, Section IV reports the experimental results and conducts a series of model analyses. Finally, the paper is summarized in Section V.

II. NOTATIONS AND PRELIMINARIES

Throughout this paper, lowercase letters, bold lowercase letters, and bold uppercase letters represent scalars, vectors, and matrices, e.g., x, $\mathbf{x}$, and $\mathbf{X}$. $\mathbf{X}'$ is the conjugate transpose of $\mathbf{X}$. $\|\mathbf{X}\|_F$, $\|\mathbf{X}\|_1$ and $\|\mathbf{X}\|_{2,1}$ denote the Frobenius norm (F-norm), l_1-norm and $l_{2,1}$-norm of $\mathbf{X}$, respectively. Third-order tensors are denoted by calligraphy letters. To make it convenient to represent the sub-modules, an indexing is devised for third-order tensors. Given $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathcal{X}(i, j, :)$ denotes the (i, j)-th tube; $\mathcal{X}(i, :, :)$, $\mathcal{X}(:, j, :)$, and $\mathcal{X}(:, :, k)$ represent the i-th horizontal, j-th lateral, and k-th frontal slices, respectively. $\mathcal{X}_{(j)}$ and $\mathcal{X}^{(k)}$ are interchangeable with $\mathcal{X}(:, j, :)$ and $\mathcal{X}(:, :, k)$ for the ease of notation. $\mathcal{X}_f \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is obtained by applying Discrete Fourier Transform (DFT) on $\mathcal{X}$ along its third dimension. $\mathcal{X}_f$ can be conveniently computed using the Matlab command $\mathcal{X}_f = fft(\mathcal{X},[],3)$. The inverse operation is $\mathcal{X} = ifft(\mathcal{X}_f,[],3)$. Tensor conjugate transpose of $\mathcal{X}$ denoted by $\mathcal{X}' \in \mathbb{R}^{n_2 \times n_1 \times n_3}$ is achieved by conjugate transposing each frontal slice of $\mathcal{X}$ and then reversing the order of the slices from two to n. $\mathcal{I}$ represents the identity tensor whose first frontal slice is an identity matrix, while the others are all zero matrices. If third-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_1 \times n_2}$ satisfies $\mathcal{X}' * \mathcal{X} = \mathcal{I}$ and $\mathcal{X} * \mathcal{X}' = \mathcal{I}$, then $\mathcal{X}$ is an orthogonal tensor. As for tensor norms, F-norm and l_1-norm of $\mathcal{X}$ are computed as $\|\mathcal{X}\|_F = (\sum_{i,j,k} |\mathcal{X}(i, j, k)|^2)^{\frac{1}{2}}$ and $\|\mathcal{X}\|_1 = \sum_{i,j,k} |\mathcal{X}(i, j, k)|$, respectively.

Next, we introduce the definitions associated with t-product. First, three commonly used third-order tensor operators with their notations are given as follows. For a third-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the block circulant operator is defined as

$$bcirc(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^{(1)} & \mathcal{X}^{(n_3)} & \dots & \mathcal{X}^{(2)} \\ \mathcal{X}^{(2)} & \mathcal{X}^{(1)} & \dots & \mathcal{X}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{X}^{(n_3)} & \mathcal{X}^{(n_3-1)} & \dots & \mathcal{X}^{(1)} \end{bmatrix} \in \mathbb{R}^{n_1 n_3 \times n_2 n_3}, \quad (1)$$

the block vectorizing operator and its inverse operator are

$$bvec(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^{(1)} \\ \mathcal{X}^{(2)} \\ \vdots \\ \mathcal{X}^{(n_3)} \end{bmatrix} \in \mathbb{R}^{n_1 n_3 \times n_2}, \quad bvfold(bvec(\mathcal{X})) = \mathcal{X},$$

$$\quad (2)$$
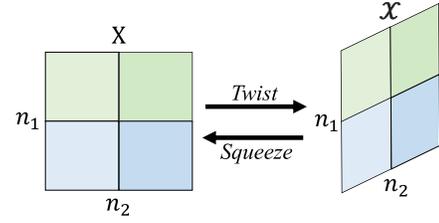

Fig. 1. *squeeze* and *twist* operators.

and the block diagonalizing operator is defined as

$$bdiag(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^{(1)} & & & \\ & \mathcal{X}^{(2)} & & \\ & & \ddots & \\ & & & \mathcal{X}^{(n_3)} \end{bmatrix} \in \mathbb{R}^{n_1 n_3 \times n_2 n_3}. \quad (3)$$

Then t-product is defined as follows.

Definition 1 (t-Product $*$ [29]): Given $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Y} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, t-product between them is defined as

$$\mathcal{Z} = \mathcal{X} * \mathcal{Y} = bvfold(bcirc(\mathcal{X})bvec(\mathcal{Y})) \in \mathbb{R}^{n_1 \times n_4 \times n_3}. \quad (4)$$

It is well-known that a block circulant matrix can be block diagonalized through Discrete Fourier Transform, thus we have the following fact:

Fact 1 (Fourier Domain Connection): For third-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the following connection exists:

$$bdiag(\mathcal{X}_f) = (F_{n_3} \otimes I_{n_1})bcirc(\mathcal{X})(F'_{n_3} \otimes I_{n_2}), \quad (5)$$

where $F_{n_3} \in \mathbb{C}^{n_3 \times n_3}$ represents the normalized Discrete Fourier Transform matrix, and $\otimes$ denotes Kronecker product.

Fact 1 establishes a connection between $\mathcal{X}$ and $\mathcal{X}_f$. Based on this fact, the following Lemma can be derived:

Lemma 1: Given two arbitrary tensors $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Y} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$. Let $\mathcal{Z} = \mathcal{X} * \mathcal{Y} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$, then

1) $\|\mathcal{Z}\|_F^2 = \frac{1}{n_3}\|bdiag(\mathcal{Z}_f)\|_F^2$;
2) $\mathcal{Z} = \mathcal{X} * \mathcal{Y} \iff bdiag(\mathcal{Z}_f) = bdiag(\mathcal{X}_f)bdiag(\mathcal{Y}_f)$.

The second clause of Lemma 1 can be further derived as

$$\mathcal{Z} = \mathcal{X} * \mathcal{Y} \iff \mathcal{Z}_f^{(k)} = \mathcal{X}_f^{(k)} \mathcal{Y}_f^{(k)}, \ k = 1, 2, \dots, n_3. \quad (6)$$

In most cases, due to the complex form of t-product, real-valued objective functions of third-order tensor variables can not be optimized directly. Fortunately, the original optimization problem in the third-order tensor domain can be transformed into an equivalent problem in the complex domain using Lemma 1, which can be solved efficiently with the help of Wirtinger derivative [35].

squeeze and *twist* operators are applied to arrange two-dimensional data samples into third-order tensors. Given a two-dimensional data sample $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$, the output of $twist(\mathbf{X})$ is a third-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times 1 \times n_2}$, while *squeeze* is its inverse operator as $squeeze(\mathcal{X}) = \mathbf{X}$. The two mutually-reversing operators are displayed in Fig. 1.

III. PROPOSED METHOD

In this section, we first introduce the proposed modules featuring different concerns. To fully utilize the inherent correlations within the image set and to avoid the computation of high-dimensional vectors, all the modules are constructed using t-product. Afterward, we formulate the Robust Discriminative t-Linear Subspace Learning model (RDtSL) based on the proposed modules. Under the framework of Alternating Direction Method of Multipliers (ADMM) [36], we develop an iterative algorithm to optimize RDtSL. Finally, the computational complexity of the proposed algorithm is analyzed.

A. Energy Preservation Module

A subspace learning method that excels in joint image feature extraction and dimensionality reduction should be capable of preserving the primary energy of samples to ensure that the discriminative information is effectively retained in the learned representations, even when the target dimension is low. To this end, we propose an energy preservation module.

Denote $\{\mathbf{X}_i \in \mathbb{R}^{m \times n}\}_{i=1}^h$ as the centralized image set. After twisting each image sample, we stack them along the second dimension to construct the data tensor as $\mathcal{X} \in \mathbb{R}^{m \times h \times n}$. Motivated by the energy preservation property of PCA, we first propose the t-product-based PCA model, which aims to maximize the post-projection variance of samples as:

$$\max_{\mathcal{R}} \sum_{i=1}^h \|\mathcal{R}' * \mathcal{X}_{(i)}\|_F^2 \qquad \text{s.t. } \mathcal{R}' * \mathcal{R} = \mathcal{I}, \qquad (7)$$

where $\mathcal{R} \in \mathbb{R}^{m \times d \times n}$ is the projection tensor, and d is the subspace dimension. Eq. (7) preserves the primary energy of samples by retaining the variance after the tensorial projection. However, the orthogonal constraint imposed on the projection tensor makes it difficult to incorporate other regularizations. To facilitate the inclusion of additional modules, we further propose a variant of Eq. (7) as follows:

$$\min_{\mathcal{P}, \mathcal{Q}} \|\mathcal{X} - \mathcal{P} * \mathcal{Q}' * \mathcal{X}\|_F^2 + \lambda \|\mathcal{Q}\|_F^2$$
$$\text{s.t. } \mathcal{P}' * \mathcal{P} = \mathcal{I}, \qquad (8)$$

where $\mathcal{Q} \in \mathbb{R}^{m \times d \times n}$ and $\mathcal{P} \in \mathbb{R}^{m \times d \times n}$ denote the projection tensor and reconstruction tensor, respectively. Eq. (8) preserves the primary energy of samples by minimizing the t-product-based reconstruction loss. By relaxing the orthogonal constraint imposed on $\mathcal{Q}$, we can easily introduce additional regularizations and modules. The following theorem demonstrates the equivalence between Eq. (7) and Eq. (8).

Theorem 1: Denote $\mathcal{X} \in \mathbb{R}^{m \times h \times n}$ as the centralized image tensor and $\hat{\mathcal{R}}$ as the optimal solution of Eq. (7). Suppose $\hat{\mathcal{P}} \in \mathbb{R}^{m \times d \times n}$ and $\hat{\mathcal{Q}} \in \mathbb{R}^{m \times d \times n}$ satisfy

$$(\hat{\mathcal{P}}, \hat{\mathcal{Q}}) = \arg\min_{\mathcal{P}, \mathcal{Q}} \|\mathcal{X} - \mathcal{P} * \mathcal{Q}' * \mathcal{X}\|_F^2 + \lambda \|\mathcal{Q}\|_F^2$$
$$\text{s.t. } \mathcal{P}' * \mathcal{P} = \mathcal{I}. \qquad (9)$$

Then, $\hat{\mathcal{R}}_{(j)}$ is proportional to $\hat{\mathcal{Q}}_{(j)}$ in t-linear space. In other words, $\hat{\mathcal{Q}}_{(j)} = \hat{\mathcal{R}}_{(j)} * \mathcal{V}_j$, where $\mathcal{V}_j \in \mathbb{R}^{1 \times 1 \times h}$ is a tubal-scalar, and $\hat{\mathcal{R}}_{(j)}$ can be obtained by normalizing $\hat{\mathcal{Q}}_{(j)}$.

The proof of Theorem 1 is provided in the Appendix. The equivalence between Eq. (7) and Eq. (8) leads us to conclude that t-product-based PCA can be defined as both maximizing post-projection variance and minimizing reconstruction loss. Besides, the projection and reconstruction can be achieved using two different projection bases. Eq. (8) effectively preserves the primary energy of samples within the learned representations and facilitates the straightforward incorporation of additional regularizations and modules.

B. Subspace Representations Classification Module

While Eq. (8) effectively preserves the primary energy of the samples, its main emphasis is on the statistical features of the samples, and the learned representations are independent of the class labels. Obviously, it is crucial to retain the distinctive features that differentiate each class within the subspace representations. To address this limitation, we propose the t-product-based subspace representations classification module, which creates a seamless link between the subspace representations and their respective labels, thus enabling the extraction of distinctive features across the classes. During the training phase, subspace representations $\mathcal{Q}' * \mathcal{X}$ are fed into the t-product-based classifier function $f(input, \mathcal{T})$ to optimize the parameter tensor $\mathcal{T} \in \mathbb{R}^{C \times d \times n}$, where C is the number of classes. $\mathcal{Q}' * \mathcal{X}$ is tightly coupled with the labels by minimizing the classification error as:

$$\min_{\mathcal{T}, \mathcal{Q}} \sum_{j=1}^h \mathcal{L}(\mathcal{Y}_{(j)}, f(\mathcal{Q}' * \mathcal{X}_{(j)}, \mathcal{T})), \qquad (10)$$

where $\mathcal{X}_{(j)}$ is the j-th twisted image sample. For clarification, 1) the classifier function is based purely on t-product, which is defined as $f(input, \mathcal{T}) = \mathcal{T} * input$. 2) Assume the label of the j-th sample is C_j, $\mathcal{Y}_{(j)}$ is constructed by first initializing it with all zeros, and then setting the first frontal slice of its C_j-th tube to 1. That is to say, set $\mathcal{Y}(C_j, j, 1)$ to 1. With this configuration, the equivalent problem of Eq. (10) in the Fourier domain is a standard matrix-based regression model, which maximizes the discriminative ability in our experiments. 3) $\mathcal{L} = \|\cdot\|_F^2$ measures the empirical classification error. Based on the above definitions, Eq. (10) is equivalent to

$$\min_{\mathcal{T}, \mathcal{Q}} \|\mathcal{Y} - \mathcal{T} * \mathcal{Q}' * \mathcal{X}\|_F^2. \qquad (11)$$

C. Class-Oriented Graph Embedding Module

Both similarity and locality information of samples have significantly influence on learning the latent pattern of data distributions [3], [4], which are not taken into account by Eq. (8) and Eq. (10). To address it, we propose the t-product-based class-oriented graph embedding module, designed to capture the affinities between image samples and consequently further enrich the learned subspace representations, as follows:

$$\min_{\mathcal{Q}, r} \sum_{i=1}^C N_i \sum_{l \neq j}^{N_i} (r_{j,l}^i)^2 \|\mathcal{Q}' * \mathcal{X}_{(j)}^i - \mathcal{Q}' * \mathcal{X}_{(l)}^i\|_F^2$$

$$\text{s.t. } \sum_{\substack{l=1; \\ l \neq j}}^{N_i} r_{j,l}^i = 1, \, r_{j,l}^i \geq 0, \quad \mathcal{Q}' * \mathcal{Q} = \mathcal{I}, \qquad (12)$$

where $\mathcal{X}^i \in \mathbb{R}^{m \times N_i \times n}$ is composed of the twisted samples from class i, N_i is the number of samples in class i, and $r_{j,l}^i$ is the learned similarity of the j-th and l-th samples from class i. The graph embedding module as formulated in Eq. (12) offers the following advantages: 1) Compared to global affinity graphs without additional informative regularization, class-oriented graphs not only capture the locality and similarity information of samples but also enhance the discriminative ability by increasing intra-class compactness. A global graph embedding module without the help of class labels might place two samples from different classes close in the learned subspace. 2) In contrast to pre-computed graphs with fixed weights, learning and embedding the graph adaptively makes the module more robust and less susceptible to noise.

D. Robust Discriminative t-Linear Subspace Learning

Based on the proposed modules, RDtSL is formulated as:

$$
\min_{\mathcal{P},\mathcal{Q},\mathcal{E},\mathcal{T},r} \|\mathcal{E}\|_1 + \overbrace{\lambda_1 \|\mathcal{Y} - \mathcal{T}*\mathcal{Q}'*\mathcal{X}\|_F^2}^{\text{Subspace Representations Classification}}
$$

$$
+ \underbrace{\lambda_2 \sum_{i=1}^{C} N_i \sum_{l \neq j}^{N_i} (r_{j,l}^i)^2 \|\mathcal{Q}'*\mathcal{X}_{(j)}^i - \mathcal{Q}'*\mathcal{X}_{(l)}^i\|_F^2}_{\text{Class-Oriented Graph Learning and Embedding}}
$$

$$
\text{s.t.} \quad \underbrace{\mathcal{X} = \mathcal{P}*\mathcal{Q}'*\mathcal{X} + \mathcal{E}}_{\text{Data Energy Preservation}}, \quad \mathcal{P}'*\mathcal{P} = \mathcal{I},
$$

$$
\sum_{l=1,l\neq j}^{N_i} r_{j,l}^i = 1, r_{j,l}^i \geq 0, \tag{13}
$$

where $\mathcal{X}$, $\mathcal{P}$, $\mathcal{Q}$, $\mathcal{Y}$, $\mathcal{T}$, r, C, and N have the same definitions as in Eqs. (8), (11) and (12); λ_1 and λ_2 serve as the trade-off parameters to balance the weights of the three modules. Tensor $\mathcal{E}$ denotes the reconstruction error to fit the noise during the learning phase. Notably, the energy preservation module in Eq. (13) is slightly different from Eq. (8). The reason for this modification is that squared F-norm is sensitive to noise, which could potentially compromise the effectiveness in preserving data energy. To enhance the robustness of RDtSL, we reformulate the t-product-based energy preservation module as demonstrated in Eq. (13). In this version, we utilize tensor l_1-norm to robustly measure the reconstruction loss.

For the ease of understanding, the primary structure of RDtSL is illustrated in Fig. 2. As shown in the figure, RDtSL connects the t-linear subspace representations $\mathcal{Q}'*\mathcal{X}$ with the label tensor $\mathcal{Y}$ and the denoised image tensor ($\mathcal{X} - \mathcal{E}$) using two learnable projection tensors, $\mathcal{T}$ and $\mathcal{P}$. Therefore, the primary energy and class-distinctive features will be effectively preserved within the subspace representations to minimize the classification and reconstruction loss. The graph embedding module learns and embeds the affinity graph of image samples within each class in an adaptive manner. As a conclusion, the mutual balance maintained between the three proposed modules helps RDtSL preserve more valuable information in the learned representations and reach the optimal solution.

E. Optimization of RDtSL

In this section, under the framework of ADMM [36], we propose an iterative algorithm to solve RDtSL. The augmented Lagrangian function of Eq. (13) is presented as

$$
\mathcal{L}(\mathcal{P},\mathcal{Q},\mathcal{E},\mathcal{T},r) = \|\mathcal{E}\|_1 + \lambda_1 \|\mathcal{Y} - \mathcal{T}*\mathcal{Q}'*\mathcal{X}\|_F^2
$$

$$
+ \lambda_2 \sum_{i=1}^{C} N_i \sum_{l \neq j}^{N_i} (r_{j,l}^i)^2 \|\mathcal{Q}'*\mathcal{X}_{(j)}^i - \mathcal{Q}'*\mathcal{X}_{(l)}^i\|_F^2
$$

$$
+ \frac{\rho}{2} \|\mathcal{X} - \mathcal{P}*\mathcal{Q}'*\mathcal{X} - \mathcal{E} + \frac{\mathcal{H}}{\rho}\|_F^2,
$$

$$
\text{s.t.} \quad \sum_{l=1,l\neq j}^{N_i} r_{j,l}^i = 1, r_{j,l}^i \geq 0, \quad \mathcal{P}'*\mathcal{P} = \mathcal{I}, \tag{14}
$$

where $\rho > 0$ is the penalty parameter, and $\mathcal{H} \in \mathbb{R}^{m \times h \times n}$ is the Lagrange multiplier. The variables will be alternatively updated to optimize Eq. (14) until convergence as follows:

1) Subproblem-$\mathcal{T}$: Focusing on $\mathcal{T}$, Eq. (14) becomes

$$
\min_{\mathcal{T}} \|\mathcal{Y} - \mathcal{T}*\mathcal{Q}'*\mathcal{X}\|_F^2. \tag{15}
$$

According to Lemma 1, Eq. (15) has an equivalent form in the Fourier domain as

$$
\min_{\mathcal{T}_f} \sum_{k=1}^{n} \|\mathcal{Y}_f^{(k)} - \mathcal{T}_f^{(k)} \mathcal{Q}_f^{(k)'} \mathcal{X}_f^{(k)}\|_F^2. \tag{16}
$$

Eq. (16) is a real-valued function of complex variables. Let the Wirtinger derivative [35] of Eq. (16) with respect to $\mathcal{T}_f^{(k)}$ be zero, then the optimal $\mathcal{T}_f^{(k)}$ is computed as

$$
\mathcal{T}_f^{(k)} = (\mathcal{Y}_f^{(k)} \mathcal{X}_f^{(k)'} \mathcal{Q}_f^{(k)})(\mathcal{Q}_f^{(k)'} \mathcal{X}_f^{(k)} \mathcal{X}_f^{(k)'} \mathcal{Q}_f^{(k)})^{-1},
$$

$$
k = 1, 2, \cdots, n. \tag{17}
$$

$\mathcal{T}$ can be recovered by inverse DFT as $\mathcal{T} = ifft(\mathcal{T}_f, [], 3)$.

2) Subproblem-$\mathcal{P}$: Focusing on the parts related to $\mathcal{P}$, Eq. (14) is reduced to

$$
\min_{\mathcal{P}} \|\mathcal{X} - \mathcal{P}*\mathcal{Q}'*\mathcal{X} - \mathcal{E} + \frac{\mathcal{H}}{\rho}\|_F^2
$$

$$
\text{s.t.} \quad \mathcal{P}'*\mathcal{P} = \mathcal{I}. \tag{18}
$$

Through calculation, we conclude that each frontal slice of $\mathcal{I}_f$ is an identity matrix. Thus, Eq. (18) is equivalent to minimize n independent subproblems, and the k-th one is shown as

$$
\min_{\mathcal{P}_f^{(k)}} \|\mathcal{X}_f^{(k)} - \mathcal{P}_f^{(k)} \mathcal{Q}_f^{(k)'} \mathcal{X}_f^{(k)} - \mathcal{E}_f^{(k)} + \frac{\mathcal{H}_f^{(k)}}{\rho}\|_F^2
$$

$$
\text{s.t.} \quad \mathcal{P}_f^{(k)'} \mathcal{P}_f^{(k)} = \mathbf{I}, \tag{19}
$$

where $\mathbf{I} \in \mathbb{R}^{d \times d}$ represents the identity matrix. Eq. (19) is a complex orthogonal Procrustes problem, which can be addressed as

$$
\mathcal{P}_f^{(k)} = \mathbf{U}\mathbf{V}', \quad k = 1, 2, \cdots, n, \tag{20}
$$

where $\mathbf{U}$ and $\mathbf{V}$ are the left-singular and right-singular vectors of $(\mathcal{X}_f^{(k)} - \mathcal{E}_f^{(k)} + \frac{\mathcal{H}_f^{(k)}}{\rho})\mathcal{X}_f^{(k)'} \mathcal{Q}_f^{(k)}$. Once $\mathcal{P}_f$ has been solved slice by slice, $\mathcal{P}$ can be recovered as $\mathcal{P} = ifft(\mathcal{P}_f, [], 3)$.
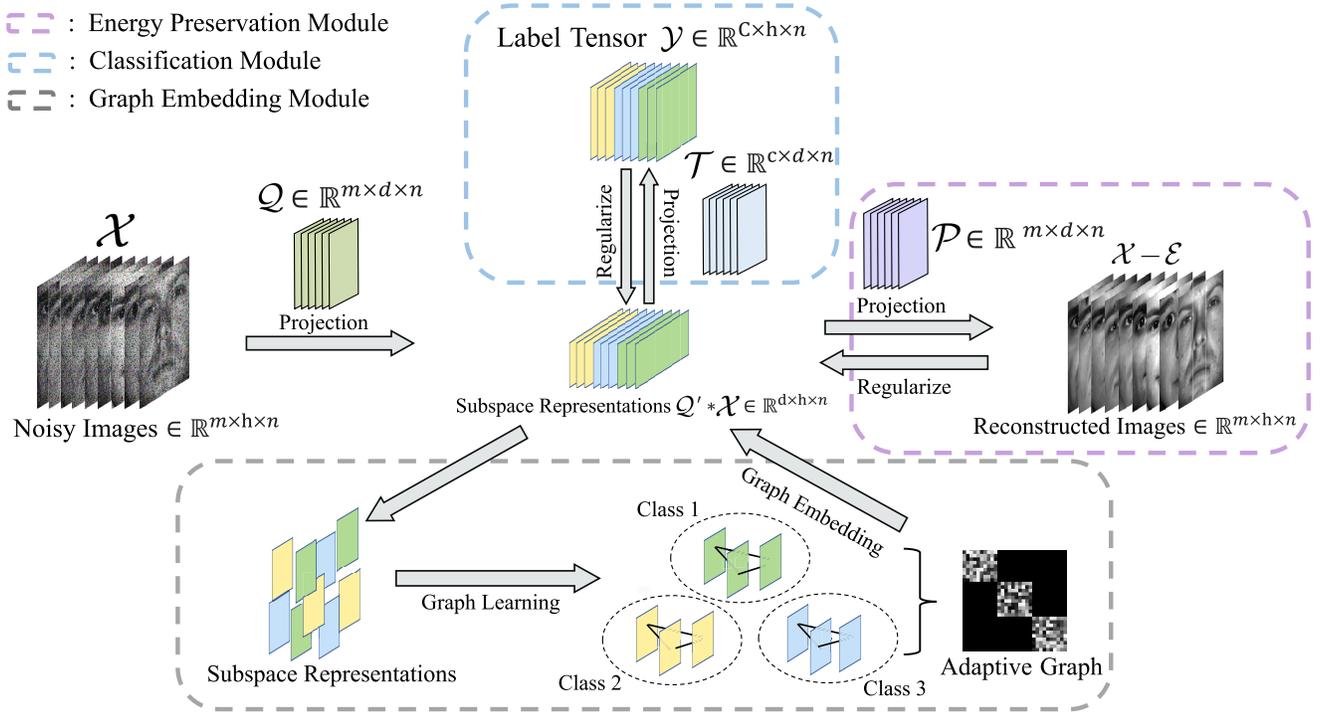
Fig. 2. Framework of the RDtSL model. Images are arranged into a tensor in advance using the *twist* operator. All the linear transformations are performed by t-product. The subspace representations $\mathcal{Q}' * \mathcal{X}$ are regularized by the proposed modules for data energy preservation, class-distinctive feature extraction, and affinity capture to search for the optimal projection tensor $\mathcal{Q}$. The mutual balance between the modules helps RDtSL reach the optimal solution.

3) Subproblem-$\mathcal{Q}$: By removing the terms that do not involve $\mathcal{Q}$, the objective function becomes

$$
\begin{aligned}
\min_{\mathcal{Q}} \quad & \lambda_1 \|\mathcal{Y} - \mathcal{T} * \mathcal{Q}' * \mathcal{X}\|_F^2 \\
& + \lambda_2 \sum_{i=1}^{C} N_i \sum_{l \neq j}^{N_i} (r_{j,l}^i)^2 \|\mathcal{Q}' * \mathcal{X}_{(j)}^i - \mathcal{Q}' * \mathcal{X}_{(l)}^i\|_F^2 \\
& + \frac{\rho}{2} \ \|\mathcal{X} - \mathcal{P} * \mathcal{Q}' * \mathcal{X} - \mathcal{E} + \frac{\mathcal{H}}{\rho}\|_F^2,
\end{aligned} \tag{21}
$$

which can also be divided into n independent subproblems. The k-th one is shown as

$$
\begin{aligned}
\min_{\mathcal{Q}_f^{(k)}} \quad & \lambda_1 \ \|\mathcal{Y}_f^{(k)} - \mathcal{T}_f^{(k)} \mathcal{Q}_f^{(k)'} \mathcal{X}_f^{(k)}\|_F^2 \\
& + \lambda_2 \ tr(\mathcal{Q}_f^{(k)'} \mathcal{M}_f^{(k)} \mathcal{Q}_f^{(k)}) \\
& + \frac{\rho}{2} \ \|\mathcal{X}_f^{(k)} - \mathcal{P}_f^{(k)} \mathcal{Q}_f^{(k)'} \mathcal{X}_f^{(k)} - \mathcal{E}_f^{(k)} + \frac{\mathcal{H}_f^{(k)}}{\rho}\|_F^2,
\end{aligned} \tag{22}
$$

where

$$
\mathcal{M}_f^{(k)} = \mathcal{X}_f^{(k)} \mathbf{L} \ \mathcal{X}_f^{(k)'}, \tag{23}
$$

where $\mathbf{L}$ is the Laplacian matrix of the adaptive affinity graph. By calculating the Wirtinger derivative with respect to $\mathcal{Q}_f^{(k)}$, and setting it to zero, we can obtain:

$$
\mathbf{A}\mathcal{Q}_f^{(k)} + \mathcal{Q}_f^{(k)}\mathbf{B} = \mathbf{C}, \tag{24}
$$

with

$$
\begin{aligned}
\mathbf{A} &= \lambda_1 \mathbf{N}^{-1}(\frac{\rho}{2}\mathbf{N} + \lambda_2 \mathcal{M}_f^{(k)}), \\
\mathbf{B} &= T_f^{(k)'} T_f^{(k)}, \\
\mathbf{C} &= \lambda_1 \mathbf{N}^{-1}(\frac{\rho}{2}\mathcal{X}_f^{(k)}\mathbf{S}'\mathcal{P}_f^{(k)} + \lambda_1 \mathcal{X}_f^{(k)}\mathcal{Y}_f^{(k)'}\mathcal{T}_f^{(k)}),
\end{aligned} \tag{25}
$$

where $\mathbf{N} = \mathcal{X}_f^{(k)} \mathcal{X}_f^{(k)'}$ and $\mathbf{S} = \mathcal{X}_f^{(k)} - \mathcal{E}_f^{(k)} + \frac{\mathcal{H}_f^{(k)}}{\rho}$. Eq. (24) is a Sylvester equation that can be solved by the Hessenberg-Schur algorithm [37]. After solving $\mathcal{Q}_f$ slice by slice, $\mathcal{Q}$ can be recovered by $\mathcal{Q} = ifft(\mathcal{Q}_f, [], 3)$.

4) Subproblem-$\mathcal{E}$: Fixing all the variables other than $\mathcal{E}$, the objective function becomes

$$
\min_{\mathcal{E}} \quad \|\mathcal{E}\|_1 + \frac{\rho}{2} \ \|\mathcal{X} - \mathcal{P} * \mathcal{Q}' * \mathcal{X} - \mathcal{E} + \frac{\mathcal{H}}{\rho}\|_F^2. \tag{26}
$$

According to [38], the closed-form solution of Eq. (26) is given as

$$
\mathcal{E} = \sigma_{\frac{1}{\rho}}[\mathcal{X} - \mathcal{P} * \mathcal{Q}' * \mathcal{X} + \frac{\mathcal{H}}{\rho}], \tag{27}
$$

where $\sigma_{\frac{1}{\rho}}[x] = \text{sign}(x)\max(|x| - (\frac{1}{\rho}), 0)$ denotes the soft-thresholding operator.

5) Subproblem-r: After removing the irrelevant terms, the objective function can be divided in to C independent

subproblems, and the i-th one is show as

$$\min_{r^i} \sum_{j,l=1,l\neq j}^{N_i} (r_{j,l}^i)^2 f_{j,l}^i$$

$$\text{s.t.} \sum_{l=1;l\neq j}^{N_i} r_{j,l}^i = 1, r_{j,l}^i \geq 0, \qquad (28)$$

where $f_{j,l}^i = \|\mathcal{Q}' * \mathcal{X}_{(j)}^i - \mathcal{Q}' * \mathcal{X}_{(l)}^i\|_F^2$. Eq. (28) can be solved using the following theorem.

Theorem 2 ([39]): Denote $\mathbf{f} = \begin{bmatrix} f_1, \cdots, f_n \end{bmatrix} \in \mathbb{R}^{1\times n}$ as a vector with all positive entries. Then the following problem

$$\min_{\mathbf{r}} \sum_{j=1}^{n} r_j^2 f_j \qquad \text{s.t.} \sum_{j=1}^{n} r_j = 1, \ r_j \geq 0, \qquad (29)$$

will reach its optimum at $r_j = \frac{1}{f_j}(\sum_o \frac{1}{f_o})^{-1}$, for j $= 1, \cdots, n$. Based on Theorem 2, the optimal $r_{j,l}^i$ is computed as

$$r_{j,l}^i = \frac{1}{f_{j,l}^i}(\sum_{o\neq j,o=1}^{N_i} \frac{1}{f_{j,o}^i})^{-1}. \qquad (30)$$

6) Multiplier and Penalty Parameter: The multiplier $\mathcal{H}$ and the penalty parameter ρ are updated as follows:

$$\mathcal{H} = \mathcal{H} + \rho\ (\mathcal{X} - \mathcal{P} * \mathcal{Q}' * \mathcal{X} - \mathcal{E}),$$
$$\rho = \min\{\mu\rho, \rho_{\max}\}, \qquad (31)$$

where μ and $\rho_{\max}$ are constants and μ is empirically set to 1.1.

When $res = \|\mathcal{X} - \mathcal{P} * \mathcal{Q}' * \mathcal{X} - \mathcal{E}\|_F/\|\mathcal{X}\|_F$ is sufficiently small, the iterative optimization algorithm will converge and stop. The entire procedure is summarized as Algorithm 1.

Algorithm 1 Algorithm for Solving RDtSL

Input: data tensor $\mathcal{X}$; label tensor $\mathcal{Y}$; parameters λ_1, λ_2
Output: projection tensor $\mathcal{Q}$

Initialisation : Initialize $\mathcal{P}$ and $\mathcal{Q}$ using the result of Eq. (7); initialize $\mathcal{T}$ and $\mathcal{E}$ to zero tensors; initialize $\rho = 0.01$, $\mu = 1.1$, $\rho_{\max} = 10^5$ and $\epsilon = 10^{-5}$.

1: **repeat**
2:     Update $\mathcal{T}$ using Eq. (17).
3:     Update $\mathcal{P}$ using Eq. (20).
4:     Update $\mathcal{Q}$ by solving Eq. (24).
5:     Update $\mathcal{E}$ using Eq. (27).
6:     Update r using Eq. (30).
7:     Update multiplier and penalty parameter using Eq. (31).
8: **until** $res \leq \epsilon$

F. Complexity Analysis

Computational complexity of basic operators is listed as follows: 1) Given tensor $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$, the computational complexity of applying (inverse) fast Fourier transform along its third dimension is $\mathcal{O}(n_1 n_2 n_3 \log(n_3))$. 2) Given $\mathcal{Y} \in \mathbb{R}^{n_2\times n_4\times n_3}$, the computational complexity of $\mathcal{X} * \mathcal{Y}$ is $\mathcal{O}(\max(n_1, n_4)n_2 n_3 \log(n_3) + n_1 n_2 n_3 n_4)$. 3) Given $\mathbf{X} \in \mathbb{C}^{n_1\times n_2}$, the Hessenberg-Schur algorithm requires the

computational complexity of $\mathcal{O}(n_1^3 + n_2^3)$ for solving an Sylvester equation. 4) The computational complexity of applying economic-sized SVD decomposition on $\mathbf{X} \in \mathbb{C}^{n_1\times n_2}$ is $\mathcal{O}(\max(n_1, n_2)\min(n_1, n_2)^2)$.

For the objective function Eq. (14), h is far greater than m, C, and n in practical applications. C, d, m, n are comparable and we assume $C > m > n > d$ in calculations to simplify the expressions. The computational complexity of each subproblem within a single iteration is computed as: 1) Subproblem-$\mathcal{T}$ consists of matrix product and matrix inverse, and the total complexity is $\mathcal{O}(hCmn + hdmn)$. 2) Economic-sized SVD, matrix product, and (inverse) FFT are essential for solving Subproblem-$\mathcal{P}$. The overall computational complexity is $\mathcal{O}(hdmn)$. 3) For updating $\mathcal{Q}$, matrix product, matrix inverse, Laplace matrix calculation, Hessenberg-Schur, and (inverse) FFT are required. The total complexity is $\mathcal{O}(hmmn + hCmn)$. 4) For Subproblem-$\mathcal{E}$, the computational cost of calculating $\mathcal{P} * \mathcal{Q}' * \mathcal{X}$ is $\mathcal{O}(mhn\log(n) + hdmn)$. Time complexity of the thresholding operator is negligible. 5) Compared with other sub-problems, the cost of updating r, the multiplier and the penalty parameter can be neglected. Above all, the complexity of a single iteration is $\mathcal{O}\ (hmn\ (m + d + C))$. Denote t as the total number of iterations, the computational complexity of the proposed iterative algorithm is $\mathcal{O}\ (thmn\ (m + d + C))$.

IV. EXPERIMENTS

In this section, we evaluate the performance of RDtSL in image feature extraction and dimensionality reduction through face, object, and handwritten digital recognition. After outlining the experimental configurations, we compare RDtSL with the state-of-the-arts in discriminative ability and robustness. Subsequently, we conduct a series of model analyses, including an ablation study, a parameter sensitivity analysis, and a convergence analysis, to comprehensively demonstrate the properties of RDtSL. All the experiments and visualizations are conducted using MATLAB R2022b.

A. Experimental Setup

1) Database: The following five databases are chosen for the model evaluation: three face databases (EYaleb,[1] ORL,[2] and UMIST[3]), an object database (COIL-100[4]), and a handwriting digital database (USPS[5]). The Extended YaleB database (EYaleb) comprises only frontal face images with minimal variations in expression and pose, with some images exhibiting changes in illumination. The ORL database is similar to EYaleb but has fewer images and more variation in facial expressions and head poses. The UMIST database offers a collection of face images with consistent expressions and poses, but captured from various angles. The COIL-100 database provides images of various kinds of objects (e.g., bottles, cups, building blocks) with a wide range of shapes. All the objects are photographed from different angles. Lastly, the

---

[1]http://vision.ucsd.edu/ iskwak/ExtYaleDatabase/ExtYaleB.html/
[2]https://www.cl.cam.ac.uk/research/dtg/attarchive/facedataset.html
[3]https://cs.nyu.edu/ roweis/data.html
[4]https://www.cs.columbia.edu/CAVE/software/softlib/
[5]http://www.cad.zju.edu.cn/home/dengcai/Data/data.html

(a) EYaleB

(b) ORL

(c) UMIST

(d) COIL-100

(e) USPS
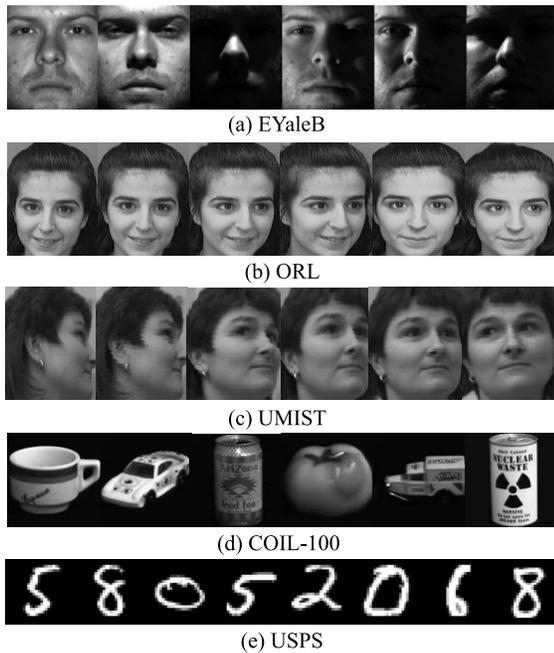
Fig. 3.   Representative samples of the databases.

TABLE I
SUMMARY OF DATABASES

Database	Category	Amount	Class	Size
Extended YaleB	face	2414	38	32×32
ORL	face	400	40	32×32
UMIST	face	575	20	28×23
COIL-100	object	7200	100	32×32
USPS	handwriting	11000	10	16×16

USPS database is a collection of handwritten digits, consisting of ten classes ranging from '0' to '9'. The selected databases are summarized in Table I, and representative samples are displayed in Fig. 3. With the exception of the USPS database, all the images are downsampled for efficiency.

2) Competing Methods: Deep Neural Networks-based methods are not adopted in this comparison since they are data-dependent and perform poorly on small-scale databases. Thirteen linear feature extraction and classification methods are selected for the evaluation. First, Kenel PCA [40] and LRE [15] are set as the baselines; RLSL [22], RSLDA [23], SALPL [27], LLRSE [24], and RALDA [28] are closely-related matrix-based feature extraction methods that contain a PCA-like data reconstruction constraint for energy preservation; The state-of-the-arts include CDPL [41], JDLSL [25], PRDR [42], DRAGD [3], and FDBRL [7]. All the competitors are supervised methods except for Kernel PCA and LRE.

3) Parameter Settings: As illustrated in Eq. (13), there are two trade-off parameters, λ_1 and λ_2, in RDtSL, which control the weights of the classification module and the graph embedding module, respectively. Since the required time of our model for single training and validation is low, the technique

grid search is well-suited for exploring the optimal trade-off parameters of RDtSL. Both parameters λ_1 and λ_2 are tuned within the range of $[10^{-3}, 10^{-2}, \cdots, 10^3]$. In addition to being related to the trade-off parameters, dimension of the target subspace (d) also has an impact on the recognition accuracy. Suppose the size of an input image is $m \times n$ and the number of classes is C. For SALPL, LLRSE, PRDR, and DRAGD, d should be fixed to C. For the remaining competitors, d is tuned within the range of $[0, \min(m \times n, 250)]$ with a step size of four. For RDtSL, d is adjusted within the range of $[2, m]$ with a step size of two. After the tuning process, each candidate is evaluated and compared using the best-performing trade-off parameters and subspace dimension.

4) Classification Configurations: The databases are partitioned into separate training and testing sets. The training set is employed to learn the projection matrix/tensor, while the testing set is used for the evaluation. Once the projection basis is computed, both training and testing samples are projected onto the learned subspace. We then employ the 1-Nearest Neighbor Classifier (1-NN) to classify the testing samples. Classification accuracy is adopted as the metric for evaluation and comparison. The process is repeated ten times to ensure statistical stability, each with a different random seed. Finally, we calculate and report the average classification accuracy, along with the standard deviation.

B. Performance Comparison

To evaluate the competing methods, we perform face, object, and handwritten digit recognition. Overall, RDtSL demonstrates superior discriminative ability when facing clean image samples, and exhibits strong robustness when handling images that contain common corruptions.

1) Face Recognition: There are three face databases. For the EYaleb database, ten images per class are randomly selected to form the training set, while the remaining samples are used for testing. For ORL, the database is randomly and evenly divided into training and testing sets, each containing five samples per class. UMIST has varying sample sizes for each class. Similar to EYaleb, ten images per class are randomly chosen to build the training set. We can notice that RDtSL outperforms all its competitors across the three face databases, and it achieves a 2.52% higher classification accuracy on the EYaleb database than the second-place method, far surpassing all other competing methods. PRDR and DRAGD alternate as the second-best performers among the face databases, and both are linear regression-based image classification methods. DRAGD focuses on learning and embedding flexible graphs to capture the local geometric structures of samples. PRDR attempts to learn latent representations for labels to mitigate the negative impacts of strict binary labels. These two classification method provide potential directions for further improving the t-linear subspace learning model.

2) Object Recognition: The COIL-100 database comprises a total of 7,200 samples. Ten samples are randomly selected from each class to construct the training set. Notably, the testing set is substantially larger, containing 6,200 samples. RDtSL achieves the highest classification accuracy among the competing methods, with DRAGD ranking second. In contrast

TABLE II

MEAN CLASSIFICATION ACCURACIES ± STANDARD DEVIATIONS (%) OF THE COMPETING METHODS ON DIFFERENT DATABASES

Method \ Database	EYaleB	ORL	UMIST	COIL-100	USPS
KPCA	50.19 ± 0.91	90.50 ± 1.97	92.10 ± 2.51	81.48 ± 0.68	69.70 ± 0.68
LRE	83.50 ± 4.04	92.20 ± 1.54	93.17 ± 1.46	82.10 ± 1.21	78.06 ± 0.97
RLSL	78.80 ± 2.00	91.25 ± 2.11	93.15 ± 1.76	77.42 ± 0.81	66.72 ± 1.33
RSLDA	84.65 ± 1.13	93.75 ± 1.88	96.61 ± 1.44	83.19 ± 0.97	82.81 ± 1.65
SALPL	88.39 ± 1.12	95.20 ± 1.80	95.28 ± 0.96	75.55 ± 0.92	83.38 ± 1.54
LLRSE	88.64 ± 0.86	93.05 ± 1.89	94.72 ± 1.80	75.98 ± 0.45	83.64 ± 0.45
RALDA	86.03 ± 1.62	94.05 ± 1.47	91.68 ± 3.42	84.09 ± 0.83	81.39 ± 1.49
CDPL	85.45 ± 0.85	94.20 ± 1.73	96.16 ± 1.67	75.78 ± 0.98	84.57 ± 0.43
JDLSL	88.91 ± 0.66	95.15 ± 1.27	96.88 ± 2.01	82.50 ± 1.31	84.07 ± 1.01
PRDR	89.62 ± 1.14	<u>96.55 ± 1.19</u>	97.22 ± 1.52	86.74 ± 1.33	<u>85.88 ± 1.23</u>
DRAGD	<u>91.14 ± 0.98</u>	96.35 ± 0.79	<u>97.63 ± 1.21</u>	<u>87.14 ± 0.92</u>	85.69 ± 0.52
FDBRL	89.35 ± 0.91	95.85 ± 1.47	97.60 ± 0.91	85.90 ± 0.87	84.98 ± 0.50
RDtSL	**93.66 ± 0.79**	**97.70 ± 0.87**	**98.85 ± 0.83**	**89.21 ± 0.71**	**87.12 ± 0.72**

Bold font indicates the highest value; <u>Underlined</u> font represents the second-best performance.
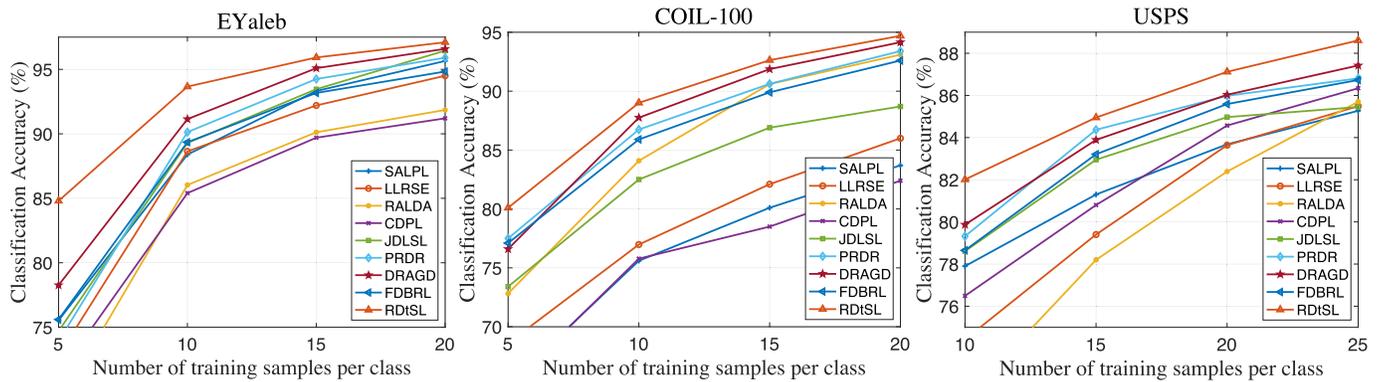

Fig. 4. Performance of the competing methods with different training sizes on the face, object, and handwriting databases.

to the face databases, objects in the COIL-100 database have a wide range of shapes and are captured from different angles, so the images are not aligned by pixels. RDtSL utilizes t-product to mitigate the negative impact of complex geometric structures on its discriminative ability.

3) Handwritten Digit Recognition: The USPS handwritten digit database is randomly divided into training and testing sets, with twenty image samples per class allocated for training and the remainder for testing. RDtSL maintains its top ranking among the competing methods, followed by PRDR. Combined with the two experiments discussed above, we can conclude that RDtSL is the best performer among all the selected databases, showing exceptional discriminative and generalization capabilities. These strengths stem from the exploitation of t-product-based operations and the harmonious balance maintained between the proposed modules.

4) Image Classification With Varying Training Sizes: To assess the robustness of the competing methods to variations in training sizes, we train the models using different numbers of samples and compare their performance. The number of training samples for each class of EYaleb, COIL-100, and USPS are chosen from the sets {5, 10, 15, 20}, {5, 10, 15, 20}, and {10, 15, 20, 25}, respectively. The experimental results

are presented in Fig. 4. To enhance the clarity of the line charts, we omit the four earlier methods with relatively poor performance. Key observations are summarized as follows: 1) RDtSL exhibits superior performance with a limited number of training samples, which underscores the advantages of employing t-product in scenarios with small training sizes. 2) RDtSL consistently secures the top position across the face, object, and handwriting digit databases, irrespective of the number of samples used for training. This phenomenon illustrates the robustness of RDtSL to variations in training sizes. 3) DRAGD, PRDR, and FDBRL also exhibit resilience to changes in training sizes. The three methods demonstrate comparable performance with RDtSL when a sufficient amount of training samples is available.

5) Contaminated Image Classification: In this experiment, we introduce Gaussian noise and block occlusion separately to the EYaleb database to evaluate the robustness of competing methods to different types of synthetic corruption. Half of the training samples are intentionally contaminated beforehand. In order to generate image corruptions at different levels of severity, we vary the variance of Gaussian noise across four distinct levels and adjust the size of block occlusion within the set {4 × 4, 8 × 8, 12 × 12, 16 × 16}. The experimental results
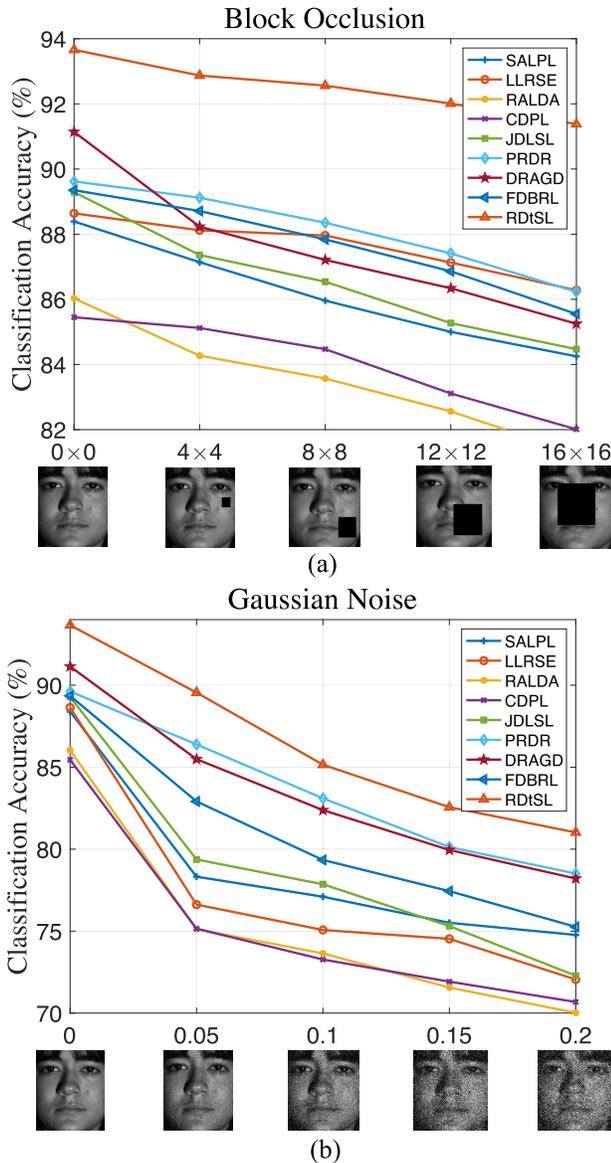
Fig. 5. Robustness of the competing methods to varying degrees of image corruptions. The x-coordinates of the two subfigures represent the area of block occlusion and the variance of Gaussian noise, respectively.

TABLE III
ABLATION STUDY ON CLASSIFICATION MODULE (CLS)
AND GRAPH EMBEDDING MODULE (GE)

CLS	GE	EYaleB	COIL-100	USPS
✓	✗	69.52 (-24.14)	84.11 (-5.10)	81.69 (-5.43)
✗	✓	83.17 (-10.49)	87.96 (-1.25)	85.52 (-1.60)
✓	✓	**93.66**	**89.21**	**87.12**

TABLE IV
ABLATION STUDY ON ENERGY PRESERVATION
MODULE (EP) WITH COIL-100

EP	$dim = 1$	$dim = 4$	$dim = 7$	$dim = 10$
✗	82.17	87.61	85.48	83.82
✓	85.73	89.61	86.23	84.11
Gain	3.56	2.00	0.75	0.29

dim represents subspace dimensionality.

modules in RDtSL. We employ the EYaleB, COIL-100, and USPS databases for this study and divide them into training and testing sets using the same configuration described in Sections IV-B.1, IV-B.2 and IV-B.3. First, we individually deactivate the classification module and the graph embedding module by setting their corresponding trade-off parameters to zero. Subsequently, we compare their performance with RDtSL. The experimental results are presented in Table III. It is evident that the performance of RDtSL significantly degrades when any of the two modules is removed, and the graph embedding module has a more pronounced impact on classification accuracy. The experimental results demonstrate that the two modules work collaboratively to achieve the optimal solution.

To illustrate the beneficial effects of the energy preservation module, we begin by removing it from RDtSL, which involves eliminating the data reconstruction constraint, the orthogonal constraint, and the term $\|\mathcal{E}\|_1$. These operations result in RDtSL being reformulated as a graph-regularized regression model. Subsequently, we compare the resulting model with RDtSL using the COIL-100 and USPS databases. To more convincingly demonstrate the effectiveness of this module, we compare their performance across different subspace dimensions. The experimental results are presented in Table IV and Table V. As evident from our observations, RDtSL consistently outperforms the graph-regularized regression model across all evaluated dimensions, demonstrating remarkable stability to variations in subspace dimension. As the dimensionality decreases, the gap in classification accuracy between the two models widens, affirming that the energy preservation module effectively preserves the discriminative information of samples in the learned subspace, even when the target dimension is low. As a conclusion, the energy preservation module equips RDtSL with an effective capability of carrying out image dimensionality reduction.

are presented in Fig. 5. Key observations are summarized as follows: 1) RDtSL demonstrates exceptional robustness against synthetic corruptions, particularly block occlusions, consistently outperforming all its competitors. 2) PRDR also exhibits high resilience to different image corruptions, maintaining stable performance even when the images are heavily contaminated. 3) While LLRSE may not be as competitive in terms of discriminative ability compared to the state-of-the-arts, it shows excellent resilience to block occlusion, highlighting the strength of low-rank self-representation learning.

C. Model Analysis

In this section, we perform a series of model analyses from different perspectives to comprehensively demonstrate the properties of RDtSL.

1) Ablation Study: The ablation study is performed to demonstrate the beneficial effects of the three proposed

TABLE V
ABLATION STUDY ON ENERGY PRESERVATION
MODULE (EP) WITH USPS

EP	$dim = 1$	$dim = 2$	$dim = 3$	$dim = 4$
✗	51.52	73.32	82.41	86.47
✓	54.40	75.08	82.82	86.68
Gain	2.88	1.76	0.41	0.21

dim represents subspace dimensionality.
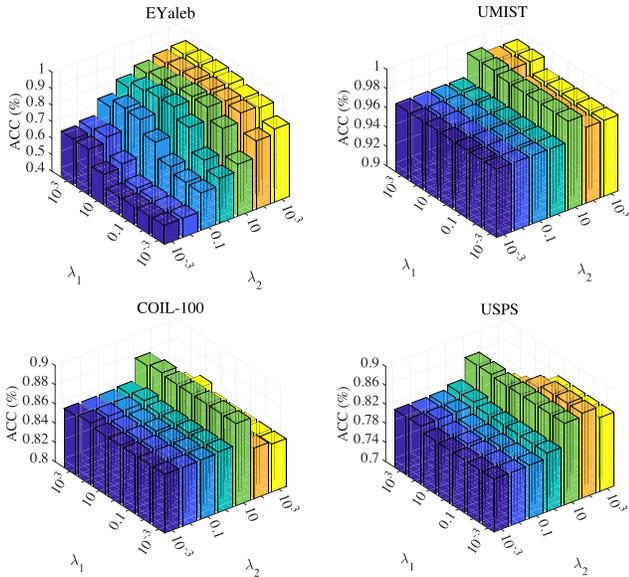
Fig. 6. Performance of RDtSL with varying trade-off parameters on face, object and handwriting databases.

Fig. 7. Classification accuracy with varying subspace dimensions.

2) Parameter Sensitive Analysis: There are two trade-off parameters λ_1 and λ_2 in RDtSL to control the weights of the classification and the graph embedding modules. Apart from being related to the parameters, subspace dimension d also impacts the recognition accuracy. In this study, we evaluate the sensitivity of RDtSL to both of these factors.

a) Trade-Off Parameters: The two trade-off parameters λ_1 and λ_2 are tuned within the range of $[10^{-3}, 10^{-2}, \cdots, 10^3]$. To evaluate the sensitivity of RDtSL to parameter variations, comprehensive experiments are conducted and the results are demonstrated in Fig. 6. This figure visualizes the impact of changes in the values of λ_1 and λ_2 on the performance of RDtSL. All the databases are divided using the same criteria outlined in Sections IV-B.1, IV-B.2 and IV-B.3. We can observe:

- With the exception of the EYaleB database, the performance of RDtSL exhibits minimal variations as the trade-off parameters change. The classification accuracy is maintained within a high range across the whole parameter range, suggesting that RDtSL is generally insensitive to variations in the trade-off parameters in most scenarios.
- As depicted in the upper left panel of Fig. 6, the performance of RDtSL is sensitive to variations in the trade-off parameters on the EYaleb database. When both trade-off parameters are small, the recognition rate drops to
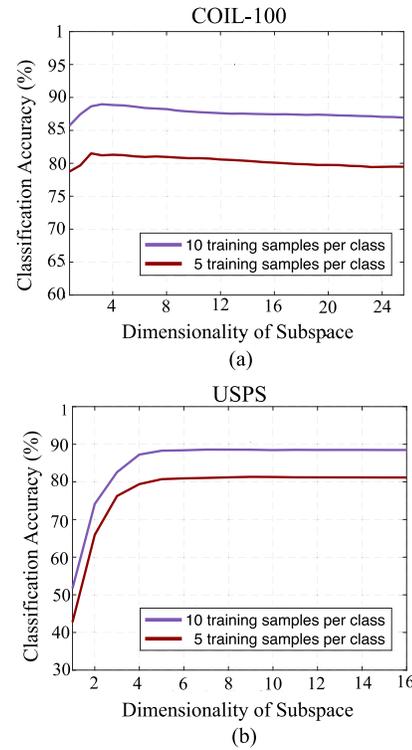
51.42%, and it experiences a rapid improvement as the trade-off parameters rise, which is entirely different from the scenarios on the other three databases. Compared to them, the EYaleb database exhibits significant variations in illumination. The huge illumination variation affects the ability of our energy preservation module to capture the discriminant information of samples, as it also tends to preserve the illumination variance. When both trade-off parameters are small, the performance of RDtSL is compromised, as it heavily relies on the energy preservation module. As the trade-off parameters increase, the performance improves rapidly due to the increased weights assigned to the other two modules. Therefore, RDtSL exhibits a significant difference in parameter sensitivity on the EYaleB database compared to the others.

- For UMIST, COIL-100, and USPS, the classification accuracy is mainly affected by λ_2, demonstrating the effectiveness of the graph embedding module. Besides, as λ_1 increases, the classification accuracy increases slowly, while it grows fast as λ_1 rises on the EYaleb database. This phenomenon arises from the fact that the EYaleb database exhibits a significant intra-class variance. Images with different illumination may display lower similarity, even if they belong to the same class. The classification module is designed to capture class-distinctive features in the learned representation. It focuses on extracting the features that differentiate one class from the others and is not concerned with similarity information of the data samples. Therefore, this module can mitigate the negative impact of high intra-class variance on the feature extraction. This explains the rapid improvement
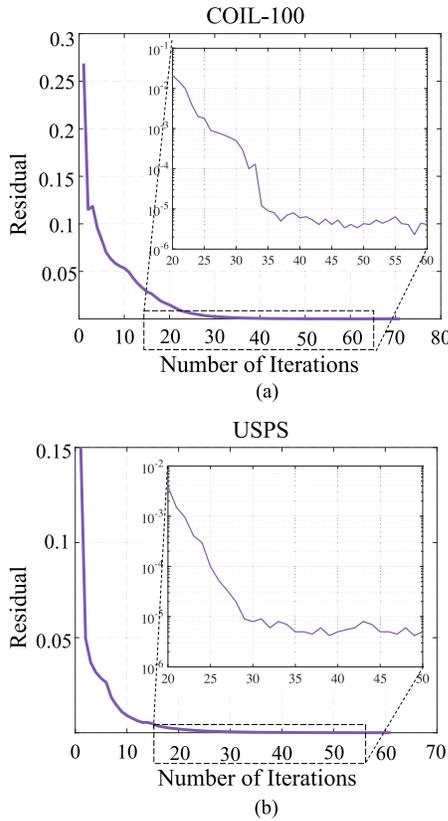
Fig. 8.   Empirical convergence curves of RDtSL.

in performance as the trade-off parameter λ_1 increases on the EYaleB database.

b) Subspace Dimension: In the ablation study, we empirically demonstrate that RDtSL effectively preserves the primary energy of samples within the learned subspace and maintains good performance even when the subspace dimension is low. To show in more detail the sensitivity of RDtSL to variations in subspace dimensions, we evaluate its performance across all potential subspace dimensions on the COIL-100 and USPS databases and record the trends. The experimental results are illustrated in Fig. 7. For COIL-100, the classification accuracy remains satisfactory even when the dimensionality is reduced to one. Regardless of the number of image samples used for training, the classification accuracy is maintained within a high range with a tendency of initially growing and then falling. For the USPS database, the classification accuracy stabilizes and remains at a high level when the target dimensionality reaches four. These findings demonstrate that the performance of RDtSL is not significantly affected by variations in subspace dimensions, attesting to its effectiveness and robustness in image dimensionality reduction. In addition, this analysis further highlight the positive impact of the proposed energy preservation module.

3) Convergence Analysis: We empirically analyze the convergence behavior of RDtSL on the COIL-100 and USPS databases. Leveraging a flexible penalty scheme to adaptively adjust the penalty parameter ρ under the ADMM framework, we can anticipate a rapid convergence rate. Fig. 8 presents the plotted residual curves, which indicate the convergence progress. It can be observed that the residuals rapidly descend to below 10 to the power of negative five and stabilize

after 30 iterations, demonstrating that RDtSL exhibits a fast convergence property in real-world scenarios.

V. CONCLUSION

This paper introduces RDtSL, a novel subspace learning method designed for joint image feature extraction and dimensionality reduction. The exceptional performance of RDtSL arises from its capability to preserve the primary energy of samples, extract class-distinctive features, and capture both locality and similarity information. Moreover, its effectiveness is further heightened by the utilization of t-product, which enables RDtSL to comprehensively exploit correlations within the image set during the learning phase.

The development of RDtSL extends the application of this tensor linear algebraic system in feature extraction and dimensionality reduction. Furthermore, each proposed module presents an independent research avenue with the potential for further investigation. For instance, in order to enhance the interpretability of the energy preservation module and further improve its effectiveness, it is highly desirable to introduce the sparse constraint on the tensor projection basis, thereby achieving the objective of feature selection. Moreover, further enhancements can also be made to our classification module through the design of appropriate label representations. This is particularly important considering that hard labels can be restrictive for regression models, potentially resulting in overfitting and diminished performance. Our future research will focus on investigating these topics.

REFERENCES

[1] W. Wang, C. Han, T. Zhou, and D. Liu, "Visual recognition with deep nearest centroids," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2023, pp. 1–30.
[2] Z. Qin, C. Han, Q. Wang, X. Nie, Y. Yin, and X. Lu, "Unified 3D segmenter as prototypical classifiers," in *Proc. 37th Conf. Neural Inf. Process. Syst.*, 2023, pp. 46419–46432.
[3] J. Wen et al., "Discriminative regression with adaptive graph diffusion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 1797–1809, Feb. 2024.
[4] L. Hu, Z. Dai, L. Tian, and W. Zhang, "Class-oriented self-learning graph embedding for image compact representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 1, pp. 74–87, Jan. 2023.
[5] X. Xiao, Y. Chen, Y.-J. Gong, and Y. Zhou, "2D quaternion sparse discriminant analysis," *IEEE Trans. Image Process.*, vol. 29, pp. 2271–2286, 2020.
[6] T. Su, D. Feng, M. Wang, and M. Chen, "Dual discriminative low-rank projection learning for robust image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 12, pp. 7708–7722, Dec. 2023.
[7] Y. Wang, Z. Li, F. Li, Y. Mi, and J. Yue, "Fuzzy discriminative block representation learning for image feature extraction," *IEEE Trans. Image Process.*, vol. 31, pp. 4994–5008, 2022.
[8] J. Yao, D. Hong, H. Wang, H. Liu, and J. Chanussot, "UCSL: Toward unsupervised common subspace learning for cross-modal image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5514212.
[9] W. Yin, Z. Ma, and Q. Liu, "Discriminative subspace learning via optimization on Riemannian manifold," *Pattern Recognit.*, vol. 139, Jul. 2023, Art. no. 109450.
[10] V. P. Kshirsagar, M. R. Baviskar, and M. E. Gaikwad, "Face recognition using eigenfaces," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* Maui, HI, USA: IEEE Computer Society, Jan. 1991, pp. 302–306.
[11] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 16, 2003, pp. 1–8.
[12] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, Oct. 2005, pp. 1208–1213.

[13] Y. Lu, Z. Lai, Y. Xu, X. Li, D. Zhang, and C. Yuan, "Low-rank preserving projections," *IEEE Trans. Cybern.*, vol. 46, no. 8, pp. 1900–1913, Aug. 2016.

[14] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.

[15] W. K. Wong, Z. Lai, J. Wen, X. Fang, and Y. Lu, "Low-rank embedding for robust image feature extraction," *IEEE Trans. Image Process.*, vol. 26, pp. 2905–2917, 2017.

[16] D. Tao, X. Li, W. Hu, S. Maybank, and X. Wu, "Supervised tensor learning," in *Proc. 5th IEEE Int. Conf. Data Mining (ICDM05)*, Nov. 2005, pp. 1–8.

[17] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.

[18] M. Li and B. Yuan, "2D-LDA: A statistical linear discriminant analysis for image matrix," *Pattern Recognit. Lett.*, vol. 26, no. 5, pp. 527–532, Apr. 2005.

[19] S. Chen, H. Zhao, M. Kong, and B. Luo, "2D-LPP: A two-dimensional extension of locality preserving projections," *Neurocomputing*, vol. 70, nos. 4–6, pp. 912–921, Jan. 2007.

[20] Z. Lai, Y. Xu, Q. Chen, J. Yang, and D. Zhang, "Multilinear sparse principal component analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 10, pp. 1942–1950, Oct. 2014.

[21] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H.-J. Zhang, "Multilinear discriminant analysis for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 212–220, Jan. 2007.

[22] X. Fang, S. Teng, Z. Lai, Z. He, S. Xie, and W. K. Wong, "Robust latent subspace learning for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2502–2515, Jun. 2018.

[23] J. Wen et al., "Robust sparse linear discriminant analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 2, pp. 390–403, Feb. 2019.

[24] Z. Ren, Q. Sun, B. Wu, X. Zhang, and W. Yan, "Learning latent low-rank and sparse embedding for robust image feature extraction," *IEEE Trans. Image Process.*, vol. 29, pp. 2094–2107, 2020.

[25] J. Zhou, B. Zhang, S. Zeng, and Q. Lai, "Joint discriminative latent subspace learning for image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 7, pp. 4653–4666, Jul. 2022.

[26] X. Zhang, Z. Tan, H. Sun, Z. Wang, and M. Qin, "Orthogonal low-rank projection learning for robust image feature extraction," *IEEE Trans. Multimedia*, vol. 24, pp. 3882–3895, 2022.

[27] X. Fang et al., "Approximate low-rank projection learning for feature extraction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5228–5241, Nov. 2018.

[28] J. Guo, Y. Sun, J. Gao, Y. Hu, and B. Yin, "Robust adaptive linear discriminant analysis with bidirectional reconstruction constraint," *ACM Trans. Knowl. Discovery Data*, vol. 14, no. 6, pp. 1–20, Dec. 2020.

[29] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra Appl.*, vol. 435, no. 3, pp. 641–658, Aug. 2011.

[30] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, 2013.

[31] P. Zhou, C. Lu, J. Feng, Z. Lin, and S. Yan, "Tensor low-rank representation for data recovery and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 5, pp. 1718–1732, May 2021.

[32] H. Wang, J. Peng, W. Qin, J. Wang, and D. Meng, "Guaranteed tensor recovery fused low-rankness and smoothness," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 10990–11007, Mar. 2023.

[33] Y. Chen, X. Xiao, and Y. Zhou, "Jointly learning kernel representation tensor and affinity matrix for multi-view clustering," *IEEE Trans. Multimedia*, vol. 22, no. 8, pp. 1985–1997, Aug. 2020.

[34] Y. Chen, X. Xiao, C. Peng, G. Lu, and Y. Zhou, "Low-rank tensor graph learning for multi-view subspace clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 1, pp. 92–104, Jan. 2021.

[35] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.

[36] C. Song, S. Yoon, and V. Pavlovic, "Fast ADMM algorithm for distributed optimization with adaptive penalty," in *Proc. AAAI Conf. Artif. Intell.*, 2016, vol. 30, no. 1, pp. 1–7.

[37] G. Golub, S. Nash, and C. Van Loan, "A Hessenberg–Schur method for the problem AX + XB= C," *IEEE Trans. Autom. Control*, vol. AC-24, no. 6, pp. 909–913, Dec. 1979.

[38] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[39] X. Li, Q. Wang, F. Nie, and M. Chen, "Locality adaptive discriminant analysis framework," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 7291–7302, Aug. 2022.

[40] S. Mika, B. Schölkopf, A. Smola, K.-R. Müller, M. Scholz, and G. Rätsch, "Kernel PCA and de-noising in feature spaces," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 11, 1998, pp. 1–7.

[41] M. Meng, M. Lan, J. Yu, J. Wu, and D. Tao, "Constrained discriminative projection learning for image classification," *IEEE Trans. Image Process.*, vol. 29, pp. 186–198, 2020.

[42] C. Zhang, H. Li, Y. Qian, C. Chen, and Y. Gao, "Pairwise relations oriented discriminative regression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 7, pp. 2646–2660, Jul. 2021.

Kangdao Liu received the B.S. degree in computer science and technology from Shandong University, China, in 2022. He is currently pursuing the Ph.D. degree with the University of Macau. His research interests include dimensionality reduction and image processing.

Xiaolin Xiao (Member, IEEE) received the B.E. degree from Wuhan University, China, in 2013, and the Ph.D. degree from the University of Macau, Macau, in 2019. She is currently an Associate Professor with the School of Computer Science, South China Normal University, China. Her research interests include multi-view learning and color image processing and understanding.

Jinkun You (Graduate Student Member, IEEE) received the B.S. degree in computer science and technology from Wuhan University of Technology, China, in 2019, and the M.E. degree in computer technology from Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, China, in 2022. He is currently pursuing the Ph.D. degree with the University of Macau. His research interests include multimedia security, super-resolution, and image processing.

Yicong Zhou (Senior Member, IEEE) received the B.S. degree in electrical engineering from Hunan University, Changsha, China, and the M.S. and Ph.D. degrees in electrical engineering from Tufts University, Medford, MA, USA. He is currently a Professor with the Department of Computer and Information Science, University of Macau, Macau, China. His research interests include image processing, computer vision, machine learning, and multimedia security.

He is a fellow of the Society of Photo-Optical Instrumentation Engineers (SPIE) and was recognized as one of Highly Cited Researchers in 2020, 2021, and 2023. He serves as an Associate Editor for IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.