

Research Article

Large Margin Graph Embedding-Based Discriminant Dimensionality Reduction

Yanjia Tian ^{1,2} and Xiang Feng ^{1,3}

¹Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai, China

²School of Electronics and Information, Shanghai Dianji University, Shanghai, China

³Shanghai Engineering Research Center of Smart Energy, Shanghai, China

Correspondence should be addressed to Xiang Feng; xfeng_ecust@163.com

Received 11 October 2021; Revised 18 November 2021; Accepted 22 November 2021; Published 3 December 2021

Academic Editor: Bai Yuan Ding

Copyright © 2021 Yanjia Tian and Xiang Feng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Discriminant graph embedding-based dimensionality reduction methods have attracted more and more attention over the past few decades. These methods construct an intrinsic graph and penalty graph to preserve the intrinsic geometry structures of intraclass samples and separate the interclass samples. However, the marginal samples cannot be accurately characterized only by penalty graphs since they treat every sample equally. In practice, these marginal samples often influence the classification performance, which needs to be specially tackled. In this study, the near neighbors' hypothesis margin of marginal samples has been further maximized to separate the interclass samples and improve the discriminant ability by integrating intrinsic graph and penalty graph. A novel discriminant dimensionality reduction named LMGE-DDR has been proposed. Several experiments on public datasets have been conducted to verify the effectiveness of the proposed LMGE-DDR such as ORL, Yale, UMIST, FERET, CMU-PIE09, and AR. LMGE-DDR performs better than other compared methods, and the corresponding standard deviation of LMGE-DDR is smaller than others. This demonstrates that the evaluation method verifies the effectiveness of the introduced method.

1. Introduction

Dimensionality reduction (DR) is more important in most fields such as machine learning and pattern recognition [1–4]. It aims to resolving the curse of dimensionality by achieving relevant low-dimensional representations of high-dimensional datasets. Linear discriminant analysis (LDA) and principal component analysis (PCA) are the most representative methods [5, 6]. PCA obtains low-dimensional space by maximizing variance. LDA can use label information to project the feature space to distinguish categories by maximizing the interclass distance and minimizing the intraclass distance. However, LDA cannot capture the local structure of data. As is known, the local structures of high-dimensional data are very important for data representation.

K near neighbor graph can better characterize the local structure of data [7]. Thus, over the past years, graph embedding-based dimensionality reduction methods have sprung up [7, 8], such as LLE [9], Isomap [10, 11], and Laplacian eigenmap [12]. However, these manifold learning methods do not directly process the new samples because they do not obtain any mapping function, which is known as the 'out-of-sample' problem [13]. Therefore, to solve the problem, a more effective method is presented to obtain the explicit projection mapping. Locality preserving projections (LPPs) are to preserve the local structure of data in the low-dimensional space, which is a famous method [2]. For its simplicity and effectiveness, its variants have been proposed [14, 15]. However, LPP performs worse in classification since it does not fully use label information, which is an unsupervised method [16]. Neighborhood preserving projection

(NPP) preserved the local neighborhood information on the data manifold [17].

To further improve the classification performance, discriminant graph embedding-based methods have gradually become a popular research topic by using label information, which aims to preserve the within-class geometrical structure while, at the same time, maximizing the between-class distances of different manifolds [18]. Thus, recently, more and more discriminant graph embedding-based methods have been studied. Marginal fisher analysis (MFA) constructs two adjacency graphs to maximize the separability between pairwise marginal data points [19]. Local discriminant embedding (LDE) [20] utilized the label information and proposed the nearest neighbor-based embedding. However, it suffers from the so-called small-sample-size (SSS) problem that it cannot directly be applied to high-dimensional data [20]. Considering the local intraclass attraction or interclass repulsion, discriminant neighborhood embedding (DNE) was proposed to make data points in the same class compacted, whereas the gaps between classes become wider in a low-dimensional subspace [21]. However, DNE does not always set the edges with its neighbors of different classes, which would reduce the interclass distance in the new space and will deteriorate the classification [22]. Thus, Ding et al. constructed double adjacency graphs to link their homogeneous and heterogeneous neighbors and introduced a more effective version of DNE termed DAG-DNE [22]. Inspired by DAG-DNE, some discriminant analysis-based methods have been proposed over the past few years [23–33].

Most dimensionality reduction methods can be unified in the graph-embedding framework [19]. The ways to construct the similarity graph and the penalty graph among these methods are different [34]. Therefore, the graph-embedding-based methods are sensitive to the weight matrix, whereas they endow the same weight for each sample (including marginal samples) in the same way. However, as stated in [35], these marginal samples located in the class margin in the high-dimensional space have been treated to achieve maximum between-class hypothesis margin and good classification performance, which is more crucial in the classification performance. Therefore, large hypothesis margins between near neighbors of these marginal samples can improve the discriminating power of embedding features and should be treated separately. In this study, for the marginal samples, the nearest neighbors' hypothesis margin of the marginal sample has been considered and maximized to improve the discriminant power, in addition to constructing double adjacency graphs. In this study, a novel large margin graph embedding-based discriminant dimensionality reduction named LMGE-DDR has been introduced. Most experimental results confirm the effectiveness of the proposed LMGE-DDR on several public datasets.

2. Methods

Firstly, the common notations in this study are presented. The high-dimensional data are denoted as $X = [x_1, x_2, \dots, x_n] \in R^{d \times n}$ with n samples in d -dimensions and include

C classes with class $c_i \in 1, 2, \dots, C$. $y_i = \mathbf{P}^T x_i$ denotes that the sample x_i is transformed by the matrix $\mathbf{P} = [p_1, p_2, \dots, p_r] \in R^{d \times r}$, where $d \gg r$, $p_i \in R^d$ is any one column vector. $N_k^+(x_i)$ ($N_k^-(x_i)$) and $N_k(x_i)$, respectively, denote the k neighbors with the same class (different class) and k neighbors of sample x_i .

2.1. DNE. Discriminant neighborhood embedding (DNE) considered the local intraclass attraction and interclass repulsion and learned the intrinsic graph F^w and penalty graph F^b as follows:

$$F_{ij}^w = \begin{cases} 1, & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \quad c_i = c_j, \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

$$F_{ij}^b = \begin{cases} 1, & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \quad c_i \neq c_j, \\ 0, & \text{otherwise.} \end{cases}$$

The objective function can be denoted as follows:

$$\max \theta(P) = \Psi(P) - \Phi(P), \quad (2)$$

$$\text{s.t. } P^T P = I.$$

Herein,

$$\Phi(P) = \sum_{x_i \in N_k^+(x_j)} \sum_{x_j \in N_k^+(x_i)} P^T x_i - P^T x_j^2 F_{ij}^w, \quad (3)$$

$$= 2P^T X (D^w - F^w) X^T P,$$

where $D_{ii}^w = \sum_j F_{ij}^w$.

$$\Psi(P) = \sum_{x_i \in N_k^-(x_j)} \sum_{x_j \in N_k^-(x_i)} P^T x_i - P^T x_j^2 F_{ij}^b, \quad (4)$$

$$= 2P^T X (D^b - F^b) X^T P,$$

where $D_{ii}^b = \sum_j F_{ij}^b$.

The constraint $P^T P = I$ can preserve the local structure and reinforce the discriminant ability [36].

The objective in (2) can be rewritten by the formal of trace as follows:

$$\begin{aligned} \theta(P) &= \Psi(P) - \Phi(P), \\ &= 2tr P^T X (D^b - F^b) X^T P - 2tr P^T X (D^w - F^w) X^T P \\ &= 2tr P^T X (D^b - F^b - D^w + F^w) X^T P \\ &= 2tr P^T X S X^T P, \end{aligned} \quad (5)$$

where $S = D^b - F^b - D^w + F^w$. Therefore, the objective function (2) can be rewritten as follows:

$$\begin{cases} \max & \theta(P) = tr\{P^T X S X P\}, \\ \text{s.t.} & P^T P = I. \end{cases} \quad (6)$$

The projection matrix P can be found by resolving the following eigenvector problem:

$$XSX^T P = \lambda P, \quad (7)$$

where λ_i is the eigenvalues, $i = 1, \dots, d$, and P_i ($i = 1, \dots, d$) is the corresponding eigenvector. Assume $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ and $P = [P_1, P_2, \dots, P_r]$. The details are presented in [21].

2.2. DAG-DNE. Double adjacency graph-based discriminant neighborhood embedding termed DAG-DNE constructed double adjacency graphs to propose a more effective version of DNE. In DAG-DNE, F^b and F^w can be defined as follows:

$$F_{ij}^w = \begin{cases} 1, & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \\ 0, & \text{otherwise,} \end{cases} \quad (8)$$

$$F_{ij}^b = \begin{cases} 1, & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \\ 0, & \text{otherwise.} \end{cases}$$

The projection matrix \mathbf{P} can be solved as in DNE as follows:

$$\begin{cases} \max \theta(\mathbf{P}) = \Psi(\mathbf{P}) - \Phi(\mathbf{P}), \\ \text{s.t. } \mathbf{P}^T \mathbf{P} = \mathbf{I}. \end{cases} \quad (9)$$

3. Proposed Method

It is revealed that the weights in the adjacency matrix have been endowed in the same way for each sample including the marginal sample, which cannot further improve the between-class hypothesis margin and deteriorates the classification performance. In this study, the marginal sample is defined in Definition 1. The hypothesis margin was studied as in [37–39].

Definition 1 (marginal sample). x_i is regarded as a marginal sample if $\exists x_j \in N_{k_0}(x_i)$ and $c_i \neq c_j$.

The marginal samples in this study are the ones located in class margin. Figure 1 is the k_0 near neighbors' graph and shows the marginal samples (i.e., {5, 6, 7, 8}).

Definition 2 (hypothesis margin). As is shown in [37], the hypothesis margin can be defined as follows:

$$H(x) = \frac{1}{2}x - \text{near miss}(x) - x - \text{near hit}(x), \quad (10)$$

where $\text{nearhit}(x)$ and $\text{nearmiss}(x)$ denote the nearest neighbors of sample x with the same class and different class, respectively. $\|\cdot\|$ represents the L_2 norm. The sample x can be accurately recognized by 1NN classifier (the nearest neighbor) when $H(x) > 0$, as illustrated in Figure 2.

Definition 3 (heterogeneous near neighbors' hypothesis margin). A marginal of sample x_1 is shown in Figure 3 to illustrate the heterogeneous near neighbors' hypothesis margin of x_1 , which is defined as follows:

$$N_{k_1}^+(x_1) = \{x_2, x_3, x_4\}. \quad (11)$$

Herein, $N_{k_2}^+(x_1) = \{x_2, x_3, x_4\}$, $N_{k_1}^-(x_1) = \{x_5, x_6, x_7\}$.

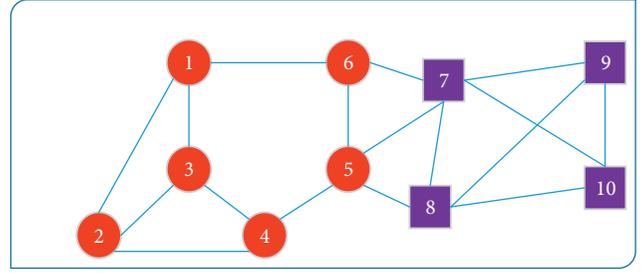


FIGURE 1: An illustration of k_0 near neighbors graph and marginal samples. Different style corresponds to different classes ($k_0 = 3$). Nodes {5 6 7 8} are the marginal samples.

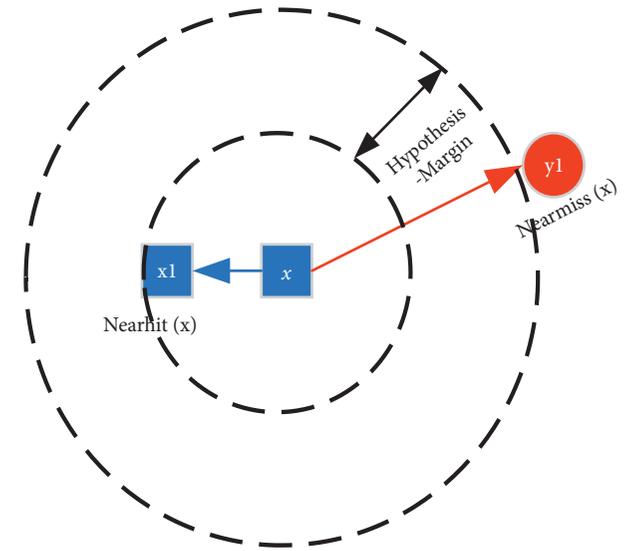


FIGURE 2: Illustration of the hypothesis margin.

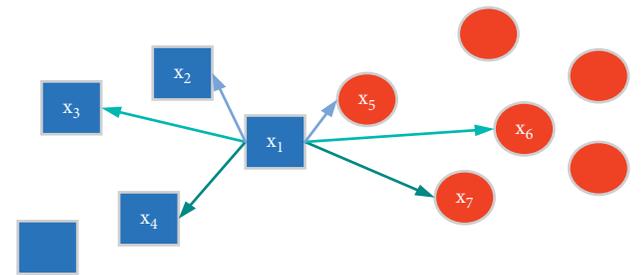


FIGURE 3: Illustration of heterogeneous near neighbors' hypothesis margin ($k_1 = k_2 = 3$).

As shown in (11), it can keep the heterogeneous samples separated and achieve a large margin between heterogeneous near neighbors when all the expressions in brackets are larger than zero, which means it can be correctly classified by the 1NN classifier.

4. LMGE-DDR

On the basis of DAG-DNE, the marginal samples in high-dimensional space are additionally treated separately by

maximizing the heterogeneous near neighbors' margin, which can improve the discriminant power. LMGE-DDR can be proposed as follows:

The interclass weight F_{ij}^b and F_{ij}^w are the same as in DAG-DNE, the objective function of LMGE-DDR can be denoted as

$$\begin{cases} \max \theta(P) = \Psi(P)\Phi(P) + M(P), \\ \text{s.t. } P^T P = I. \end{cases} \quad (12)$$

$$\begin{aligned} M(P) &= \sum_{x_i \in MS} \sum_{x_t \in N_{k_1}^-(x_i)} \sum_{x_j \in N_{k_1}^-(x_i)} \sum_{x_t \in N_{k_2}^+(x_i)} \left((1-\alpha) \|P^T x_i - P^T x_t\| - \alpha \|P^T x_i - P^T x_j\| \right), \\ &= \sum_{x_i \in MS} \left((1-\alpha) \sum_{x_t \in N_{k_1}^-(x_i)} k_2 \|P^T x_i - P^T x_t\| - \alpha \sum_{x_j \in N_{k_2}^+(x_i)} k_1 \|P^T x_i - P^T x_j\| \right), \\ &= \sum_{x_i \in MS} \left((1-\alpha) \sum_{x_t \in N_{k_1}^-(x_i)} k_2 (x_i - x_t) - PP^T (x_i - x_t) - \alpha \sum_{x_j \in N_{k_2}^+(x_i)} k_1 (x_i - x_t) - PP^T (x_i - x_t) \right). \end{aligned} \quad (13)$$

The solution of (12) is easily obtained by solving the maximum eigenvalue problem.

$$XSX^T P + \sum_{x_i \in MS} ((1-\alpha)G_i - \alpha H_i)P = \lambda P. \quad (14)$$

$$\begin{aligned} \text{Here, } G_i &= \sum_{x_t \in N_{k_1}^-(x_i)} k_2 (x_i - x_t)^T (x_i - x_t) \\ H_i &= G_i = \sum_{x_j \in N_{k_1}^-(x_i)} k_1 (x_i - x_t)^T (x_i - x_t), \end{aligned} \quad (15)$$

where λ_i is the eigenvalues, $i=1, \dots, d$, and P_i is the corresponding eigenvector, $i=1, \dots, d$. Assume $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ and $\mathbf{P} = [P_1, P_2, \dots, P_r]$.

The details of LMGE-DDR can be seen in Algorithm 1.

5. Analysis of LMGE-DDR

In this section, LMGE-DDR will be analyzed to illustrate the effectiveness in preserving the geometrical and discriminant structures.

Although LMGE-DDR is similar to DAG-DNE in constructing an adjacency graph, for the marginal samples in high-dimensional space, LMGE-DDR maximizes the heterogeneous near neighbors' hypothesis margin to achieve a large between-class margin in low-dimensional subspace and discriminate the local structure of neighbors, improving the discriminant power compared to DAG-DNE.

The performances of LMGE-DDR in a Toy data are illustrated in Figure 4.

As shown in Figure 4(a), for the sample x_1 , $\text{nearhit}(x_1)$ is x_2 and $\text{nearmiss}(x_1)$ is x_3 . Thus, based on (12), the hypothesis margin of x_1 is denoted as follows:

Here, $\Psi(P), \Phi(P)$ are the same as in DAG-DNE. MS denotes the marginal samples set in the high-dimensional space. α is a trade-off parameter and here $\alpha \in [0, 1]$.

This objective function is transformed into two parts as follows:

$$\Psi(P) - \Phi(P) = 2trP^T XSX^T P \text{ based on (5).}$$

$$\begin{aligned} H(x_1) &= \frac{1}{2} (x_1 - \text{nearmiss}(x_1) - x_1 - \text{nearhit}(x_1)), \\ &= \frac{1}{2} (x_1 - x_3 - x_1 - x_2) = -0.83. \end{aligned} \quad (16)$$

Based on Definition 2, the sample x_1 will be recognized by mistake because its hypothesis margin is less than zero.

The embedded results and hypothesis margins in one-dimensional space are illustrated in Figures 4(b)–4(e). It can be seen that the hypothesis margins of sample x_1 in the low-dimensional space are less than zero in MFA and DAG-DNE, which is the opposite situation in MNMDP, DNE, and LMGE-DDR. In LMGE-DDR, the hypothesis margin of sample x_1 is larger ($H(x_1) = 0.39$) than that in DAG-DNE, which is useful for the classification.

Overall, maximizing the heterogeneous near neighbors' hypothesis margin of marginal samples can further improve the discriminant power in low-dimensional space.

6. Experiments

In this section, compared with several popular methods such as DAG-DNE, DNE, MNMDP, and MFA, LMGE-DDR is conducted on several experiments systematically to verify its effectiveness. Specifically, the performance of LMGE-DDR is illustrated on the experiments of face recognition and 2-dimensional visualization. The randomly selected l images from each person constitute the training data, and the remaining are the testing data. The nearest neighbor parameters k , k_l , and k_2 in constructing adjacency graphs are set as $l-1$ for all methods as in [40]. PCA is taken to reduce

Input: a training set $\{(x_i, c_i)\}_{i=1}^N$, α, k_1, k_2, k and the dimensionality of discriminant subspace r .
 Output: projection matrix \mathbf{P} ;
 (1) Construct the intraclass adjacency graph \mathbf{F}^w by:

$$F_{ij}^w = \begin{cases} 1, & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\ 0, & \text{otherwise} \end{cases}$$
 and interclass adjacency graph \mathbf{F}^b by:

$$F_{ij}^b = \begin{cases} 1, & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \\ 0, & \text{otherwise,} \end{cases}$$

 (2) computing the MS, G_i, H_i based on (15)
 (3) Eigendecompose the matrix, where $XSX^T + \sum_{x_i \in MS} ((1 - \alpha)G_i - \alpha H_i)$
 (4) Choose the r largest eigenvalues corresponding eigenvectors: $\mathbf{P} = [P_1, P_2, \dots, P_r]$.

ALGORITHM 1: LMGE-DDR.

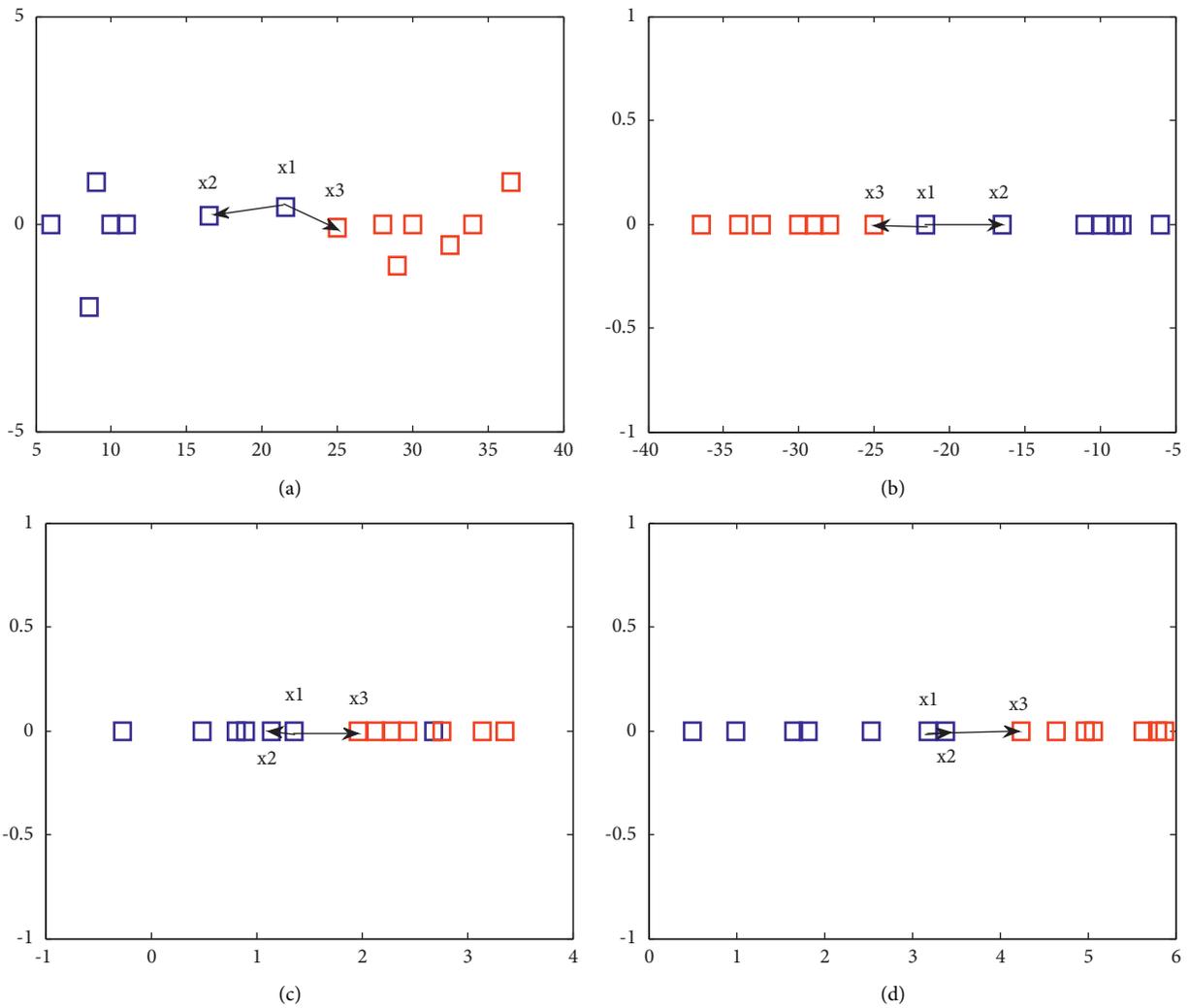


FIGURE 4: Continued.

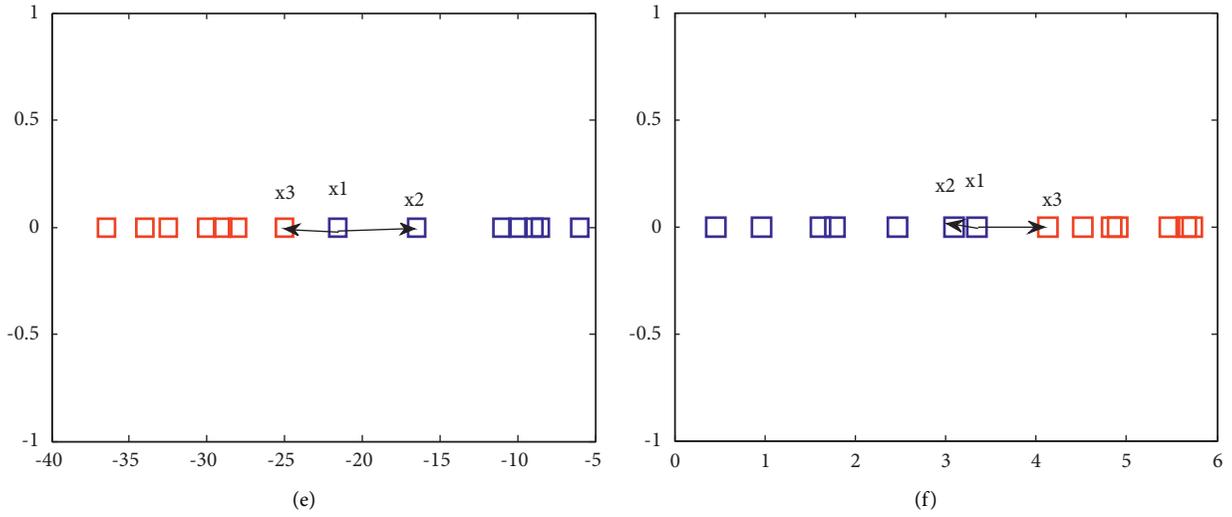


FIGURE 4: Illustration of embedding space and hypothesis margins: (a) Toy data ($H(x_I) = -0.83$); (b) MFA ($k_1 = k_2 = 5$, $H(x_I) = -0.84$); (c) MNMDP ($(k) = 5$, $H(x_I) = 0.19$); (d) DNE ($(k) = 5$, $H(x_I) = 0.43$); (e) DAG-DNE ($(k) = 5$, $H(x_I) = -0.84$); (f) LMGE-DDR ($(k) = k_1 = k_2 = 3$, $(a) = 0.8$, $H(x_I) = 0.39$).

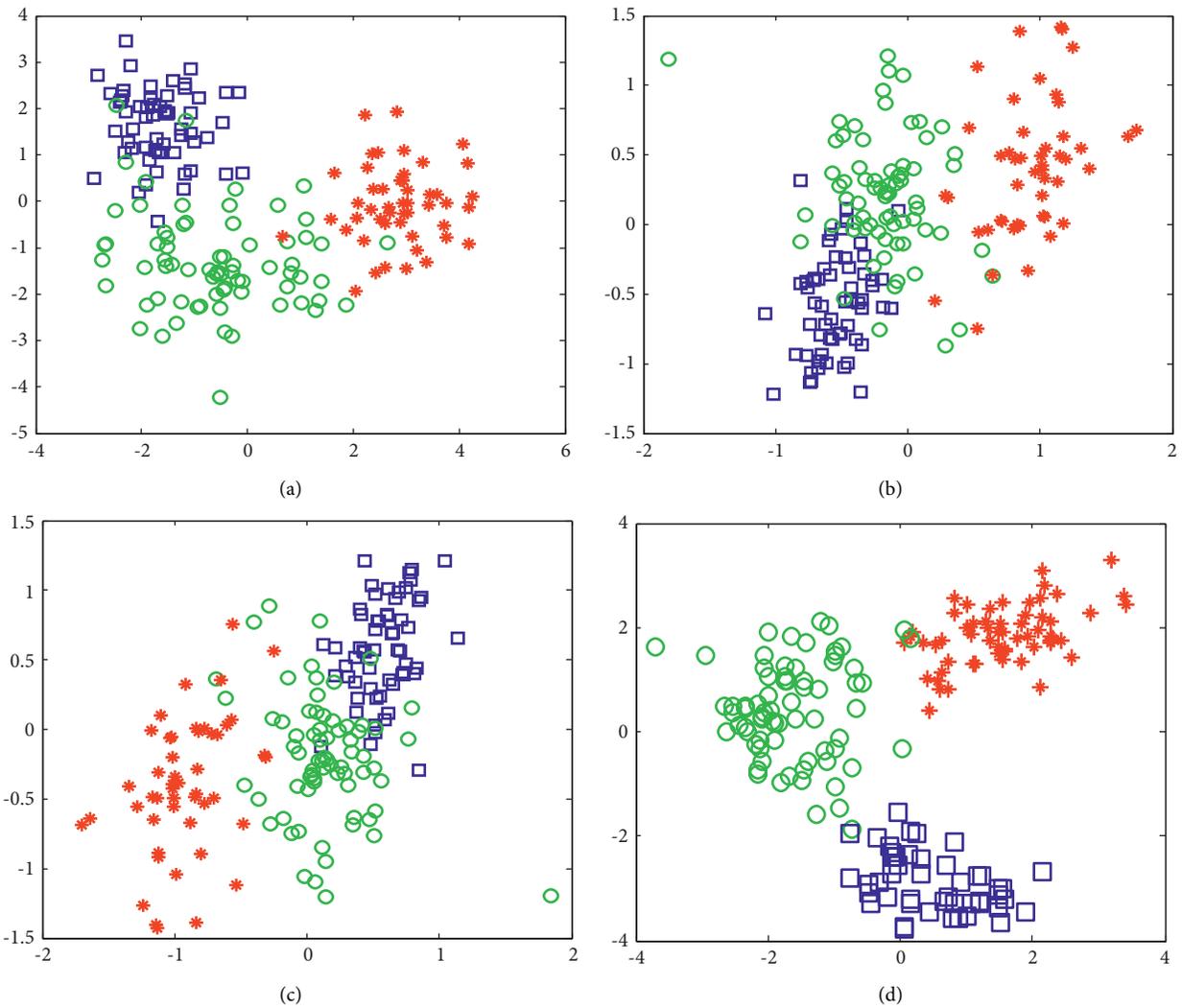


FIGURE 5: Continued.

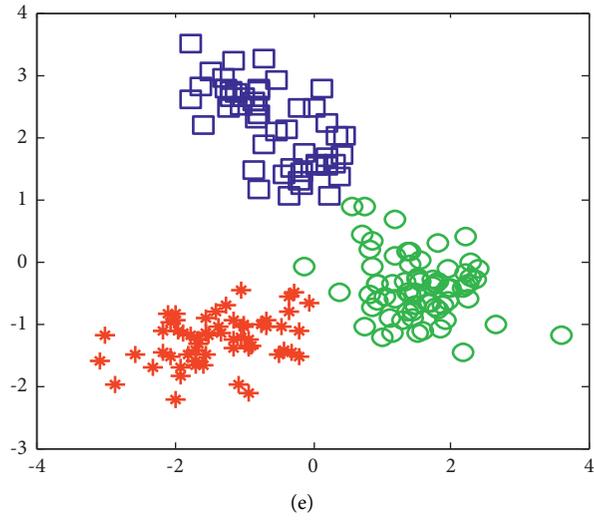


FIGURE 5: Illustration of visualization on wine dataset: (a) MFA ($k_1 = 15, k_2 = 25$); (b) DNE ($(k) = 15$); (c) MNMDDP ($(k) = 15$); (d) DAG-DNE ($(k) = 15$); (e) LMGE-DDR ($(k) = k_1 = k_2 = 15, k_0 = 9, (a) = 0.8$).

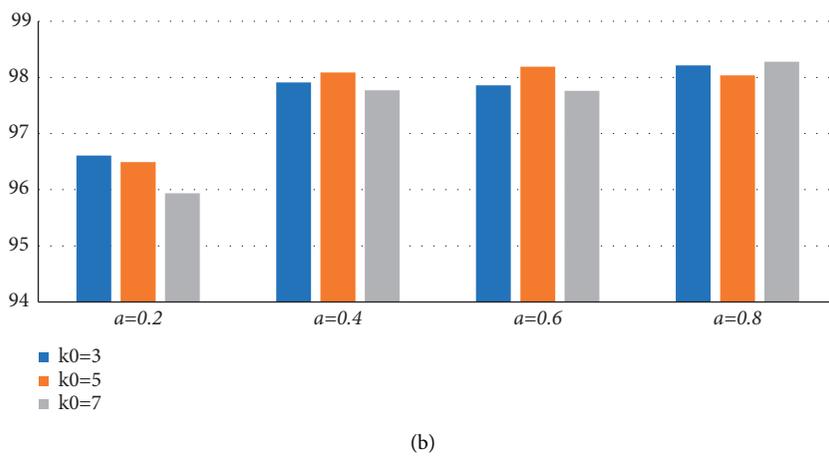
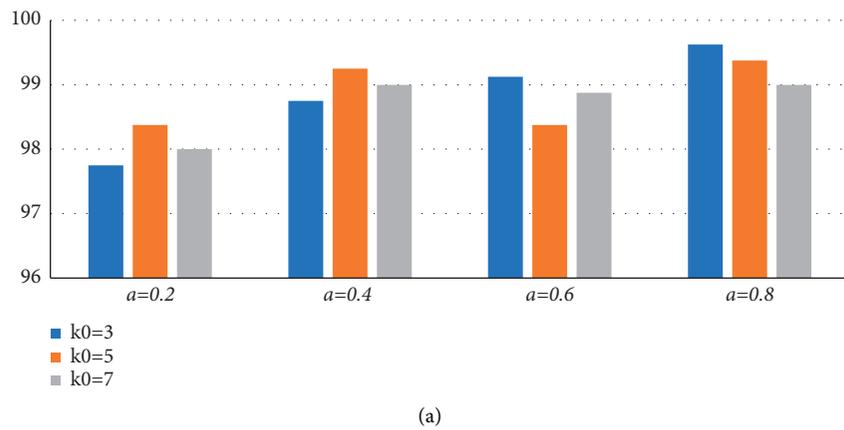


FIGURE 6: Continued.

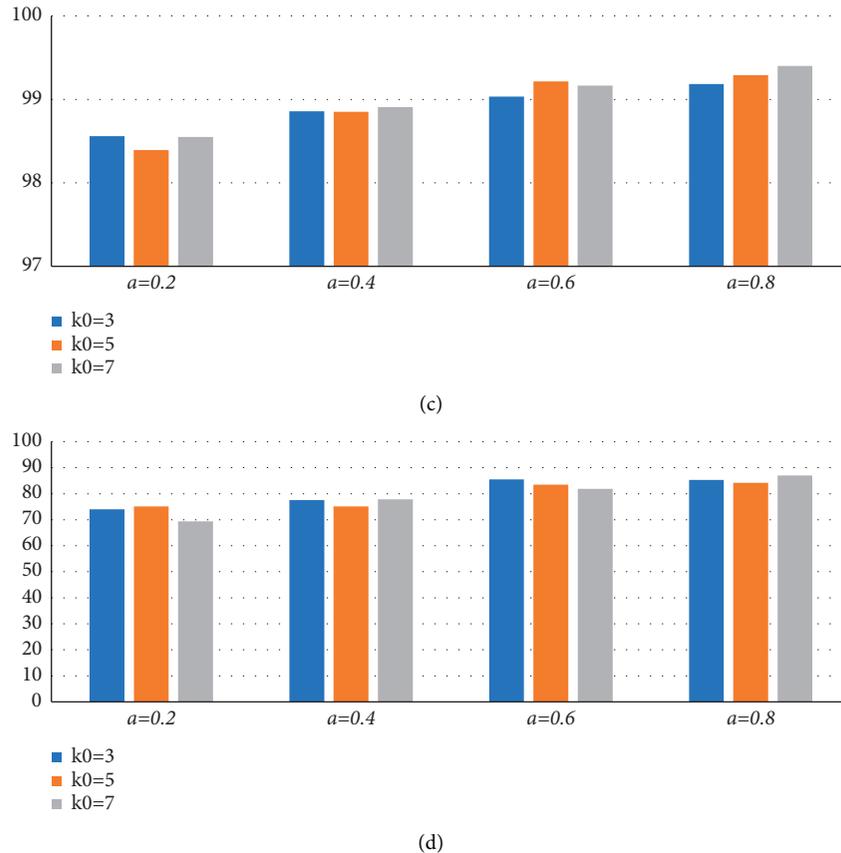


FIGURE 6: Recognition accuracy (%) of LMGE-DDR with variations of a , k_0 : (a) ORL (1 = 9); (b) UMIST (1 = 10); (c) AR (1 = 9); (d) Yale (1 = 9).

the dimensions by nearly 98% of the image energy. The 1NN classifier is applied to perform the classification. The average result of 20 runs is regarded as the classification result.

6.1. 2D Visualization. Wine dataset is taken to perform the 2D visualization as shown in Figure 5 [41]; from Figure 5, it can be seen clearly that the sample points in the low-dimensional space learned by LMGE-DDR are separated compared to DAG-DNE.

6.2. Face Recognition. LMGE-DDR is evaluated on the ORL (<http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html>), FERET [42], AR [43], Yale†, UMIST (<https://www.sheffield.ac.uk/eee/research/iel/research/face>), and CMU-PIE09† face datasets to evaluate the classification performance and it has been systematically compared with several popular methods such as MFA, MNMDP, DNE, and DAG-DNE.

6.2.1. Parameter Analysis. The sensitivity of parameter k_0 , a in LMGE-DDR is analyzed on several face datasets when parameters k_1 , k_2 have been set as 1. Figure 6 presents the best recognition rates of LMGE-DDR with the different values of k_0 , a . The results in Figure 6 reveal that the recognition accuracy of LMGE-DDR fluctuates up and

down. In total, the best recognition accuracy can be achieved when a and k_0 are larger. The reason is that large a can make marginal samples tightly clustered toward the class center. The large k_0 is, the more marginal samples are. That is to say, heterogeneous near neighbors' margin of more marginal samples can be maximized and achieve large between-class margin, which is favorable for classification. Thus, the values of k_0 and a in LMGE-DDR on different datasets are adopted by cross-validation in face recognition experiments.

6.2.2. Experiments Results. In this section, several experiments on public datasets have been conducted to verify the effectiveness of the proposed LMGE-DDR, such as ORL, Yale, UMIST, FERET, CMU-PIE09, and AR, whose example images are shown in Figure 7. Each image in ORL is first aligned and cropped to 32×32 . Each image in Yale is first aligned and cropped to 32×32 . Each image in UMIST is first aligned and cropped to 40×50 . All the images in FERET are cropped to 80×80 . All the images in CMU-PIE09 are cropped to 64×64 . All the images in AR are cropped to 50×40 . Tables 1–6 are the best recognition results on different datasets. Figure 8 are the recognition results on different dimensions.

As shown in Figure 8 and Tables 1–6, we can see that in most experiments, LMGE-DDR performs better than other

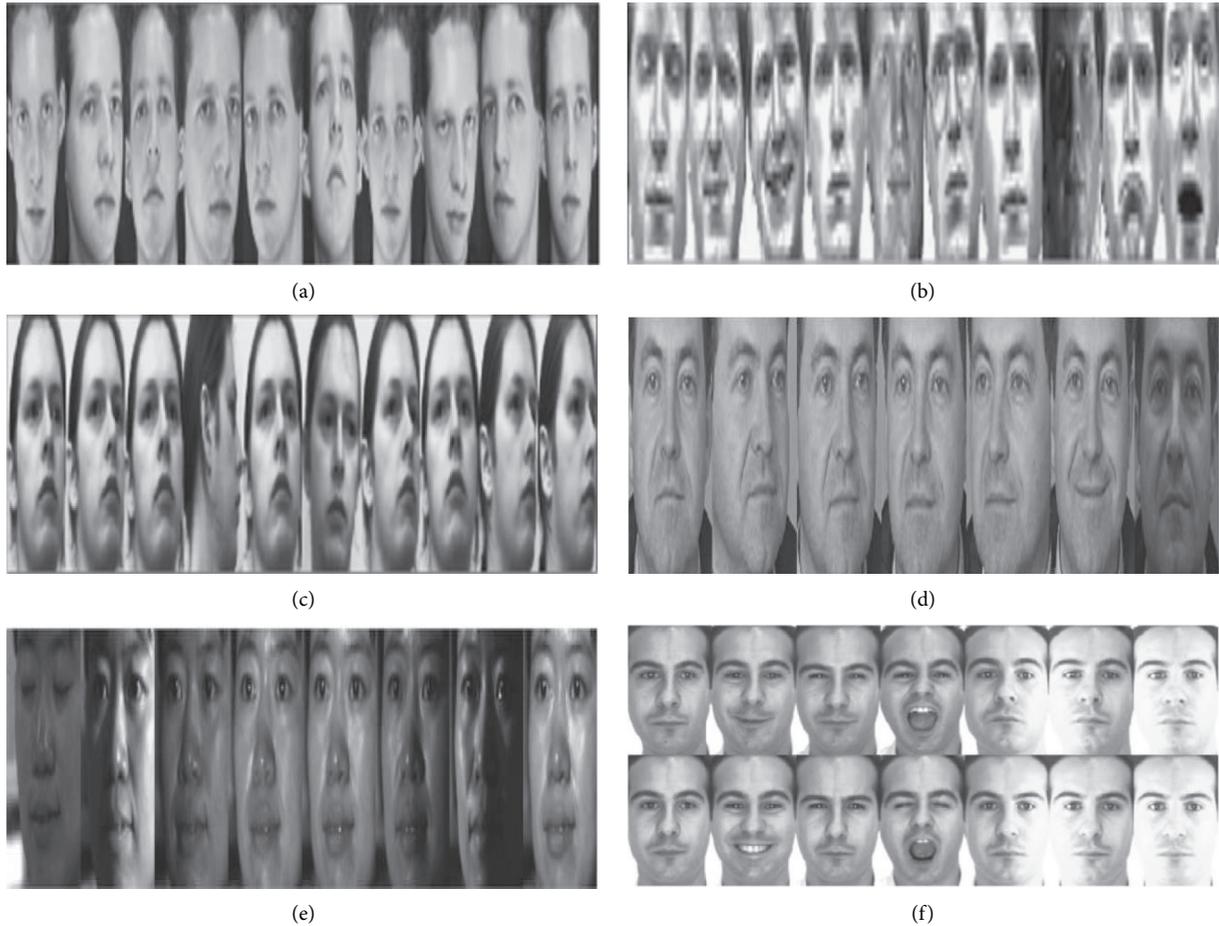


FIGURE 7: Illustration of some images on different datasets: (a) ORL; (b) Yale; (c) UMIST; (d) FERET; (e) CMU PIE; (f) AR.

TABLE 1: The classification performance on the ORL dataset (%).

Method	$l=5$ ($k_0=3$, $a=0.8$)	$l=6$ ($k_0=3$, $a=0.8$)	$l=7$ ($k_0=5$, $a=0.8$)
MFA	88.10 ± 2.36 (55)	91.19 ± 1.98 (61)	92.79 ± 1.78 (54)
DNE	89.08 ± 1.88 (39)	91.53 ± 1.21 (43)	93.46 ± 1.76 (39)
MNMDP	95.45 ± 1.51 (20)	96.01 ± 1.21 (27)	97.61 ± 1.11 (31)
DAG-DNE	96.67 ± 0.94 (40)	97.66 ± 0.97 (59)	98.29 ± 1.16 (34)
LMGE-DDR	96.95 ± 1.14 (24)	98.03 ± 0.89 (22)	98.88 ± 0.98 (30)

TABLE 2: The classification performance on the Yale dataset (%).

Method	$l=7$ ($k_0=3$, $a=0.8$)	$l=8$ ($k_0=5$, $a=0.8$)	$l=9$ ($k_0=7$, $a=0.8$)
MFA	62.34 ± 4.24 (45)	63.63 ± 5.61 (50)	68.23 ± 5.71 (45)
DNE	67.36 ± 4.17 (21)	69.23 ± 5.33 (26)	73.23 ± 5.63 (28)
MNMDP	71.23 ± 3.62 (26)	72.20 ± 5.39 (21)	76.23 ± 5.29 (21)
DAG-DNE	77.58 ± 4.76 (21)	77.87 ± 5.56 (18)	81.17 ± 5.26 (16)
LMGE-DDR	83.17 ± 3.21 (40)	83.78 ± 4.15 (28)	86.17 ± 4.55 (22)

compared methods and the corresponding standard deviation of LMGE-DDR is smaller than others.

6.2.3. Time Cost Analysis. In this section, the time cost of different methods is evaluated on several datasets

including ORL, Yale, UMIST, and FERET. In Table 7, it is computed by the time of running one time where $l=5$, $d=20$.

It can be concluded that LMGE-DDR is comparable with other methods in time cost; however, some perform better than others.

TABLE 3: The classification performance on the UMIST dataset (%).

Method	$l=10 (k_0=7, a=0.8)$	$l=11 (k_0=7, a=0.8)$	$l=12 (k_0=7, a=0.8)$
MFA	95.81 ± 1.42 (35)	96.54 ± 1.01 (40)	95.81 ± 1.42 (35)
DNE	97.45 ± 1.11 (12)	97.98 ± 1.17 (10)	97.45 ± 1.11 (12)
MNMDP	98.16 ± 0.84 (12)	98.52 ± 1.01 (10)	98.16 ± 0.84 (12)
DAG-DNE	98.43 ± 0.68 (26)	98.91 ± 0.76 (18)	99.13 ± 0.62 (24)
LMGE-DDR	98.85 ± 0.51 (24)	99.04 ± 0.54 (22)	99.35 ± 0.59 (24)

TABLE 4: The classification performance on the FERET dataset (%).

Method	$l=4 (k_0=3, a=0.8)$	$l=5 (k_0=3, a=0.8)$	$l=6 (k_0=3, a=0.8)$
MFA	36.68 ± 1.55 (60)	60.00 ± 2.32 (65)	44.53 ± 2.79 (60)
DNE	69.47 ± 1.43 (22)	56.58 ± 1.97 (12)	60.38 ± 2.57 (16)
MNMDP	99.03 ± 0.34 (76)	72.47 ± 1.66 (72)	77.33 ± 2.78 (42)
DAG-DNE	84.75 ± 1.28 (30)	87.88 ± 1.04 (36)	89.63 ± 1.76 (48)
LMGE-DDR	88.05 ± 0.86 (32)	89.71 ± 0.67 (40)	90.73 ± 1.68 (36)

TABLE 5: The classification performance (%) on the CMIU-PIE09 dataset.

Method	$l=10 (k_0=5, a=0.8)$	$l=8 (k_0=5, a=0.8)$	$l=6 (k_0=5, a=0.8)$
MFA	72.45 ± 2.67 (60)	63.47 ± 2.35 (60)	53.17 ± 1.65 (60)
DNE	72.60 ± 1.79 (60)	64.56 ± 1.81 (60)	54.89 ± 1.67 (60)
MNMDP	93.24 ± 0.89 (60)	91.11 ± 1.01 (60)	87.51 ± 0.97 (60)
DAG-DNE	93.92 ± 0.73 (56)	92.65 ± 1.02 (59)	90.56 ± 0.88 (59)
LMGE-DDR	94.29 ± 0.61 (60)	93.31 ± 0.83 (52)	91.35 ± 0.83 (54)

TABLE 6: The classification performance (%) on the AR dataset.

Method	$l=5 (k_0=3, a=0.8)$	$l=7 (k_0=3, a=0.8)$	$l=9 (k_0=7, a=0.8)$
MFA	76.59 ± 1.67 (200)	83.38 ± 1.19 (201)	87.15 ± 1.65 (196)
DNE	78.19 ± 1.58 (190)	86.13 ± 1.49 (194)	95.04 ± 1.15 (126)
MNMDP	97.55 ± 0.75 (72)	99.03 ± 0.34 (76)	99.51 ± 0.19 (64)
DAG-DNE	97.47 ± 0.84 (196)	98.81 ± 0.37 (132)	99.02 ± 0.39 (56)
LMGE-DDR	97.81 ± 0.72 (60)	99.07 ± 0.22 (56)	99.42 ± 0.27 (38)

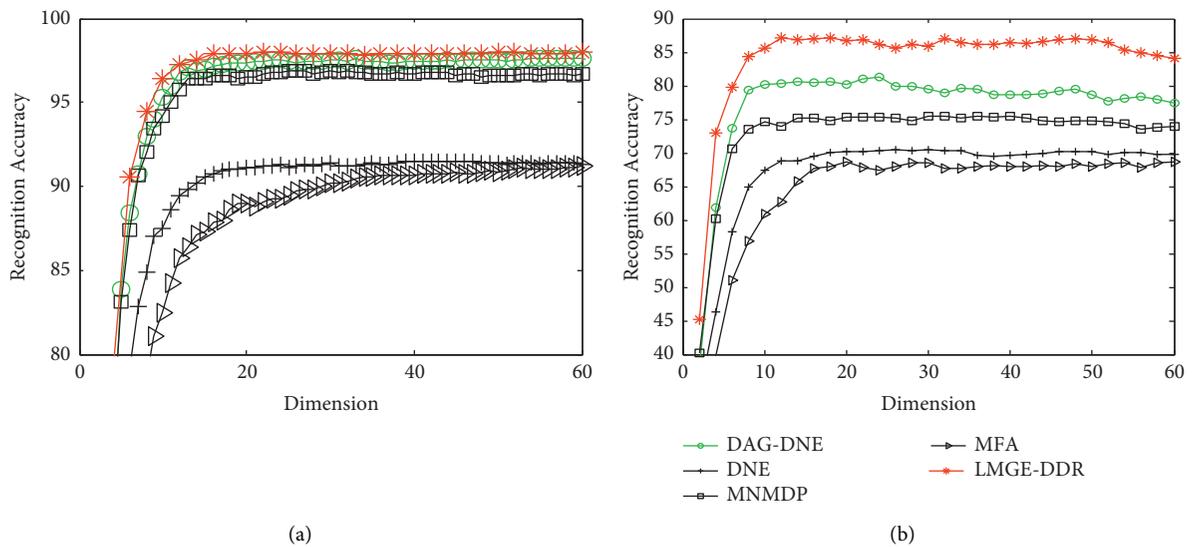


FIGURE 8: Continued.

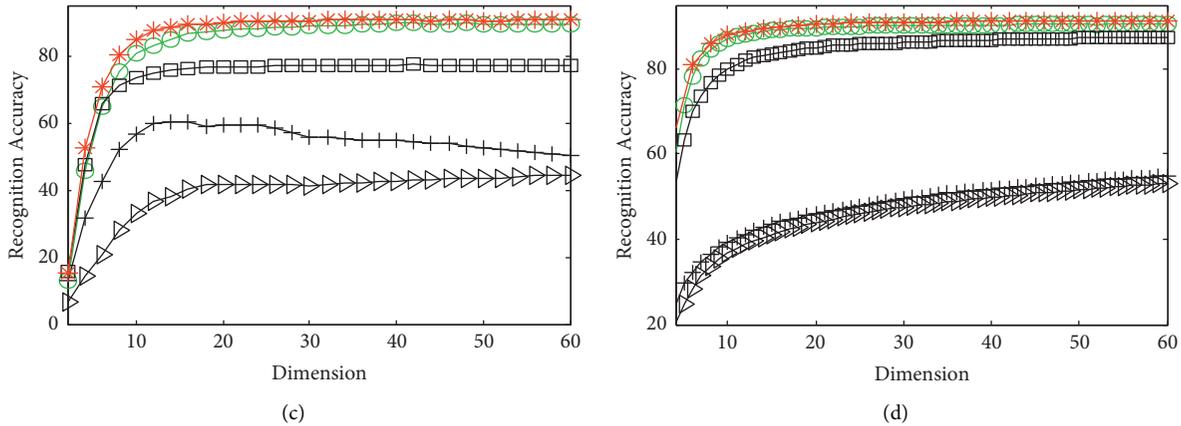


FIGURE 8: Recognition rate curves of different methods versus dimensions: (a) ORL ($l=6$); (b) Yale ($l=9$); (c) FERET ($l=6$); (d) CMU PIE ($l=6$).

TABLE 7: The time cost (seconds) on different datasets for different methods.

Method	ORL	Yale	UMIST	FERET
LMGE-DDR	0.0529	0.0285	0.0471	0.6080
DAG-DNE	0.0448	0.0171	0.0325	0.6118
MNMDP	0.0495	0.0139	0.0428	0.7186
DNE	0.0613	0.0153	0.0491	0.9115
MFA	0.0950	0.0464	0.1410	1.6723

7. Conclusions and Future Works

In this study, we propose a novel graph embedding-based dimensionality reduction approach named LMGE-DDR, which is based on heterogeneous near neighbors' hypothesis margin. Different from other discriminant learning methods, for marginal samples in high-dimensional space, we additionally maximize the heterogeneous near neighbors' hypothesis margin to achieve a large between-class margin, excluding learning two kinds of adjacency graphs for each same equally. This is very crucial for classification of the experiment results. Experimental results illustrate the effectiveness of LMGE-DDR. In this paper, we also employed several evaluation methods to evaluate the proposed model. The results show that on several public datasets such as ORL, Yale, UMIST, FERET, CMU-PIE09, and AR, the proposed model outperformed other benchmark models. However, in constructing adjacency graphs and marginal samples, it will be influenced by the noise, which is not completely avoided. In the future works, how to evaluate the reliability of neighborhood will be studied by introducing an adaptive adjacency factor as in [44].

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

- [1] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," *Computer Vision Pattern Recognition*, vol. 27, no. 3, pp. 586–591, 1991.
- [2] X. F. He, S. C. Yan, Y. X. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using Laplacianfaces," *Pattern Anal. Mach. Intell.* vol. 27, pp. 328–340, 2005.
- [3] Y. Y. Lin, T. L. Liu, and H. T. Chen, "Semantic manifold learning for image retrieval," *Proc. ACM Int. Conf. Multimedia*, pp. 249–258, 2005.
- [4] X. F. He, D. Cai, and J. W. Han, "Learning a maximum margin subspace for image retrieval," *Knowl. Data Eng.* vol. 20, pp. 189–201, 2008.
- [5] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [6] S. Deng, C. Wang, Z. Fu, and M. Wang, "An intelligent system for insider trading identification in Chinese security market," *Computational Economics*, vol. 57, no. 6, 2020.
- [7] H. Cai, V. Zheng, and K. Chang, "A comprehensive survey of graph embedding: problems, techniques and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 9, pp. 1616–1637, 2017.
- [8] P. Goyal and E. Ferrara, "Graph embedding techniques, applications, and performance: a survey," *Knowl-Based Syst.* vol. 151, pp. 78–94, 2018.

- [9] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
- [10] J. B. Tenenbaum, V. de-Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [11] H. C. Qu, L. Li, Z. N. Li, and J. Zheng, "Supervised discriminant isomap with maximum margin graph regularization for dimensionality reduction," *Expert Systems with Applications*, vol. 180, no. 15, Article ID 115055, 2021.
- [12] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [13] Y. Bengio, J. Paiement, P. Vincent, O. Delalleau, N. L. Roux, and M. Ouimet, "Out-of sample extensions for LLE, isomap, MDS, eigen-maps, and spectral clustering," in *Proceedings of the International Conference Neural Information Processing Systems*, pp. 177–184, Whistler, Canada, December 2004.
- [14] D. Lu, C. Ding, J. Xu, and S. Wang, "Hierarchical discriminant analysis," *Sensors*, vol. 18, no. 1, p. 279, 2018.
- [15] X. Yang, G. Liu, Q. Yu, and R. Wang, "Stable and orthogonal local discriminant embedding using trace ratio criterion for dimensionality reduction," *Multimedia Tools and Applications*, vol. 77, no. 3, pp. 3071–3081, 2018.
- [16] W. Yu, X. Teng, and C. Liu, "Face recognition using discriminant locality preserving projections," *Image and Vision Computing*, vol. 24, no. 3, pp. 239–248, 2006.
- [17] X. He, D. Cai, S. Yan, and H. J. Zhang, "Neighborhood preserving embedding," in *Proceedings of the 2005, International Conference on Computer Vision*, vol. 2, pp. 1208–1213, ICCV, Cagliari, Italy, September 2005.
- [18] X. F. He, D. Cai, S. C. Yan, and H. J. Zhang, "Neighborhood preserving embedding," in *Proceedings of the IEEE International Conference Computer Vision (ICCV'05)*, vol. 1, pp. 1208–1213, Beijing, China, October 2005.
- [19] S. Yan, D. Xu, B. Zhang, H. J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, 2007.
- [20] H. T. Chen, H. W. Chang, and T. L. Liu, "Local discriminant embedding and its variants," in *Proceedings of the International Conference Computer Vision and Pattern Recognition*, San Diego, CA, USA, June 2005.
- [21] W. Zhang, X. Y. Xue, H. Lu, and Y. F. Guo, "Discriminant neighborhood embedding for classification," *Pattern Recognition*, vol. 39, pp. 2240–2243, 2006.
- [22] C. T. Ding and L. Zhang, "Double adjacency graphs-based discriminant neighborhood embedding," *Pattern Recognition*, vol. 48, pp. 1734–1742, 2015.
- [23] Q. You, N. Zheng, S. Du, and Y. Wu, "Neighborhood discriminant projection for face recognition," *Pattern Recognition Letters*, 2007.
- [24] M. Wan, Z. Lou, and Z. Jin, "Margin maximum embedding discriminant (MMED) for feature extraction and classification," in *Proceedings of the IEEE Chinese Conf. Pattern Recognition (CCPR)*, pp. 1–5, Nanjing, China, November 2009.
- [25] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," *Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, 2007.
- [26] D. Cai, X. He, K. Zhou, J. Han, and H. Bao, "Locality sensitive discriminant analysis," in *Proceedings of the Joint Conference Artificial Intelligence*, pp. 1713–1726, Gold Coast, Australia, December 2007.
- [27] P. Huang, C. Chen, Z. Tang, and Z. Yang, "Feature extraction using local structure preserving discriminant analysis," *Neurocomputing*, vol. 140, pp. 104–113, 2014.
- [28] J. P. Gou, Y. Z. Zhan, M. Wan, X. Shen, J. Chen, and L. Du, "Maximum neighborhood margin discriminant projection for classification," *Science World Journal*, vol. 16, 2014.
- [29] J. Gou, Y. Yang, Z. Yi, J. Lv, and Y. Zhan, "Discriminative globality and locality preserving graph embedding for dimensionality reduction," *Expert Systems with Applications*, vol. 144, Article ID 113079, 2019.
- [30] S. Wang, C. Ding, C. H. Hsu, and F. Yang, "Dimensionality reduction via preserving local information," *Future Generation Computer Systems*, vol. 108, no. 22, 2018.
- [31] J. P. Gou and Z. Yi, "Locality-based discriminant neighborhood embedding," *The Computer Journal*, vol. 9, pp. 1063–1082, 2013.
- [32] H. S. Hu, D. Z. Feng, and Q. Y. Chen, "A novel dimensionality reduction method: similarity order preserving discriminant analysis," *Signal Processing*, vol. 182, Article ID 107933, 2020.
- [33] Z. Liu, K. Shi, K. Zhang, W. Ou, and L. Wang, "Discriminative sparse embedding based on adaptive graph for dimension reduction," *Engineering Applications of Artificial Intelligence*, vol. 94, Article ID 103758, 2020.
- [34] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, "Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image," *IEEE Transactions on Cybernetics*, vol. 49, pp. 2406–2419, 2018.
- [35] G. Zhao and Y. Wu, "Efficient large margin-based feature extraction," *Neural Processing Letters*, vol. 50, no. 12, 2019.
- [36] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, John Wiley & Sons, Hoboken, NJ, USA, 2nd edition, 2000.
- [37] R. G. Bachrach, A. Navot, and N. Tishby, "Margin based feature selection-Theory and algorithms," in *Proceedings of the Of the 21st International Conference on Machine Learning*, pp. 43–50, ACM Press, Banff, Canada, July 2004.
- [38] K. Crammer, R. Gilad-Bachrach, A. Navot, and N. Tishby, "Margin analysis of the l₁q algorithm," in *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, USA, 2003.
- [39] G. D. Zhao, Z. Y. Zhou, and J. M. Zhang, "Theoretical framework in graph embedding-based discriminant dimensionality reduction," *Signal Processing*, vol. 189, Article ID 108289, 2021.
- [40] J. Yang, D. Zhang, J. Yang, and B. Niu, "Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 650–664, 2007.
- [41] Wine Database: <http://www.resnet.cn/databases> (Accessed on 2021-7-01).
- [42] P. J. Phillips, "The facial recognition technology (FERET) database," 2004, http://www.itl.nist.gov/iad/humanid/feret/feret_master.html.
- [43] Q. Gao, J. Liu, H. Zhang, X. Gao, and K. Li, "Joint global and local structure discriminant analysis," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 4, pp. 626–635, 2013.
- [44] Y. Gao, Y. Zhang, J. Pan, S. Luo, and C. Yeng, "Discriminant analysis based on reliability of local neighborhood," *Expert Systems with Applications*, vol. 175, Article ID 114790, 2021.