IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 April 2024, accepted 24 May 2024, date of publication 31 May 2024, date of current version 7 June 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3407781*

RESEARCH ARTICLE

# Discriminative Regression With Latent Label Learning for Image Classification

**LANG LANG[1], XIAO QIN CHEN[1], SHA LIU[2], AND QIANG ZHOU[3]**
[1]School of Intelligent Manufacturing, Chongqing Three Gorges Vocational College, Chongqing 404155, China
[2]National Key Laboratory of Wireless Communications (NKLWC), University of Electronic Science and Technology of China, Chengdu 611731, China
[3]School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Corresponding author: Qiang Zhou (zhouqiang0715@foxmail.com)

**ABSTRACT** As one of the popular and effective supervised classification methods, linear regression is extensively used in image classification. However, the zero-one labeling matrix is too strict to be conducive to linear regression methods for learning labeling information. In addition, the linear regression focuses only on the fit of the input features to the corresponding output labels and ignores the distinctiveness between the samples. To address these two issues, this paper proposes a new method, namely, discriminative regression with latent label learning, for image classification. In contrast to the other methods, the proposed method learns labeling information in the latent label space instead of the input zero-one labeling space, doing so has the advantage that the proposed method can learn the labeling information in the data more flexibly. To guide the transform matrix to learn the discriminative information in the data, a regularization term with the idea of shortening the distance between samples within a class and lengthening the distance between samples between classes is integrated into the objective function of the proposed method. To obtain the solution of the proposed model, an iterative optimization algorithm is developed. Comprehensive experiments show that the classification performance of the proposed method outperforms the current state-of-the-art methods and deep learning methods on public image datasets with small sample sizes.

**INDEX TERMS** Linear regression, latent label learning, supervised learning, classification.

## I. INTRODUCTION

Image classification plays an important role in the field of pattern recognition [1]. However, the high dimensionality of images poses challenges for image classification. To overcome this challenge, researchers use suitable methods to find low dimensional representations of images to improve the classification performance of images. In other words, appropriate image representation can improve classification performance [2]. Recent years, researchers have proposed some better methods in face recognition and image classification. For face recognition, Leng et al. proposed dynamic weighted discrimination power analysis (DWDPA) [3] to enhance the discrimination power of the selected discrete cosine transform coefficients properties for improving recognition performance. For image classification, Polap et al. presented a new convolutional neural network (CNN) model [4] for building systems for vehicle image analysis. The model achieves a classification accuracy of 94.78% in public datasets, exceeding the metrics of known migration learning models. Although the advantages of deep learning methods are obvious when dealing with large-scale datasets, they are less obvious when dealing with image datasets with smaller sample sizes.

For the classification of image data with small sample sizes, linear regression (LR) are one of the many simple and effective methods. LR is an important statistical analysis method that can help to better understand the nature of data and phenomena, and to make more accurate predictions and decisions. LR is commonly used in the fields of machine learning and pattern recognition, with applications such as natural language processing [5], image classification [6], [7], [8], and speech recognition [9]. LR tries to find an optimal linear mapping relationship that allows features in the source data to predict the target data. In a specific implementation, LR finds the optimal transform matrix by minimizing the

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo.

sum of squares of the residuals between the predicted and actual values [10]. This process is usually solved using the least squares method, where the value of the sum of squared residuals is gradually reduced by iteratively updating the parameters of the transform matrix until an optimal solution is found.

In recent years, researchers have proposed a large number of methods based on linear regression, e.g., support vector machine (SVM) [11], linear regression based classification (LRC) [12], for image classification and face recognition. To improve the classification accuracy, some LRC-based variant methods have been proposed. To fully utilize the information in the train space, a classifier based on LRC, called global linear regression coefficient (GLRC) [13] classifier, is proposed for recognition. To address the problem of partial face occlusion, a novel approach based on LRC and combined with chunked sampling, an improved distance-based evidence fusion method is presented [14]. Recently, subspace learning techniques and sparse techniques are also often used to improve the classification performance of LRC. To simultaneously accommodate both image representation and attribution prediction, a joint optimization framework of LR and nonnegative matrix factorization (LR-NMF) [15] based on the self-organized graph is proposed. Based on sparse and collaborative representation, kernel pairwise linear regression classification (KPLRC) [16] is used for image classification. After analysis, it can be found that the improvement of the linear regression method can effectively improve the classification accuracy.

Under certain conditions, the objective function of LR is equivalent to the objective function of linear discriminant analysis (LDA) [17]. Notice that LDA also seeks a transfer matrix that pulls samples of the same class together and pushes samples of different classes apart in the discriminant subspace [18]. In comparison to LDA, LR is more flexible and efficient. However, there are some limitations that restrict the practical application of some LR-related methods. To deal with this limitation that binary labeling matrices are stringent for regression may lead to over-fitting [19], various soft labeling techniques have been used to relax the labeling matrix [20], [21], [22]. Zhang et al. proposed reoriented least squares regression (ReLSR) [20] based on the idea of learning regression targets directly from data. Han et al. adapted the strict zero-one labeling matrix and graph regularization term to a more flexible form to avoid the over-fitting problem [21]. Zhang et al. proposed pairwise relations oriented discriminative regression (PRDR) [22] to avoid the over-fitting problem by preserving pairwise labeling relations in the latent space. To better deal with the problem of classification of noisy images, Su et al. proposed regularized denoising latent subspace based linear regression (RDLSLR) [23]. RDLSLR is used to learn the discriminative information in the data by following two steps. The first step is to obtain clean data by adding a denoised latent space between the visual space and the labelling space; the second step is to learn the regression target with clean data using another transformation matrix. Wang et al. proposed robust double relaxed regression (RDRR) [24] for noisy image classification. The idea of RDRR is to mitigate the conflict between increasing model flexibility and exacerbating the over-fitting problem through noise reduction.

To improve classification accuracy, researchers have added various regularization terms to the linear regression [25], [26], [27], [28]. By utilizing the $L_{2,1}$-norm to model the correlation between samples, Wen et al. proposed inter-class sparsity based discriminative least square regression (ICSDLR) [25]for image classification. Zhan et al. proposed group low-rank representation-based discriminant linear regression (GLRRDLR) by introducing a group low-rank constraint on the transformed data. Although these methods improve the discriminative ability of the regression to some extent, they lead to a tedious and time-consuming optimization effort. In view of the advantages of manifold learning in enhancing model discrimination, Shi et al. integrated graph embedding and sparse regression into one unified model [29], which improves the recognition performance well.

In this paper, we simultaneously study two issues discussed about LR, i.e., 1) the zero-one labeling matrix is too strict to be conducive to linear regression methods for learning labeling information; 2) how to improve the discrimination of transformed data. To address these two issues, this paper proposes a new method, namely, discriminative regression with latent label learning (DRLLL), for image classification. To relax the zero-one labeling matrix (i.e., to address the first problem), DRLLL learns labeling information in the latent label space instead of the input zero-one labeling space. To improve the discrimination of transformed data (i.e., to address the second problem), a discriminative regularization term is used to guide the transform matrix to learn the discriminative information in the data. Contributions of our paper are as follows:

1) The proposed method learns labeling information in the latent labeling space, which was found to avoid the strict regression targets problem.
2) The label-approximating manifold regularization term in the model can effectively avoid the over-fitting problem.
3) The discriminative regularization term can effectively improve the classification performance of the proposed method.
4) An iterative optimization algorithm is developed to obtain the solution of the DRLLL model.

The next sections are listed below. In Section II, the related work of DRLLL, e.g., LR, to be introduced. In Section III, the learning model of DRLLL and the developed iterative optimization will be designed. In Section IV, several groups of classification experiments on image datasets are used to verify the validity of the proposed DRLLL method. Section V presents brief conclusions.

## II. RELATED WORK

A review on the related work on linear regression is presented in this section. Before describing the related work, we provide a description of some commonly used symbols. For a training matrix $X = [x_1, x_2, \ldots, x_n] \in \mathrm{R}^{m\times n}$, where $m$ is the number of features in the training samples and $n$ is the number of training samples. The label matrix corresponding to the training samples is $Y = [y_1, y_2, \ldots, y_n] \in R^{c\times n}$, where $c$ is the number of classes. The goal of linear regression is in finding an optimal transform matrix $W \in R^{c\times m}$ to express the relationship between the training data and its corresponding labeling matrix. $I$ in the paper denotes the identity matrix of the appropriate size. For a matrix $X$, $\mathrm{Tr}(X)$ and $X^T$ denote the trace of the matrix $X$ and the transpose of the matrix $X$, respectively. $\|X\|_F = \sqrt{\sum_i \sum_j x_{i,j}^2}$ denotes the Frobenius norm of $X$, and $\|X\|_{2,1} = \sum_i \sqrt{\sum_j x_{i,j}^2}$ denotes the $L_{2,1}$-norm of $X$. In addition, the nuclear norm of $\|X\|_*$ is $\|X\|_* = \sum_i \lambda_i(X)$, where $\lambda_i(X)$ is the $i$-th singular value of $X$.

### A. STANDARD LINEAR REGRESSION

Normally, a projection matrix $W \in R^{c\times m}$ is used by linear regression to establish a connection between a training data $X \in R^{m\times n}$ and its corresponding class labels $Y \in R^{c\times n}$. To find the transform matrix $W$, the following objective function is established:

$$\min_W \|Y - WX\|_F^2 + \alpha\|W\|_F^2 \tag{1}$$

where $\alpha$ is the regularization parameter. There is a closed solution to problem (1) as $W = YX^T(XX^T + \alpha I)^{-1}$. However, linear regression method is insufficient to fulfill the needs in the face of datasets generated in different scenarios. As a result, researchers have proposed a large number of variant methods [20], [25] on the objective function of problem (1). To take advantage of the correlation in high-dimensional data, a low-rank constraint $\|W\|_*$ is used in the objective function (1) of LRLR [30] to replace the $\|W\|_F^2$ regularization term.

### B. LINEAR REGRESSION WITH REGULARIZATION TERMS

To be able to better explore the similarities in the data, the researchers incorporated the objective function (1) with different types of regularization terms. The objective function for these types of methods are:

$$\min_W \|Y - WX\|_F^2 + \alpha\psi(W) + \beta\phi(WX) \tag{2}$$

where $\psi(W)$ and $\phi(WX)$ are regularization terms. To preserve the local manifold structure in the data, a graph regularization term is assembled into $\phi(WX)$ by the method in the literature [27]. To improve intra-class compactness in data to facilitate classification, sparse and low-rank techniques are also frequently used to constrain the regularization term $\phi(WX)$. The $L_{2,1}$ norm technique is used by ICSDLR to constrain $WX$, i.e., $\phi(WX) = \sum_{i=1}^{c} \|WX_i\|_{2,1}$, in a way that the extracted features contain natural discriminability between samples. Similarly, the low-rank technique, i.e., $\phi(WX) = \sum_{i=1}^{c} \|WX_i\|_*$, is used by GLRRDLR [26] to explore the correlation between classes. However, most of the linear regression-based methods are generally time-consuming in that they require learning multiple transform matrices. In addition, the problem of strict regression targeting still exists.

## III. PROPOSED METHOD

In this section, the objective function of the proposed method, the developed optimization algorithm, and the convergence and computational complexity analysis of the optimization algorithm are presented in detail.

### A. DEVELOPMENT OF LEARNING MODEL

From the discussion of related work, it can be found that the empirical loss function and regularization term on the transform matrix in the linear regression method can improve the discriminative property of the extracted features. Therefore, to improve the classification accuracy, it is extremely important to design the discriminant regularization term and the empirical loss function. However, the zero-one labeling matrix may be too strict and its not suitable for empirical loss function modeling. To deal with this problem, we introduce a matrix of latent label $V$ in problem (3) as follows:

$$\begin{aligned} &\min_{W,V} \|V - WX\|_F^2 + \alpha\|W\|_F^2 \\ &\text{s.t.} \sum_{j=1}^{n} v_{ij} = 1. \end{aligned} \tag{3}$$

where $v_{ij}$ is the ij-th element of the matrix $V \in R^{c\times n}$ and $W \in R^{c\times m}$ is a transform matrix. The constraint $\sum_{j=1}^{n} v_{ij} = 1$ is imposed to relax the zero-one labeling matrix. In doing so, the goal is to make the model more flexible in learning the label information in the data. It is clearer that the latent label matrix is the crucial to guide the projection matrix $V$ to learn the label information in the data. In addition, the objective function (3) faces an over-fitting problem in learning latent label information. To enhance the differentiation between latent label in the data, the manifold learning theory is introduced. According to the manifold learning assumption, i.e., when two samples are located in a small local neighborhood in the manifold, they have similar class label, a manifold regularization term on the latent label is added to the objective function (3). Then, we have

$$\begin{aligned} &\min_{W,V} \|V - WX\|_F^2 + \alpha\|W\|_F^2 + \beta \sum_{i,j=1}^{c} \|v_i - y_j\|_2^2 s_{ij} \\ &\text{s.t.} \sum_{j=1}^{n} v_{ij} = 1. \end{aligned} \tag{4}$$

where $v_i$ and $y_j$ are the elements in $V$ and $Y$, respectively. Here, the label matrix $Y$ is used to avoid over-fitting of the latent label matrix $V$ in the update process. $s_{ij}$ is used to

penalize latent label in the portion of the latent label matrix that does not match the local structure of the data point. Usually, $s_{ij}$ is pre-computed based on the a priori information given in the data. In this paper, the settings for $s_{ij}$ is shown below:

$$s_{ij} = \begin{cases} 1, & \text{if } y_i = y_j, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Then, the third term in the objective function (4) can be transformed into the following trace form:

$$\sum_{i,j=1}^{c} \|v_i - y_j\|_2^2 s_{ij} = Tr(VDV^T - 2VSY^T + YDY^T) \tag{6}$$

where $D \in R^{n\times n}$ is the diagonal matrix and its i-th diagonal element is $d_{ii} = \sum_{j=1} s_{ij}$. The definition of $S \in R^{n\times n}$ is the similarity matrix used to describe the similarity of two labels. Notice that the incorporation of label-approximating manifold regularization ensures that similar data points in the original data space have similar latent label, while local structural information in the latent label space can be maintained to avoid over-fitting.

For classification purposes, one always hopes to reduce the distance between similar samples while increasing the distance between dissimilar samples. To accomplish this, a discriminant regularization term inspired by LDA is introduced as follow:

$$\min_{W} Tr(W(S_b - \mu S_w)W^T) \tag{7}$$

where $S_b \in R^{m\times m}$ and $S_w \in R^{m\times m}$ are pre-computed inter-class scatter matrices and intra-class scatter matrices, respectively. Specifically, the computation of $S_b$ and $S_w$ is as follows:

$$S_b = \frac{1}{n}\sum_{k=1}^{c} n_k(u_k - u)(u_k - u)^T \tag{8}$$

and

$$S_w = \frac{1}{n}\sum_{k=1}^{c}\sum_{i=1}^{n_k}(x_i^k - u_k)(x_i^k - u_k)^T \tag{9}$$

where $u_k$ denotes the average features of the k-th class of samples, which is calculated as $u_k = \frac{1}{n_k}\sum_{i=1}^{n_k} x_i^k$. $u$ is defined as the average feature of all samples, which is calculated as $u = \frac{1}{n}\sum_{k=1}^{c}\sum_{i=1}^{n_k} x_i^k$. By combining the descriptions in the objective function (4), Eq. (6) and Eq. (7), the overall objective function of discriminative regression with latent label learning (DRLLL) can be obtained.

$$\begin{aligned} \min_{W,V} \|V - WX\|_F^2 + \alpha\|W\|_F^2 + \beta Tr(VDV^T - 2VSY^T + YDY^T) \\ + \gamma Tr(W(S_b - \mu S_w)W^T) \\ \text{s.t.} \sum_{j=1}^{n} v_{ij} = 1. \end{aligned} \tag{10}$$

where $\alpha$, $\beta$ and $\gamma$ are balance parameters. As shown in the objective function (10), the proposed DRLLL method optimizes only the transform matrix $W$ and the latent label matrix $V$. To optimize these two variables simultaneously, an iterative optimization algorithm is introduced.

### B. EFFECTIVE SOLUTION OF THE LEARNING MODEL

In this section, a solution scheme of the proposed learning model is introduced. To facilitate the optimization of problem (10), an optimization algorithm based on the theory of iterative optimization is developed. We solve problem (10) by the following two steps: 1) by fixing $W$, updating $V$; 2) by fixing $V$,updating $W$. The procedure for solving problem (10) is as follows.

#### 1) UPDATE $V$

By fixing the variable $W$, we can update $V$ with the following sub-problem:

$$\begin{aligned} \min_{V} \|V - WX\|_F^2 + \beta Tr(VDV^T - 2VSY^T) \\ \text{s.t.} \sum_{j=1}^{n} v_{ij} = 1. \end{aligned} \tag{11}$$

Due to the constraints $\sum_{j=1}^{n} v_{ij} = 1$, we use a method similar to that used in the literature [22] to solve for $V$ through the following two steps. The first step is to remove the constraints to find the intermediate variable $\bar{V}$. The suboptimization problem (11) without the constraints is a convex optimization problem. By setting its derivative with respect to $\bar{V}$ to zero, that is

$$\bar{V} - WX + \beta(\bar{V}D - YS^T) = 0$$

Then, the closed-form solution of the optimization problem (11) without the constraints is

$$\bar{V} = (WX + \beta YS^T)(I + \beta D)^{-1} \tag{12}$$

In the second step, these constraints $\sum_{j=1}^{n} v_{ij} = 1$ are handled through the following methods. Specifically, the optimal $V$ can be calculated as:

$$V = [v_1, v_2, \cdots, v_n] \tag{13}$$

where $v_i = v_i / \sqrt{\sum_{j=1}^{c} \bar{v}_{ij}^2}$.

#### 2) UPDATE $W$

By fixing the variable $V$, we can update $W$ with the following sub-problem:

$$\min_{W} \|V - WX\|_F^2 + \alpha\|W\|_F^2 + \gamma Tr(W(S_b - \mu S_w)W^T) \tag{14}$$

In a similar way to updating $V$, by setting its derivative with respect to $W$ to zero, that is

$$WXX^T + \gamma W(S_b - \mu S_w) + \alpha W - VX^T = 0$$

Then, the closed-form solution of the optimization problem (14) can be calculated as:

$$W = VX^T(XX^T + \gamma(S_b - \mu S_w) + \alpha I)^{-1} \tag{15}$$

By executing the first and second steps alternately, the objective function value of problem (10) can be gradually minimized until convergence. It can be noted that the $(XX^T + \gamma(S_b - \mu S_w) + \alpha I)$ in (15) is fixed in the iteration, so we can compute its inverse outside the loop. The algorithm used to solve the model (10) is presented in Algorithm 1. To perform the multi-classification task, the transformed data, i.e., $WX$, is fed into a K-nearest neighbor classifier, and in this experiment $K$ is set to 1.

**Algorithm 1** DRLLL (Solving the Objective Function (10))

**Input:** Data matrix $X \in R^{m\times n}$, class label $Y \in R^{c\times n}$, parameters $\alpha, \beta, \gamma$.

**Initialization:** $W$ with random values, $\mu = 10^{-3}$;

1: Compute $S$ by using (5);
2: Compute $S_b$ by using (8);
3: Compute $S_w$ by using (9);
4: **while** not converged **do**
5: Update $V$ by solving (11);
6: Update $W$ by using (15);
7: **end while**

**Output:** Transform matrix $W \in R^{c\times m}$.

### C. COMPLEXITY AND CONVERGENCE ANALYSIS

#### 1) COMPUTATIONAL COMPLEXITY

The computational complexity of Algorithm 1 will be analyzed in this part. The computational cost of Algorithm 1 is mainly divided into outside the loop and inside the loop. On the outside of the iterative loop, the calculation cost comes from the inverse calculation of matrix $(XX^T + \gamma(S_b - \mu S_w) + \alpha I) \in R^{m\times m}$, which has a calculation cost of $O(m^3)$. In addition, the cost of computing $S$ is $O(n^2)$. In the iterative loop, the complexity of the inverse computation of the matrix $(I + \beta D)$ can be neglected since it is a diagonal matrix. Thus, the calculation of updating $V$ and $W$ costs mainly the multiplication of matrices. In summary, the total computational complexity of Algorithm 1 is $O(n^2 + m^3 + \tau \max\{cmn, cn^2\})$, where $\tau$ is the number of loop iterations.

#### 2) CONVERGENCE ANALYSIS

According to the previous introduction, all variables can simply be computed using closed-form solutions in their respective sub-problems. Also, the corresponding sub-problems corresponding to updating $W$ and $V$ are convex.

*Theorem 1:* The optimization method outlined in Algorithm 1 monotonically reduces the objective value of the problem (10).

*Proof of Theorem 1:* Let $L(V^t, W^t)$ to be the value of the objective function of the problem (10) at the t-th iteration, that is

$$\begin{aligned} L(V^t, W^t) = &\|V^t - W^t X\|_F^2 + \gamma Tr(W^t(S_b - \mu S_w)(W^t)^T) \\ &+ \alpha\|W^t\|_F^2 + \beta Tr(V^t D(V^t)^T \\ &- 2V^t S Y^T + YDY^T) \end{aligned} \tag{16}$$

In $(t+1)$-th iterations, the optimal solution of $v^{t+1}$ is first obtained by tackling the convex subproblem (11). Thus, we have:

$$L(V^{t+1}, W^t) \le L(V^t, W^t) \tag{17}$$

By similar means, the subproblem corresponding to the update $W$ is convex. Thus, we can get:

$$L(V^{t+1}, W^{t+1}) \le L(V^{t+1}, W^t) \tag{18}$$

In combination with (17) and (18), we have:

$$L(V^{t+1}, W^{t+1}) \le L(V^t, W^t) \tag{19}$$

Thus, the proof of Theorem 1 is complete. ■

Under the condition i.e., given a suitable $\mu$ such that the third term $Tr(W(S_b - \mu S_w)W^T) \ge 0$ holds, the value of the objective function in the problem (10) has a lower bound. With the above conditions satisfied, the proposed method is eventually going to find a locally optimal solution when the value of the objective function of the problem (10) converges.

## IV. EXPERIMENTS

In this section, several groups of comprehensive experiments were set up to validate the effectiveness of the proposed method. In particular, some commonly correlated methods, e.g., LDA [17], robust discriminant regression (RDR) [27], robust latent subspace learning (RLSL) [31], ICSDLR [25], robust sparse linear discriminant analysis (RSLDA) [32], PRDR [22], RDLSLR [23], RDRR [24], are used to do the classification comparison. All the methods were used to perform classification experiments on six real datasets, and the classification experiments of all the methods were repeated 10 times in a random combination of training and test samples. The average classification accuracy and bias of them are then reported for comparison.

### A. DESCRIPTION OF THE EXPERIMENTAL DATASETS

The six image datasets commonly used to test linear regression methods were used for the experiments in this paper. The general statistics for these datasets are presented in Table 1. Some sample pictures are shown in Figure 1. The sources and descriptions of these experimental datasets are given below.

1) **ORL** dataset[1]: The ORL dataset was sampled from 10 different images attached to each of 40 different objects. These images varied in time and each image was unique in terms of lighting, facial expression (e.g., eyes open and closed vs. smiling or not), and facial details (e.g., whether or not glasses were worn). In this

[1] http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html

**TABLE 1. Some details of the six datasets used for the experiments.**

| No. | Datasets | # samples | # features | # classes |
|---|---|---|---|---|
| 1 | ORL | 400 | 1024 | 40 |
| 2 | LFW | 1251 | 1024 | 86 |
| 3 | AR | 1300 | 1024 | 100 |
| 4 | YaleB | 2114 | 1024 | 38 |
| 5 | COIL100 | 7200 | 1024 | 100 |
| 6 | PIE | 11554 | 1024 | 68 |

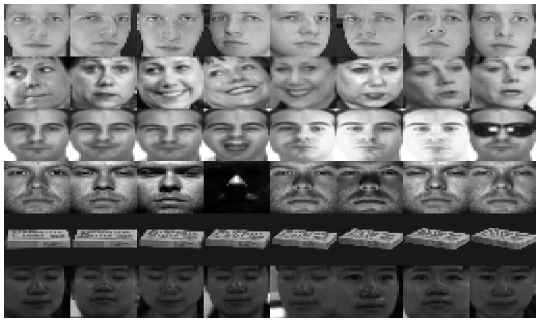

**FIGURE 1. Sample images of the datasets used in the experiment, from top to bottom sample images from, the ORL, LFW, AR, YaleB, COIL100, and PIE datasets.**

experiment, the pixels of each image were scaled to 32 × 32. There are 3, 4, 5 and 6 images randomly selected for training and the remaining images for testing for each class, respectively.

2) **LFW** dataset[2]: The LFW dataset is a challenging face dataset. The reason is that all the images in the dataset are collected directly from the web, and these samples present different poses, backgrounds, expressions, lighting, and are captured by different image acquisition devices. All the images were resized to 32 × 32 beforehand. There are 5, 6, 7 and 8 images randomly selected for training and the remaining images for testing for each class, respectively.
3) **AR** dataset[3]: The samples in the AR face dataset have the following differences, e.g., the samples have different facial expressions, the samples have different lighting conditions, have sunglasses, and the faces are obscured by scarves. All the images were resized to 32 × 32 beforehand. There are 6, 8, 10 and 12 images randomly selected for training and the remaining images for testing for each class, respectively.
4) **YaleB** dataset[4]: The YaleB dataset is a face dataset that is commonly used for classification tests. The samples for this dataset are collected from 38 different people, where each class has 59-64 frontal images with different lighting. Similarly, all the images in this dataset were resized to 32 × 32 beforehand. There are 8, 13, 18 and 23 images randomly selected for training and the remaining images for testing for each class, respectively.
5) **COIL100** dataset[5]: The COIL100 dataset is an object image dataset that is commonly used for classification tests and contains 7200 images of 100 objects. All the images in this dataset are pre-sized to 32 × 32 pixels. There are 8, 13, 18 and 23 images randomly selected for training and the remaining images for testing for each class, respectively.
6) **PIE** dataset[6]: The PIE dataset, which contains more than 11554 face images with different poses, lighting conditions and expressions, is one of the challenging databases in the field of face recognition. The samples for this dataset were collected from 68 individuals, each with 5 different poses, and nearly 170 samples were collected. Similarly, all the images in this dataset are pre-sized to 32 × 32 pixels. There are 8, 13, 18 and 23 images randomly selected for training and the remaining images for testing for each class, respectively.

### B. PARAMETER SETTING AND ANALYSIS

For linear regression based on regularization terms the methods, a suitable combination of parameters can improve the classification accuracy. In the proposed method, three parameters, e.g., $\alpha$, $\beta$ and $\gamma$, need to be tuned. To analyze the effect of these three parameters on the model, we first set up a candidate set (i.e., {0.0001, 0.001, 0.01, 0.1, 1, 10, 100}) for these three parameters. Here, the LFW, AR, YaleB and COIL100 datasets were used for this experiment. Specifically, in the LFW dataset, 5 samples in each class were randomly selected as the training set and the remaining samples were used for testing; for the AR dataset, 6 samples in each class were randomly selected as the training set and the remaining samples were used for testing; and for the YaleB and COIL100 datasets, 8 samples in each class were each selected as the training set and the remaining samples were used for testing. Figures 2-3 presents the trend of classification accuracy with different combinations of the three parameters.

As shown in Figures 2-3, adaptively selecting optimal regularization parameters for different datasets is very difficult and remains an open problem. In this study, we use the following strategy [25], [31], [32] to find the optimal parameter combination for the proposed method. The analysis shows that the proposed method is insensitive to the choice of parameter $\gamma$. Therefore, we can first fix the parameter $\gamma$ and select the optimal combination of parameters $\alpha$ and $\beta$ in the provided candidate set. Then, we fix the values of parameters $\alpha$ and $\beta$ to the values corresponding to the optimal classification accuracy obtained and perform the same experiment to obtain the optimal parameter $\gamma$. Finally, the best combination of the found parameters $\alpha$, $\beta$ and $\gamma$ was

[2]http://vis-www.cs.umass.edu/lfw/
[3]http://www2.ece.ohio-state.edu/ aleix/ARdatabase.html
[4]http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html
[5]https://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php
[6]https://www.ri.cmu.edu/project/pie-database/

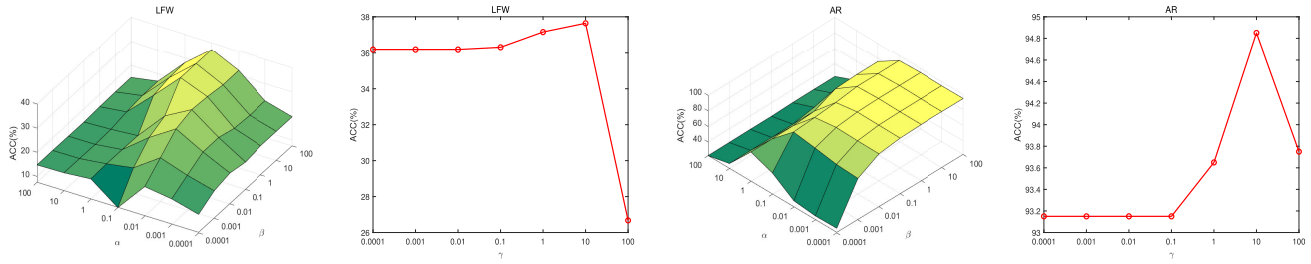


**FIGURE 2.** Classification accuracy of DRLLL versus its different combinations of parameters on the LFW and AR datasets.

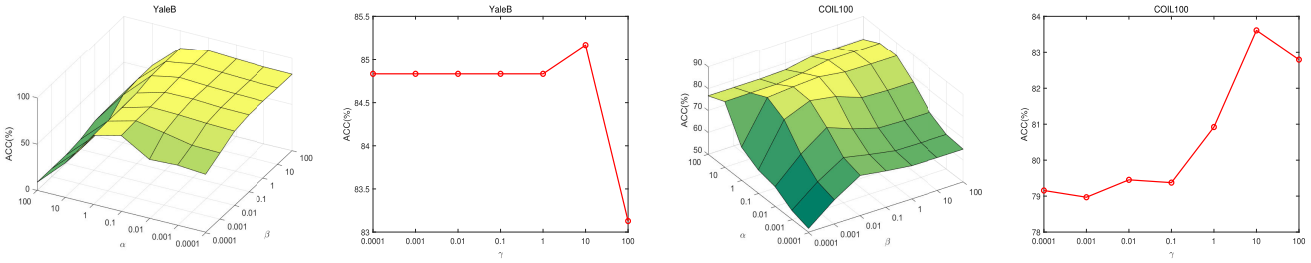


**FIGURE 3.** Classification accuracy of DRLLL versus its different combinations of parameters on the YaleB and COIL100 datasets.

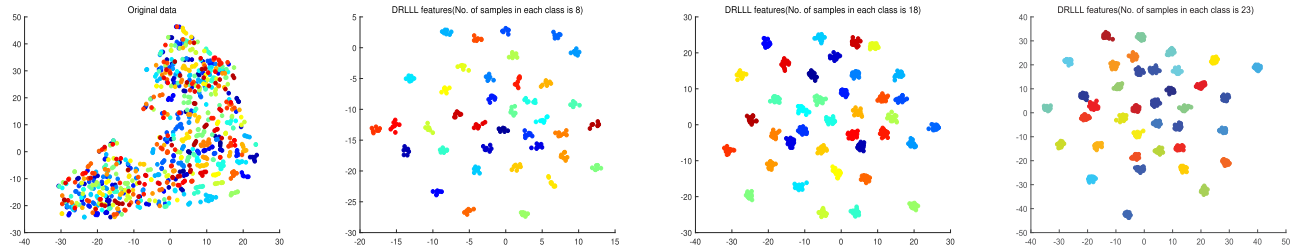


**FIGURE 4.** T-SNE visualization of original samples and learned features of DRLLL on YaleB dataset.

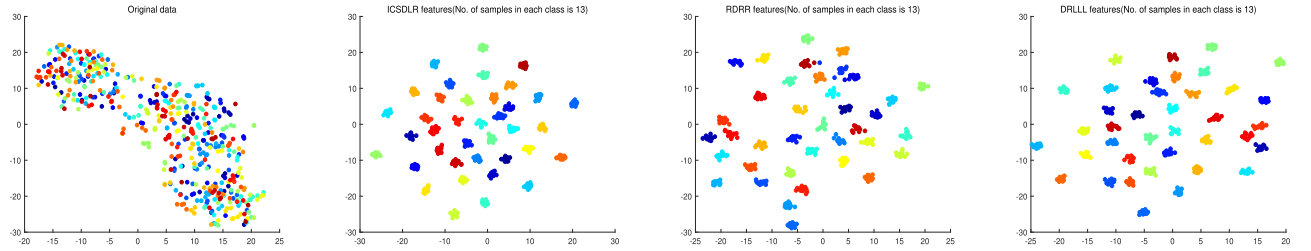


**FIGURE 5.** T-SNE visualization of original data and ICSDLR, RDRR and DRLLL features on YaleB dataset with 13 images in each class for training.

used for the subsequent experiments. For all the compared methods, their optimal parameter value ranges are referred to as given in the corresponding papers.

### C. CLASSIFICATION EXPERIMENTS

#### 1) EXPERIMENTAL SETUP

To validate the effectiveness of the proposed DRLLL method for image classification, we conducted image classification experiments on the above six datasets. In particular, some commonly correlated methods, e.g., LDA [17], RDR [27], RLSL [31], ICSDLR [25], RSLDA [32], PRDR [22], RDLSLR [23], RDRR [24], are used to do the classification comparison. Noting that the RDR, RLSL, ICSDLR, PRDR, RDLSLR, and RDRR are LR-based methods. Among these methods, RLSL and ICSDLR are learning the labeling information in the data by direct regression on the zero-one labeling matrix. In the RDR and PRDR methods,

**TABLE 2.** Acc(%) (mean±std) of the eight methods on ORL dataset.

| No. | LDA [17] | RDR [27] | RLSL [31] | ICSDLR [25] | RSLDA [32] | PRDR [22] | RDLSLR [23] | RDRR [24] | DRLLL(our) |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 55.54±3.11 | 77.86±1.86 | 84.89±2.01 | 91.07±1.21 | 84.93±1.59 | 82.46±1.69 | 91.84±1.43 | 91.25±1.78 | **92.21**±1.30 |
| 4 | 67.63±6.60 | 86.33±1.35 | 89.83±1.67 | 94.17±1.58 | 88.00±1.42 | 88.79±1.71 | 94.74±1.57 | 93.96±1.86 | **95.75**±1.12 |
| 5 | 71.45±2.34 | 91.40±1.60 | 92.35±1.78 | 96.20±0.90 | 91.85±1.12 | 92.65±2.35 | 96.95±1.56 | 96.66±1.31 | **97.20**±0.86 |
| 6 | 74.31±3.81 | 95.31±1.40 | 96.00±1.40 | 97.63±1.03 | 92.56±1.19 | 95.50±1.18 | 98.21±1.12 | 98.01±1.32 | **98.69**±0.96 |

**TABLE 3.** Acc(%) (mean±std) of the eight methods on LFW dataset.

| No. | LDA [17] | RDR [27] | RLSL [31] | ICSDLR [25] | RSLDA [32] | PRDR [22] | RDLSLR [23] | RDRR [24] | DRLLL(our) |
|---|---|---|---|---|---|---|---|---|---|
| 5 | 17.64±0.89 | 32.58±1.04 | 35.20±1.05 | 36.54±0.97 | 33.23±1.14 | 37.06±0.98 | 37.76±1.56 | 37.45±1.78 | **38.15**±1.49 |
| 6 | 21.32±1.30 | 33.60±1.35 | 39.81±1.20 | 39.82±1.21 | 35.06±1.31 | 39.01±1.62 | 39.49±1.20 | 39.55±1.36 | **40.49**±0.87 |
| 7 | 23.13±0.79 | 34.88±1.28 | 41.84±1.52 | 42.23±1.40 | 36.76±1.06 | 42.29±1.46 | 43.01±1.66 | 42.98±1.72 | **43.14**±1.48 |
| 8 | 26.25±1.04 | 35.79±1.14 | 42.88±0.98 | 43.06±2.06 | 37.37±1.51 | 43.95±1.86 | 44.10±1.78 | 44.05±1.91 | **44.14**±1.49 |

**TABLE 4.** Acc(%) (mean±std) of the eight methods on AR dataset.

| No. | LDA [17] | RDR [27] | RLSL [31] | ICSDLR [25] | RSLDA [32] | PRDR [22] | RDLSLR [23] | RDRR [24] | DRLLL(our) |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 91.77±0.44 | 83.07±0.94 | 80.18±1.64 | 91.40±0.24 | 87.59±0.53 | 87.89±0.44 | 91.83±0.57 | 91.89±0.66 | **92.65**±0.20 |
| 8 | 94.72±0.51 | 88.69±1.14 | 84.19±1.33 | 94.90±0.52 | 88.36±0.76 | 92.51±0.67 | 95.31±0.87 | 95.51±0.65 | **95.62**±0.54 |
| 10 | 96.11±0.39 | 93.45±1.07 | 87.16±1.33 | 96.42±0.32 | 89.48±1.12 | 94.87±0.62 | 96.45±0.65 | 96.62±0.45 | **96.81**±0.33 |
| 12 | 96.94±0.18 | 94.44±0.53 | 94.15±1.17 | 97.19±0.25 | 92.27±0.88 | 95.94±0.37 | 97.02±0.45 | 97.11±0.44 | **97.39**±0.24 |

**TABLE 5.** Acc(%) (mean±std) of the eight methods on YaleB dataset.

| No. | LDA [17] | RDR [27] | RLSL [31] | ICSDLR [25] | RSLDA [32] | PRDR [22] | RDLSLR [23] | RDRR [24] | DRLLL(our) |
|---|---|---|---|---|---|---|---|---|---|
| 8 | 82.83±1.01 | 80.91±0.94 | 80.37±1.41 | **84.88**±0.66 | 80.36±0.69 | 79.14±1.07 | 84.24±1.02 | 84.14±1.13 | 84.48±0.79 |
| 13 | 90.96±0.91 | 86.21±0.84 | 82.30±1.38 | 91.39±0.86 | 84.30±0.69 | 88.19±1.06 | 91.23±1.34 | 91.32±1.45 | **91.82**±0.94 |
| 18 | 92.18±0.62 | 91.91±0.71 | 84.47±1.37 | 94.19±0.66 | 90.91±0.51 | 92.21±0.74 | 94.32±0.88 | 94.21±0.79 | **94.57**±0.69 |
| 23 | 93.17±0.44 | 93.65±1.04 | 94.81±0.75 | 96.08±0.45 | 86.32±0.55 | 94.49±0.42 | 96.22±0.47 | 96.31±0.53 | **96.42**±0.40 |

**TABLE 6.** Acc(%) (mean±std) of the eight methods on COIL100 dataset.

| No. | LDA [17] | RDR [27] | RLSL [31] | ICSDLR [25] | RSLDA [32] | PRDR [22] | RDLSLR [23] | RDRR [24] | DRLLL(our) |
|---|---|---|---|---|---|---|---|---|---|
| 8 | 67.66±0.69 | 79.75±0.74 | 71.90±0.78 | 74.40±0.79 | 76.19±0.67 | 84.03±0.78 | 84.94±0.76 | 84.98±0.61 | **85.31**±0.52 |
| 13 | 78.49±0.59 | 86.51±0.50 | 81.55±0.45 | 82.66±0.56 | 83.38±0.41 | 87.29±0.55 | 88.65±0.52 | 88.78±0.67 | **88.96**±0.50 |
| 18 | 83.71±0.53 | 90.18±0.35 | 89.45±0.39 | 87.27±0.50 | 91.01±0.50 | 91.71±0.49 | 91.02±0.63 | 91.41±0.65 | **92.02**±0.27 |
| 23 | 87.61±0.22 | 92.89±0.35 | 91.71±0.65 | 90.48±0.35 | 92.62±0.42 | **94.06**±0.40 | 92.68±0.54 | 92.94±0.58 | 93.90±0.34 |

**TABLE 7.** Acc(%) (mean±std) of the eight methods on PIE dataset.

| No. | LDA [17] | RDR [27] | RLSL [31] | ICSDLR [25] | RSLDA [32] | PRDR [22] | RDLSLR [23] | RDRR [24] | DRLLL(our) |
|---|---|---|---|---|---|---|---|---|---|
| 8 | 84.71±0.68 | 83.42±0.51 | 84.81±1.59 | 84.99±0.61 | 83.63±0.81 | 84.43±0.70 | 84.89±0.51 | 84.42±0.78 | **86.94**±0.53 |
| 13 | 90.84±0.48 | 89.12±0.96 | 90.21±0.62 | 91.22±0.35 | 84.66±0.65 | 91.51±0.36 | 90.98±0.42 | 90.66±0.61 | **92.08**±0.29 |
| 18 | 93.17±0.36 | 92.42±0.54 | 92.59±0.77 | 93.59±0.28 | 88.84±0.42 | 93.67±0.38 | 92.59±0.32 | 92.47±0.45 | **94.09**±0.24 |
| 23 | 94.58±0.18 | 94.44±0.85 | 94.71±0.70 | 94.98±0.20 | 91.76±0.43 | 94.57±0.23 | 94.65±0.24 | 94.43±0.36 | **95.31**±0.19 |

**TABLE 8.** Classification results (%) (mean±std) of the nine methods on all the datasets in term of F1-score, where the number of samples used for training in each dataset is 3, 5, 6, 8, 8, and 8, respectively.

| Datasets | LDA [17] | RDR [27] | RLSL [31] | ICSDLR [25] | RSLDA [32] | PRDR [22] | RDLSLR [23] | RDRR [24] | DRLLL(our) |
|---|---|---|---|---|---|---|---|---|---|
| ORL | 52.60±3.31 | 81.75±2.27 | 84.49±2.09 | 90.97±1.31 | 84.02±1.97 | 90.77±2.74 | 90.35±2.40 | 91.59±1.92 | **92.94**±1.33 |
| LFW | 24.27±1.06 | 32.30±0.87 | 34.86±0.84 | 36.26±1.26 | 30.33±1.00 | 36.29±0.62 | 36.71±0.76 | 34.82±1.17 | **37.97**±0.95 |
| AR | 91.53±0.91 | 83.20±1.17 | 86.38±1.06 | 91.49±0.60 | 88.33±1.25 | 89.58±0.77 | 92.22±1.17 | 91.14±0.67 | **93.97**±0.62 |
| YaleB | 83.46±1.02 | 80.59±2.31 | 82.56±1.40 | 85.39±0.86 | 80.62±1.90 | 84.48±1.16 | 84.36±2.08 | 86.06±0.88 | **86.53**±0.73 |
| COIL100 | 67.09±0.76 | 80.23±0.56 | 74.80±0.63 | 73.49±0.54 | 75.53±0.63 | 84.34±0.44 | 84.17±0.81 | 82.34±0.47 | **85.76**±0.86 |
| PIE | 85.07±0.51 | 83.70±0.94 | 84.10±0.68 | 85.34±0.79 | 83.13±0.85 | 85.22±0.55 | 85.24±0.56 | 85.16±0.58 | **87.88**±0.72 |

the graph regularization term employed to explore the similarity between data. LDA and RSLDA are multi-class discriminant analysis methods which are related to the discriminant regularization term of the proposed method.

For all the comparison methods the dimension of the transformed data were set as follows. All the methods are set to $c$ except the dimension of LDA which is set to $c-1$, where $c$ is the number of classes of input data. For The experimental

**TABLE 9.** Classification results (%) (mean±std) of the nine methods on all the datasets in term of MCC, where the number of samples used for training in each dataset is 3, 5, 6, 8, 8, and 8, respectively.

| Datasets | LDA [17] | RDR [27] | RLSL [31] | ICSDLR [25] | RSLDA [32] | PRDR [22] | RDLSLR [23] | RDRR [24] | DRLLL(our) |
|---|---|---|---|---|---|---|---|---|---|
| ORL | 50.12±9.28 | 80.12±2.21 | 83.08±2.16 | 89.85±1.97 | 83.14±1.75 | 90.89±2.47 | 90.26±2.87 | 90.16±1.87 | **91.76**±1.98 |
| LFW | 24.69±1.02 | 33.81±1.34 | 34.74±1.47 | 36.15±1.61 | 30.65±0.90 | 36.74±1.72 | 36.67±1.64 | 37.30±2.13 | **37.52**±1.52 |
| AR | 92.17±0.61 | 84.93±0.96 | 86.83±0.54 | 91.46±0.68 | 89.95±1.30 | 90.65±1.37 | 92.65±1.54 | 91.91±0.81 | **94.39**±0.57 |
| YaleB | 83.21±0.98 | 85.05±1.83 | 82.54±1.25 | 85.40±0.58 | 80.68±0.69 | 84.34±0.47 | 84.92±1.26 | 85.88±0.69 | **86.66**±0.75 |
| COIL100 | 68.04±0.35 | 80.87±0.72 | 75.62±0.75 | 73.86±0.57 | 76.67±0.55 | 84.68±0.63 | 84.79±0.47 | 82.86±0.59 | **86.17**±0.49 |
| PIE | 85.61±0.75 | 83.80±1.51 | 84.49±1.35 | 85.62±0.79 | 84.59±0.99 | 85.35±0.84 | 85.34±1.64 | 85.61±0.75 | **88.15**±0.69 |

process for the classification experiment is as follows. Firstly, we input the training data into all the comparison methods to output the transform matrix. The transformed data is then fed into a 1-NN classifier for classification. To evaluate the classification accuracy obtained by all the methods, the commonly used classification accuracy metric (ACC(%) ), i.e., $ACC(\%) = N_{correct}/N_{total} \times 100\%$, is used, where $N_{correct}$ denotes the number of correctly assigned samples, and $N_{total}$ denotes the total number of samples involved in the classification. Furthermore, to increase the credibility of the experiments, we added two commonly used metrics, i.e., F1-score [33] and Matthews correlation coefficient (MCC) [33], to test the performance of all the methods. The classification experiments of all the methods were repeated 10 times in a random combination of training and test samples. All the classification results are presented in Tables 2-7.

#### 2) CLASSIFICATION RESULTS AND ANALYSIS

From the results output from the experiment, we are able to make some interesting observations. As shown in Tables 2-7, we had the following discussion.

1) In comparison with the multi-class discriminant analysis methods (e.g., LDA and RSLDA), the majority of the linear regression-based methods (e.g., RLSL, ICSDLR, PRDR and the proposed DRLLL) have achieved good classification accuracy in the ORL, LFW, YaleB, COIL100 and PIE datasets. This implies that labeling regression can effectively facilitate the transform matrix to learn the labeling information in the data.
2) By comparing DRLLL with RLSL, ICSDLR, PRDR, we can find that the classification accuracy of ICSDLR, PRDR, and DRLLL always outperforms RLSL in most of the scenarios. This is due to the fact that DRLLL, ICSDLR, and PRDR all impose regularization terms on the transform matrix to enhance the discrimination of the transformed data. This shows that imposing the regularization term on the transform matrix can improve the image classification accuracy.
3) Figure 5 shows the visualization of the DRLLL features where the number of images in each class in the training set is 8, 18 and 23 respectively. At a training size of 23 per subject, our DRLLL can achieve a classification accuracy of 96.42%. Obviously, the intra-class compactness and inter-class separability of DRLLL are greatly improved in the studied features. The discriminative power of the projection increases as the training sample increases. Figure 5 shows the visualization of original data and ICSDLR, RDRR and DRLLL features on YaleB dataset with 13 images in each class for training. The features of ICSDLR and RDRR focus on expanding the distance between different classes, and we use them as a comparison. It is clear that the features of DRLLL have greater inter-class scatter, and this suggests that DRLLL can learn more discriminative features.
4) Revisiting the conclusions of the experiments in Tables 2-7, we can find that the proposed DRLLL method always obtains optimal or sub-optimal classification accuracy in all the classification scenarios. As shown in Tables 8-10, the performance of the proposed method is also better than all the comparison methods. In addition, from Figure 4, we can clearly observe that DRLLL can ensure a close distribution of samples of the same class, and try to keep samples of different classes as far away as possible. The proposed DRLLL method obtains such good results due to the following two aspects: 1) the proposed DRLLL method to learn labeling information in the latent labeling space can effectively avoid the strict regression target problem. Also, the label-approximating manifold regularization term in the model can effectively avoid the over-fitting problem; 2) the discriminant regularization term is helpful in guiding the transform matrix in the DRLLL model to explore similar relationships in the data.

### D. COMPARISON WITH DEEP LEARNING METHODS

In this subsection, the proposed method is compared with some popular deep learning methods such as VGG16 [34], ResNet50 [35], Xception [36], and MobileNet [37]. The conclusions of all the methods on the AR and PIE datasets are displayed in Table 10.

From the Table 10, it can be observed that as the number of labels in the training set increases the performance of the deep learning methods is also improved. However, the performance of the proposed method is better than the deep learning methods. The reason for this is that deep learning based methods rely heavily on large scale data and deep learning methods require a large number of labeled datasets to train the network. Therefore, the advantages of the proposed method are more obvious compared to deep learning methods on small sample size datasets.

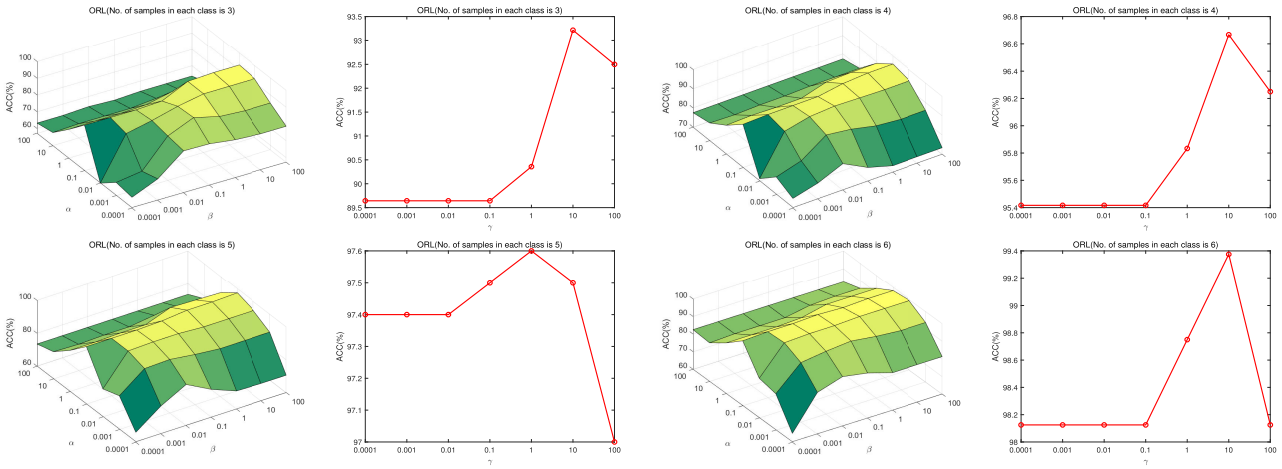


**FIGURE 6.** Classification accuracy of DRLLL versus its different combinations of parameters on the ORL dataset.

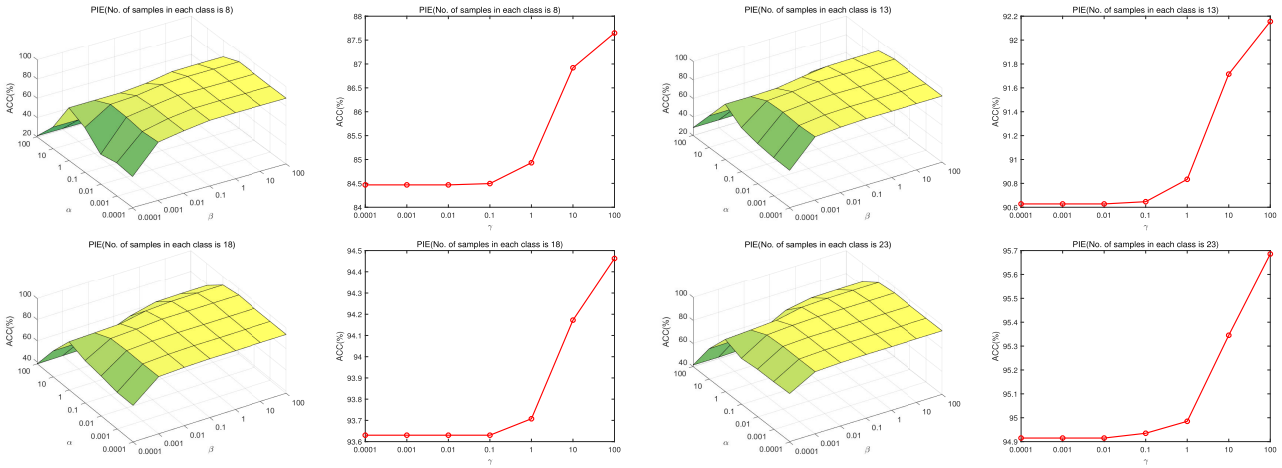


**FIGURE 7.** Classification accuracy of DRLLL versus its different combinations of parameters on the PIE dataset.

**TABLE 10.** Acc(%) (mean±std) of the methods on the AR and PIE datasets.

| AR | 8 | 10 | 12 |
|---|---|---|---|
| VGG16 [34] | 34.06±0.71 | 40.45±0.69 | 45.87±0.70 |
| ResNet50 [35] | 56.66±0.53 | 62.45±1.06 | 65.86±0.89 |
| Xception [36] | 44.78±0.49 | 58.49±0.70 | 68.19±0.72 |
| MobileNet [37] | 54.23±0.72 | 61.81±0.68 | 62.98±0.74 |
| DRLLL(our) | 95.62±0.54 | 96.81±0.33 | 97.39±0.24 |
| PIE | 13 | 18 | 23 |
| VGG16 [34] | 39.23±0.74 | 46.75±0.99 | 50.21±1.03 |
| ResNet50 [35] | 49.54±0.90 | 55.89±0.73 | 66.54±0.73 |
| Xception [36] | 38.23±0.85 | 46.23±0.77 | 55.41±0.53 |
| MobileNet [37] | 46.89±0.65 | 51.31±0.59 | 59.56±0.54 |
| DRLLL(our) | 92.08±0.29 | 94.09±0.24 | 95.31±0.19 |

### E. HYPERPARAMETER ANALYSIS

The proposed DRLLL method has three hyperparameters, i.e., $\alpha$, $\beta$ and $\gamma$. In this subsection, the three hyperparameters can be tested and explained. The AR and PIE datasets are used to tested. Figures 6-7 show the relationship between hyperparameters and classification accuracy. We can find that the classification accuracy of the proposed DRLLL method is insensitive to the parameter of the number of samples of each class in the training set. For the hyperparameter $\gamma$, we can observe that a better classification accuracy is obtained for DRLLL when the value of the hyperparameter $\gamma$ is around 10. When the value of the hyperparameter $\beta$ is in the set $\{0.001, 0.01, 0.1, 1, 10, 100\}$, DRLLL can obtain a better classification accuracy. We can clearly know that DRLLL is sensitive to the hyperparameter $\alpha$. However, we can observe that better classification accuracy is obtained for values of the hyperparameter $\alpha$ around 1 for DRLLL.

### F. CONVERGENCE ANALYSIS

The convergence theory of the proposed DRLLL method is provided. In this subsection, convergence experiments of the proposed DRLLL method with $\mu = 10^{-3}$ are continued to be studied. Figure 8 shows the convergence curves of the proposed method on ORL, LFW, YaleB and COIL100 datasets. We can find that the objective function value of the

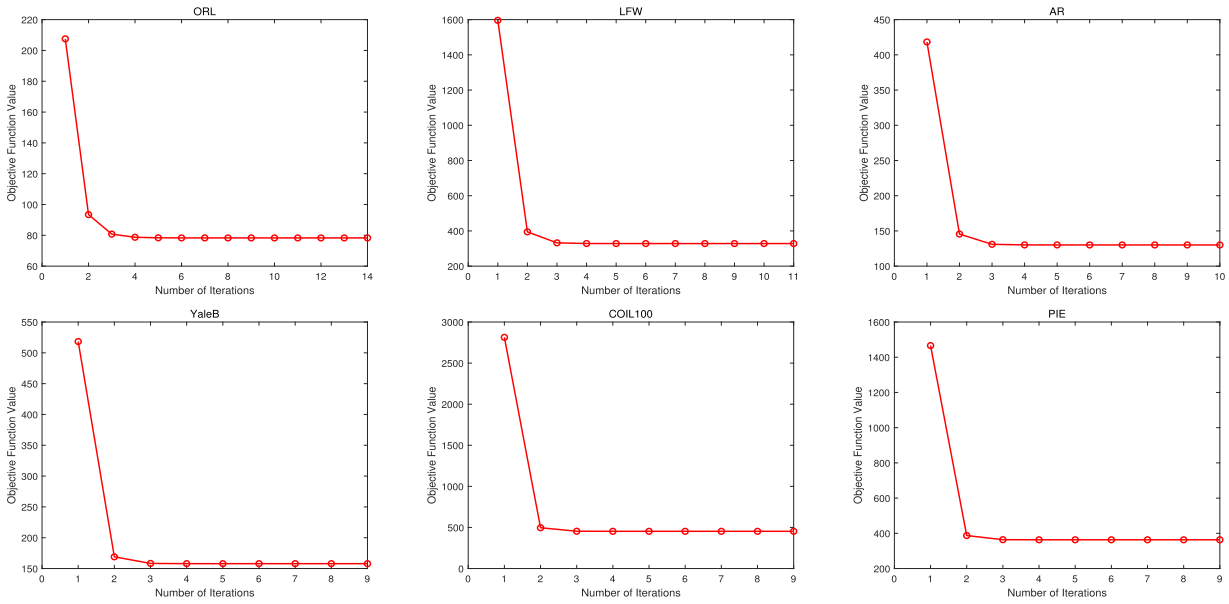


**FIGURE 8.** Convergence curve versus the number of iterations on all the datasets.

proposed method can reach the steady state very quickly. This implies that the proposed iterative optimization algorithm in fast and effective.

## V. CONCLUSION

In this paper, we proposed a new linear regression based methods, namely DRLLL, for image classification. The objective function of the proposed DRLLL method can be viewed as a framework which can be used to simultaneously learn both latent labeling information and discriminative information from the data. On several real-world classification experiments on image datasets, the classification accuracy obtained by the proposed method outperforms that of the related methods. The reasons why the proposed method obtains such good image classification results are the following: 1) the target regressed by DRLLL is the latent label matrix rather than the zero-one label matrix; 2) the introduced discriminant regularization term can effectively guide the transformation matrix to explore the discriminant information in the data. In addition, the techniques devised in this paper can be extended to other regression-based methods. However, similar to many LR-related methods, the proposed method has two drawbacks, i.e., 1) it may be sensitive to anomalies or noise features in the data and 2) it requires more time to adjust the three hyperparameters in the method. Therefore, we will continue to investigate robust linear regression methods with fewer hyperparameters in the future.

## REFERENCES

[1] K. Yu, Y. Zhu, X. Yin, T. Shu, Y. Wang, and E. Hu, "Structure-aware preserving projections with applications to medical image clustering," *Appl. Soft Comput.*, vol. 158, Jun. 2024, Art. no. 111576.

[2] 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.

[3] L. Leng, J. Zhang, J. Xu, M. K. Khan, and K. Alghathbar, "Dynamic weighted discrimination power analysis in DCT domain for face and palmprint recognition," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Nov. 2010, pp. 467–471.

[4] D. Połap, A. Jaszcz, K. Prokop, and G. Srivastava, "Lightweight CNN based on spatial features for a vehicular damage detection system," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2023, pp. 5097–5101.

[5] S. P. G. P. L. Raja and R. V. Venkatesan, "The grammatical structure used by a Tamil lyricist: A linear regression model with natural language processing," *Soft Comput.*, vol. 27, no. 23, pp. 18215–18225, Oct. 2023.

[6] S. T. H. Shah, S. A. Qureshi, A. U. Rehman, S. A. H. Shah, A. Amjad, A. A. Mir, A. Alqahtani, D. A. Bradley, M. U. Khandaker, M. R. I. Faruque, and M. Rafique, "A novel hybrid learning system using modified breaking ties algorithm and multinomial logistic regression for classification and segmentation of hyperspectral images," *Appl. Sci.*, vol. 11, no. 16, p. 7614, Aug. 2021.

[7] X. Shu, J. Tang, G.-J. Qi, Z. Li, Y.-G. Jiang, and S. Yan, "Image classification with tailored fine-grained dictionaries," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 2, pp. 454–467, Feb. 2018.

[8] X. Fang, N. Han, G. Zhou, S. Teng, Y. Xu, and S. Xie, "Dynamic double classifiers approximation for cross-domain recognition," *IEEE Trans. Cybern.*, vol. 52, no. 4, pp. 2618–2629, Apr. 2022.

[9] D. K. Kim and M. J. F. Gales, "Noisy constrained maximum-likelihood linear regression for noise-robust speech recognition," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 2, pp. 315–325, Feb. 2011.

[10] A. Guan, C.-H. Lin, and P.-W. Chi, "Secure collaborative computing for linear regression," *Appl. Sci.*, vol. 14, no. 1, p. 227, Dec. 2023.

[11] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Netw.*, vol. 17, no. 1, pp. 113–126, Jan. 2004.

[12] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2106–2112, Nov. 2010.

[13] Q. Feng, X. Zhu, and J.-S. Pan, "Global linear regression coefficient classifier for recognition," *Optik*, vol. 126, no. 21, pp. 3234–3239, Nov. 2015.

[14] W. Xu, S. H. Lee, and E. J. Lee, "A robust method for partially occluded face recognition," *KSII Trans. Internet Inf. Syst.*, vol. 9, no. 2, pp. 2667–2682, Jan. 2015.

[15] W. Zhu and Y. Yan, "Joint linear regression and nonnegative matrix factorization based on self-organized graph for image clustering and classification," *IEEE Access*, vol. 6, pp. 38820–38834, 2018.

[16] X. Gao, Q. Sun, H. Xu, and J. Gao, "Sparse and collaborative representation based kernel pairwise linear regression for image set classification," *Expert Syst. Appl.*, vol. 140, Feb. 2020, Art. no. 112886.
[17] J. Ye, "Least squares linear discriminant analysis," in *Proc. 24th Int. Conf. Mach. Learn.*, Jun. 2007, p. 1087.
[18] Q. Wang, Z. Meng, and X. Li, "Locality adaptive discriminant analysis for spectral–spatial classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2077–2081, Nov. 2017.
[19] S. Xiang, F. Nie, G. Meng, C. Pan, and C. Zhang, "Discriminative least squares regression for multiclass classification and feature selection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 11, pp. 1738–1754, Nov. 2012.
[20] X.-Y. Zhang, L. Wang, S. Xiang, and C.-L. Liu, "Retargeted least squares regression algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 9, pp. 2206–2213, Sep. 2015.
[21] N. Han, J. Wu, X. Fang, W. K. Wong, Y. Xu, J. Yang, and X. Li, "Double relaxed regression for image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 2, pp. 307–319, Feb. 2020.
[22] 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.
[23] Z. Su, W. Wenbo, and W. Zhang, "Regularized denoising latent subspace based linear regression for image classification," *Pattern Anal. Appl.*, vol. 26, no. 3, pp. 1027–1044, Aug. 2023.
[24] W. Wang, L. Fang, and W. Zhang, "Robust double relaxed regression for image classification," *Signal Process.*, vol. 203, Feb. 2023, Art. no. 108796.
[25] J. Wen, Y. Xu, Z. Li, Z. Ma, and Y. Xu, "Inter-class sparsity based discriminative least square regression," *Neural Netw.*, vol. 102, pp. 36–47, Jun. 2018.
[26] S. Zhan, J. Wu, N. Han, J. Wen, and X. Fang, "Group low-rank representation-based discriminant linear regression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 760–770, Mar. 2020.
[27] Z. Lai, D. Mo, W. K. Wong, Y. Xu, D. Miao, and D. Zhang, "Robust discriminant regression for feature extraction," *IEEE Trans. Cybern.*, vol. 48, no. 8, pp. 2472–2484, Aug. 2018.
[28] Y. Zhang, W. Li, H.-C. Li, R. Tao, and Q. Du, "Discriminative marginalized least-squares regression for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3148–3161, May 2020.
[29] X. Shi, Z. Guo, Z. Lai, Y. Yang, Z. Bao, and D. Zhang, "A framework of joint graph embedding and sparse regression for dimensionality reduction," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1341–1355, Apr. 2015.
[30] A. Argyriou, T. Evgeniou, and M. Pontil, "Convex multi-task feature learning," *Mach. Learn.*, vol. 73, no. 3, pp. 243–272, Jan. 2008.
[31] 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.
[32] J. Wen, X. Fang, J. Cui, L. Fei, K. Yan, Y. Chen, and Y. Xu, "Robust sparse linear discriminant analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 2, pp. 390–403, Feb. 2019.
[33] D. Chicco, M. J. Warrens, and G. Jurman, "The Matthews correlation coefficient (MCC) is more informative than Cohen's Kappa and brier score in binary classification assessment," *IEEE Access*, vol. 9, pp. 78368–78381, 2021.
[34] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent.*, 2015, pp. 1–11.
[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[36] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.
[37] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

**LANG LANG** is currently the Chief Judge of national-level professional skills competitions, a Chongqing Science and Technology Envoy, a member of Chongqing Vocational Education Industry Guidance Committee, and the Leader of the Exemplary Curriculum and Distinguished Teacher Team of ideological and political education courses in Chongqing. Additionally, he is an evaluator and training Lecturer of sensor network application development, a backbone Teacher, a dual-qualified Teacher, and the person in charge of high-quality blended courses for universities in Chongqing. His research interests include machine learning, image processing, and agricultural anomaly detection. He is a member of Chinese Electronics Society and holds various important roles and responsibilities in the field of education and technology.

**XIAO QIN CHEN** received the master's degree. She is currently a Lecturer. Her research interests include machine vision technology and applications and sensor application technology.

**SHA LIU** is currently pursuing the Ph.D. degree with the National Key Laboratory of Wireless Communications (NKLWC), University of Electronic Science and Technology of China (UESTC). Her current research interests include artificial intelligence and wireless communications.

**QIANG ZHOU** received the Ph.D. degree from the School of Computer Science and Technology, Chongqing University of Posts and Telecommunications. His research interests include machine learning, image processing, and defect detection.