# Face Image Recognition Algorithm Based on Label Complementation

*Jiakang Tang, Lin Cui*, Zhiwei Zhang*

*Suzhou University, China*
*tangjk1212@ahszu.edu.cn, cl@ahszu.edu.cn, zhiwei.zhang@ahszu.edu.cn*

## Abstract

In face image recognition, labels play a fairly important role in recognition and classification, and rich and perfect labels can greatly improve the accuracy rate. However, it is almost impossible for the labels in the image to be recognized to describe the image completely and accurately. At the same time, the data obtained when feature extraction is performed on an image inevitably extracts a large amount of redundant and useless information at the same time, which affects the generalization performance of the model. Accordingly, we propose a face image recognition algorithm based on label completion in multi label learning. First, the SVD algorithm is used to remove redundant and useless information from the features of the original data by dimensionality reduction operation to obtain simplified sample attribute information, and the label completion algorithm is used to supplement the labels of the images using the extracted feature information. Finally the obtained label data as complete as possible is put into the extreme learning machine to construct the face recognition model and give the prediction results of the images. Experiments on the ORL dataset demonstrate that the algorithm can achieve good recognition results.

**Keywords:** Multi label, Label complementation, Singular value decomposition, Face image recognition

## 1 Introduction

This paper is further extended and modified from our paper Face Image Recognition Algorithm based on Singular Value Decomposition [1] presented at the DSA National Conference. Face recognition is the use of computer processing problems to compare different face images to find the same image, a process that requires matching externally acquired data with known data. Such face recognition technology has a wide range of applications in today's society, including unlocking electronic devices and finding the trajectory of a specific person.

Computer recognition of faces is achieved by comparing the features of the acquired data with the known image features, so that the ability to obtain high quality features is the key issue for efficient recognition. Generally speaking, the quality of the features directly determines the accuracy of recognition, but careful feature advancement of the image will greatly enhance the feature dimension, and high latitude features are extremely likely to induce dimensional disasters. At the same time, the really effective part of a large number of features does not increase linearly, and then we need to select the extracted features which is to reduce the dimensionality of the features. In order to reduce the invalid and redundant features in the face image and achieving an improvement in stability of the features [2]. The Singular Value Decomposition (SVD) algorithm is employed to apply dimensionality reduction on face image features and transform the original features to improve the validity of the data.

No matter what kind of feature extraction method can obtain the features always can not cover all the sample information, which will cause the sample label missing problem, and solving the label missing problem can greatly improve the algorithm performance. Currently, some researchers propose that if two instance spaces are similar, then the label space between instances will also have high similarity. Also feature-to-feature correlation should be considered. If two features are correlated in the feature space, it means that these two features can share the related label subsets. Here it is assumed that $d_i$ features and $d_j$ features are correlated, $y_i$ labels have di features and $d_j$ features with high probability, and $y_j$ labels do not have $d_i$ features and do not have $d_j$ features with high probability. The above correlation principle is used to complete the labels.

In this paper we use the ORL face image dataset for face image recognition research experiments, and we label the images in this dataset with multivariate labels [3]. By labeling all known labels we convert the dataset into a multi-label dataset so that we can easily use the evaluation criteria of multi-label algorithms for comparing the experimental results [4].

Based on the above analysis, we propose an algorithm that first performs the missing label complementation and then recognition. In the first step, the SVD [5] algorithm is used to perform the feature dimensionality reduction of the image, in which the singular value operation of the original data is performed first, and then only the first K singular values are selected as valid features [6]. A label-completion algorithm using feature correlation, instance correlation, and label correlation is also proposed to input the results of the obtained reduced-dimensional valid feature data to complete the missing labels in the face image. In this way we get as many labels as possible for a face image, and next a classification model is built using an extreme learning mechanism, and this model can classify and recognize face images. This label-completion algorithm also compares to the existing multi-label algorithms ML-KNN [3] and ML-

*Corresponding Author: Lin Cui; E-mail: cl@ahszu.edu.cn
DOI: 10.53106/160792642023112406007

RBF [7], and the performance of the label-completion face image recognition algorithm is tested using the ORL dataset. The results obtained from the algorithm runs show that the algorithm is effective in both predicting image classes and performing multi-label learning.

# 2 Multi-label Learning

### 2.1 Algorithm Definition

So called multi-labeling learning [8] is the process of assigning a sample data to as many labels as possible, and these labels are used to accurately describe that sample. Suppose there exists an input-output pair $D = \{(x_i, Y_i | 1 \le i \le n)\}$ in the sample space $X = R^D$, where the $P$ dimensional feature vector of the $i$ sample is represented by $x_i = [x_{i1}, x_{i2}, ..., x_{ip}]$ a set of labeled vectors corresponding to $x_i$ is represented by $Y_i = [y_{i1}, y_{i2}, ..., y_{ij}]$, and the $j$ label of the $i$ sample is expressed as $y_{ij}$, if the corresponding label exists in the input sample then $y_{ij} = 1$, otherwise $y_{ij} = 0$ [9].

In multi-label learning we need to construct a multi-label classifier $f: X \to 2^Y$, let the sample to be classified be X so that its attributes are $x_i \in X$, and use the classifier $f$ to give the set of category tokens belonging to the sample $f(x) \subseteq Y$.

### 2.2 Label Complementation Algorithm

Considering that if there is a strong correlation between two features in an instance, then the instance has one of the features with a high probability of having the other feature, otherwise it has a high probability of not having the other feature. It can be seen that the use of correlation between features is beneficial for extracting class attributes，Based on this, we propose a label-completion algorithm that combines multi category correlation information(LCMCC). The following model is obtained:

$$\min_{w,c} \frac{1}{2}\|XWC - Y\|_F^2 + \lambda_1 \|W\|_1 + \lambda_2 tr(WLW^T) + \lambda_3 \|C\|_* + \frac{\lambda_4}{2} tr((XW)^T L_1 (XW)). \tag{1}$$

$\boldsymbol{X}$ denotes the training feature matrix $X \in i^{n \times d}$, W denotes the weight matrix $W \in i^{d \times l}$, Y: denotes the label matrix $Y \in i^{n \times l}$ of the input training set. C denotes the label correlation matrix $C \in i^{l \times l}$. n denotes the number of instances, l denotes the number of labels, and d denotes the number of features. tr(g) denotes the number of traces of the matrix, $\|\cdot\|_*$ denotes the kernel parametrization, L denotes the graph Laplacian matrix $L=D-C$, $D$ is the diagonal matrix, the diagonal elements are the row sums of $C$. $L_1$ denotes the graph Laplacian matrix about the instance correlation, $L_1=D_1-S$, $D_1$ is the diagonal matrix, the diagonal elements are the row sums of $S$, $S$ is:

$$s_{i,j} = \begin{cases} \exp(-\frac{\|x_i, x_j\|^2}{\sigma^2}) & x_i \in N_k(x_j) \quad or \quad x_j \in N_k(x_i) \\ 0 & otherwise \end{cases}. \tag{2}$$

In the first term of the model a classification objective function is used to perform classification learning, the second term $l_1$ parametric sparsity is used for class attribute learning [10], the third term is a low-rank label relevance matrix for missing labels for completing, and the fourth term is instance relevance introduction for ensuring that labels and instances maintain structural consistency. The model works by first constructing the base model using least squares loss, while considering the extraction of label-specific features using $l_1$ parametric to constrain the weights. Secondly label relevance can effectively improve the performance of the multi-label classifier, however, since the labeled samples may not be sufficient, label relevance is learned by using the features of the training data and the known labels.

In this paper $C$ is randomly initialised and used as a trainable parameter for optimisation. Considering the problem that the correlation matrix is of low rank in the absence of labels, approximate optimisation is performed by introducing a kernel parametrization. Considering that features between strongly correlated labels are also strongly correlated, constraint $tr(WLW^{\mathrm{T}})$ is introduced. Considering that the more correlated instances are, the more similar their features are, instance correlation is introduced to provide additional information. Constraint $tr((XW)^{\mathrm{T}}\mathrm{L}_1(XW))$ is added according to the streamwise regularity.

The model is solved using alternating iterations, first fixing the parameters $C$ to solve for $W$ and then fixing the parameters $W$ to solve for C until the iteration termination condition is satisfied. Here the objective function is divided into two parts $f$(x) and $g$(x) [11].

$$\begin{aligned} f(x) &= \min_w \frac{1}{2}\|XWC - Y\|_F^2 + \lambda_2 tr(WLW^T) \\ &\quad + \frac{\lambda_4}{2} tr((XW)^T L_1 (XW)) \\ g(x) &= \|W\|_1 . \end{aligned} \tag{3}$$

From equation (3), The proximal gradient descent method is used for the optimal solution of the variable $W$. We can obtain:

$$\begin{aligned} \frac{\partial f(x)}{\partial W} &= X^T XWCC^T - X^T YC^T \\ &\quad + \lambda_2 W(L + L^T) + \lambda_4 X^T L_4 XW. \end{aligned} \tag{4}$$

The next step is to solve for C, which leads to the following equation:

$$\min_c \frac{1}{2}\|XWC - Y\|_F^2 + \lambda_3 \|C\|_* . \tag{5}$$

$\|g\|_*$ denotes the kernel parametrization, which forces the label correlation matrix C to be of low rank. To optimise the above equation directly, which is more difficult, we introduce an auxiliary variable Z in place of C, transforming equation xx into the following equivalent problem.

$$\min_{Z.C.\Lambda} \frac{1}{2}\|XWC - Y\|_F^2 + \lambda_s \|Z\|_* + \Phi(\Lambda, C - Z). \quad (6)$$

$$\Phi(\Lambda, C - Z) = \frac{\mu}{2}\|C - Z\|_F^2 + \langle \Lambda, \Lambda - C \rangle. \quad (7)$$

where $\langle \cdot, \cdot \rangle$ defines the inner prodct of matrices, $\Lambda$ is the Lagrange multiplier matrix and μ is the positive penalty scalar. This allows us to decompose the above problem into the following two subproblems.

Sub Question C, Fixing other variables, we update C. by solving the following problems:

$$C^* = \arg\min_C \frac{1}{2}\|XWC - Y\|_F^2 + \Phi(\Lambda, C - Z). \quad (8)$$

Taking the derivative of C and setting it to zero, we can update C with the following rule.

$$C^* = (W^T X^T XW + \mu I)^{-1}(W^T X^T Y + \Lambda + \mu Z). \quad (9)$$

Subproblem Z: In order to update the auxiliary variable Z, we have solved the following problem.

$$\begin{aligned} Z^* &= \arg\min_Z \lambda_3 \|Z\|_* + \Phi(\Lambda, C - Z) \\ &= \arg\min_Z \frac{\lambda_3}{\mu}\|Z\|_* + \frac{1}{2}\|C + \frac{\Lambda}{\mu} - Z\|_F^2. \end{aligned} \quad (10)$$

### 2.3 Multi-label Learning Evaluation Metric

Since the multilabel learning framework is trained using known features and labeled samples, based on this feature researchers have proposed the following five commonly used evaluation metrics in multilabel learning evaluation: the Average Precision (AP), Coverage (CV), Hamming Loss (HL), Ranking Loss (RL), and One-Error (OE), [12] to judge the effectiveness of the algorithm.

The Average Precision (AP) is used to determine the probability that the marker before any relevant marker is also a relevant marker. It is optimal when the value is 1, i.e., the higher the value, the better the performance. The formula is:

$$AP_D(f) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}\sum_{y \in Y}\frac{\left|\{rank_f(x_i, y') \le rank_f(x_i, y), y' \in Y_i\right|}{rank_f(x_i, y)}. \quad (11)$$

The Coverage (CV) is an indication of whether all markers in a sample sequence are labelged, and a smaller value indicates the best performance. The equation is:

$$CV_D(f) = \frac{1}{n}\sum_{i=1}^{n}\max_{y \in Y_i} rank_f(x_i, y) - 1. \quad (12)$$

The Hamming Loss (HL) is used to represent the case when there are no tokens that can be correctly classified. When the value of this indicator is zero, the algorithm results are optimal, i.e., the smaller the value of this indicator, the better. The equation is:

$$HL_D(h) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y}\left|h(x_i) \ne Y_i\right|. \quad (13)$$

The Ranking Loss (RL) is used to examine the relevance of tokens and samples in multi-token learning. Ideally, the ranking of tokens related to the samples must be higher than that of unrelated tokens. It is optimal when the value of this item is zero. The equation is:

$$RL_D(f) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|Y_i||\overline{Y}_i|}\left|\left\{(y_1, y_2)\middle| \begin{matrix} f((x_i, y_1), (y_1, y_2)) \\ \in Y_i \times \overline{Y}_i \end{matrix}\right\}\right|. \quad (14)$$

The One-Error (OE) is an indication of the case where the topmost marker of the sample sequence is not correctly marked. When the value of this indicator is zero, the performance is optimal. The equation is:

$$OE_D(f) = \frac{1}{n}\sum_{i=1}^{n}\left[\left[\arg\max_{y \in Y} f(x_i, y)\right] \notin Y_i\right]. \quad (15)$$

## 3 SVD Algorithm

Then there exists a decomposition that can be implemented to decompose a training matrix $A \in R^{m \times n}$, which is assumed to have a training matrix containing m samples and n features, into the following expression:

$$A_{m \times n} = U_{m \times m}\Sigma_{m \times n}V_{n \times n}^T. \quad (16)$$

Therefore, the so-called singular value decomposition is the decomposition of the above matrix A into the form of the product of three matrices U, Σ, $V^T$ multiplied by each other [13].

The content of information contained in the features can be determined from the SVD calculation results, and in general the larger the calculated value is the more information content [14]. During the calculation, we always find the result that the calculation after a certain large singular value (in this case k) yields a small singular value. We can assume that the first k features in this data set are much more important for the sample than the other features. Thus using the feature of singular value decomposition we can complete the extraction of the main features in the data [15] removing the unimportant and redundant features.

# 4 Label Completion for Face Image Recognition Algorithm

## 4.1 SVD Based Data Dimensionality Reduction Process

The SVD decomposition method can be used to convert a known matrix of order $m \times n$, matrix $R$, into a product of three matrices, as shown in Equation (17):

$$R_{m\times n} = U_{m\times r} \cdot S_{r\times r} \cdot V_{r\times n}. \tag{17}$$

Here, the orthogonal matrix of $m \times r$ is $U$, the orthogonal matrix of $r \times n$ is $V$, the diagonal matrix is $S$, and the singular values are the elements on the diagonal. The result of the first of these singular values is the largest, and the direction it represents contains the most information. The amount of useful information contained in the data can be derived from the change in the size of the singular values. Therefore, using the first k values can effectively describe the sample data set and achieve dimensionality reduction of the data features.

Data feature dimensionality reduction using SVD algorithm is the feature dimensionality reduction of data information by extracting local features from high-dimensional features using the characteristics of matrix singular values [16]. Singular value decomposition has the ability to process data under the global optimum, and the top 20% to 30% of the feature values can effectively describe most of the information in the dataset, so this paper uses SVD for multi label datasets to achieve dimensionality reduction of data by calculating the feature values of multi-labeled feature data and retaining the top 20 percent of the largest feature values to achieve dimensionality reduction of multi-labeled data by SVD [17].

Let $R_D$ be the result of dimensionality reduction of $m \times n$ original order matrix R. Then $R_D$ is a matrix of order $m \times k$, where k<<n.

## 4.2 Face Image Recognition Algorithm Based on Label Complementation

Here we propose a Face Image Recognition Algorithm Based on Label Complementation (FIRLC). Extreme learning machine is an efficient classification framework, and this paper is based on the classification performance of ELM for classification and recognition of face images.

For any one sample, in the multi-label learning algorithm $x_i = [x_{i1}, x_{i2}, ..., x_{ip}]$ eigenvectors with $p$ dimensions, feature dimensionality reduction is performed using SVD, and The sample with reduced dimensionality can be written as $x'_i = [x'_{i1}, x'_{i2}, ..., x'_{iq}]$, It can be found that $q$ is the reduced dimension, and $q$ is smaller than $p$. Then the original dataset $D = \{(x_i Y_i | 1 \le i \le n)$ and the dimensionality reduction dataset is $D' = \{(x_i' Y_i | 1 \le i \le n)$.

The matrix is represented in the extreme learning machine [18] as

$$H \cdot \beta = T. \tag{18}$$

We can obtain an extreme learning machine where $H$ is the implicit layer node, $\beta$ is the output weight, and $T$ is the output expectation:

$$\beta = H^+ \cdot T. \tag{19}$$

The output model of the extreme learning machine can be deduced from the above expression:

$$f(X_i) = \sum_{i=1}^{N} h(X_i)\beta. \tag{20}$$

The $h(x_i)$ vector is an L-dimensional feature vector, which is obtained by mapping the original data from the input space to the L-dimensional space, $\beta$ is the weight vector. By KKT condition, it is possible to derive the output weights, described as follows:

$$\beta = H^T \left(\frac{1}{C} + HH^T\right)^{-1} T. \tag{21}$$

$C$ is a regular function, so the multi label output function can be expressed as:

$$f(x) = HH^T \left(\frac{1}{C} + HH^T\right)^{-1} T. \tag{22}$$

The traditional ELM algorithm requires setting the number of hidden layer nodes, and the weights and biases are also given randomly by the algorithm itself, which can lead to unstable calculation results, and the effect of the randomly given values can be amplified by the algorithm itself. The introduction of kernel function [19] can solve this drawback. Here the kernel matrix Ω is used instead of the random matrix $H$ in ELM, where the kernel matrix is:

$$\Omega_{ELM} = h(x_i) \cdot h(x_j) = K(x_i, x_j). \tag{23}$$

$K(x_i, x_j)$ is the kernel function of the extreme learning machine, such that we can obtain a kernel extreme learning machine using the kernel function:

$$f(x) = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}^T \left(\frac{1}{C} + \Omega\right)^{-1} T. \tag{24}$$

The algorithm takes the mapping relationship between known features and known labels obtained from the training dataset and uses it to predict the output labels of the test set. Figure 1 is the structural flowchart of the algorithm.

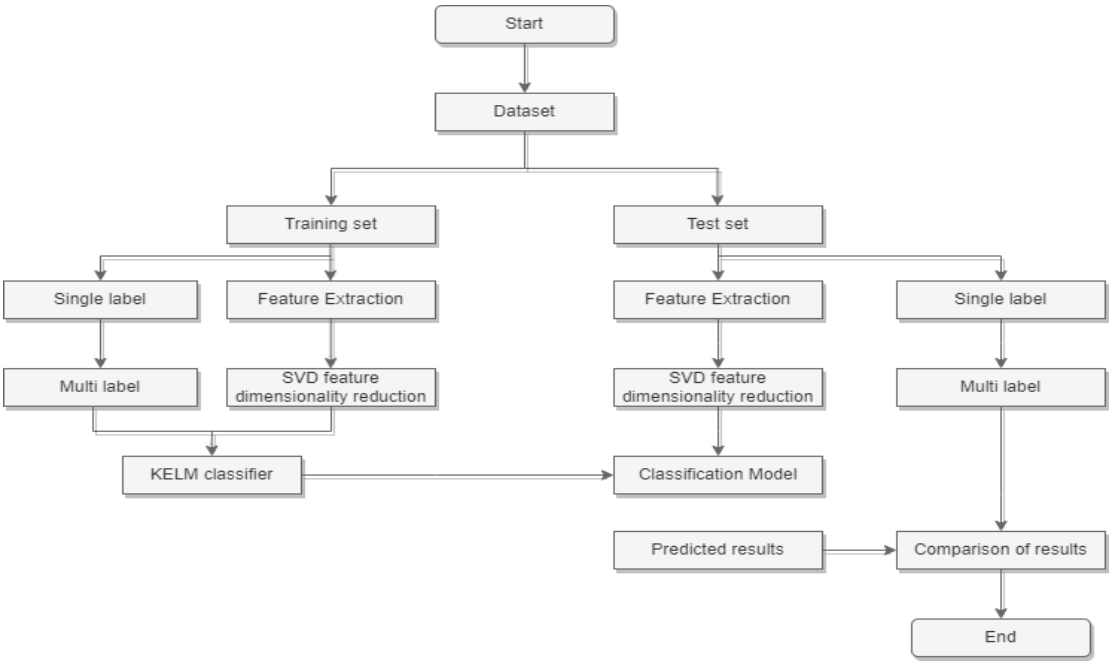


**Figure 1.** FIRLC framework work process

# 5 Experiments and Analysis of Results

## 5.1 Experimental Data Set

We experimentally illustrate the performance of the algorithm in this paper, four commonly used multi label datasets, namely Arts, Business, Recreation and Education, were chosen for the experiments in this paper. These datasets can help us to verify the validity of LCMCC. The details are shown in Table 1. In this paper, face recognition experiments are conducted using the dataset of ORL to verify whether the algorithm can recognize the images correctly, which consists of forty images of faces with different genders and ages, 10 pictures of a human with different facial expressions form a group, and Figure 2 shows a part of the images in the dataset.

**Table 1.** Multi-label datasets

| Dataset | Training set | Test set | Label | Attribute | Domain |
|---|---|---|---|---|---|
| Arts | 2000 | 3000 | 26 | 462 | Text |
| Business | 2000 | 3000 | 30 | 438 | Text |
| Recreation | 2000 | 3000 | 22 | 606 | Text |
| Education | 2000 | 3000 | 33 | 550 | Text |

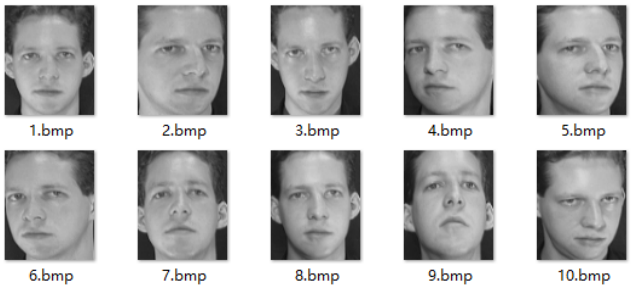


**Figure 2.** ORL dataset (partial)

## 5.2 Experimental Comparison Scheme and Analysis

The experiments in this paper evaluate the performance of the algorithm by a total of 5 evaluation metrics, AP, HL, CV, OE, and RL. Under the same conditions and evaluation metrics, the performance differences of a total of three algorithms, LCMCC, ML-RBF, and ML-KNN, in terms of multi-label learning were first compared and evaluated, as shown in Table 2 to Table 5. Table 6 gives the performance comparison results of our algorithm FIRLC and the comparison algorithm on face recognition. We write the best one of the results in bolded font.

The results of the LCMCC algorithm on the ORL dataset for multi-label experiments are placed in Table 2. The symbol "↑" indicates that the larger the value of the metric, the better the performance that the algorithm has, and the symbol "↓" indicates that the smaller the value of the metric, the better the performance that the algorithm has.

**Table 2.** Business data set test results

| Business | LCMCC | ML-RBF | ML-KNN |
|---|---|---|---|
| HL↓ | **0.0240** | 0.1978 | 0.1980 |
| OE↓ | **0.1018** | 0.2372 | 0.2345 |
| CV↓ | 6.5351 | 6.4963 | 6.4144 |
| RL↓ | **0.0305** | 0.1736 | 0.1715 |
| AP↑ | **0.8966** | 0.7586 | 0.7585 |

**Table 3.** Arts data set test results

| Arts | LCMCC | ML-RBF | ML-KNN |
|---|---|---|---|
| HL↓ | **0.0529** | 0.0542 | 0.0593 |
| OE↓ | **0.4468** | 0.4759 | 0.5503 |
| CV↓ | **4.1702** | 5.7788 | 4.8050 |
| RL↓ | **0.1025** | 0.1498 | 0.1279 |
| AP↑ | **0.6350** | 0.6066 | 0.5697 |

**Table 4.** Recreation data set test results

| Recreation | LCMCC | ML-RBF | ML-KNN |
|---|---|---|---|
| HL↓ | **0.0518** | 0.0547 | 0.0594 |
| OE↓ | **0.4340** | 0.4712 | 0.5617 |
| CV↓ | **0.1625** | 4.5297 | 4.2997 |
| RL↓ | **0.1198** | 0.1549 | 0.1550 |
| AP↑ | **0.6523** | 0.6201 | 0.5613 |

**Table 5.** Health data set test results

| Health | LCMCC | ML-RBF | ML-KNN |
|---|---|---|---|
| HL↓ | **0.0309** | 0.0355 | 0.0367 |
| OE↓ | **0.2356** | 0.2602 | 0.3040 |
| CV↓ | **0.0796** | 3.7748 | 2.7877 |
| RL↓ | **0.0394** | 0.0587 | 0.0473 |
| AP↑ | **0.8030** | 0.7812 | 0.7562 |

**Table 6.** Test results

| Algorithm | RL↓ | RL↓ | OE↓ | CV↓ | AP↑ |
|---|---|---|---|---|---|
| ML-KNN | 0.9900 | 0.0137 | 0.1825 | 0.5275 | 0.8897 |
| ML-RBF | 0.9750 | 0.0028 | 0.0500 | 0.1075 | 0.9691 |
| SVD-KELM | 0.9750 | 0.0022 | 0.0325 | 0.0875 | 0.9802 |
| FIRLC | **0.9543** | **0.0019** | **0.0317** | **0.0843** | **0.9872** |

Table 2 to Table 5 present data on the results of the LMCC algorithm on 4 different dataset. Results of the experiment of LCMCC algorithm on label complementation outperform the comparison algorithm on most of the metrics, and only the CV metrics on the Business dataset are weaker than the comparison algorithm. Table 6 shows the results of the FIRLC algorithm and the experiments of the comparison algorithm on the face dataset. The analysis results found that the FIRLC algorithm improved each metric over the comparison algorithm in multi-label evaluation indicators, and the results obtained by this paper's method also improved compared to the SVD-KELM algorithm, a research work at the DSA International Conference. The algorithm can realize feature dimensionality reduction for data with lots of features, and the labeling of the descended data is supplemented with relatively excellent results. Through comparison, it is found that FIRLC algorithm can achieve better results in face image classification, which proves that FIRLC is advantageous in helping to enhance image classification and improve the accuracy of face recognition.

### 5.3 Analysis of Real Data Results

The label complementation and classification performance of the algorithm have been experimented on the multi-label dataset and the face dataset previously, and the recognition experiments are performed on the images in the face dataset next. In the experiments, the training set is four randomly selected images and the test set is the remaining images, and the results are shown in Figure 3 to Figure 6.

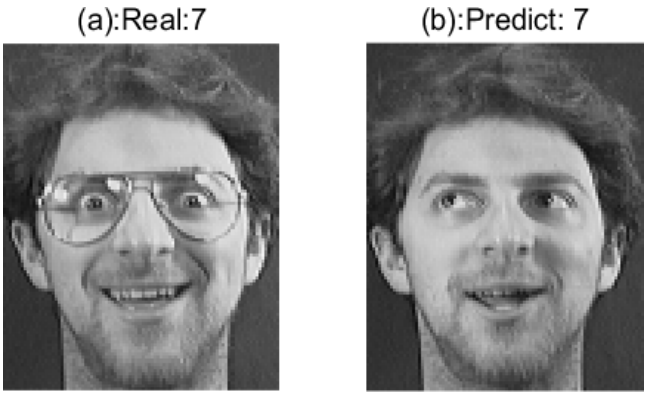


**Figure 3.** Group 7 real portrait recognition results

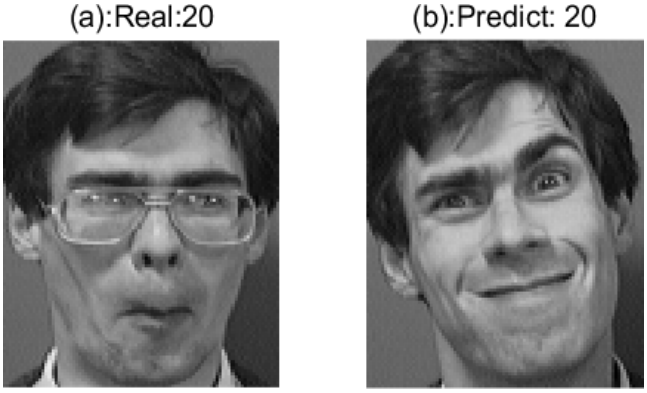


**Figure 4.** Group 20 real portrait recognition results

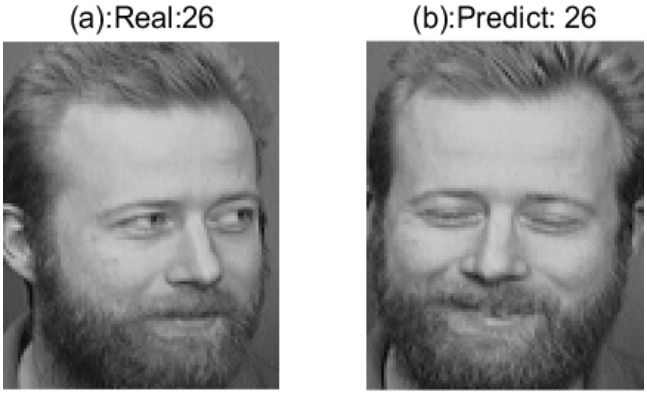


**Figure 5.** Group 26 real portrait recognition results

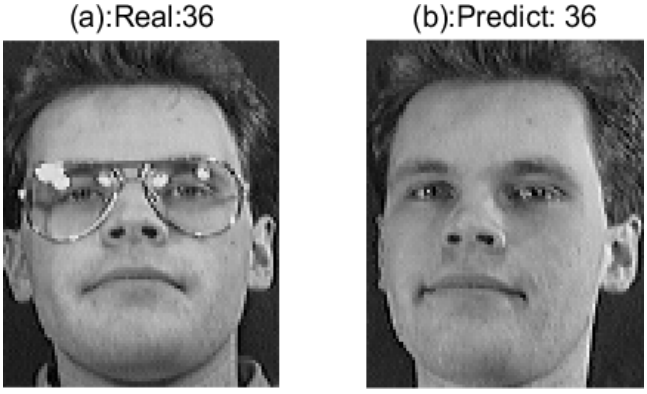


**Figure 6.** Group 36 real portrait recognition results

The following conclusions can be drawn from the analysis of the experimental results in Figures 3 through 6:

Four randomly selected images from the face dataset are used as the test set, and the part a of each figure is the image to be tested and the category it belongs to. The b part of each figure is the category group number predicted by the algorithm. In order to illustrate the experimental results more clearly, the images of the predicted group numbers are shown here directly, it is easy to find out whether the results obtained by this algorithm are correct by directly viewing the images. By observing the results of different groups, it can be found that the FIRLC algorithm can get the correct prediction result based on the original image in terms of the category to which the test image belongs. All the above experimental results show that the algorithm in this paper can improve the recognition effect of face images.

## 6 Conclusion

In this paper, we propose a face recognition algorithm based on label completion. The algorithm starts with a feature dimensionality reduction of the original data by SVD to eliminate useless and redundant features, then the extracted processed features are put into the LMCC algorithm to complete the labels of the images, and then the completed face image data are put into the FIRLC algorithm to recognize the images. The algorithm achieves relatively excellent results on common multi-label data sets and gives correct recognition results on real image data sets ORL, and the experimental results prove that the LMCC and FIRLC algorithms are feasible and effective.

## Acknowledgment

This work was supported in part by the Key Research and Development Project of Anhui Province in China under grant 202004b11020023, in part by the Academic support project for top-notch talents in disciplines (majors) of colleges and universities at Anhui Province in China under grant gxbjZD21081, Scientific research platform open project of Suzhou University under grant 2022ykf03, in part by the school-level key disciplines of computer science and technology at Suzhou University in China under grant 2019xjzdxk1, School Scientific Research Platform of Suzhou University under grant 2021XJPT50, Innovative and Entrepreneurial Teaching and Learning Research Project of Suzhou University under grant szxy2022ccjy04.

## References

[1] J. Tang, L. Cui, Z. Pan, C. Tan, S. Li, W. Wang, Face Image Recognition Algorithm based on Singular Value Decomposition, *2022 9th International Conference on Dependable Systems and Their Applications (DSA)*, Wulumuqi, China, 2022, pp. 693-697, doi: 10.1109/DSA56465.2022.00098.

[2] J. Miao, K. I. Kou, D. Cheng, W. Liu, Quaternion higher-order singular value decomposition and its applications in color image processing, *Information Fusion*, Vol. 92, pp. 139-153, April, 2023.

[3] M.-L. Zhang, Z.-H. Zhou, A review on multi-label learning algorithms, *IEEE transactions on knowledge and data engineering*, Vol. 26, No. 8, pp. 1819-1837, August, 2014.

[4] Z. A. Kakarash, F. Mardukhia, P. Moradi, Multi-label feature selection using density-based graph clustering and ant colony optimization, *Journal of Computational Design and Engineering*, Vol. 10, No. 1, pp. 122-138, February, 2023.

[5] T. Zhu, W. Qu, W. Cao, An optimized image watermarking algorithm based on SVD and IWT, The Journal of Supercomputing, Vol. 78, No. 1, pp. 222-237, January, 2022.

[6] H. Lv, Optical coherence tomography image despeckling based on edge feature-guided higher-order singular value decomposition, *Optics and Lasers in Engineering*, Vol. 160, Article No. 107314, January, 2023.

[7] X. Xu, D. Shan, S. Li, T. Sun, P. Xiao, J. Fan, Multi-label learning method based on ML-RBF and laplacian ELM, *Neurocomputing*, Vol. 331, pp. 213-219, February, 2019.

[8] D. Zhao, H. Li, Y. Lu, D. Sun, D. Zhu, Q. Gao, Multi-label weak-label learning via semantic reconstruction and label correlations, *Information Sciences*, Vol. 623, pp. 379-401, April, 2023.

[9] K. Qian, X.-Y. Min, Y. Cheng, F. Min, Weight matrix sharing for multi-label learning, *Pattern Recognition*, Vol. 136, Article No. 109156, April, 2023.

[10] T. T. Nguyen, T. T. T. Nguyen, A. V. Luong, Q. V. H. Nguyen, A. W.-C. Liew, B. Stantic, Multi-label classification via label correlation and first order feature dependance in a data stream, *Pattern recognition*, Vol. 90, pp. 35-51, June, 2019.

[11] W. Weng, B. Wei, W. Ke, Y. Fan, J. Wang, Y. Li, Learning label-specific features with global and local label correlation for multi-label classification, *Applied Intelligence*, Vol. 53, No. 3, pp. 3017-3033, February, 2023.

[12] J. Zhang, H. Wu, M. Jiang, J. H. Liu, S. Z. Li, T. Yang, J. Y. Long, Group-preserving label-specific feature selection for multi-label learning, *Expert Systems with Applications*, Vol. 213, Article No. 118861, March, 2023.

[13] L. Li, P. Wang, P. Gu, Q. Xie, Distributed singular value decomposition recommendation algorithm based on LU decomposition and alternating least square, *Pattern Recognition and Artificial Intelligence*, Vol. 33, No. 1, pp. 32-40, January, 2020

[14] M. Zhang, W. Ding, Y. Li, J. Sun, Z. Liu, Color image watermarking based on a fast structure-preserving algorithm of quaternion singular value decomposition, *Signal Processing*, Vol. 208, Article No. 108971, July, 2023.

[15] M. Begum, J. Ferdush, M. S. Uddin, A Hybrid robust watermarking system based on discrete cosine transform, discrete wavelet transform, and singular value decomposition, *Journal of King Saud University-*

*Computer and Information Sciences*, Vol. 34, No. 8, pp. 5856-5867, September, 2022.

[16] Y.-W. Wen, K. Li, H. Chen, Accelerated matrix completion algorithm using continuation strategy and randomized SVD, *Journal of Computational and Applied Mathematics*, Vol. 429, Article No. 115215, September, 2023.

[17] M. Diwakar, P. Kumar, P. Singh, A. Tripathi, L. Singh, An efficient reversible data hiding using SVD over a novel weighted iterative anisotropic total variation based denoised medical images, *Biomedical Signal Processing and Control*, Vol. 82, Article No. 104563, April, 2023.

[18] X. Liu, Y. Zhou, W. Meng, Q. Luo, Functional extreme learning machine for regression and classification, *Mathematical Biosciences and Engineering*, Vol. 20, No. 2, pp. 3768-3792, January, 2023.

[19] L. Tao, T. Cao, Q. Wang, D. Liu, O. Bai, J. Sun, Application of self-adaptive multiple-kernel extreme learning machine to improve MI-BCI performance of subjects with BCI illiteracy, *Biomedical Signal Processing and Control*, Vol. 79, Article No. 104183, January, 2023.

## Biographies

**Jiakang Tang** received his master's degree in statistics from Anqing Normal University and is currently working in Suzhou University. His current research interests include data mining, machine learning and statistical analysis.

**Lin Cui** received her M.S. degree in computer application technology from Hefei University of Technology, and she received her Ph.D. degree in computer science and technology from Nanjing University of Aeronautics and Astronautics. Her current research interests include personalized recommendation, data mining and analysis etc.

**Zhiwei Zhang** received his B.S. degree in computer science and technology from Suzhou University, and he received his M.S. degree in computer application technology from Kunming University of Science and Technology. he received his Ph.D. degree in computer science and technology from South China University of Technology. His current research interests include social network analysis, data mining and analysis.