https://doi.org/10.2352/ISSN.2470-1173.2018.10.IMAWM-339


# Learn a Hybrid Collaborative Representation for Fine-Grained Image Classification

***Wen-Yang Xie[1]; College of Information and Control Engineering, China University of Petroleum; Qingdao, 266580, China***
***Bao-Di Liu; College of Information and Control Engineering, China University of Petroleum; Qingdao, 266580, China***
***Xue Li; Department of Electronic Engineering, Tsinghua University; Beijing 100084, China***
***Yan-Jiang Wang; College of Information and Control Engineering, China University of Petroleum; Qingdao, 266580, China***

## Abstract

*Image classification has attracted more and more interest over the recent years. Consequently, a number of excellent non-parametric classification algorithms, such as collaborative representation based classification (CRC), have emerged and achieved superior performance to parametric classification algorithms. However, for fine-grained image classification task, both the class specific attributes and the shared attributes play significant roles in describing the image. CRC scheme does not consider the characteristics and merely utilizes all attributes without separation to represent an image. In this paper, we propose a hybrid collaborative representation based classification method to describe an image from perspective of the shared features, as well as the class specific features. Moreover, to reduce the representation error and obtain precise description, we learn a dictionary for hybrid collaborative representation with the training samples. We conduct extensive experiments on fine-grained image datasets to verify the superior performance of our proposed algorithm compared with the conventional approaches.*

## I Introduction

Image classification [1, 2, 3] is one of the most popular topics recently and has been attracting more and more attention. In the past decades, a sea of visual recognition methods emerged. Generally speaking, the conventional visual recognition methods can be categorized into two types. One is parametric methods and the other is non-parametric methods. Recently, non-parametric methods attract thousands of scholars and researchers due that it is easy to implement, avoid over-fitting, and superior performance compared with parametric classification algorithms.

One of the classical non-parametric classifiers are the nearest subspace methods. The principle of such classifier is to assign a test sample to the class which is closest to it. Richard *et al.* introduced a nearest neighbor method [4] to predict the label of a test image using its nearest neighbors in the training samples.Then the nearest subspace method [5] which assigns the label of a test image by comparing its reconstution error for each category was proposed by Tao *et al.*. Wei *et al.* [6] proposed to classify images through comparing the reconstruction error of each category. Wright *et al.* [7] described a sparse representation based classification (SRC) system and achieved impressive performance for face recognition. Given a test sample, the sparse representation technique [2] represents it as a sparse linear combination of the train samples. The predicted label is determined by the residual error from each class. After that, Zhang *et al.* [8] put forward an efficient face classification scheme, which is the widely-used collaborative representation based classification (CRC). A little different from SRC, CRC represents a test sample as the linear combination of almost all the training samples. Moreover, they demonstrated that it was the collaborative representation rather than the sparse representation that makes the nearest subspace method powerful for classification. After the SRC and CRC are introduced, a variety of nearest subspace method [5] was proposed to enhance the visual recognition performance. Yang *et al.*[9] learned a dictionary for each class with sparse coefficients and applied it for face recognition. Wang *et al.* [10] introduced a modified sparse model and a supervised class-specific representation method for classification. Liu *et al.* [11, 12, 13] proposed a class specific dictionary learning (CSDL) based representation algorithm which can find the intrinsic relationship between the base vectors and the original image features. Wang *et al.*[14] proposed a label constrained specific representation approach to preserve the structural information in the feature space. Cai *et al.* [15] proposed a probabilistic collaborative representation based classification (ProCRC) method. The probabilistic collaborative representation based classification method employed a probabilistic collaborative representation framework to jointly maximize the probability that a test sample belongs to each class.

In this paper, we propose a hybrid collaborative representation learning method (Hybrid-CRC), which characterizes a test sample with both shared representation and class specific representation under a learnt hybrid dictionary, to classify fine-grained image. **The main contribution** is listed in three aspects:

- We propose a hybrid collaborative representation based classification method to describe an image from perspective of the shared features, as well as the class specific features. Such representation can increase the accuracy because of the inherent attributes of fine-grained image.
- We also learn a hybrid dictionary with the training samples to reduce the reconstruction error and obtain precise description.
- We conduct extensive experiments on fine-grained image datasets to verify the superior performance of our proposed algorithm compared with the conventional approaches.

The rest of the paper are organized as follows. Section II overviews the two classical visual recognition algorithms. Section III proposes our hybrid collaborative representation based classification. Then, experimental results and analysis are shown in

[1]2958674461@qq.com

**Algorithm 1** Algorithm for CRC

**Require:** Training samples $X \in \mathbb{R}^{D \times N}$, $\beta$, and test sample $y$

Code $y$ with the dictionary $X$ via collaborative representation Eqn. (1).

**for** $c = 1; c \leq C; c{+}{+}$ **do**

  Compute the residuals $e^c(y) = \|y - X^c \hat{s}^c\|_2^2$

**end for**

$id(y) = \arg\, min_c \{e^c\}$

**return** $id(y)$

section IV. Finally, discussions and conclusions are drawn in section V.

## II Overview of CRC and CSDL

In this section, we overview two related algorithms, including collaborative representation based classification (CRC) and class specific dictionary learning based collaborative representation (CSDL).

### Overview of CRC

Collaborative representation algorithm can be considered as method of rearranging the structure of the original data in order to make the representation compact and discriminative under non-orthogonal bases. Hence, the data vector is represented as a linear combination of active basis vectors.

Zhang *et al.* proposed the collaborative representation based classification (CRC) algorithm [8] for robust image recognition. Specifically, given the training samples $X = [X^1, X^2, \cdots X^C] \in \mathbb{R}^{D \times N}$, $X^c \in \mathbb{R}^{D \times N_c}$ represents the training samples from the $c_{th}$ class, $C$ represents the number of classes, $N_c$ represents the number of training samples in the $c_{th}$ class ($N = \sum_{c=1}^{C} N_c$), and $D$ represents the dimensions of the samples. Supposing that $y \in \mathbb{R}^{D \times 1}$ is a test sample, the collaborative representation algorithm aims to solve the following objective function,

$$\hat{s} = \arg\, min_s \left\{ \|y - Xs\|_2^2 + \beta \|s\|_2^2 \right\}. \tag{1}$$

Here, $\beta$ is the regularization parameter to control the trade-off between fitting goodness and collaborative property (i.e., multiple entries in $X$ participating in representing the test sample).

The collaborative representation based classifier is to find the minimum value of the residual error for each class:

$$id(y) = \arg\, min_c \|y - X^c \hat{s}^c\|_2^2. \tag{2}$$

The procedure of CRC is shown in Algorithm 1. The residual error $e_c$ in Algorithm 1 is associated with most of the images in class $c$.

### Overview of CSDL

CRC algorithm directly uses the training samples as the dictionary and encode the test sample $y$ as

$$y \approx XWs, \tag{3}$$

where $W \in \mathbb{R}^{N \times N}$ is an identity matrix. This means that the training samples contribute equally for constructing the dictionary $B = XW$ when representing the test sample $y$. To make $W$ more adaptive, it would be of great benefit to impose that the training samples of the same class have different weights when constructing bases in the corresponding dictionary while the training samples have no contribution when constructing bases in the different classes of dictionary. Liu *et al.* [12] proposed the class specific dictionary learning based collaborative representation algorithm for robust image recognition.

The objective function of CSDL becomes

$$\begin{aligned} &\mathscr{G}(W^1, \cdots, W^C, S^1, \cdots, S^C) = \sum_{c=1}^{C} \left\{ \|X^c - X^c W^c S^c\|_F^2 + \beta \|S^c\|_F^2 \right\} \\ &s.t.\, \|X^c W^c_{\bullet k}\|_2^2 \leq 1, \forall k = 1, 2, \cdots, K, \forall c = 1, 2, \ldots, C. \end{aligned} \tag{4}$$

where $\|\bullet\|_F^2$ represents the Frobenius norm. $B_{\bullet i}$ and $B_{j \bullet}$ denote the $i_{th}$ column and $j_{th}$ row vectors of matrix $B$, respectively. $W$ is the learned weight coefficient for constructing the dictionary and $S$ is the corresponding collaborative representation.

## III Our Approach

In this section, we propose a hybrid collaborative representation learning method, which characterizes a test sample with both shared representation and class specific representation under a learned hybrid dictionary.

### Learn a Hybrid Collaborative Representation

The motivation of our proposed hybrid collaborative representation are as follows.

#### Class specific collaborative representation

For collaborative representation based classification algorithm, all training samples are concatenated together as the base vectors to form a training set space. That is to say, all the training samples participate in describing the test sample, whatever classes in the training set. The shared part(e.g., facial attributes in common) can be well approximated by all the training samples, while the class specific part (e.g., facial attributes belongs to specific person) can be well represented by the training samples in the same class as the test sample. Therefore, it is necessary to obtain the collaborative representation with specific class for the test sample, as follows,

$$\hat{s} = \arg\, min_s \sum_{c=1}^{C} \left\{ \|y - X^c s^c\|_2^2 + \gamma \|s^c\|_2^2 \right\} \tag{5}$$

#### Hybrid collaborative representation

Class specific collaborative representation is advantageous to obtain the description of the common part and the distinctive part for the test sample. However, such description will generate high residual error and instability, especially for the description by the training samples that are from the different class with the test sample. To reduce the residual error, enhance the instability, and increase the distinctiveness, we propose a hybrid collaborative representation algorithm, as follows,

$$\hat{s} = \arg\, min_s \left\{ \|y - Xs\|_2^2 + \lambda \|s\|_2^2 + \tau \sum_{c=1}^{C} \left\{ \|y - X^c s^c\|_2^2 + \gamma \|s^c\|_2^2 \right\} \right\} \tag{6}$$

In Eqn.(6), the first two terms are the conventional collaborative representation and the latter two terms are the class specific collaborative representation. The conventional collaborative representation guarantees the residual error and robustness, while the class specific collaborative representation obtains the distinctiveness via different classes. Eqn. (6) can be further arranged as follows,

$$\hat{s} = \arg min_s \left\{ \|y - Xs\|_2^2 + \beta \|s\|_2^2 + \tau \sum_{c=1}^{C} \left\{ \|y - X^c s^c\|_2^2 \right\} \right\} \tag{7}$$

Here, $\beta = \lambda + \tau \times \gamma$.

***Learn a dictionary for hybrid collaborative representation***

The approach mentioned above performs classification without training procedures (i.e., the training samples directly use for predicting the labels). By contrast, our approach compensates this deficiency by introducing a hybrid dictionary learning and assuming that different samples contribute unevenly in constructing the corresponding dictionary. The objective function of our proposed hybrid dictionary learning then becomes

$$\begin{aligned} f(W,S) &= \|X - XWS\|_F^2 + \beta \|S\|_F^2 \\ &+ \tau \textstyle\sum_{c=1}^{C} \|X - X^c W^c [0 \ \cdots \ S^c \ \cdots \ 0]\|_F^2 \\ &s.t. \|X^c W_{\bullet k}^c\|_2^2 \le 1, \forall k = 1,2,\cdots,K, \forall c = 1,2,\ldots,C. \end{aligned} \tag{8}$$

In Eqn.(8), the first term represents the fitness term with the whole training samples. The second term is a regularizer term to enable more training samples to participate in describing the $X$. The third term refers to the fitness term with class specific training samples. $W = \begin{bmatrix} W^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W^c \end{bmatrix}$ is the learned weight matrix for constructing the dictionary. $S = \begin{bmatrix} S^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & S^C \end{bmatrix}$ is the corresponding collaborative representation.

After removing the constant term, Eqn.(8) can be simplified as the following objective function,

$$\begin{aligned} &f(W^1,...,W^c,S^1,...,S^c) = \\ &(1+\tau) \textstyle\sum_{c=1}^{C} \{\|X^c - X^c W^c S^c\|_F^2 + \frac{\beta}{1+\tau} \|S^c\|_F^2\} \\ &s.t. \|X^c W_{\bullet k}^c\|_2^2 \le 1, \forall k = 1,2,\cdots,K, \forall c = 1,2,\ldots,C. \end{aligned} \tag{9}$$

## Optimization of the objective function

In this section, we focus on solving the optimization of the objective functions proposed in the last section.

***Learning a dictionary for hybrid collaborative representation***

The optimization problem of learning a dictionary is not jointly convex in both $W^c$ and $S^c$, but is separately convex in either $W^c$ or $S^c$ with $S^c$ or $W^c$ fixed. So the objective function can be optimized by alternating minimization to two optimization subproblems as follows.

- With fixed $W^c$, the objective function of finding collaborative representation $S^c$ can be written as an $\ell_2-$ regularized least-squares ($\ell_2 - ls$) minimization subproblem:

$$\min_{S^c} f(S^c) = \|X^c - X^c W^c S^c\|_F^2 + \frac{\beta}{1+\tau} \|S^c\|_F^2 . \tag{10}$$

Eqn. (10) can be easily solved by derivation and its analytical solutions is

$$S^c = \left( W^{cT} X^{cT} X^c W^c + \frac{\beta}{1+\tau} I \right)^{-1} W^{cT} X^{cT} X^c \tag{11}$$

- With fixed $S^c$, the objective function of learning weight $W^c$ can be written as an $\ell_2-$ constrained least-squares ($\ell_2 - ls$) minimization subproblem:

$$\begin{aligned} &\min_{W^c} f(W^c) = \|X^c - X^c W^c S^c\|_F^2 \\ &\qquad s.t. \left\|X^c W_{\bullet k}^g\right\|_2^2 \le 1, \ \forall k = 1,2,\cdots,K. \end{aligned} \tag{12}$$

Ignoring the unrelated terms, Eqn. (12) can be simplified as

$$\begin{aligned} f(W^c) &= -2 \sum_{k=1}^{K} [S^c (X^{cT} X^c)]_{k\bullet} W_{\bullet k}^c + \sum_{k=1}^{K} W_{\bullet k}^{cT} [X^{cT} X^c W^c S^c S^{cT}]_{\bullet k} \\ &s.t. \|X^c W_{\bullet k}^c\|_2^2 \le 1, \forall k = 1,2,\cdots,K. \end{aligned} \tag{13}$$

We optimize each column of $W^c$ alternately. Specifically, the Lagrangian is

$$\begin{aligned} \mathscr{L}(W^c,\lambda_k) &= \sum_{k=1}^{K} W_{\bullet k}^{cT} [X^T X W^c S^c S^{cT}]_{\bullet k} - 2 \sum_{k=1}^{K} [S^c X^T X]_{k\bullet} W_{\bullet k}^c \\ &+ \lambda_k (1 - [W^{cT} X^T X W^c]_{kk}). \end{aligned} \tag{14}$$

The partial derivative with respect to $W_{\bullet k}^c$ is

$$\begin{aligned} &(1): \frac{\partial \mathscr{L}(W^c,\lambda_k)}{\partial W_{\bullet k}^c} = 0 \\ &(2): 1 - [W^{cT} X^{cT} X^c W^c]_{kk} = 0 \\ &(3): \lambda_k > 0 \end{aligned} \tag{15}$$

Hence, the solution to $W_{\bullet k}^c$ is obtained as

$$W_{\bullet k}^c = \frac{S_{k\bullet}^{c\ T} - [\widetilde{W^c}^k F]_{\bullet k}}{\pm \sqrt{(S_{k\bullet}^{c\ T} - [\widetilde{W^c}^k F]_{\bullet k})^T X^T X (S_{k\bullet}^{c\ T} - [\widetilde{W^c}^k F]_{\bullet k})}} . \tag{16}$$

Where, $F = S^c S^{cT}$ and $\overline{W^c}^k = \begin{cases} W_p^c, p \ne k \\ 0, p = k \end{cases}$

From Eqn. (16), two solutions are obtained with $\pm$ signs. The sign of $W_{\bullet k}^c$ is not essential since it can be easily absorbed by converting between $S_{k\bullet}^c$ and $-S_{k\bullet}^c$.

***Optimizing hybrid collaborative representation***

The optimization problem of obtaining the hybrid collaborative representation in Eqn. (7) can be rewritten as follows,

$$\begin{aligned} f(s) &= \|y - XWs\|_2^2 + \beta \|s\|_2^2 \\ &+ \tau \sum_{c=1}^{C} \left\{ \|y - [0,\cdots,X^c W^c,\cdots,0]s\|_2^2 \right\} \end{aligned} \tag{17}$$

Here, let $[0, \cdots, X^c, \cdots, 0]$ be $\hat{X}$. The Eqn. (17) can be simplified as follows,

$$f(s) = (1+\tau C)y^T y - 2(1+\tau)y^T X s + s^T \left( X^T X + \beta I + \tau \sum_{c=1}^{C} \hat{X}^T \hat{X} \right) s \quad (18)$$

It is convenient to obtain the optimum $\hat{s}$ for Eqn.(18), as follows,

$$\hat{s} = \left( X^T X + \beta I + \tau \sum_{c=1}^{C} \hat{X}^T \hat{X} \right)^{-1} \{(1+\tau) y^T X\} \quad (19)$$

## IV Experimental results

In this section, we show our experimental results on four datasets, including two handwritten recognition datasets MNIST dataset [16] and USPS dataset [17], and two face recognition datasets Extended YaleB [18] and CMU PIE dataset [19]. We compare our method (Hybrid-CRC) with some state-of-the-art methods to illustrate the significance of our approach. In the following section, we firstly introduce experimental environment settings. Then we illustrate the experimental results on each dataset, and finally, we present the analysis of results.

### Experimental settings

We test our method on four datasets. The proposed Hybrid-CRC algorithm is compared with other classification algorithms, including nearest neighbor classification (NN), Support Vector Machine (LIBSVM) [20], Collaborative representation based classification (CRC) [8], class specific dictionary learning algorithm (CSDL) [12] and Probabilistic collaborative representation based classification (ProCRC) [15].

There are two parameters in the objective function of the Hybrid-CRC algorithm which need to be specified. $\beta$ is an important parameter in Hybrid-CRC algorithm, which is used to adjust the trade-off between the reconstruction error and the collaborative representation. We increase $\beta$ from $2^{-16}$ to $2^2$ in each experiment and find the best $\beta$ in our experiments. And $\tau$ is another important factor in the algorithm, which is used to control the trade-off between the shared collaborative representation and the class specific collaborative representation. We increase $\tau$ from $2^{-11}$ to $2^2$ and find the best $\tau$ in all of our experiments. $K$ is the size of the dictionary for each class. $K$ is set to be twice of the size of the training samples per class.

For all the datasets, we randomly select 5 images as the training samples and 10 images as the testing samples from each category. To eliminate the randomness, we randomly (repeatable) split the dataset into the train set and test set 10 times, respectively. The average accuracy is recorded.

### Experiment on Face Recognition datasets

For both benchmark datasets, each face image is cropped to $32 \times 32$, pulled into a column vector, and performed a $\ell_2$ normalization to form the raw feature.

For the Extended Yale B dataset, it involves 38 categories and 2,414 frontal-face images in total, where all the images are taken under different illumination conditions. Figure 1 shows some sample images from the dataset. We set $\beta$ and $\tau$ to different values in order to achieve the best accuracy of different methods. Parameter $\beta$ ranges from $2^{-16}$ to $2^{-4}$ and $\tau$ ranges from $2^{-7}$ to $2^{-3}$. The optimal parameter $\beta$ is $2^{-10}$, $2^{-6}$, $2^{-15}$, $2^{-10}$ for CRC, CSDL, ProCRC and Hybrid-CRC,respectively. Parameter $\tau$ appears in Pro-

**Figure 1.** *Examples of the Extended YaleB dataset*

**Figure 2.** *Examples of the CMU PIE dataset*

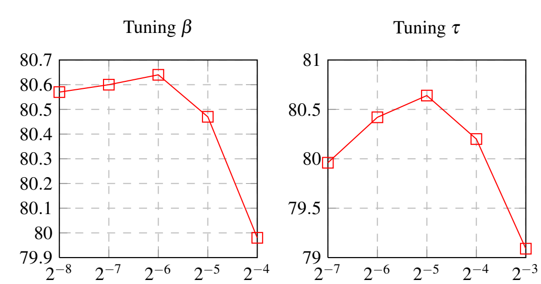


**Figure 3.** *Parameter tuned on accuracy on CMU PIE dataset. The left figure is for tuning $\beta$ with $\tau = 2^{-5}$. The right figure is for tuning $\tau$ with $\beta = 2^{-6}$*

| Methods\datasets | Extended YaleB | CMU PIE |
|---|---|---|
| NN | 43.29 | 40.05 |
| SVM | 71.57 | 67.56 |
| CRC | 80.74 | 76.46 |
| CSDL | 81.11 | 79.10 |
| ProCRC | **83.65** | 80.32 |
| Hybrid-CRC | 83.19 | **80.64** |

**Table 1: Recognition rate on the face recognition datasets (%).**

| Methods\datasets | MNIST | USPS |
|---|---|---|
| NN | 67.16 | 75.86 |
| SVM | 65.35 | 77.90 |
| CRC | 69.90 | 80.49 |
| CSDL | 70.18 | 81.35 |
| ProCRC | 69.90 | 81.72 |
| Hybrid-CRC | **72.22** | **82.60** |

**Table 2: Recognition rate on the handwritten recognition datasets (%).**

CRC algorithm and Hybrid-CRC algorithm with optimal value of $2^{-1}$ and $2^{-5}$, respectively.

The recognition accuracy is shown in Table 1. From Table 1, we can clearly see that Hybrid-CRC algorithm achieving accuracy of 83.19%, while the method CSDL arrives at 81.11%. Table 1 illustrates the effectiveness and robustness of Hybrid-CRC for classifying images with illumination variations, since lighting of images are quite different in the dataset.

For the CMU-PIE dataset, there are 41,368 pieces of pictures, captured under different lighting, poses and expressions. The CMU-PIE dataset includes 68 individuals totally, and each person has 43 different illumination conditions with 13 different poses. We choose two types of them to finish our experiment: five near frontal poses and all different illuminations, including 11,554 images in total. Each individual contains approximately 170 images. Figure 2 shows some sample images from the dataset. Parameter $\beta$ ranges from $2^{-11}$ to $2^{-2}$ and $\tau$ ranges from $2^{-7}$ to $2^{2}$. We set $\beta$ to different values to achieve the best accuracy of different methods, with $2^{-7}$ and $2^{-4}$ for CRC and CSDL respectively. For the ProCRC algorithm, $\beta$ and $\tau$ are set to $2^{-9}$ and $2^{0}$, respectively. For the Hybrid-CRC algorithm, $\beta$ and $\tau$ are set to $2^{-6}$ and $2^{-5}$, respectively.

The parameter tuning for $\tau$ and $\beta$ is reported in Figure 3. The recognition accuracy is shown in Table 1. From Table 1, we can clearly see that Hybrid-CRC algorithm outperforms other conventional methods, achieving accuracy of 80.64%, while accuracy of ProCRC arrives at 80.32%, CSDL arrives at 79.10% and CRC only arrives at 76.46% From these experimental results, we further confirm the effectiveness and robustness of CSDL algorithm for image classification with illumination and expression changes.

## Experiment on Handwritten Recognition datasets

MNIST dataset includes 70,000 images of handwritten numbers. Figure 4 shows some sample images from the dataset. The image size is $28 \times 28$. We pull each image into a column vector, and perform a $\ell_2$ normalization to form the raw feature.

*Figure 4. Examples of the MNIST dataset*

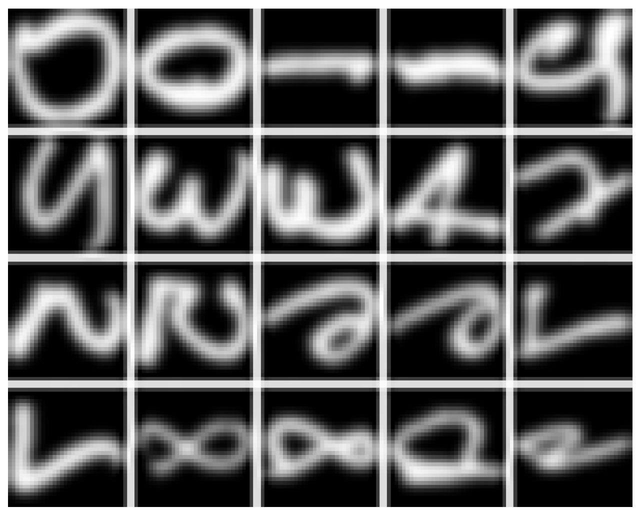

*Figure 5. Examples of the usps dataset*

We set different $\beta$ to achieve the highest accuracy of different methods, with $2^{-1}$ and $2^{0}$ for CRC and CSDL algorithm, respectively. For the ProCRC algorithm, $\beta$ and $\tau$ are set to the value $2^{-1}$ and $2^{-9}$ respectively. For the Hybrid-CRC algorithm, $\beta$ and $\tau$ are set to the value $2^{0}$ and $2^{0}$, respectively.

The recognition accuracy is shown in Table 2. From Table 2, we can clearly see that Hybrid-CRC algorithm is better than other methods,reaching the highest accuracy of 72.22%. And it is about 2% higher than accuracy of CSDL algorithm. From these experimental results, we further verify the effectiveness of Hybrid-CRC algorithm for image classification.

For USPS dataset, there are 9,298 images of handwritten numbers in total. Figure 5 shows some sample images from the dataset. The image size is $16 \times 16$. We pull each image into a column vector, and perform a $\ell_2$ normalization to form the raw feature. Parameter $\beta$ ranges from $2^{-5}$ to $2^{0}$ and $\tau$ ranges from $2^{-2}$ to $2^{2}$. The optimal parameter $\beta$ is $2^{-3}$, $2^{-2}$, $2^{-3}$, $2^{-2}$ for CRC, CSDL, ProCRC and Hybrid-CRC,respectively. Parameter $\tau$ is used in ProCRC algorithm and Hybrid-CRC algorithm with optimal value of $2^{-1}$ and $2^{0}$, respectively.

The recognition accuracy is shown in Table 2. From Table 2, we can clearly illustrate that our proposed Hybrid-CRC algorithm is more accurate than other approaches for image classification. Our algorithm arrivers at the best accuracy of 82.60%, while accuracy of CSDL reaches 81.35% and CRC only reaches 80.49%. From these experimental results, the effectiveness of Hybrid-CRC algorithm for classification is further confirmed.

### *Analysis of experimental results*

From the experimental results, we can obtain the following conclusions.

(1) The classification accuracy on these four datasets with our Hybrid-CRC method is higher than that using most of other methods,including NN, LIBSVM, CRC, ProCRC, and CSDL.

(2) Our proposed Hybrid-CRC algorithm can efficiently improve the performance of fine-grained image classification.

## V Conclusion

In this paper, we mainly focus on improving conventional class specific dictionary learning for fine-grained image classification. On one hand, we propose a hybrid collaborative representation based classification method to describe an image. On the other hand, we learn a dictionary for hybrid collaborative representation based classification to reduce the representation error and obtain precise description with the training samples. These enhancements extremely improve the performance of fine-grained image classification accuracy. Experimental results demonstrate our proposed Hybrid-CRC algorithm for visual recognition tasks.

## VI Acknowledgment

This paper is supported partly by the National Natural Science Foundation of China (Grant No. 61402535, No. 61271407), the Natural Science Foundation for Youths of Shandong Province, China (Grant No. ZR2014FQ001) , the Natural Science Foundation of Shandong Province, China(Grant No. ZR2017MF069), Qingdao Science and Technology Project (No. 17-1-1-8-jch), and the Fundamental Research Funds for the Central Universities, China University of Petroleum (East China) (Grant No. 16CX02060A).

## References

[1] Bao-Di Liu, Yu-Xiong Wang, Yu-Jin Zhang, and Bin Shen. Learning dictionary on manifolds for image classification. *Pattern Recognition*, 46(7):1879–1890, 2013.

[2] Bao-Di Liu, Yu-Xiong Wang, Bin Shen, Yu-Jin Zhang, and Martial Hebert. Self-explanatory sparse representation for image classification. In *ECCV*, pages 600–616. Springer, 2014.

[3] Bao-Di Liu, Yu-Xiong Wang, Bin Shen, Xue Li, Yu-Jin Zhang, and Yan-Jiang Wang. Blockwise coordinate descent schemes for efficient and effective dictionary learning. *Neurocomputing*, 178:25–35, 2015.

[4] O Duda Richard, E Hart Peter, and G Stork David. Pattern classification. *A Wiley-Interscience*, pages 373–378, 2001.

[5] Dacheng Tao, Xuelong Li, Xindong Wu, and Stephen J Maybank. Geometric mean for subspace selection. *IEEE Trans. PAMI*, 31(2):260–274, 2009.

[6] Chia-Po Wei, Yu-Wei Chao, Yi-Ren Yeh, and Yu-Chiang Frank Wang. Locality-sensitive dictionary learning for sparse representation based classification. *Pattern Recognition*, 46(5):1277–1287, 2013.

[7] John Wright, Allen Y Yang, Arvind Ganesh, Shankar S Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Trans. PAMI*, 31(2):210–227, 2009.

[8] Lei Zhang, Meng Yang, and Xiangchu Feng. Sparse representation or collaborative representation: Which helps face recognition? In *ICCV*, pages 471–478. IEEE, 2011.

[9] Meng Yang, Lei Zhang, Jian Yang, and Dejing Zhang. Metaface learning for sparse representation based face recognition. In *ICIP*, pages 1601–1604. IEEE, 2010.

[10] Haoran Wang, Chunfeng Yuan, Weiming Hu, and Changyin Sun. Supervised class-specific dictionary learning for sparse modeling in action recognition. *Pattern Recogntiion*, 45(11):3902–3911, 2012.

[11] Bao-Di Liu, Bin Shen, and Yu-Xiong Wang. Class specific dictionary learning for face recognition. In *Security, Pattern Analysis, and Cybernetics (SPAC), 2014 International Conference on*, pages 229–234. IEEE, 2014.

[12] Bao-Di Liu, Bin Shen, Liangke Gui, Yu-Xiong Wang, Xue Li, Fei Yan, and Yan-Jiang Wang. Face recognition using class specific dictionary learning for sparse representation and collaborative representation. *Neurocomputing*, 204:198–210, 2016.

[13] Bao-Di Liu, Liangke Gui, Yuting Wang, Yu-Xiong Wang, Bin Shen, Xue Li, and Yan-Jiang Wang. Class specific centralized dictionary learning for face recognition. *Multimedia Tools and Applications*, 76(3):4159–4177, 2017.

[14] Wei Wang, Yan Yan, Stefan Winkler, and Nicu Sebe. Category specific dictionary learning for attribute specific feature selection. *IEEE Transactions on Image Processing*, 25(3):1465–1478, 2016.

[15] Sijia Cai, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. A probabilistic collaborative representation based approach for pattern classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2950–2959, 2016.

[16] Yann LeCun, Corinna Cortes, and Christopher JC Burges. Mnist handwritten digit database. *AT&T Labs [Online]. Available: http://yann. lecun. com/exdb/mnist*, 2, 2010.

[17] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[18] Athinodoros S Georghiades, Peter N Belhumeur, and David J Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. PAMI*, 23(6):643–660, 2001.

[19] Terence Sim, Simon Baker, and Maan Bsat. The cmu pose, illumination, and expression (pie) database. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 46–51. IEEE, 2002.

[20] Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

## Author Biography

*Wen-Yang Xie received her bachelor's degree in Electronic and Information Engineering from China University of Petroleum (2013). Currently, she is studying for her master degree in Information and Communication Engineering in China University of Petroleum. Her main research interests is deep learning.*