Journal Pre-proof

Weighted discriminative collaborative competitive representation for robust image classification

Jianping Gou, Lei Wang, Zhang Yi, Yunhao Yuan, Weihua Ou, Qirong Mao

PII: S0893-6080(20)30022-8
DOI: https://doi.org/10.1016/j.neunet.2020.01.020
Reference: NN 4383

To appear in: *Neural Networks*

Received date : 15 July 2019
Revised date : 31 October 2019
Accepted date : 17 January 2020

Weighted Discriminative Collaborative Competitive Representation for Robust Image Classification

Jianping Gou[a], Lei Wang[a], Zhang Yi[b], Yunhao Yuan[c], Weihua Ou[d], Qirong Mao[a]

[a]*School of Computer Science and Communication Engineering and Jiangsu Key Laboratory of Security Tech. for Industrial Cyberspace, Jiangsu University, Zhenjiang, Jiangsu, 212013, China*
[b]*School of Computer Science, Sichuan University, Chengdu, 610065 China*
[c]*School of Computer Science and Technology, Yangzhou University, Yangzhou, Jiangsu, 225127, China.*
[d]*School of Big Data and Computer Science, Guizhou Normal University, Guiyang, Guizhou, 550025, China.*

Abstract

Collaborative representation-based classification (CRC) is a famous representation-based classification method in pattern recognition. Recently, many variants of CRC have been designed for many classification tasks with the good classification performance. However, most of them ignore the inter-class pattern discrimination among the class-specific representations, which is very critical for strengthening the pattern discrimination of collaborative representation (CR). In this article, we propose a novel CR approach for image classification, called weighted discriminative collaborative competitive representation (WDCCR). The proposed WDCCR designs the discriminative and competitive collaborative representation among all the classes by fully considering the class information. On the one hand, we incorporate two discriminative constraints into the unified WDCCR model. Both constraints are the competitive class-specific representation residuals and the pairs of class-specific representations for each query sample. On the other hand, the constraint of the weighted categorical representation coefficients is introduced into the proposed model for further enhancing the power of discriminative and competitive representation. In the weighted constraint, we assume that the different classes of each query sample should have less contribution to the representation with the small representation coefficients, and then two types of weight factors are designed to constrain the representation coefficients. Furthermore, the robust WDCCR (R-WDCCR) is proposed with l_1-norm representation fidelity for recognizing noisy images. Extensive experiments on six image data sets demonstrate the effective and robust superiorities of the proposed WDCCR and R-WDCCR over the related state-of-the-art representation-based classification methods.

Keywords: Collaborative representation-based classification, Collaborative representation, Representation-based classification, Image classification, Pattern recognition

1. Introduction

Collaborative representation (CR) (Zhang et al., 2011) that is regarded as the general form of sparse representation has drawn a considerable interest in the fields of computer vision and pattern recognition recently. A great many CR-based classification methods as a typical type of representation-based classification have been developed for various image classification tasks, such as face recognition (Zhang et al., 2011; Wang et al., 2015; Gui et al., 2016; Deng et al., 2018),

Email address: goujianping@ujs.edu.cn (Jianping Gou)

scene classification (Akhtar et al., 2017; Chi et al., 2018), object recognition (Cai et al., 2016; Lan and Zhou, 2017; Lan et al., 2018) and hyperspectral image classification (Li et al., 2015a; Su et al., 2018; Li et al., 2015b; Gui et al., 2014).

In the representation-based classification, sparse representation-based classification (SRC) (Wright et al., 2009) and collaborative representation-based classification (CRC) (Zhang et al., 2011) are the most representative linear ones. Through sparse representation or collaborative representation, a query sample can be linearly approximated by all the training samples and then be classified according to their representation residuals. In SRC, the sparsity nature of l_1-norm similar to l_0-norm (Donoho, 2006) is held and a query sample is reconstructed by a small quantity of training samples. Since SRC was first introduced, it has gained much attention and its many extensions have been developed (Zhang et al., 2015). The l_1-norm regularization-based sparse representation highly relies on the distribution of data. A query sample could be sparsely represented by one single training sample when the pair-wise correlations of the training samples are very high. To overcome this issue, a group sparsity promoting classifier is designed by constraining the class-specific sum over l_2-norm regularization of the representation coefficients. As a weighted extension of group sparsity promoting classifier, the similarity between a query sample and each training sample as the weight constrains the $l_{2,1}$ mixed-norm of representation coefficients (Tang et al., 2014). To enhance the efficiency and robustness of SRC, a discriminative sparse classifier (DSRC) was proposed in (Xu et al., 2017). DSRC exploits l_2-norm regularization of the pairs of the class-specific representations, which can degrade the correlations among different classes and obtain good sparsity for discriminative representation. To solve the poor performance of SRC on uncontrolled and imbalanced classification, the sparse supervised representation-based classifier was proposed in (Shu et al., 2018), where a query sample is linearly represented by each class and the class-specific weights on representation fidelity are learned according to categorical variances. Nowadays some deep sparse representation-based classification methods on the basis of the idea of deep learning have been developed to learn better features of complex data for improving classification performance (Tariyal et al., 2016; Fan et al., 2017; Sharma et al., 2017).

Due to the time-consuming optimization of l_1-norm and the doubt whether or not the sparsity plays an essential role in SRC, it has been argued that it is the collaboration of l_2-norm instead of sparsity of l_1-norm that works on the representation-based classification. Accordingly, collaborative representation-based classification (CRC) with an efficient closed-form solution was proposed by l_2-norm regularization of representation coefficients (Zhang et al., 2011). CRC collaboratively represents a query sample with all the training samples and achieves the promising classification performance. However, the latent classification mechanism of CRC is still unclear. This mechanism was explained from the perspective of probability, and accordingly a probabilistic collaborative representation-based classifier (ProCRC) was proposed in (Cai et al., 2016). ProCRC maximizes the probability that a query sample belongs to each class by minimizing the deviation between reconstruction of all the classes and that of each class. ProCRC has good classification performance on some challenging data sets with features learned by deep learning (Simonyan and Zisserman, 2014). Using the prior information of data, an extended version of ProCRC (EProCRC) was proposed in (Lan and Zhou, 2017). Then, the full works of EProCRC were reported and four strategies of using prior knowledge for improving the classification performance of ProCRC were designed in (Lan et al., 2018). Because of good classification performance of ProCRC, several robust variants of ProCRC were recently proposed by using coarse to fine representation (Gou et al., 2019a) and extending the representation fidelity (Gou et al., 2019b), respectively. The natural discrimination of CR was further analyzed in (Deng et al., 2018) and a superposed linear representation classifier was proposed to reduce the misleading representation coefficients from wrong classes. Using the advantages from nearest subspace classifier (NSC) (Lee et al., 2005) and CRC, the collaborative representation optimized classifier (CROC) was proposed to strike a balance between the objective functions of CRC and NSC (Chi and Porikli, 2014). To get more efficient of CRC, collaborative representation with the chosen k-nearest classes was proposed for classification (Zheng and Wang, 2019). Besides them, there are some CR-based dictionary learning methods for classification (Zheng and Sun, 2019; Song et al., 2019; Liu et al., 2015a).

Among many extensions of CRC, the weighted constraints of the collaborative representation coefficients are often the simple and effective ways to enhance the pattern discrimination (Timofte and Van Gool, 2014; Waqas et al., 2013; Gou et al., 2018a; Liu et al., 2015b; Lu et al., 2013). In these weighted CRC methods, the weights constrained on collaborative representation coefficients are defined by the local similarities between the given query sample and the training samples. And the weighted CRC (Liu et al., 2015b) is also a two-phase CRC method using the coarse to fine collaborative representation (Xu et al., 2011). Currently, the competitive and collaborative representation for classification has been proposed lately in (Chi et al., 2018; Yuan et al., 2018). It strengthens the power of each class to competitively and discriminatively represent each query sample. In the competitive and collaborative representation-based classifier (Chi et al., 2018), the classification contribution of the true class of each query sample is enhanced by the l_2-norm regularization of the representation from all the other classes with the designed competitive weight. In (Yuan et al., 2018), the collaborative-competitive representation-based classifier (CCRC) was proposed by integrating the procedures of the representations of all the training samples from all the classes and each class. CCRC can well achieve the tradeoff between collaborative and competitive representation, and unify the phases of classification and representation. It is regarded as the combination of CRC and NSC, and ProCRC can also be integrated into it.

In this article, by borrowing the superiorities of discriminative, competitive and weighted collaborative representations, we propose a novel weighted discriminative collaborative competitive representation (WDCCR) for image classification. The proposed WDCCR not only considers the competitive representation of each class, but also enhances the inter-class discrimination, which is meaningful for improving the power of competitive representation from the true class. The discriminative and competitive representation is carried out by the l_2-norm regularizations of the pairs of class-specific representations and the class-specific reconstructive residuals. To further enhance the power of discrimination and competitive representation from each class, the class-specific representation coefficients are weighted by using two types of considering the intrinsic relationship between one query sample and each class. Due to the weak classification performance of CRC on noisy data, we extend the proposed WDCCR to robust WDCCR (R-WDCCR) by using l_1-norm instead of l_2-norm representation fidelity for classifying images with random corruptions or block occlusions. To demonstrate the proposed WDCCR methods, we conduct extensive experiments on image data sets in comparison with the state-of-the-art representation-based classification methods. The experimental results show that the proposed methods perform better than the competing methods. To sum up, several main contributions of this article are summarized as follows:

1. The new collaborative representation entitled WDCCR is developed for classification. It integrates the inter-class discriminative and competitive representation into a unified model.
2. The robust WDCCR (R-WDCCR) was proposed by using l_1-norm representation fidelity for the noisy image classification.
3. Two different types of weight definitions are designed for constraining the class-specific representation coefficients.

This article is the full extension and improvement of our previous work that has been published in 20th IEEE International Conference on Multimedia and Expo (ICME) (Gou et al., 2019c).

Compared to the previous work in our conference paper, we also propose several WDCCR methods by designing two types of weighted constraints on representation coefficients and using l_1-norm representation fidelity. Moreover, we present the derivation of the proposed models in details and give in-depth analysis of their rationale, advantages and computational complexities. And also, we conduct extensive experiments to fully evaluate the proposed methods.

The remaining part of this article is organized as follows. Some related works are introduced in Section 2. The proposed WDCCR and R-WDCCR are detailedly presented in Section 3. Section 4 reports extensive experiments to evaluate the proposed methods. Section 5 further analyzes the power of pattern discrimination and the computational complexities of the proposed methods. Finally, the conclusions are given in Section 6.

2. Related Works

Before briefly reviewing related works, we first introduce some common used notations in this article. Let the training set be $X=[X_1, X_2, \ldots, X_C] \in \mathbb{R}^{m\times N}$ that contains N training samples belonging to C classes and each column of X is one training sample with m dimensionalities. $X_i \in \mathbb{R}^{m\times n_i}$ denotes a subset of X from the ith class and $N=\sum_{i=1}^{C} n_i$. In the representation-based classification, a query sample $y\in\mathbb{R}^m$ can be linearly approximated as $y \approx X\alpha = \sum_{i=1}^{C} X_i\alpha_i$, where $\alpha=[\alpha_1^T, \alpha_2^T, \ldots, \alpha_C^T]^T \in \mathbb{R}^N$ is the vector of the representation coefficients corresponding to all the training samples and $\alpha_i \in \mathbb{R}^{n_i}$ is a subvector of α with regard to class i.

2.1. SRC and CRC

Up to now, many applications based on the extended SRC methods (Shu et al., 2018; Qu et al., 2018; Gou et al., 2018b; Ou et al., 2018; Benuwa et al., 2019) and CRC methods (Li et al., 2015a; Zhang and Sun, 2018; Peng et al., 2018; Zeng et al., 2018, 2017) have been reported to verify their good classification performance. Actually, SRC (Wright et al., 2009) and CRC (Zhang et al., 2011) are the linear representation-based classifiers and impose a regularization term to representation vector α. The general model is determined as follows:

$$\arg\min_{\alpha}\{\| y - X\alpha \|_2^2 + \lambda\psi(\alpha)\} , \tag{1}$$

where λ is a positive parameter and $\psi(\alpha)$ is a regularized norm of α. For different $\psi(\alpha)$, different solutions of α can be resolved, such as a sparse solution of SRC when $\psi(\alpha) = \| \alpha \|_1$ and a dense solution of CRC when $\psi(\alpha) = \| \alpha \|_2^2$. Here, $\|.\|_1$ denotes the l_1-norm that calculates the sum of the absolute values of all elements of a vector, and to minimize $\|.\|_1$ can enhance the sparsity of representation in Eq. (1). In SRC, a query sample y is sparsely represented by a few training samples and classified into the class with the minimum reconstructive residual by $\| y - X_i\alpha_i \|_2^2$ among all the classes. Note that we use OMP (Pati et al., 1993) to quickly solve SRC model in our works. In CRC, a query sample y is collaboratively represented by all the training samples and classified into the class with the most reconstructive contribution according to $\frac{\|y-X_i\alpha_i\|_2^2}{\|\alpha_i\|_2^2}$. And CRC has the simple and efficient closed-form solution.

2.2. DSRC

Considering the complex and time-consuming computation of SRC, DSRC (Xu et al., 2017) develops an efficient SRC model with novel discriminative l_2-norm regularization as follows:

$$\arg\min_{\alpha}\left\{ \| y - X\alpha \|_2^2 + \lambda \sum_{i=1}^{C}\sum_{j=1}^{C} \| X_i\alpha_i + X_j\alpha_j \|_2^2 \right\} . \tag{2}$$

The second term in Eq. (2) can be rewritten as $\|X_i\alpha_i\|_2^2 + \|X_j\alpha_j\|_2^2 + 2(X_i\alpha_i)^T(X_j\alpha_j)$, where the minimization of $(X_i\alpha_i)^T(X_j\alpha_j)$ means the decorrelation of representations between classes i and j. After degrading the correlation among classes, the inter-class discrimination is well enhanced. In the procedure of the classification decision, the query sample y is classified into the class with the minimum residual by $\|y - X_i\alpha_i\|_2^2$ among all the classes.

2.3. CCRC

CCRC (Yuan et al., 2018) integrates the phases of both collaborative representation and the classification decision into the unified model that results in competitive representation from each class. The model of CCRC is defined as

$$\arg\min_{\alpha}\left\{\| y - X\alpha \|_2^2 + \lambda_1 \sum_{i=1}^{C} \| y - X_i\alpha_i \|_2^2 + \lambda_2 \|\alpha\|_2^2 \right\}, \tag{3}$$

where the term $\sum_{i=1}^{C} \| y - X_i\alpha_i \|_2^2$ is the sum of categorical reconstruction residuals that makes each class have competitive contribution to reconstructing the query sample as far as possible. Through Eq. (3), each class competitively represents each query sample, and the class label of the query sample y is determined by $\|y - X_i\alpha_i\|_2^2$.

3. Proposed Methods

In this section, the proposed WDCCR and R-WDCCR approaches are detailed. We first give the motivation of the proposed approaches and then present their mathematical models as well as their derivations of optimizing solutions. Finally, two types of weight definitions are designed for constraining the categorical representation coefficients.

3.1. Motivation

The good classification performance of CCRC benefits from integrating both representation and classification of CRC into an unified model. Each class in CCRC learns competitive representation as far as passible to classify each query sample. The competitive term $\|y - X_i\alpha_i\|_2^2$ in the CCRC model can be reformulated as:

$$\| y - X_i\alpha_i \|_2^2 = \| y \|_2^2 + \| X_i\alpha_i \|_2^2 - 2y^T(X_i\alpha_i) . \tag{4}$$

Wherein, $y^T(X_i\alpha_i)$ is the dot product between y and $X_i\alpha_i$ that reflects their geometrical and spatial correlation. If $X_i\alpha_i$ is very similar to y, the correlation $y^T(X_i\alpha_i)$ is very large. To minimize $\|y–X_i\alpha_i\|_2^2$ is to enhance the probability that y belongs to class i. Particulary, minimizing $\|y–X_i\alpha_i\|_2^2$ is equivalent to minimize $\| y \|_2^2$ and $\| X_i\alpha_i \|_2^2$, and to maximize the correlation $y^T(X_i\alpha_i)$ simultaneously. The correlations between the query sample and the class-specific representations are maximized to achieve the competitive representation. However, several classes close to the query sample y could have very similar competitive representation with similar correlation $y^T(X_i\alpha_i)$, so as to difficultly identify the true class label of y.

To visually illustrate this issue in CCRC, we first define the degree of correlation between y and the representation $X_i\alpha_i$ from the ith class as

$$Corr(y, X_i\alpha_i) = y^T(X_i\alpha_i) . \tag{5}$$

Here, $Corr(y, X_i\alpha_i)$ is called correlation coefficient between y and the ith class. Then, the examples of correlation coefficient between y and the representation $X_i\alpha_i$ and its corresponding representation residual $\|y - X_i\alpha_i\|_2^2$ for the given query sample y are shown in Fig. 1. The given query image

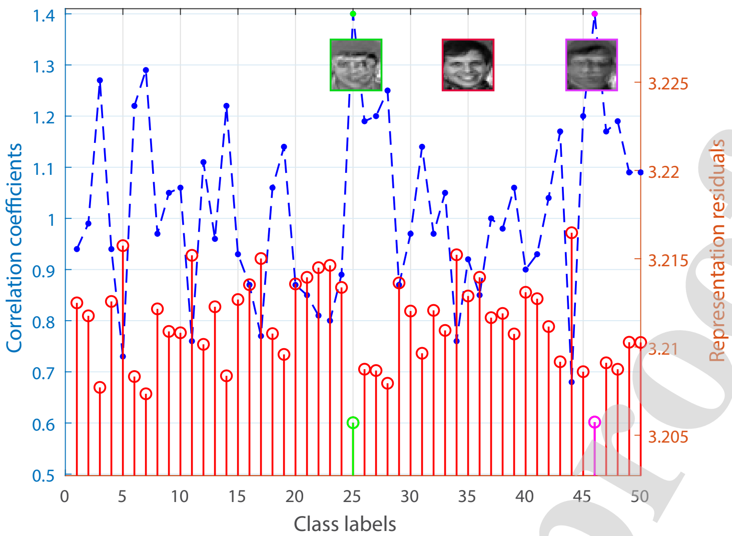


Figure 1: The examples of the correlation coefficients (blue dash-dot line) and representation residuals (red solid line) obtained by CCRC for a query sample from class 46 on GT. Note that the given query image is within red pane, its reconstructive images from classes 25 and 46 are within green and magenta panes, respectively.

sample in Fig. 1 is chosen from class 46 in the GT face database[1], which has 750 image samples within 50 classes. Note that the first 8 samples per class are chosen for training and the rest 7 samples for testing. The green and magenta lines in Fig. 1 denotes the representation residuals of the wrongly classified class and true class of the given query sample, respectively. We can clearly see the fact that each class-specific representation residual is always inversely proportional to its corresponding correlation coefficient. That is, one class with larger correlation to y always has its corresponding smaller representation residual. In Fig. 1, $Corr(y, X_{25}\alpha_{25})$ and $Corr(y, X_{46}\alpha_{46})$ have the identical largest value, but the representation residual from classes 25 is slightly smaller than the one from class 46. So that CCRC mistakenly classifies the given query sample into class 25. In such case, the better way is to further enhance the inter-class discrimination in the competitive representation.

To overcome the issue in CCRC above, we introduce the discriminative term $\| X_i\alpha_i + X_j\alpha_j \|_2^2$ from the DSRC model. $\| X_i\alpha_i + X_j\alpha_j \|_2^2$ can be reformulated as $\| X_i\alpha_i \|_2^2 + \| X_j\alpha_j \|_2^2$ $+2(X_i\alpha_i)^T(X_j\alpha_j)$. As argued in (Xu et al., 2017), $(X_i\alpha_i)^T(X_j\alpha_j)$ can be reviewed as the correlation between class i and class i. The minimization of $\| X_i\alpha_i + X_j\alpha_j \|_2^2$ contains the minimization of $(X_i\alpha_i)^T(X_j\alpha_j)$, which can strengthen the discrimination between class i and class j. Moreover, if the given query sample y is from class i, the decorrelation between $X_i\alpha_i$ and $X_j\alpha_j$ through the minimization of $\| X_i\alpha_i + X_j\alpha_j \|_2^2$ can make $Corr(y, X_i\alpha_i)$ larger and $Corr(y, X_j\alpha_j)$ smaller. So the inter-class discrimination can be enhanced and each query sample can obtain the discriminative and competitive representation for favorable classification.

To learn more discriminative and competitive representation for classification, we further consider the class information on the collaborative representation coefficients and design the weighted constraint of class-specifical representation coefficients. As stated in (Gou et al., 2018b; Ma et al., 2017), the training samples with the larger representation coefficients are very similar to the given query sample and has more classification contribution to classifying the given query sample. We

[1]http://www.anefian.com/research/face_reco.htm

assume that the true class of the given query sample should have larger representation coefficients with smaller penalized weights, and vice versa. That is, large penalized weights can make the different class of the given query sample have small representation coefficients. According to this assumption, we design two strategies of defining the weights that can discriminatively constrain the collaborative representation coefficients.

Thus, inspired by the idea of learning discriminative and competitive collaborative representation, we design the weighted discriminative collaborative competitive representation (WDCCR), and accordingly several WDCCR-based classification methods are proposed for image recognition.

3.2. Proposed WDCCR

To learn discriminative and competitive representation for classification, the model of the weighted discriminative collaborative competitive representation (WDCCR) is defined as follows:

$$\arg\min_{\alpha}\left\{ \| y - X\alpha \|_2^2 + \sum_{i=1}^{C} \left(\gamma \| y - X_i\alpha_i \|_2^2 + \beta \sum_{j\neq i}^{C} \| X_i\alpha_i + X_j\alpha_j \|_2^2 + \lambda w_i \| \alpha_i \|_2^2 \right) \right\}, \tag{6}$$

where γ, β, λ are positive constants for balancing each term, w_i is a class-specific weight factor which constrains the representation coefficients for favorable classification.

Through the model in Eq. (6), if the query sample y is from the kth class, WDCCR can always learn the dominant representation from class k among all the classes by minimizing the sum of $\| y - X_k\alpha_k \|_2^2$ and $\sum_{j\neq k}^{C} \| X_k\alpha_k + X_j\alpha_j \|_2^2$. To minimize $\sum_{i=1}^{C} \| y - X_k\alpha_k \|_2^2$ can make the kth class have more contribution to competitively representing y and to minimize $\sum_{i=1}^{C}\sum_{j\neq k}^{C} \| X_k\alpha_k + X_j\alpha_j \|_2^2$ can degrade the correlations among all the classes to make the kth class more discriminative to the other classes. The combination of the competitive term $\sum_{i=1}^{C} \| y - X_k\alpha_k \|_2^2$ and the discriminative term $\sum_{i=1}^{C}\sum_{j\neq k}^{C} \| X_k\alpha_k + X_j\alpha_j \|_2^2$ in the proposed WDCCR can make the true class of the query sample y competitively and discriminatively represent y. Finally, the classification decision of the WDCCR-based classifier is defined as

$$l(y) = \arg\min_{i} \| y - X_i\alpha_i \|_2^2 , \tag{7}$$

where $l(y)$ is the determined class label of y. Clearly, the competitive term directly contains the classification decision and the discriminative term indirectly reflects the classification decision, which are very beneficial to classification.

In fact, each class can have equal probability to represent and determine the class label of the given query sample in collaborative representation. But it is expected that the true class of the query sample should have dominant representation among all the classes, and especially the similar classes should have more discriminative representation. In the proposed WDCCR, we assume that the different classes of the given query sample should have less contribution to representing the query sample. To ensure this assumption, the representation coefficients from each class are constrained by the weight scalar w_i. The weight scalar w_i for each class is defined by acquiring the inherent similarity information between y and the ith class. Using the defined class-specific weight scalar, the different classes of the query sample can be penalized to have less contribution to representing and classifying the query sample, and the true class of the query sample could have more discriminative and competitive representation. Two strategies of defining weight factor w_i will be given in Section 3.5.

3.3. Proposed R-WDCCR

The CRC methods with l_2-norm regularization of collaborative representation coefficients generally have efficient and effective classification performance, but they may have week classification performance on some noisy data, such as image data sets with random corruptions or block occlusions (Wright et al., 2009; Zhang et al., 2012; Zeng et al., 2017; Jin et al., 2017). It has been well known that some CRC methods with l_1-norm regularization of the representation fidelity can enhance pattern discrimination and have more robustness to noise data (Zhang et al., 2012; Cai et al., 2016; Gou et al., 2018a, 2019b). Thus, instead of l_2-norm regularization of the representation fidelity in collaborative representation, we extend the proposed WDCCR with l_1-norm regularization of the representation fidelity for more robust classification performance on the noisy image data with corruptions and occlusions.

The robust version of WDCCR (R-WDCCR) with l_1-norm regularization of the representation fidelity is defined as follows:

$$\arg\min_{\alpha}\Big\{ \| y - X\alpha \|_1 + \sum_{i=1}^{C} \big(\gamma \| y - X_i\alpha_i \|_2^2 + \beta \sum_{j\neq i}^{C} \| X_i\alpha_i + X_j\alpha_j \|_2^2 + \lambda w_i \| \alpha_i \|_2^2 \big)\Big\} . \tag{8}$$

The used classification decision of R-WDCCR is the same as that of R-WDCCR in Eq. (7).

3.4. Solutions to WDCCR and R-WDCCR

3.4.1. Solution to WDCCR

Since the objective function of WDCCR in Eq. (6) is convex and differentiable, the collaborative representation coefficients can be resolved efficiently. Nevertheless, not only the vector α of the representation coefficients from all the classes but also the vector α_i of the class-specific representation coefficients are contained in Eq. (6). We cannot directly achieve the solution of α by the derivative of Eq. (6) with respect to α. For simplification, we first calculate the derivative of each term of Eq. (6) and then merge the derivatives to finally obtain the solution of collaborative representation coefficients.

The first term of the proposed model in Eq. (6) is denoted as $G_0 = \| y - X\alpha \|_2^2$, and its derivative with respect to α is calculated as :

$$\frac{dG_0}{d\alpha} = 2X^T(X\alpha - y) . \tag{9}$$

Let the second term to be denoted as $G_1 = \gamma \sum_{i=1}^{C} \| y - X_i\alpha_i \|_2^2$. Considering the fact α is not directly included in G_1, we can calculate all partial derivations $\partial G_1/\partial\alpha_k$ $(k = 1, 2, \ldots, C)$ to obtain $dG_1/d\alpha = [\partial G_1/\partial\alpha_1, \partial G_1/\partial\alpha_2, ..., \partial G_1/\partial\alpha_C]^T$. For the simple derivation, G_1 is first reformulated as follows:

$$G_1 = \gamma \sum_{i\neq k}^{C} \| y - X_i\alpha_i \|_2^2 + \gamma \| y - X_k\alpha_k \|_2^2 . \tag{10}$$

Then, the partial derivative of G_1 with respect to α_k is calculated as:

$$\frac{\partial G_1}{\partial\alpha_k} = 2\gamma X_k^T(X_k\alpha_k - y) . \tag{11}$$

Using $\partial G_1/\partial\alpha_k$, $dG_1/d\alpha$ is calculated as

$$\begin{aligned}\frac{dG_1}{d\alpha} &= \begin{bmatrix} 2\gamma X_1^T(X_1\alpha_1 - y) \\ 2\gamma X_2^T(X_2\alpha_2 - y) \\ \vdots \\ 2\gamma X_C^T(X_C\alpha_k - y) \end{bmatrix} \\ &= 2\gamma \begin{bmatrix} X_1^T X_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & X_C^T X_C \end{bmatrix} \alpha - 2\gamma \begin{bmatrix} X_1^T \\ \vdots \\ X_C^T \end{bmatrix} y \\ &= 2\gamma(M\alpha - X^T y) \ , \end{aligned} \tag{12}$$

where M is a block diagonal matrix, each main diagonal element of which is $X_i^T X_i$, and O denotes the zero matrix.

The third term of Eq. (6) is denoted as $G_2 = \beta\sum_{i=1}^{C}\sum_{j\neq i}^{C} \| X_i\alpha_i + X_j\alpha_j \|_2^2$. Just like G_1, the derivative of G_2 with respect to α is achieved by first computing each partial derivative of G_2 with respect to α_k. Using the similar derivations in (Xu et al., 2017; Gou et al., 2018a), G_2 can be first reformulated as follows:

$$\begin{aligned} G_2 &= \beta\Big(\sum_{i\neq k}^{C}\sum_{j\neq i}^{C} \| X_i\alpha_i + X_j\alpha_j \|_2^2 + \sum_{j\neq k}^{C} \| X_k\alpha_k + X_j\alpha_j \|_2^2 \Big) \\ &= \beta\Big(\sum_{i\neq k}^{C}\sum_{j\neq i,k}^{C} \| X_i\alpha_i + X_j\alpha_j \|_2^2 + 2\sum_{i\neq k}^{C} \| X_i\alpha_i + X_k\alpha_k \|_2^2 \Big) \ . \end{aligned} \tag{13}$$

And then, the partial derivation of G_2 with aspect to α_k is easily obtained as

$$\begin{aligned} \frac{\partial G_2}{\partial \alpha_k} &= 4\beta X_k^T \sum_{i\neq k}^{C} (X_i\alpha_i + X_k\alpha_k) \\ &= 4\beta\big(X_k^T X\alpha + (C-2)X_k^T X_k\alpha_k\big) \ . \end{aligned} \tag{14}$$

Accordingly, we get $dG_2/d\alpha$:

$$\begin{aligned} \frac{dG_2}{d\alpha} &= \begin{bmatrix} 4\beta(X_1^T X\alpha + (C-2)X_1^T X_1\alpha_1) \\ 4\beta(X_2^T X\alpha + (C-2)X_2^T X_2\alpha_2) \\ \vdots \\ 4\beta(X_C^T X\alpha + (C-2)X_C^T X_C\alpha_C) \end{bmatrix} \\ &= 4\beta \begin{bmatrix} X_1^T \\ \vdots \\ X_C^T \end{bmatrix} X\alpha + 4\beta(C-2) \begin{bmatrix} X_1^T X_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & X_C^T X_C \end{bmatrix} \alpha \\ &= 4\beta\big(X^T X\alpha + (C-2)M\alpha\big) \ , \end{aligned} \tag{15}$$

where M is the same as in Eq. (12). Most interestedly, the derivations of G_1 and G_2 have the same block diagonal matrix M that can improve the correlations among the intra-classes for good classification (Xu et al., 2017). As such, the compactness of the intr-class samples can be enhanced.

Let the last term in Eq. (6) to be denoted as $G_3 = \lambda \sum_{i=1}^{C} w_i \parallel \alpha_i \parallel_2^2$. To get the derivative of G_3 with respect to α , we first rewrite G_3 as follows:

$$\begin{aligned} G_3 =& \lambda \sum_{i=1}^{C} w_i \parallel \alpha_i \parallel_2^2 & (16)\\ =& \lambda \sum_{i=1}^{C} \parallel W_i \alpha_i \parallel_2^2 & (17)\\ =& \lambda \parallel W\alpha \parallel_2^2 \ , & (18) \end{aligned}$$

where W_i and W are the diagonal weight matrix for each class and all the classes, respectively. The diagonal weight matrix W for all the classes is defined as

$$W = \begin{bmatrix} W_1 & \dots & O \\ \vdots & \ddots & \vdots \\ O & \dots & W_C \end{bmatrix}, \tag{19}$$

where $W_i = diag(\sqrt{w_i}, \sqrt{w_i}, \dots, \sqrt{w_i}) \in \mathbb{R}^{n_i \times n_i}$ $(i = 1, 2, \dots, C)$. The definition of w_i is detailed in Subsection 3.5. And then, the derivative of G_3 with respect to α is calculated as follows:

$$\begin{aligned} \frac{dG_3}{d\alpha} &= \frac{d(\lambda \sum_{i=1}^{C} w_i \parallel \alpha_i \parallel_2^2)}{d\alpha} \\ &= \frac{d(\lambda \parallel W\alpha \parallel_2^2)}{d\alpha} \\ &= 2\lambda W^T W\alpha \ . \end{aligned} \tag{20}$$

Finally, let $\frac{dG_0}{d\alpha} + \frac{dG_1}{d\alpha} + \frac{dG_2}{d\alpha} + \frac{dG_3}{d\alpha} = 0$. Using Eqs. (9), (12), (15) and (20), we can easily achieve the solution of the collaborative representation coefficients in the proposed WDCCR model of Eq. (6) as follows:

$$\alpha = (1+\gamma)\Big((1+2\beta)X^T X + (2\beta(C-2)+\gamma)M + \lambda W^T W\Big)^{-1} X^T y \ . \tag{21}$$

As discussed above, the proposed WDCCR for classification is briefly summarized in Algorithm 1

3.4.2. Solution to R-WDCCR

The R-WDCCR model cannot directly get a closed-form solution, because of the non-smooth convex problem (Chang et al., 2016; Chen et al., 2012; Parikh et al., 2014) in Eq. (8) with the l_1-norm representation fidelity. Thus, we can convert Eq. (8) to a smooth convex function by using the iterative reweighted least square (IRLS) (Chartrand and Wotao Yin, 2008) technique. Firstly, a reweighted matrix $A \in \mathbb{R}^{m \times m}$ is defined as follows:

$$A = \begin{bmatrix} \frac{1}{|X(1,:)\alpha - y_1|} & \dots & O \\ \vdots & \ddots & \vdots \\ O & \dots & \frac{1}{|X(m,:)\alpha - y_m|} \end{bmatrix}, \tag{22}$$

where $\mathrm{X}(i,:) \in \mathbb{R}^{1 \times N}$ stands for the ith row of the training set matrix X and y_i is the ith element of the given query sample y. Using IRLS, the model of the proposed R-WDCCR can be reformulated

Algorithm 1 The proposed WDCCR method for classification

Input: A training sample set $X \in \mathbb{R}^{m\times N}$, a query sample $y \in \mathbb{R}^m$ and positive constants λ, β, γ.
Output: The class label $l(y)$ of y.

1. Normalize each column of X.
2. Calculate the weight matrix W using Algorithm 3 or 4.
3. Calculate the vector α of the representation coefficients by
$$\alpha = (1+\gamma)\Big((1+2\beta)X^TX + (2\beta(C-2)+\gamma)M + \lambda W^TW\Big)^{-1} X^Ty \ .$$
4. Calculate the class-specific residuals:
$$r_i = \| y - X_i\alpha_i \|_2^2 \ \ , \ i = 1, 2, \ldots, C$$
5. Classify the query sample y by $l(y) = \arg\min_i r_i$.

as follows:

$$\arg\min_\alpha \{(y-X\alpha)^TA(y-X\alpha)+\sum_{i=1}^{C}(\gamma \| y-X_i\alpha_i \|_2^2+\beta\sum_{j\neq i}^{C} \| X_i\alpha_i+X_j\alpha_j \|_2^2+\lambda w_i \| \alpha_i \|_2^2)\} \ . \tag{23}$$

To solve Eq. (23), here we denote the first term as $G_4 = (y-X\alpha)^TA(y-X\alpha)$ and only calculate the derivative of G_4 with respect to α as follows:

$$\begin{aligned}\frac{dG_4}{d\alpha} &= \frac{d}{d\alpha}\big((y-X\alpha)^TA(y-X\alpha)\big) \\ &= -2X^TA(y-X\alpha) \ .\end{aligned} \tag{24}$$

And the derivations of the other three terms of Eq. (23) with respect to α can be calculated by the same way as in Eq. (6). Finally, using the derivations of all the terms of Eq. (23) with respect to α. we can simply achieve the solution of the representation coefficients α in R-WDCCR as follows:

$$\alpha = \Big((2\beta(C-2)+\gamma)M + 2\beta X^TX + X^TAX + \lambda W^TW\Big)^{-1}(\gamma X^T + X^TA)y \ . \tag{25}$$

Since the reweighted matrix A is determined by α, we can iteratively update A and α until reaching the default iterations or convergence. Here, the initial α is determined by the solution of Eq. (21) to speed up convergence. To sum up, the main steps of the proposed R-WDCCR for classification are summarized in Algorithm 2.

3.5. Design of Weight Factor

As stated in Section 3.2, w_i in Eq. (6) is a class-specific weight factor that constrains the representation coefficients of the ith class. The definition of weight factor is satisfied by the assumption that the different classes of the query sample should have less contribution to representing the query sample. Specifically speaking, the ith class that is very different from the true class of the query sample should be penalized by the larger weight factor w_i in the minimization of $w_i \| \alpha_i \|_2^2$ to degrade the representation contributions, and vice versa. Meanwhile, the discrimination of the competitive representations among the different classes should be further enhanced, and the true class of the query sample could dominantly represent and classify the query sample. Thus, under such assumption, the weight factor corresponding to each class for each query sample is designed by considering the intrinsic information of data between the query sample and each class, and here we design two strategies of defining weight factor: distance-based weight factor and representation-based weight factor.

Algorithm 2 The proposed R-WDCCR method for classification

Input: A training sample set $X \in \mathbb{R}^{m\times N}$, a query sample $y \in \mathbb{R}^m$ and positive constants λ, β, γ.
Output: The class label $l(y)$ of y.

1. Normalize each column of X.
2. Calculate the weight matrix W using Algorithm 3 or 4.
3. Initialize the representation coefficients α^0 by WDCCR with Eq. (21).

$$\alpha^0 = (1+\gamma)\big((1+2\beta)X^TX + (2\beta(C-2)+\gamma)M + \lambda W^TW\big)^{-1}X^Ty \ .$$

4. Initialize iterative number t with 0.
5. Initialize the reweighted matrix A^0 by

$$A^0 = \begin{bmatrix} \frac{1}{|X(1,:)\alpha^0 - y_1|} & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & \frac{1}{|X(m,:)\alpha^0 - y_m|} \end{bmatrix} .$$

6. Iteratively compute the reweighted matrix A and representation coefficient α:

while $\| A^{t+1} - A^t \|_F \leq \delta_A$ or $t \leq T-1$ **do**
update iterative number $t = t+1$.
update reweighted matrix A^t by

$$A^t = \begin{bmatrix} \frac{1}{|X(1,:)\alpha^{t-1} - y_1|} & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & \frac{1}{|X(m,:)\alpha^{t-1} - y_m|} \end{bmatrix} .$$

update representation coefficient α^t by

$$\alpha = \Big((2\beta(C-2)+\gamma)M + 2\beta X^TX + X^TAX + \lambda W^TW\Big)^{-1}(\gamma X^T + X^TA)y \ .$$

Note that $\| \cdot \|_F$ denotes the Frobenius norm of a matrix, A^t is the weight matrix in the tth iteration, δ_A is the threshold value and T is the default number of iterations.
end while

7. Calculate the class-specific residuals:

$$r_i = \| y - X_i\alpha_i \|_2^2 \ , \ i = 1, 2, \ldots, C$$

8. Determine the label $l(y) = \arg\min_i r_i$.

3.5.1. Distance-Based Weight Factor

To design distance-based weight factor, we further suppose that the classes with small distances to each query sample could have more contribution to the representation. In the collaborative competitive representation, each class can competitively represent each query sample, but some classes with strong correlations represented by Eq. (5) could degrade the classification performance. The possible reason is that some training samples from these classes are close to the given query sample y and some samples from the true class of y is far away from y. In view of this, the mean distance between y and all the training samples from each class can be well used for reflecting the intrinsic similarity between y and each class. The class with a smaller mean distance could have larger probability that y belongs to it. Of course, in the collaborative representation, the class with the smaller mean distance could have the larger representation coefficients. If the smaller class-specific mean distance as the corresponding weights are employed to constrain the representation coefficients, its corresponding representation coefficients can be further enlarged for favorable representation. Thus, the weight factor w_i for class i is defined as the mean distance

$$w_i = \frac{1}{n_i} \cdot \sum_{x \in X_i} \| y - x \|_2^2 \ , \tag{26}$$

where n_i is the number of the training samples from class i and x is a training sample from the ith class.

Using the distance-based weight factor w_i in the minimization of $w_i \| \alpha_i \|_2^2$, the similar class i with small weight could have large collaborative representation coefficients to represent the query sample y for favourable classification. And the larger weights for different classes of y could penalize their representations with smaller representation coefficients. The computation procedure of distance-based weight factor is summarized in Algorithm 3. For ease of distinction, it should be noted that the proposed WDCCR and R-WDCCR with distance-based weight factor are denoted as WDCCR-BD and R-WDCCR-BD, respectively.

Algorithm 3 Distance-Based Weight Matrix

Input: A training sample set $X \in \mathbb{R}^{m \times N}$, a query sample $y \in \mathbb{R}^m$.
Output: The weight matrix $W \in \mathbb{R}^{\mathbb{N} \times \mathbb{N}}$.

1. Compute the weight w_i $(i = 1, 2, \ldots, C)$ between the query sample y and each class:
for each i in $\{1, 2, \ldots, C\}$ **do**
  Let $s = 0$.
  for each x in X_i **do**
    Let $s = s + \| y - x \|_2^2$.
  end for
  Compute the weight factor for class i as $w_i = s/n_i$.
end for
2. Achieve the weight matrix for all the classes as $W = diag(W_1, W_1, \ldots, W_C)$, where $W_i = diag(\sqrt{w_i}, \sqrt{w_i}, \ldots, \sqrt{w_i}) \in \mathbb{R}^{n_i \times n_i}$ $(i = 1, 2, \ldots, C)$.

3.5.2. Representation-Based Weight Factor

To well design the class-specific weight factor w_i, we employ the fundamental idea that the samples from a class should be well represented by a linear subspace of the class (Naseem et al., 2010). Under this fundamental concept, the class-specific nearest subspaces for each query sample can be learned as far as possible, and the class-specific representation-based weight factor w_i could be well defined by using the representations of the class-specific nearest subspaces. In fact, the design of representation-based weight factor assumes that the dominant representations of the class-specific nearest subspaces could have more contribution to representing each query sample.

For the given query sample y, it is approximately represented by each class as $y \approx X_i\bar{\alpha}_i$ without the influence of the other classes. Then, $\bar{\alpha}_i$ can be solved as

$$\arg\min_{\bar{\alpha}_i} \| y - X_i\bar{\alpha}_i \|_2^2 \ . \tag{27}$$

The solution can be directly obtained as $\bar{\alpha}_i = (X_i^T X_i)^{-1} X_i^T y$ in general. However, if $X_i^T X_i$ is a ill-conditioned matrix, a small l_2-norm regularization of $\bar{\alpha}_i$ is imposed to (27) for avoiding the inverse of singular matrix. After the class-specific representation, the representation-based weight factor w_i is defined as

$$w_i = \frac{r_i}{\sum_{j=1}^{C} r_j}, \tag{28}$$

where r_i is the representation residual $\| y - X_i\bar{\alpha}_i \|_2^2$ between y and class i. Clearly, if class i is very similar to the query sample y, it could have smaller residual r_i corresponding to smaller w_i. Accordingly, the similar class i with smaller w_i can well represent the query sample y with larger representation coefficients, and vice versa. The computation procedure of representation-based weight factor is summarized in Algorithm 4. Noted that the proposed WDCCR and R-WDCCR with representation-based weight factor are denoted as WDCCR-BR and R-WDCCR-BR, respectively.

Algorithm 4 Representation-Based Weight Matrix

Input: A training sample set $X \in \mathbb{R}^{m\times N}$, a query sample $y \in \mathbb{R}^m$.
Output: The weight matrix $W \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$.

1. Calculate the representation residual r_i $(i = 1, 2, \ldots, C)$ between the query sample y and each class:
   Let $s = 0$.
   for each i in $\{1, 2, \ldots, C\}$ **do**
   Calculate the representation coefficient $\bar{\alpha}_i$ of the ith class by
   $$\bar{\alpha}_i = (X_i^T X_i)^{-1} X_i^T y \ .$$
   Calculate the representation residual of class i:
   $$r_i = \| y - X_i\bar{\alpha}_i \|_2^2 \ .$$
   Let $s = s + r_i$.
   end for
2. Calculate the weight factor for class i as $w_i = r_i/s$, $(i = 1, 2, \ldots, C)$.
3. Achieve the weight matrix for all the classes as $W = diag(W_1, W_1, \ldots, W_C)$, where $W_i = diag(\sqrt{w_i}, \sqrt{w_i}, \ldots, \sqrt{w_i}) \in \mathbb{R}^{n_i\times n_i}$ $(i = 1, 2, \ldots, C)$.

---

4. Experiments

In this section, the extensive experiments on several image data sets are performed to verify the classification performance of the proposed WDCCR-BD, WDCCR-BR, R-WDCCR-BD and

R-WDCCR-BR, compared to the state-of-the-art representation-based classification methods. We first describe the six used benchmark image data sets, and then evaluate the classification performance of the proposed methods with varying the values of three regularized parameters, finally conduct the comparative experiments among the competing methods on image data without noises and on image data with noises including random corruptions and occlusions.

4.1. Data Sets

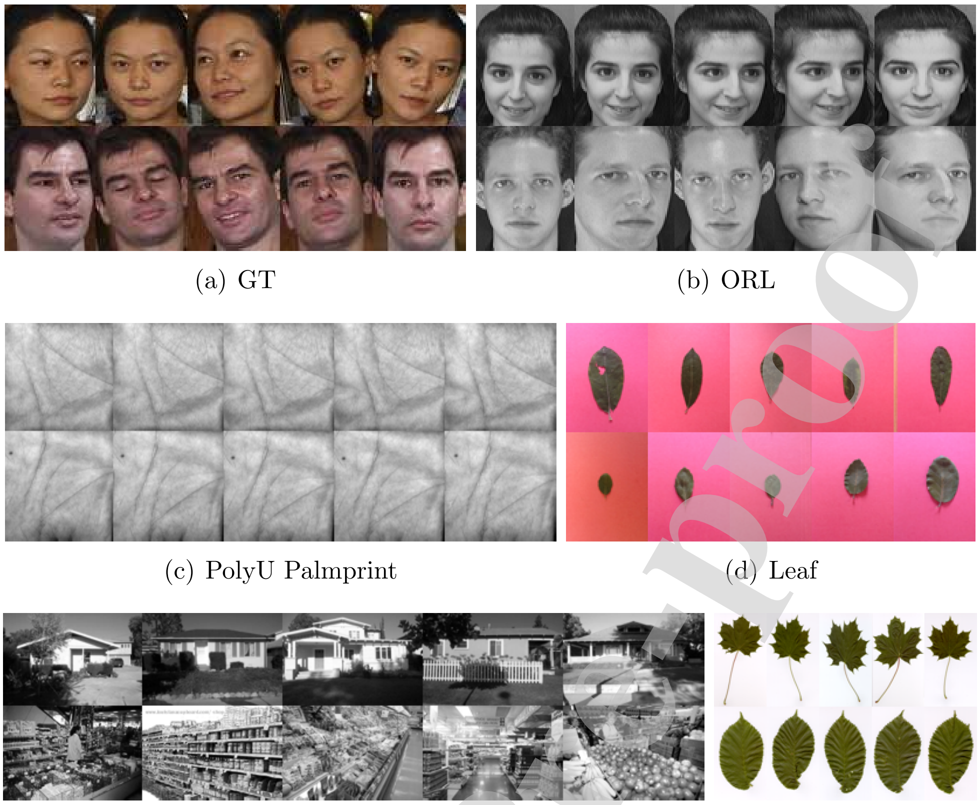
In this subsection, the used six image data sets including Georgia Tech (GT) data set[2], Olivetti Research Laboratory (ORL) data set[3], PolyU Palmprint data set (Zhang et al., 2010), Leaf data set (Silva et al., 2013), Fifteen Scene Category data set (Lazebnik et al., 2006) and Swedish Leaf data set (Söderkvist, 2001) are briefly described as follows.

1. GT: There are 750 frontal or titled face images from 50 persons taken in two or three sessions. Each person has 15 RGB images with various positions, expressions and lighting conditions. In the experiments, each face image is grayed and resized into 32×32 pixels. Some images of two persons on GT are illustrated in Fig. 2(a).
2. ORL: There are 400 face images taken from 40 persons, each of which has 10 images. The images were taken by varying the different lighting conditions and expressions. Each face image with 92×112 pixels is resized into 32×32 pixels in the experiments. Some images of two persons on ORL are illustrated in Fig. 2(b).
3. PolyU Palmprint: A subset of this data set is used in the experiments. This subset is composed of 1200 palm images with 20 images per person. The original size of each image is 191×191 pixels and we resize each image into 32×32 pixels. Fig. 2(c) illustrates some images from two persons on PolyU Palmprint.
4. Leaf: There are 40 diverse plant species containing 443 color leaf images totally. The original resolution of each image is 720×960 pixels taken by an Apple iPad2 device. Since each plant species has 5$\sim$16 images, we choose 36 species, each of which has more than 9 images in the experiments, and the total number of images is 413. Each image is grayed and resized into 32×32 pixels. Fig. 2(d) illustrates some images from two plant species on Leaf.
5. Fifteen Scene Category: There are 4485 images taken from 15 nature scenes. Each scene contains 200$\sim$400 images with the average size 250×300 pixels. For each image, a 1000 dimensional spatial pyramid feature is extracted by the method in (Lazebnik et al., 2006). Fig. 2(e) illustrates some images of two scenes on Fifteen Scene Category.
6. Swedish Leaf: There are 1125 color images taken from 15 tree classes, each of which has 75 images. Each image is grayed and resized into 32×32 pixels in the experiments. Fig. 2(f) illustrates some images of two classes on Swedish Leaf.

4.2. Experimental Parameter Selections

As stated in the proposed models, there are three parameters λ, γ and β that simultaneously balance their corresponding constraint term in the collaborative representation. Particularly speaking, λ adjusts the weighted collaborative representation coefficients, γ and β adjust the abilities of the competitive term and the discrimination term, respectively. In the experiments, we investigate the classification performance of the proposed WDCCR and R-WDCCR with varying the values of these three parameters, and their values of three parameters are from the preseted

[2]http://www.anefian.com/research/face_reco.htm
[3]http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.



(a) GT (b) ORL

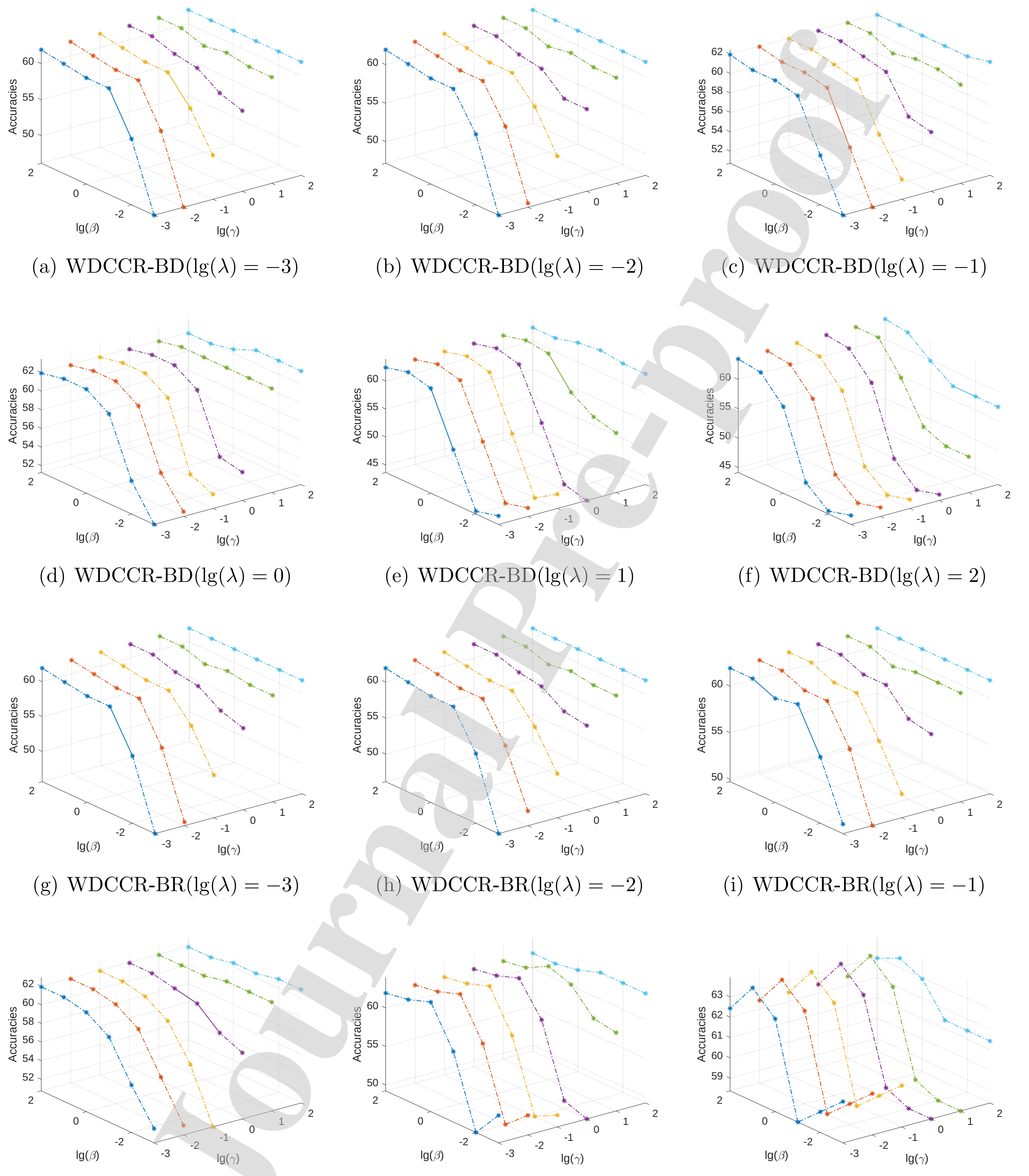
(c) PolyU Palmprint (d) Leaf

(e) Fifteen Scene Category (f) Swedish Leaf

Figure 2: Some image samples for two classes from different data sets.

range $\{0.001, 0.01, 0.1, 1, 10, 100\}$. For clearly expounding the influence of these parameters on the performance of the proposed methods and optimally choosing values of these parameters, we conduct the experiments on the Leaf data set (Silva et al., 2013) with the random 6 samples per class as the training samples and the remaining samples as the testing ones. Fig. 3 illustrates the classification accuracies of the proposed WDCCR-BD and WDCCR-BR with the variations of β, γ and λ on Leaf. In this figure, each color line in the subfigures denotes the accuracy variation curve with different values of β and a fixed value of γ. For simple illustrations in Fig. 3, it should be noted that the values of $\lg(\lambda)$, $\lg(\beta)$ and $\lg(\gamma)$ in the experiments are used for corresponding to the values of λ, β and γ, respectively.

We can see the most interesting experimental fact from Fig. 3 that the proposed WDCCR-BD and WDCCR-BR almost have the similar pattern with varying the values of the parameters λ, γ and β on Leaf. It is clear that both WDCCR-BD and WDCCR-BR have very similar pattern of the classification accuracies with varying the values of γ and β when $\lg(\lambda) = -3, -2, -1$. This experimental phenomenon implies that the smaller values of λ have less impact on the classification performance by adjusting weighted representation coefficients. When $\lg(\lambda) = 0, 1, 2$, the proposed WDCCR-BD still has the similar pattern, but its classification accuracies have been increased in comparison with them when $\lg(\lambda) = -3, -2, -1$. We can also see that the classification accuracies of WDCCR-BR are significantly varied when $\lg(\lambda) = 0, 1, 2$, compared to the similar pattern of WDCCR-BR when $\lg(\lambda) = -3, -2, -1$. The better classification accuracies of the proposed methods when $\lg(\lambda) = 0, 1, 2$ means the large value of λ can well adjust the weighted representation coefficients for favorable classification. For further illustrating the classification performance of



(a) WDCCR-BD($\lg(\lambda) = -3$) (b) WDCCR-BD($\lg(\lambda) = -2$) (c) WDCCR-BD($\lg(\lambda) = -1$)

(d) WDCCR-BD($\lg(\lambda) = 0$) (e) WDCCR-BD($\lg(\lambda) = 1$) (f) WDCCR-BD($\lg(\lambda) = 2$)

(g) WDCCR-BR($\lg(\lambda) = -3$) (h) WDCCR-BR($\lg(\lambda) = -2$) (i) WDCCR-BR($\lg(\lambda) = -1$)

(j) WDCCR-BR($\lg(\lambda) = 0$) (k) WDCCR-BR($\lg(\lambda) = 1$) (l) WDCCR-BR($\lg(\lambda) = 2$)

Figure 3: The classification accuracies (%) of WDCCR-BD and WDCCR-BR with different values of the parameters λ, γ and β on Leaf.

Table 1: The best classification accuracies (%) for one given λ on different data sets

$\lg(\lambda)$		-3	-2	-1	0	1	2
Leaf ($l = 6$)	WDCCR-BD	61.93	61.93	62.44	63.45	**63.96**	63.45
	WDCCR-BR	61.93	61.93	61.93	62.94	**63.96**	**63.96**
Swedish Leaf ($l = 20$)	WDCCR-BD	81.94	81.94	81.94	82.06	**82.18**	81.82
	WDCCR-BR	81.94	81.94	81.94	81.94	82.30	**82.67**
GT ($l = 10$)	WDCCR-BD	84.40	84.40	84.80	84.80	**85.60**	85.20
	WDCCR-BR	84.40	84.40	84.80	**85.20**	84.80	84.40

the proposed methods with the variations of the parameter λ, we also list the best classification accuracies of WDCCR-BD and WDCCR-BR at each value of λ on Leaf, Swedish Leaf and GT in Table 1. Note that 20 training samples per class on Swedish Leaf and 10 training samples per class on GT are chosen as the training samples and the remaining ones per class are used as the testing samples. The best classification results among the range of λ are indicated in bold-face. It can be seen from Table 1 that the proposed methods nearly obtain the same classification results on each data set when $\lg(\lambda) = -3, -2, -1$, and the better classification results are achieved when $\lg(\lambda) = 0, 1, 2$. Thus, the values of λ are empirically set as $\{0.001, 0.01, 0.1, 1, 10, 100\}$ in our experiments.

Furthermore, the classification performance of the proposed methods with varying the values of β and γ is also clearly displayed at each value of λ in Fig. 3. We can observe that the classification accuracies of the proposed WDCCR-BD and WDCCR-BR increase quickly with the increase of the values of β and then increase slowly or tend to be stable when $\lg(\gamma) = -3, -2, -1, 0$. And when $\lg(\gamma) = 1, 2$, the classification accuracies of the proposed methods almost increase slowly and finally tends to stable. It is obvious that more effectiveness and robustness of the proposed methods are always achieved when $\lg(\gamma) = 1, 2$ and $\lg(\beta) = 2, 1, 0, -1$. As discussed above, the values of λ, β and γ play the different roles in adjusting the corresponding constraints in the proposed models on the Leaf data set for good classification. Therefore, we empirically set the ranges of λ, β and γ as $\{0.001, 0.01, 0.1, 1, 10, 100\}$ for our proposed methods in the next experiments.

In fact, these three parameters dynamically adjust and balance their corresponding constraint term using their different values in the proposed models. From the experimental results in Fig. 3, we can also see that the competitive term, discriminative term and the weighted term with their own appropriate regularized parameters can obtain good classification in the proposed WDCCR-BD and WDCCR-BR.

4.3. Experimental Results on Data without Noises

In this subsection, we conduct the experiments to verify the classification performance of the proposed WDCCR-BD and WDCCR-BR methods on the GT, ORL, PolyU Palmprint, Leaf, Fifteen Scene Category and Swedish Leaf image data sets without the added noises. The proposed WDCCR-BD and WDCCR-BR are compared with the very related state-of-the-art representation-based classification methods including NSC (Lee et al., 2005), CCRC (Yuan et al., 2018), CRC (Zhang et al., 2011), CROC (Chi and Porikli, 2014), DSRC (Xu et al., 2017), EProCRC (Lan and Zhou, 2017), ProCRC (Cai et al., 2016), SRC (Wright et al., 2009) and CNRC (Waqas et al., 2013). In the experiments, each data set is randomly divided into the training and testing sets ten times, and the number l of the image samples per class are randomly chosen as the training samples and the rest are the testing samples each time. The classification results of each method are the averages of ten best accuracies on each data set and the corresponding standard deviations

Table 2: The best classification accuracies of each method with the corresponding standard deviations on GT

l	2	4	6	8	10
NSC	50.49±1.31	64.69±1.26	73.11±1.58	77.77±1.47	81.12±2.43
CCRC	51.88±1.24	65.85±1.83	75.07±1.22	78.97±2.41	82.80±2.83
CRC	47.53±0.70	60.25±1.12	67.06±1.52	71.37±1.46	74.08±1.77
CROC	50.49±1.31	64.69±1.26	73.11±1.58	77.77±1.47	81.12±2.43
DSRC	52.95±1.10	66.11±1.41	75.02±2.03	78.69±2.65	82.96±2.85
EProCRC	44.68±1.63	55.67±1.68	63.29±1.44	64.46±1.74	67.20±1.88
ProCRC	49.14±1.23	61.13±1.38	69.82±0.71	71.49±1.48	74.88±2.09
SRC	48.34±1.89	60.65±1.60	68.76±1.38	71.20±1.80	72.40±2.53
CNRC	51.38±0.75	63.24±2.09	71.20±1.38	73.26±2.34	76.64±1.82
WDCCR-BD	**54.65±1.34**	**66.69±1.77**	**75.87±1.54**	**79.31±2.72**	**83.28±2.66**
WDCCR-BR	**53.20±0.88**	**66.33±1.53**	**75.47±1.75**	**79.20±2.68**	**83.12±2.75**

Table 3: The best classification accuracies of each method with the corresponding standard deviations on ORL

l	2	3	4	5	6
NSC	79.94±1.49	87.86±2.33	94.67±1.54	94.70±1.48	96.63±1.02
CCRC	84.19±0.65	91.57±1.78	96.17±1.36	96.40±1.52	97.75±0.64
CRC	81.50±0.75	89.00±1.91	93.58±1.88	94.30±0.91	96.63±0.75
CROC	82.31±0.72	89.29±1.75	94.67±1.36	94.70±1.48	97.13±0.85
DSRC	84.56±1.14	91.71±1.22	96.00±1.27	96.30±1.44	**97.88±0.64**
EProCRC	79.75±1.02	87.86±1.94	93.42±1.80	93.20±1.82	94.88±0.83
ProCRC	82.88±1.28	91.79±1.41	95.92±1.08	95.70±1.44	96.50±0.50
SRC	81.31±1.20	89.14±2.22	93.83±1.26	94.70±1.30	95.75±0.61
CNRC	82.19±1.06	90.07±1.22	94.67±1.43	94.70±1.35	96.50±0.31
WDCCR-BD	**85.00±0.80**	**92.07±1.59**	**96.58±1.08**	**96.70±1.35**	**98.13±0.56**
WDCCR-BR	**84.81±0.90**	**92.21±1.68**	**96.58±0.95**	**96.70±1.35**	**97.88±0.64**

on each data set are also calculated. The preseted values of l are in the ranges of 2 to 10 with a step 2 on GT, 2 to 6 with a step 1 on ORL, 5 to 9 with a step 1 on PolyU Palmprint, 2 to 6 with a step 1 on Leaf, 14 to 38 with a step 6 on Fifteen Scene Category and 18 to 16 with step 2 on Swedish Leaf.

The best classification results with the corresponding standard deviations of all the competing methods are listed in Table 2 on GT, Table 3 on ORL, Table 4 on PolyU Palmprint, Table 5 on Leaf, Table 6 on Fifteen Scene Category and Table 7 on Swedish Leaf. Note that the classification accuracies of the proposed methods are denoted in bold-face and the classification accuracies of the state-of-the-art methods are also denoted in bold-face if they are better than that in one of the propose methods. As can be seen from these six tables, the classification accuracies of each method significantly increase with increasing the number l of the training samples from each class on all the data sets. We can obviously observe that the proposed WDCCR-BD and WDCCR-BR methods nearly perform better than the other competing representation-based methods on each data set. Although the proposed methods have the same competitive term as CCRC and the similar discriminative term to DSRC, the classification results in these tables clearly show that the proposed methods almost outperform CCRC and DSRC. This experimental fact implies that the integration of the competitive and discriminative terms with the weighted constraints of the

Table 4: The best classification accuracies of each method with the corresponding standard deviations on PolyU Palmprint

l	5	6	7	8	9
NSC	95.18±0.44	96.34±0.36	95.57±2.80	98.16±0.20	98.35±0.26
CCRC	95.55±0.46	97.05±0.41	97.30±0.30	98.07±0.21	98.31±0.27
CRC	95.13±0.67	95.52±0.37	96.73±0.52	97.27±0.21	97.62±0.29
CROC	95.20±0.54	96.52±0.31	96.73±0.52	98.16±0.20	98.35±0.26
DSRC	95.58±0.46	95.43±3.83	96.14±2.62	98.03±0.25	96.65±2.29
EProCRC	94.58±0.52	96.57±0.41	97.23±0.23	98.06±0.25	**98.40±0.28**
ProCRC	92.97±0.52	95.44±0.46	96.32±0.38	97.42±0.14	97.92±0.25
SRC	92.15±0.62	93.94±0.64	94.85±0.53	96.31±0.24	96.76±0.17
CNRC	90.37±0.52	92.84±0.63	93.79±0.52	95.17±0.44	95.44±0.30
WDCCR-BD	**95.70±0.46**	**97.18±0.41**	**97.41±0.29**	**98.18±0.21**	**98.41±0.27**
WDCCR-BR	**95.69±0.46**	**97.17±0.41**	**97.44±0.32**	**98.19±0.22**	**98.39±0.26**

Table 5: The best classification accuracies of each method with the corresponding standard deviations on Leaf

l	2	3	4	5	6
NSC	41.88±2.38	49.77±2.95	55.76±2.34	58.71±1.57	61.22±2.90
CCRC	44.22±2.41	51.02±2.26	57.17±2.08	60.17±1.39	62.84±2.62
CRC	36.83±1.24	41.18±3.06	46.99±2.10	49.79±1.27	49.24±2.00
CROC	41.88±2.38	49.77±2.95	55.76±2.34	58.71±1.57	61.22±2.90
DSRC	43.05±2.44	50.82±3.17	56.06±1.99	59.31±1.39	63.15±1.78
EProCRC	34.55±2.44	38.56±2.72	39.85±2.15	39.23±0.75	39.09±2.85
ProCRC	36.54±1.79	41.64±3.55	45.50±2.93	46.35±1.92	46.90±2.74
SRC	34.49±2.96	40.59±2.43	42.68±3.89	44.72±2.86	44.67±3.26
CNRC	43.23±1.86	47.67±2.37	52.49±1.75	52.36±1.86	53.30±2.34
WDCCR-BD	**44.40±1.97**	**51.87±2.18**	**58.36±1.39**	**61.03±1.13**	**64.47±2.15**
WDCCR-BR	**44.87±1.92**	**51.67±2.56**	**57.99±1.90**	**61.12±1.71**	**64.37±1.77**

representation coefficients can well enhance the power of pattern discrimination. In addition, we can see the proposed WDCCR-BD obtains the similar but different classification performance with WDCCR-BR on each data sets. This means that both the weighted constraints of the representation coefficients with the distance-based weight factor and the representation-based weight factor can strengthen the ability of competitive and discriminative representation under the integration of both competitive and discriminative terms. Therefore, it can be concluded that the proposed WDCCR-BD and WDCCR-BR are the promising classification methods.

4.4. Experimental Results on Data with Noises

In this subsection, the robust classification performance of the proposed R-WDCCR-BD and R-WDCCR-BR is mainly explored by conducting the experiments on GT, ORL and PolyU Palmprint image data sets with the noises by adding the random corruptions and block occlusions. In the experiments, we compare both R-WDCCR-BD and R-WDCCR-BR with NSC (Lee et al., 2005), CCRC (Yuan et al., 2018), CRC (Zhang et al., 2011), CROC (Chi and Porikli, 2014), DSRC (Xu et al., 2017), EProCRC (Lan and Zhou, 2017), ProCRC (Cai et al., 2016), SRC (Wright et al., 2009) and CNRC (Waqas et al., 2013). For further doing the good comparisons, we also compare both R-WDCCR-BD and R-WDCCR-BR with the robust ProCRC (R-ProCRC) (Cai

Table 6: The best classification accuracies of each method with the corresponding standard deviations on Fifteen Scene Category

l	14	20	26	32	38
NSC	90.45±0.59	95.18±0.44	96.34±0.36	95.57±2.80	98.06±0.20
CCRC	92.16±0.49	95.55±0.46	97.05±0.41	97.30±0.30	98.07±0.21
CRC	92.02±0.50	95.13±0.67	95.52±0.37	96.73±0.52	97.27±0.21
CROC	92.03±0.46	95.20±0.54	96.52±0.31	96.73±0.52	98.06±0.20
DSRC	92.09±0.45	95.58±0.46	95.43±3.83	96.14±2.62	98.03±0.25
EProCRC	89.90±0.71	94.58±0.52	96.57±0.41	97.23±0.23	98.06±0.25
ProCRC	88.03±0.83	92.97±0.52	95.44±0.46	96.32±0.38	97.42±0.14
SRC	87.49±0.69	92.15±0.62	93.94±0.64	94.85±0.53	96.31±0.24
CNRC	85.71±0.76	90.37±0.52	92.84±0.63	93.79±0.52	95.17±0.44
WDCCR-BD	**92.39±0.50**	**95.70±0.46**	**97.18±0.41**	**97.41±0.29**	**98.18±0.21**
WDCCR-BR	**92.38±0.50**	**95.69±0.46**	**97.17±0.41**	**97.44±0.32**	**98.19±0.22**

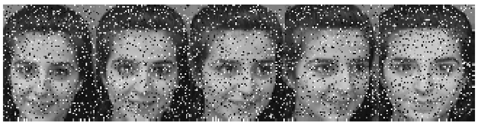
Table 7: The best classification accuracies of each method with the corresponding standard deviations on Swedish Leaf

l	18	20	22	24	26
NSC	81.68±1.82	82.50±1.21	84.53±1.02	84.37±0.59	83.86±0.43
CCRC	82.08±1.86	82.81±1.18	84.58±1.00	84.52±0.51	84.65±0.37
CRC	77.52±1.10	78.50±1.37	79.85±1.61	80.18±0.96	79.59±0.89
CROC	81.68±1.82	82.50±1.21	84.53±1.02	84.37±0.59	83.86±0.43
DSRC	80.30±3.79	73.16±18.85	80.88±4.28	82.56±2.68	81.28±2.34
EProCRC	72.14±0.84	72.51±1.60	73.74±1.14	74.56±1.75	73.88±1.04
ProCRC	74.67±1.11	76.41±1.51	77.53±1.36	78.90±0.56	77.69±0.44
SRC	74.53±1.29	74.79±1.41	76.45±1.20	77.20±1.43	76.00±1.56
CNRC	75.77±1.12	76.00±1.24	77.31±1.45	78.38±1.64	77.66±1.35
WDCCR-BD	**82.18±1.87**	**83.01±1.14**	**84.65±1.00**	**84.68±0.46**	**84.82±0.44**
WDCCR-BR	**82.50±1.70**	**83.22±1.15**	**84.91±0.96**	**84.89±0.53**	**84.84±0.33**

et al., 2016), because R-ProCRC uses the same l_1-norm representation fidelity as R-WDCCR-BD and R-WDCCR-BR to obtain the robust classification performance. Meanwhile, we also verify the classification performance of the proposed WDCCR-BD and WDCCR-BR under the random corruptions and block occlusions. In our experiments, the first l training samples per class are chosen as the training samples and the remaining ones are as the testing samples on each used data set. The values of l are set as $l = 6, 9$ on GT, $l = 4, 6$ on ORL and $l = 6, 10$ on PolyU Palmprint.

In the experiments, the noises are added into all the testing images in the used three data sets by the random corruptions and block occlusions. For random corruptions, each testing image is corrupted by randomly replacing the original pixels with uncertain gray scale values between 0 and 255. The rates of the corrupted size to the original size of each image are from 0.1 to 0.6 with a step 0.1. For block occlusions, a block size of each testing image is randomly replaced by the image of panda. The rates of the block size to the original size of each image are also from 0.1 to 0.6 with a step 0.1. As an example, Fig. 4 shows some testing images with 20% random corruption rate and 20% block occlusion rate on ORL data set.

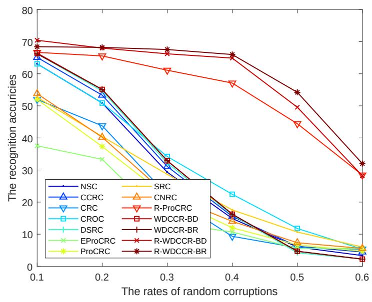
The recognition accuracies of the competing methods on the GT, ORL and PolyU Palmprint



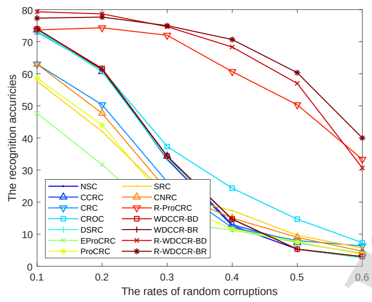
(a) Random corruptions

(b) Block occlusions

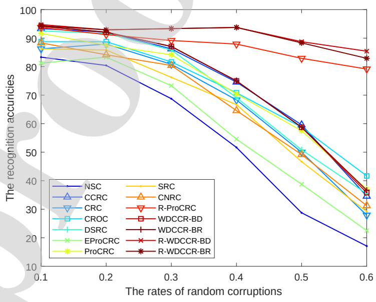
Figure 4: Some image samples with random corruptions or block occlusions on ORL.



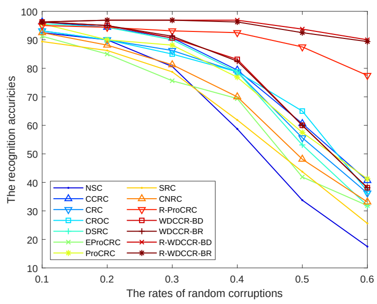
(a) GT ($l = 6$)



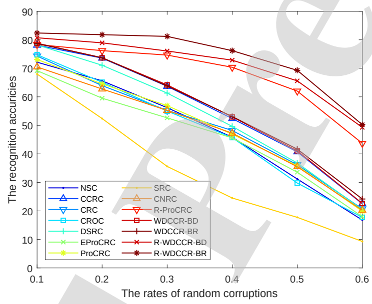
(b) GT ($l = 9$)



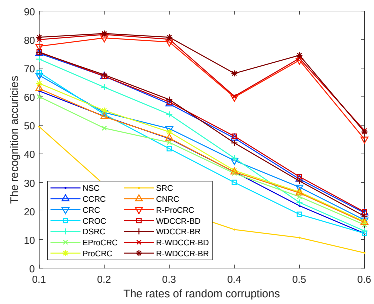
(c) ORL ($l = 4$)



(d) ORL ($l = 6$)



(e) PolyU Palmprint ($l = 6$)



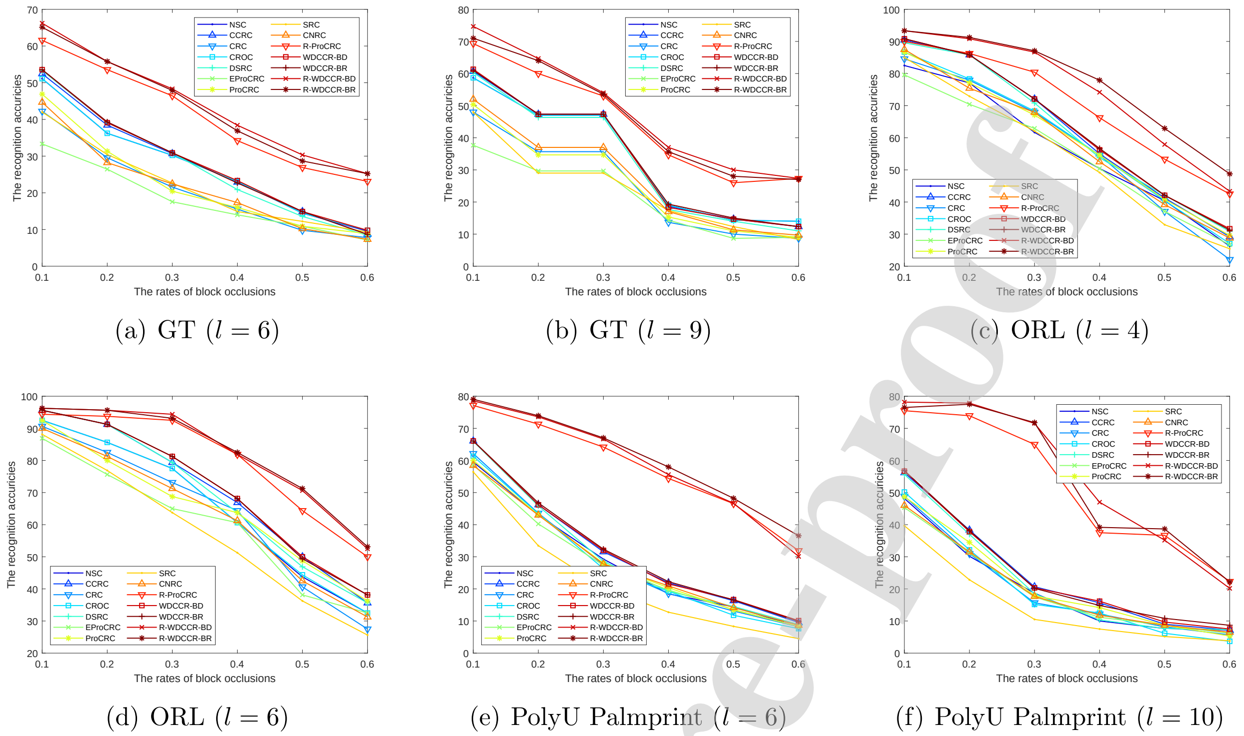
(f) PolyU Palmprint ($l = 10$)

Figure 5: Recognition accuracies (%) of the competing methods on each data set with random corruptions.

image data sets with random corruptions and block occlusions are illustrated in Figs. 5 and 6, respectively. It can be seen that the proposed R-WDCCR-BD and R-WDCCR-BR obtain more robust and effective classification performance than the other competing methods at each value of the corrupted and occluded ratios. And the R-ProCRC method achieves the similar pattern as R-WDCCR-BD and R-WDCCR-BR, and performs better than NSC, CCRC, CRC, CROC, DSRC, EProCRC, ProCRC, SRC and CNRC. Moreover, it is clear that the R-WDCCR-BD, R-WDCCR-BR and R-ProCRC methods are very more robust than NSC, CCRC, CRC, CROC, DSRC, EProCRC, ProCRC, SRC and CNRC with increasing the corrupted and occluded rates of the testing samples. The more robust and effective classification performance of R-WDCCR-BD, R-WDCCR-BR and R-ProCRC means that the representation-based classification methods with the l_1-norm representation fidelity can enhance the more power of pattern discrimination for classification. From the experimental results in Figs. 5 and 6, we can see that the proposed WDCCR-BD and WDCCR-BR methods also achieve the competitive classification performance.

In summary, several important experimental observations from the extensive experiments above can be concluded as follows:

1. The designed two weight factors to constrain the representation coefficients can enhance



(a) GT ($l = 6$) (b) GT ($l = 9$) (c) ORL ($l = 4$)

(d) ORL ($l = 6$) (e) PolyU Palmprint ($l = 6$) (f) PolyU Palmprint ($l = 10$)

Figure 6: Recognition accuracies (%) of the competing methods on each data set with block occlusions.

pattern discrimination.

2. The joint of the competitive and discriminative terms with the weighted constraints in the proposed unified models can significantly improve the classification performance.
3. The parameter selections of the proposed models can be easily realized.
4. The proposed methods achieve the better classification performance than the state-of-the-art representation-based classification methods on data with and without noises.

Therefore, the proposed four WDCCR methods have good robustness and effectiveness for classification, and the proposed weighted competitive and discriminative collaborative representation is also one of the promising collaborative representation methods.

5. Discussions

In this section, we further discuss the effectiveness and efficiency of the proposed methods by analyzing the computational complexity of the proposed methods and the differences between the proposed WDCCR and the very related methods including CCRC (Yuan et al., 2018) and DSRC (Xu et al., 2017).

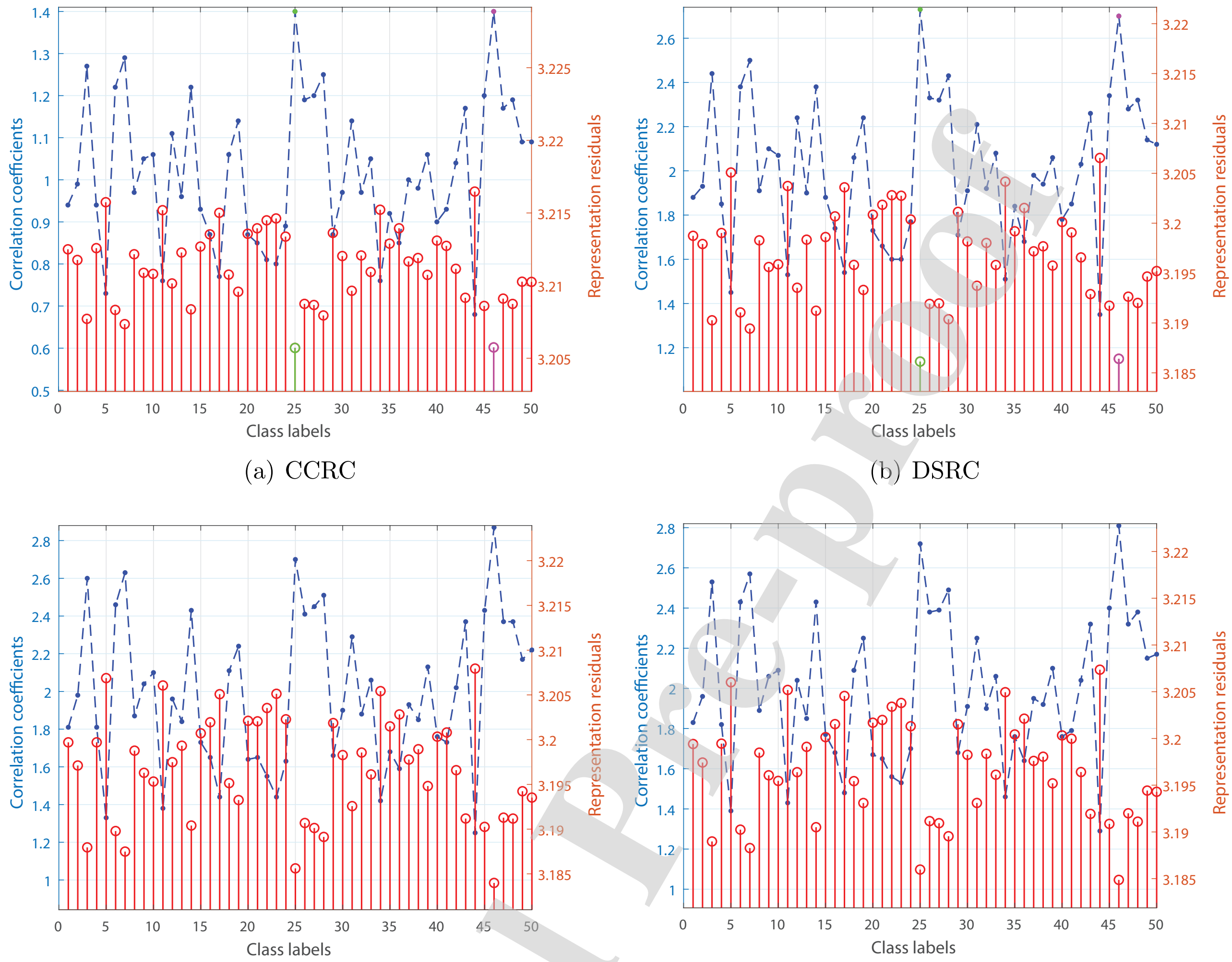
5.1. Analysis of Differences between WDCCR and CCRC, DSRC

In this subsection, we analyze the differences between the proposed WDCCR and CCRC, DSRC for further highlighting the good pattern discrimination for classification. As stated in the proposed WDCCR, the collaborative representation for each query sample is competitively and discriminatively learned by all the training samples with fully using the class information. The competitive and discriminative collaborative representation for favorable classification mainly benefits from the competitive term $\sum_{i=1}^{C} \gamma \parallel y - X_i\alpha_i \parallel_2^2$, the discriminative term $\sum_{i=1}^{C}\sum_{j\neq i}^{C} \beta \parallel X_i\alpha_i + X_j\alpha_j \parallel_2^2$ and the weighted constraint term $\sum_{i=1}^{C} \lambda w_i \parallel \alpha_i \parallel_2^2$. According to the CCRC model in Eq. (3), WDCCR and CCRC have the same competitive term. Although CCRC has good ability of competitive representation from each class, the similar classes could competitively have similar representation to degrade the classification performance, which has been stated in subsection 3.1. To overcome the issue, apart from the competitive constraint in the proposed WDCCR, WDCCR not only introduces the discriminative constraint of the different class-specific representations, but also designs the weighted constraint of class-specific representation coefficients. So that the proposed WDCCR can learn more pattern discrimination than CCRC. In contrast to DSRC, the proposed WDCCR adopts the similar discriminative term in the DSRC model in Eq. (2). Unlike DSRC, the proposed WDCCR also takes both the competitive constraint of the class-specific residuals and the weighted constraint of class-specific representation coefficients into account. Moreover, WDCCR with the competitive constraint contains the classification decision in the representation, but DSRC doesn't contain the classification decision. Accordingly, WDCCR further competitively obtains more pattern discrimination among the different classes for favorable classification than DSRC. As a consequence, the proposed WDCCR well holds the superiorities of both CCRC and DSRC, and possesses the strong inter-class pattern discrimination through competitive and discriminative collaborative representation.

For visually demonstrating the superior competitive and discriminative representation of WDCCR over CCRC and DSRC for classification, we give an experimental example of a given query sample y by using the correlation coefficient $Corr(y, X_i\alpha_i) = y^T(X_i\alpha_i)$ in Eq. (5) and the representation residual $\|y - X_i\alpha_i\|_2^2$ between y and $X_i\alpha_i$. In the example, the given query sample is from class 46 in the GT data set, and the first 8 samples and last 7 samples per class are chosen as the training and testing samples, respectively. Note that the experimental settings in this example are the same as in Fig. 1. The correlation coefficients and the representation residuals from all the classes obtained by CCRC, DSRC, WDCCR-BD and WDCCR-BR for the query sample are displayed in Fig. 7. Note that the green solid line denotes the class that CCRC and DSRC wrongly classify y into and the magenta solid line denotes the true class that y belongs to in Figs. 7(a) and 7(b). We can see that CCRC obtains the very similar or even same correlation coefficients and the representation residuals from classes 25 and 46, and DSRC also obtains the similar results, so both fail to classify the given query sample. The possible reason is that the inter-class pattern discrimination in CCRC and the competitive representation for classification in DSRC are not always good. However, the proposed WDCCR considers the competitive and discriminative representation simultaneously with the constraints of weighted class-specific representation coefficients. As shown in Figs. 7(c) and 7(d), the proposed WDCCR-BD and WDCCR-BR increase differences of the correlation coefficients and the representation residuals between classes 25 and 46. This implies that the proposed WDCCR can hold the more power of pattern discrimination among all the classes than CCRC and DSRC. Thus, the effective classification performance of the proposed methods can be further well verified

5.2. Computational Complexity of WDCCR

In this subsection, the computational complexities of the proposed WDCCR and R-WDCCR methods are analyzed for embodying their efficiency for classification. As stated in Section 3, the main computational time is to calculate Eq. (21) in WDCCR and Eq. (25) in R-WDCCR. For easily calculating the computational complexities, some used denotations are first given. These denotations are the training set of all the C classes $X \in \mathbb{R}^{m\times N}$, the training subset of class i $X_i \in \mathbb{R}^{m\times n_i}$, the given query sample $y \in \mathbb{R}^{m\times 1}$, $M \in \mathbb{R}^{N\times N}$ in Eqs. (21) and (25) and $A \in \mathbb{R}^{m\times m}$



(a) CCRC (b) DSRC

(c) WDCCR-BD (d) WDCCR-BR

Figure 7: The examples of the correlation coefficients (blue dash-dot line) and representation residuals (red solid line) obtained by CCRC, DSRC, WDCCR-BD and WDCCR-BR for a query sample from class 46 on GT.

in Eq. (25). Besides, there is the weight matrix $W \in \mathbb{R}^{N\times N}$ in Eqs. (21) and (25). To clearly differentiate two types of weighted constraints, let $W_{BD} = W^T W \in \mathbb{R}^{N\times N}$ denote the distance-based weight matrix used in WDCCR-BD and R-WDCCR-BD, and $W_{BD} = W^T W \in \mathbb{R}^{N\times N}$ denote the representation-based weight matrix used in WDCCR-BR and R-WDCCR-BR.

To get the computational complexities of the proposed methods, the computational complexities of W_{BD} and W_{BR} are first calculated. The calculation of W_{BD} is to mainly compute distance between y and each training sample in Eq. (26), and its main complexity is $O(m^2N)$. To calculate W_{BR} is to mainly compute $\bar{\alpha}_i = (X_i^T X_i)^{-1} X_i^T y$ for C classes in Eq. (27). Since $F_1 = X_i^T X_i$ has the complexity $O(mn_i^2)$, $(F_1)^{-1}$ has $O(n_i^3)$ and $F_1^{-1} X_i^T y$ has $O(mn_i^2 + mn_i)$, the total complexity of $\bar{\alpha}_i = (X_i^T X_i)^{-1} X_i^T y$ for C classes is $O(\sum_{i=1}^{C}(mn_i^2 + n_i^3 + mn_i))$. For simplicity, we use $O(mN^2 + N^3 + mN)$ [4] to replace $O(\sum_{i=1}^{C}(mn_i^2 + n_i^3 + mn_i))$. Thus, the main complexity of W_{BR}

[4]Because $\sum_{i=1}^{C} mn_i^2$ is less than $m(\sum_{i=1}^{C} n_i)^2 = mN^2$, $\sum_{i=1}^{C} n_i^3$ is less than $(\sum_{i=1}^{C} n_i)^3 = N^3$ and $\sum_{i=1}^{C} mn_i$ is equal to mN, $O(\sum_{i=1}^{C}(mn_i^2 + n_i^3 + mn_i)$ is less than $O(mN^2 + N^3 + mN)$.

is $O(mN^2 + N^3 + mN)$.

The complexity of WDCCR is to mainly calculate $\alpha = (1+\gamma)((1+2\beta)X^TX + (2\beta(C-2)+\gamma)M + \lambda W^TW)^{-1}X^Ty$. It is easy to get the complexity $O(mN^2)$ of X^TX, $O(\sum_{i=1}^{C} mn_i^2)$ of M, $O(N^3)$ of F_2^{-1} where $F_2 = (1+2\beta)X^TX + (2\beta(C-2)+\gamma)M + \lambda W^TW$, and $O(mN^2 + mN)$ of $F_2^{-1}X^Ty$. Note that we simply use $O(mN^2)$ to replace $O(\sum_{i=1}^{C} mn_i^2)$. Thus, the complexity of $(1+\gamma)((1+2\beta)X^TX+(2\beta(C-2)+\gamma)M+\lambda W^TW)^{-1}X^Ty$ is $O(mN^2+N^3+mN)$. Although W_{BD} and W_{BR} have different complexities, their corresponding proposed WDCCR-BD and WDCCR-BR have the similar computational complexities, and their main complexities are $O(mN^2+N^3+mN)$.

The computational complexity of the proposed R-WDCCR is to iteratively calculate $\alpha = \big((2\beta(C-2)+\gamma)M + 2\beta X^TX + X^TAX + \lambda W^TW\big)^{-1}(\gamma X^T + X^TA)y$ until the preseted iteration number T is reached. To obtain the complexity of R-WDCCR, we should calculate the complexities $O(Nm^2+mN^2)$ of X^TAX and $O(Nm^2+mN)$ of X^TAy with the other similar complexities to WDCCR. Thus, the main complexities of the proposed R-WDCCR-BD and R-WDCCR-BR methods are $O(TNm^2+TmN^2+TN^3+TmN)$. In fact, since most of collaborative representation-based classification methods have the closed-form solution of the representation coefficients, they have similar computational complexities for efficient classification.

Table 8: The average computational time (milli-seconds) of each competing method for representing and classifying any one testing image sample on ORL and GT

ORL	NSC	CCRC	CRC	CROC	DSRC	WDCCR-BD	WDCCR-BR
	14.41	16.61	7.48	22.51	10.07	14.12	14.00
	SRC	CNRC	ProCRC	EProCRC	R-ProCRC	R-WDCCR-BD	R-WDCCR-BR
	17.71	7.43	12.23	16.23	107.53	119.78	117.37
GT	NSC	CCRC	CRC	CROC	DSRC	WDCCR-BD	WDCCR-BR
	20.08	20.90	11.00	29.00	14.00	19.00	19.50
	SRC	CNRC	ProCRC	EProCRC	R-ProCRC	R-WDCCR-BD	R-WDCCR-BR
	24.90	13.30	15.20	19.90	190.30	184.60	213.00

After analyzing the computational complexities of four proposed methods, we experimentally give their computational time on ORL and GT, in comparison with NSC (Lee et al., 2005), CCRC (Yuan et al., 2018), CRC (Zhang et al., 2011), CROC (Chi and Porikli, 2014), DSRC (Xu et al., 2017), ProCRC and R-ProCRC (Cai et al., 2016), EProCRC (Lan and Zhou, 2017), SRC (Wright et al., 2009) and CNRC (Waqas et al., 2013). Note that all our experiments have been conducted by using MATLAB 2019a and Windows 10 professional operation system with Intel Core i7-8565U CPU and 8 GB RAM. Here, the numbers of training samples from each class are set as $l = 6$ on ORL and $l = 8$ on GT, and the remaining ones are the testing samples. To objectively reflect the computational time of each competing method, the computational time for representing and classifying each testing image sample is the average of the overall running time for all the testing image samples. The average computational time (milli-seconds) of each competing method for representing and classifying any one testing image sample on ORL and GT is shown in Table 8. From the computational time in the table, we can observe three facts: NSC, CCRC, ProCRC, EProCRC and the proposed WDCCR-BD and WDCCR-BR have the similar running time; CRC, DSRC and CNRC have the similar running time; R-ProCRC and the proposed R-WDCCR-BD and R-WDCCR-BR have the similar running time. Thus, the experimental facts of the running time imply that the theoretical calculation time of the method determine its running time.

6. Conclusions

In this article, with the aim of improving the pattern discrimination, we proposed the weighted discriminative collaborative competitive representation (WDCCR) for image classification. In the proposed WDCCR, the discriminative constraint of pairs of categorical representations, the competitive constraint of categorical reconstructive residuals and the weighted constraint of the categorical representation coefficients are integrated into unified model by fully using class information in each constraint. In the weighted constraint, two types of defining class-specific weights are also designed for constraining the categorical representation coefficients. Through the discriminative and competitive collaborative representation, the inter-class pattern discrimination are significantly enhanced for favorable image classification. Moreover, using the robustness of l_1-norm representation fidelity on the representation-based classification, we also propose robust WDCCR (R-WDCCR) with l_1-norm instead of l_2-norm representation fidelity. The effective and robust classification performance of the proposed methods is verified on six public image data sets in comparison with the state-of-the-art representation-based classification methods. The extensive experiments show that the proposed methods perform very well for image classification.

In the future work, we will mainly plan to study the discriminative competitive collaborative dictionary learning for classification. Moreover, since the representation coefficients has more discrimination information, the proposed weighted discriminative collaborative competitive representation can be used for designing other types of classifiers. For example, we can use the representation coefficients and residuals to determine the nearest samples of each query sample and accordingly develop the k-nearest neighbor classification. In addition, the robustness and effectiveness of the proposed methods for image classification means that they can be well applied for many real-world classification tasks in computer vision, such as object recognition, speech recognition and image retrieval.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 61976107, U1836220, 61962010, 61672268 and 61502208), the Natural Science Foundation of Jiangsu Province of China (Grant No. BK20150522), International Postdoctoral Exchange Fellowship Program of China Postdoctoral Council (No. 20180051), the Research Foundation for Talented Scholars of JiangSu University (Grant No. 14JDG037), Open Foundation of Artificial Intelligence Key Laboratory of Sichuan Province (Grant No. 2017RYJ04), and the Sichuan Science and Technology Program (2018RZ0072).

Compliance with ethical standards

Conflict of Interests

All the authors declare that there are no conflicts of interests regarding the publication of this article.

References

Akhtar, N., Shafait, F., Mian, A., 2017. Efficient classification with sparsity augmented collaborative representation. Pattern Recognit. 65, 136–145. doi:10.1016/j.patcog.2016.12.017.

Benuwa, B.B., Zhan, Y., Liu, J., Gou, J., Ghansah, B., Ansah, E.K., 2019. Group sparse based locality–sensitive dictionary learning for video semantic analysis. Multimed. Tools Appl. 78, 6721–6744. doi:10.1007/s11042-018-6417-3.

Cai, S., Zhang, L., Zuo, W., Feng, X., 2016. A probabilistic collaborative representation based approach for pattern classification, in: IEEE Int. Conf. Comput. Vis. Pattern Recognit., pp. 2950–2959. doi:10.1109/cvpr.2016.322.

Chang, X., Nie, F., Yang, Y., Zhang, C., Huang, H., 2016. Convex sparse pca for unsupervised feature learning. ACM Trans. Knowl. Discovery Data 11, 3. doi:10.1145/2910585.

Chartrand, R., Wotao Yin, 2008. Iteratively reweighted algorithms for compressive sensing, in: 2008 IEEE Int. Conf. Acoust. Speech Signal Process., pp. 3869–3872. doi:10.1109/ICASSP.2008.4518498.

Chen, J., Liu, J., Ye, J., 2012. Learning incoherent sparse and low-rank patterns from multiple tasks. ACM Trans. Knowl. Discovery Data 5, 22. doi:10.1145/2086737.2086742.

Chi, H., Xia, H., Zhang, L., Zhang, C., Tang, X., 2018. Competitive and collaborative representation for classification. Pattern Recognit. Lett. , 1–10doi:10.1016/j.patrec.2018.06.019.

Chi, Y., Porikli, F., 2014. Classification and boosting with multiple collaborative representations. IEEE Trans. Pattern Anal. Mach. Intell. 36, 1519–1531. doi:10.1109/TPAMI.2013.236.

Deng, W., Hu, J., Guo, J., 2018. Face recognition via collaborative representation: its discriminant nature and superposed representation. IEEE Trans. Pattern Anal. Mach. Intell. 40, 2513–2521. doi:10.1109/TPAMI.2017.2757923.

Donoho, D.L., 2006. Compressed sensing. IEEE Trans. Inf. Theory 52, 1289–1306. URL: 10.1109/TIT.2006.871582.

Fan, J., Chen, T., Lu, S., 2017. Superpixel guided deep-sparse-representation learning for hyperspectral image classification. IEEE Trans. Circuits Syst. Video Technol. 28, 3163–3173. doi:10.1109/TCSVT.2017.2746684.

Gou, J., Wang, L., Hou, B., Lv, J., Yuan, Y., Mao, Q., 2019a. Two-phase probabilistic collaborative representation-based classification. Expert Syst. Appl. 133, 9–20. doi:10.1016/j.eswa.2019.05.009.

Gou, J., Hou, B., Ou, W., Mao, Q., Yang, H., Liu, Y., 2019b. Several robust extensions of collaborative representation for image classification. Neurocomputing 348, 120–133. doi:10.1016/j.neucom.2018.06.089.

Gou, J., Wang, L., Yi, Z., Lv, J., Mao, Q., Yuan, Y.H., 2018a. A new discriminative collaborative neighbor representation method for robust face recognition. IEEE Access 6, 74713–74727. doi:10.1109/ACCESS.2018.2883527.

Gou, J., Wang, L., Yi, Z., Yuan, Y., Ou, W., Mao, Q., 2019c. Discriminative group collaborative competitive representation for visual classification, in: Proc. IEEE Int. Conf. Multimedia Expo. pp. 1474–1479. doi:10.1109/ICME.2019.00255

Gou, J., Xu, Y., Zhang, D., Mao, Q., Du, L., Zhan, Y., 2018b. Two-phase linear reconstruction measure-based classification for face recognition. Inf. Sciences 433, 17–36. doi:10.1016/j.ins.2017.12.025.

Gui, J., Liu, T., Tao, D., Sun, Z., Tan, T., 2016. Representative Vector Machines: A Unified Framework for Classical Classifiers. IEEE Trans. Cybernetics 46(8), 1877–1888. doi:10.1109/TCYB.2015.2457234.

Gui, J., Tao, D., Sun, Z., Lou, Y., You, X., Tang, Y., 2014. Group sparse multiview patch alignment framework with view consistency for image classification. IEEE Trans. Image Process. 23(7), 3126–3137. doi:10.1109/TIP.2014.2326001.

Jin, T., Liu, Z., Yu, Z., Min, X., Li, L., 2017. Locality preserving collaborative representation for face recognition. Neural Process. Lett. 45, 967–979. doi:10.1007/s11063-016-9558-2.

Lan, R., Zhou, Y., 2017. An extended probabilistic collaborative representation based classifier for image classification, in: Proc. IEEE Int. Conf. Multimedia Expo, pp. 1392–1397. doi:10.1109/ICME.2017.8019308.

Lan, R., Zhou, Y., Liu, Z., Luo, X., 2018. Prior knowledge-based probabilistic collaborative representation for visual recognition. IEEE Trans. Cybernetics. , 1–11doi:10.1109/TCYB.2018.2880290.

Lazebnik, S., Schmid, C., Ponce, J., 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. Proc. IEEE Comput. Vis. Pattern Recognit. 2, 2169–2178. doi:10.1109/CVPR.2006.68.

Lee, K.C., Ho, J., Kriegman, D.J., 2005. Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. Pattern Anal. Mach. Intell. 27, 684–698. doi:10.1109/TPAMI.2005.92.

Li, W., Du, Q., Xiong, M., 2015a. Kernel collaborative representation with tikhonov regularization for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 12, 48–52. doi:10.1109/LGRS.2014.2325978.

Li, W., Du, Q., Zhang, F., Hu, W., 2015b. Collaborative-representation-based nearest neighbor classifier for hyperspectral imagery. IEEE Geosci. Remote Sens. Lett. 12, 389–393. doi:10.1109/LGRS.2014.2343956.

Liu, W., Yu, Z., Lu, L., Wen, Y., Li, H., Zou, Y., 2015a. KCRC-LCD: Discriminative kernel collaborative representation with locality constrained dictionary for visual categorization. Pattern Recognit. 48, 3076–3092. doi:10.1016/j.patcog.2015.04.014.

Liu, Z., Pu, J., Xu, M., Qiu, Y., 2015b. Face recognition via weighted two phase test sample sparse representation. Neural Process. Lett. 41, 43–53. doi:10.1007/s11063-013-9333-6.

Lu, C., Min, H., Gui, J., Zhu, L., Lei, Y., 2013. Face recognition via weighted sparse representation. J. Vis. Commun. Image R. 3, 111–116. doi:10.1016/j.jvcir.2012.05.003.

Ma, H., Gou, J., Wang, X., Ke, J., Zeng, S., 2017. Sparse Coefficient-Based k-Nearest Neighbor Classification. IEEE Access 5, 16618–16634. doi:10.1109/ACCESS.2017.2739807.

Naseem, I., Togneri, R., Bennamoun, M., 2010. Linear regression for face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 32, 2106–2112. doi:10.1109/TPAMI.2010.128.

Ou, W., Luan, X., Gou, J., Zhou, Q., Xiao, W., Xiong, X., Zeng, W., 2018. Robust discriminative nonnegative dictionary learning for occluded face recognition. Pattern Recognit. Lett. 107, 41–49. doi:10.1016/j.patrec.2017.07.006.

Parikh, N., Boyd, S., et al., 2014. Proximal algorithms. Foundations and Trends® in Optimization 1, 127–239. doi:10.1561/2400000003.

Pati, Y.C., Rezaiifar, R., Krishnaprasad, P.S., 1993. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, in: Proceedings of 27th Asilomar Conference on signals, systems and computers, pp. 40–44. doi:10.1109/ACSSC.1993.342465.

Peng, Y., Li, L., Liu, S., Lei, T., Wu, J., 2018. A new virtual samples-based crc method for face recognition. Neural Process. Lett. 48, 313–327. doi:10.1007/s11063-017-9721-4.

Qu, X., Wang, W., Lu, K., Zhou, J., 2018. In-air handwritten chinese character recognition with locality-sensitive sparse representation toward optimized prototype classifier. Pattern Recognit. 78, 267–276. doi:10.1016/j.patcog.2018.01.021.

Sharma, P., Abrol, V., Sao, A.K., Sharma, P., Abrol, V., Sao, A.K., 2017. Deep-sparse-representation-based features for speech recognition. IEEE/ACM Trans. Audio Speech Lang. Process. 25, 2162–2175. doi:10.1109/TASLP.2017.2748240.

Shu, T., Zhang, B., Tang, Y.Y., 2018. Sparse supervised representation-based classifier for uncontrolled and imbalanced classification. IEEE Trans. Neural Netw. Learn. Syst. , 1–10doi:10.1109/TNNLS.2018.2884444.

Silva, P.F., Marcal, A.R., da Silva, R.M.A., 2013. Evaluation of features for leaf discrimination, in: International Conference Image Analysis and Recognition, Springer. pp. 197–204. doi:10.1007/978-3-642-39094-4_23.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition, in: Int. Conf. Learn. Representations. URL: https://arxiv.org/pdf/1409.1556.pdf.

Söderkvist, O., 2001. Computer vision classification of leaves from swedish trees. Master's thesis. Linköping University. URL: http://www.diva-portal.org/smash/get/diva2:303038/FULLTEXT01.pdf.

Song, X., Chen, Y., Feng, Z.H., Hu, G., Zhang, T., Wu, X.J., 2019. Collaborative representation based face classification exploiting block weighted LBP and analysis dictionary learning. Pattern Recognit. 88, 127–138. doi:10.1016/j.patcog.2018.11.008.

Su, H., Zhao, B., Du, Q., Du, P., Xue, Z., 2018. Multifeature dictionary learning for collaborative representation classification of hyperspectral imagery. IEEE Trans. Geosci. Remote Sens. 56, 2467–2484. doi:10.1109/TGRS.2017.2781805.

Tang, X., Feng, G., Cai, J., 2014. Weighted group sparse representation for undersampled face recognition. Neurocomputing 145, 402–415. doi:10.1016/j.neucom.2014.05.012.

Tariyal, S., Majumdar, A., Singh, R., Vatsa, M., 2016. Deep dictionary learning. IEEE Access 4, 10096–10109. doi:10.1109/ACCESS.2016.2611583.

Timofte, R., Van Gool, L., 2014. Adaptive and weighted collaborative representations for image classification. Pattern Recognit. Lett. 43, 127–135. doi:10.1016/j.patrec.2013.08.010.

Wang, D., Lu, H., Yang, M.H., 2015. Kernel collaborative face recognition. Pattern Recognit. 48, 3025–3037. doi:10.1016/J.PATCOG.2015.01.012.

Waqas, J., Yi, Z., Zhang, L., 2013. Collaborative neighbor representation based classification using l2-minimization approach. Pattern Recognit. Lett. 34, 201–208. doi:10.1016/j.patrec.2012.09.024.

Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y., 2009. Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. 31, 210–227. doi:10.1109/TPAMI.2008.79.

Xu, Y., Zhang, D., Yang, J., Yang, J.Y., 2011. A two-phase test sample sparse representation method for use with face recognition. IEEE Trans. Circuits Syst. Video Technol. 21, 1255–1262. doi:10.1109/TCSVT.2011.2138790.

Xu, Y., Zhong, Z., Yang, J., You, J., Zhang, D., 2017. A new discriminative sparse representation method for robust face recognition via l_2 regularization. IEEE Trans. Neural Netw. Learn. Syst. 28, 2233–2242. doi:10.1109/TNNLS.2016.2580572.

Yuan, H., Li, X., Xu, F., Wang, Y., Lai, L.L., Tang, Y.Y., 2018. A collaborative-competitive representation based classifier model. Neurocomputing 275, 627–635. doi:10.1016/j.neucom.2017.09.022.

Zeng, S., Gou, J., Deng, L., 2017. An antinoise sparse representation method for robust face recognition via joint l1 and l2 regularization. Expert Syst. Appl. 82, 1–9. doi:10.1016/j.eswa.2017.04.001.

Zeng, S., Gou, J., Yang, X., 2018. Improving sparsity of coefficients for robust sparse and collaborative representation-based image classification. Neural Comput. and Appl. 30, 2965–2978. doi:10.1007/s00521-017-2900-4.

Zhang, D., Guo, Z., Lu, G., Zhang, L., Zuo, W., 2010. An online system of multispectral palmprint verification. IEEE Trans. Instrum. Meas. 59, 480–490. doi:10.1109/TIM.2009.2028772.

Zhang, L., Yang, M., Feng, X., 2011. Sparse representation or collaborative representation: Which helps face recognition?, in: Proc. IEEE Int. Conf. Comput. Vis., pp. 471–478. doi:10.1109/ICCV.2011.6126277.

Zhang, L., Yang, M., Feng, X., Ma, Y., Zhang, D., 2012. Collaborative representation based classification for face recognition. arXiv preprint URL: https://arxiv.org/ftp/arxiv/papers/1204/1204.2358.pdf.

Zhang, Q., Sun, H., 2018. Probabilistic collaborative representation based orthogonal discriminative projection for image set classification. J. Vis. Commun. Image R. 55, 106–114. doi:10.1016/j.jvcir.2018.05.016.

Zhang, Z., Xu, Y., Yang, J., Li, X., Zhang, D., 2015. A survey of sparse representation: algorithms and applications. IEEE Access 3, 490–530. doi:10.1109/ACCESS.2015.2430359.

Zheng, C., Wang, N., 2019. Collaborative representation with k-nearest classes for classification. Pattern Recognit. Lett. 117, 30–36. doi:10.1016/j.patrec.2018.11.005.

Zheng, Z., Sun, H., 2019. Jointly discriminative projection and dictionary learning for domain adaptive collaborative representation-based classification. Pattern Recognit. doi:10.1016/j.patcog.2019.01.004.

1. Propose two WDCCR methods with l_2-norm fidelity.
2. Propose two robust WDCCR methods with l_1-norm fidelity.
3. Design two weight factors in the proposed WDCCR methods.

Conflict of Interest

All the authors declare that there are no conflicts of interests regarding the publication of this article.