Discriminative non-negative representation based classifier for image recognition

Kai-Jun Hu¹, He-Feng Yin² and Jun Sun²

Abstract
During the past decade, representation based classification method has received considerable attention in the community of pattern recognition. The recently proposed non-negative representation based classifier achieved superb recognition results in diverse pattern classification tasks. Unfortunately, discriminative information of training data is not fully exploited in non-negative representation based classifier, which undermines its classification performance in practical applications. To address this problem, we introduce a decorrelation regularizer into the formulation of non-negative representation based classifier and propose a discriminative non-negative representation based classifier for pattern classification. The decorrelation regularizer is able to reduce the correlation of representation results of different classes, thus promoting the competition among them. Experimental results on benchmark datasets validate the efficacy of the proposed discriminative non-negative representation based classifier, and it can outperform some state-of-the-art deep learning based methods. The source code of our proposed discriminative non-negative representation based classifier is accessible at https://github.com/yinhefeng/DNRC.

Keywords
Image recognition, non-negative representation, representation based classification, decorrelation regularizer

Received 1 February 2021; Revised received 15 April 2021; accepted 23 August 2021

Introduction
Recent years have witnessed the success of representation based classification method (RBCM) in a variety of classification tasks. In face recognition, the most influential work is the sparse representation based classification (SRC).¹ SRC treats all the training samples as a dictionary, and a test sample is sparsely coded over the dictionary, then the classification is performed by checking which class yields the least reconstruction error. Naseem et al.² proposed a linear regression classification (LRC) algorithm which represents the test sample as a linear combination of class-specific training samples. The essence of LRC is nearest subspace classifier (NSC)³ with down-sampled features. Due to the principle of ℓ1-minimization, the coefficients obtained by SRC may vary a lot even for similar test samples. Therefore, SRC tends to lose locality information. Yu et al.⁴ argued that locality is more essential than spars-ity. Inspired by this finding, Lu et al.⁵ presented a weighted SRC (WSRC) method which imposes the locality on the ℓ1 regularization. Similarly, Wang et al.⁶ proposed a locality-constrained linear coding (LLC) scheme. Different from the ℓ1-minimization used in WSRC, LLC employs the ℓ2-norm based regularizer, thus LLC has closed-form solution. In SRC, all the training samples are used to construct the over-complete dictionary; however, not all the training samples have positive contributions in representing the test sample. Xu et al.⁷ developed a two-phase test sample representation (TPTSR) method which determines a certain number of nearest neighbors for the test sample. Zhang et al.⁸ argued that it is the collaborative representation mechanism rather than the ℓ1-norm sparsity that makes SRC successful for face classification, and they presented a collaborative representation based classification (CRC) algorithm. Similar to the idea of WSRC, Timofte and Van Gool⁹ designed a weighted CRC (WCRC)

¹School of IoT Engineering, Jiangsu Vocational College of Information Technology, China
²School of Artificial Intelligence and Computer Science, Jiangnan University, China

Corresponding author:
Kai-Jun Hu, Jiangsu Vocational College of Information Technology, No. 1, Qianou Road, Huishan District, Wuxi, 214153, China.
Email: 7141905017@vip.jiangnan.edu.cn
method. To fully exploit multiple collaborative representations in their formulations, Chi and Porikli proposed a collaborative representation optimized classifier (CROC) which strikes a balance between NSC and CRC. The label information of training data is not considered in SRC. To alleviate this problem, Lai and Jiang developed a class-wise sparse representation (CSR) method which minimizes the number of training classes in representing a class-wise sparse representation (CSR) method which.

Although Zhang et al.8 offered a geometric interpretation of CRC from a probabilistic perspective and proposed a probabilistic CRC (ProCRC). Cai et al.13 analyzed the classification mechanism of CRC and proposed a discriminative CRC (ProCRC). Xu et al.14 presented a discriminative collaborative representation optimized classification method, Xu et al.12 presented a discriminative representation mechanism rather than the ℓ1-norm constraint in CRC with the ℓ1-norm sparsity that makes SRC powerful for classification. And they presented CRC algorithm, which replaces the ℓ1-norm in SRC with the ℓ2-norm constraint, the objective function of CRC is formulated as follows:

$$\min_{\|c\|_1, \text{s.t.} \|y - Xc\|_2^2 \leq \varepsilon}$$

where \(\varepsilon\) is a given error tolerance. When we obtain the coefficient vector \(c\) of \(y\), the test sample \(y\) is classified according to the following formulation:

$$\text{id}(y) = \arg\min_{c} \|y - Xc\|_2$$

where \(c\) is the coefficient vector that corresponds to the \(i\)-th class.

### Collaborative representation based classification

SRC and its variants have achieved encouraging results in diverse pattern classification tasks. Nevertheless, Zhang et al.8 argued that it is the collaborative representation mechanism rather than the ℓ1-norm sparsity that makes SRC powerful for classification. And they presented CRC algorithm, which replaces the ℓ1-norm in SRC with the ℓ2-norm constraint, the objective function of CRC is formulated as follows:

$$\min_{\|c\|_1, \text{s.t.} \|y - Xc\|_2^2 + \lambda\|c\|_2^2}$$

CRC has the following closed-form solution:

$$c = (X^TX + \lambda\mathbf{I})^{-1}X^Ty$$

where \(\mathbf{I}\) is the identity matrix. Let \(P = (X^TX + \lambda\mathbf{I})^{-1}X^T\), we can see that \(P\) is only determined by the training data matrix \(X\). Therefore, when given all the training data, \(P\) can be pre-computed, which makes CRC very efficient. CRC employs the following regularized residual for classification:

$$\text{id}(y) = \arg\min_{c} \frac{\|y - Xc\|_2}{\|c\|_2}$$

### Non-negative representation based classification

SRC and CRC have become two representative methods in RBCM. However, the coding vector of conventional RBCM contains negative entries. The test sample should be better expressed by homogeneous samples with non-
negative representation coefficients. Moreover, Lee and Seung\textsuperscript{15} pointed out that it is unsuitable to approximate the test sample by allowing the training samples to cancel each other out with complex additions and subtractions. Therefore, Xu et al.\textsuperscript{14} proposed the following non-negative representation model by imposing the non-negative constraint on the coding vector:

\[
e_{\text{min}} \|y - Xc\|^2_2, \text{ s.t. } c \geq 0
\]  

(6)

Similar to SRC, NRC employs the class specific residual to classify the test sample, that is, identity(y) = \text{arg min} \|y - Xc_i\|_2.

**Discriminative non-negative representation based classifier**

In this section, first we introduce the formulation of our proposed DNRC, then we present its optimization procedures. Finally, we give the classification scheme of DNRC.

**Proposed model**

According to the mechanism of RBCM, we know that \(Xc_i\) denotes the representation result of the test sample by using the training samples from the \(i\)-th class, and \((Xc_i)^T(Xc_j)\) measures the correlation between the representation results of the training samples from the \(i\)-th and \(j\)-th classes in expressing the test sample. Minimization of \((Xc_i)^T(Xc_j)\) means that the representation results of the \(i\)-th and \(j\)-th classes have the lowest correlation, which encourages the representation results of distinct classes to be discriminative.

By minimizing \(\sum_{i=1}^{K} \sum_{j=1}^{K} (Xc_i)^T(Xc_j)\), the decorrelation effect for different classes can be achieved. Moreover, we have \(\sum_{i=1}^{K} \sum_{j=1}^{K} (Xc_i)^T(Xc_j) = \|Xc\|^2_2\). To promote the competition between different classes by decorrelating the representation results of them in NRC, we incorporate \(\|Xc\|^2_2\) into the formulation of NRC and propose the following DNRC model:

\[
e_{\text{min}} \|y - Xc\|^2_2 + \lambda \|Xc\|^2_2, \text{ s.t. } c \geq 0
\]  

(7)

where \(\lambda > 0\) is a balancing parameter. The first term in equation (7) is the reconstruction error term and the second term is the decorrelation term. One can see that when \(\lambda = 0\), DNRC is degenerated to NRC. Consequently, NRC can be viewed as a special version of DNRC.

**Optimization**

We adopt an alternative strategy to solve the DNRC model. By introducing an auxiliary variable \(z\), equation (7) can be reformulated as

\[
c, z_{\text{min}} \|y - Xc\|^2_2 + \lambda \|Xc\|^2_2, \text{ s.t. } z = c, z \geq 0
\]  

(8)

Equation (8) can be solved by the ADMM\textsuperscript{16} algorithm, and the Lagrangian function of equation (8) is

\[
\mathcal{L}(c, z, \delta, \mu) = \|y - Xc\|^2_2 + \lambda \|Xc\|^2_2 + \langle \delta, z - c \rangle + \frac{\mu}{2} \|z - c\|^2_2
\]  

(9)

where \(\delta\) is the Lagrange multiplier and \(\mu > 0\) is a penalty parameter. The optimization of equation (9) can be solved iteratively by updating \(c\) and \(z\) once at a time. The detailed updating procedures are presented as follows.

**Update c:** Fix the other variables and update \(c\) by solving the following problem:

\[
e_{\text{min}} \|y - Xc\|^2_2 + \lambda \|Xc\|^2_2 + \frac{1}{2} \|z - c + \frac{\delta}{\mu}\|^2_2
\]  

(10)

Setting the partial derivative of equation (10) with respect to \(c\) to zero, we can obtain the following closed-form solution to \(c\):

\[
c_{t+1} = \left(1 + \lambda \mu \right) X^T X + \frac{\mu}{2} I \right)^{-1} \left( X^T y + \frac{\delta}{\mu} \right)
\]  

(11)

**Update z:** To update \(z\), we fix variables other than \(z\) and solve the following problem accordingly:

\[
z_{\text{min}} \|y - (c_{t+1} - \frac{\delta}{\mu})\|^2_2, \text{ s.t. } z \geq 0
\]  

(12)

The solution to \(z\) is given by

\[z_{t+1} = \max \left(0, c_{t+1} - \frac{\delta}{\mu}\right)
\]  

(13)

where the max operator performs element-wisely.

**Update \(\delta\):** The Lagrange multiplier \(\delta\) is updated according to the following formulation:

\[
\delta_{t+1} = \delta_t + \mu (z_{t+1} - c_{t+1})
\]  

(14)

The detailed procedures of solving equation (7) are summarized in Algorithm 1.

**Classification**

For the test sample \(y \in \mathbb{R}^d\), first we obtain its coding vector \(c\) over the entire training data \(X\), then the test sample is classified into the class that yields the least residual,

**Algorithm 1 Solve Equation (7) via ADMM.**

**Input:** Test sample \(y\), training data matrix \(X\), balancing parameter \(\lambda, \text{ tol} > 0, \mu > 0\) and the maximum iteration number \(T\).

1: Initialize \(z_0 = c_0 = \delta_0 = \mathbf{0}\);
2: while not converged do
3: \quad Update \(c\) by equation (11);
4: \quad Update \(z\) by equation (13);
5: \quad Update \(\delta\) by equation (14);
6: end while

**Output:** Coding vectors \(z\) and \(c\).
Algorithm 2 Our proposed discriminative non-negative representation based classifier (DNRC) algorithm.

Input: Training data matrix $X = [X_1, X_2, \ldots, X_K] \in \mathbb{R}^{d \times n}$, test data $y \in \mathbb{R}^d$ and balancing parameter $\lambda$.
1: Normalize the columns of $X$ and $y$ to have unit $\ell_2$ norm;
2: Obtain the coding vector $c$ of $y$ on $X$ by solving the DNRC model in (7);
3: Compute the class-specific residuals $r_i = \|y - X_i c_i\|_2$;
Output: $\text{label}(y) = \arg\min_i r_i$.

i.e., identity($y$) = $\arg\min_i \|y - X_i c_i\|_2$, where $c_i$ is the coding vector that belongs to the $i$-th class. The complete process of our proposed DNRC is summarized in Algorithm 2.

Rationale of DNRC

Our proposed DNRC model (7) introduces a decorrelation regularizer into the formulation of NRC. As mentioned earlier, the decorrelation regularizer $\|X c\|_2^2$ can be rewritten as $\sum_{i=1}^K \sum_{j=1}^K (X_i c_i)^T (X_j c_j)$, which is actually the pairwise class competition term. Therefore, DNRC takes the non-negative constraint and pairwise class competition into consideration simultaneously. Compared with (6), the second term in (7) impacts all pairs of classes, which promotes the competition among them and further enhances the discrimination of representation. Since the test sample is represented as a weighted sum (i.e., a linear superposition) of the training samples of all classes, each class has a contribution to expressing the test sample. Competition means that when training samples of one class have a great contribution to representing the test sample, other classes have relatively less contribution.

Since DNRC is closely related to NRC, we make a comparison between the two approaches. On the AR database, seven non-occluded images per subject in Session 1 are used for training, and one face image in Session 2 from the 98th class is used for testing. Figures 1 and 2 show the representation results (i.e., $X_i c_i$) of all classes produced by NRC and DNRC, respectively. As can be seen in Figure 1, the representation result of the correct class is not dominant. In contrast, from Figure 2, we can observe that DNRC significantly depresses the irrelevant classes and achieves more discriminative representation, which is beneficial for classification.

Experiments

In this section, we assess the classification performance of DNRC on diverse benchmark datasets: two face databases including AR and Extended Yale B databases, Stanford 40 Actions dataset for action recognition, three fine-grained object datasets which includes the Oxford 102 Flowers dataset, the Aircraft dataset and the Cars dataset, the details of these datasets are summarized in Table 1. We compare the classification accuracy of DNRC with NSC, linear SVM, SRC, CRC, CROC, ProCRC and NRC. In addition, on the Aircraft and Cars datasets, we also compare DNRC with some state-of-the-art deep methods, such as Symbiotic, FV-FGC and B-CNN. The parameter of our method, that is, $\lambda$ in equation (7), is tuned to achieve the best performance via fivefold cross validations from the candidate set {$1 \times 10^{-3}$, $1 \times 10^{-2}$, 0.1, 1}. Recognition accuracy is
used as the evaluation metric, which is the ratio of correctly classified samples to the total number of test samples, and it gives the value in percentage.

**Experiments on the AR database**

The AR database\(^{17}\) contains more than 4000 color images of 126 subjects (70 men and 56 women), these images have variations in facial expressions, illumination conditions and occlusions, example images from this database are shown in Figure 3. Following the experimental settings by Xu et al.,\(^{14}\) in our experiments, we use a subset with only illumination and expression changes that contains 50 male subjects and 50 female subjects from the AR database. For each subject, seven images from Session 1 are used as training samples, and the other seven images from Session 2 as test samples. All the images are firstly cropped to 60×43 pixels and projected to a subspace of dimensions 54, 120 and 300 by PCA. Experimental results are summarized in Table 2, the balancing parameter \(\lambda\) of DNRC under dimensions 54, 120 and 300 is set to be 0.1, 0.1 and 1, respectively. We can observe that our proposed DNRC consistently outperforms the others under all the three reduced dimensions.

In order to demonstrate the statistical significance of our proposed DNRC compared with the other methods, we conduct a significance test, McNemar’s test (non-parametric),\(^{26,27}\) for the results shown in Table 2. The significance level, that is, \(p\)-value is set as 0.05, which means that the performance difference between two methods is statistically significant, if the estimated \(p\)-value is lower than 0.05. Table 3 lists the \(p\)-values between DNRC and the other methods. From Table 3, one can see that the performance differences between DNRC and the methods (NSC, SVM, SRC, CROC and ProCRC) are statistically significant in all cases. The performance differences between DNRC and CRC/NRC are not statistically significant; however, the recognition accuracy of DNRC is higher than that of CRC/NRC. The above experimental results validate the effectiveness of our proposed DNRC.

**Experiments on the extended yale B database**

The Extended Yale B database\(^{18}\) consists of 2414 face images from 38 individuals, each having 59–64 images. These images have illumination variations, example images from this database are shown in Figure 4. The original images are of 192×168 pixels. In our experiments, all the images are resized to 54×48 pixels. 32 images per subject are randomly chosen for training and the remaining for testing. The resized images are projected to a subspace...
of dimensions 84, 150 and 300 by PCA. Table 4 lists the classification accuracy of the competing approaches, the balancing parameter \( \lambda \) of DNRC under dimensions 84, 150 and 300 is set to be 1, 1 and 0.1, respectively. It can be seen that DNRC achieves better performance than all the comparison methods in all dimensionalities.

**Experiments on the stanford 40 actions dataset**

The Stanford 40 Actions dataset\(^1\) contains 40 different classes of human actions, for example, brushing teeth, cleaning the floor, reading book and throwing a Frisbee, example images from this dataset are shown in Figure 5. It contains 9532 images in total, 180–300 images per action. Following the training-test split settings scheme in Yao et al.,\(^1\) we randomly select 100 images per class as the training images and employ the remaining images as the testing set. Features are extracted by using the pre-trained VGG19 network,\(^2\) and the dimension of the feature for each image is 4096. Experimental results are shown in Table 5, the balancing parameter \( \lambda \) of DNRC is \( 1 \times 10^{-3} \). One can see that DNRC achieves higher accuracy than previous RBCM, that is, SRC, CRC and NRC. This indicates that by introducing the decorrelation regularizer, discriminative information of the training data can be explored. As a result, our proposed DNRC is more effective than previous RBCM.

**Experiments on the oxford 102 flowers dataset**

The Oxford 102 Flowers dataset\(^3\) includes 8189 images of 102 different flowers. Each flower class has over 40 images. These flower images are captured under diverse lighting conditions, flower poses, and image scales, example images from this dataset are shown in Figure 6. Features are extracted by employing the pre-trained VGG19 network. Table 6 summarizes the classification accuracy of comparison methods, the balancing parameter \( \lambda \) of DNRC is 0.1. It can be seen that DNRC achieves better performance when compared to the other competing methods.
Experiments on the aircraft dataset

The Aircraft dataset\(^{21}\) includes 10,000 images of 100 different aircraft model variants, 100 images for each class. These aircrafts appear at diverse appearances, scales, and design structures, making this dataset very challenging for visual recognition. Example images from this dataset are shown in Figure 7. We adopt the training-testing split protocol provided by Maji et al.\(^{21}\) to design our experiments, features are extracted via a pre-trained VGG-16 network.\(^{28}\) We also compare with the methods of Symbiotic,\(^{23}\) FV-FGC,\(^{24}\) B-CNN.\(^{25}\) Recognition accuracy of competing approaches is presented in Table 7, the balancing parameter \(\lambda\) of DNRC is \(1 \times 10^{-3}\). We can observe that our proposed DNRC achieves higher accuracy than NRC and B-CNN. This demonstrates that DNRC can outperform not only the traditional RBCM such as SRC, CRC and NRC, but also the CNN based approaches.

Experiments on the cars dataset

The Cars dataset\(^{22}\) has 16,185 images of 196 car classes. Each car class contains about 80 images at different scales and heavy clutter background, making this dataset very challenging for visual recognition. Example images from this dataset are shown in Figure 8. We use the same split scheme provided by Krause et al.,\(^{22}\) in which 8144 images are employed as the training samples and the other 8041 images are employed as the testing samples. Features are extracted via a pre-trained VGG-16 network. We also compare with Symbiotic,\(^{23}\) fisher vector for fine-grained classification (FV-FGC),\(^{24}\) bilinear CNN (B-CNN).\(^{25}\) Recognition accuracy of competing approaches on this dataset is presented in Table 8, the balancing parameter \(\lambda\) of DNRC is \(1 \times 10^{-3}\). Again, our proposed DNRC performs the best among all the competing approaches.
methods, and it makes an improvement of 0.3% and 0.2% over B-CNN and NRC, respectively.

Parameter analysis

In this subsection, we investigate how the balancing parameter $\lambda$ influences the classification performance of DNRC. Experiments are conducted on the AR database, and the dimensionality of PCA is 300. Figure 9 shows the variation of recognition accuracy with $\lambda$. It can be seen that the classification performance of our proposed DNRC is stable in quite a wide range.
Figure 7. Example images from the Aircraft dataset.

Table 7. Recognition accuracy (%) of competing approaches on the Aircraft dataset.

| Methods      | NSC$^3$ | SRC$^4$ | CRC$^8$ | CROC$^{10}$ | ProCRC$^{13}$ |
|--------------|---------|---------|---------|-------------|--------------|
| Accuracy     | 85.5    | 86.1    | 86.7    | 86.9        | 86.8         |
| Methods      | Symbiotic$^{23}$ | FV-FGC$^{24}$ | B-CNN$^{25}$ | NRC$^{14}$ | DNRC         |
| Accuracy     | 72.5    | 80.7    | 84.1    | 87.3        | 87.4         |

SRC: sparse representation based classification; NSC: nearest subspace classifier; CRC: collaborative representation based classification; CROC: collaborative representation optimized classifier; DNRC: discriminative non-negative representation based classifier; ProCRC: probabilistic collaborative representation based classifier; FV-FGC: fisher vector for fine-grained classification; B-CNN: bilinear CNN; NRC: non-negative representation based classifier.

Figure 8. Example images from the Cars dataset.

Table 8. Recognition accuracy (%) of competing approaches on the Cars dataset.

| Methods      | NSC$^3$ | SRC$^4$ | CRC$^8$ | CROC$^{10}$ | ProCRC$^{13}$ |
|--------------|---------|---------|---------|-------------|--------------|
| Accuracy     | 88.3    | 89.2    | 90.0    | 90.3        | 90.1         |
| Methods      | Symbiotic$^{23}$ | FV-FGC$^{24}$ | B-CNN$^{25}$ | NRC$^{14}$ | DNRC         |
| Accuracy     | 78.0    | 82.7    | 90.6    | 90.7        | 90.9         |

SRC: sparse representation based classification; NSC: nearest subspace classifier; CRC: collaborative representation based classification; CROC: collaborative representation optimized classifier; DNRC: discriminative non-negative representation based classifier; ProCRC: probabilistic collaborative representation based classifier; FV-FGC: fisher vector for fine-grained classification; B-CNN: bilinear CNN; NRC: non-negative representation based classifier.
Conclusions

In this paper, we designed a DNRC for pattern classification. By incorporating the decorrelation regularizer into the formulation of NRC, training samples of different classes are enforced to compete in representing the test sample. Therefore, the coefficient vector obtained by DNRC contains more discriminative information than that of NRC. Our proposed DNRC is solved elegantly via the ADMM algorithm. Experimental results on face databases, human action dataset and fine-grained datasets demonstrate that DNRC is superior to NRC and traditional RBCM, and it also outperforms some deep learning based approaches.

DNRC is an improvement of NRC, and they both belong to shallow model. In our future work, we will compare DNRC with graph convolutional network and design the deep version of DNRC.

Acknowledgements

The authors thank the editor and the anonymous reviewers for their constructive and valuable comments and suggestions, and thank Dr. Jun Xu for providing the source code of NRC at https://github.com/csjunxu/NSCR.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Key Project of Jiangsu Vocational College of Information Technology (grant no. JSITKY201905), project of high-level specialty group construction of higher vocational education of Jiangsu province (grant no. [2021] 1), general project fund for natural science research of colleges and universities in Jiangsu province (grant no. 18KJD510011), project of high-level backbone specialty construction of Jiangsu province (grant no. [2017] 17), and in part by the National Natural Science Foundation of China (grant no. 61672263).

ORCID iDs

Kai-Jun Hu https://orcid.org/0000-0001-6795-2928
He-Feng Yin https://orcid.org/0000-0001-5831-1475

References

1. Wright J, Yang AY, Ganesh A, et al. Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 2009; 31: 210–227.
2. Naseem I, Togneri R and Bennamoun M. Linear regression for face recognition. IEEE Trans Pattern Anal Mach Intell 2010; 32: 2106–2112.
3. Lee K, Ho J and Kriegman D. Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans Pattern Anal Mach Intell 2005; 27: 684–698.
4. Yu K, Zhang T and Gong Y. Nonlinear learning using local coordinate coding. In: NeurIPS, Vancouver, Canada, 6–12 December 2009, pp.2223–2231.
5. Lu C, Min H, Gui J, et al. Face recognition via weighted sparse representation. J Visual Commun Image Represent 2013; 24: 111–116.
6. Wang J, Yang J, Yu K, et al. Locality-constrained linear coding for image classification. In: Conference on computer vision and pattern recognition (CVPR), San Francisco, California, 13–18 June 2010, pp.3360–3367.
7. Xu Y, Zhang D, Yang J, et al. A two-phase test sample sparse representation method for use with face recognition. IEEE Trans Circuits Syst Video Technol (TCSVT) 2011; 21: 1255–1262.
8. Zhang L, Yang M and Feng X. Sparse representation or collaborative representation: Which helps face recognition?. In: International conference on computer vision (ICCV), Barcelona, Spain, 6–13 November 2011, pp.471–478.
9. Timofte R and Van Gool L. Weighted collaborative representation and classification of images. In: IJCV, Tsukuba, Japan, 11–15 November 2012, pp.1606–1610.
10. Chi Y and Porikli F. Classification and boosting with multiple collaborative representations. IEEE Trans Pattern Anal Mach Intell 2013; 36: 1519–1531.
11. Lai J and Jiang X. Classwise sparse and collaborative patch representation for face recognition. IEEE Trans Image Process 2016; 25: 3261–3272.
12. Xu Y, Zhong Z, Yang J, et al. A new discriminative sparse representation method for robust face recognition via L2 regularization. IEEE Trans Neural Netw Learn Syst 2016; 28: 2233–2242.
13. Cai S, Zhang L, Zuo W, et al. A probabilistic collaborative representation based approach for pattern classification. In: Conference on computer vision and pattern recognition (CVPR), Las Vegas, USA, 27–30 June 2016, pp.2950–2959.
14. Xu J, An W, Zhang L, et al. Sparse, collaborative, or non-negative representation: Which helps pattern classification?. Pattern Recognit 2019; 88: 679–688.
15. Lee D and Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature 1999; 401: 788–791.
16. Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Mach Learn 2011; 3: 1–122.
17. Martinez A and Benavente R. The AR face database, CVC Tech. Rep. no. 24, June, 1998.
18. Georgiades A, Belhumeur P and Kriegman D. From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans Pattern Anal Mach Intell 2001; 23: 643–660.
19. Yao B, Jiang X, Khosla A, et al. Human action recognition by learning bases of action attributes and parts. In: International conference on computer vision (ICCV), Barcelona, Spain, 6–13 November 2011, pp.1331–1338.
20. Nilsback M and Zisserman A. Automated flower classification over a large number of classes. In: Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP), Bhubaneswar, India, 16–19 December 2008, pp.722–729.
21. Maji S, Rahtu E, Kannala J, et al. Fine-grained visual classification of aircraft. arXiv preprint arXiv:1306.5151, 2013.
22. Krause J, Stark M, Deng J, et al. 3D object representations for fine-grained categorization. In: International conference on computer vision workshop (ICCVW), Sydney, Australia, 1–8 December 2013, pp.554–561.
23. Chai Y, Lempitsky V and Zisserman A. Symbiotic segmentation and part localization for fine-grained categorization. In: International conference on computer vision (ICCV), Sydney, Australia, 1–8 December 2013, pp.321–328.
24. Gosselin P, Murray N, Jégou H, et al. Revisiting the fisher vector for fine-grained classification. Pattern Recogn Lett 2014; 49: 92–98.
25. Lin T, Roy Chowdhury A and Maji S. Bilinear CNN models for fine-grained visual recognition. In: International conference on computer vision (ICCV), Santiago, Chile, 7–13 December 2015, pp.1449–1457.
26. Dietterich T. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Comput 1998; 10: 1895–1923.
27. Wen J, Fang X, Xu Y, et al. Low-rank representation with adaptive graph regularization. Neural Netw 2018; 108: 83–96.
28. Simonyan K and Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.