Sparse representation-based dictionary learning with CNN for image classification

Shuai Yu, Tao Zhang, and Jie Yang*

Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China.
{yushuai9471,zhb827,jieyang}@sjtu.edu.cn

Abstract. In this paper, we propose a novel framework for image recognition based on a extended sparse model. First, inspired by the impressive results of CNN over different tasks in computer vision, we use the CNN models pre-trained on large datasets to extract features. Then we propose a extended sparse model which learns a dictionary for classification by incorporating the representation-constrained term and the coefficients incoherence term. With this learned dictionary, not only the representation residual but also the representation coefficients will be. Experiments on Caltech101 and PASCAL VOC 2012 datasets show the effectiveness of both our sparse model and our classification scheme on image classification.

Keywords: image classification,CNN,sparse model,supervised dictionary learning

1 Introduction

As one of the most active research areas in computer vision, image classification has been widely studied. Conventional approaches for image classification use carefully designed hand-crafted features, e.g., SIFT and HOG. Recently, in contrast to the hand-crafted features, features with deep network architectures, represented by deep convolutional neural networks (CNN) [13] have got impressive results in image classification, i.e., ILSVRC on ImageNet Dataset [6]. Specifically, deep learning attempts to model the visual data of high level abstract structural composites using multivariate nonlinear transformations. And several works [7, 18, 21] show that the pre-trained CNN models on colossal datasets with data diversity can be transferred to extract discriminative features for other tasks.

For sparse representation-based classification (SRC) Wright et al. [22] proposed a general classification scheme based on sparse representation and applied it on robust face recognition (FR). Since the SRC scheme achieves competitive performance in face recognition (FR), it triggers the researchers' interest in

* Corresponding author: Jie Yang, jieyang@sjtu.edu.cn

sparsity-based pattern classification [3, 12]. How to learn a discriminative dictionary for both sparse data representation and classification is still an open problem.

According to predefined relationship between dictionary atoms and class labels, we can divide current supervised dictionary learning into three categories: shared dictionary learning, class-specific dictionary learning and hybrid dictionary learning. In the shared dictionary learning, a dictionary shared by all classes is learned but also the discriminative power of the representation coefficients is mined. It is popular to learn a shared dictionary and simultaneously training a classifier using the representation coefficients. In [16], Marial et al. proposed a scheme which learned discriminative dictionaries while training a linear classifier over coding coefficients. Inspired by KSVD [1], Zhang and Li [25] proposed discriminative KSVD (DKSVD) learning algorithm on FR. Following the work in [25], Jiang et al. [9] proposed to enhance the discriminative power via adding a label consistent term. Recently, Mairal et al. [14] proposed to minimize different risk functions over the coding coefficients for different tasks, called a task-driven DL. In generally, in this scheme, a shared dictionary and a classifier over the representation coefficients are together learned. However, there is no relationship between the dictionary atoms and the class labels, and thus no class-specific representation residuals are introduced to perform classification task.

In the class-specific dictionary learning, a dictionary whose atoms are predefined to correspond to subject class labels is learned and thus the class-specific reconstruction error could be used to perform classification . Via adding a discriminative reconstruction penalty term in the KSVD model [1], Mairal et al. [15] proposed to a dictionary learning algorithm for texture segmentation and scene analysis. Yang et al. [23] proposed to learn a structural dictionary and impose the Fisher discrimination criterion on both the sparse coding coefficients to enhance class discrimination power. In [4], via adding non-negative penalty on both dictionary atoms and representation coefficients, Castrodad and Sapiro proposed to learn a set of action-specific dictionaries. In [17], Ramirez et al. introduced an incoherence promoting term to the DL model for ensuring the dictionaries representing different classes to be as independent as possible. H. Wang et al. [20] learned a dictionary with similarity constrained term and the dictionary incoherence term and applied it to human action recognition. Based on each atom in the learned dictionary is fixed to a single class label, the representation residual associated with each class-specific dictionary could be used to perform classification.

Very recently, the hybrid dictionary models which combines shared dictionary atoms and class-specific dictionary atoms have been proposed. Using a Fisher-like penalty term on the coding coefficients, Zhou et al. [26] learned a hybrid dictionary, while introducing a coherence penalty term on different sub-dictionaries, Kong et al. [11] learned a hybrid dictionary. Although the shared dictionary atoms could encourage learned hybrid dictionary compact to some extent, how to balance the shared part and class-specific part in the hybrid dictionary is not a trivial task.

We propose a extended sparse framework to learn a class-specific dictionary with input features extracted from CNN, i.e., the dictionary atoms correspond to the class labels. In this proposed framework, two terms named the representation-constrained term and the coefficients incoherence term, are introduced to ensure the learned dictionary with the powerful discriminative ability. The representation-constrained term is utilized to enforce that class-specific sub-dictionary has good reconstruction capability for the training samples from the same class. The coefficients incoherence term is utilized to enforce that class-specific sub-dictionaries have poor reconstruction capability for training samples from different classes. Therefore, both the representation residual and the representation coefficients of a query sample will be discriminative, and a corresponding classification scheme is proposed to exploit such information. Then we test our classification scheme on the widely used datasets (Caltech-101 [8] and VOC 2012 [8]).

The remainder of this paper is organized as follows. In Section 2, we introduce the proposed extended sparse framework and a supervised class-specific dictionary learning method for classification. In Section 3, we demonstrate experimental results. In Section 4, we make conclusions about our method.

2 Methodology

Since the previous works show that pre-trained CNN models on colossal datasets with data diversity, can be transferred to extract CNN features for other image datasets [18]. We use the pre-trained VGG-Net [19] model on ImageNet to extract features for sparse representation-based dictionary learning. As for the selection of the features from CNN net, the shallow layers have features with too much dimensions and they are too sparse to get effective results for classification.However, some of the deepest layer are totally corresponding to the original data set for CNN training. So we choose some deep but not the deepest layers of CNN net to get features for classification.

2.1 Sparse representation and dictionary learning

Very recently, Wright et al. [22] proposed the sparse representation based classification (SRC) method for robust face recognition (FR). Obviously, SRC is imaginable that a test sample can be represented by a weighted linear combination of those training samples belonging to the same class. Impressive results have been reported in [22].

The model We consider that CNN features of the samples in different image classes have different are discriminative. We adopt the class-specific dictionary. In the class-specific dictionary learning (DL), each dictionary atom in the learned dictionary $D = [D_1, D_2, \ldots, D_K]$ have class label correspondence to the subject classes, where D_i is the sub-dictionary corresponding to class i. By representing a

test sample over the learned dictionary D, the representation residual associated with each class can be naturally employed to classify it, as in the SRC method.

Given training samples $a_{i,j}, i = 1, \dots K$, $j = 1, \dots, n_i$ denotes a feature got from CNN in class i. Which K is the sum of classes, and n_i is the number of samples in class i. We form $A_i = [a_{i,1}, a_{i,2}, \dots, a_{i,n_i}]$. The dictionary D can be learned by the following extended sparse model:

$$< D, Z >= \arg\min_{D,Z} \sum_{i=1}^{K} \{\|A_i - DZ_i\|_F^2 + \quad \lambda_1 \|Z_i\|_1 + \lambda_2 \|A_i - D_i Z_i^i\|_F^2 + \kappa \sum_{j \neq i} \|\tilde{Z}_j^T Z_i\|_F^2\} \quad (1)$$

$$s.t. \|d_n\|_2 = 1, \forall n$$

where Z_i is the sub-matrix containing the coding coefficients of A_i over D. Z_i can be written as $Z_i = [Z_i^1; \dots; Z_i^j; \dots; Z_i^K]$, where Z_i^j represents the coefficients of A_i over D_j; and $\tilde{Z}_j$ is $\tilde{Z}_j = [\tilde{Z}_{j,1}, \tilde{Z}_{j,2}, \dots, \tilde{Z}_{j,n}]$, where $\tilde{Z}_{j,i} = Z_{j,i}/\|Z_{j,i}\|$ is normalized coefficients of the ith sample in A_i over D.

Different from the conventional sparse model SRC in[22], the representation-constrained term ($\|A_i - D_i Z_i^i\|_F^2$) and coefficients incoherence term ($\sum_{j \neq i} \|\tilde{Z}_j^T Z_i\|_F^2$) are introduced in Eq.(1).

Representation-constrained term For A_i, it should be well represented by the dictionary D, hence there is $A_i \approx DZ_i$. Since A_i is associated with the class i, it is expected that A_i could be represented further well by D_i. This implies that Z_i should have some significant coefficients Z_i^i such that $\|A_i - D_i Z_i^i\|_F^2$ is small.

Coefficients incoherence term In the SRC scheme proposed by Wright et al. [22], given a test sample, the accurate classification can be conducted based on that the largest coefficients are associated with the training samples that belong to the same class as the test sample. It implies that the reconstruction error is minimized when test sample are sparsely represented by its own training samples. Likewise, in the class-specific dictionary learning, it is expected that the largest coefficients of A_i are associated with the sub-dictionary D_i. In Eq. (1), minimizing the coefficients incoherence term $\sum_{j \neq i} \|\tilde{Z}_j^T Z_i\|_F^2$ encourages that for the A_i and A_j, the largest coefficients are associated with the corresponding different sub-dictionary D_i and D_j as illustrated in Figure 1. This means that similar samples over dictionary D have similar coefficients and samples belonging to different classes over dictionary D have absolutely different coefficients. Therefore, the value of the object function Eq.(1) is minimized when samples are sparsely represented by dictionary atoms in their own sub-dictionaries.

Overall, minimizing the representation-constrained term $\|A_i - D_i Z_i^i\|_F^2$ guarantees that class-specific sub-dictionary has good representation power to the

samples from the corresponding class and minimizing the coefficients incoherence term $\sum_{j\neq i}\|\tilde{Z}_j^T Z_i\|_F^2$ encourages samples from different classes are reconstructed by different class-specific sub-dictionaries. By incorporating the representation-constrained term and coefficients incoherence term, our proposed sparse representation algorithm is more effective for classification.

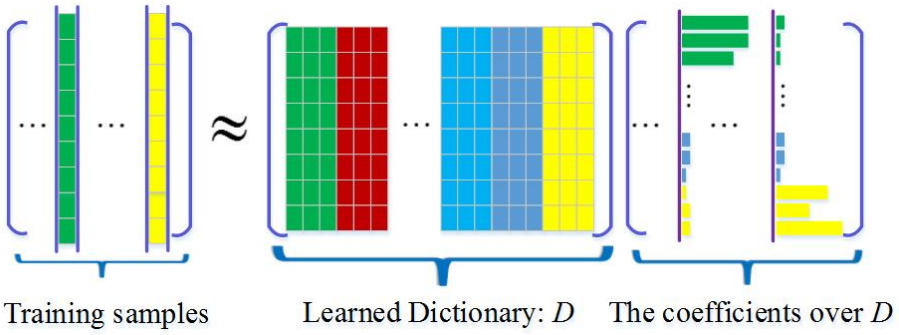

Fig. 1. Sparse representation of training samples using the learned dictionary D. The green and yellow training samples are belong to class i and j; the green and yellow atoms in D have class labels correspondences to class i and j. The sparse coefficients of green and yellow training samples recovered are plotted in the coefficients matrix with the green and yellow largest values associated with the green and yellow atoms in D which have class labels correspondences to class i and j.

The optimization Although the objective function in Eq.(1) is not jointly convex to (D, Z). Like other authors [20, 24] have done when trying to solve similar optimization problems, here we divide the objective function in Eq.(1) into two sub-problems by optimizing D and Z alternatively: updating coefficient matrix Z while fixing the dictionary D, and updating dictionary D while fixing the coefficient matrix Z.

Update of Z When we fix the dictionary D, the objective function in Eq.(1) is reduced to a sparse representation problem to compute $Z = [Z_1, Z_2, \ldots, Z_K]$. We can compute Z_i class by class by fixing $Z_j, j \neq i$. The objective function in Eq.(1) is further reduced to:

$$\min_{Z_i}\{\|A_i - DZ_i\|_F^2 + \lambda_1\|Z_i\|_1 + \lambda_2\|A_i - D_i Z_i^i\|_F^2 + \kappa\sum_{j\neq i}\|\tilde{Z}_j^T Z_i\|_F^2\} \qquad (2)$$

It can be proved that $\varphi_i(Z_i) = \|A_i - DZ_i\|_F^2 + \lambda_2\|A_i - D_i Z_i^i\|_F^2 + \kappa\sum_{j\neq i}\|\tilde{Z}_j^T Z_i\|_F^2$ is convex with Lipschitz continuous gradient. Hence, in this work we adopt a new fast iterative shrinkage-thresholding algorithm (FISTA) [2] to solve Eq.(2), as described in Algorithm 1.

Algorithm 1 Learning sparse code Z_i.

Input:
A training subset A_i from class i; the dictionary D; the parameters $\rho, \tau, > 0$.
Initialize:
$\hat{Z}_i^{(1)} \leftarrow 0$ and $t \leftarrow 0$;
while convergence or the maximal iteration step is not reached do **do**
$t \leftarrow t+1$; $u^{t-1} \leftarrow \hat{Z}_i^{(t-1)} - 1/2\rho\nabla\varphi_i(\hat{Z}_i^{(t-1)})$,
where $\nabla\varphi_i(\hat{Z}_i^{(t-1)})$is the derivative of $\varphi_i(\hat{Z}_i^{(t-1)})$ w.r.t. $\hat{Z}_i^{(t-1)}$;
$\hat{Z}_i^t \leftarrow soft(u^{(t-1)}, \tau/\rho)$, where $soft(u^{(t-1)}, \tau/\rho)$ is defined by Eq.(4) [10]:
end while
Output:
$\hat{Z}_i = \hat{Z}_i^{(t)}$.

Update of *D* In this subsection we describe how to update $D = [D_1, D_2, \ldots, D_K]$, while fixing the coefficient matrix Z. When updating D_i, all $D_j, j \neq i$, are fixed and $D_i = [d_1, d_2, \ldots, d_{p_i}]$is updated class by class. We can reduce objective function in Eq.(1) as:

$$\min_{D_i}\{\|\bar{A}_i - DZ_i\|_F^2 + \lambda_2\|A_i - D_i Z_i^i\|_F^2\} \quad s.t.\|d_l\|_2 = 1, l = 1, \ldots, p_i \tag{3}$$

Algorithm 2 Learning dictionary D_i.

Input:
A training subset A_i from class i; the coefficients Z_i; the dictionary D_i^o.
Let $Z_i = [z_1; z_2; \ldots; z_{p_i}]$ and $D_i^o = [d_1; d_2; \ldots; d_{p_i}]$, where z_j, $j = 1, 2, \ldots, p_i$, is the row vector of Z_i and d_j is the jth column vector of D_i^o;
$\hat{Z}_i^{(1)} \leftarrow 0$ and $t \leftarrow 0$;
for $j = 1$ to p_i **do**
Fix all d_l, lj update d_j. Let $X = \Lambda_i - \sum_{l \neq j} d_l z_l$. The minimization of Eq.(3) becomes:
$\min_{d_j} \|X - d_j z_j\|_F^2 \text{s.t.} \|d_j\|_2 = 1$
By solving this objective function,we could get the solution
$d_j = X z_j^T / \|X z_j^T\|_2$.
end for
Output:
The updated version of D_i^o:D_i.

$$[soft(u^{(t-1)}, \tau/\rho)]_j = \begin{cases} 0 & |u_j| \leq \tau/\rho \\ u_j - sign(u_j)\tau/\rho & \text{otherwise} \end{cases} \tag{4}$$

where $\bar{A} = A - \sum_{j=1, j\neq i}^{K}$; Z^irepresent the coefficient matrix of A over D_i. Eq.(3) can be efficiently solved by updating each dictionary atom one by one via the algorithm like [24], as presented in Algorithm 2.

Complete dictionary D learning algorithm The complete algorithm is summarized in Algorithm 1. The algorithm converges since the cost function in Eq.(1) is lower bounded and can only decrease in the two alternative minimization stages (i.e., updating Z and updating D).

Algorithm 3 The complete algorithm of dictionary D learning.

Initialize D.
We initialize the atoms of Di as the eigenvectors of A.
Update coefficients Z.
Fix D and solve Z_i, $i = 1, 2, \ldots, K$, one by one by solving Eq.(2) with Algorithm 1.
Update coefficients D.
Fix Z and update each D_i, $i = 1, 2, \ldots, K$, by solving Eq.(3) with Algorithm 2.
return
Update D and Z when the objective function values between adjacent iterations are not close enough or the maximum number of iterations is not reached.
Output:
Z and D

The classification scheme Once the dictionary D have been trained, it could be adopted to represent a query sample y and do a classification task. According to different schemes for learning the dictionary D, different information can be utilized to perform the classification task.

In our proposed sparse representation model, not only the desired dictionary D is learned from the training dataset A, but also the normalized representation matrix $\tilde{Z}_i$ of each class A_i is computed. Considering both the representation residual and the representation coefficients are discriminative, we can make use of both of them to achieve more accurate classification results. Hence, we propose the following representation model:

$$\hat{\alpha} = \arg\min_{\alpha}\{\|y - D\alpha\|_2^2 + \gamma\|\alpha\|_1\} \tag{5}$$

where,γ constant.

Denote by $\hat{\alpha} = [\hat{\alpha}^1, \hat{\alpha}^2, \ldots, \hat{\alpha}^K]$, where $\hat{\alpha}^i$ is the coefficient sub-vector associated with sub-dictionary D_i. In the training stage, we have enforced the class-specific representation residual to be discriminative. Therefore, if y is from class i, the residual $\|y - D_i\hat{\alpha}\_i\|_2^2$ should be small while $\|y - D_i\hat{\alpha}^j\|_2^2, j \neq i$, should be big. In addition, the representation sub-vector $\hat{\alpha}^i$ should be far different from the representation vector of other classes. By considering the discrimination capability of both representation residual and representation vector, we could define the following metric for classification:

$$e_i = \|y - D_i\hat{\alpha}_2\| + w\sum_{j\neq i}\|\tilde{Z}_j^T\hat{\alpha}\|/n_j \tag{6}$$

where w is preset weight to balance the contribution of the two terms for classification. The classification rule is simply set as $identity(y) = \arg\min_i\{e_i\}$.

3 Experimental Results

3.1 Datasets and experiments settings

In the VGG-Net [19] model , we choose the 18th layer of 4096 dimensions as the feature for classification, as we describe in the beginning of Section 2.

In our proposed sparse representation model, there are two stages: dictionary learning (DL) stage and classification stage. In DL stage we set $\lambda_1 = 0.005$, $\lambda_2 = 1$,$\kappa = 0.01$; in classification stage we set $\gamma = 1$,$w = 0.05$. In the proposed model, the number of atoms in D_i, denoted by p_i, is important and it is set as the number of training samples by default. All of the experiments are executed on a workstation with Intel 2.8GHz CPU and 16GB RAM.

3.2 Experiments on Caltech-101

To verify the effectiveness of our proposed sparse model for image classification, we make comparisons with other classifiers. We use the same features exacted from CNN as the input of SRC [22], SVM and our sparse model incorporating representation-constrained and coefficients incoherence terms (SDRCI).

We evaluate our algorithm on Caltech-101 dataset with cross-validation: 5-30 random images are used for training, the remaining for testing; for each size of training images, we process 10 times with our method and the results are averaged. The SVM we use to compare with our method is LIBSVM [5] fine-tuned on each training data. The result is shows in Table.1.

The accuracy of SVM is higher than that of SRC which only uses the original training samples as dictionary, and our SDRCI achieves the highest accuracy. It proves that the proposed sparse model with supervised dictionary learning method is discriminative for image classification. It proves that by incorporating the representation-constrained term and coefficients incoherence term, our proposed sparse representation model is more effective for classification.

Table 1. The SDRCI performance comparison on Caltech-101(Accuracy)

Training images	5	10	15	20	25	30
SRC	61.65	68.63	72.28	76.35	78.72	81.60
SVM	61.35	69.28	73.65	76.59	80.62	83.01
SDRCI	63.63	71.37	75.28	80.39	82.39	84.88

3.3 Experiments on VOC 2012

Further verify the effectiveness of our method,which combined the CNN features and sparse representation-based dictionary learning, we make experiments on VOC 2012 for classification task and compare with the results of state-of-art

methods.And we do not use the ground-truth bounding box information of the annotation of the dataset.

Table.5 shows the results shows that our method get convictive result comparing with other work based on CNN .

Table 2. Our method performance comparison on VOC 2012(AP)

Category	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow
Category	table	dog	horse	motor	person	plant	sheep	sofa	train	tv	mAp

4 Conclusion

In this paper, we propose a extended sparse model to learn a discriminative dictionary for classification. We adopt the pre-trained CNN model on large datasets to exact input features. In the proposed sparse model, the representation-constrained term and the coefficients incoherence term are introduced to ensure the learned dictionary to obtain powerful discriminative ability. With this learned dictionary, both the representation residual and the representation coefficients are discriminative. Finally, we present a corresponding classification scheme by exploiting such information. The experiments show that both our proposed sparse model incorporating supervised dictionary learning and our entire method are effective for classification.

Reference

1. Michal Aharon, Michael Elad, and Alfred Bruckstein. ¡ img src=”/images/tex/484.gif” alt=”\ rm k”¿-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, 2006.
2. Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.
3. Matteo Bregonzio, Shaogang Gong, and Tao Xiang. Recognising action as clouds of space-time interest points. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1948–1955. IEEE, 2009.
4. Alexey Castrodad and Guillermo Sapiro. Sparse modeling of human actions from motion imagery. *International journal of computer vision*, 100(1):1–15, 2012.
5. Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.
6. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.

7. Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. *arXiv preprint arXiv:1310.1531*, 2013.
8. M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.
9. Zhuolin Jiang, Zhe Lin, and Larry S Davis. Label consistent k-svd: Learning a discriminative dictionary for recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(11):2651–2664, 2013.
10. Alexander Klaser, Marcin Marszałek, and Cordelia Schmid. A spatio-temporal descriptor based on 3d-gradients. In *BMVC 2008-19th British Machine Vision Conference*, pages 275–1. British Machine Vision Association, 2008.
11. Shu Kong and Donghui Wang. A dictionary learning approach for classification: separating the particularity and the commonality. In *Computer Vision–ECCV 2012*, pages 186–199. Springer, 2012.
12. Adriana Kovashka and Kristen Grauman. Learning a hierarchy of discriminative space-time neighborhood features for human action recognition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2046–2053. IEEE, 2010.
13. B Boser Le Cun, John S Denker, D Henderson, Richard E Howard, W Hubbard, and Lawrence D Jackel. Handwritten digit recognition with a back-propagation network. In *Advances in neural information processing systems*. Citeseer, 1990.
14. Julien Mairal, Francis Bach, and Jean Ponce. Task-driven dictionary learning. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(4):791–804, 2012.
15. Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman. Discriminative learned dictionaries for local image analysis. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
16. Julien Mairal, Jean Ponce, Guillermo Sapiro, Andrew Zisserman, and Francis R Bach. Supervised dictionary learning. In *Advances in neural information processing systems*, pages 1033–1040, 2009.
17. Ignacio Ramirez, Pablo Sprechmann, and Guillermo Sapiro. Classification and clustering via dictionary learning with structured incoherence and shared features. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3501–3508. IEEE, 2010.
18. Ali Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 806–813, 2014.
19. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.
20. Haoran Wang, Chunfeng Yuan, Weiming Hu, and Changyin Sun. Supervised class-specific dictionary learning for sparse modeling in action recognition. *Pattern Recognition*, 45(11):3902–3911, 2012.
21. Yunchao Wei, Wei Xia, Junshi Huang, Bingbing Ni, Jian Dong, Yao Zhao, and Shuicheng Yan. Cnn: Single-label to multi-label. *arXiv preprint arXiv:1406.5726*, 2014.

22. John Wright, Allen Y Yang, Arvind Ganesh, Shankar S Sastry, and Yi Ma. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210–227, 2009.
23. Meng Yang, Lei Zhang, Xiangchu Feng, and David Zhang. Fisher discrimination dictionary learning for sparse representation. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 543–550. IEEE, 2011.
24. Meng Yang, Lei Zhang, Xiangchu Feng, and David Zhang. Sparse representation based fisher discrimination dictionary learning for image classification. *International Journal of Computer Vision*, 109(3):209–232, 2014.
25. Qiang Zhang and Baoxin Li. Discriminative k-svd for dictionary learning in face recognition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2691–2698. IEEE, 2010.
26. Ning Zhou, Yi Shen, Jinye Peng, and Jianping Fan. Learning inter-related visual dictionary for object recognition. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3490–3497. IEEE, 2012.