Class-Aware Analysis Dictionary Learning for Pattern Classification

Jiujun Wang, Yanqing Guo, *Member, IEEE*, Jun Guo, Xiangyang Luo, and Xiangwei Kong, *Member, IEEE*

***Abstract*—Dictionary learning (DL) plays an important role in pattern classification. However, learning a discriminative dictionary has not been well addressed in analysis dictionary learning (ADL). This letter proposes a Class-aware Analysis Dictionary Learning (CADL) model to improve the classification performance of conventional ADL. The objective function of CADL mainly includes two parts to promote the discriminability. The first part aims to learn a discriminative analysis subdictionary for each class instead of a global dictionary for all classes. The learned analysis dictionary is class-aware, generating a block-diagonal coding coefficient matrix. The second part aims to enhance the discrimination of coding coefficients by integrating a max-margin regularization term into our proposed framework. This term ensures the coefficients of different classes to be separated by a max-margin, which boosts the confidence of classification. A theoretical analysis is also given to support the max-margin regularization term from the perspective of preserving the pairwise relations of samples in coding space. We employ an alternating minimization algorithm to iteratively find the convergent solution. By evaluating our method on four pattern classification datasets, we demonstrate the superiority of our CADL method to the state-of-the-art DL methods.**

***Index Terms*—Analysis dictionary learning (ADL), max-margin.**

I. Introduction

THE synthesis dictionary learning (SDL) methods have been widely used in pattern classification [1]–[4] in the past decades. However, the inefficiency of SDL limits its development. On the one hand, the use of sparsity constraint like l_0-norm or l_1-norm makes the training and testing stage time-consuming. On the other hand, the matrix factorization to obtain the coefficients in testing stage is computationally complex. Therefore, ADL [5], [6], as the dual model of SDL, begins to attract great interests in recent years for its high efficiency. The analysis dictionary usually has the closed-form solution, which contributes to reducing the computational burden.

The analysis K-SVD [7] frame was proposed to learn the analysis dictionary from a training set. It is the dual viewpoint of K-SVD algorithm adopted by synthesis models. However, the conventional ADL model cannot have a wide range of applications because the learned dictionary is merely used for data transformation. To circumvent this problem, Ravishankar and Bresler [8] presented well-conditioned square transforms for image representation and denoising. Shekhar *et al.* [9] put a full column or row rank constraint on the analysis dictionary to ensure a well-conditioned solution for classification. After that, Rubinstein and Elad [10] incorporated the thresholding operator into the synthesis-analysis dictionary learning to keep the coding coefficients sparse for small-kernel image deblurring. In addition, Gu *et al.* [11] integrated structured analysis dictionaries and synthesis dictionaries together for better representation and discrimination.

The aforementioned ADL algorithms mainly pay attention to the higher efficiency, but they do not exploit the discrimination of analysis dictionary in classification problem and give little attention to its discrimination promotion. Guo *et al.* [12] focused on this problem and proposed DADL algorithm to improve the classification performance of ADL. They imposed a code consistent term and a local topology preserving term on the basic analysis model, leading to a discriminative global dictionary. However, few researchers learn class-aware analysis dictionaries to enhance discrimination for classification. Hence, there is still much scope to further develop the classification performance of ADL by learning subdictionaries.

In this letter, we propose a novel Class-Aware Analysis Dictionary Learning (CADL) model by fully exploiting the class information. CADL improves the classification performance of analysis dictionary from two main aspects. One aspect is to learn the discriminative class-aware analysis subdictionary. The other aspect is to employ a max-margin regularizer to enhance the discrimination of coding coefficients.

Our work has three major contributions:

1) We learn a discriminative analysis subdictionary for each class. Each subdictionary is constrained to transform the samples from the same class into coding coefficients, meanwhile, the samples from different classes into a nearly null space. This constraint generates a block-diagonal coding coefficient matrix, which provides an efficient alternative of the time-consuming sparsity constraint.
2) We impose a max-margin regularization term on coding coefficients. This term ensures the coefficients of different

Manuscript received March 15, 2017; revised July 25, 2017; accepted July 27, 2017. Date of publication August 1, 2017; date of current version October 27, 2017. This work was supported in part by the National Natural Science Foundation of China under Grants 61402079 and 61379151, in part by the Foundation for Innovative Research Groups of the NSFC under Grant 71421001, and in part by the Open Project Program of the National Laboratory of Pattern Recognition (NLPR, No. 201600022). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Sebastien Marcel. *(Corresponding author: Yanqing Guo.)*

J. Wang, Y. Guo, and X. Kong are with the School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China (e-mail: jiujunwang@mail.dlut.edu.cn; guoyq@dlut.edu.cn; kongxw@dlut.edu.cn).

J. Guo is with the Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, Shenzhen 518055, China (e-mail: eeguojun@outlook.com).

X. Luo is with the State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou Science and Technology Institute, Zhengzhou 450001, China (e-mail: xiangyangluo@126.com).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2017.2734860

classes to be separated by a max-margin, which boosts the confidence of classification. In addition, we offer a theoretical analysis to support the max-margin term from the perspective of preserving the pairwise relations of samples in coding space.

3) We employ an alternating minimization algorithm to iteratively find the convergent solution. Experimental results on four different classification tasks show that our class-aware method can improve the classification performance of ADL and outperform the state-of-the-art DL methods, such as DADL and DPL.

II. Brief Introduction of DL

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbb{R}^{d\times n}$ represent n original samples. $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n] \in \mathbb{R}^{m\times n}$ denotes the coding coefficient matrix of $\mathbf{X}$ over the learned dictionary. m is the length of each coding coefficient. In SDL, a synthesis dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_m] \in \mathbb{R}^{d\times m}$ (usually, $d < m$) and coding coefficient matrix $\mathbf{A}$ are often solved by minimizing the reconstruction error

$$\min_{\mathbf{D},\mathbf{A}} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 \,, \text{ s.t. } \mathbf{D} \in \mathcal{D},\ \|\mathbf{a}_i\|_0 \le T,\ i = 1, 2, \cdots, n$$

where $\mathcal{D}$ is a set to constrain $\mathbf{D}$ to obtain a well-regularized solution. $\|\cdot\|_0$ stands for the number of nonzero entries and T is a positive integer constraining the sparsity level.

ADL, as the dual model of SDL, offers us a pretty intuitive illustration for encoding like feature transformation and leads to a higher testing efficiency. The analysis dictionary $\mathbf{\Omega} \in \Re^{m\times d}$ (usually, $m > d$) can be learned by solving

$$\min_{\mathbf{\Omega},\mathbf{A}} \|\mathbf{A} - \mathbf{\Omega}\mathbf{X}\|_F^2 \,, \text{ s.t. } \mathbf{\Omega} \in \Gamma,\ \|\mathbf{a}_i\|_0 \le T,\ i = 1, 2, \cdots, n.$$

$\|\mathbf{A} - \mathbf{\Omega}\mathbf{X}\|_F^2$ can be regarded as the error term in transform domain. The set Γ often constrains the analysis dictionary with unity rowwise norm and relatively small Frobenius norm to make sure a solvable and well-regularized solution, which is indicated by [9]. Notably, the ADL model differs from the general transform model [8] where no limits are imposed on the transfer matrix.

III. Proposed CADL Method

In this section, we present the proposed CADL method, which integrates the discriminative class-aware analysis dictionary term and a max-margin regularization term together to improve the discriminability of the conventional ADL.

A. Class-Aware Analysis Dictionary Term

The conventional ADL is short of discrimination and cannot be well applied to classification tasks. To enhance its discrimination, we make full use of the class priors of data to learn the class-aware analysis subdictionary.

Let $\mathbf{X} = [\mathbf{X}_1, \cdots, \mathbf{X}_k, \cdots, \mathbf{X}_K]$ denote a training dataset from K classes, where $\mathbf{X}_k \in \mathbb{R}^{d\times n_k}$ represents that there are n_k samples in class k. We propose to learn the class-aware analysis dictionary $\mathbf{\Omega} = [\mathbf{\Omega}_1; \cdots; \mathbf{\Omega}_k; \cdots; \mathbf{\Omega}_K]$ by solving the following discrimination term:

$$\min_{\mathbf{\Omega},\mathbf{A}} \sum_{k=1}^{K} \|\mathbf{A}_k - \mathbf{\Omega}_k\mathbf{X}_k\|_F^2 + \lambda_1\|\mathbf{\Omega}_k\mathbf{X}_{\hat{k}}\|_F^2, \text{ s.t. } \mathbf{\Omega}_k \in \Gamma \quad \forall k \tag{1}$$

where $\mathbf{A}_k \in \mathbb{R}^{m_k\times n_k}$ is the coding coefficient matrix of $\mathbf{X}_k$ over the analysis subdictionary $\mathbf{\Omega}_k \in \mathbb{R}^{m_k\times d}$ associated with class k. λ_1 is a positive scalar constant parameter. $\mathbf{X}_{\hat{k}}$ is the complementary set of $\mathbf{X}_k$ in the whole set $\mathbf{X}$. It means that $\mathbf{X}_k \cup \mathbf{X}_{\hat{k}} = \mathbf{X}$ and $\mathbf{X}_k \cap \mathbf{X}_{\hat{k}} = \varnothing$. According to Shekhar *et al.* [9], Γ is the constraint set to ensure the matrixes with simultaneous unity rowwise norm and relatively small Frobenius norm for a nontrivial and stable solution.

In (1), the first term means that each subdictionary can only transform the samples from the same class to coding coefficients. The second term constrains the subdictionary $\mathbf{\Omega}_k$ to transform the data from different class $i\,(i \neq k)$, to a nearly null space. That is to say, $\mathbf{\Omega}_k\mathbf{X}_i \approx \mathbf{0}\,\forall i \neq k$, leading to the block-diagonal structure of $\mathbf{A}$. It is an efficient alternative of the time-consuming sparsity constraint, which greatly reduces the time complexity and computational burden.

B. Max-Margin Regularization Term

Enforcing coding coefficients to be discriminative can indirectly promote the discrimination of dictionary. To make the coefficients as discriminative as possible, we employ a max-margin regularization term that ensures the coefficients of different classes to be separated by a max-margin. Intuitively, when the coefficients, as new representations of original data, are separated by a hyperplane, the large margin between different classes can boost the confidence of classification.

In a binary classification problem, we assume that $\mathbf{u}$ and b are the normal vector and bias determining the hyperplane of support vector machine (SVM). For coding coefficient matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n] \in \mathbb{R}^{m\times n}$, its corresponding label vector is $\mathbf{y} = [y_1, y_2, ..., y_n]$, where $y_i \in \{-1, +1\}$. The max-margin regularization term can be formulated as

$$\min_{\mathbf{u},b} f(\mathbf{A}, \mathbf{y}, \mathbf{u}, b) = \tfrac{1}{2}\|\mathbf{u}\|_2^2 + C\sum_{i=1}^{n} l(\mathbf{a}_i, y_i, \mathbf{u}, b) \tag{2}$$

where $l(\mathbf{a}_i, y_i, \mathbf{u}, b)$ is the hinge loss function indicating the classification error. Only if the margin is large enough, i.e., $y_i(\mathbf{u}^T\mathbf{a}_i + b) > 1$, the loss $l(\mathbf{a}_i, y_i, \mathbf{u}, b)$ is equal to 0. $\frac{1}{2}\|\mathbf{u}\|_2^2$ is the regularization term. $C > 0$ is the penalty parameter.

Integrating the objective function (2) into the class-aware analysis dictionary model, the classifier $< \mathbf{u}, b >$ will be learned by iteration algorithm. The basic ADL + SVM (ADL+SVM) method introduced in [9] also uses SVM classifier in testing stage. However, it separates the conventional ADL and the classifier training procedure. Compared with that, the SVM need be learned by training in our model, which not only contributes to enhancing discrimination of coding coefficients but also leads to an optimized SVM for classification.

C. Overall Formulation of CADL

Merging the class-aware analysis dictionary term and a max-margin regularizer, we can obtain the integrated CADL model. For multiclass tasks, we adopt the one-versus-all strategy to learn *K* hyperplanes (*K* is the number of classes), where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_K]$ and $\mathbf{b} = [b_1, b_2, ..., b_K]$. In this situation, CADL model can be formulated as

$$\min_{\boldsymbol{\Omega},\mathbf{A},\mathbf{U},\mathbf{b}} \sum_{k=1}^{K} \|\mathbf{A}_k - \boldsymbol{\Omega}_k \mathbf{X}_k\|_F^2 + \lambda_1 \|\boldsymbol{\Omega}_k \mathbf{X}_{\hat{k}}\|_F^2 + \lambda_2 f(\mathbf{A}, \mathbf{y}^k, \mathbf{u}_k, b_k),$$
$$\text{s.t.} \quad \boldsymbol{\Omega}_k \in \Gamma \quad \forall k \tag{3}$$

where λ_1 and λ_2 are the positive tradeoff parameters, $\mathbf{y}^k = [y_1^k, y_2^k, ..., y_{n_k}^k]$ is the label vector of samples in kth class, n_k is the number of the samples in kth class, and $y_i^k = 1$ if the ith sample belongs to the kth class, otherwise $y_i^k = -1$.

D. Discussion

The max-margin regularizer (2) can be transformed into the primal optimization problem of soft-margin SVM as follows:

$$\min_{\mathbf{u},b} \frac{1}{2}\|\mathbf{u}\|_2^2 + C\sum_{i=1}^{n} \xi_i$$
$$\text{s.t. } 1 - y_i(\mathbf{u}^T \mathbf{a}_i + b) \le \xi_i,\ \xi_i \ge 0\ \forall i \tag{4}$$

where ξ_i is the slack variable. The soft-margin SVM can better control the outliers. Its dual problem is formulated as

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j \mathbf{a}_i^T \mathbf{a}_j$$
$$\text{s.t.} \sum_{i=1}^{n} y_i \alpha_i = 0,\ 0 \le \alpha_i \le C \quad \forall i \tag{5}$$

where $\alpha = [\alpha_1, \alpha_2, ..., \alpha_n]^T$ is the Lagrange multiplier vector. $\sum_{i=1}^{n} \alpha_i$ can be regarded as a regularization term to avoid the trivial solution of α.

With the constraint $\sum_{i=1}^{n} y_i \alpha_i = 0$, we can obtain

$$-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j \mathbf{a}_i^T \mathbf{a}_j = \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\mathbf{a}_i - \mathbf{a}_j\|_2^2 y_i y_j \alpha_i \alpha_j. \tag{6}$$

Let $\mathbf{W}_{ij}(\alpha) = y_i y_j \alpha_i \alpha_j$. The objective function of the max-margin regularization term (2) can be rewritten as

$$\min_{\alpha} f(\mathbf{A}, \mathbf{W}_{ij}(\alpha)) = \sum_{i=1}^{n}\sum_{j=1}^{n} \|\mathbf{a}_i - \mathbf{a}_j\|_2^2 \mathbf{W}_{ij}(\alpha) \tag{7}$$

where the constraint $\sum_{i=1}^{n} y_i \alpha_i = 0, \alpha_i \ge 0\ \forall i$ is still satisfied. $\mathbf{W}_{ij}(\alpha)$ can be taken as the parameterization weight.

The function in (7), as the weighted sum of the squared Euclidean distance of coding vector pairs, usually serves for preserving the pairwise relations of samples in coding space to promote the discrimination of coding coefficients. In the process of learning the discriminative dictionary and coding coefficients, some coding vector pairs may play more important roles than others, which causes that the predefined weight assignment [12] may have a bad effect on final classification performance. The parameterized weight method has been proved to be more effective in [13], which is based on SDL. In this viewpoint, the max-margin regularization term (2) is proved to be valid in terms of enhancing the discriminative ability of our proposed CADL model.

IV. Optimization Procedure

Solving (3) is a challenging task because the CADL model in (3) is not a jointly convex optimization problem, but it is fortunate that it is convex for one variable when the other variables are fixed. Hence, we divide the model (3) into three subproblems by updating $\boldsymbol{\Omega}$, $\mathbf{A}$, and $\langle \mathbf{U}, \mathbf{b}\rangle$ alternatively.

1) When $\mathbf{A}$ and $\langle \mathbf{U}, \mathbf{b}\rangle$ are fixed, the optimization problem of each analysis subdictionary is formulated as

$$\arg\min_{\boldsymbol{\Omega}_k} \sum_{k=1}^{K} \|\mathbf{A}_k - \boldsymbol{\Omega}_k \mathbf{X}_k\|_F^2 + \lambda_1 \|\boldsymbol{\Omega}_k \mathbf{X}_{\hat{k}}\|_F^2,\ \text{s.t. } \boldsymbol{\Omega}_k \in \Gamma. \tag{8}$$

Referring to [9], we define the set Γ to be matrixes with relatively small Frobenius norm and unity rowwise norm, which can make the solution nontrivial and stable. Then, the problem (8) can be written as

$$\arg\min_{\boldsymbol{\Omega}_k} \sum_{k=1}^{K} \|\mathbf{A}_k - \boldsymbol{\Omega}_k \mathbf{X}_k\|_F^2 + \lambda_1 \|\boldsymbol{\Omega}_k \mathbf{X}_{\hat{k}}\|_F^2 + \lambda_3 \|\boldsymbol{\Omega}_k\|_F^2,$$

where λ_3 is the scalar parameter. The solution of $\boldsymbol{\Omega}$ can be easily obtained by setting the first-order derivative of above formula to zero, which is followed by normalizing each row of $\boldsymbol{\Omega}_k^*$ to unit norm.

2) When $\boldsymbol{\Omega}$ and $\langle \mathbf{U}, \mathbf{b}\rangle$ are fixed, each column of the coding coefficient matrix $\mathbf{A}$ can be optimized by

$$\mathbf{a}_i^* = \arg\min_{\mathbf{a}_i} \|\mathbf{a}_i - \boldsymbol{\Omega}\mathbf{x}_i\|_F^2 + \lambda_2 C \sum_{k=1}^{K} l(\mathbf{a}_i, y_i^k, \mathbf{u}_k, b_k).$$

The constant terms $\|\boldsymbol{\Omega}_k \mathbf{X}_{\hat{k}}\|_F^2$ and $\|\mathbf{u}\|_2^2$ are omitted since they have no impact on the suboptimization of $\mathbf{A}$.

Here, we adopt the quadratic hinge loss function for its computational simplicity and better smooth property, which can generate similar results with the common hinge loss function [14]. The definition of quadratic hinge loss function is $l(\mathbf{a}_i, y_i^k, \mathbf{u}_k, b_k) = \|1 - y_i^k(\mathbf{u}_k^T \mathbf{a}_i + b_k)\|^2$ if $1 - y_i^k(\mathbf{u}_k^T \mathbf{a}_i + b_k) > 0$, otherwise $l(\mathbf{a}_i, y_i^k, \mathbf{u}_k, b_k) = 0$. Taking the derivation of $\mathbf{a}_i$ can lead to the closed solution.

3) When $\mathbf{A}$ and $\boldsymbol{\Omega}$ are fixed, the optimization of $\langle \mathbf{U}, \mathbf{b}\rangle$ is a multiclass SVM problem, which can be solved by the SVM solver proposed in [14].

In each iteration, we have the closed-solution for $\boldsymbol{\Omega}$ and a conditional solution for $\mathbf{A}$. According to [14], we can know that the linear SVM solver can make the quadratic loss function converge finally. Besides, the objective function value of class-aware analysis dictionary term is also decreasing gradually. Therefore, the CADL algorithm will converge to a stable point. The convergence curve of our algorithm on the AR face dataset [15] is shown in Fig. 1.

After learning the class-aware analysis dictionary $\boldsymbol{\Omega}$ and classifier $\langle \mathbf{U}, \mathbf{b}\rangle$ in the training stage, we can perform classification on the testing set. Given a testing sample $\mathbf{x}_{\text{test}}$, we first need to obtain its coding coefficient by $\mathbf{a}_{\text{test}} = \boldsymbol{\Omega}\mathbf{x}_{\text{test}}$. Note that our CADL does not need sparsity constraint because of the block-diagonal structure of $\mathbf{A}$, which is indicated in Section III-A. In the coding coefficient space, we adopt the learned K one-against-all classifiers $\langle \mathbf{u}_k, b_k\rangle, k \in \{1, 2, ..., K\}$, to predict the label y_{test} of $\mathbf{x}_{\text{test}}$.

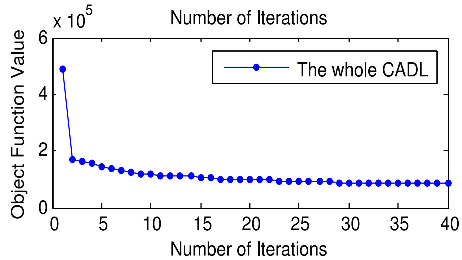

Fig. 1. Convergence curve on AR database.

TABLE I
OPTIMAL PARAMETERS IN DIFFERENT DATABASES

	Caltech 101	Scene 15	UCF 50	AR
λ_1	5e-4	1e-4	1e-2	4e-3
λ_2	5e-4	1e-2	3e-5	4e-4
λ_3	5e-3	1e-4	1e-3	4e-4

V. EXPERIMENTS

In this section, we conduct experiments on four widely used classification datasets, i.e., Caltech 101, Scene 15, UCF 50, and AR face dataset, to evaluate the performance of CADL. All these datasets are widely used to evaluate DL algorithms in previous works. Their features are provided by Jiang [16][1] and Corso [17],[2] respectively. To highlight the advantages of CADL, we compare CADL with different kinds of state-of-the-art DL methods including: 1) sparse representation-based classifier (SRC) [18], collaborative representation-based classifier (CRC) [19] and representative discriminative SDL methods, such as DL with structured incoherence (DLSI) [2], Fisher Discrimination Dictionary Learning (FDDL) [20], and Label Consistent KSVD (LC-KSVD) [16], etc. 2) two ADL based methods: the basic ADL + SVM [9] and discriminative analysis dictionary learning (DADL) [12]; 3) the recently proposed projective dictionary pair learning (DPL) [11] with SDL and ADL integrated. Moreover, we also perform experiments that utilize SVM classifier in DADL and DPL, named as DADL+SVM and DPL + SVM, to further testify the superiority of CADL.

There are four parameters in our model, i.e., λ_1, λ_2, λ_3, and C, where $C = 0.2$ is preset and the optimal λ_1, λ_2, and λ_3 are obtained by cross validation. The parameters we set in each database are listed in Table I.

A. Datasets

Caltech 101 dataset [21] contains 9144 images from 101 object classes and one background class. The sample size of each class varies from 30 to 800. Scene 15 [22] has 15 natural scene categories. Each category has about 200 to 400 image samples. UCF 50 [23] is a large-scale and challenging action recognition database. It has 50 action categories and 6680 realistic human action videos collected from YouTube. AR face database [15] contains illumination, expression, and occlusions variations. We choose a subset consisting of 2600 face images from 50 males and 50 females. The detailed experimental settings can refer to [16] and [12].

[1]http://www.umiacs.umd.edu/∼ zhuolin/projectlcksvd.html.
[2]http://www.cse.buffalo.edu/∼ jcorso/r/actionbank.

TABLE II
CLASSIFICATION ACCURACIES (%) ON FOUR DATASETS

	Caltech 101	Scene 15	UCF 50	AR
SRC	70.7	91.8	75.0	97.5
CRC	68.2	92.0	75.6	98.0
DLSI	73.1	91.7	75.4	97.5
FDDL	73.2	92.3	76.5	97.5
LC-KSVD	73.6	92.9	70.1	97.8
ADL+SVM	64.5	90.1	72.3	96.1
DADL	74.6	98.3	**78.0**	98.7
DADL+SVM	74.3	97.9	77.3	97.3
DPL	73.9	97.7	77.4	98.3
DPL+SVM	74.8	98.3	75.0	98.3
CADL	**75.0**	**98.6**	**78.0**	**98.8**

B. Experiment Results and Analyses

The classification accuracies and running times of all the competing methods are listed in Table II. By contrast, it can be observed that our proposed CADL achieves the best classification performance on four different classification tasks. The obviously higher accuracy than the basic ADL+SVM shows that the CADL method succeeds in promoting the discrimination of the conventional ADL. It also demonstrates the effectiveness of the max margin term and class-aware dictionary term. The better performance than DADL+SVM and DPL+SVM proves the validity of the class-aware analysis dictionary in CADL. DADL improves ADL model by a different but efficient strategy. In spite of this, CADL is still competitive.

As for the running efficiency, CADL has higher testing efficiency than DADL and DPL that have been proved to be pretty effective in [11] and [12], but the training efficiency is low for incorporating SVM into training model. There is no sparsity constraint on the coding coefficients like l_0-norm or l_1-norm, so the computation complexities in training and testing stages are greatly reduced. However, the training times are still consumed for training SVM iteratively. The tradeoff between training times and accuracy cannot be avoided. Due to page limit, we cannot show the competing results of running efficiency. For fair comparison, all the experiments are implemented in MATLAB R2012a on a computer with 32 GB memory and 2.6 GHz Intel CPU.

VI. CONCLUSION

This letter proposed a novel Class-Aware Analysis Dictionary Learning (CADL) method for classification tasks by fully exploiting class priors. CADL mainly focuses on two points to promote the discriminability of ADL. One point is to learn a discriminative analysis subdictionary for each class instead of a global dictionary for all classes. The other point is to enhance the discrimination of coding coefficients by adding a max-margin regularization term. We also give a theoretical analysis to support the max-margin regularization term from the perspective of preserving the pairwise relations of samples in coding space. To efficiently optimize CADL, we employ an alternating minimization algorithm to find the convergent solution. The comparative experimental results on different classification tasks verify the superiority of our CADL method to state-of-the-art DL methods.

References

[1] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. Comput. Vis. Pattern Recognit.*, pp. 1794–1801, Jun. 2008.

[2] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. Comput. Vis. Pattern Recognit.*, pp. 3501–3508, Jun. 2010.

[3] G. Zhang, M. Zhang, R. He, and Z. Sun, "Jointly learning dictionaries and subspace structure for video-based face recognition," in *Proc. Asian Conf. Comput. Vis.*, 2014, pp. 97–111.

[4] S. Zhang, M. Zhang, R. He, and Z. Sun, "Transform-invariant dictionary learning for face recognition," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 348–352.

[5] J. Wang, Y. Guo, J. Guo, M. Li, and X. Kong, "Synthesis linear classifier based analysis dictionary learning for pattern classification," *Neurocomputing*, vol. 238, pp. 103–113, 2017.

[6] P. Sprechmann, R. Litman, T. B. Yakar, A. M. Bronstein, and G. Sapiro, "Supervised sparse analysis and synthesis operators," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 908–916.

[7] R. Rubinstein, T. Peleg, and M. Elad, "Analysis k-SVD: A dictionary-learning algorithm for the analysis sparse model," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 661–677, Feb. 2013.

[8] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1072–1086, Mar. 2013.

[9] S. Shekhar, V. M. Patel, and R. Chellappa, "Analysis sparse coding models for image-based classification," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, pp. 5207–5211.

[10] R. Rubinstein and M. Elad, "Dictionary learning for analysis-synthesis thresholding," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5962–5972, Nov. 2014.

[11] S. H. Gu, L. Zhang, W. Zuo, and X. Feng, "Projective dictionary pair learning for pattern classification," in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 793–801, Dec. 2014.

[12] J. Guo, Y. Guo, X. Kong, M. Zhang, and R. He, "Discriminative analysis dictionary learning," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 1617–1623.

[13] S. Cai, W. Zuo, L. Zhang, X. Feng, and P. Wang, "Support vector guided dictionary learning," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 624–639.

[14] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1794–1801.

[15] A. Martinez and R. Benavente, "The AR face database," CVC, Universitat Autònoma de Barcelona, Tech. Rep., vol. 24, 1998.

[16] Z. L. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov. 2013.

[17] S. Sadanand and J. J. Corso, "Action bank: A high-level representation of activity in video," in *Proc. Comput. Vis. Pattern Recognit.*, pp. 1234–1241, Jun. 2012.

[18] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[19] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 471–478.

[20] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 543–550.

[21] F. F. Li, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," *Comput. Vis. Image Understanding*, vol. 106, no. 1, pp. 59–70, 2007.

[22] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. Comput. Vis. Pattern Recognit.*, vol. 2, pp. 2169–2178, Jun. 2006.

[23] K. K. Reddy and M. Shah, "Recognizing 50 human action categories of web videos," *Mach. Vis. Appl.*, vol. 24, no. 5, pp. 971–981, Jul. 2013.