IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ADVANCED OPTICAL IMAGING FOR EXTREME ENVIRONMENTS

Received March 24, 2019, accepted April 8, 2019, date of publication April 23, 2019, date of current version May 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2912932

Graph-Regularized Discriminative Analysis-Synthesis Dictionary Pair Learning for Image Classification

HEYOU CHANG[1,2], HUI TANG[2], FANLONG ZHANG[3], YANG CHEN[2], (Senior Member, IEEE), AND HAO ZHENG[1]

[1]Key Laboratory of Trusted Cloud Computing and Big Data Analysis, Nanjing Xiaozhuang University, Nanjing 211171, China
[2]Laboratory of Image Science and Technology, Southeast University, Nanjing 211171, China
[3]School of Technology, Nanjing Audit University, Nanjing 211171, China

Corresponding author: Heyou Chang (cv_hychang@126.com)

This work was supported in part by the National Natural Science Fund of China under Grant 61806098, 61603192, 81530060 and 81471752, in part by the State's Key Project of Research and Development Plan under Grant 2017YFA0104302, 2017YFC0109202 and 2017YFC0107900, in part by Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant 18KJB520029, and in part by the Nanjing Xiaozhuang University under Grant 2017NXY49.

ABSTRACT Analysis-synthesis dictionary pair learning, which can provide a comprehensive view of data representation, has been applied in various computer vision tasks. Although good performance has been reported in image denoising, discriminative dictionary pair learning for image classification remains unsolved. In this paper, we propose a novel model of graph-regularized discriminative analysis-synthesis dictionary pair learning (GDASDL), in which a graph-regularized term and a discriminative term are incorporated into dictionary pair learning. By taking advantage of graph constraints, the proposed GDASDL can preserve the local geometry structure of the data. Global information is introduced by associating label information with dictionary atoms. In this paper, an iteration algorithm is presented to efficiently solve the proposed GDASDL. We extensively conduct experiments on three public image datasets and one face dataset in comparison with the existing dictionary learning approaches, and the experimental results show that the proposed model achieves superior performance using a simple linear classifier.

INDEX TERMS Representation learning, dictionary learning, image classification, local geometry structure.

I. INTRODUCTION

In recent years, sparse representation has been successfully applied to various vision tasks, including image denoising [1]–[3], face recognition [4], [5] and image classification [6]–[8]. In sparse representation, the dictionary (i.e., a set of representation bases) plays a significant role, which can be set as off-the-shelf bases (wavelets) or learned from training data. It has been shown that a learned dictionary is more powerful than a prespecified dictionary in some specific applications(such as image classification).

According to the way of encoding input signals, the dictionary learned in sparse representation can be classified into two categories: *synthesis dictionary* and *analysis dictionary*. A synthesis dictionary is used to approximate an input signal by a linear combination of a few dictionary atoms. Specifically, $\mathbf{x} = \mathbf{DZ}$, where $\mathbf{x} \in \mathcal{R}^{m\times 1}$ is the input signal, $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_k] \in \mathcal{R}^{m\times k}$ is a dictionary with atoms as its columns, and $\mathbf{z} \in \mathcal{R}^{k\times 1}$ is the sparse representation. An analysis dictionary $\Omega = [\boldsymbol{\omega}_1^T; \ldots; \boldsymbol{\omega}_k^T] \in \mathcal{R}^{k\times m}$ with atoms as its rows aims to represent a signal with $\mathbf{z} = \Omega\mathbf{x}$, where $\mathbf{z}$ contains mostly zeros. Some recent studies [9], [10] have shown compelling relations between the analysis dictionary and the synthesis dictionary, and further research in this direction is pending. According to the property of the dictionary, current prevailing dictionary learning approaches can be mainly divided into three groups: *synthesis dictionary learning*(SDL), *analysis dictionary learning*(ADL), and *analysis-synthesis dictionary pair learning*(ASDL).

SDL has been well studied, and most existing dictionary learning methods belong to this category. Among these methods, the learned synthesis dictionary can be class-shared, where the dictionary is shared with all classes [1], [4], [11], or class-specific [6], [12], where each dictionary atom has a single class label. To separate common patterns from the overall dictionary and improve the classification performance

The associate editor coordinating the review of this manuscript and approving it for publication was Huimin Lu.

of sparse representation, some hybrid dictionary learning methods, which contains a shared sub-dictionary and a set of class-specific sub-dictionary, have been proposed [8], [13]. Although good performance has been reported, the inherent high time-consumption of SDL is unavoidable.

Recently, ADL has attracted considerable attention. Compared to the synthesis dictionary, the analysis dictionary representation has some unique merits, e.g., generalization to feature transformation and image convolution. The representative ASL approach is analysis K-SVD [14]. In [14], Rubinstein *et al.* proposed to learn an analysis dictionary and adopted an algorithm similar to the K-SVD algorithm [1] to solve the analysis dictionary. Analysis K-SVD has good performance on image processing but is incompetent for classification. Later, Shekhar *et al.* [15] adopted ADL for a recognition task, and demonstrated that ADL was robust and efficient and performed competitively with SDL. By integrating the local topological structures and discriminative sparse labels into the ADL, Guo *et al.* [16] proposed discriminative ADL (DADL). Tang *et al.* [17] incorporated the class structural information in ADL and proposed structured ADL (SADL).

To learn the data representation in a faster and more comprehensive way, ASDL, which inherits the advantages of both SDL and ASL, has been proposed. In [10], Rubinstein *et al.* proposed an ASDL model for image denoising, in which a pair of class-shared synthesis dictionary and analysis dictionary was learned from an image patch set. Unlike ASDL, Gu *et al.* [18] tired to learn a class-specific dictionary pair and proposed projective dictionary pair learning (PDPL) model. By requiring both the synthesis class-specific dictionary and analysis class-specific dictionary to represent the corresponding class well but to represent the other classes poorly, PDPL achieved good performance in image classification. Later, Gu *et al.* [19] introduced analysis sparse representation and synthesis sparse representation into convolutional sparse representation learning and proposed joint convolutional analysis and synthesis (JCAS) sparse representation models for different image processing tasks. By embedding Fisher-like discrimination information into the analysis representation, the analysis dictionary and the synthesis representation, Yang *et al.* [20] proposed a model of Fisher discrimination dictionary pair learning (FDDPL). Since ASDL ignores the introduction of discriminative information into the dictionary learning process, its application for classification tasks is limited. By incorporating a discriminative term into the objective function, Yang *et al.* [21] proposed a discriminative analysis-SDL (DASDL) model to enhance the discriminative ability of ASDL. Although improved performance has been reported in existing ASDL approaches, the discriminative information embedded in the data (e.g., structural information between data points) are not fully exploited, and methods of learning a more discriminative analysis-synthesis pair dictionary are still being considered.

In practice, it is rational to assume that the representations of signals in a new feature space should be similar if the signals come from the same class. Some works have revealed that the geometric and discriminating structures of signals are essential to image classification tasks. To preserve local geometric structures embedded in the original data space, numerous manifold learning methods have been proposed, such as locally linear embedding (LLE) [23], locality preserving projection (LPP) [24] and Laplacian Eigenmaps (LE) [25]. Liu *et al.* [26] proposed a locality-sensitive dictionary learning algorithm with global consistency and smoothness constraints. Li *et al.* [27] proposed a locality-constrained and label embedding dictionary learning (LCLE-DL) algorithm for image classification, where a graph Laplacian matrix of the learned dictionary and a label embedding term were constructed to preserve the locality information and label information. Jiang *et al.* [28] proposed a joint dictionary learning algorithm for face sketch synthesis by utilizing the locality and manifold geometry of a data space. Zheng *et al.* [29] considered the local manifold structures of data and proposed a graph-regularized sparse coding method.

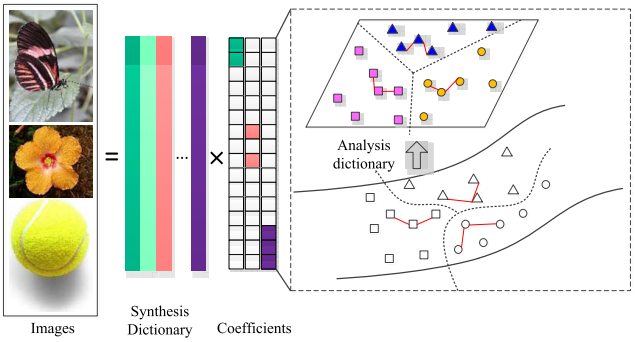


FIGURE 1. Illustration of our methods. The synthesis dictionary is trained to represent the images well, and the structure information is transformed from the original image space to a more discriminative space using an analysis dictionary.

Motivated by the aforementioned works, in this paper, we propose a novel model of GDASDL for image classification. In the proposed model, the structure information is modeled by a nearest neighbor graph and then transformed from the original image space to a new space, as shown in Fig.1. Label information is also incorporated to encourage the optimal representation to be close to the block-diagonal. Benefiting from regularization, the within-class representation scatter is small, and the between-class representation scatter is large, resulting in good performance even when using a simple linear classifier. To evaluate the proposed model, a series of experiments on three image datasets and one face dataset are conducted. The experimental results demonstrate that the proposed GDASDL achieves very promising performance in image classification task.

The remainder of this paper is organized as follows. Section 2 provides a brief review of related works. Section 3 presents the proposed GDASDL model. Section 4 describes the optimization of GDASDL. Section 5 discusses experiments on image classification, and Section 6 concludes the paper.

II. RELATED WORKS

In this section, we will briefly review prior studies on SDL, ADL, and ASDL.

A. SDL

Given an original signal, SDL methods focus on the minimization of

$$\min_{\mathbf{D},\mathbf{z}} \|\mathbf{x}-\mathbf{D}\mathbf{z}\|_F^2+\lambda\|z\|_1+f(\mathbf{D})+h(\mathbf{z}) \tag{1}$$

where $f(\mathbf{D})$ and $h(\mathbf{z})$ are functions on $\mathbf{D}$ and $\mathbf{z}$, respectively. For example, label consistent constraint [4], Fisher discrimination term [6], locality-constrained term [27]. The functions are designed according to the tasks.

B. ADL

The conventional ADL problem aims to obtain an analysis dictionary via

$$\begin{aligned}&\min_{\Omega,\mathbf{z}} \ |\mathbf{z}-\Omega\mathbf{x}\|_F^2+\lambda\|z\|_1+h(\mathbf{z})\\ &\text{s.t. } \Omega\in\Phi\end{aligned} \tag{2}$$

where Φ is a set of constraints on Ω to make the solution non-trivial.

To improve the model's performance for image classification, Guo *et al.* [16] constructed a code consistent term $\|\mathbf{Z}-\mathbf{H}\|_F^2$ and a local topology preserving loss function $\sum_{i=1}^n\sum_{j=1}^n\{\mathbf{W}_{i,j}\|\mathbf{z}_i-\mathbf{z}_j\|_2^2\}$, where $\mathbf{H}$ is a target code and $\mathbf{W}$ is a weighting matrix. Instead of incorporating the local structure information, Tang *et al.* [17] embedded a classification error $\|\mathbf{L}-\mathbf{W}\mathbf{H}\|_F^2$ into (2), where $\mathbf{L}$ is a label matrix.

C. ANALYSIS-SYNTHESIS DICTIONARY LEARNING

Recently, Rebinstein and Elad [10] proposed an ASDL algorithm by learning a pair of an analysis dictionary and a synthesis dictionary via

$$\begin{aligned}&\min_{\Omega,\mathbf{D},\boldsymbol{\lambda}} \ |\mathbf{X}-\mathbf{D}S_{\boldsymbol{\lambda}}(\Omega\mathbf{X})\|_F^2\\ &\text{s.t. } |\boldsymbol{\omega}_i\|_2=1\ \forall i\end{aligned} \tag{3}$$

where $\mathbf{X}=[\mathbf{x}_1,\ldots,\mathbf{x}_n]\in\mathcal{R}^{m\times n}$ is an observation data matrix, $\mathbf{D}=[\mathbf{d}_1,\ldots,\mathbf{d}_k]\mathcal{R}^{m\times k}$ is a synthesis dictionary, $\Omega=[\boldsymbol{\omega}_1^T;\ldots;\boldsymbol{\omega}_k^T]\in\mathcal{R}^{k\times m}$ is an analysis dictionary, $\boldsymbol{\lambda}=[\lambda_1,\ldots,\lambda_k]$ is a threshold vector and λ_i is the threshold for the i-th row of Ω. $S_{\boldsymbol{\lambda}}(\cdot)$ is a function operating on matrices with m rows and defined as follows

$$S_{\boldsymbol{\lambda}}([\Omega\mathbf{X}]_{i,j})=\begin{cases}[\Omega\mathbf{X}]_{i,j}, & if\ |[\Omega\mathbf{X}]_{i,j}|\geq\lambda_i\\ 0, & otherwise\end{cases}$$

where $[\mathbf{A}]_{i,j}$ denotes the element in the ith row and jth column of matrix $\mathbf{A}$. The l_2 norm of $\boldsymbol{\omega}_i$ is required to avoid a trivial solution.

An approach similar to K-SVD and analysis K-SVD is adopted to optimize problem (3). Specifically, only one pair of atoms is updated at each step by keeping all other pairs fixed. For the optimization of each pair of dictionary atoms, a rank-one approximation approach is used.

Since no discriminative information is considered, ASDL is not qualified for classification. To address this limitation, Yang *et al.* [21] proposed learning a discriminative analysis-synthesis dictionary by incorporating a discriminative term. The objective function of DASDL can be formulated as

$$\begin{aligned}&\min_{\Omega,\mathbf{D},\mathbf{P},\boldsymbol{\lambda}} \ |\mathbf{X}-\mathbf{D}S_{\lambda}(\Omega\mathbf{X})\|_F^2+\alpha\|\mathbf{Y}-\mathbf{P}S_{\lambda}(\Omega\mathbf{X})\|_F^2\\ &\text{s.t. } |\mathbf{d}_i\|_2=1\ \forall i;\ \|\mathbf{P}\|_F^2\leq\sigma\end{aligned} \tag{4}$$

where $\mathbf{P}\in\mathcal{R}^{C\times k}$ and $\mathbf{Y}=[\mathbf{y}_1,\ldots,\mathbf{y}_n]\in\mathcal{R}^{C\times n}$ are a linear projection matrix and a label matrix, respectively. The second term in (4) is a classification error term which aims to project the i-th class coefficients to only the i-th dimension of the label space. Problem (4) is effectively solved by a similar algorithm to problem (3).

III. GRAPH-REGULARIZED DISCRIMINATIVE ANALYSIS-SYNTHESIS DICTIONARY PAIR LEARNING

In this section, a novel ASDL model with graph regularization and discrimination constraint is proposed. The coefficients calculated by the new model can capture the local geometric structures among the data and consider label information. Thus, the proposed model can effectively determine the representations by relying on a small analysis dictionary.

A. GRAPH-REGULARIZED TERM FOR COEFFICIENTS

Since a nonlinear space can always be locally approximated by several linear subspaces, it is valid to assume that there is a linear relationship between a signal and its neighbors. Based on this assumption, the local geometry of the signals can be characterized by the linear coefficients that reconstruct each signal from its neighbors. Let $\mathbf{X}=[\mathbf{x}_1,\ldots,\mathbf{x}_n]$ be the data set,; we first construct a weighted undirected graph $G=(V,E;\mathbf{A})$, where $V=\{v_i\}_{i=1}^n$, $E=\{e_{i,j}\}$, and $\mathbf{A}\in\mathcal{R}^{n\times n}$ are the vertex set, edge set and weight matrix, respectively. In V, each node v_i corresponds to a data point $\mathbf{x}_i$. Edge $e_{i,j}$ represents the association between nodes v_i and v_j. $\mathbf{A}$ is a symmetric weight matrix, and the value of $\mathbf{A}_{i,j}$ is formulated as follows

$$\mathbf{A}_{i,j}=\begin{cases}1, & if\ \mathbf{x}_i\in KNN(\mathbf{x}_j)\ or\ \mathbf{x}_j\in KNN(\mathbf{x}_i)\\ 0, & otherwise\end{cases}$$

where $KNN(\mathbf{x}_j)$ denotes the set of K nearest neighbors of $\mathbf{x}_j$.

Assume that, if two signals are close in the original space, then their coefficients are also close to each other in a new space. Then, the graph-regularized term can be formulated as

$$\min_{\mathbf{Z}}\sum_{i=1}^n\sum_{j=1}^n\|\mathbf{z}_i-\mathbf{z}_j\|_2^2\mathbf{A}_{i,j} \tag{5}$$

where $\mathbf{z}_i$ and $\mathbf{z}_j$ are the coefficients of $\mathbf{x}_i$ and $\mathbf{x}_j$, respectively, under the learned dictionary. This strategy has been widely used in many applications [27], [28]. The graph-regularized term can preserve the local affinity between signals and has a positive impact on the classification performance.

B. DISCRIMINATION CONSTRAINT TERM FOR COEFFICIENTS

For image classification task, label information is significant. To utilize the label information effectively, we construct a block-diagonal matrix $\mathbf{Q} \in \mathcal{R}^{k\times n}$ shown as follows

$$\mathbf{Q} = \begin{bmatrix} 1 & 1 & \dots & 0 & 0 \\ 1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & \dots & 1 & 1 \end{bmatrix}$$

where $\mathbf{Q}_{i,j} = 1$, if $\mathbf{d}_i$ and $\mathbf{x}_j$ belong to the same class; otherwise, $\mathbf{Q}_{i,j} = 0$. Instead of encouraging the representations to be close to $\mathbf{Q}$, we propose transforming the representations to be the most discriminative representations in feature space $\mathcal{R}^k$. Then, the discrimination constraint term can be written as

$$\min_{\mathbf{P}} \|\mathbf{Q} - \mathbf{P}\mathbf{Z}\|_F^2 \tag{6}$$

where $\mathbf{P}$ is a linear transformation. The discrimination constraint term enforces the signals from the same class to have similar representations and those from different classes to have dissimilar representations. Since this term improves the consistency of the representations within a class and enhances the divergence among different classes, good performance can be achieved using a simple linear classifier.

C. GDASDL MODEL

By jointly taking the graph regularization and the discrimination constraint into consideration, the proposed GDASDL model can be formulated as

$$\begin{aligned} \min_{\Omega,\mathbf{D},\mathbf{P},\boldsymbol{\lambda}} \; & \|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2 + \alpha\|\mathbf{Q} - \mathbf{P}\mathbf{Z}\|_F^2 \\ & + \beta \sum_{i=1}^{n}\sum_{j=1}^{n} \|\mathbf{z}_i - \mathbf{z}_j\|_2^2 \mathbf{A}_{i,j} \\ \text{s.t. } & \mathbf{Z} = S_\lambda(\Omega\mathbf{X});\ \|\mathbf{d}_i\|_2 = 1\ \forall i;\ \|\mathbf{P}\|_F^2 \le \sigma \end{aligned} \tag{7}$$

where α and β are scalar parameters. In problem (7), the first term is a reconstruction error term, which ensures that the learned dictionary pair ($\mathbf{D}$,Ω) can represent the data well. The last two terms represent the discrimination constraint term and the graph-regularized term, which are used to improve the classification ability of the proposed model.

IV. OPTIMIZATION

To efficiently solve problem (7), we first reformulate problem (7) as

$$\begin{aligned} \min_{\Omega,\mathbf{D},\mathbf{P},\boldsymbol{\lambda}} \; & \left\| \begin{pmatrix} \mathbf{X} \\ \sqrt{\alpha}\mathbf{Q} \end{pmatrix} - \begin{pmatrix} \mathbf{D} \\ \sqrt{\alpha}\mathbf{P} \end{pmatrix} S_\lambda(\Omega\mathbf{X}) \right\|_F^2 \\ & + \beta \sum_{i=1}^{n}\sum_{j=1}^{n} \|S_\lambda(\Omega\mathbf{x}_i) - S_\lambda(\Omega\mathbf{x}_j)\|_2^2 \mathbf{A}_{i,j} \\ \text{s.t. } & \|\mathbf{d}_i\|_2 = 1\ \forall i;\ \|\mathbf{P}\|_F^2 \le \sigma \end{aligned} \tag{8}$$

Let $\hat{\mathbf{X}} = \begin{pmatrix} \mathbf{X} \\ \sqrt{\alpha}\mathbf{Q} \end{pmatrix}$ and $\hat{\mathbf{D}} = \begin{pmatrix} \mathbf{D} \\ \sqrt{\alpha}\mathbf{P} \end{pmatrix}$. The regularization penalty term $\|\mathbf{P}\|_F^2 \le \sigma$ can be dropped since $\hat{\mathbf{D}}$ is normalized column-wise on subsequent passes. In [21], Yang *et al.* verified that the accuracy of DASDL with a predefined threshold has a performance similar to that with a learning threshold. For the convenience of solving problem (8), we simply set $\lambda_1 = \lambda_2 = \cdots = \lambda_k = c$, where c is a constant. Then, problem (8) can be reformulated as

$$\begin{aligned} \min_{\Omega,\hat{\mathbf{D}}} \; & \|\hat{\mathbf{X}} - \hat{\mathbf{D}} S_\lambda(\Omega\mathbf{X})\|_F^2 \\ & + \beta \sum_{i=1}^{n}\sum_{j=1}^{n} \|S_\lambda(\Omega\mathbf{x}_i) - S_\lambda(\Omega\mathbf{x}_j)\|_2^2 \mathbf{A}_{i,j} \\ \text{s.t. } & \|\hat{\mathbf{d}}_i\|_2 = 1\ \forall i \end{aligned} \tag{9}$$

To efficiently solve problem (9), we adopt a strategy that updates the dictionary pair atom by atom. Specifically, at the mth step, all but the mth pair of atoms are kept fixed. Then, we can isolate the dependence on the mth atom pair, and rewrite problem (9) as

$$\begin{aligned} \min_{\boldsymbol{\omega}_m,\hat{\mathbf{d}}_m} \; & \|\hat{\mathbf{E}} - \hat{\mathbf{d}}_m S_{\lambda_m}(\boldsymbol{\omega}_m^T\mathbf{X})\|_F^2 \\ & + \beta \sum_{i=1}^{n}\sum_{j=1}^{n} \|S_{\lambda_m}(\boldsymbol{\omega}_m^T\mathbf{x}_i) - S_{\lambda_m}(\boldsymbol{\omega}_m^T\mathbf{x}_j)\|_2^2 \mathbf{A}_{i,j} + C \\ \text{s.t. } & \|\hat{\mathbf{d}}_m\|_2 = 1 \end{aligned} \tag{10}$$

where $\hat{\mathbf{E}} = \hat{\mathbf{X}} - \sum_{l\neq m} \hat{\mathbf{d}}_l S_{\lambda_l}(\boldsymbol{\omega}_l^T\mathbf{X})$, and C is a constant. Although problem (10) is not joint convex to $(\boldsymbol{\omega}_m, \hat{\mathbf{d}}_m)$, it is convex with respect to both $\boldsymbol{\omega}_m$ and $\hat{\mathbf{d}}_m$ when the other is fixed. Therefore, we adopt an alternative optimization algorithm to solve problem (10).

Let $S_{\lambda_m}(\cdot)$ partition $\mathbf{X}$ in two sets, i.e., $\mathbf{X}^J \in \mathcal{R}^{m\times n_1}$ and $\mathbf{X}^{\bar{J}} \in \mathcal{R}^{m\times n_2}$, with current $\boldsymbol{\omega}_m$, where $n = n_1 + n_2$, $S_{\lambda_m}(\boldsymbol{\omega}_m^T\mathbf{X}^J) = \boldsymbol{\omega}_m^T\mathbf{X}^J$, and $S_{\lambda_m}(\boldsymbol{\omega}_m^T\mathbf{X}^{\bar{J}}) = 0$. $\hat{\mathbf{E}}$ is similarly split into submatrices $\hat{\mathbf{E}}^J$ and $\hat{\mathbf{E}}^{\bar{J}}$. Then, we approximate problem (10) as

$$\begin{aligned} \min_{\boldsymbol{\omega}_m,\hat{\mathbf{d}}_m} \; & \|\hat{\mathbf{E}}^J - \hat{\mathbf{d}}_m\boldsymbol{\omega}_m^T\mathbf{X}^J\|_F^2 + \beta(2\sum_{i=1}^{n_1}\|\boldsymbol{\omega}_m^T\mathbf{x}_i^J\|_2^2 \sum_{j=1}^{n_2}\mathbf{A}_{i,j}^{\bar{J}} \\ & + \sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\|\boldsymbol{\omega}_m^T\mathbf{x}_i^J - \boldsymbol{\omega}_m^T\mathbf{x}_j^J\|_2^2\mathbf{A}_{i,j}^J) \\ \text{s.t.} \; & \|\hat{\mathbf{d}}_m\|_2 = 1 \end{aligned} \tag{11}$$

Let $\Gamma(\boldsymbol{\omega}_m, \hat{\mathbf{d}}_m) = \arg\min_{\boldsymbol{\omega}_m,\hat{\mathbf{d}}_m} \|\hat{\mathbf{E}}^J - \hat{\mathbf{d}}_m\boldsymbol{\omega}_m^T\mathbf{X}^J\|_F^2 + \beta(2\sum_{i=1}^{n_1}\|\boldsymbol{\omega}_m^T\mathbf{x}_i^J\|_2^2\sum_{j=1}^{n_2}\mathbf{A}_{i,j}^{\bar{J}} + \sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\|\boldsymbol{\omega}_m^T\mathbf{x}_i^J - \boldsymbol{\omega}_m^T\mathbf{x}_j^J\|_2^2\mathbf{A}_{i,j}^J)$, then the partial derivatives with respect to $\Gamma(\boldsymbol{\omega}_m, \hat{\mathbf{d}}_m)$ are calculated by

$$\frac{\partial\Gamma}{\partial\hat{\mathbf{d}}_m} = \hat{\mathbf{E}}^J\mathbf{X}^{J^T}\boldsymbol{\omega}_m^T(\boldsymbol{\omega}_m\mathbf{X}^J\mathbf{X}^{J^T})^{-1} \tag{12}$$

$$\frac{\partial\Gamma}{\partial\boldsymbol{\omega}_m} = ((\hat{\mathbf{d}}_m^T\hat{\mathbf{d}}_m + \beta\mathbf{L}_m + 2\beta\Lambda)\mathbf{X}^{J^T})^{-1}\hat{\mathbf{E}}^{J^T}\hat{\mathbf{d}}_m \tag{13}$$

where $\mathbf{L}_m = \mathbf{S}^J - \mathbf{A}^J_{i,j}$, $\mathbf{S}^J$ is a diagonal matrix and the ith diagonal entry $\mathbf{S}^J_{i,i} = \sum_j \mathbf{A}^T_{i,j}$. Λ is a diagonal matrix with $\Lambda_{i,i} = \sum_j \mathbf{A}^{\bar{J}}_{i,j}$.

The whole algorithm for solving GDASDL is summarized in Alg.1. In our experiments, $\hat{\mathbf{D}} = [\hat{\mathbf{D}}_1, \ldots, \hat{\mathbf{D}}_c]$ is initialized class by class via K-SVD [1] from $\hat{\mathbf{X}}$, and Ω is initialized as $(\hat{\mathbf{D}}^T\hat{\mathbf{D}})^{-1}\hat{\mathbf{D}}^T\hat{\mathbf{X}}\hat{\mathbf{X}}^T$. Fig. 2 shows a convergence curve of the solving algorithm on the CIFAR-10 dataset. We can see that the value of the loss function (7) decreases gradually with the increase of iteration number, and Alg. 1 converges to a satisfactory local optimal solution after several iterations.

Algorithm 1 Solving the GDASDL model algorithm

Input: Training data $\mathbf{X}$, discriminative information $\mathbf{Q}$, the number of nearest neighbors K, parameters α, β and $\boldsymbol{\lambda}$

1: **Initialization**: Compute weight matrix $\mathbf{A}$, dictionary pair $\hat{\mathbf{D}}$ and Ω
2: **while** not converged **do**
3: **for** $m = 1, 2, \ldots, k$ **do**
4: update $\hat{\mathbf{d}}_m$ via (12)
5: update $\boldsymbol{\omega}_m$ via (13)
6: $\hat{\mathbf{d}}_m := \hat{\mathbf{d}}_m / \|\hat{\mathbf{d}}_m\|_2$
7: $\boldsymbol{\omega}_m := \boldsymbol{\omega}_m \cdot \|\hat{\mathbf{d}}_m\|_2$
8: $\hat{\mathbf{D}}(:, m) := \hat{\mathbf{d}}_m$
9: $\Omega(m, :) := \boldsymbol{\omega}_m^T$
10: **end for**
11: **end while**

Output: dictionary pair $\hat{\mathbf{D}}$ and Ω

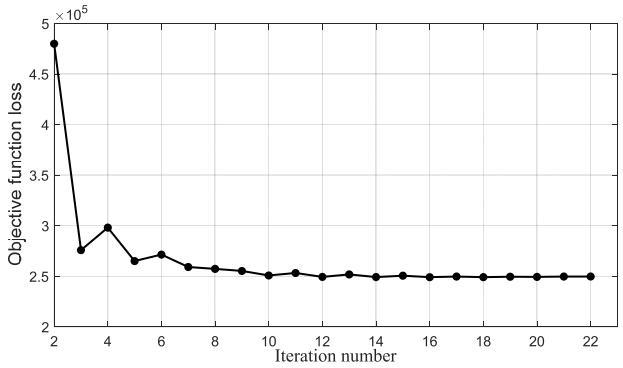


FIGURE 2. Convergence of the solving algorithm for image classification on the CIFAR-10 database.

When the training process is finished, we can learn $\hat{\mathbf{D}}$ and Ω. Since the synthesis dictionary $\mathbf{D}$ and projection matrix $\mathbf{P}$ are jointly normalized in the learning process, we should renormalize $\mathbf{D}$ and $\mathbf{P}$, because we require that $\|\mathbf{d}_i\| = 1$. Normalization can be performed via

$$\begin{aligned} \mathbf{d}_m &= \hat{\mathbf{d}}_m(1:k)./\|\hat{\mathbf{d}}_m(1:k)\|_2 \\ \mathbf{p}_m &= \hat{\mathbf{d}}_m(k+1:end)./\|\hat{\mathbf{d}}_m(1:k)\|_2 \end{aligned} \tag{14}$$

In classification, we adopt a simple linear classifier $\hat{\mathbf{W}} = \mathbf{Y}_{tr}\mathbf{Z}_{tr}^T(\mathbf{Z}_{tr}\mathbf{Z}_{tr}^T + \varsigma\mathbf{I})^{-1}$, where $\mathbf{Z}_{tr}$ and $\mathbf{Y}_{tr}$ are the coefficient matrix and label matrix of training data, respectively. ς is the weight of the regularization term (empirically set as $1e^{-4}$) and $\mathbf{I}$ is an identity matrix.

Given a test sample $\mathbf{x}_t$, the representations can be effectively calculated by $\mathbf{z}_t = S_{\boldsymbol{\lambda}}(\Omega\mathbf{x}_t)$. Then, the label for test sample $\mathbf{x}_t$ is given by $l = \arg\max_l(\mathbf{y}_l = \hat{\mathbf{W}}\mathbf{z}_t)$, where $\mathbf{y}_l$ is the class label vector of $\mathbf{x}_t$.

V. EXPERIMENTAL RESULTS

We perform experiments on image classification, face recognition and gender classification to evaluate the proposed GDASDL model. We compare GDASDL with several of the latest dictionary learning algorithms, such as label-consistent K-SVD (LC-KSVD) [4], Fisher discrimination dictionary learning (FDDL) [6], PDPL [18], structured ADL (SADL) [17], analysis-SDL with a support vector machine (SVM) as a classifier (ASDL-SVM) [10], and DASDL [21]. In all competing methods, LC-KSVD and FDDL belong to SDL, SADL is an ADL method, and the other three methods belong to ASDL. In experiments, the number of synthesis dictionary atoms for all competing methods is set as 20, 102 and 257 on the CIFAR-10 [30], Caltech-101 [31] and Caltech-256 [32] datasets, respectively. The dictionary size is set to 100 for face recognition and gender classification. To obtain reliable results, we repeat each experiment 5 times with different random selection of the training and testing images, and present the mean accuracy on each dataset.

FIGURE 3. Samples from the CIFAR-10 dataset.

A. IMAGE CLASSIFICATION

CIFAR-10: The CIFAR-10 dataset contains a total of 60,000 32×32 color images in 10 classes (i.e., airplance, automobile, bird, cat, deer, dog, frog, horse, ship and truck), with 6,000 images per class. Some examples are listed in Fig.3. There are 50,000 training images and 10,000 testing images. The training images are divided into five training batches, each with 10,000 images. Since the training batches contain the images in random order, some training batches may contain more images from one class than another. Different from the training batches, the test batch contains exactly 1,000 randomly selected images for each class. In each experiment, we randomly choose 10,000 images in all five training batches for training, with 1,000 samples per class, and the test batch for testing.

Caltech-101: The Caltech-101 dataset contains 9,144 images from 102 classes (i.e., 101 object classes and a "background" class). The samples from each category have significant shape variability, and the number of images in

FIGURE 4. Samples from the Caltech-101 dataset.

FIGURE 5. Samples from the Caltech-256 dataset.

TABLE 1. Image classification accuracy (%) and training/testing time (s) on the CIFAR-10, Caltech-101, and Caltech-256 datasets.

Method	CIFAR-10	Caltech-101	Caltech-256	Training time	Testing time
LC-KSVD [4]	24.5	65.7	24.5	10.7	4e-5
FDDL [6]	60.1	75.2	34.1	854.4	1e-1
PDPL [18]	34.0	74.8	34.0	2.3	2e-5
SADL [17]	60.7	73.6	33.7	357.4	1.3e-4
ASDL-SVM [10]	33.8	71.3	33.8	1205.0	4e-4
DASDL [21]	61.8	75.5	34.2	551.1	1.1e-4
GDASDL	66.5	76.1	34.6	709.5	1.0e-4

each category varies from 31 to 800. Fig. 4 provides some examples from the Caltech-101 dataset. Following the popular experimental settings, we randomly select 3,060 images for training, with 30 images per category, and the remainder 6,084 images for testing.

Caltech-256: The Caltech-256 dataset contains more than 30,000 images of 256 classes, and each class contains 80 to 827 images. Compared to the Caltech-101 dataset, it is more challenging due to the large categories and variations in object shape and size. In each experiment, 30 images per class are randomly selected for training and the rest for testing. Some examples are shown in Fig.5.

The feature descriptor of each image used in the CIFAR-10 and Caltech-101 databases is extracted by LSAE [33]. For the Caltech-256 dataset, the feature descriptors are extracted as [34], where the scale-invariant feature transform (SIFT) descriptors [35] from 16×16 patches are densely extracted with a step-size of 6 pixels, followed by learning a codebook with 1,024 atoms using the standard k-means clustering. Then, based on the SIFT features and codebook, the spatial pyramid matching (SPM) feature [36] for each image is extracted with three grids of size 1×1, 2×2 and 4×4. Since the dimension of feature vectors is large, the image features on the three datasets are reduced to 500 dimensions by PCA. Parameters α, β and λ are set as 100, 10, and 0.001, respectively. K is set as 5. The classification accuracy is reported in Table 1.

	airplane	automobile	bird	cat	deer	dog	frog	horse	ship	truck
airplane	67.4	4.2	5.6	1.8	1.6	0.4	1.4	1.4	11.0	5.2
automobile	1.1	84.2	1.0	0.2	0.4	0.5	1.3	0.5	1.7	9.1
bird	7.7	2.0	46.5	5.1	7.5	7.3	12.5	5.8	2.9	2.7
cat	3.5	3.2	6.6	38.7	5.1	19.5	13.1	3.3	2.1	4.9
deer	3.6	1.5	4.4	4.4	52.6	3.2	13.5	12.5	2.5	1.8
dog	2.4	2.6	6.5	10.5	4.4	60.8	4.7	4.1	1.5	2.5
frog	1.0	0.6	4.1	2.2	2.6	2.8	85.0	0.5	0.8	0.4
horse	1.8	1.6	4.1	3.0	4.1	9.7	2.7	69.3	1.3	2.4
ship	5.8	4.9	1.6	1.6	0.6	0.8	1.1	0.6	78.4	4.6
truck	2.5	8.4	1.1	0.6	0.5	1.0	1.3	0.7	3.3	80.6

FIGURE 6. Confusion matrix on the CIFAR-10 dataset.

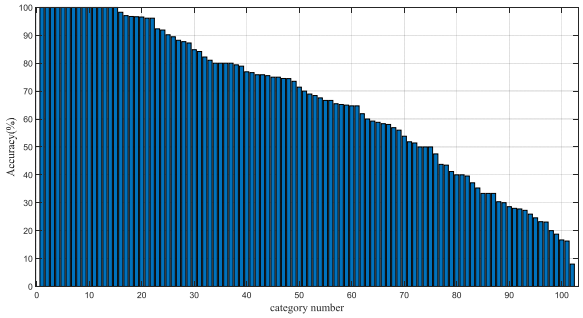


FIGURE 7. Classification accuracy of each category in the Caltech-101 dataset.

From Table 1, we can see that the proposed GDASDL achieves better performance than all other methods. The advantages of using the proposed method, compared to LC-KSVD, on the three datasets are highlighted. For ADL, the accuracy of GDASDL is 5.8%, 2.5% and 0.9% higher than that of SADL on the three datasets. Since GDASDL considers both graph-structure information and discriminative information in dictionary pair learning, it outperforms DASDL by 4.7%, 0.6% and 0.4% on the three datasets. The confusion matrix for GDASDL in the CIFAR-10 dataset is shown in Fig.6. We can observe that the accuracies of cat and bird are lower than those of others since the object is small and the background is relatively complex. Fig.7 shows the classification accuracy of each category in the Caltech-101 dataset. It can be seen that there are 15 categories with 100% accuracy.

The time-consumption of each method on the CIFAR-10 database, including training time and testing times, is shown in Table 1. Note that the time is obtained under the MATLAB 2014a programming environment and a CPU server of 16 CPUs with 1.3 GHz and 16GB RAM. It can be seen that GDASDL is more efficient than FDDL and ASDL-SVM when training but less efficient than other methods. This is mainly due to numerous amounts of matrix inversion computation in the proposed Alg.1. For testing, GDASDL is very effective. Although GDASDL is slower than LC-SVD and PDPL, the classification accuracy of GDASDL is more than 32.5% higher than that of LC-KSVD and PDPL on the CIFAR-10 dataset.

B. FACE RECOGNITION

The AR face database [37] contains more than 4,000 face images of 126 people. The images include different

(a) (b)

FIGURE 8. (a) Some samples of the AR database. (b) Visualization of $PS_\lambda(\Omega \mathbf{X})$ for 100 samples on the AR database.

TABLE 2. Face recognition accuracy (%) on the AR database.

Method	Accuracy	Method	Accuracy
LC-KSVD [4]	73.8	ASDL-SVM [10]	91.8
FDDL [6]	68.3	DASDL [21]	96.3
PDPL [18]	95.2	GDASDL	98.5
SADL [17]	95.1		

illuminations, different expressions and different disguise(sunglasses and scarves), as shown in Fig.8(a). Each person has 26 image images in two sessions. Following the popular experimental settings, a subset of the dataset consisting of 2,600 images from 50 females and 50 males are used. For each person, 20 images are randomly selected for training, and the other 6 images are used for testing. The 540-d features provided by [4] are used as the image features. Parameters α, β and λ are set as 100, 10, and 0.005, respectively.

Table 2 lists the recognition accuracy of all competing methods. From Table 2, it can be seen that GDASDL achieves the highest accuracy of all methods. Compared to SDL and ASL approaches (i.e. LC-KSVD, FDDL and SADL), the accuracy of GDASDL is at least 3.4% higher. Compared to other ASDL approaches (i.e., PDPL, ASDL-SVM and DASDL), the accuracy is improved by over 2.2%.

C. GENDER CLASSIFICATION

In this experiment, we select a nonoccluded subset of the AR database consisting of 50 males and 50 females, with 14 images per person, to conduct experiments of gender classification. We train on the images of first 25 males and 25 females, and test on the images of the remaining 25 males and 25 females. The image is resized from 165×120 to 60×43, and rearranged into a feature vector. Then, PCA is adopted to reduce the dimensionality of the feature vector to 100. Parameters α, β and λ are set as 10, 1, and 0.0005, respectively.

The classification performances are summarized in Table 3. The proposed GDASDL is better than all other methods, with at least 0.8% improvement. Specifically, the accuracy of GDASDL is 2.7% higher than that of LC-KSVD, FDDL and PDPL. Compared to DASDL and ASDL-SVM, GDASDL jointly incorporates graph-structure information and discriminative information into dictionary pair learning, which could enhance the discriminative ability of the representations. We also visualized $\mathbf{P}S_\lambda(\Omega\mathbf{X})$ for 100 samples, as shown in Fig. 8(b). It can be seen that the transformed representation coefficients has a class-specific block-diagonal structure.

TABLE 3. Gender classification accuracy (%) on the AR database.

Method	Accuracy	Method	Accuracy
LC-KSVD [4]	84.4	ASDL-SVM [10]	88.1
FDDL [6]	88.4	DASDL [21]	90.3
PDPL [18]	86.0	GDASDL	91.1
SADL [17]	89.7		

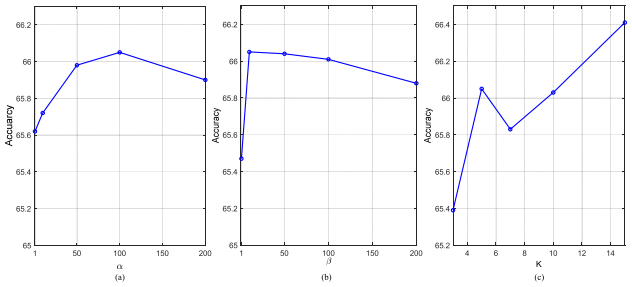


FIGURE 9. Classification accuracy of GDASDL with different parameters in the CIFAR-10 dataset. (a) α, (b) β, (c) K.

D. DISCUSSION OF PARAMETERS

In this subsection, we verify the effects of parameters (i.e., α, β and K) on the performance of the proposed GDASDL on the CIFAR-10 dataset. In the experiments, we change only one parameter at a time while fixing the others. The classification accuracy of GDASDL with different parameter values is shown in Fig. 9.

From Fig. 9(a), it can be seen that the proposed GDASDL achieves increasing classification accuracy as the value of α increase from 1 to 100. When the value of α is larger than 100 (i.e., 200), the performance of GDASDL worsens. From Fig. 9(b), we can see that the best performance is achieved when parameter $\beta = 10$. For parameter K, a higher the value of K does not indicate higher accuracy. From Fig. 9(c), we can see that, when $K = 15$, the classification accuracy is highest; however, the time consumption is also biggest. Considering the time consumption of the proposed model, we set $K = 5$ in all experiments.

E. DISCUSSION ON GDASDL WITH DEEP LEARNING METHODS

Deep learning has been widely studied in the past ten years and achieved excellent performance in various computer vision tasks [38]–[40]. Lu *et al.* [41] proposed a light field imaging approach for solving underwater imaging problems using deep convolutional neural networks with depth estimation. Cen and Wang [42] proposed a deep dictionary representation-based classification scheme for occluded faces. Tariyal *et al.* [43] analyzed the relevance between dictionary learning and deep learning, and proposed deep dictionary learning.

The proposed method will feasibly perform better in image classification using deep features, because the deep features have stronger representative and discriminative ability than the hand-crafted features. In fact, the features of image can be extracted through deep learning, followed by dictionary pair learning via (7). In this work, we mainly focus on the

design of the structure constraint and discriminative dictionary learning. We will exploit a framework that combine the GDASDL model with deep features in the future.

VI. CONCLUSION

This paper has presented a novel GDASDL model based on the local geometric structures and label information of the data for image classification. The locality is used as a graph-regularization term to restrain the representations. Meanwhile, label information is also used as a discrimination constraint to enhance the discriminative ability of the learned dictionary pair. By embedding the two terms into the objective function, the proposed model can achieve superior performance using a simple linear classifier. We also presented an iterative algorithm for solving the proposed model. The experimental results demonstrate that the proposed GDASDL model yields good classification results on four well-known public datasets, and outperforms the six state-of-the-art dictionary learning methods, i.e., LC-KSVD, FDDL, PDPL, SADL, ASDL and DASDL.

In the future, we will design a more efficient algorithm for solving GDASDL and apply this method for other problems, such as speech processing and biometrics. We also plan to extend ASDL to multilayer dictionary pair learning by combining dictionary learning and deep learning to capture high-level image features on small sample data set.

REFERENCES

[1] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[2] S. Serikawa and H. Lu, "Underwater image dehazing using joint trilateral filter," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 41–50, Jan. 2014.

[3] F. Alonso-Fernandez, R. A. Farrugia, J. Bigun, J. Fierrez, and E. Gonzalez-Sosa, "A survey of super-resolution in iris biometrics with evaluation of dictionary-learning," *IEEE Access*, vol. 7, pp. 6519–6544, Dec. 2018.

[4] Z. 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.

[5] J. Yang, D. Chu, L. Zhang, Y. Xu, and J. Yang, "Sparse representation classifier steered discriminative projection with applications to face recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 7, pp. 1023–1035, Jul. 2013.

[6] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse representation based Fisher discrimination dictionary learning for image classification," *Int. J. Comput. Vis.*, vol. 109, no. 3, pp. 209–232, Sep. 2014.

[7] J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, and F. R. Bach, "Supervised dictionary learning," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, Dec. 2008, pp. 1033–1040.

[8] H. Chang, M. Yang, and J. Yang, "Learning a structure adaptive dictionary for sparse representation based classification," *Neurocomputing*, vol. 190, pp. 124–131, May 2016.

[9] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Appl. Comput. Harmon. Anal.*, vol. 34, no. 1, pp. 30–56, Jan. 2013.

[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] Y. Yuan, D. Ma, and Q. Wang, "hyperspectral anomaly detection via sparse dictionary learning method of capped norm," *IEEE Access*, vol. 7, pp. 16132–16144, Jan. 2019.

[12] H. Zheng, H. Chang, and F. Zhang, "Structured discriminative dictionary learning based on schatten-p norm low-rank representation" in *Proc. 4th IAPR Asian Conf. Pattern Recognit.*, Nanjing, China, Nov. 2017, pp. 688–693.

[13] S. Kong and D. Wang, "A dictionary learning approach for classification: Separating the particularity and the commonality," in *Proc. Eur. Conf. Comput. Vis.*, Florence, Italy, vol. 7572, Oct. 2012, pp. 186–199.

[14] 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.

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

[16] J. Guo, Y. Guo, X. Kong, M. Zhang, and R. He, "Discriminative analysis dictionary learning," in *Proc. 30th AAAI Conf. Artif. Intell.*, Phoenix, AZ, USA, Feb. 2016, pp. 1617–1623.

[17] W. Tang, A. Panahi, H. Krim, and L. Dai, "Structured analysis dictionary learning for image classification," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Calgary, AB, Canada, Apr. 2018, pp. 2181–2185.

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

[19] S. Gu, D. Meng, W. Zuo, and L. Zhang, "Joint convolutional analysis and synthesis sparse representation for single image layer separation," in *Proc. IEEE Int. Conf. Comput. Vis.,* Venice, Italy, Oct. 2017, pp. 1717–1725.

[20] M. Yang, H. Chang, W. Luo, and J. Yang, "Fisher discrimination dictionary pair learning for image classification," *Neurocomputing*, vol. 269, pp. 13–20, Dec. 2017.

[21] M. Yang, H. Chang, and W. Luo, "Discriminative analysis-synthesis dictionary learning for image classification," *Neurocomputing*, vol. 219, pp. 404–411, Jan. 2017.

[22] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2010.

[23] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[24] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacian faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[25] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[26] B.-D. Liu, B. Shen, and X. Li, "Locality sensitive dictionary learning for image classification," in *Proc. Int. Conf. Image Process.*, Quebec City, QC, Canada, Sep. 2015, pp. 3807–3811.

[27] Z. Li, Z. Lai, Y. Xu, J. Yang, and D. Zhang, "A locality-constrained and label embedding dictionary learning algorithm for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 2, pp. 278–293, Feb. 2017.

[28] J. Jiang, Y. Yu, Z. Wang, X. Liu, and J. Ma, "Graph-regularized locality-constrained joint dictionary and residual learning for face sketch synthesis," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 628–641, Feb. 2019.

[29] M. Zheng *et al.*, "Graph regularized sparse coding for image representation," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1327–1336, May 2011.

[30] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Univ. Toronto, Toronto, ON, USA, Tech. Rep. 3, Jan. 2009.

[31] L. Fei-Fei, 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 Understand.*, vol. 106, no. 1, pp. 59–70, Apr. 2007.

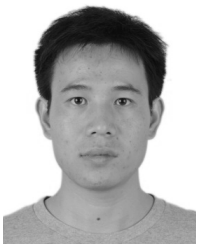
[32] G. Griffin, A. Holub, and P. Perona, "Caltech-256 Object category dataset," California Inst. Technol., Pasadena, CA, USA, Tech. Rep. CNS-TR-2007-001, Apr. 2007.

[33] W. Luo, J. Yang, W. Xu, and T. Fu, "Locality-constrained sparse auto-encoder for image classification," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1070–1073, Aug. 2015.

[34] 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.*, Miami, FL, USA, Jun. 2009, pp. 1794–1801.

[35] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, Kerkyra, Greece, Sep. 1999, pp. 1150–1157.

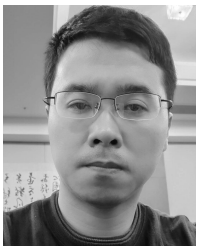
[36] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, Jun. 2006, pp. 2169–2178.
[37] A. M. Martínez and R. Benavente, "The AR face database," CVC Tech. Rep. #24, Jun. 1998. [Online]. Available: http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html
[38] H. Lu *et al.*, "CONet: A cognitive ocean network," *IEEE Wireless Commun.*, to be published.
[39] H. Lu, Y. Li, M. Chen, H. Kim, and S. Serikawa, "Brain intelligence: Go beyond artificial intelligence," *Mobile Netw. Appl.*, vol. 23, no. 2, pp. 368–375, Apr. 2018.
[40] W. Yang *et al.*, "Improving low-dose CT image using residual convolutional network," *IEEE Access*, vol. 5, pp. 24698–24705, 2017.
[41] H. Lu, Y. Li, T. Uemura, H. Kim, and S. Serikawa, "Low illumination underwater light field images reconstruction using deep convolutional neural networks," *Future Gener. Comput. Syst.*, vol. 82, pp. 142–148, May 2018.
[42] F. Cen and G. Wang, "dictionary representation of deep features for occlusion-robust face recognition," *IEEE Access*, vol. 7, pp. 26595–26605, Feb. 2019.
[43] S. Tariyal, A. Majumdar, R. Singh, and M. Vatsa, "Deep dictionary learning," *IEEE Access*, vol. 4, pp. 10096–10109, Dec. 2016.



HEYOU CHANG received the B.S. degree in computer science and technology from the Henan University of Science and Technology, Luoyang, China, in 2011, and the Ph.D. degree from Nanjing University of Science and Technology (NUST), Nanjing, China, in 2017. He is currently a Postdoctoral Fellow of Southeast University. He is also an Assistant Professor with Nanjing Xiaozhuang University, Nanjing. His research interests include sparse representation, machine learning, and medical image processing.

HUI TANG received the B.S. and M.S. degrees in biomedical engineering from Southeast University, Nanjing, China, in 2003 and 2005, respectively, the Ph.D. degree in biomedical engineering from Southeast University, and the Ph.D. degree in signal processing and telecommunication from the University of Rennes 1, France, in 2008. She is currently an Associate Professor with the Library of Image Science and Technology (LIST), School of Computer Science and Engineering, Southeast University. Her current research interests include image processing, image analysis, and scientific visualization applied in medical areas.

FANLONG ZHANG received the B.S. and M.S. degrees in applied mathematics from Liaocheng University, in 2007 and 2010, respectively, and the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology (NUST), in 2015. He is currently an Assistant Professor with the School of Technology, Nanjing Audit University (NAU). His current research interests include pattern recognition and optimization.



YANG CHEN received the M.S. and Ph.D. degrees in biomedical engineering from First Military Medical University, Guangzhou, China, in 2004 and 2007, respectively. He is currently a Professor with the Department of Computer Science and Engineering, Southeast University, China. His recent work concentrates on the medical image reconstruction, image analysis, pattern recognition, and computerized-aid diagnosis.

HAO ZHENG received the B.S. degree from Southeast University, in 1998, the M.S. degree from the Nanjing University Posts and Telecommunications, in 2005, and the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology, in 2013. He is currently a Professor with Nanjing Xiaozhuang University. His research interests include pattern recognition, image processing, face recognition, computer vision.