



Jianqiang Song <sup>1,2</sup>, Lin Wang <sup>1</sup>, Zuozhi Liu <sup>3</sup>, Muhua Liu <sup>2</sup>, Mingchuan Zhang <sup>1</sup>, and Qingtao Wu <sup>1,\*</sup>

- <sup>1</sup> School of Information Engineering, Henan University of Science and Technology, No. 263 Kaiyuan Avenue, Luoyang 471023, China; songjiebang@163.com (J.S.); linwang@haust.edu.cn (L.W.); zhang\_mch@haust.edu.cn (M.Z.)
- <sup>2</sup> Control Science and Engineering Postdoctoral Mobile Station, Henan University of Science and Technology, No. 263 Kaiyuan Avenue, Luoyang 471023, China; lxk0379@126.com
- <sup>3</sup> School of Mathematics and Statistics, Guizhou University of Finance and Economics, Guiyang 550025, China; liuzuo.zhi@163.com
- \* Correspondence: wqt8921@haust.edu.cn

**Abstract:** Dictionary learning has been an important role in the success of data representation. As a complete view of data representation, hybrid dictionary learning (HDL) is still in its infant stage. In previous HDL approaches, the scheme of how to learn an effective hybrid dictionary for image classification has not been well addressed. In this paper, we proposed a locality preserving and label-aware constraint-based hybrid dictionary learning (LPLC-HDL) method, and apply it in image classification effectively. More specifically, the locality information of the data is preserved by using a graph Laplacian matrix based on the shared dictionary for learning the commonality representation, and a label-aware constraint with group regularization is imposed on the coding coefficients corresponding to the class-specific dictionary for learning the particularity representation. Moreover, all the introduced constraints in the proposed LPLC-HDL method are based on the  $l_2$ -norm regularization, which can be solved efficiently via employing an alternative optimization strategy. The extensive experiments on the benchmark image datasets demonstrate that our method is an improvement over previous competing methods on both the hand-crafted and deep features.

Keywords: locality preserving; label-aware constraint; hybrid dictionary learning; image classification

# 1. Introduction

Due to the insufficiency of data representation, Dictionary Learning (DL) has aroused considerable interest in the past decade and achieved much success in the various applications, such as image denoising [1,2], person re-identification [3,4] and vision recognition [5–8]. Generally speaking, the DL methods are developed based on a basic hypothesis, which are that a test signal can be well approximated using the linear combination of some atoms in a dictionary. Thus the dictionary usually plays an important role in the success of these applications. Traditionally, the DL methods can be roughly divided into two categories: The unsupervised DL methods and the supervised DL methods [9,10].

In the unsupervised DL methods, a dictionary is optimized to reconstruct all the training samples without any label assignment; hence, there is no class information in the learned dictionary. By further integrating the label information into dictionary learning, the supervised DL methods can achieve better classification performance than the unsupervised ones for image classification. Supervised dictionary learning encodes the input signals using the learned dictionary, then utilizes the representation coefficients or the residuals for classification. Thus the discriminative ability of the dictionary and the representative ability of the coding coefficients play the key roles in this kind of approach. According to the types of the dictionary, the supervised DL methods can be further divided into three categories: The class-shared DL methods, the class-specific DL methods and the hybrid DL methods [11].



Citation: Song, J.; Wang, L.; Liu, Z.; Liu, M.; Zhang, M.; Wu, Q. Locality Preserving and Label-Aware Constraint-Based Hybrid Dictionary Learning for Image Classification. *Appl. Sci.* 2021, *11*, 7701. https:// doi.org/10.3390/app11167701

Academic Editor: Manuel Graña

Received: 5 July 2021 Accepted: 17 August 2021 Published: 21 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). The class-shared DL methods generally force the coding coefficients to be discriminative via learning a single dictionary shared by all the classes. Based on the K-SVD algorithm, Zhang and Li [5] proposed the Discriminative K-SVD (D-KSVD) method to construct a classification error term for learning a linear classifier. Jiang et al. [6] further proposed the Label Consistent K-SVD (LC-KSVD) which encourages the coding coefficients from the same class to be as similar as possible. Considering the characteristics of atoms, Song et al. [12] designed an indicator function to regularize the class-shared dictionary to improve the discriminative ability of coding coefficients. In general, the computation of the test stage is very efficient in the class-shared DL methods, but it is hard to improve the coefficients' discriminativity for better classification performance as the class-shared dictionary is not enough for fitting the complex data.

In the class-specific dictionary learning, each sub-dictionary is assigned to a single class and the sub-dictionaries with the different classes are encouraged to be as independent as possible. As a representative class-specific DL method, Fisher Discrimination Dictionary Learning (FDDL) [13] employs Fisher discrimination criterion on the coding coefficients, then utilizes the representation residual of each class to establish the discriminative term. Using with the incoherence constraint, Ramirez et al. [14] proposed a structured dictionary learning scheme to promote the discriminative ability of the class-specific sub-dictionaries. Akhtar et al. [15] developed a Joint discriminative Bayesian Dictionary and Classifier learning (JBDC) model to associate the dictionary atoms by the class labels with Bernoulli distributions. The class-specific DL methods usually associate a dictionary atom to a single class directly; hence, the reconstruction error with respect to each class can be used for classification. However, the test stage of this category often requires the coefficient computation of test data over many sub-dictionaries.

In the hybrid dictionary learning, a dictionary is designed to have a set of class-shared atoms in addition to the class-specific sub-dictionaries. Wang and Kong [16] proposed a hybrid dictionary dubbed DL-COPAR to explicitly separate the common and particular features of the data, which also encourages the class-specific sub-dictionaries to be incoherent. Vu et al. [17] developed a Low-Rank Shared Dictionary Learning (LRSDL) method to preserve the common features of samples. Gao et al. [18] developed a Category-specific and Shared Dictionary Learning (CSDL) approach for fine-grained image classification. Wang et al. [19] designed a structured dictionary consisting of label-particular atoms corresponding to some class and shared atoms commonly used by all the classes, and introduced a Cross-Label Suppression for Dictionary Learning (CLSDL) to generate approximate sparse coding vectors for classification. To some extent, the hybrid dictionary is very effective at preserving the complex structure of the visual data. However, it is nontrivial to design the class-specific and shared dictionaries with the proper number of atoms, which often has a severe effect on classification performance.

In addition to utilizing the class label information, more and more supervised DL approaches have been proposed to incorporate the locality information of the data into the learned dictionary. By calculating the distances between the bases (atoms) and the training samples, Wang et al. [20] developed a Locality-constrained Linear Coding (LLC) model to select the k-nearest neighbor bases for coding, and set the coding coefficients of other atoms to zero. Wei et al. [21] proposed locality-sensitive dictionary learning to enhance the power of discrimination for sparse coding. Song et al. [22] integrated the locality constraints into the multi-layer discriminative dictionary to avoid the appearance of over-fitting. By coupling the locality reconstruction and the label reconstruction, the LCLE-DL method [7] ensures that the locality-based and label-based coding coefficients are as approximate to each other as possible. It is noted that the locality constraint in LCLE-DL may cause the dictionary atoms from the different classes to be similar, which weakens the discriminative ability of the learned dictionary.

It is observed that the real-world object categories are not only a marked difference, but are also strongly correlated in terms of the visual property, e.g., faces from the different persons often share similar illumination and pose variants; objects in the Caltech 101 dataset [23] have the correlated background. These correlations are not very helpful to distinguish the different categories, but without them the data with common features cannot be well represented. Thus, the dictionary learning approach should learn the distinctive features with the class-specific dictionary, and simultaneously exploit the common features of the correlated classes by learning a commonality dictionary. To this end, we proposed the locality preserving and label-aware constraint-based hybrid dictionary learning (LPLC-HDL) method for image classification, which is composed of a label-aware constraint, a group regularization and a locality constraint. The main contributions are summarized as follows.

- [1]. The proposed LPLC-HDL method learns the hybrid dictionary by fully exploiting the locality information and the label information of the data. In this way, the learned hybrid dictionary can not only preserve the complex structural information of the data, but also have strong discriminativity for image classification.
- [2]. In LPLC-HDL, a locality constraint is constructed to encourage the samples from different classes with similar features to have similar commonality representation; then, a label-aware constraint is integrated to make the class-specific dictionary sparsely represent the samples from the same class, so that the robust particularity– commonality representation can be obtained by the proposed LPLC-HDL.
- [3]. In a departure from the competing methods which impose the  $l_0$ -norm or  $l_1$ -norm on the coefficients, LPLC-HDL consists of  $l_2$ -norm constraints that can be calculated efficiently. The objective function is solved elegantly by employing an alternative optimization technique.

The rest of this paper is outlined as follows. Section 2 reviews the related work on our LPLC-HDL method. Then Section 3 presents the details of LPLC-HDL and an effective optimization is introduced in Section 4. To verify the efficiency of our method for image classification, the experiments are conducted in Section 5. Finally, the conclusion is summarized in Section 6.

### 2. Notation and Background

In this section, we first provide the notation used in this paper, then review the LCLE-DL algorithm and the objective function of hybrid dictionary learning (DL), which can be taken as the theoretical background of our LPLC-HDL method.

## 2.1. Notation

Let  $X = [X_1, \dots, X_C] \in \mathbb{R}^{m \times N}$  be a set of N training samples in an m dimension with class labels  $Y_i \in [1, \dots, C]$ ; here, C is the class number of the training samples and  $X_i$  is a matrix consisting of  $N_i$  training samples of the *i*th class. Suppose  $D_0 \in \mathbb{R}^{m \times k_0}$  and  $D_p = [D_1, \dots, D_C] \in \mathbb{R}^{m \times K_p}$  make up the learned hybrid dictionary  $D = [D_0, D_p]$  from the training samples X, where  $k_0$  is the atom number of shared dictionary,  $K_p = \sum_{i=1}^{C} k_i$  denotes the atom number of class-specific dictionary and  $k_i$  is the atom number of *i*th class sub-dictionary. Let  $Z = [Z_1, \dots, Z_C] \in \mathbb{R}^{(k_0 + K_p) \times N}$  be the coding coefficients of training samples X over the hybrid dictionary D; then,  $Z_0 = [Z_0^1, \dots, Z_0^C] \in \mathbb{R}^{k_0 \times N}$  and  $Z_i = [Z_i^1, \dots, Z_i^C] \in \mathbb{R}^{k_i \times N}$  represent the coding coefficients over the shared dictionary and the *i*th class sub-dictionary, respectively.

According to [24], a row vector of coefficient matrix *Z* can be defined as a profile of the corresponding dictionary atom. Therefore, we can define a vector  $z^r = [z_1^r; z_2^r; \cdots; z_C^r] \in \mathbb{R}^{N \times 1}$  ( $r = 1, \cdots, K$ ) as the profile of atom  $d_r$  for all the training samples, where the subvector  $z_c^r = [z_1^r, z_2^r, \cdots, z_{N_c}^r]^T \in \mathbb{R}^{N_c \times 1}$  is the sub-profile for the training samples of the *c*th class.

## 2.2. The LCLE-DL Algorithm

To improve the classification performance, the LCLE-DL algorithm [7] takes both the locality and label information of dictionary atoms into account in the learning process. This algorithm firstly constructs the locality constraint to ensure that similar profiles have

similar atoms, then establishes the label embedding constraint to encourage the atoms of the same class to have similar profiles. The objective function of LCLE-DL is defined as follows.

$$\min_{D_p, Z_p, V_p, L} \|X - D_p Z_p\|_2^2 + \alpha Tr(Z_p^T L Z_p) + \|X - D_p V_p\|_2^2 + \beta Tr(V_p^T U V_p) 
+ \gamma \|Z_p - V_p\|_2^2, \quad s.t. \|d_j\|^2 = 1, \ j = 1, \cdots, K$$
(1)

where  $Z_p \in \mathbb{R}^{K_p \times N}$  and  $V_p \in \mathbb{R}^{K_p \times N}$  denote the locality-based and the label-based coding coefficients,  $L \in \mathbb{R}^{K_p \times K_p}$  is the graph Laplacian matrix that is constructed by the atom's similarity in the dictionary  $D_p \in \mathbb{R}^{m \times K_p}$ ,  $U \in \mathbb{R}^{K_p \times K_p}$  is the scaled label matrix which is constructed using the label matrix of the dictionary  $D_p$ .  $||X - D_p Z_p||_2^2$  combined with the second term encodes the reconstruction under the locality constraint;  $||X - D_p V_p||_2^2$  combined with fourth term encodes the reconstruction under the label embedding;  $||Z_p - V_p||_2^2$  is used to transfer the label constraint to the locality constraint.  $\alpha$ ,  $\beta$  and  $\gamma$  are the regularization parameters; the constraint on the atoms can avoid the scaling issue.

The LCLE-DL algorithm first exploits the K-SVD algorithm to learn sub-dictionaries  $D_i(i = 1, \dots, C)$  using with the training samples  $X_i$ . Similar to the label matrix of the training samples Y, the label matrix of the dictionary  $D_p$  can be obtained as  $B = [b_1, \dots, b_{K_p}] \in \mathbb{R}^{K_p \times C}$ . Then a weighted label matrix is constructed by  $G = B(B^TB)^{-\frac{1}{2}} \in \mathbb{R}^{K_p \times C}$ . Next, the label embedding of atoms is defined as

$$\min_{V_p} Tr(V_p^{\mathrm{T}} G G^{\mathrm{T}} V_p) = \min_{V_p} Tr(V_p^{\mathrm{T}} U V_p),$$
(2)

where  $U = GG^{T} \in \mathbf{R}^{K_{p} \times K_{p}}$  is the scaled label matrix of the dictionary  $D_{p}$ ; the above terms make the coding coefficients  $V_{p}$  have a block-diagonal structure with strong discriminative information.

The learned dictionary inherits the manifold structure of the data via using the derivedgraph Laplacian matrix *L*, and the optimal representation of the samples can be obtained with the label embedding of dictionary atoms. By combining the double reconstructions, LCLE-DL ensures the label-based and the locality-based coding coefficients are as approximate to each other as possible. However, the locality constraint is imposed on the class-specific dictionary in the LCLE-DL algorithm, which may cause the dictionary atoms from the different classes to be similar, thus the discriminative ability of the dictionary is weakened.

## 2.3. The Objective Function of Hybrid DL

In recent years, the hybrid DL [16,19,25,26] has been getting more and more attention in the classification problem. The hybrid dictionary has been shown to perform better than the other types of dictionaries, as it can preserve both the class-specific and common information of the data. To learn such a dictionary, we can define the objective function of Hybrid DL as follows.

$$\min_{D,Z} \sum_{i=1}^{C} \|X_i - D_0 Z_0^i - D_i Z_i^i\|_F^2 + \lambda_1 \psi(D) + \lambda_2 \phi(Z)$$

$$s.t. \|d_j\|^2 = 1, \ j = 1, \cdots, k_0, \cdots, K$$
(3)

where  $D = [D_0, D_1, \dots, D_C] \in \mathbb{R}^{m \times K}$ ,  $D_0$  is the dictionary shared by all the classes,  $D_i$   $(i = 1, \dots, C)$  is the *i*th class sub-dictionary,  $Z_0^i$  and  $Z_i^i$  are the coding coefficients of the samples from the *i*th class over  $D_0$  and  $D_i$ , and  $\psi(D)$  and  $\phi(Z)$  denote the functions about the hybrid dictionary and the coding coefficients, respectively. The constraint of  $\phi(Z)$  typically adopts  $l_1$  [26] norm or  $l_{2,1}$  [25,27] norm for sparse coding.

### 3. The Proposed Method

By learning the shared dictionary  $D_0$ , the previous hybrid DL algorithms can capture the common features of the data, but they do not concern the correlation among these features, which will reduce the robustness of the learned dictionary. In this section, we firstly utilize the locality information of the atoms in  $D_0$  to construct a locality constraint, then impose it on the coefficients  $Z_0$ , so that the correlation of the common features is captured explicitly and the learned dictionary  $D_0$  is very robust concerning commonality representation. Moreover, once the correlation is discarded, the classification of a query sample will be dominated by the class-specific sub-dictionary corresponding to the correct class to reach minimized data fidelity.

To obtain the discriminative ability of the class-specific dictionary, we further introduce a label-aware constraint as well as group regularization on the distinctive coding coefficients for the particularity representation. Since this constraint integrated with the locality constraint to reconstruct the input data, they will reinforce each other in the learning process, which results in a discriminative hybrid dictionary for image classification.

Accordingly, the objective function of the proposed LPLC-HDL method can be formulated as follows.

$$\min_{D_{0},L_{0},Z_{0},D_{p},Z_{p}} \sum_{i=1}^{C} \left\{ \|X_{i} - D_{0}Z_{0}^{i} - D_{p}Z_{p}^{i}\|_{F}^{2} + \lambda \|P_{i}Z_{p}^{i}\|_{F}^{2} \right\} + \gamma \left( \|Z_{p}\|_{F}^{2} + Tr(Z_{p}L_{p}Z_{p}^{T}) \right) + \eta \left( \|Z_{0}\|_{F}^{2} + Tr(Z_{0}^{T}L_{0}Z_{0}) \right),$$

$$s.t. \|d_{0}^{k}\|_{2}^{2} = 1, \ (k = 1, 2, \cdots, k_{0}); \ \|d_{p}^{l}\|_{2}^{2} = 1, \ (l = 1, 2, \cdots, K_{p}),$$

$$(4)$$

where  $\lambda$ ,  $\gamma$  and  $\eta$  are the regularization parameters, which can adjust the weights of the label-aware constraint, the group regularization and the the locality constraint, respectively. Here we set the Euclidean length of all the atoms in the shared dictionary  $D_0$  and the class-specific dictionary  $D_p$  to be 1, which can avoid the scaling issue.

### 3.1. The Locality Constraint for Commonality Representation

Locality information of the data has played an important role in many real applications. By incorporating locality information for learning a dictionary, we can ensure that the samples with common features tend to have similar coding coefficients [7]. Further more, the dictionary atoms measure the similarity of the samples, which are more robust to the noise and outliers than the original samples. Hence, we use the atoms of the shared dictionary to capture the correlation among the common features and construct a locality constraint.

Based on the shared dictionary  $D_0 \in \mathbb{R}^{m \times k_0}$ , we can construct a nearest neighbor graph  $M_0 \in \mathbb{R}^{k_0 \times k_0}$  as follows.

$$M_0^{r,s} = \begin{cases} \exp(-\frac{\|d_r - d_s\|_2}{\delta}), & \text{if } d_r \in kNN(d_s); \\ 0, & \text{else,} \end{cases}$$
(5)

where  $\delta$  is a parameter to control the exponential function,  $kNN(d_s)$  denotes k nearest neighbors of atom  $d_s$ ,  $M_0^{r,s}$  indicates the similarity between the atoms  $d_r$  and  $d_s$ . For convenience of calculation, we invariably set the parameters  $\delta = 4$  and k = 1, as they have the stable values in the experiments.

Once  $M_0$  is calculated, we construct a graph Laplacian matrix  $L_0$  as follows.

$$L_0 = G_0 - M_0, \ G_0 = \operatorname{diag}(g_0^1, \dots, g_0^{k_0}), \text{ and } g_0^r = \sum_{s=1}^{k_0} M_0^{r,s}.$$
 (6)

Since  $L_0$  is constructed based on the dictionary  $D_0$ , it will be updated in coordination with  $D_0$  in the learning process.

By now, we can obtain a locality constraint term for choosing graph Laplacian matrix  $L_0$  as follows.

$$\frac{1}{2}\sum_{r=1}^{k_0}\sum_{s=1}^{k_0} (z^r - z^s)^2 M_0^{r,s} = Tr(Z_0^T L_0 Z_0) = \sum_{i=1}^C Tr(Z_0^{i T} L_0 Z_0^i).$$
(7)

Because the profile  $z^r/z^s$  and the atom  $d^r/d^s$  have a one-to-one correspondence, the above equation ensures that similar atoms encourage similar profiles [7]. Hence, the correlated information of the common features can be inherited by the coefficient matrix  $Z_0$  and the graph Laplacian matrix  $L_0$ .

### 3.2. The Constraints for Particularity Representation

### 3.2.1. The Label-Aware Constraint

To obtain the particularity representation for the classification, we will assign the labels to the atoms of the class-specific dictionary, as presented in [7,19]. If an atom  $d^t \in D_i$  ( $i = 1, \dots, C$ ), the label i will be assigned to the atom  $d^t$  and kept invariant in the iterations. We take  $S_i$  as the index set for the atoms of ith class sub-dictionary  $D_i$ , and S as the index set for all the atoms of the class-specific dictionary. For the particularity representation of samples  $X_i$ , it is desirable that the large coefficients should mainly occur on the atoms in  $S_i$ . In other words, the sub-profiles associated with the atoms in  $S \setminus S_i$  need to be suppressed to some extent.

For the *i*th class samples, we construct a matrix  $P_i \in \mathbf{R}^{K \times K}$  to pick up the sub-profiles from the representation  $Z_i^T$ , which locate at the atoms in  $S \setminus S_i$  rather than the atoms in  $S_i$  ( $i = 1, \dots, C$ ), so that we can define a label-aware constraint term as follows.

$$|Z_p^{i\,\mathrm{T}}P_i^{\mathrm{T}}||_F^2 = ||P_i Z_p^i||_F^2,\tag{8}$$

and the matrix  $P_i$  can be written as

$$P_i(t_1, t_2) = \begin{cases} 1, & t_1 = t_2 \text{ and } t_2 \in \mathcal{S} \setminus \mathcal{S}_i; \\ 0, & \text{else,} \end{cases}$$
(9)

where  $P_i(t_1, t_2)$  is the  $(t_1, t_2)$ th entry of matrix  $P_i$ . For the particularity representation  $Z_p^i$ , minimizing the label-aware constraint can suppress the large values in the sub-profiles associated with the atoms in  $S \setminus S_i$ , as well as encourage the large ones in the sub-profiles associated with *i*th class atoms. Therefore, it is expected that this constraint with a proper scalar can make the particularity representation approximately sparse. Besides, once a series of matrices  $P_i$  ( $i = 1, \dots, C$ ) are constructed, they will be unchanged in the iterations. Thus it is very efficient for coding over the class-specific dictionary.

## 3.2.2. The Group Regularization

Furthermore, to promote the consistency of particularity representation from the same class, we introduce the group regularization on  $Z_p^i$ . In light of the label information of training samples, assume one sample is related to one vertex; the vertices corresponding to the same class samples are connected and neighboring each other; thus, each class forms a densely connected sub-graph. Considering the training samples  $X_i$  and their coding coefficients  $Z_p^i$ , we first define  $k_i$  graph maps with mapping each graph to a line that consists of  $N_i$  points, as follows

$$f_{k}^{i} = \begin{bmatrix} z_{1}^{i}(k) \\ z_{2}^{i}(k) \\ \vdots \\ z_{N_{i}}^{i}(k) \end{bmatrix} \in \mathbf{R}^{N_{i} \times 1}, \ \forall \ k = 1, 2, \cdots, k_{i}; \ i = 1, 2, \cdots, C.$$
(10)

Here  $z_n^i(k)$   $(n = 1, 2, \dots, N_i)$  is the *k*th component of  $z_n^i$ , which corresponds to the *k*th atom in the *i*th class sub-dictionary  $D_i$ .

Then, we can calculate the variation for these  $k_i$  graph maps as follows.

$$V_{i} = \sum_{k=1}^{k_{i}} f_{k}^{i\,\mathrm{T}} L_{i} f_{k}^{i} = Tr \left( \begin{bmatrix} f_{1}^{i\,\mathrm{I}} \\ \vdots \\ f_{k_{i}}^{i\,\mathrm{T}} \end{bmatrix} L_{i} [f_{1}^{i}, \cdots, f_{k_{i}}^{i}] \right) = Tr(Z_{p}^{i} L_{i} Z_{p}^{i\,\mathrm{T}}), \tag{11}$$

where  $L_i \in \mathbf{R}^{N_i \times N_i}$  denotes the normalized Laplacian of the overall graph for the *i*th class, which can be derived as

$$L_{i} = \frac{1}{N_{i} - 1} \begin{bmatrix} N_{i} - 1 & -1 & \cdots & -1 \\ -1 & N_{i} - 1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots \\ -1 & -1 & \cdots & N_{i} - 1 \end{bmatrix}.$$
 (12)

For the different classes, the vertices related to their samples should not be connected; thus, the graphs of *C* classes are isolated from each other. Therefore, we can obtain the total variation for  $K_p$  graph maps of all the *C* classes as  $\sum_{i=1}^{C} Tr(Z_p^i L_i Z_p^{iT}) = Tr(Z_p L_p Z_p^T)$ , where  $L_p = \text{diag}(L_1, \dots, L_C)$ . Keeping this group regularization small will promote the consistency of the representation for the same class samples. Moreover, by combing it with the label-aware constraint, the coding coefficients for the different classes will be remarkably distinct with the large coefficients locating in the different areas, which is very favorable for the classification task.

### 4. Optimization Strategy and Classification

4.1. Optimization Strategy of LPLC-HDL

As the objective function in Equation (4) is not a jointly convex optimization problem for the variables  $(D_0, L_0, Z_0, D_p, Z_p)$ , it will be divided into two sub-problems by learning the shared and class-specific dictionaries alternatively, that is, updating variables  $(D_0, L_0, Z_0)$  by fixing  $(D_p, Z_p)$ , then updating variables  $(D_p, Z_p)$  by fixing  $(D_0, L_0, Z_0)$ .

We firstly use k-means algorithm to initialize shared dictionary  $D_0 \in \mathbb{R}^{m \times k_0}$  by using all the training samples, then initialize *i*th class sub-dictionary  $D_i \in \mathbb{R}^{m \times k_i}$  by using the training samples  $X_i$  and concatenate all these sub-dictionaries as the class-specific dictionary  $D_p \in \mathbb{R}^{m \times K_p}$ . Next, we compute the initialized coefficients  $Z_0$  and  $Z_p$  with the Multi-Variate Ridge Regression (MVRR) algorithm, and obtain the initialized matrix  $L_0$  by Equation (6). In line with the corresponding sub-dictionaries, the serial of matrices  $P_i$  and  $L_i$  ( $i = 1, 2, \dots, C$ ) are constructed with Equations (9) and (12), respectively. After finishing the initialization, we can optimize the objective function as the following steps.

### 4.1.1. Shared Dictionary Learning

By fixing the variables  $(D_p, Z_p)$ , the objective function of our LPLC-HDL for learning the shared dictionary becomes

$$\min_{D_0, L_0, Z_0} \sum_{i=1}^{C} \|X_i' - D_0 Z_0^i\|_F^2 + \eta \left( \|Z_0\|_F^2 + Tr(Z_0^T L_0 Z_0) \right),$$

$$s.t. \|d_0^k\|_2^2 = 1, \ (k = 1, 2, \cdots, k_0),$$
(13)

where  $X'_i = X_i - D_p Z^i_p$ . To update the variables  $(D_0, L_0, Z_0)$ , we turn to an iterative scheme, i.e., updating  $D_0$  by fixing  $(L_0, Z_0)$ , constructing  $L_0$  based on  $D_0$ ; updating  $Z_0$  by fixing  $(D_0, L_0)$ . The detailed steps are elaborated as below.

(a) Update the shared dictionary  $D_0$  and matrix  $L_0$ 

Without loss of generality, we first concentrate on optimizing the shared dictionary  $D_0$  by fixing the variables ( $L_0$ ,  $Z_0$ ). The function (13) for  $D_0$  becomes

$$\min_{D_0} \|X' - D_0 Z_0\|_F^2, \ s.t. \ \|d_0^k\|_2^2 = 1, \ (k = 1, 2, \cdots, k_0), \tag{14}$$

where  $X' = \sum_{i=1}^{C} X'_i$  and  $Z_0 = \sum_{i=1}^{C} Z_0^i$ . The quadratic constraint is introduced on each atom in  $D_0$  to avoid the scaling issue risk. The above minimization can be solved by adopting the Lagrange dual algorithm presented as in [28].

Based on the obtained dictionary  $D_0$ , the graph Laplacian matrix  $L_0$  can be constructed using (5) and (6).

(b) Update the commonality representation  $Z_0$ 

By fixing the variables  $(D_0, L_0)$ , the optimization of the commonality representation  $Z_0$  in Equation (13) can be formulated as

$$\min_{Z_0} \|X' - D_0 Z_0\|_F^2 + \eta \left( \|Z_0\|_F^2 + Tr(Z_0^{\mathsf{T}} L_0 Z_0) \right), \tag{15}$$

where  $X' = \sum_{i=1}^{C} X'_i$  and  $Z_0 = \sum_{i=1}^{C} Z_0^i$ ; there is a closed-form solution for  $Z_0$  as all the above terms are based on the  $l_2$  norm regularization.

By setting the derivative of (15) to zero, the optimal representation  $Z_0$  can be obtained as  $Z_0 = (D_0^T D_0 + 2\eta (L_0 + I_0)) D_0^T X'$ , where  $I_0 \in \mathbf{R}^{k_0 \times k_0}$  is an identity matrix.

## 4.1.2. Class-Specific Dictionary Learning

Assuming the variables  $(D_0, L_0, Z_0)$  are fixed, the objective function (4) for learning the class-specific dictionary can be formulated as follows.

$$\min_{D_p, Z_p} \sum_{i=1}^{C} \left\{ \|X_i'' - D_p Z_p^i\|_F^2 + \lambda \|P_i Z_p^i\|_F^2 + \gamma \left( \|Z_p^i\|_F^2 + \operatorname{Tr} \left( Z_p^i L_i Z_p^{iT} \right) \right) \right\}, \quad (16)$$

$$s.t. \|d_p^l\|_2^2 = 1, \ (l = 1, 2, \cdots, K_p),$$

where  $X_i'' = X_i - D_0 Z_0^i$ . To update the variables  $(D_p, Z_p)$ , we follow the iterative scheme presented in Section 4.1.1, i.e., updating  $D_p$  by fixing  $Z_p$ ; then updating  $Z_p$  by fixing  $D_p$ . The updating steps are elaborated as below.

(c) Update the class-specific dictionary  $D_p$ 

By fixing the variable  $Z_p$ , the optimization of dictionary  $D_p$  can be simplified as the optimization of each atom  $d_p^k$  ( $k = 1, 2, \dots, K_p$ ). Provided  $k \in S_i$  and the other atoms in  $D_p$  are fixed, to update the atom  $d_p^k$  we can solve the optimization problem as follows.

$$\min_{d_p^k} \|X'' - \sum_{t \notin S_i} d_p^t \bar{z}_p^t - \sum_{t \in S_i, t \neq k} d_p^t \bar{z}_p^t - d_p^k \bar{z}_p^k \|_F^2,$$
(17)

where  $X'' = \sum_{i=1}^{C} X''_i$ , and  $\bar{z}^t_p$  denotes the *t*th row vector of coefficient matrix  $Z_p$ . By introducing a variable  $\tilde{E} = X'' - \sum_{t \notin S_i} d^t_p \bar{z}^t_p - \sum_{t \in S_i, t \neq k} d^t_p \bar{z}^t_p$ , the solution of atom  $d^k_p$  can be easily derived as

$$\tilde{d}_{p}^{k} = \frac{1}{\|\bar{z}_{p}^{t}\|_{2}^{2}} \tilde{E} \bar{z}_{p}^{t\mathrm{T}}.$$
(18)

Consider the energy of each atom as being constrained in (16), the solution can be further normalized as  $\tilde{}$ 

$$\hat{d}_{p}^{k} = \frac{E\bar{z}_{p}^{r_{1}}}{\|\tilde{E}\bar{z}_{p}^{tT}\|_{2}}.$$
(19)

Because the atoms with indices out of  $S_i$  are fixed when updating the atoms with indices in  $S_i$ , we can compute  $E = X'' - \sum_{t \notin S_i} d_p^t \overline{z}_p^t$  in advance and take it into the calculation

of  $\tilde{E}$  to accelerate the update. Likewise, we successively update the atoms corresponding to  $S_i$  with ( $i = 1, 2, \dots, C$ ) and obtain the overall class-specific dictionary  $D_p$ .

(d) Update the particularity representation  $Z_p$ 

For the particularity representation of the *i*th class  $Z_{p}^{i}$ , the optimization problem depending on it becomes

$$\min_{Z_p^i} \|X_i'' - D_p Z_p^i\|_F^2 + \lambda \|P_i Z_p^i\|_F^2 + \gamma \Big( \|Z_p^i\|_F^2 + Tr(Z_p^i L_i Z_p^{iT}) \Big).$$
(20)

The solution of the above equation can be obtained by computing each code in  $Z_p^i$  as follows

$$\min_{z_{p}^{i}(n)} \|x_{i}^{\prime\prime}(n) - D_{p}z_{p}^{i}(n)\|_{2}^{2} + \lambda \|P_{i}z_{p}^{i}(n)\|_{2}^{2} + \gamma \|z_{p}^{i}(n)\|_{2}^{2} 
+ \gamma \sum_{n_{1},n_{2}} L_{i}(n_{1},n_{2})z_{p}^{i}(n_{2})^{\mathrm{T}}z_{p}^{i}(n_{1}),$$
(21)

where  $x_i''(n)$   $(n = 1, 2, \dots, N_i)$  is the *n*th training sample of the *i*th class,  $z_p^i(n)$  is the code of  $x_i''(n)$  and  $L_i(n_1, n_2)$  denotes the  $(n_1, n_2)$ th entry of matrix  $L_i$ . Since all the terms in Equation (21) are based on the quadratic form, the solution of  $z_p^i(n)$  can be easily obtained by setting its derivation to zero.

By using (12) and denoting  $\bar{L}_i$  as  $L_i - I_i$  with the identity matrix  $I_i \in \mathbb{R}^{N_i \times N_i}$ , we can further obtain a matrix version for  $Z_p^i$  update as:

$$\hat{Z}_p^i = \left(D_p^{\mathrm{T}} D_p + \lambda P_i^{\mathrm{T}} P_i + 2\gamma I_i\right)^{-1} \left(D_p^{\mathrm{T}} X_i^{\prime\prime} - \gamma Z_p^i \bar{L}_i\right).$$
<sup>(22)</sup>

By repeating the steps (a)–(d), we can iteratively obtain all the optimal variables  $(D_0, L_0, Z_0, D_p, Z_p)$ . The overall optimization of our LPLC-HDL is summarized in Algorithm 1. Moreover, the objective function value of Equation (4) decreases as the number of iterations increases on a given dataset. For example, the convergence curve of LPLC-HDL on the Yale face dataset [29] is illustrated in Figure 1. From it we can see that the proposed algorithm converges quickly, no more than 20 iterations.



**Figure 1.** The convergence curve of the proposed LPLC-HDL with the number of iterations on the Yale face dataset.

In Algorithm 1, the time complexity of our LPLC-HDL mainly comes from these parts:  $O(k_0^3mN)$  for updating the shared dictionary  $D_0$ ,  $O(k_0^2)$  for updating the graph Laplacian matrix  $L_0$ ,  $O(k_0^3)$  for updating the commonality representation  $Z_0$ ,  $O(K_pmN)$  for updating the class-specific dictionary  $D_p$  and  $O(K_p^3)$  for updating the particularity representation  $Z_p$ . Thus the final time complexity is  $O((k_0^3 + K_p)mN + K_p^3)$  in each iteration.

#### Algorithm 1 Optimization procedure of the proposed LPLC-HDL

1: **Input:** Train samples *X*, class label *Y*, the parameters  $\lambda$ ,  $\gamma$ ,  $\eta$ 

- 2: Initialization: Using the k-means algorithm, initialize D<sub>0</sub> by all the training samples, compute the matrix L<sub>0</sub> with Equation (6); Initialize D<sub>i</sub> (i = 1, 2, ..., C) by the training samples X<sub>i</sub>, concatenate these sub-dictionaries as the initialized D<sub>p</sub>; Initialize the coefficients Z<sub>0</sub> and Z<sub>p</sub> with the MVRR algorithm; Construct the serial of matrices P<sub>i</sub> and L<sub>i</sub> with Equations (9) and (12), respectively.
  3: while converge condition no satisfied do
- 4: (a) Update  $D_0$  by solving Equation (14), update  $L_0$  by using Equations (5) and (6);
- 5: **(b)** Update  $Z_0 = (D_0^T D_0 + 2\eta (L_0 + I_0)) D_0^T X';$
- 6: (c) Update  $D_p$  as follows:
- 7: **for** i = 1 to *C* **do**
- 8: Compute  $E = X'' \sum_{t \notin S_i} d_p^t \bar{z}_p^t$ ;
- 9: **for**  $k \in S_i$  **do**
- 10: Update the atom  $d_p^k$  by solving Equation (19);
- 11: end for
- 12: **end for**
- 13: **(d)** Update  $Z_p$  as follows:
- 14: **for** i = 1 to C **do**
- 15: Update the coding coefficients  $Z_p^i$  by solving Equation (22);
- 16: **end for**
- 17: end while
- 18: **Output:** The variables  $D_0$ ,  $L_0$ ,  $Z_0$ ,  $D_p$ ,  $Z_p$ .

## 4.2. Classification Procedure

After completing Algorithm 1, we can use the learned hybrid dictionary  $D = [D_0, D_p]$  to represent a test sample and predict its label. Firstly, we find the commonality–particularity representation  $\hat{z}_t = [(z_t^0)^T, (z_t^p)^T]^T$  of a test sample  $x_t$  by solving the following equation:

$$\hat{z}_t = \arg\min_{t} \|x_t - Dz_t\|_2^2 + \gamma \|z_t\|_2^2.$$
(23)

Secondly, we compute  $x'_t = x_t - D_0 z_t^0$  based on  $\hat{z}_t$  to exclude the contribution of the shared dictionary. Thus, the class label of  $x_t$  can be determined by:

$$label(x_t) = \arg \min \|x_t' - D_i z_t^i\|_2^2, \ i = 1, 2, \cdots, C.$$
(24)

## 5. Experiments

In this section, we compare LPLC-HDL with the representative dictionary learning methods including D-KSVD [5], LC-KSVD [6], LCLE-DL [7], FDDL [13], DL-COPAR [16] and CLSDDL [19] on the Yale face dataset [29], the Extended YaleB face dataset [30], the Labeled Faces in the Wild (LFW) dataset [31] for face recognition, the Caltech-101 object dataset [23] and the Oxford 102 Flowers dataset [32] for object classification and flower classification, respectively. Moreover, we further compare it with the Sparse Representation based Classification (SRC) [33], the Collaborative Representation based Classifier (CRC) [34], the Probabilistic CRC (ProCRC) [35], the Sparsity Augmented Collaborative Representation (SACR) [36] and some other state-of-the-art methods on the particular datasets.

In the proposed LPLC-HDL method, the three parameters  $\lambda$ ,  $\gamma$  and  $\eta$  are selected by five-fold cross validation on the training set, and their optimal values on each dataset are shown in Table 1. In addition, the atom numbers of the shared dictionary and the class-specific dictionary for each dataset are elaborated as detailed in the following experiments.

	Yale	Ext. YaleB	LFW	Caltech-101	102 Flowers
λ	$1  imes 10^{-1}$	$1  imes 10^{-1}$	$1 \times 10^{-2}$	3	$2  imes 10^{-1}$
$\gamma$	$1 imes 10^4$	$2 imes 10^3$	$5  imes 10^3$	$3 \times 10^2$	$1 imes 10^4$
η	$1 \times 10^{-3}$	$1 \times 10^{-3}$	$1  imes 10^{-3}$	$1 \times 10^1$	$1  imes 10^{-1}$

 Table 1. The parameters of the proposed LPLC-HDL method on each dataset.

## 5.1. Experiments on the Yale Face Dataset

In the experiments, we first consider the Yale face dataset which contains 165 gray scale images for 15 individuals with 11 images per category. Each individual contains one different facial expression or configuration: Left-light, center-light, right-light, w/glasses, w/no glasses, happy, sad, normal, surprised, sleepy and wink, as shown in Figure 2a. Each image has  $24 \times 24$  pixel resolution, and is resized to 576-dimensional vectors with normalization for representation. Following the setting in [19], six images of each individual are randomly selected for training and the rest are used for testing. In addition, the number of dictionary atoms is set to  $15 \times 4 + 60 = 120$ , which means four class-specific atoms for each individual and 60 shared atoms for all the individuals. The parameters of our LPLC-HDL on the Yale face dataset can be seen in Table 1.



(a) The Yale face dataset.



(b) The Extended YaleB face dataset.



(c) The LFW face dataset.

Figure 2. The example images from the face datasets.

To acquire a stable recognition accuracy, we operate LPLC-HDL over 30 times rather than 10 times with independent training/testing splits. The comparison results on the Yale face dataset are listed in Table 2. Since the variations in terms of facial expression are complex, the variations in the testing samples cannot be well represented via directly using the training data. Hence SRC has worse accuracy than the other dictionary learning methods. We can also see that the hybrid dictionary learning methods including DL-COPAR [16], LRSDL [17] and CLSDDL [19] outperform the remaining competing approaches, and the proposed LPLC-HDL method achieves the best recognition accuracy of 97.01%, which illustrates that our method can distinguish the shared and class-specific information of face images more appropriately.

Method Acc. (%) Method Acc. (%) LLC [37]  $82.1\pm2.6$ LC-KSVD (120) [6]  $83.6\pm2.7$  $94.4\pm2.8$  $91.2 \pm 3.1$ SVM [38] LCLE-DL (120) [7]  $81.3\pm2.5$ SRC [33] DL-COPAR (120) [16]  $92.3\pm3.7$  $89.8 \pm 1.9$ CRC [34]  $92.7\pm3.3$ LRSDL (120) [17] ProCRC [35]  $91.7\pm2.1$  $95.9\pm2.2$ CLSDDL-LC (120) [19] D-KSVD (120) [5]  $82.3\pm4.5$ CLSDDL-GC (120) [19]  $95.3\pm2.9$ FDDL (120) [13]  $89.6\pm2.6$ LPLC-HDL (120)  $97.01 \pm 2.6$ 

**Table 2.** The recognition rates (%) on the Yale face dataset.

We further illustrate the performance of LPLC-HDL on face recognition accuracy with different sizes of the shared and class-specific dictionaries, as shown in Figure 3. From it we can see that a higher recognition accuracy on the Yale face dataset can be obtained by increasing the atom numbers of both the shared and class-specific dictionaries. Besides, Figure 3 also shows that when the number of the class-specific atoms is few, the recognition accuracy is sensitive to the number of the shared atoms, as the large shared dictionary is harmful to the discriminative power of the class-specific dictionary in this case. However, increasing the number of the class-specific atoms beyond the particular number of training samples of each class brings no notable increase in recognition accuracy. Considering that the large dictionary will slow down both the training time and the testing time, the atom numbers of the shared and class-specific dictionaries are not set to big values in the experiments.



**Figure 3.** The recognition accuracy versus different sizes of the shared dictionary and the class-specific dictionary on the Yale face dataset.

## 5.2. Experiments on the Extended YaleB Face Dataset

The Extended YaleB face dataset contains 2414 frontal face images of 38 people. For each person, there are about 64 face images and the original images have  $192 \times 168$  pixels. This dataset is challenging due to varying poses and illumination conditions, as displayed in

Figure 2b. For comparison, we use the normalized  $32 \times 32$  images instead of the original pixel information. In addition, we randomly select 20 images per category for training and take the rest as the testing images. The parameters of LPLC-HDL on the Extended YaleB face dataset are also shown in Table 1. The hybrid dictionary size is set to  $38 \times 13 + 76 = 570$ , which denotes 13 class-specific atoms for each person and 76 shared atoms for all the persons, with the same structure adopted by the other hybrid dictionary learning methods.

We repeatedly run LPLC-HDL and all the comparison methods 10 times for reliable accuracy, and the average recognition rates are listed in Table 3. As shown in Table 3, compared with the K-SVD, D-KSVD, FDDL and LC-KSVD methods, LCLE-DL achieves a better recognition result with the same dictionary size. The reason for this behavior is that the LCLE-DL method can effectively utilize the locality and label information of the atoms in the dictionary learning. It is also shown that the hybrid dictionary learning methods, including DL-COPAR, LRSDL, CLSDDL and LPLC-HDL, generally outperform the other DL methods, which demonstrates the discriminative ability of the hybrid dictionary. By integrating the locality and label information into the hybrid dictionary, our LPLC-HDL method obtains the best recognition rate of 97.25%, which outperforms the second best approach CLSDDL-GC by 0.8%, and at least 1.2% higher than the other competing methods.

Table 3. The recognition rates (%) on the Extended YaleB face dataset.

Method	Acc. (%)	Method	Acc. (%)
LLC [37]	$88.9\pm0.57$	LC-KSVD (570) [6]	$92.7\pm0.62$
SVM [38]	$92.5\pm0.82$	LCLE-DL (570) [7]	$95.8\pm0.42$
SRC [33]	$95.3\pm0.65$	DL-COPAR (570) [16]	$95.8\pm0.88$
CRC [34]	$95.0\pm0.48$	LRSDL (570) [17]	$95.7\pm0.51$
K-SVD (570) [39]	$94.0\pm0.71$	CLSDDL-LC (570) [19]	$96.05\pm0.56$
D-KSVD (570) [5]	$94.3\pm0.63$	CLSDDL-GC (570) [19]	$96.45\pm0.52$
FDDL (570) [13]	$93.4\pm0.76$	LPLC-HDL (570)	$97.25\pm0.43$

### 5.3. Experiments on the LFW Face Dataset

The LFW face dataset has more than 13,000 images with the name of the person pictured, and all of them are collected from the web for unconstrained face recognition and verification. Following the prior work [7], we use a subset of the LFW face dataset which consists of 1215 images from 86 persons. In this subset, there are around 11–20 images for each person, and all the images are resized to be  $32 \times 32$ . Some samples from this face dataset are shown in Figure 2c. For each person, eight samples are randomly selected for training and the remaining samples are taken for testing. The parameters of LPLC-HDL on the LFW face dataset are also shown in Table 1. In addition, the hybrid dictionary size is set to  $86 \times 4 + 86 = 430$ , which means four class-specific atoms for each individual and 86 shared atoms for all the individuals.

We repeatedly run LPLC-HDL and the comparison methods 10 times, the average recognition rates are reported in Table 4, where the symbol  $\pm$  denotes the standard deviation of average recognition rates. Similar to the results on the Extended YaleB face dataset, the LCLE-DL method achieves a higher recognition rate than the other shared or class-specific DL methods, and the hybrid dictionary learning methods, e.g., DL-COPAR, LRSDL, CLSDDL and LPLC-HDL, have a super performance in general. The proposed LPLC-HDL method obtains the best result with an average recognition rate of 42.39%, which is significantly better than all the comparison methods.

Method	Acc. (%)	Method	Acc. (%)	
LLC [37]	$34.6\pm0.88$	LC-KSVD (430) [6]	$34.1 \pm 1.32$	
SVM [38]	$37.73 \pm 1.5$	LCLE-DL (430) [7]	$39.1 \pm 1.19$	
SRC [33]	$36.5\pm1.92$	DL-COPAR (430) [16]	$38.63 \pm 1.6$	
CRC [34]	$36.7\pm1.55$	LRSDL (430) [17]	$40.23 \pm 1.4$	
K-SVD (430) [39]	$32.8 \pm 1.27$	CLSDDL-LC (430) [19]	$41.05 \pm 1.5$	
D-KSVD (430) [5]	$33.2\pm1.25$	CLSDDL-GC (430) [19]	$38.45 \pm 1.8$	
FDDL (430) [13]	$34.8 \pm 1.76$	LPLC-HDL (430)	$42.39\pm1.6$	

Table 4. The recognition rates (%) on the LFW face dataset.

# 5.4. Object Classification

In this subsection, we evaluate LPLC-HDL on the Caltech-101 dataset [23] for object classification. This dataset contains a total of 9146 images from 101 object categories and a background category. The number of images for per category varies from a minimum of 31 to a maximum of 800 images. The resolution of each image is about  $300 \times 200$ , as shown in Figure 4. Following the settings [6,40], we perform the Spatial Pyramid Features (SPFs) on this dataset. Firstly, we partition each image into  $2^L \times 2^L$  sub-regions with different spatial scales L = 0, 1, 2, then extract SIFT descriptors over a sub-region with a spacing of 8 pixels and concatenate them as the SPFs. Next, we encode the SPFs with a codebook of size 1024. Finally, the dimension of the features is reduced to 3000 using the the Principal Component Analysis (PCA) algorithm.

For this dataset, 10 samples of each category are selected as the training data and the remaining are for testing. The parameters of LPLC-HDL on this dataset can be seen in Table 1. The hybrid dictionary size is set to  $102 \times 9 + 100 = 1018$ , which denotes nine class-specific atoms for each category and 100 shared atoms for all the categories. The proposed LPLC-HDL and the comparison methods are carried out 10 times, the average classification rates are reported in Table 5. As can be seen in Table 5, our LPLC-HDL method achieves the best classification result again, with improvement margins of at least 1.3% compared with the comparison methods.



Figure 4. Examples from the Caltech-101 object dataset.

Table 5. The classification results (%) on the Caltech101 object dataset.

Method	Acc. (%)	Method	Acc. (%)
SVM [38]	$59.1\pm0.80$	LC-KSVD (1020) [6]	$63.1\pm0.45$
LLC [37]	$59.8\pm0.61$	LCLE-DL (1020) [7]	$62.7\pm0.31$
SRC [33]	$60.1\pm0.51$	JBDC (1020) [15]	$61.3\pm0.35$
CRC [34]	$59.4 \pm 0.48$	DL-COPAR (1018) [16]	$63.0\pm0.57$
SACR [36]	$60.7\pm0.40$	LRSDL (1018) [17]	$63.5\pm0.55$
K-SVD (1020) [39]	$59.1\pm0.65$	CLSDDL-LC (1018) [19]	$63.7\pm0.48$
D-KSVD (1020) [5]	$59.5\pm0.87$	CLSDDL-GC (1018) [19]	$64.0\pm0.43$
FDDL (1020) [13]	$63.6\pm0.59$	LPLC-HDL (1018)	$65.3\pm0.38$

### 5.5. Flower Classification

We finally evaluate the proposed LPLC-HDL on the Oxford 102 Flowers dataset [32] for fine-grained image classification, which consists of 8189 images from 102 categories, and each category contains at least 40 images. This dataset is very challenging because there exist large variations within the same category but small differences across several categories. The flowers appear at different scales, poses and lighting conditions; some flower images are shown in Figure 5. For each category, 10 images are used for training, 10 for validation, and the rest for testing, as in [32]. For ease of comparison, we take the convolutional neural network (CNN) features provided by Cai et al. [35] as the image-level features.

The parameters of LPLC-HDL on the Oxford 102 Flowers dataset can be seen in Table 1. The size of hybrid dictionary is set to be  $102 \times 9 + 100 = 1018$ , which means nine atoms for each category and 100 atoms as the shared amount. We compare LPLC-HDL with different kinds of representative methods for flower classification, such as the basic classifiers (Softmax and linear SVM [38]), the recent ProCRC [35], the related DDL methods and the deep learning models. Following the common measurement [41,42], we evaluate the comparison methods by the average classification accuracy of all the categories, and the results are reported in Table 6. From it we can see that, taking the CNN features as the image-level features, the proposed LPLC-HDL method outperforms the basic classifiers and the competing DL methods. Compared with the other kinds of methods, LPLC-HDL is greatly superior to SparBases [43] and SMP [44]. Moreover, our method outperforms the recent deep learning models including GoogLeNet-GAP [45], ASPD [46] and AugHS [47], which need to design the special CNN architectures for flower classification.



Figure 5. Example images from the Oxford 102 Flowers dataset.

Method	Acc. (%)	Method	Acc. (%)	
Softmax	87.3	LC-KSVD (1020) [6]	93.60	
SVM [38]	90.9	DL-COPAR (1025) [16]	91.62	
SRC [33]	93.2	JBDC (1002) [15]	93.12	
NSC [48]	90.10	CLSDDL-LC (1018) [19]	94.01	
CRC [34]	93.0	CLSDDL-GC (1018) [19]	94.47	
ProCRC [35]	94.8	OverFeat [41]	86.8	
K-SVD (1020) [39]	92.41	NAC [49]	95.3	
D-KSVD (1020) [5]	93.55	DAS [50]	80.66	
FDDL (1020) [13]	94.60	LPLC-HDL (1018)	95.7	

 Table 6. The classification results (%) on the Oxford 102 Flowers dataset.

### 5.6. Parameter Sensitivity

In the proposed LPLC-HDL method, there are three key parameters, i.e.,  $\lambda$ ,  $\gamma$  and  $\eta$ , which are used to balance the importance of the label-aware constraint, the group regularization and the locality constraint. To analyze the sensitivity of the parameters, we define a candidate set {10<sup>-5</sup>, 10<sup>-4</sup>, 10<sup>-3</sup>, 10<sup>-2</sup>, 10<sup>-1</sup>, 10<sup>0</sup>, 10<sup>1</sup>, 10<sup>2</sup>, 10<sup>3</sup>, 10<sup>4</sup>, 10<sup>5</sup>} for them and then perform LPLC-HDL with different combinations of the parameters on the Yale face dataset and the Caltech 101 dataset. By fixing the parameter  $\eta$ , the classification accuracy versus

different values of the parameters  $\lambda$  and  $\gamma$  are shown in Figures 6a and 7a. As can be seen in the figures, the best classification result can be obtained when the parameters  $\lambda$  and  $\gamma$ locate in a feasible range. When  $\gamma$  is very small, the effect of the group regularization is limited, leading to weak discrimination of the class-specific dictionary. On the other hand, when  $\gamma$  becomes very large, the classification accuracy drops as the remaining terms in (4) become less important, which decreases the representation ability of the hybrid dictionary.

By fixing the parameters  $\lambda$  and  $\gamma$ , the classification accuracy versus different values of the parameter  $\eta$  are shown in Figures 6b and 7b. From them we can see that the classification accuracy is insensitive to the parameter  $\eta$  when its value located in a certain range, e.g.,  $10^{-5} \leq \eta \leq 10^{-1}$  on the Yale face dataset. It should be noted that the incoherence between the shared and class-specific dictionaries is increased with increasing parameter  $\eta$ , which influences the reconstruction of the test data and decreases the classification accuracy.

Due to the diversity of the datasets, it is still an open problem to adaptively select the optimal parameters for the different datasets. In the experiments, we use an effective and simple way to find the optimal values for the parameters  $\lambda$ ,  $\gamma$  and  $\eta$ . Based on the previous analysis, we first fix the parameter  $\eta$  to a small value such as 0.01, then search the candidate combination of the parameters  $\lambda$  and  $\beta$  from the coarse set of {10<sup>-5</sup>, 10<sup>-4</sup>, 10<sup>-3</sup>, 10<sup>-2</sup>, 10<sup>-1</sup>, 10<sup>0</sup>, 10<sup>1</sup>, 10<sup>2</sup>, 10<sup>3</sup>, 10<sup>4</sup>, 10<sup>5</sup>}. According to the best coarse combination of them, we can further define a fine candidate set where their optimal values may exist. Then we perform the proposed LPLC-HDL again with different combinations of the parameters  $\lambda$  and  $\beta$  selected from the fine candidate set. This way, we can obtain the optimal values of the parameters for all the experimental datasets: hence, the best classification results are guaranteed.



**Figure 6.** The classification accuracy versus parameters  $\lambda$ ,  $\gamma$  and  $\eta$  on the Yale face dataset; (a) parameter  $\eta$  is fixed; (b) parameters  $\lambda$  and  $\gamma$  are fixed.

Besides the key parameters  $\lambda$ ,  $\gamma$  and  $\eta$  in our LPLC-HDL method, there are also the parameters  $\delta$  and k in the proposed locality constraint. In the experiments, we find that these two parameters have stable values on the experimental datasets. For example, Figure 8a,b show the classification accuracies of our LPLC-HDL method with respect to the parameter  $\delta$  by fixing the remaining parameters on the Yale face dataset and the Caltech 101 dataset. From the subfigures, we can see that the classification accuracy is insensitive to the parameter  $\delta$ , and the approximate best result can be obtained when  $\delta = 4$ . For the parameter k, this similar phenomenon can be observed on the experimental datasets. Thus we can set  $\delta = 4$  and k = 1 in our LPLC-HDL method for convenience of calculation.



**Figure 7.** The classification accuracy versus parameters  $\lambda$ ,  $\gamma$  and  $\eta$  on the Caltech 101 object dataset; (a) parameter  $\eta$  is fixed; (b) parameters  $\lambda$  and  $\gamma$  are fixed.



**Figure 8.** The classification accuracy versus parameter  $\delta$  on (**a**) the Yale face dataset, and (**b**) the Caltech 101 object dataset by fixing the parameters  $\lambda$ ,  $\gamma$ ,  $\eta$  and k.

# 5.7. Evaluation of Computational Time

We also conducted experiments to evaluate the running time of the proposed LPLC-HDL and other representative DL methods on the two face datasets and the Caltech-101 dataset, the comparison results are listed in Table 7. "Train" denotes the running time of each iteration, and "Test" is the average processing time for classifying one test sample. All the experiments are conducted on a 64-bit computer with Intel i7-7700 3.6 GHz CPU and 12 GB RAM under the MATLAB R2019b programming environment. From Table 7, we can see that, although slower than CLSDDL-LC [19], the training efficiency of LPLC-HDL is obviously higher than that of D-KSVD [5], LC-KSVD [6], FDDL [13] and DL-COPAR [16]. In the testing stage, the proposed LPLC-HDL has similar testing efficiency as D-KSVD and LC-KSVD, and the testing process of LPLC-HDL is much faster than that of FDDL and DL-COPAR. Specifically, the average time for classifying a test image by LPLC-HDL is always less than that of CLSDDL-LC on the experiments.

	Ext. YaleB			LFW Caltech-1		ltech-101
	Train	Test	Train	Test	Train	Test
D-KSVD [5]	46.7	$3.36  imes 10^{-4}$	57.3	$1.70  imes 10^{-3}$	303.4	$2.89 \times 10^{-3}$
LC-KSVD [6]	47.8	$3.37 imes10^{-4}$	59.1	$1.57  imes 10^{-3}$	311.2	$3.12  imes 10^{-3}$
FDDL [13]	455.8	$9.40  imes 10^{-2}$	558.3	$2.90 imes10^{-1}$	3537.0	1.59
DL-COPAR [16]	8.54	$1.01 \times 10^{-2}$	38.9	$6.21 \times 10^{-2}$	120.5	$1.12  imes 10^{-1}$
CLSDDL-LC [19]	1.26	$1.53 imes10^{-4}$	1.42	$9.18 imes10^{-4}$	21.37	$5.10 imes10^{-3}$
LPLC-HDL	3.62	$1.50 imes10^{-4}$	3.82	$6.25 imes10^{-4}$	22.19	$3.80 imes10^{-3}$

 Table 7. The computational times (second) on the experimental datasets.

# 6. Conclusions

In this paper, we propose a novel hybrid dictionary learning (LPLC-HDL) method by taking advantage of the locality and label information of the data, which can solve the image classification task effectively. The LPLC-HDL method incorporates the locality constraint with the label-aware constraint to more appropriately distinguish the shared and particularity information. More specifically, the locality constraint on the shared dictionary is used to model the similar features of the images from different classes; the label-aware constraint and the group regularization are coupled to make the class-specific dictionary more discriminative. An effective alternative strategy is developed to solve the objective function of LPLC-HDL. After training the hybrid dictionary, the class label of the test image can be predicted more accurately by excluding the contribution of the shared dictionary. The experimental results on three face datasets, the object dataset and the flower dataset demonstrate that our method has effectiveness and superiority for image classification.

**Author Contributions:** Methodology, J.S. and Z.L.; Software, L.W. and M.L.; Project administration, M.Z. and Q.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the National Natural Science Foundation of China (Nos. 62002102, 62176113), in part by the Leading talents of science and technology in the Central Plain of China under Grants no. 214200510012, the Key Technologies R & D Program of Henan Province (Nos. 212102210088, 202102210169), the Basic Research Projects in the University of Henan Province (No. 19zx010), Guizhou Provincial Science and Technology Foundation ([2020]1Y253), Henan Postdoctoral Foundation, in part by the Luoyang Major Scientific and Technological Innovation Projects under Grant No.2101017A, and Foundation from the Postdoctoral Research Station of Control Science and Engineering at Henan University of Science and Technology.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

## References

- 1. Elad, M.; Aharon, M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* **2006**, *15*, 3736–3745. [CrossRef] [PubMed]
- Yan, R.; Shao, L.; Liu, Y. Nonlocal Hierarchical Dictionary Learning Using Wavelets for Image Denoising. *IEEE Trans. Image Process.* 2013, 22, 4689–4698. [CrossRef]
- 3. Li, S.; Shao, M.; Fu, Y. Person Re-identification by Cross-View Multi-Level Dictionary Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, 40, 2963–2977. [CrossRef]
- Li, K.; Ding, Z.; Li, S.; Fu, Y. Toward Resolution-Invariant Person Reidentification via Projective Dictionary Learning. *IEEE Trans.* Neural Netw. Learn. Syst. 2019, 30, 1896–1907. [CrossRef]
- Zhang, Q.; Li, B. Discriminative K-SVD for dictionary learning in face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2691–2698.
- 6. Jiang, Z.; Lin, Z.; Davis, L.S. Label Consistent K-SVD: Learning a Discriminative Dictionary for Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2651–2664. [CrossRef]

- Li, Z.; Lai, Z.; Xu, Y.; Yang, J.; Zhang, D. A Locality-Constrained and Label Embedding Dictionary Learning Algorithm for Image Classification. *IEEE Trans. Neural Netw. Learn. Syst.* 2017, 28, 278–293. [CrossRef]
- Gu, S.; Zhang, L.; Zuo, W.; Feng, X. Projective dictionary pair learning for pattern classification. In Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 793–801.
- Mairal, J.; Ponce, J.; Sapiro, G.; Zisserman, A.; Bach, F.R. Supervised dictionary learning. In Proceedings of the 23th International Conference on Neural Information Processing Systems, Whistler, BC, Canada, 6–11 December 2009; pp. 1033–1040.
- Mairal, J.; Bach, F.; Ponce, J. Task-Driven Dictionary Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 2012, 34, 791–804. [CrossRef] [PubMed]
- 11. Song, J.; Xie, X.; Shi, G.; Dong, W. Exploiting class-wise coding coefficients: Learning a discriminative dictionary for pattern classification. *Neurocomputing* **2018**, *321*, 114–125. [CrossRef]
- Song, Y.; Zhang, Z.; Liu, L.; Rahimpour, A.; Qi, H. Dictionary Reduction: Automatic Compact Dictionary Learning for Classification. In Proceedings of the Asian Conference on Computer Vision, Taipei, Taiwan, 21–23 November 2016; pp. 305–320.
- Yang, M.; Zhang, L.; Feng, X.; Zhang, D. Fisher Discrimination Dictionary Learning for sparse representation. In Proceedings of the International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 543–550.
- Ramirez, I.; Sprechmann, P.; Sapiro, G. Classification and clustering via dictionary learning with structured incoherence and shared features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3501–3508.
- Akhtar, N.; Mian, A.; Porikli, F. Joint discriminative bayesian dictionary and classifier learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3919–3928.
- 16. Wang, D.; Kong, S. A classification-oriented dictionary learning model: Explicitly learning the particularity and commonality across categories. *Pattern Recognit.* **2014**, *47*, 885–898. [CrossRef]
- 17. Vu, T.H.; Monga, V. Fast low-rank shared dictionary learning for image classification. *IEEE Trans. Image Process.* 2017, 26, 5160–5175. [CrossRef] [PubMed]
- Gao, S.; Tsang, I.H.; Ma, Y. Learning Category-Specific Dictionary and Shared Dictionary for Fine-Grained Image Categorization. *IEEE Trans. Image Process.* 2014, 23, 623–634. [PubMed]
- Wang, X.; Gu, Y. Cross-label Suppression: A Discriminative and Fast Dictionary Learning with Group Regularization. *IEEE Trans. Image Process.* 2017, 26, 3859–3873. [CrossRef] [PubMed]
- 20. Wang, J.; Yang, J.; Kai, Y.; Lv, F.; Gong, Y. Locality-constrained Linear Coding for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3360–3367.
- 21. Wei, C.P.; Chao, Y.W.; Yeh, Y.R.; Wang, Y.C.F. Locality-sensitive dictionary learning for sparse representation based classification. *Pattern Recognit.* **2013**, *46*, 1277–1287. [CrossRef]
- 22. Song, J.; Xie, X.; Shi, G.; Dong, W. Multi-layer Discriminative Dictionary Learning with Locality Constraint for Image Classification. *Pattern Recognit.* **2019**, *91*, 135–146. [CrossRef]
- Fei-Fei, L.; Fergus, R.; Perona, P. Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; pp. 59–70.
- 24. Sadeghi, M.; Babaie-Zadeh, M.; Jutten, C. Learning overcomplete dictionaries based on atom-by-atom updating. *IEEE Trans. Signal Process.* **2014**, *62*, 883–891. [CrossRef]
- Sun, Y.; Liu, Q.; Tang, J.; Tao, D. Learning Discriminative Dictionary for Group Sparse Representation. *IEEE Trans. Image Process.* 2014, 23, 3816–3828. [CrossRef]
- Wen, Z.; Hou, B.; Jiao, L. Discriminative Dictionary Learning With Two-Level Low Rank and Group Sparse Decomposition for Image Classification. *IEEE Trans. Cybern.* 2017, 47, 3758–3771. [CrossRef]
- 27. Wang, C.P.; Wei, W.; Zhang, J.S.; Song, H.B. Robust face recognition via discriminative and common hybrid dictionary learning. *Appl. Intell.* **2017**, *48*, 156–165. [CrossRef]
- 28. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers; Now Publishers Inc.: Delft, The Netherlands, 2011; Volume 3, pp. 1–122.
- 29. Georghiades, A.S.; Belhumeur, P.N.; Kriegman, D.J. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 643–660. [CrossRef]
- Lee, K.C.; Ho, J.; Kriegman, D.J. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intell.* 2005, 27, 684–698.
- 31. Huang, G.B.; Ramesh, M.; Berg, T.; Learned-Miller, E. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments; Technical Report 07-49; University of Massachusetts: Amherst, MA, USA, 2007.
- 32. Nilsback, M.E.; Zisserman, A. Automated Flower Classification over a Large Number of Classes. In Proceedings of the Sixth Indian Conference on Computer Vision, Graphics & Image Processing, Bhubaneswar, India, 16–19 December 2008; pp. 722–729.
- 33. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust Face Recognition via Sparse Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227. [CrossRef] [PubMed]
- Zhang, L.; Yang, M.; Feng, X. Sparse representation or collaborative representation: Which helps face recognition? In Proceedings of the International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 471–478.

- Cai, S.; Zhang, L.; Zuo, W.; Feng, X. A probabilistic collaborative representation based approach for pattern classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2950–2959.
- Akhtar, N.; Shafait, F.; Mian, A. Efficient classification with sparsity augmented collaborative representation. *Pattern Recognit.* 2017, 65, 136–145. [CrossRef]
- Yang, J.; Yu, K.; Gong, Y.; Huang, T. Linear spatial pyramid matching using sparse coding for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1794–1801.
- Fan, R.E.; Chang, K.W.; Hsieh, C.J.; Wang, X.R.; Lin, C.J. LIBLINEAR: A library for large linear classification. J. Mach. Learn. Res. 2008, 9, 1871–1874.
- Aharon, M.; Elad, M.; Bruckstein, A. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE Trans. Signal Process.* 2006, 54, 4311–4322. [CrossRef]
- Lazebnik, S.; Schmid, C.; Ponce, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006; pp. 2169–2178.
- Razavian, A.S.; Azizpour, H.; Sullivan, J.; Carlsson, S. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 512–519.
- Azizpour, H.; Razavian, A.S.; Sullivan, J.; Maki, A.; Carlsson, S. From Generic to Specific Deep Representations for Visual Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 36–45.
- Yao, B.; Jiang, X.; Khosla, A.; Lin, A.L.; Guibas, L.; Fei-Fei, L. Human action recognition by learning bases of action attributes and parts. In Proceedings of the International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 1331–1338.
- 44. Khan, F.S.; van de Weijer, J.; Anwer, R.M.; Felsberg, M.; Gatta, C. Semantic pyramids for gender and action recognition. *IEEE Trans. Image Process.* 2014, 23, 3633–3645. [CrossRef]
- 45. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2921–2929.
- Khan, F.S.; Xu, J.; Van De Weijer, J.; Bagdanov, A.D.; Anwer, R.M.; Lopez, A.M. Recognizing actions through action-specific person detection. *IEEE Trans. Image Process.* 2015, 24, 4422–4432. [PubMed]
- 47. Qi, T.; Xu, Y.; Quan, Y.; Wang, Y.; Ling, H. Image-based action recognition using hint-enhanced deep neural networks. *Neurocomputing* **2017**, *267*, 475–488. [CrossRef]
- 48. Chien, J.-T.; Wu, C.-C. Discriminant waveletfaces and nearest feature classifiers for face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, 24, 1644–1649. [CrossRef]
- 49. Simon, M.; Rodner, E. Neural Activation Constellations: Unsupervised Part Model Discovery with Convolutional Networks. In Proceedings of the International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1143–1151.
- Angelova, A.; Zhu, S. Efficient Object Detection and Segmentation for Fine-Grained Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 811–818.