

# A Broad Sparse Fine-Grained Image Classification Model Based on Dictionary Selection Strategy

Jianjie Zheng <sup>ID</sup>, Pengpeng Liang <sup>ID</sup>, Huimin Zhao <sup>ID</sup>, Member, IEEE, and Wu Deng <sup>ID</sup>, Member, IEEE

**Abstract**—When the fine-grained recognition problems of image classification processed, broad learning system (BLS) is more efficient in classification, but has difficulty in distinguishing features with large similarities. Sparse representation classification (SRC) is more capable of handling similarity features, but is more computationally expensive. To better use the BLS model to tackle the fine-grained recognition problem and improve the ability to handle similarity features, this article combines the advantages of BLS and SRC, and proposes a broad sparse fine-grained image classification model based on dictionary selection strategy, dictionary broad sparse representation classification (DBSRC). First, to solve the parameter selection problem of the BLS model, leave one out cross validation (LOO) is introduced to quickly find the better regularization parameters and build a BLS-LOO coarse-grained classification model. Then propose reliability criteria based on a threshold selection strategy for the selection of fine-grained images. Next, an adaptive dictionary selection strategy is designed based on the output of the BLS-LOO to construct a sparse subdictionary for each fine-grained image that is not distinguished by the BLS-LOO. Finally, a sparse subdictionary based SRC model is used to classify fine-grained images. Experimental results show that DBSRC achieves good classification performance on three image datasets with different complex dimensions, ImageNet, USPS, and Pavia, and has strong processing capability for fine-grained features.

**Index Terms**—Broad sparse classification models, fine-grained identification, image classification, leave-one-out cross validation, threshold selection strategy.

## I. INTRODUCTION

IMAGE classification is a popular research area in machine learning [1], [2]. Image classification methods [3], [4] aim to assign the correct category to an image by distinguishing differences in image features. However, in the practical application of image classification, there are similar textures or shapes between some different classes of images with high similarity features (as shown in Fig. 1), resulting in fine-grained recognition problems [5], [6], [7], [8]. This poses a performance challenge for image classification methods. The core

Manuscript received 10 September 2022; revised 18 February 2023 and 15 July 2023; accepted 19 August 2023. Date of publication 7 September 2023; date of current version 5 March 2024. This work was supported by the National Natural Science Foundation of China under Grant U2133205, Grant 52372436, and Grant 51879027. Associate Editor: L. Ma. (Corresponding authors: Huimin Zhao; Wu Deng.)

Jianjie Zheng and Pengpeng Liang are with the School of Psychology, Capital Normal University, Beijing 100048, China (e-mail: zheng853796151@126.com; ppliang@cnu.edu.cn).

Huimin Zhao and Wu Deng are with the College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China (e-mail: hm\_zhao1977@126.com; wdeng@cauc.edu.cn).

Digital Object Identifier 10.1109/TR.2023.3309278

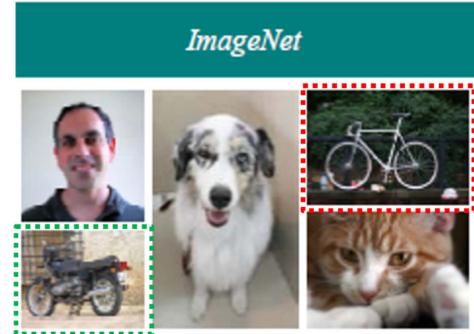


Fig. 1. Fine-grained image recognition problems [high similarity in the shape features of the bicycle (in red) and the motorbike (in green)].

of solving the fine-grained recognition problem lies in finding images with a high degree of similarity and distinguishing their features. Traditional classification methods often use manual annotation to find and distinguish similarity features, such as scale invariant feature transformations [9], directional gradient histograms [10], and speeded-up robust features [11]. Although manual features can enhance image classification, their high dependence on human experience limits their application. Due to their powerful autonomous feature analysis and extraction capabilities, deep classification networks, such as convolutional neural networks (CNNs) [12], [13] and generating adversarial networks [14], have now been successfully used to distinguish similarity features, solve fine-grained recognition problems and improve classification results. However, the performance of deep networks can be limited by the amount of training samples, and the modeling process is extremely time-consuming.

Broad learning system (BLS) [15] consists of feature nodes, enhancement nodes, and output coefficients. BLS can provide a method for image classification. BLS first extracts and analyses the image features, then calculates the probability that the image belongs to each class, and then selects the class with the higher classification probability as the image class. Compared to deep classification networks such as CNN and deep belief networks [16], BLS is more effective in modeling. BLS uses a ridge regression algorithm that requires only one iteration to solve for the output coefficients, while the deep network uses a gradient descent method that requires multiple iterations to solve for the parameters. Furthermore, the performance of deep networks is easily restricted by the number of training samples, whereas BLS performance is less affected by the number of samples.

In contrast to shallow classification methods such as extreme learning machine (ELM) [17], support vector machine (SVM) [18], BLS is able to obtain features from the data by constructing node combination (feature and enhancement nodes) to remain the effectiveness of the data. BLS is used to tackle image classification problems in many fields, such as disease diagnosis [19], [20], [21], geological exploration [22], [23], intelligent transport [24], [25], [26], and so on. Although BLS has achieved good results in many fields, for the fine-grained recognition problem in image classification, it is difficult for BLS to distinguish features with high similarity between samples, and there is a small difference between the output category probability values, resulting in unsatisfactory classification results.

Sparse representation classification (SRC) is a representation-based classification method proposed by Wright et al. [27]. SRC has been successfully applied to solve classification problems for a variety of images, such as grey-scale images [28], [29], color images [30], [31], and hyperspectral images [32], [33], [34], [35]. When there is a fine-grained recognition problem between some classes of images in the dataset, SRC is still able to use linear approximation to distinguish features with high similarity and achieve better classification results [36], [37], [38]. However, when the size of the linear combination is large, the test sample will be trapped in an immensely time-consuming query process. The reason for this is that SRC suffers from the problem of paradigm optimization and the problem of dictionary redundancy in the classification process. Luo and Zhang [39] proposed a cascade classification model, ELMSRC, by combining the efficiency advantages of ELM with the robustness advantages of SRC.

In practice, image classification tasks often suffer from fine-grained recognition problems, and BLS has difficulty in distinguishing similar features between samples, resulting in unsatisfactory classification results. Compared to BLS, SRC is able to handle similar texture or shape features between samples better, but it is more computationally expensive. Therefore, this article proposes a broad sparse fine-grained image classification model based on a dictionary selection strategy, dictionary broad sparse representation classification (DBSRC), based on the advantages of BLS and SRC. DBSRC aims to implement a simple and efficient cascade classification model to improve the effectiveness and efficiency of image classification. First, leave one out (LOO) is introduced into the BLS to select the optimal regularization parameters so that BLS can obtain a better classification hyperplane under different conditions and still maintain a high efficiency. Then, a reliability criterion based on a threshold selection strategy is proposed for the selection of fine-grained images. And an adaptive subdictionary selection strategy is designed to construct a subdictionary for each fine-grained image. Next, a sparse representation of the samples is performed using the subdictionary. Finally, the performance of the DBSRC model for classification and its ability to handle fine-grained features is verified with three image datasets of different complex dimensions.

The innovations of this article are described as follows.

- 1) A broad sparse fine-grained classification model based on dictionary selection strategy, called DBSRC, is proposed.
- 2) LOO is used in the BLS modeling process to find better regularization parameters, which results in better classification hyperplanes under different conditions while maintaining better efficiency.
- 3) A reliability criterion based on a threshold selection strategy is proposed for fine-grained image selection.
- 4) The output of BLS-LOO is employed to design an adaptive dictionary selection strategy to construct a sparse subdictionary for each fine-grained image for reducing dictionary redundancy and improving sample query efficiency.
- 5) Sparse subdictionary based SRC is employed to classify fine-grained images.
- 6) DBSRC achieves good classification performance on three image datasets of different complex dimensions, ImageNet, USPS, and Pavia, with strong processing power for fine-grained features.
- 7) Compared with SVM, ELM, CNN, SRC, BLS, ELM-SRC, and BLS-LOO, DBSRC is more capable of processing fine-grained images and can achieve better classification results with better classification efficiency.

## II. BASIC METHODS

### A. Broad Learning System

BLS is a neural network based on horizontal expansion (as shown in Fig. 2). The network can provide methods for solving classification problems. The classification process of BLS is split into training and testing stages.

1) *Training Stages:* The linear function converts input data  $X = [x_1, x_2, \dots, x_N]$  into a mapping feature and stores it in the feature node

$$Z_i = \phi_i (XW_{ei} + \beta_{ei}) \quad (1)$$

where  $X \in R^{(N \times S)}$ ,  $N$  is the total number of input data,  $S$  is the dimension of each input data,  $w_{ei}$  and  $\beta_{ei}$  represent randomly generated weight coefficients,  $i = 1, 2, \dots, n$ .  $n$  is the number of feature node groups,  $k$  is the number of feature nodes in each group, and  $\phi(\cdot)$  denotes a linear function, integrate all the mapping features and recorded it as  $Z^n$

$$Z^n \equiv [Z_1, \dots, Z_n] \quad (2)$$

where  $Z^n \in R^{(N \times nk)}$ .

The nonlinear function further transforms the characteristic node  $Z^n$  into an enhancement node

$$H_j = \xi_j (Z^n W_{h_j} + \beta_{h_j}) \quad (3)$$

where  $W_{h_j}$  and  $\beta_{h_j}$  represent randomly generated weight coefficients.  $j = 1, \dots, m$ ,  $m$  represents the number of enhancement nodes.  $\xi(\cdot)$  is activation function.

All enhancement nodes are defined as  $H^m$

$$H^m \equiv [H_1, \dots, H_m] \quad (4)$$

where  $H^m \in R^{(N \times m)}$ .

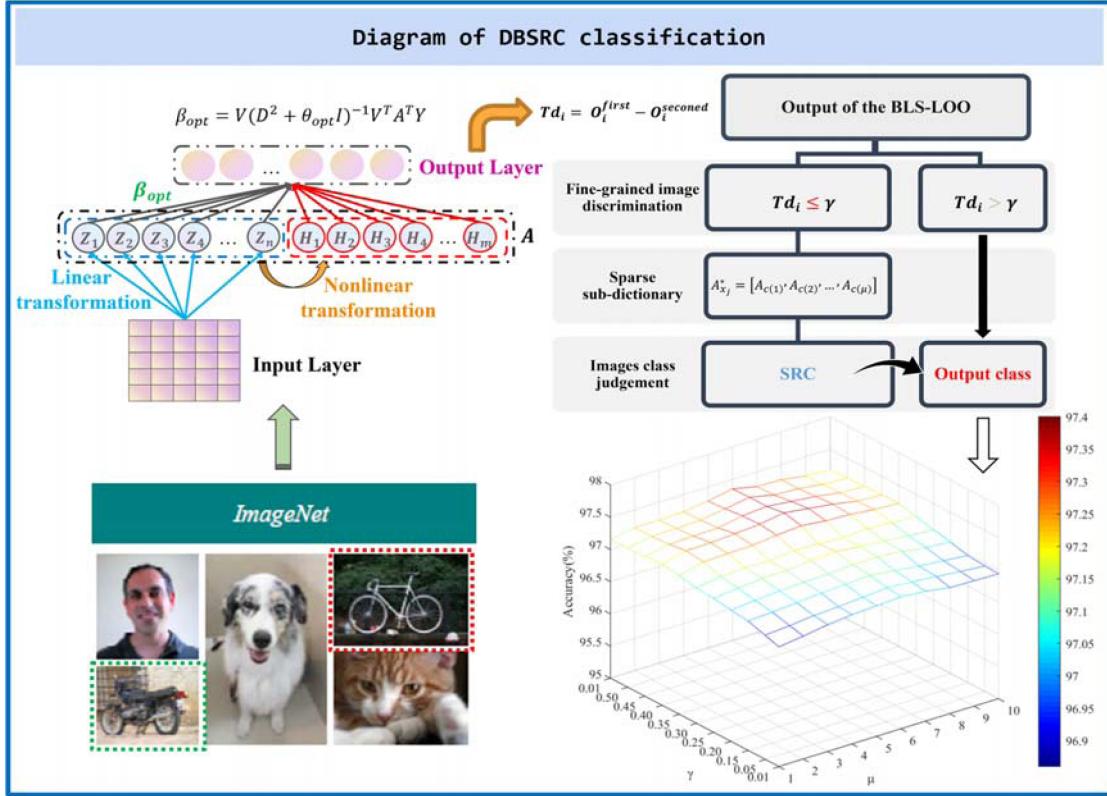


Fig. 2. Schematic diagram of image classification by DBSRC.

Integrate  $Z^n$  and  $H^m$

$$A = [Z^n | H^m] \quad (5)$$

where  $A \in R^{N \times (nk+m)}$ .

The objective function of BLS is as follows:

$$f(\beta) = \arg \min_{\beta} \theta \|\beta\|^2 + \|A\beta - Y\|^2 \quad (6)$$

where  $Y = [y_1, y_2, \dots, y_N]$  is the label corresponding to  $X$ ,  $\beta$  is the target coefficient, and  $\theta$  is the regularization parameter.

The target coefficient  $\beta$  is solved by the ridge regression

$$\beta = (A^T A + \theta I)^{-1} A^T Y \quad (7)$$

where  $I$  is the identity matrix,  $\beta \in R^{(nk+m) \times C}$  and  $C$  is the total number of categories of labeled data.

2) *Testing Stages*: Use the feature nodes weight coefficients  $w_{ei}$ ,  $\beta_{ei}$ , linear function  $\phi(\cdot)$ , enhancement nodes weight coefficients  $W_{hj}$ ,  $\beta_{hj}$ , and nonlinear function  $\xi(\cdot)$  generated by the BLS training process to construct the nodes matrix  $A_{test}$  of the test data  $[x_1, x_2, \dots, x_L]$

$$A_{test} = [Z_{test}^n | H_{test}^m] \quad (8)$$

where  $A_{test} \in R^{L \times (nk+m)}$ ,  $L$  is the number of test data.

Use the target coefficient  $\beta$  in (7) and the node matrix  $A_{test}$  in (8), the classification hyperplane  $O$  can be obtained

$$O = A_{test} \beta \quad (9)$$

where  $O \in R^{L \times C}$ .

According to the classification hyperplane  $O$ , the category of the test data can be calculated

$$Y^{test} = \arg \max_{c \in \{1, \dots, C\}} (O) \quad (10)$$

where  $Y^{test} = [y_1^{test}, \dots, y_L^{test}]$ ,  $L$  is the number of test data.

### B. Leave-One-Out Cross Validation

LOO [40], [41] is one of the useful methods for parameter optimization in the field of machine learning. First, the method divides the data set  $D = [X, Y]$  into  $N$  training data, that is  $\{(x_1, y_1), \dots, (x_N, y_N)\}$ . Next, uses  $N - 1$  data sets for training each time. finally, each time uses the remaining data to evaluate the generalization performance of the training model. Mean square error (MSE) is often used as an indicator to evaluate the generalization performance of a model. The predicted residual sum of square (PRESS) is a statistic, which is often used to calculate the MSE of the model. The PRESS calculation based on MSE ( $MSE^{PRESS}$ ) is as follows:

$$MSE^{PRESS} = \frac{1}{N} \sum_{i=1}^N \left( \frac{y_i - \hat{y}_i}{1 - HAT_{ii}} \right)^2 \quad (11)$$

where  $N$  is the total number of data sets,  $y_i$  and  $\hat{y}_i$  are the actual output and expected output of the  $i$ th data, respectively.  $HAT_{ii}$  represents the  $i$ th value on the diagonal of the HAT matrix. The HAT matrix can be calculated

$$HAT = XX^+ = X(X^T X)^{-1} X^T \quad (12)$$

where  $X^+$  is the pseudoinverse of  $X$ .

### C. Sparse Representation Classification

SRC has been extensively used in the field of classification. First, SRC evaluates the sparse representation coefficient of the query data by means of a complete dictionary (each column in the dictionary is a feature extracted from known data, called a base atom). Then, use the coefficient to calculate the minimum residual, so as to obtain the label of the query sample.

For data set  $\{A_c\}_{c=1}^C$ ,  $C$  is the total number of data categories  $A_c \in R^{n \times k_c}$ ,  $n$  is the data dimension, and  $k_c$  represents the total number of data belonging to class  $c$ .  $A = [A_1, A_2, \dots, A_C]$  is a dictionary. The optimal sparsity coefficient for the query data  $y$  can be solved by the following equation:

$$\hat{x} = \arg \min_x \|x\|_0 \text{ s.t. } Ax = y \quad (13)$$

where  $\|\cdot\|_0$  is the norm of  $\ell_0$ , and  $\hat{x}$  is the optimal sparsity coefficient.

Since the  $\ell_0$  norm is difficult to solve, (13) can be approximated as follows:

$$\hat{x} = \arg \min_x \|x\|_1 \text{ s.t. } \|Ax - y\|_2^2 < \varepsilon \quad (14)$$

where  $\|\cdot\|_1$  is the norm of  $\ell_1$ ,  $\|\cdot\|_2$  is the norm of  $\ell_2$ , and  $\varepsilon$  is the error value.

The optimal sparsity coefficient is used to calculate residual

$$r_c(y) = \|y - A\delta_C(\hat{x})\|_2^2 \quad (15)$$

where  $\delta_C(\cdot)$  is the mapping function,  $c$  is the category, and  $\delta_C(\hat{x})$  means that the characteristic function maps  $\hat{x}$  to a new vector. The nonzero elements of this vector are the items in  $\hat{x}$  that are associated with category  $c$ .

The classification is achieved by the smallest residual error

$$\text{Label}(y) = \arg \min_{c \in \{1, 2, \dots, C\}} r_c(y). \quad (16)$$

## III. BROAD SPARSE CLASSIFICATION MODEL BASED ON DICTIONARY SELECTION STRATEGY

To address the problem of fine-grained recognition in image classification, a broad sparse fine-grained image classification model based on dictionary selection strategy, DBSRC is proposed.

### A. BLS Model Based on LOO Parameter Optimization

Due to the differences between different data sets, it is difficult to reduce the error and complexity of BLS in different environments by using only empirical regularization parameters, which makes it difficult for the model to obtain a better classification hyperplane. This will limit the generalization performance of the model. It is both time and labor intensive to use manual tuning or grid search to find the appropriate parameters. LOO can assess the performance of the model under each parameter and realize the optimal regularization parameter selection. More importantly, LOO can use the PRESS statistic (only need to train the model once) to evaluate the MSE of the model under different regularization parameters, which can reduce the computational

complexity. Therefore, this article introduces the LOO into the BLS modeling process to find the optimal regularization parameters, optimize the classification hyperplane, and enhance the generalization performance of the model. The BLS classification process based on LOO (BLS-LOO) can be divided into training and testing phases.

For the training data set  $\{(x_i, y_i)\}_{i=1}^N$  containing  $C$  categories, the training stages of BLS-LOO are as follows.

Construct the nodes matrix of BLS-LOO by (1)–(5)

$$A = [Z^n | H^m] \quad (17)$$

where  $A \in R^{N \times (nk+m)}$ ,  $N$  is the total number of input data,  $n$  is the number of feature nodes groups, the number of feature nodes is  $k$ ,  $m$  is the total number of enhancement nodes.

According to (7), the pseudoinverse of nodal matrix  $A$  needs to be solved by ridge regression algorithm. Therefore, the HAT matrix  $HAT_{BLS}$  in BLS-LOO can be calculated

$$HAT_{BLS} = A(A^T A + \theta I)^{-1} A^T \quad (18)$$

where  $\theta$  is the regularization parameter and  $A \in R^{N \times (nk+m)}$  is the nodes matrix.

To enhance the calculation efficiency, the singular value decomposition is used to decompose the matrix  $A$ , denoted as  $A = UDV^T$ , and the (18) can be transformed into the following:

$$\begin{aligned} HAT_{BLS} &= A(VDU^T UDV^T + \theta I)^{-1} A^T \\ &= AV(D^2 + \theta I)^{-1} V^T A^T \end{aligned} \quad (19)$$

where  $U \in R^{N \times N}$  is unitary matrix,  $D \in R^{N \times (nk+m)}$  is diagonal matrix,  $V^T \in R^{(nk+m) \times (nk+m)}$  is unitary matrix,  $\theta$  is regularization parameter, and  $A \in R^{N \times (nk+m)}$  is node matrix.

From (11) and (19), the  $MSE^{PRESS}$  of BLS-LOO can be expressed

$$MSE^{PRESS} = \frac{1}{N} \sum_{i=1}^N \left( \frac{y_i - \hat{y}_i}{1 - (HAT_{BLS})_{ii}} \right)^2 \quad (20)$$

where  $y_i$  and  $\hat{y}_i$  are the actual output and expected output of the  $i$ th data.  $(HAT_{BLS})_{ii}$  represents the  $i$ th value on the diagonal of the  $HAT_{BLS}$  matrix.

Calculate the  $MSE^{PRESS}$  value of each  $\theta$  in the regularization parameter candidate set  $[\theta_{\min}, \theta_{\max}]$  according to (19) and (20). Choose the  $\theta$  that minimizes the value of  $MSE^{PRESS}$  as the optimal regularization parameter and recorded it as  $\theta_{opt}$

$$\min MSE^{PRESS}$$

$$= \frac{1}{N} \sum_{i=1}^N \left( \frac{y_i - \hat{y}_i}{1 - (AV(D^2 + \theta_{opt} I)^{-1} V^T A^T)_{ii}} \right)^2 \quad (21)$$

where  $\theta_{opt} \in [\theta_{\min}, \theta_{\max}]$ .

Use the optimal regularization parameter  $\theta_{opt}$  to transform the (6)

$$\underset{\beta_{opt}}{\operatorname{argmin}} \theta_{opt} \|\beta_{opt}\|^2 + \|A\beta_{opt} - Y\|^2 \quad (22)$$

where  $Y = [y_1, y_2, \dots, y_N]$  is the label corresponding to the train data.

The optimal target coefficient  $\beta_{opt}$  can be solved by ridge regression

$$\beta_{opt} = (A^T A + \theta_{opt} I)^{-1} A^T Y \quad (23)$$

where  $I$  represents the identity matrix.

Similar to (19), (23) can be changed as follows:

$$\beta_{opt} = V(D^2 + \theta_{opt} I)^{-1} V^T A^T Y. \quad (24)$$

The testing phases of BLS-LOO are as follows.

Similar to (8), construct nodes matrix  $A_{test}$  with test data  $[x_1, x_2, \dots, x_L]$

$$A_{test} = [Z_{test}^n | H_{test}^m] \quad (25)$$

where  $A_{test} \in R^{L \times (nk+m)}$ ,  $L$  is the number of test data.

According to the target coefficient  $\beta_{opt}$  in (22) and the nodes matrix  $A_{test}$  in (8), the classification hyperplane  $O$  can be obtained

$$O = A_{test} \beta_{opt} \quad (26)$$

where  $O \in R^{L \times C}$ ,  $C$  is the total number of data categories.

According to the classification hyperplane  $O$ , the category of the test data can be calculated

$$Y^{test} = \arg \max_{c \in \{1, \dots, C\}} (O) \quad (27)$$

where  $Y^{test} = [y_1^{test}, \dots, y_L^{test}]$ ,  $L$  is the number of test data and  $C$  is the total number of data categories.

The classification process based on BLS-LOO is shown in Algorithm 1.

### B. Broad Sparse Fine-Grained Image Classification Model Based on Dictionary Selection Strategy

To better use the BLS model to tackle the fine-grained recognition problem and improve the ability to handle similarity features, a reliability criterion based on a threshold selection strategy is proposed for the selection of fine-grained images. Then, an adaptive dictionary selection strategy is designed based on the output of BLS-LOO to construct a sparse subdictionary for each fine-grained image that is not distinguished by BLS-LOO. Finally, an SRC model based on the sparse subdictionary is used to classify fine-grained images.

According to (25)–(27), the output matrix  $O$  of the test data set  $X = [x_1, x_2, \dots, x_L]$  can be obtained. For each test data output  $O_i$ ,  $i = 1, \dots, L$ , its label is determined by the position of the maximum value in columns of the output matrix. In other words, in the output vector  $O_i$ ,  $i = 1, \dots, L$ , each column value represents the probability of each category, and the category of the data is consistent with the index of the maximum position.

Therefore, the difference between the first and second largest terms of each output vector can reflect the classification boundary of each sample

$$Td_i = O_i^{\text{first}} - O_i^{\text{second}} \quad (28)$$

---

**Algorithm 1: BLS-LOO.**


---

Input

Training data set  $\{(x_i, y_i)\}_{i=1}^N$  has  $C$  categories, the number of feature node groups  $n$ , the number of feature nodes in each group  $k$ , the number of enhancement nodes  $m$ , and regularization parameter candidate value  $\theta_j \in [\theta_{min}, \theta_{max}]$ . Test data set  $\{(x_i)\}_{i=1}^L$ .

Output

Class label of  $Y^{test} = \{(x_i)\}_{i=1}^L$

**Step 1.** Calculate the nodes matrix  $A = [Z^n | H^m]$ .

**Step 2. For**  $\theta_j \in [\theta_{min}, \theta_{max}]$  **do**

$$HAT_{BLS} = V(D^2 + \theta_j I)^{-1} V^T A^T$$

$$MSE^{PRESS} = \frac{1}{N} \sum_{i=1}^N \left( \frac{y_i - \hat{y}_i}{1 - (HAT_{BLS})_{ii}} \right)^2$$

**End for**

**Step 3.** Obtain  $\theta_{opt}$ .

$$\min MSE^{PRESS} = \frac{1}{N} \sum_{i=1}^N \left( \frac{y_i - \hat{y}_i}{1 - (AV(D^2 + \theta_{opt} I)^{-1} V^T A^T)_{ii}} \right)^2.$$

**Step 4.** Calculate  $\beta_{opt} = V(D^2 + \theta_{opt} I)^{-1} V^T A^T Y$ .

**Step 5.** Calculate  $A_{test} = [Z_{test}^n | H_{test}^m]$ .

**Step 6.** Calculate  $O = A_{test} \beta_{opt}$ .

**Step 7.** Calculate the label of  $Y^{test} = \arg \max_{c \in \{1, \dots, C\}} (O)$ .

---

where  $i = 1, \dots, L$ ,  $O_i^{\text{first}}$  and  $O_i^{\text{second}}$  are the first and second largest values of the output vector  $O_i$ , and  $Td_i$  is the classification boundary. If the  $Td_i$  value is larger, the classification boundary is better, and the accuracy of data classification is higher. If the value is smaller, the classification boundary is worse, and the accuracy of data classification is lower. Therefore, the threshold  $\gamma$  can be set to select noise samples and further improve the effect of data classification. Specifically, if  $Td_i > \gamma$ , BLS-LOO is selected for classification, if  $Td_i \leq \gamma$ , SRC is further selected for classification.

To make full use of the output vector information and adaptively select the correct category for the data with poor boundary, this article further considers the category information of the output vector. Specifically, for the test data assigned to the SRC, this article first obtains the index of the first  $\mu$  largest elements in the output vector. Then select the base atoms with the same label as the first  $\mu$  largest items from the training samples, and adaptively construct suitable sparse representation dictionary for the test data

$$A_{x_i}^* = [A_{c(1)}, A_{c(2)}, \dots, A_{c(\mu)}] \quad (29)$$

where  $A_{x_i}^*$  is an adaptive sparse representation dictionary built for test samples.  $x_i$  is a sample assigned to SRC classification,  $i = 1, \dots, L_{SRC}$  and  $L_{SRC} \leq L$ ,  $L$  is the total number of all test samples.  $c(\mu)$  is the category of the  $\mu$ th largest item position index,  $c(\mu) \leq C$ ,  $C$  is the largest category of data.  $A_{c(\mu)}$  is the data set belonging to the  $c(\mu)$ th category in the training dataset.

**Algorithm 2:** Classification process based on DBSRC.

Input

Training data set  $\{(x_i, y_i)\}_{i=1}^N$  has  $C$  categories, the number of feature node groups  $n$ , the number of feature nodes in each group  $k$ , the number of enhancement nodes  $m$ , and regularization parameter candidate value  $\theta_j \in [\theta_{min}, \theta_{max}]$ . Test data set  $\{(x_i)\}_{i=1}^L$ .

Output

Class label of  $\mathbf{Y}^{test} = \{(x_i)\}_{i=1}^L$

**Step 1.** Use **Algorithm 1** to obtain the output matrix of the test data:  $\mathbf{O} = A_{test}\beta_{opt}$ .

**Step 2. For**  $i = 1$  to  $L$  **do**

```

If  $Td_i = O_i^{first} - O_i^{second} > \gamma$  then
     $\mathbf{y}_i^{test} = \arg \max_c (O)$ 
     $c \in \{1, \dots, C\}$ 
End if
Else if  $Td_i = O_i^{first} - O_i^{second} \leq \gamma$  then
    Get the index of  $\mu$  maxima in  $\mathbf{O}_i$ 
    Construct an adaptive sparse sub-dictionary  $A_{x_i}^* = [A_{c(1)}, A_{c(2)}, \dots, A_{c(\mu)}]$ 
    for  $x_i$ 
    Solve  $\hat{x} = \arg \min_x \|A_{x_i}^* x - x_i\|_2^2 + \tau \|x\|_1$ 
    Calculate  $r_c(x_i) = \|x_i - A_{x_i}^* \delta_c(\hat{x})\|_2^2$ 
    Calculate  $\mathbf{y}_i^{test} = \arg \min_{c \in \{1, 2, \dots, C\}} r_c(x_i)$ 
End else if
End for

```

**Step 3.** Calculate the label of  $\mathbf{Y}^{test} = [y_1^{test}, \dots, y_L^{test}]$ .

The sparsity coefficient of  $x_j$  can be solved

$$\hat{x} = \arg \min_x \|A_{x_i}^* x - x_i\|_2^2 + \tau \|x\|_1 \quad (30)$$

where  $\tau$  is the regularization parameter of SRC and  $\hat{x}$  is the sparse coefficient.

Calculate the residual according to the optimal sparsity coefficient  $\hat{x}$

$$r_c(x_i) = \|x_i - A_{x_i}^* \delta_c(\hat{x})\|_2^2 \quad (31)$$

where  $\delta_c(\hat{x})$  means that the mapping function maps  $\hat{x}$  to a new vector. The nonzero elements of this vector are the items in  $\hat{x}$  that are associated with category  $c$ ,  $c \in \{1, \dots, C\}$ .

Classification is achieved by finding the smallest residual

$$\text{Label}(x_i) = \arg \min_{c \in \{1, 2, \dots, C\}} r_c(x_i). \quad (32)$$

The classification process based on DBSRC is shown in Algorithm 2 and Fig. 2.

**IV. PERFORMANCE EVALUATION OF THE DBSRC**

To assess the performance of the presented DBSRC, three image datasets, ImageNet, USPS, and Pavia, were selected in this article. Also, SVM, ELM, CNN, SRC, BLS, BLS-LOO, ELMSRC, and DBSRC models were selected for comparison.

The experiments were done under the MATLAB 2018b.

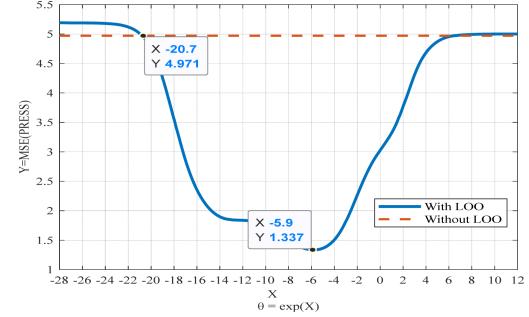


Fig. 3.  $MSE^{PRESS}$  values for different  $\theta$ .

**A. Experiments Results on the ImageNet Dataset**

To evaluate the classification accuracy and efficiency of DBSRC, this article selects the ImageNet dataset.

**1) Dataset Description:** ImageNet<sup>1</sup> is a large-scale image classification data set. Here, we use its subset, including 5 categories (people, motorcycles, dogs, bicycles, and cats), and a total of 7341 samples, as shown in Fig. 1. In addition, each data uses DeCaf [42] for feature extraction, and takes the sixth layer feature, namely DeCaf 6 feature. Each sample has total of 4096 dimensions. In the experiment, 250 samples are taken from the dataset as training set, and the rest are used as test set.

**2) Baseline Selection:** The relevant parameters of SVM<sup>2</sup> are the default values in Lib-SVM. The number of ELM<sup>3</sup> nodes is 2000. The regularization parameter of SRC is 0.1, and the sparse coefficient is calculated by Luo and Zhang [39]. The parameters of the CNN are the default values in the deep learning toolbox. The planar error value of ELMSRC is 0.5 and the rest of the parameters are the same as for ELM and SRC. The regularization parameters of BLS are empirical values,  $e^{-20.7}(2^{-30})$ , and the nodes combination is 300–1500. The regularization parameters of DBSRC are the same as BLS-LOO.

**3) Model Parameter Selection:** The regularization parameters of BLS-LOO are selected in  $[e^{-28}, e^{-27.9}, e^{-27.8}, \dots, e^{12}]$ . The selection process is shown in Fig. 3.

As can be seen from Fig. 3, the regularization parameter of traditional BLS (without LOO) is  $e^{-20.7}$ , and  $MSE^{PRESS}$  is 4.971. The regularization parameter of traditional BLS (without LOO) is  $e^{-5.9}$ , and  $MSE^{PRESS}$  is 1.337. The results show that the  $MSE^{PRESS}$  of BLS-LOO is smaller than that of BLS. Therefore, the experiment sets the regularization parameter of BLS-LOO to  $e^{-5.9}$ . The results show that under the ImageNet dataset, BLS-LOO can find better regularization parameters and has better generalization performance.

The nodes combination of DBSRC is the same as that of BLS-LOO. The nodes selection process of BLS-LOO is shown in Table I.

As can be seen from Table I, when the node combinations are 350–1500 and 300–1750, the test accuracy of BLS-LOO

1.<https://github.com/jindongwang/transferlearning/blob/master/data/dataset.md#imagenet>.

2.[http://www.ntu.edu.sg/home/egbhuang/elm\\_codes.html](http://www.ntu.edu.sg/home/egbhuang/elm_codes.html).

3.<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

TABLE I  
EXPERIMENT RESULTS OF THE BLS-LOO UNDER DIFFERENT NUMBER OF NODES

Feature nodes	Enhancement nodes	Test accuracy(%)	Classification time(s)	STD
100	1500	85.36	0.98	0.42
150	1500	86.04	1.21	0.38
200	1500	86.26	1.44	0.35
250	1500	86.33	1.68	0.28
300	1500	86.46	1.97	0.26
350	1500	86.48	2.30	0.23
400	1500	86.45	2.57	0.24
500	1500	86.42	3.12	0.25
300	500	86.20	0.54	0.35
300	750	86.32	0.84	0.32
300	1000	86.44	1.16	0.24
300	1250	86.42	1.52	0.28
300	1750	86.48	2.32	0.28
300	2000	86.46	2.75	0.28
300	2500	86.41	3.20	0.24
300	3000	86.35	4.07	0.25

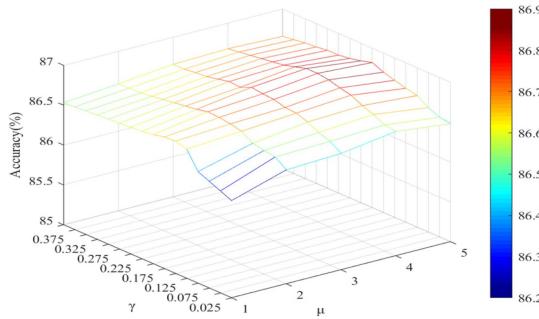


Fig. 4. Test accuracy of DBSRC under different combinations of  $(\gamma, \mu)$ .

is 86.48%, which is better than the other node combinations. When the node combination is 300–500, the BLS-LOO time is the shortest. BLS-LOO has the smallest standard deviation (STD) value and the best stability when the node combination is 350–1500. Comprehensive analysis of the experimental results, the experiment sets the nodes combination of BLS-LOO to 300–1500.

In order to analyze the influence of the combination of threshold  $\gamma$  and adaptive dictionary dimension  $\mu$  on DBSRC test accuracy and classification time, the experiment set the range of  $\gamma$  to  $[0.025, 0.05, 0.075, \dots, 0.40]$ , and the range of  $\mu$  to  $[1, 2, \dots, 5]$ . The results of test accuracy and classification time are shown in Figs. 4 and 5, respectively.

As can be seen from Figs. 4 and 5, when the threshold  $\gamma$  is fixed, as the dimension  $\mu$  of the adaptive subdictionary increases, the classification accuracy will first increase and then decrease. Whether increasing the threshold  $\gamma$  or increasing the dictionary dimension  $\mu$ , the classification time of DBSRC will increase. So DBSRC can flexibly select the combination of  $\gamma$  and  $\mu$  to obtain the best classification accuracy or better classification efficiency. The combined effect of threshold  $\gamma$  and dictionary dimension  $\mu$  on the performance of DBSRC, the experiment selects  $(\gamma, \mu)$  as  $(0.1, 4)$ .

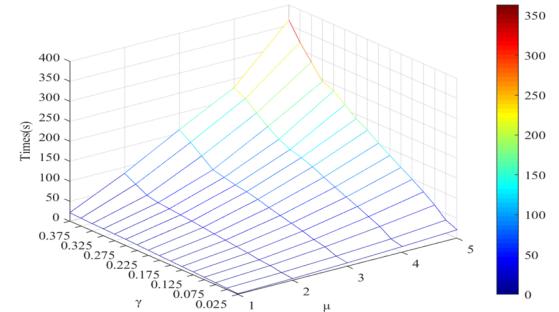


Fig. 5. Classification time of DBSRC under different combinations of  $(\gamma, \mu)$ .

TABLE II  
CLASSIFICATION RESULTS FOR FINE GRAINED IMAGES

Methods	Number of fine-grained images	Number of correct predictions	Classification accuracy(%)
BLS-LOO	377	316	83.82
DBSRC	377	334	88.59

TABLE III  
EXPERIMENT COMPARISON RESULTS OF EACH METHOD ON IMAGENET DATA

Methods	Test accuracy(%)	Classification time(s)	STD
SVM	84.05	14.91	<b>0.00</b>
ELM	81.63	3.81	0.37
CNN	74.27	276.56	0.50
SRC	86.67	550.90	0.19
BLS	82.33	1.06	0.48
BLS-LOO	86.46	1.97	0.26
ELMSRC	85.92	78.58	0.25
<b>DBSRC</b>	<b>86.70</b>	52.63	0.22

4) *Experimental Results and Analysis:* The effectiveness of DBSRC and BLS-LOO for classifying fine-grained images was experimentally compared at  $(\gamma, \mu)$  combinations of  $(0.1, 4)$ . The results are shown in Table II.

As can be seen from Table II, the number of correct predictions is 334 for DBSRC and 316 for BLS-LOO for the combination of  $(\gamma, \mu)$  as  $(0.1, 4)$ . The results show that DBSRC can better enhance the classification accuracy of fine-grained images.

The performance of SVM, ELM, CNN, SRC, BLS, BLS-LOO, ELMSRC, and DBSRC was compared under the same conditions. Each method was running 10 times independently. The experiment selects three indicators of average test accuracy, average classification time and average test accuracy STD for comparison. The experimental results are shown in Table III.

As can be seen from Table III that the classification time of BLS is 1.06 s, which is faster than other classification models. The stability of SVM is better than other classification models. The classification accuracy of DBSRC is 86.70%, which is better than SVM, ELM, SRC, BLS, BLS-LOO, and ELMSRC. The classification time of DBSRC is 52.63 s, which is 526.54 s less than SRC. The STD value of DBSRC is 0.22, which is smaller than ELM, BLS, and BLS-LOO. The experimental results show that DBSRC has better classification effect and efficiency on ImageNet dataset.



Fig. 6. USPS dataset (the number 0 in the green box and the number 6 in the red box have a high similarity feature).

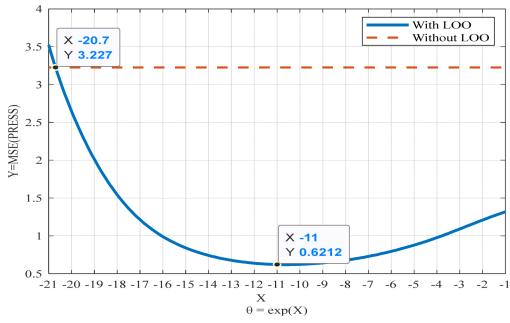


Fig. 7.  $MSE_{PPRESS}$  values for different  $\theta$ .

### B. Experiments Results on the USPS Dataset

The USPS data set is selected to verify the effectiveness of DBSRC.

1) *Dataset Description*: The USPS<sup>4</sup> dataset consists of 0–9 handwritten digital images, with a total of 9298 images (as shown in Fig. 6). Each number is a grayscale image with  $16 \times 16$  pixels. In the experiment, 2000 image data (200 in each category) are selected as the training set, and the remaining 7298 image data are used as the test set. Here, each column in the original USPS image will be directly used as a feature.

2) *Baseline Selection*: The relevant parameters of SVM are the default values in Lib-SVM toolbox. The number of ELM nodes is 3000. The regularization parameter of SRC is 0.1, and the sparse coefficient is calculated by Luo and Zhang [39]. The relevant parameters for the CNN are empirical values from the deep learning toolbox. The plane error value for ELMSRC is 0.4 and the rest of the parameters are the same as for ELM and SRC. The regularization parameters of BLS are empirical values,  $e^{-20.7}(2^{-30})$ , and the node combination is 100–2000. The regularization parameters of DBSRC are the same as BLS-LOO.

3) *Model Parameter Selection*: The regularization parameters of BLS-LOO are selected in  $[e^{-21}, e^{-20.9}, e^{-20.8}, \dots, e^{-1}]$ . The selection process is shown in Fig. 7.

4.<https://cs.nyu.edu/~roweis/data.html>.

TABLE IV  
EXPERIMENT RESULTS OF THE BLS-LOO UNDER DIFFERENT NUMBER OF NODES

Feature nodes	Enhancement nodes	Test accuracy(%)	Classification time(s)	STD
50	2000	96.49	2.75	0.44
100	2000	96.57	3.22	0.27
150	2000	96.59	3.75	0.28
200	2000	96.60	4.26	0.30
250	2000	96.58	4.78	0.33
350	2000	96.55	5.64	0.26
500	2000	96.52	7.16	0.29
100	500	96.21	0.56	0.35
100	1000	96.44	1.13	0.32
100	1500	96.52	1.84	0.33
100	2500	96.60	5.86	0.30
100	3000	96.63	8.39	0.36
100	3500	96.61	11.74	0.28
100	4000	96.58	14.33	0.31
100	5000	96.55	19.78	0.35

As can be seen from Fig. 7, with the increase of  $\theta$ , the  $MSE_{PPRESS}$  of BLS-LOO (with LOO) will decrease first and then increase. When  $\theta$  is  $e^{-11}$ ,  $MSE_{PPRESS}$  is 0.6212, which is less than other values  $\theta$ . Therefore, under the USPS data set, the optimal  $\theta$  value of BLS-LOO is selected as  $e^{-11}$ . The experimental results show that LOO can optimize the  $\theta$  of BLS and enhance the generalization performance of the model.

The nodes combination of DBSRC is the same as that of BLS-LOO. The nodes selection process of BLS-LOO is shown in Table IV.

As can be seen from Table IV that both feature and enhancement nodes will affect the performance of BLS-LOO. If the test accuracy of BLS-LOO is high, the node combination is selected as 100–3000. If the classification time of BLS-LOO is required to be short, the combination of feature node and enhancement node is 100–500. If the model is required to be stable, 350–2000 (node combination) is preferred. According to the influence of each node combination on BLS-LOO, 100–2000 (node combination) is chosen in the experiment.

To analyze the influence of the combination of threshold  $\gamma$  and adaptive dictionary dimension  $\mu$  on DBSRC test accuracy and classification time, the experiment set the range of  $\gamma$  to  $[0.01, 0.05, 0.10, 0.15, \dots, 0.45, 0.50]$ , and the range of  $\mu$  to  $[1, 2, \dots, 10]$ . The results of test accuracy and classification time are shown in Figs. 8 and 9, respectively.

As can be seen from Figs. 8 and 9, when  $\gamma$  is fixed and the value of  $\mu$  is 1–6, the test accuracy of DBSRC will increase, when  $\gamma$  is fixed and  $\mu$  value is greater than 6, the test accuracy of DBSRC will show a steady trend. This also verifies the importance of adaptive subdictionary dimension selection. When the value of  $\mu$  is fixed and the value of  $\gamma$  is 0.35, the test accuracy of DBSRC is the best. When the values of  $\gamma$  and  $\mu$  increase at the same time, the classification time of DBSRC will be significantly improved. According to the influence of the combination of  $\gamma$  and  $\mu$  on the accuracy and efficiency of DBSRC, the experiment sets  $\gamma$  and  $\mu$  to 0.35 and 2, respectively.

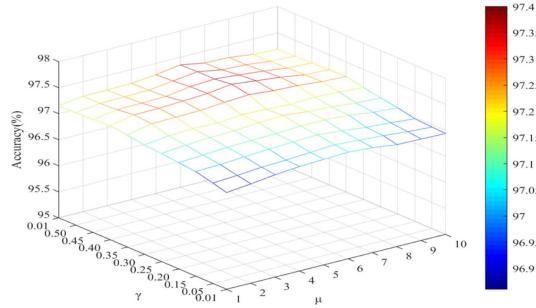


Fig. 8. Test accuracy of DBSRC under different combinations of  $(\gamma, \mu)$ .

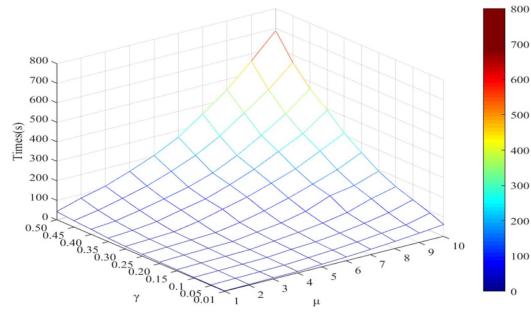


Fig. 9. Classification time of DBSRC under different combinations of  $(\gamma, \mu)$ .

TABLE V  
CLASSIFICATION RESULTS FOR FINE GRAINED IMAGES

BLS-LOO	494	407	82.38
DBSRC	494	458	92.71

*4) Experimental Results and Analysis:* The effectiveness of DBSRC and BLS-LOO for classifying fine-grained images was experimentally compared at  $(\gamma, \mu)$  combinations of  $(0.35, 2)$ . The results are shown in Table V.

As can be seen from Table V that the classification accuracy of DBSRC is 10.33% higher than that of BLS-LOO among the fine-grained images selected by the threshold selection strategy. The experimental results show that DBSRC can efficiently enhance the classification accuracy of fine-grained images.

The performance of SVM, ELM, CNN, SRC, BLS, BLS-LOO, ELMSRC, and DBSRC was compared under the same conditions. Each method was running 10 times independently. The experiment selects three indicators of average test accuracy, average classification time and average test accuracy STD for comparison. The experimental results are shown in Table VI.

As can be seen from Table VI that the STD value of BLS-LOO is 0.27, which is higher than SVM, SRC, and DBSRC. The classification time of SRC is the longest, 994.33 s, while the classification time of DBSRC is only 9.79 s. The CNN was restricted by the number of training samples, resulting in the highest test accuracy of 82.68%. The classification accuracy of ELMSRC was 96.60%, which was 3.97% better than that of ELM. However, the classification time of ELMSRC was 106.45 s, which was much higher than ELM. The classification accuracy of DBSRC is slightly better than that of SRC. The

TABLE VI  
EXPERIMENT COMPARISON RESULTS ON USPS DATA

Methods	Test accuracy(%)	Classification time(s)	STD
SVM	92.27	3.36	<b>0.00</b>
ELM	92.63	1.51	1.40
CNN	82.68	528.67	1.58
SRC	97.19	994.33	0.10
BLS	93.31	0.72	1.35
BLS-LOO	96.57	2.22	0.27
ELMSRC	96.60	106.45	0.37
DBSRC	<b>97.28</b>	9.79	0.12

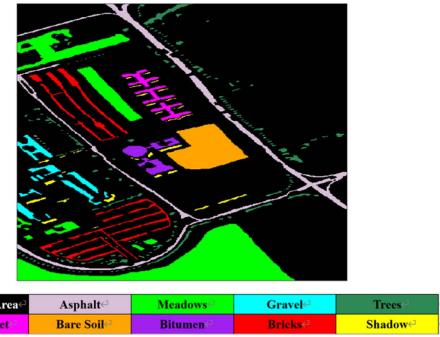


Fig. 10. Ground truth map of Pavia (asphalt, bricks, and gravel have high similarity features).

classification time of BLS is the shortest, only 0.72 s, but the test accuracy of BLS is 3.26% and 3.97% lower than that of BLS-LOO and DBSRC. The results show that DBSRC has good classification performance under USPS data.

### C. Experiments Results on the Pavia Dataset

To further verify the effectiveness of DBSRC, Pavia dataset is selected in this section.

*1) Dataset Description:* Pavia<sup>5</sup> is the scene data of Pavia University collected by the reflective optical system imaging spectrometer. The data set includes 9 classes, with a total of about 42776 samples. Pavia ground reference information is shown in Fig. 10. The experiment in this section uses the combination of the principal component analysis and the local binary pattern proposed by Li et al. [43] to perform dimensionality reduction and feature extraction on the data. After processing, each sample has 1770 dimensions. The division of data training set and test set is shown in Table VII.

*2) Baseline Selection:* The relevant parameters of SVM are the default values in Lib-SVM. The number of ELM nodes is 3000. The regularization parameter of SRC is 0.1, and the sparse coefficient is calculated by Luo and Zhang [39]. The relevant parameters for the CNN are empirical values from the deep learning toolbox. The plane error value for ELMSRC is 0.3 and the rest of the parameters are the same as for ELM and SRC. The regularization parameters of BLS are empirical values,  $e^{-20.7}(2^{-30})$ , and the nodes combination is 400–2000.

<sup>5</sup>[https://www.ehu.eus/ccwintco/index.php/Hyperspectral\\_Remote\\_Sensing\\_Scenes#Pavia\\_University\\_scene](https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes#Pavia_University_scene)

TABLE VII  
CLASS LABELS AND TRAIN-TEST DISTRIBUTION OF SAMPLES FOR THE PAVIA DATASET

Class		Numbers of samples	
No.	Name	Training-set	Testing-set
1	Asphalt	150	6481
2	Meadows	150	18499
3	Gravel	150	1949
4	Trees	150	2914
5	Metal Sheets	150	1195
6	Bare Soil	150	4879
7	Bitumen	150	1180
8	Bricks	150	3532
9	Shadow	150	797
Total		1350	41426

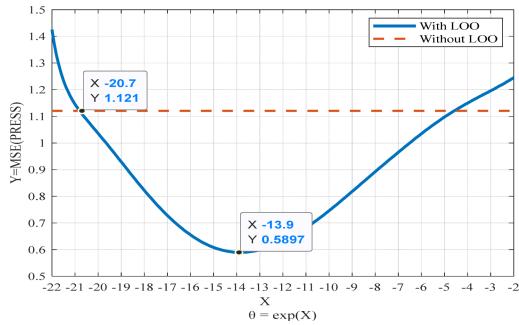


Fig. 11.  $\text{MSE}^{\text{PRESS}}$  values for different  $\theta$ .

3) *Model Parameter Selection*: The regularization parameters of DBSRC are the same as BLS-LOO. The regularization parameters of BLS-LOO are selected in  $[e^{-22}, e^{-21.9}, e^{-21.8}, \dots, e^{-2}]$ . The selection process is shown in Fig. 11.

As can be seen from Fig. 11, as the value of  $\theta$  is  $e^{-13.9}$ , the  $\text{MSE}^{\text{PRESS}}$  of BLS-LOO is 0.5897, which is much lower than the baseline of BLS (without LOO). The results show that BLS-LOO can optimize the parameters and obtain the regularization parameter that minimized  $\text{MSE}^{\text{PRESS}}$ . Therefore, the optimal regularization parameter  $\theta$  of BLS-LOO is selected as  $e^{-13.9}$ .

The nodes combination of DBSRC is the same as that of BLS-LOO. The nodes selection process of BLS-LOO is shown in Table VIII.

As can be seen from Table VIII that with the increase of feature or enhancement nodes, the classification time of BLS-LOO will increase. When the feature nodes are fixed and the number of enhancement nodes is greater than 2250, the increase of test accuracy is limited. When the enhancement node is fixed, with the increase of feature nodes, the test accuracy will increase first and then flatten. According to the influence of different node combinations on the performance of BLS-LOO, the experiment selects 400 feature nodes and 2000 enhancement nodes, i.e., 400–2000.

TABLE VIII  
EXPERIMENT RESULTS OF THE BLS-LOO UNDER DIFFERENT NUMBER OF NODES

Feature nodes	Enhancement nodes	Test accuracy(%)	Classification time(s)	STD
100	2000	95.21	7.14	0.36
150	2000	95.33	7.72	0.33
200	2000	95.54	8.27	0.27
250	2000	95.72	8.68	0.28
300	2000	95.90	9.22	0.25
350	2000	96.06	9.84	0.20
400	2000	96.20	10.33	0.19
450	2000	96.22	11.07	0.24
500	2000	96.20	12.13	0.26
400	500	95.74	4.44	0.24
400	750	95.88	5.02	0.21
400	1000	96.02	5.63	0.18
400	1250	96.08	6.27	0.25
400	1500	96.11	6.72	0.26
400	1750	96.15	7.28	0.28
400	2250	96.25	8.42	0.24
400	2500	96.26	9.23	0.20
400	3000	96.25	11.07	0.19
400	5000	96.24	18.79	0.25

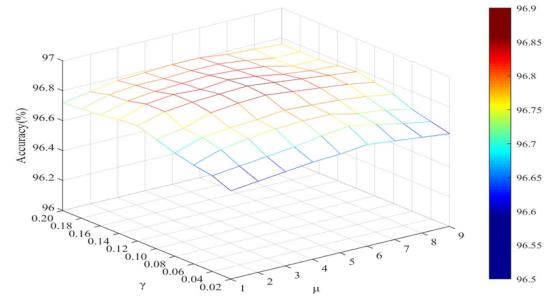


Fig. 12. Test accuracy of DBSRC under different combinations of  $(\gamma, \mu)$ .

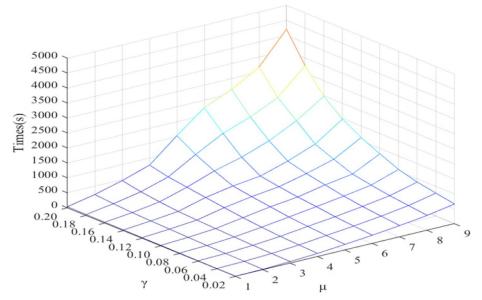


Fig. 13. Classification time of DBSRC under different combinations of  $(\gamma, \mu)$ .

To analyze the influence of the combination of threshold  $\gamma$  and adaptive dictionary dimension  $\mu$  on DBSRC test accuracy and classification time, the experiment set the range of  $\gamma$  to  $[0.02, 0.04, \dots, 0.20]$ , and the range of  $\mu$  to  $[1, 2, \dots, 9]$ . The results of test accuracy and classification time are shown in Figs. 12 and 13, respectively.

As can be seen from Figs. 12 and 13, with the combination of  $\gamma$  and  $\mu$  changes, the performance of DBSRC will be affected.

TABLE IX  
CLASSIFICATION RESULTS FOR FINE GRAINED IMAGES

Methods	Number of fine-grained images	Number of correct predictions	Classification accuracy(%)
BLS-LOO	1012	709	70.06
DBSRC	1012	956	94.47

TABLE X  
EXPERIMENT COMPARISON RESULTS OF EACH METHOD ON PAVIA DATA

Methods	Test accuracy(%)	Classification time(s)	STD
SVM	94.77	36.20	<b>0.00</b>
ELM	91.31	14.56	0.36
CNN	74.65	850.02	0.21
SRC	96.73	13985.39	0.18
BLS	93.25	3.87	0.35
BLS-LOO	96.20	10.33	0.27
ELMSRC	95.94	1639.94	0.26
DBSRC	<b>96.80</b>	107.60	0.24

With the increase of  $\gamma$  and  $\mu$ , the classification time of DBSRC will increase. When  $\gamma$  is fixed and  $\mu$  is greater than 4, the DBSRC classification time increases even more. When  $\gamma$  is fixed and  $\mu$  is 5, the test accuracy of DBSRC is better than other values. When  $\gamma$  and  $\mu$  are 0.12 and 5, respectively, the classification accuracy of DBSRC is better, which is 96.88%. According to the experimental results,  $\gamma$  and  $\mu$  of DBSRC are selected as 0.14 and 2, respectively.

4) *Experimental Results and Analysis:* The effectiveness of DBSRC and BLS-LOO for classifying fine-grained images was experimentally compared at  $(\gamma, \mu)$  combinations of (0.14, 2). The results are shown in Table IX.

As can be seen from Table IX that the number of correctly classified samples for BLS-LOO is 709 and the classification accuracy is 70.06%, while the number of correctly classified samples for DBSRC is 956 and the classification accuracy is 94.47%. The results show that DBSRC can perform the classification task of fine-grained images better.

The performance of SVM, ELM, CNN, SRC, BLS, BLS-LOO, ELMSRC, and DBSRC was compared under the same conditions. Each method was running 10 times independently. The experiment selects three indicators of average test accuracy, average classification time and average test accuracy STD for comparison. The experimental results are shown in Table X.

As can be seen from Table X that the STD value of ELM is 0.36, which is higher than other classification models. The classification time of BLS is 3.87 s, which is shorter than other classification models. DBSRC has the highest test accuracy, reaching 96.80%. The test accuracy of BLS-LOO is 96.20%, which is 1.43%, 4.89%, 2.95%, 21.55%, and 0.26% higher than SVM, ELM, BLS, CNN, and ELMSRC, respectively. The test accuracy of SRC is 96.73%, which is only slightly lower than that of DBSRC. But the classification time of SRC is 13985.39 s, which is much longer than DBSRC. The results show that DBSRC can complete the classification task under the Pavia dataset with the best test accuracy, better efficiency, and stability.

#### D. Analysis of Experiment Results

As can be seen from Figs. 3, 7, and 11, BLS-LOO can be used to optimize the regularization parameters with LOO under three different datasets, ImageNet, USPS, and Pavia. Compared with the empirical regularization parameters of BLS, the regularization parameters selected by LOO will get smaller MSE<sup>PRESS</sup>, so that the model has better performance.

As can be seen from Tables I, IV, and VIII that with the increase of feature or enhancement nodes, the classification accuracy of BLS-LOO will increase first and then flatten. Therefore, in practical application, the number of nodes can be flexibly selected according to the specific needs, the model structure can be formed reasonably, and the ideal classification can be obtained.

As can be seen from Figs. 4, 8, and 12, the  $(\gamma, \mu)$  combinations of DBSRC were chosen as (0.14), (0.35, 2), and (0.14, 2) under the three datasets of ImageNet, USPS, and Pavia, respectively. The results show that under the classification tasks of different data sets, DBSRC can select the appropriate  $\gamma$  and  $\mu$  according to the characteristics of the data to obtain the ideal classification accuracy. In addition, when  $\gamma$  is fixed, as the value of  $\mu$  increases, the redundancy of the dictionary will increase, it difficult to continuously improve the accuracy.

As can be seen from Figs. 5, 9, and 13, when  $\gamma$  is fixed, as  $\mu$  increases, the classification time of DBSRC will increase. The reason is that as the  $\mu$  increases, the dimensionality of the subdictionary increases, and the sample query time also increases. When  $\mu$  is fixed, as  $\gamma$  increases, the classification time of DBSRC will continue to increase. The main reason is that the increase of  $\gamma$  value will lead to the increase of the number of samples requiring SRC for further classification. The results show that DBSRC can use adaptive subdictionary selection strategy to reduce dictionary redundancy and obtain ideal classification results.

As can be seen from Tables II, V, and IX that the classification accuracy of DBSRC is 4.77%, 10.33%, and 24.41% higher than BLS-LOO for fine-grained images selected from the ImageNet, USPS, and Pavia datasets. The main reason for this is that DBSRC is able to classify fine-grained images selected based on the threshold selection strategy by constructing an SRC model based on the subdictionary selection strategy, thus, better performing the classification task in fine-grained images.

As can be seen from Tables III, VI, and X that DBSRC achieves better classification performance than other methods for three datasets with different complex dimensions (4096 for ImageNet, 256 for USPS and 1771 for Pavia). This indicates that DBSRC has better stability and better adaptability. Compared to the shallow models ELM and SVM, BLS has higher test accuracy and greater stability due to its ability to extract data features by constructing feature nodes and enhancement nodes to reduce data redundancy. Compared to deep network CNNs, BLS has higher classification accuracy and better efficiency. This is because CNNs are easily restricted by the number of samples in the modeling process, and the parameters of CNNs need to be calculated by gradient descent in multiple iterations. In contrast, the performance of BLS is less affected by the number of samples

and its output weights are obtained using a ridge regression algorithm that performs one iteration. Compared with ELM and SRC, ELMSRC achieves good classification results with better classification efficiency because the method effectively combines the efficiency advantage of ELM with the high robustness of SRC. Compared with BLS, BLS-LOO is able to complete the classification task with higher testing accuracy. This is because BLS-LOO is able to optimize the regularization parameters of the model by LOO, which results in a better classification hyperplane and improves the classification performance of the model. Compared to BLS-LOO, DBSRC has better processing power for fine-grained features, resulting in better classification performance. Compared with ELMSRC, DBSRC can not only obtain a better classification hyperplane through regularized parameter search, but also reduce dictionary redundancy through dictionary selection strategy, thus achieving fast and accurate query of test samples and improving classification effectiveness and efficiency.

## V. CONCLUSION

Aiming at the problem of fine-grained recognition in image classification, a broad sparse fine-grained image classification model based on dictionary selection strategy, DBSRC, is proposed. LOO is introduced in the BLS modeling process to find the better regularization parameters to obtain a better classification hyperplane. A reliability criterion based on the threshold selection strategy is proposed for fine-grained image selection. An adaptive dictionary selection strategy is designed based on the output of BLS-LOO to construct a sparse subdictionary for each fine-grained image. A sparse subdictionary based SRC model is used to classify fine-grained images.

The effectiveness of DBSRC was verified on three image datasets of different complex dimensions, ImageNet, USPS, and Pavia, respectively. The results show that DBSRC achieves good classification performance on all three datasets, with strong adaptability. Compared with SVM, ELM, CNN, SRC, BLS, ELM-SRC, and BLS-LOO, DBSRC is more capable of processing fine-grained images and can achieve better classification results with better classification efficiency.

Future work includes further improvements to DBSRC, enhancing DBSRC's ability to distinguish similarity features and interpreting the extracted features.

## REFERENCES

- [1] R. Thoreau, V. Achard, L. Risser, B. Berthelot, and X. Briottet, "Active learning for hyperspectral image classification: A comparative review," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 3, pp. 256–278, Sep. 2022, doi: [10.1109/MGRS.2022.3169947](https://doi.org/10.1109/MGRS.2022.3169947).
- [2] H. Chen, T. Wang, T. Chen, and W. Deng, "Hyperspectral image classification based on fusing S3-PCA, 2D-SSA and random patch network," *Remote Sens.*, vol. 15, no. 13, 2023, Art. no. 3402.
- [3] C. Yu, B. Gong, M. Song, E. Zhao, and C.-I. Chang, "Multiview calibrated prototype learning for few-shot hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2022, Art. no. 5544713, doi: [10.1109/TGRS.2022.3225947](https://doi.org/10.1109/TGRS.2022.3225947).
- [4] Y. Song, G. Zhao, B. Zhang, H. Chen, W. Q. Deng, and W. Deng, "An enhanced distributed differential evolution algorithm for portfolio optimization problems," *Eng. Appl. Artif. Intell.*, vol. 121, 2023, Art. no. 106004.
- [5] X.-S. Wei et al., "Fine-grained image analysis with deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 8927–8948, Dec. 2022, doi: [10.1109/TPAMI.2021.3126648](https://doi.org/10.1109/TPAMI.2021.3126648).
- [6] Y. Zhang et al., "Weakly supervised fine-grained categorization with part-based image representation," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1713–1725, Apr. 2016, doi: [10.1109/TIP.2016.2531289](https://doi.org/10.1109/TIP.2016.2531289).
- [7] S. Min, H. Yao, H. Xie, Z.-J. Zha, and Y. Zhang, "Multi-objective matrix normalization for fine-grained visual recognition," *IEEE Trans. Image Process.*, vol. 29, pp. 4996–5009, Mar. 2020, doi: [10.1109/TIP.2020.2977457](https://doi.org/10.1109/TIP.2020.2977457).
- [8] X. Li, H. Zhao, and W. Deng, "BFOD: Blockchain-based privacy protection and security sharing scheme of flight operation data," *IEEE Internet Things J.*, to be published, doi: [10.1109/IJOT.2023.3296460](https://doi.org/10.1109/IJOT.2023.3296460).
- [9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [10] N. He, J. Cao, and L. Song, "Scale space histogram of oriented gradients for human detection," in *Proc. Int. Symp. Inf. Sci. Eng.*, 2008, pp. 167–170, doi: [10.1109/ISISE.2008.308](https://doi.org/10.1109/ISISE.2008.308).
- [11] A. J. Siddiqui, A. Mammeri, and A. Boukerche, "Real-time vehicle make and model recognition based on a bag of surf features," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 11, pp. 3205–3219, Nov. 2016.
- [12] T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, and Z. Zhang, "The application of two-level attention models in deep convolutional neural network for fine-grained image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 842–850.
- [13] Y. Wang, V. I. Morariu, and L. S. Davis, "Learning a discriminative filter bank within a CNN for fine-grained recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4148–4157.
- [14] X. Chen, H. Shao, Y. Xiao, S. Yan, B. Cai, and B. Lius, "Collaborative fault diagnosis of rotating machinery via dual adversarial guided unsupervised multi-domain adaptation network," *Mech. Syst. Signal Process.*, vol. 198, 2023, Art. no. 110427.
- [15] C. L. P. Chen and Z. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 10–24, Jan. 2018.
- [16] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.
- [17] E. Cambria et al., "Extreme learning machines [trends & controversies]," *IEEE Intell. Syst.*, vol. 28, no. 6, pp. 30–59, Nov./Dec. 2013, doi: [10.1109/MIS.2013.140](https://doi.org/10.1109/MIS.2013.140).
- [18] U. Maulik and D. Chakraborty, "Remote sensing image classification: A survey of support-vector-machine-based advanced techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 33–52, Mar. 2017, doi: [10.1109/MGRS.2016.2641240](https://doi.org/10.1109/MGRS.2016.2641240).
- [19] J. Lin et al., "PDBL: Improving histopathological tissue classification with plug-and-play pyramidal deep-broad learning," *IEEE Trans. Med. Imag.*, vol. 41, no. 9, pp. 2252–2262, Sep. 2022, doi: [10.1109/TMI.2022.3161787](https://doi.org/10.1109/TMI.2022.3161787).
- [20] R. Ali et al., "Optic disk and cup segmentation through fuzzy broad learning system for glaucoma screening," *IEEE Trans. Ind. Inform.*, vol. 17, no. 4, pp. 2476–2487, Apr. 2021, doi: [10.1109/TII.2020.3000204](https://doi.org/10.1109/TII.2020.3000204).
- [21] Y. Zhou, J. Du, K. Guan, and T. Wang, "Multi-modal broad learning system for medical image and text-based classification," in *Proc. IEEE 43rd Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2021, pp. 3439–3442, doi: [10.1109/EMBC46164.2021.9630854](https://doi.org/10.1109/EMBC46164.2021.9630854).
- [22] J. Fan, X. Wang, X. Wang, J. Zhao, and X. Liu, "Incremental wishart broad learning system for fast PolSAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1854–1858, Dec. 2019, doi: [10.1109/LGRS.2019.2913999](https://doi.org/10.1109/LGRS.2019.2913999).
- [23] Y. Ma, Z. Liu, and C. L. P. Chen, "Multiscale random convolution broad learning system for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, May 2022, Art. no. 5503605, doi: [10.1109/LGRS.2021.3060876](https://doi.org/10.1109/LGRS.2021.3060876).
- [24] L. Guo, R. Li, and B. Jiang, "An ensemble broad learning scheme for semisupervised vehicle type classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 12, pp. 5287–5297, Dec. 2021, doi: [10.1109/TNNLS.2021.3083508](https://doi.org/10.1109/TNNLS.2021.3083508).
- [25] C. L. P. Chen and B. Wang, "Random-positioned license plate recognition using hybrid broad learning system and convolutional networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 444–456, Jan. 2022, doi: [10.1109/TITS.2020.3011937](https://doi.org/10.1109/TITS.2020.3011937).
- [26] C. Huang, X. Zhou, X. Ran, J. Wang, H. Chen, and W. Deng, "Adaptive cylinder vector particle swarm optimization with differential evolution for UAV path planning," *Eng. Appl. Artif. Intell.*, vol. 121, 2023, Art. no. 105942.

- [27] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, Jul. 2009.
- [28] M. Li, J. Zhang, J. Song, Z. Li, and S. Lu, "A clinical-oriented non severe depression diagnosis method based on cognitive behavior of emotional conflict," *IEEE Trans. Comput. Social Syst.*, vol. 10, no. 1, pp. 131–141, Feb. 2023, doi: [10.1109/TCSS.2022.3152091](https://doi.org/10.1109/TCSS.2022.3152091).
- [29] Q. Gao, Q. Wang, Y. Huang, X. Gao, X. Hong, and H. Zhang, "Dimensionality reduction by integrating sparse representation and fisher criterion and its applications," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5684–5695, Dec. 2015, doi: [10.1109/TIP.2015.2479559](https://doi.org/10.1109/TIP.2015.2479559).
- [30] C. Zou, K. I. Kou, and Y. Wang, "Quaternion collaborative and sparse representation with application to color face recognition," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3287–3302, Jul. 2016, doi: [10.1109/TIP.2016.2567077](https://doi.org/10.1109/TIP.2016.2567077).
- [31] B. Zhang, B. V. K. Vijaya kumar, and D. Zhang, "Noninvasive diabetes mellitus detection using facial block color with a sparse representation classifier," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1027–1033, Apr. 2014, doi: [10.1109/TBME.2013.2292936](https://doi.org/10.1109/TBME.2013.2292936).
- [32] T. Dundar and T. Ince, "Sparse representation-based hyperspectral image classification using multiscale superpixels and guided filter," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 246–250, Feb. 2019, doi: [10.1109/LGRS.2018.2871273](https://doi.org/10.1109/LGRS.2018.2871273).
- [33] C. Li, Y. Ma, X. Mei, C. Liu, and J. Ma, "Hyperspectral image classification with robust sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 641–645, May 2016, doi: [10.1109/LGRS.2016.2532380](https://doi.org/10.1109/LGRS.2016.2532380).
- [34] C. Yu, S. Zhou, M. Song, B. Gong, E. Zhao, and C.-I. Chang, "Unsupervised hyperspectral band selection via hybrid graph convolutional network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jun. 2022, Art. no. 5530515.
- [35] B. Song, P. Li, and X. Jia, "New feature selection methods using sparse representation for one-class classification of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 10, pp. 1761–1765, Oct. 2021, doi: [10.1109/LGRS.2020.3006830](https://doi.org/10.1109/LGRS.2020.3006830).
- [36] C. Sun, "Fine-grained image classification of cashmere wool based on sparse dictionary learning," in *Proc. Int. Conf. Commun., Inf. Syst. Comput. Eng.*, 2021, pp. 326–329, doi: [10.1109/CISCE52179.2021.9446019](https://doi.org/10.1109/CISCE52179.2021.9446019).
- [37] M. Srinivas, Y.-Y. Lin, and H.-Y. M. Liao, "Learning deep and sparse feature representation for fine-grained object recognition," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2017, pp. 1458–1463, doi: [10.1109/ICME.2017.8019386](https://doi.org/10.1109/ICME.2017.8019386).
- [38] C. Zhang, C. Liang, L. Li, J. Liu, Q. Huang, and Q. Tian, "Fine-grained image classification via low-rank sparse coding with general and class-specific codebooks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 7, pp. 1550–1559, Jul. 2017, doi: [10.1109/TNNLS.2016.2545112](https://doi.org/10.1109/TNNLS.2016.2545112).
- [39] M. X. Luo and K. Zhang, "A hybrid approach combining extreme learning machine and sparse representation for image classification," *Eng. Appl. Artif. Intell.*, vol. 27, pp. 228–235, 2014.
- [40] X. Hong, S. Chen, J. Gao, and C. J. Harris, "Nonlinear identification using orthogonal forward regression with nested optimal regularization," *IEEE Trans. Cybern.*, vol. 45, no. 12, pp. 2925–2936, Dec. 2015.
- [41] Z. Shao, M. J. Er, and N. Wang, "An efficient leave-one-out cross-validation-based extreme learning machine (ELOO-ELM) with minimal user intervention," *IEEE Trans. Cybern.*, vol. 46, no. 8, pp. 1939–1951, Aug. 2016.
- [42] J. Donahue et al., "DeCAF: A deep convolutional activation feature for generic visual recognition," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 647–655.
- [43] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, Jul. 2015.