Hindawi
Complexity
Volume 2019, Article ID 6876173, 23 pages
https://doi.org/10.1155/2019/6876173

Research Article

A Novel Semi-Supervised Learning Method Based on Fast Search and Density Peaks

Fei Gao,[1] **Teng Huang**,[1] **Jinping Sun**,[1] **Amir Hussain**,[2] **Erfu Yang**,[3] **and Huiyu Zhou**[4]

[1] *School of Electronic and Information Engineering, Beihang University, Beijing 101191, China*
[2] *Cognitive Big Data and Cyber-Informatics (CogBID) Laboratory, School of Computing, Edinburgh Napier University, Edinburgh EH10 5DT, Scotland, UK*
[3] *Department of Design, Manufacture and Engineering Management, University of Strathclyde, Glasgow G1 1XJ, UK*
[4] *Department of Informatics, University of Leicester, Leicester LE1 7RH, UK*

Correspondence should be addressed to Teng Huang; huangteng1220@buaa.edu.cn and Jinping Sun; sunjinping@buaa.edu.cn

Received 5 October 2018; Revised 7 December 2018; Accepted 23 December 2018; Published 3 February 2019

Guest Editor: David Tomás

Radar image recognition is a hotspot in the field of remote sensing. Under the condition of sufficiently labeled samples, recognition algorithms can achieve good classification results. However, labeled samples are scarce and costly to obtain. Our major interest in this paper is how to use these unlabeled samples to improve the performance of a recognition algorithm in the case of limited labeled samples. This is a semi-supervised learning problem. However, unlike the existing semi-supervised learning methods, we do not use unlabeled samples directly and, instead, look for safe and reliable unlabeled samples before using them. In this paper, two new semi-supervised learning methods are proposed: a semi-supervised learning method based on fast search and density peaks (S^2DP) and an iterative S^2DP method (IS^2DP). When the labeled samples satisfy a certain requirement, S^2DP uses fast search and a density peak clustering method to detect reliable unlabeled samples based on the weighted kernel Fisher discriminant analysis (WKFDA). Then, a labeling method based on clustering information (LCI) is designed to label the unlabeled samples. When the labeled samples are insufficient, IS^2DP is used to iteratively search for reliable unlabeled samples for semi-supervision. Then, these samples are added to the labeled samples to improve the recognition performance of S^2DP. In the experiments, real radar images are used to verify the performance of our proposed algorithm in dealing with the scarcity of the labeled samples. In addition, our algorithm is compared against several semi-supervised deep learning methods with similar structures. Experimental results demonstrate that the proposed algorithm has better stability than these methods.

1. Introduction

Radar image recognition is a popular research area in the field of remote sensing [1–3]. With the development of imaging technologies and the expansion of radar image data, the requirement of real-time and accuracy of data processing becomes higher and higher. Under the condition where the number of the labeled samples is sufficient, a recognition algorithm can generally achieve satisfactory classification results with a strong sample representation ability [2, 4]. However, the labeled radar images are scarce compared to the case of optical images, and the cost of labeling is also very expensive. They can usually be interpreted by an experienced expert [5, 6]. Therefore, it is unrealistic to obtain a large number of labeled samples by manual annotation.

This paper focuses on how to use these unlabeled samples to improve the performance of a recognition algorithm in the case of limited labeled samples. This is a semi-supervised learning problem. Currently, semi-supervised deep learning achieves promising recognition performance, such as Ladder Network [7] and Temporal Ensembling [8]. However, unlike those existing semi-supervised learning methods, we do not use unlabeled samples directly and, instead, look for safe and reliable unlabeled samples and then use these unlabeled samples to enhance the performance of the recognition algorithm. This is because the unlabeled radar

images need to go through the detection stage in the process of acquisition [9, 10]. These samples may deteriorate the semi-supervised algorithms' learning, especially when the number of the labeled samples and that of the unlabeled samples are somehow unbalanced. This will influence the performance of the semi-supervised algorithm. The negative effects of these unreliable and unlabeled samples on semi-supervised algorithms are analyzed comprehensively in [11, 12]. Therefore, it is very important for a semi-supervised algorithm to identify reliable unlabeled samples before we learn unlabeled samples' features.

Effective use of unlabeled samples is a new and interesting topic for semi-supervised methods. These emerging semi-supervised methods are mainly divided into two categories: semi-supervision based on integrated resources and safe semi-supervision based on weights. Semi-supervised methods, based on integration resources, usually combine multiple semi-supervised models, comprehensively analyse the predictions of unlabeled samples, and choose reliable unlabeled samples to improve the recognition performance of the system. For example, Li et al. [13] proposed the S^3VM-us method, which consists of a semi-supervised support vector machine (S^3VM) [14] and a standard support vector machine (SVM) [15]. The confidence of unlabeled samples is determined by both classifiers. If the evaluation results are consistent, the unlabeled samples are identified. Li et al. [16] also proposed a safe S^3VM method (S^4VM). We understand that the S^3VM is based on the low-density hypothesis in order to detect a significant interval along the low-density boundary from the feature space to identify unlabeled samples. Unlike S^3VM, S^4VM was based on the fact that there may be more than one low-density boundary in the feature space. This approach considers all the possible situations, equivalently, integrating multiple S^3VMs to pinpoint reliable unlabeled samples. Wang et al. [17] proposed a safety-aware semi-supervised method. It consists of a semi-supervised model and a supervised model, which minimized the square loss between the two models in order to detect reliable unlabeled samples. Similar to [17], Gan et al. [18] proposed a safe semi-supervised method which added a Laplace regularization term to the square loss function to enhance the reliability of unlabeled sample selection. Persello et al. [19] proposed a progressive S^3VM with diversity (PS^3VM-D) method. On the basis of multiple confidence measurements, reliable unlabeled samples were obtained by querying the samples nearby the margin band.

Weight-based semi-supervisory is based on the fact that the more unlabeled samples with similar weights to the labeled samples, the more reliable the system becomes. Therefore, the influence of unreliable unlabeled samples on the algorithmic performance is suppressed by reducing their weights. For example, [20] considered the unlabeled samples nearby the classification plane and suppressed their influence on the system performance by reducing their weights. In addition, [21–23] controlled the weights by density estimation, weighted likelihood maximization, and graph modelling.

The above semi-supervised methods use unlabeled samples to some extent, however, they also ignore the number of the labeled samples. If the labeled samples are too few, the performance of these algorithms is difficult to be guaranteed, which will inevitably affect the evaluation of the reliability of unlabeled samples. In addition, they lack investigating variability and similarity between unlabeled and labeled samples, which makes it difficult to understand the dynamics and interaction of unlabeled samples. Therefore, in this paper, two new semi-supervised learning methods are proposed: a semi-supervised learning method based on fast search and density peaks (S^2DP) and an iterative S^2DP method (IS^2DP).

When the labeled samples satisfy a certain number, S^2DP is used directly to identify reliable unlabeled samples. For one thing, it works with a new sample weighted kernel Fisher discriminant analysis (WKFDA) supervision method. Using the difference between the samples, the WKFDA method extracts the features of the labeled samples to help formulate the distribution of the unlabeled samples' features, solving the problem of mismatch between them. And for another, it is combined with a clustering method: fast search and determination of density peaks (DP) proposed by Rodriguez and Laio in 2014 [24]. Then, unlabeled sample features are further investigated so that the reliable unlabeled sample features are identified. Finally, an unlabeled sample labeling method based on clustering information (LCI) is designed to retrieve the labels of the unlabeled sample features.

When the labeled samples are insufficient, IS^2DP is used to iteratively render reliable unlabeled samples. Since the labeled and the unlabeled samples may be uneven in numbers, the unreliable unlabeled samples tend to deteriorate the semi-supervised algorithm. The IS^2DP first divides the unlabeled learning set into different subsets according to the size of the labeled sample set. This not only prevents the deterioration of the semi-supervised algorithm by a large number of unreliable samples but also speeds up the processing of the semi-supervised algorithm. Then, the S^3VM is exploited to go through the semi-supervised samples which may be away from the hyperplane of the S^3VM as the reliable semi-supervised samples are added to the labeled samples to improve the performance of the semi-supervised algorithm.

The rest of this paper is organized as follows. Section 2 gives a brief review of the approaches involved. Section 3 describes the proposed method in detail. Section 4 presents the experiments for the SAR images targets recognition. The conclusion is drawn in Section 5.

2. Preliminary

2.1. DP Algorithm. Clustering by fast search and detection of density peaks (DP)[24] can quickly realize accurate detection and clustering of various shapes. Moreover, it is used to evaluate each cluster membership so as to determine reliable cluster members. The DP algorithm is mainly divided into the following three steps.

(1) Determination of Cluster Centers. In the DP, it is assumed that the cluster centers are surrounded by the neighbors with the lower local density and they are at a relatively large distance from any points with a higher local density. Based on the above cluster center assumption, for each sample i, two

quantities are calculated: the local density ρ_i of the sample and the distance σ_i from a sample to the other with a high local density. In the decision map with ρ_i and σ_i as the horizontal and vertical coordinates, respectively, their product is

$$\gamma_i = \rho_i \sigma_i \tag{1}$$

where the sample point with the larger γ_i is more likely to be the cluster center. Therefore, only γ_i is sorted in a descending order, and several corresponding samples are selected as the clustering center from the largest value.

(2) Clustering of Samples. After the clustering center has been determined, all the samples are assigned to be the nearest cluster centers. Compared with the other clustering algorithms, DP clustering process is simple and does not require iterative optimization of the loss function.

(3) Automated Evaluation of Cluster Members. In the clustering results, it is important to quantitatively evaluate the credibility of each sample cluster. The DP algorithm has this capability, compared to other clustering algorithms. It firstly defines a neighbourhood for each cluster. Then, the maximum value ρ_b of the local density of the samples is found in each neighbourhood. Finally, in each cluster, all the samples with local density greater than ρ_b are considered as the cluster core candidates, otherwise, they are considered as the cluster halo of the cluster. The samples in the cluster core are very similar to the central samples and belong to reliable samples. The samples in the cluster halo have a certain distance from the central sample, which is very likely to be noise and belongs to unreliable samples. In addition, there are some cross-clustering and isolated samples that are also unreliable.

In summary, after having clustered by the DP, the samples located at the cluster core are considered to be reliable cluster samples, whilst the others are unreliable samples. Compared to the conventional clustering algorithms, such as Clara [25] or Fanny [26], the DP has lower computational complexity and less computational time. It also well characterizes the distribution of the samples and achieves more accurate clustering results. Besides, the reliability of the clustering results can be provided, which makes the DP easy to be interactive with other algorithms. However, only considering the distance between the sample points can insufficiently characterize the data because it cannot accurately describe the samples with small difference between two categories. When the sample dimension is high, the distance matrix is large, which can reduce the efficiency of the algorithm. Therefore, choosing the appropriate feature extraction method is a key in the DP.

2.2. S^3VM Method. The S^3VM is the extension of the support vector machine (SVM). A standard SVM is based on the structural risk minimization to classify the learning set by extracting the support vectors from the training set to find the optimal hyperplane. In case of the binary SVM, given the training set **L** and the testing set **U**, we have the following constriction optimization problem:

$$\begin{aligned} \min \quad & \Phi(\boldsymbol{w}) = \frac{1}{2}\left(\boldsymbol{w}^T \boldsymbol{w}\right) + \sum_{i=1}^{n} c_i \xi_i \\ s.t. \quad & \mathrm{y}_i \left[\boldsymbol{w}^T \Phi(x_i) + b\right] \geq 1 - \xi_i, \\ & \xi_i \geq 0, \ \mathrm{i} = 1, 2, \cdots, \mathrm{n} \end{aligned} \tag{2}$$

where x_i is the training sample and y_i is the corresponding label, $(x_i, y_i) \in$ **L**; $\Phi(\cdot)$ maps the data into the feature space; $\boldsymbol{w}$ is the orthogonal vector between x_i and the hyperplane; b is the bias to measure the distance between **L** and the hyperplane; ξ_i is the slack variable to represent the offset of x_i; c_i is the cost factor to measure the weight between the optimal hyperplane and the minimum offset; n is the number of the training samples.

For the S^3VM, the iterative process is operated and the semi-labeled samples (selected from **U** in the previous step) are added to **L**. Their confidence is diverse in different iterative steps and they are given different cost factors, leading to the following function:

$$\begin{aligned} \min \quad & \Phi(\boldsymbol{w}) = \frac{1}{2}\left(\boldsymbol{w}^T \boldsymbol{w}\right) + \sum_{i=1}^{n} c_i \xi_i + \sum_{i=1}^{m} c_j \varepsilon_j \\ s.t. \quad & \mathrm{y}_i \left[\boldsymbol{w}^T \Phi(x_i) + b\right] \geq 1 - \xi_i, \\ & \xi_i \geq 0, \ \mathrm{i} = 1, 2, \cdots, \mathrm{n} \\ & \hat{y}_j \left[\boldsymbol{w}^T \Phi\left(\hat{x}_j\right) + b\right] \geq 1 - \varepsilon_j, \\ & \varepsilon_j \geq 0, \ \mathrm{j} = 1, 2, \cdots, \mathrm{m} \end{aligned} \tag{3}$$

where $\hat{x}_j$ is the semi-labeled sample selected from **U**, with the slack variable (ε_j), cost factor (c_j) and semi-label ($\hat{y}_j$) and m is the number of the semi-labeled samples.

The S^3VM can deal with the nonlinear problem using the kernel methods and its semi-supervised samples with the bigger c_j. But when the sample dimension is high, the computation speed would decrease. Therefore, the dimension reduction and effective semi-supervised samples are the critical aspects to the S^3VM.

3. Proposed Methods

This paper presents two methods: S^2DP and IS^2DP. When the labeled samples exceeds a certain number, S^2DP directly performs screening and classification of the reliable unlabeled samples. When the labeled sample is insufficient, IS^2DP is used to continuously query reliable unlabeled samples and generate necessary samples to be added to the labeled samples in order to improve the recognition performance of S^2DP. The S^2DP and IS^2DP are described below, respectively.

3.1. S^2DP. Figure 1 shows the flowchart of the proposed S^2DP. First, we use WKFDA to extract the labeled sample

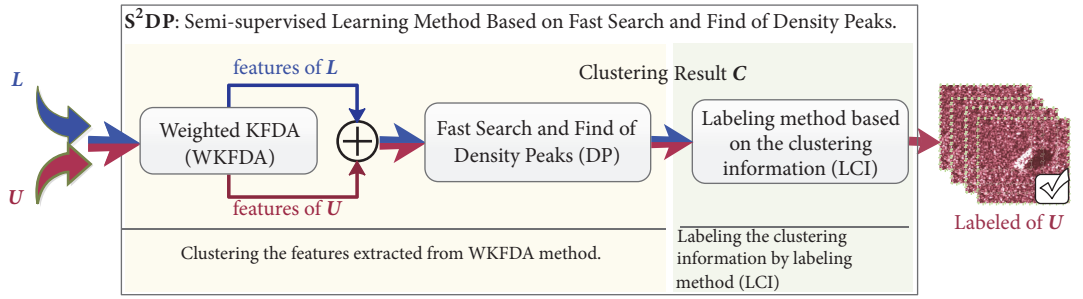


FIGURE 1: Basic flowchart of the S²DP.

features to build a new space. New features are obtained by projecting unlabeled samples into this new space. In this space, the new feature distributions are as close as possible between the intraclass features with a certain weight, and the interclass features are as far apart as possible to enhance the separability between the features. Secondly, the DP is used to cluster the generated features. Finally, the unlabeled samples are identified by the labeling method based on the DP clustering information (LCI). In Figure 1, $\boldsymbol{L}$ represents a set of the labeled samples, and $\boldsymbol{U}$ represents a set of the unlabeled samples, which respectively generate features with the labeled information (i.e., labeled features) and features without labeled information (i.e., unlabeled features) after going through WKFDA; $\boldsymbol{C}$ represents the clustering results of the DP. The WKFDA and LCI methods are described in the following section.

(1) WKFDA. Assume that $\boldsymbol{X}$ represents all the samples of $\mathbf{L}$ and the ith category $\boldsymbol{X_i} = [x_1^i, x_2^i, \ldots, x_{N_i}^i]$ is the subset of $\boldsymbol{X}$, where N_i is the samples' number of $\boldsymbol{X_i}$. $\boldsymbol{v_i} = [v_1^i, v_2^i, \ldots, v_{N_i}^i]$ is the sample weight vector of $\boldsymbol{X_i}$. It is used to control intraclass samples as close as possible with certain weights.

In case of the binary classification, it cannot simply multiply the weight by the corresponding sample. Firstly, the weight matrices $\boldsymbol{V_i}$ and $\boldsymbol{H_i}$ are generated:

$$\begin{aligned} \boldsymbol{V_i} &= \operatorname{diag}\left(v_1^i, v_2^i, \ldots, v_{N_i}^i\right)_{N_i * N_i} \\ \boldsymbol{H_i} &= \left[\boldsymbol{v_i}, \boldsymbol{v_i}, \ldots, \boldsymbol{v_i}\right]_{N_i * N_i} \end{aligned} \tag{4}$$

Secondly, the weight vector and the weight matrix are normalized using

$$\begin{aligned} \boldsymbol{v_i} &= \frac{\boldsymbol{v_i}}{sum\left(\boldsymbol{v_i}\right)} \\ \boldsymbol{V_i} &= \frac{N_i \boldsymbol{V_i}}{sum\left(\boldsymbol{V_i}\right)} \\ \boldsymbol{H_i} &= \frac{\boldsymbol{H_i}}{sum\left(\boldsymbol{H_i}\right)} \end{aligned} \tag{5}$$

where $sum(\cdot)$ represents the summation. The above weight matrix can be used to measure the information of the sample itself. Although $\boldsymbol{V_i}$ and $\boldsymbol{H_i}$ are made up of $\boldsymbol{v_i}$, their elements are different. The sum of each column's elements of $\boldsymbol{H_i}$ is equal to 1, and the trace of $\boldsymbol{V_i}$ is equal to N_i. Thirdly, the projection direction $\boldsymbol{w}$ is calculated by Equation (6):

$$\begin{aligned} \boldsymbol{w} &= \sum_{j=1}^{N} a_j \Phi\left(x_j\right) v_j = \sum_{j=1}^{N} a_j v_j \Phi\left(x_j\right) = \sum_{j=1}^{N} \beta_j \Phi\left(x_j\right) \\ &= \Phi\left(\boldsymbol{X}\right) \boldsymbol{\beta} \end{aligned} \tag{6}$$

where $\beta_j = a_j v_j$. $\Phi(\cdot)$ is nonlinear mapping that maps the samples to a new feature space. In this new space, the sample's mean, before and after the projection has been made, can be calculated by

$$\begin{aligned} \boldsymbol{m_i}^{\phi} &= \frac{1}{N_i} \sum_{j=1}^{N_i} \Phi\left(x^i{}_j\right) v^i{}_j = \frac{1}{N_i} \Phi\left(\boldsymbol{X_i}\right) \boldsymbol{v_i} \\ \boldsymbol{w}^T \boldsymbol{m_i}^{\phi} &= \frac{1}{N_i} \sum_{j=1}^{N} \sum_{k=1}^{N_i} \beta_j \Phi\left(x_j\right) \Phi\left(x^i{}_k\right) v^i{}_k = \frac{1}{N_i} \boldsymbol{\beta}^T \boldsymbol{K_i} \boldsymbol{v_i} \end{aligned} \tag{7}$$

The interclass scatter matrix $\boldsymbol{w}^T \boldsymbol{S}^{\phi}{}_b \boldsymbol{w}$ and intraclass scatter matrix $\boldsymbol{w}^T \boldsymbol{S}^{\phi}{}_w \boldsymbol{w}$, after the projection has been achieved, are calculated by

$$\begin{aligned} &\boldsymbol{w}^T \boldsymbol{S}^{\phi}{}_b \boldsymbol{w} = \boldsymbol{w}^T \left(\boldsymbol{m}^{\phi}{}_1 - \boldsymbol{m}_2{}^{\phi}\right)\left(\boldsymbol{m}^{\phi}{}_1 - \boldsymbol{m}_2{}^{\phi}\right)^T \boldsymbol{w} \\ &= \boldsymbol{\beta}^T \boldsymbol{M} \boldsymbol{\beta} \\ &\boldsymbol{w}^T \boldsymbol{S}^{\phi}{}_w \boldsymbol{w} \\ &= \sum_{i=1}^{2} \sum_{j=1}^{N_i} \boldsymbol{w}^T \left[\Phi\left(\boldsymbol{x}^i{}_j\right) v^i{}_j - \boldsymbol{m}_i{}^{\phi}\right]\left[\Phi\left(\boldsymbol{x}^i{}_j\right) v^i{}_j - \boldsymbol{m}_i{}^{\phi}\right]^T \boldsymbol{w} \\ &= \boldsymbol{\beta}^T \boldsymbol{G} \boldsymbol{\beta} \end{aligned} \tag{8}$$

where $\boldsymbol{M} = (\boldsymbol{K}_1 \boldsymbol{v}_1 / \boldsymbol{N}_1 - \boldsymbol{K}_2 \boldsymbol{v}_2 / \boldsymbol{N}_2)(\boldsymbol{K}_1 \boldsymbol{v}_1 / \boldsymbol{N}_1 - \boldsymbol{K}_2 \boldsymbol{v}_2 / \boldsymbol{N}_2)^{\mathrm{T}}$ and $\boldsymbol{G} = \sum_{i=1}^{2} \boldsymbol{K}_i (\boldsymbol{V}_i - \boldsymbol{H}_i)(\boldsymbol{V}_i - \boldsymbol{H}_i)^{\mathrm{T}} \boldsymbol{K}_i{}^T$. In order to satisfy the requirements of the maximum interclass interval and the minimum intraclass interval, this goal can be expressed as follows:

$$\max J\left(\boldsymbol{w}\right) = \frac{\boldsymbol{\beta}^T \boldsymbol{M} \boldsymbol{\beta}}{\boldsymbol{\beta}^T \boldsymbol{G} \boldsymbol{\beta}} \tag{9}$$

which is called the generalized Rayleigh quotient. Then $\boldsymbol{\beta}$ can be calculated according to the flowchart of the KFDA by solving the following optimization problem:

$$\begin{aligned} \max \quad & \boldsymbol{\beta}^T \boldsymbol{M} \boldsymbol{\beta} \\ s.t. \quad & \boldsymbol{\beta}^T \boldsymbol{G} \boldsymbol{\beta} = c \neq 0 \end{aligned} \tag{10}$$

where c is the constant. By introducing the Lagrange multiplier, the function can be transformed to a Lagrange unconstrained extremum problem:

$$L(\boldsymbol{w}, \lambda) = \boldsymbol{\beta}^T \boldsymbol{M} \boldsymbol{\beta} - \lambda \left(\boldsymbol{\beta}^T \boldsymbol{G} \boldsymbol{\beta} - c \right) \tag{11}$$

Let $\partial L(\boldsymbol{w}, \lambda)/\partial \boldsymbol{\beta} = 0, \partial(\cdot)$ represent the partial derivative. This function solution $\boldsymbol{\beta}$ is the eigenvector of $\boldsymbol{G}^{-1}\boldsymbol{M}$. Once solving $\boldsymbol{\beta}$, for any sample $\boldsymbol{x}$, its projection is

$$y = \langle \boldsymbol{w}, \Phi(\boldsymbol{x}) \rangle = \sum_{i=1}^{N} a_j v_j k(\boldsymbol{x}_i, \boldsymbol{x}) = \boldsymbol{\beta}^T \boldsymbol{K}_x \tag{12}$$

where $\boldsymbol{K}_x$ is the kernel matrix of all the training samples and $\boldsymbol{x}$.

Adding weights to KFDA algorithm is a common way to improve the KFDA algorithm. The aim is to make the WKFDA algorithm better learn sample features. However, different ways of adding weights make the WKFDA algorithm focus on learning sample features differently. For example, [27] added weights to each kernel function. The purpose was to introduce the prior knowledge of samples to enhance the learning of sample features in the WKFDA algorithm. Reference [28] added weights to the within-class scatter matrix. The purpose was to make the WKFDA algorithm not only learn the features of different types of samples but also learn the features of same types of samples in the process of finding the best vector. Unlike these algorithms, the WKFDA algorithm in this paper adds weights to samples, and these weights can be calculated by using the similarity or iterative difference of the samples. The purpose is to make the intraclass samples close to a certain distance, so that the WKFDA algorithm can not only suppress overfitting due to the small number of labeled samples but also facilitate the absorption of spectral information of samples to improve the learning of sample features. Although the binary WKFDA is shown, the multi-WKFDA can be obtained in accordance with the promotion of the kernel Fisher discriminant analysis (KFDA) [29].

(2) LCI. After the labeled sample set$\boldsymbol{L}$ and the unlabeled sample set$\boldsymbol{U}$ have been extracted by the WKFDA method, the labeled and unlabeled features are obtained. Next, the labeled and unlabeled features go into the DP to produce a clustering result$\boldsymbol{C}$. The clustering result$\boldsymbol{C}$ includes features such as cluster center, clustering core, clustering halo, and cross-clustering, but is insufficient to determine the labels of the unlabeled features. To solve this problem, we develop the LCI by using the clustering results and labeling information of the labeled features. LCI is able to label the clustering results of the unlabeled features. Because the unlabeled features are generated from the unlabeled samples, the unlabeled features and the unlabeled samples share the same labels. The basic flowchart of the LCI is shown in Figure 2.

We know that the features of clustering halo and cross-clustering are unreliable. Therefore, in Figure 2, the interference features in$\boldsymbol{C}$ need to be cleared to ensure that the subsequent unlabeled features are reliable. The$\boldsymbol{C}$ clearing the interference is processed separately according to whether the labeled features are included in the cluster core. If there are labeled features in a certain cluster core, the unlabeled features of the cluster core are very similar to the labeled features. These unlabeled features are regarded as the best learning features, combined with the corresponding labeled features, for training the S^3VM. At this time, in each iteration of the S^3VM, the labeled features from the unlabeled features are added to the next iteration to improve the robustness of the S^3VM algorithm. For the clustering cores which do not contain any labeled feature, the cluster centers are extracted and sent to the trained S^3VM to obtain their labels. Once the unlabeled cluster centers are labeled, the unlabeled features of the corresponding clustering core will be assigned the label. In this way, all the clustering cores' features are labeled, and the features that are not labeled are removed as noise. Finally, the unlabeled samples corresponding to the unlabeled features also have corresponding labels.

3.2. IS^{2}DP. When the number of the labeled samples is small but reaches a certain amount, the S^2DP uses the labeled features to investigate the distribution of the unlabeled features and also use the labeled features and the clustering result of the DP to obtain reliable unlabeled samples. However, when the number of the labeled samples is small, after the DP clustering has been achieved, the labeled features are not necessary in the cluster core, resulting in a low correlation between the labeled and unlabeled features. At this time, the labelled samples are difficult to represent the unlabeled samples, and S^2DP is no longer applicable. In this case, the common solution is that the semi-labeled sample from the unlabeled set is queried in order to increase the number of the original labeled samples. In order to obtain the reliable semi-labeled samples, the S^2DP needs to be modified iteratively.

The iterative semi-supervised method of the S^2DP, namely, the IS^{2}DP, is shown in Figure 3, where$\boldsymbol{U}$ is the unlabeled learning set to query the semi-labeled samples,$\boldsymbol{L}$ is the labeled training set, $\boldsymbol{L}^*$ is the semi-labeled samples set in each iteration, $\boldsymbol{L}'$ represents the final labeled training set, and$\boldsymbol{T}$ is the testing set.

The IS^{2}DP specific process is described as follows. Firstly, in each iteration,$\boldsymbol{U}$ is randomly divided into several subsets $(\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots, \boldsymbol{U}_n)$, which are combined with$\boldsymbol{L}$ to obtain $(\boldsymbol{U}_1, \boldsymbol{L}), (\boldsymbol{U}_2, \boldsymbol{L}), \ldots, (\boldsymbol{U}_n, \boldsymbol{L})$ as the input of the S^2DP. Secondly, the cluster cores are selected from $(\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots, \boldsymbol{U}_n)$ after the S^2DP as the candidate semi-labeled samples, and their cluster centers are added to the training set as the labeled samples to train S^3VM. For one thing, the number of the labeled samples sets is increased. And for another, it ensures that the labeled samples match the unlabeled samples since the cluster center represents the features of all the other samples in the

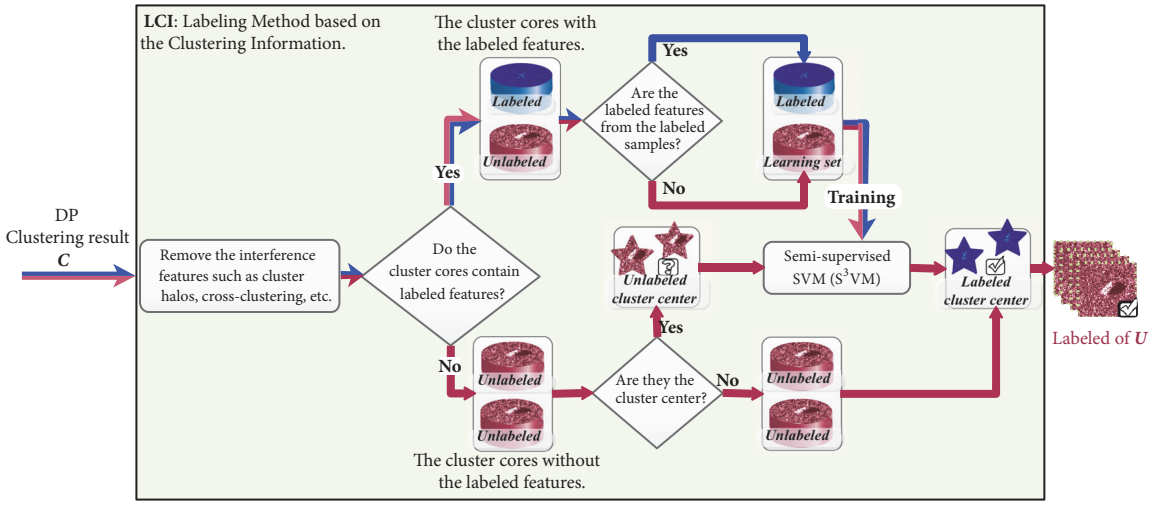


FIGURE 2: Basic flowchart of the LCI.

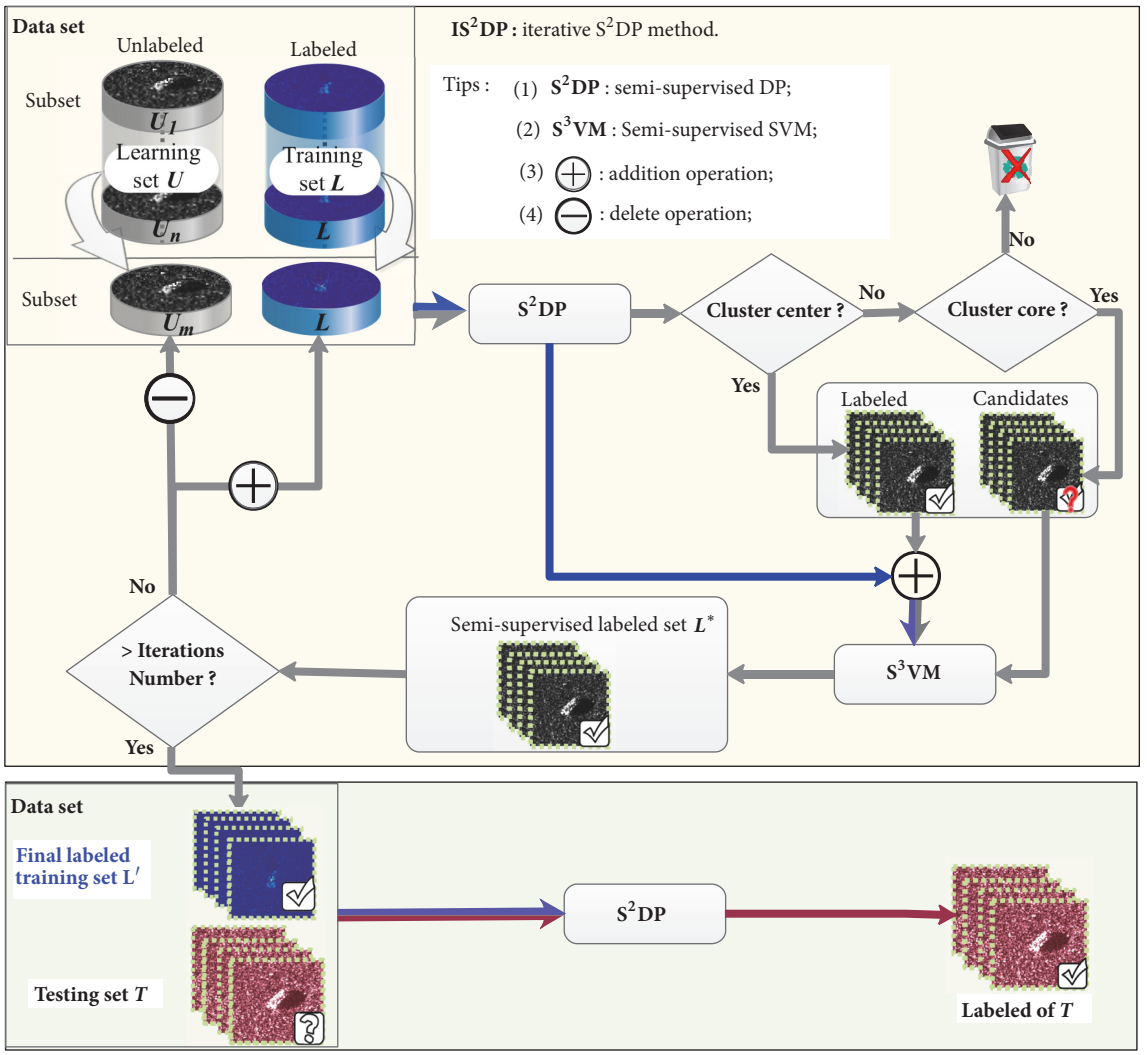


FIGURE 3: Flowchart of the IS^{2}DP.

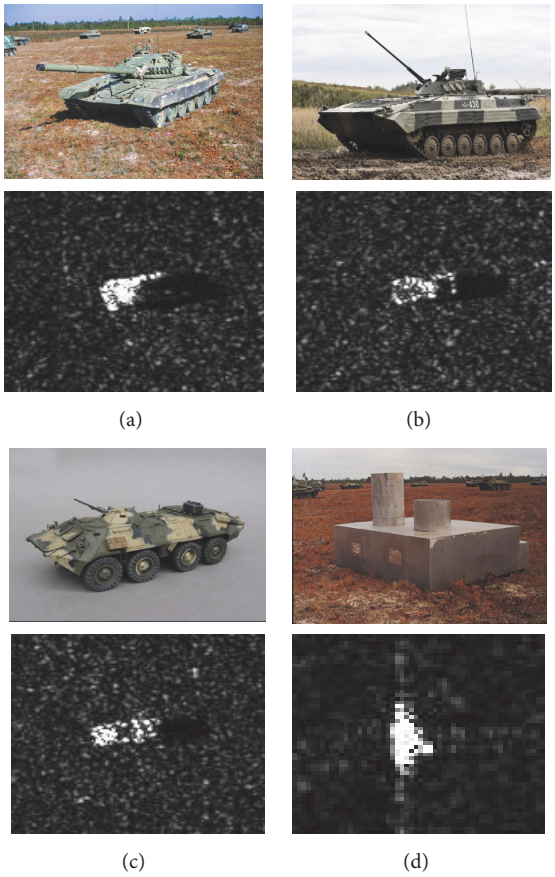


FIGURE 4: Optical images of four types of targets and corresponding SAR images: (a) T72, (b) BMP2, (c) BTR70, and (d) SLICY.

cluster core. Therefore, the robustness of the S^3VM is ensured. Thirdly, the semi-supervised sample $\boldsymbol{L}^*$ of each iteration is obtained by S^3VM. Finally, it needs to determine whether or not the iteration's termination condition is met so that the number of the iteration is greater than the threshold. If not, $\boldsymbol{L}$ is updated, $\boldsymbol{U}$ is reduced, and the iteration process continues. Otherwise, the final labeled training $\boldsymbol{L}'$ is undertaken to classify the testing set $\boldsymbol{T}$ by the S^2DP.

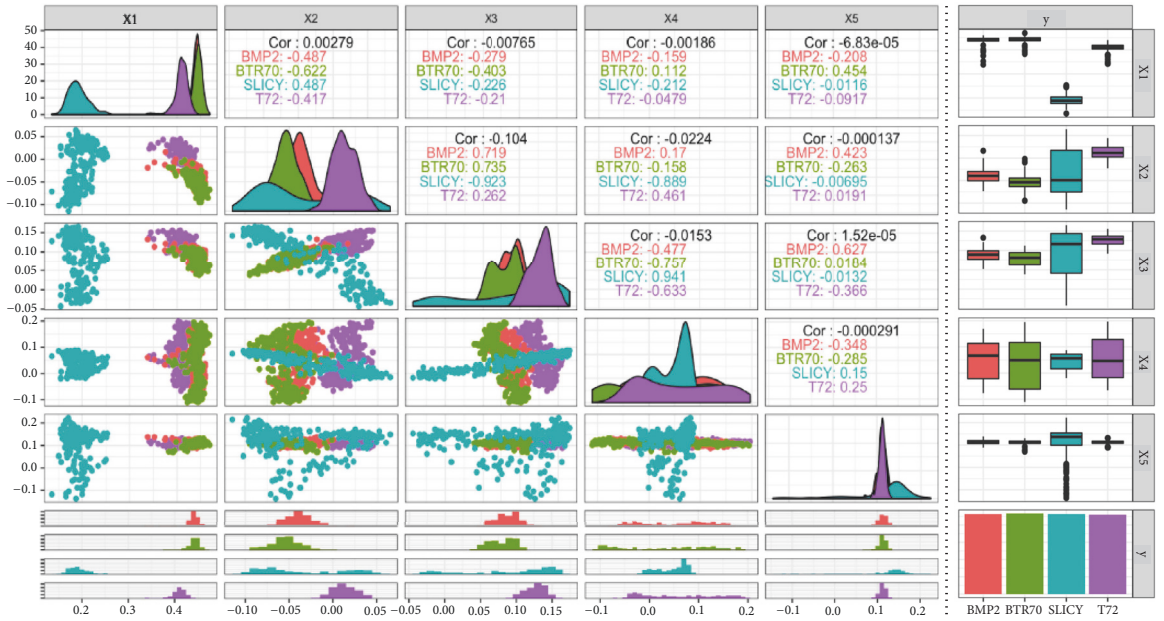
When the labeled sample is insufficient and necessary to query the semi-supervised samples, the IS^2DP can query the reliable semi-supervised samples and classify the unlabeled samples. In fact, the IS^2DP is equivalent to the S^2DP when the labeled samples reach a certain number.

4. Experiments

Our experiments use the SAR images from the Moving and Stationary Target Acquisition and Recognition (MSTAR) database, cofounded by National Defense Research Planning Bureau and the US Air Force Research Laboratory. The military targets contained in the database are collected at 15° and 17° depression angles, covering 360° azimuth angles. To display the intermediate experimental results in geometric space and highlight the significance and effectiveness of our method, the experiments in this paper use three types of military targets and one type of interference targets, which are T72, BMP2, BTR70, and SLICY. Of course, you can also choose other targets. Among these three types of military targets, BMP2 and T72 also contain different version variants. These variants have the same design blueprint, but from different manufacturers, they are slightly different in color and shape.

The optical images of the T72, BMP2, BTR70, and SLICY targets and the corresponding SAR images are shown in Figure 4. From optical images, the difference between these four types of targets is significant. However, the corresponding SAR images are difficult to distinguish by human vision due to speckle noise and similar spatial and spectral characteristics. The original resolution of these SAR image slices are 128∗128 and 45∗45. To facilitate the processing, we only take the 32∗32 resolution that contains the target and flatten these 2D images into one dimension. In order to show the separability of these data, we perform covariance operations on them in order to establish correlations between two-dimensional features. Figure 5 shows the correlation and box plot of the first 5-dimensional features. Figure 5(a) is the correlation of two dimension features, and Figure 5(b) is the corresponding box plot. In Figure 5(a), the lower left corner part is the scatter plot of two-dimensional features, and the upper right part is the correlation coefficient corresponding to the two-dimensional features. *Cor* represents the total correlation coefficient of the relevant two-dimensional features. Positive numbers indicate positive correlations and negative numbers indicate negative correlations. The greater the absolute value of these numbers, the more relevant the features of the corresponding two dimensions. From the correlation coefficient, *Cor*'s absolute value is small which shows that the correlation is low, indicating that they are independent of each other. From the scatter plot, we observe that they are very similar, which increases the difficulty of the recognition algorithm. In addition, from the box plot, there are abnormal points in the upper and lower bounds of the data. If these points are not removed in the learning set features, the performance of the algorithm will be affected.

In order to evaluate the performance of the proposed method, we design three sets of experiments: the evaluation experiment of effectiveness, the evaluation experiment of generalization ability, and the experiment compared with the semi-supervised deep learning method. Among them, the first set of the experiments will be carried out under standard operating conditions (SOC), the latter two sets under different extended operating conditions (EOC). The SOC mean that the testing and the training conditions are very similar. For example, the target types of the training, the learning, and the test sets are the same. On the basis of SOC, the gap between the training and testing conditions is gradually extended to form different EOC. For example, the target types of training set, learning set, and test set are different variants. Even the learning set contains other interfering targets. Compared with SOC, EOC significantly increases the recognition difficulty of the algorithm. We will set up one SOC and two EOCs (EOC_1 and EOC_2) to carry out the above three sets of experiments. The specific configuration of these conditions is as follows.



(a) Correlation between every two dimensions (b) Box plot

FIGURE 5: The correlation between the first 5-dimensional features after flattening SAR images of T72, BMP2, BTR70 and SLICY. (a)Correlation between every two dimensions and (b) box plot.

(1) Data Configuration of the SOC. Table 1 shows the data configuration of SOC. It contains two sets of data: data and test sets. The data set is used for the algorithm training. According to the label of the samples, the data set is divided into the labeled and unlabeled samples. The labeled samples, also known as the training set, have a number ranging from 3 to 40 per class. The unlabeled samples, also known as the learning set, have 190 samples per class. Regardless of the training or learning set, their target depression angle is 17°. The testing set is used for algorithm testing. Its target depression angle is 15°. Regardless of data or test set, we use the same variants of the targets, that is, T72 series sn_132 tanks, BMP2 series sn_c21 armored vehicles, and BTR70 series sn_c71 armored vehicles. We will verify the effectiveness of the S^2DP and IS^2DP under these conditions in Section 4.1, including their core components (LCI and WKFDA).

(2) Data Configuration of the EOC_1. Table 2 shows the data configuration of EOC_1. In Table 2, the training set is the same as that of Table 1. And the testing set is not the same version variants as the training set and the learning set. For example, the T72 is the sn_s7 version in the test, but it is the sn_132 and sn_812 versions in the training and the learning sets, respectively. These conditions will help increase the recognition difficulty of the algorithm. Other conditions shown in Table 2, such as the number of data sets, the depression angle of data sets, and the depression angle of the test set, are the same as those shown in Table 1 and are not described here. We will verify the generalization ability of the S^2DP and IS^2DP under these conditions presented in Section 4.2.

(3) Data Configuration of the EOC_2. Table 3 shows the data configuration of EOC_2. It is formed by adding the interference target SLICY to the learning set of Table 2, further increasing the recognition difficulty of the algorithm. To highlight the advantages of the proposed algorithm, we will compare the S^2DP based IS^2DP algorithm with the semi-supervised depth learning method under EOC_2 in Section 4.3.

4.1. Effectiveness Evaluation Experiment

4.1.1. The Effectiveness of the WKFDA Feature Extraction. To verify the effectiveness of the WKFDA feature extraction, it is compared with the KFDA, kernel local linear discriminant analysis (KLFDA) [30], semi-supervised KLFDA (Semi-KLFDA) [31] and kernel principal component analysis (KPCA) [32]. After these algorithms have extracted features, they all use the standard SVM as the final classifier. The experimental data configuration is shown in Table 1, and with the change of the number of the labeled samples, the overall accuracy rates (OA) of different methods are obtained, as shown in Figure 6. The horizontal axis represents the number of each type of target labeled samples corresponding to different experiments and the vertical axis represents the overall accuracy rate.

In Figure 6, the classification accuracy difference between the different algorithms is very clear. The WKFDA and KFDA both show higher accuracy, followed by the KLFDA and Semi-KLFDA, and finally KPCA. For the WKFDA and KFDA, when the number of the labeled samples is less than

Table 1: Data configuration of the SOC.

	Data set						Testing set		
	Training set (Labeled samples)			Learning set (Unlabeled samples)					
Target	T72	BMP2	BTR70	T72	BMP2	BTR70	T72	BMP2	BTR70
Type	sn_132	sn_c21	sn_c71	sn_132	sn_c21	sn_c71	sn_132	sn_c21	sn_c71
Quantity	3~40	3~40	3~40	190	190	190	196	195	196
Depression	17°			17°			15°		

Table 2: Data configuration of the EOC_1.

	Data set						Testing set		
	Training set (Labeled samples)			Learning set (Unlabeled samples)					
Target	T72	BMP2	BTR70	T72	BMP2	BTR70	T72	BMP2	BTR70
Type	sn_132	sn_c21	sn_c71	sn_812	sn_9566	sn_c71	sn_s7	sn_9563	sn_c71
Quantity	3~40	3~40	3~40	190	190	190	196	195	196
Depression	17°			17°			15°		

Table 3: Data configuration of the EOC_2.

	Data set							Testing set		
	Training set (Labeled samples)			Learning set (Unlabeled samples)						
Target	T72	BMP2	BTR70	T72	BMP2	BTR70	SLICY	T72	BMP2	BTR70
Type	sn_132	sn_c21	sn_c71	sn_812	sn_9566	sn_c71	—	sn_s7	sn_9563	sn_c71
Quantity	3~40	3~40	3~40	190	190	190	190	196	195	196
Depression	17°			17°				15°		

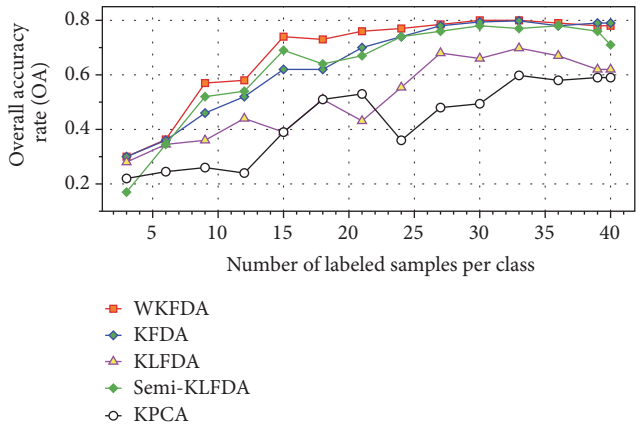


Figure 6: The OA trend chart of different feature extraction algorithms with the number of labeled samples changes in SOC experiment. Here, these feature extraction algorithms use SVM as classifier.

24, the WKFDA's classification results are better than the KFDA. When the number of the labeled samples is greater than 24, their classification results are almost the same. It shows that KFDA and WKFDA have good feature extraction capabilities, while the WKFDA is suitable for dealing with a small quantities of labeled samples. For the KFDA and Semi-KLFDA, when the number of the labeled samples is less than 20, the Semi-KLFDA's classification results are better than the KFDA. When the number of labeled samples is greater than 20, their classification results are almost the same. For the KPCA, as the number of the labeled samples increases, its classification results are always poor.

In order to understand the above experimental results, we take a close look at the projections of the learning samples under the condition that the same number of the labeled samples is taken. Figures 7(a), 7(b), 7(c), 7(d), and 7(e) shows the projection of the learning set for KPCA, KLFDA, Semi-KLFDA, KFDA, and WKFDA algorithms respectively when the number of the labeled samples is 20. As can be seen from Figure 7, the projection result shown in (e) is the best, where we can classify the three targets, second best is (d) and then (c), (b), and (a). The quality of the projection results mainly depends on whether or not the feature extraction algorithm can effectively extract features from the SAR images.

For the KPCA algorithm, it only reduces the original features of the SAR images. As the number of the labeled samples increases, the classification accuracy of the KPCA features continues to increase. The original features of the SAR images are difficult to identify. Therefore, the classification accuracy of SVM based on the KPCA features is poor, shown in Figure 7(a).

For the KLFDA and Semi-KLFDA algorithms, they take advantage of the difference between the sample classes and extract features that are easily identifiable from the SAR images to certain extent. Therefore, their projection looks better than the KPCA algorithm. However, in the case where the overall features are not separable, the KLFDA and Semi-KLFDA algorithms overemphasize the local features,

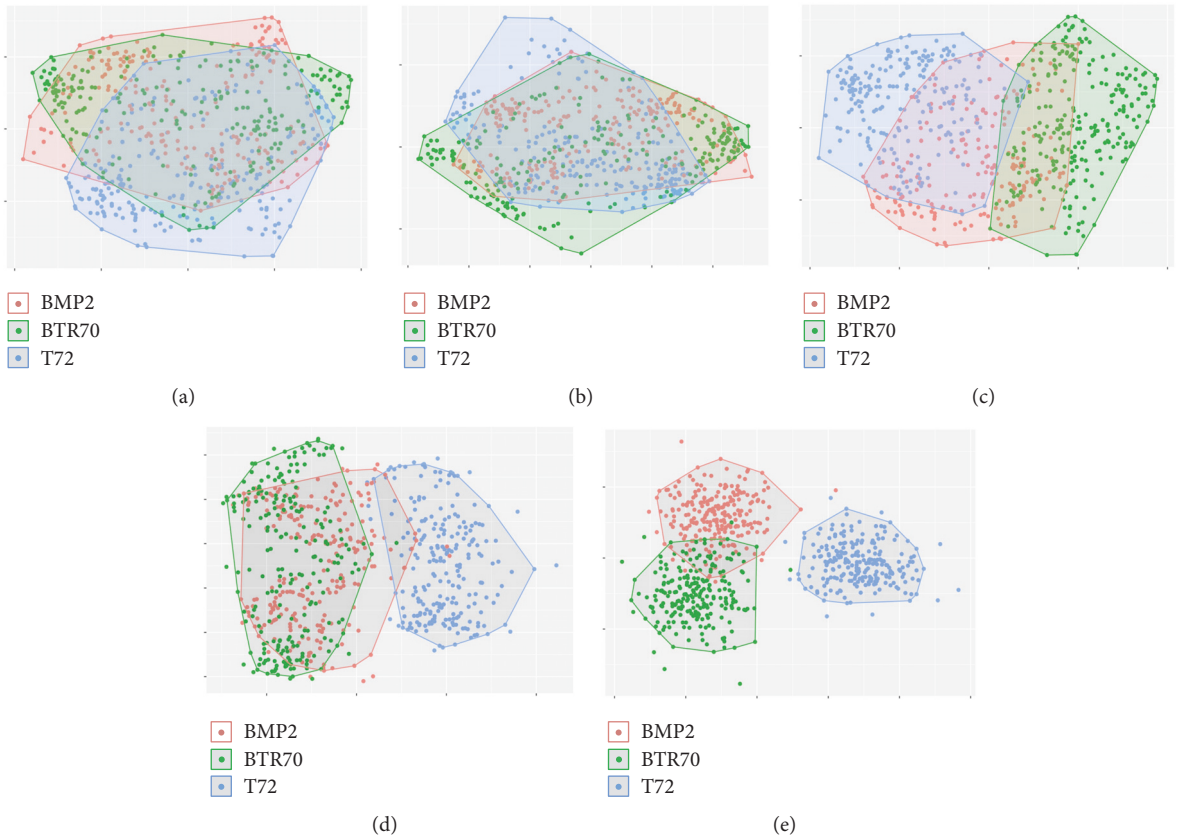


Figure 7: The projections of the learning set for KLFDA, Semi-KLFDA, and KPCA, respectively, when the number of the labeled samples is 20. (a)KPCA; (b)KLFDA; (c)Semi-KLFDA; (d)KFDA; (e)WKFDA.

resulting in more confusing clutters in the projection space. This is observed from Figures 7(b) and 7(c).

For the KFDA and WKFDA algorithms, both of them well use the difference between different classes and the similarities in the same classes. Therefore, the projections shown in Figures 7(d) and 7(e) are better than those of the other methods. We know that the KFDA and WKFDA algorithms are supervised algorithms which guide the projection of the unlabeled sample features based on the features of the labeled samples. Therefore, whether or not these algorithms are good at learning the labeled sample features will affect the quality of the projection of the unlabeled sample features. In the process of the labeled sample feature learning, the KFDA algorithm forces the interclass samples to be as far apart as possible in addition to forcing the samples intraclasses to be as close as possible. At the same time, it may cause the algorithm to overfit and is difficult to guide the unlabeled sample features to be projected onto the optimal direction. The WKFDA algorithm is able to give the samples different weights so that the intraclass samples are close to each other with a certain weight. This can balance the concentration characteristics of the samples (the intraclass samples aggregate with each other and have a certain spatial structure) and can fully utilise the spectral information of the samples and reduces the algorithm's overfitting. Therefore, the results shown in Figure 7(e) seem better than those of Figure 7(d).

To further explore the impact of weighting on the WKFDA algorithm, taking the same labeled samples, Figure 8 shows the results of the KFDA and WKFDA algorithms for learning the features of the labeled samples. Figure 8(a) shows the KFDA features. Although the interclass distance is significant, the intraclass samples are concentrated, almost grouping to a point, which is easy to cause overfitting of the algorithm. Figure 8(b) shows the WKFDA features. Under the condition that the interclasses is separable, the intraclass distance is relatively large, which is easy to learn the sample information and suppress the overfitting of the algorithm. Here, Figure 8(b) also shows that different weights can result in different intraclass distance. Compared with the weight of 100, the intraclass sample space is larger when the weight is 50, and the interclasses can be well spaced, which makes it easier to learn sample information.

4.1.2. The Effectiveness of the LCI for Labeling Unlabeled Samples. To verify the effectiveness of the LCI for labeling unlabeled samples, using the WKFDA features of the Section 4.1.1 experiments, LCI is compared with the SVM and S^3VM classifiers under the DP clustering conditions. Using Table 1 as the experimental data, the same number of the labeled samples is selected from each type of targets in the training set. With the change of the number of the labeled samples, the OA trend chart of three methods is obtained, as

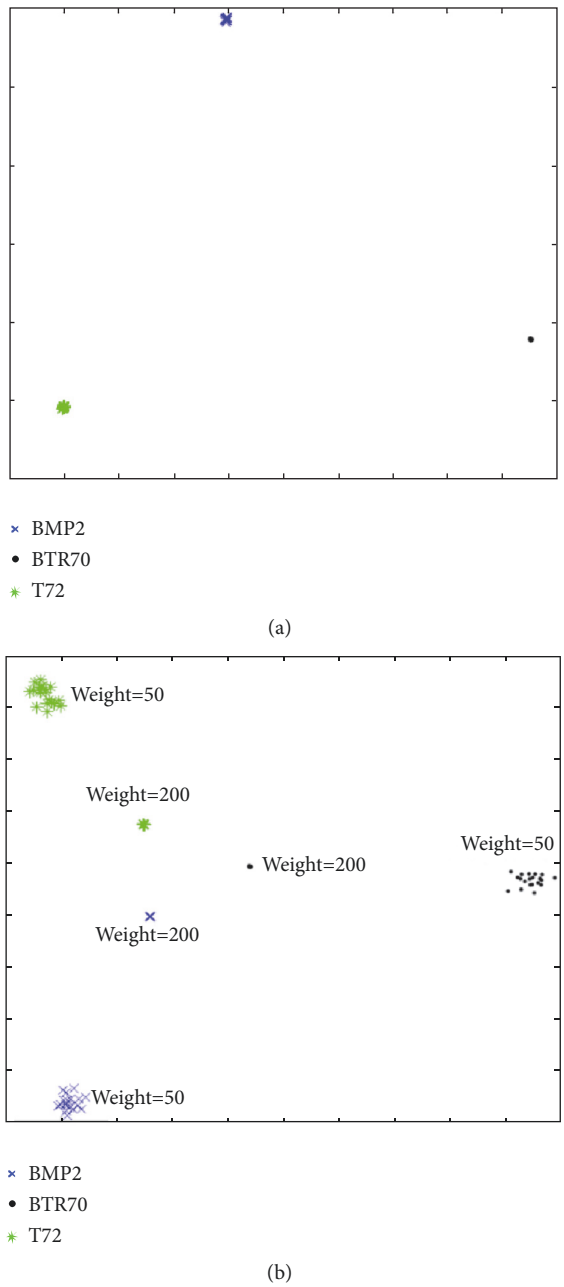


Figure 8: The projection of the KFDA and WKFDA with the same number of the labeled samples. (a)KFDA and (b)WKFDA.

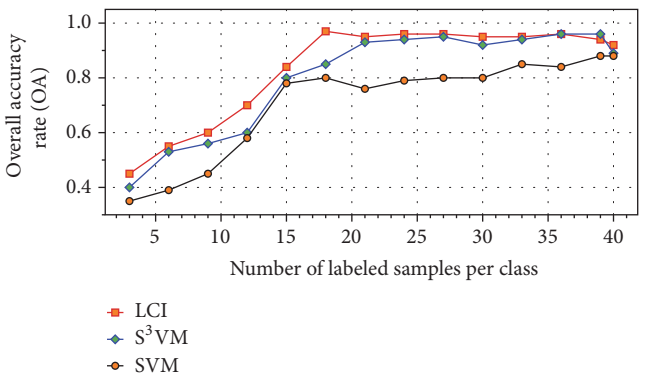


Figure 9: The OA trend chart of the LCI, S^3VM, and SVM algorithms as the number of labeled samples changes in the SOC experiment.

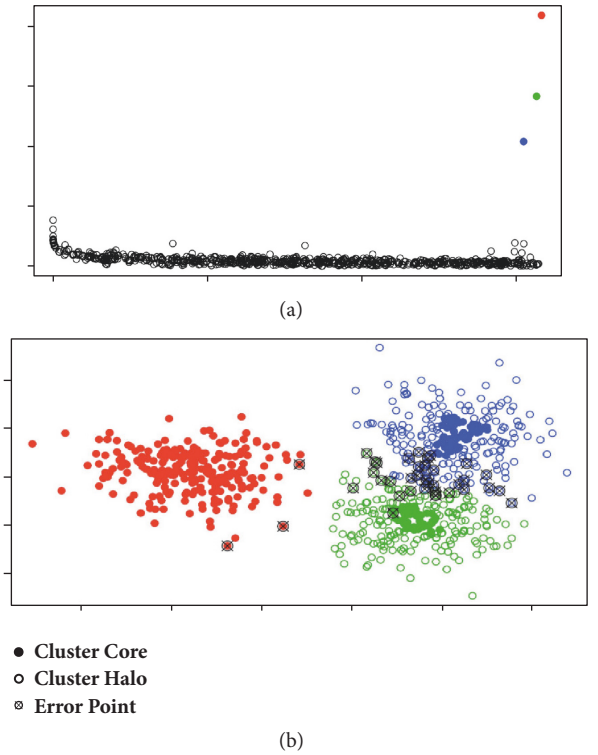


Figure 10: The clustering results of DP on the learning set when the number of the labeled samples is 21. (a) Red circle, green circle, and blue circle are the cluster centers selected by the DP; (b)the clustering results of the DP. • are the cluster cores, ◦ are clustered halos, and ⊗ are clustering error samples.

shown in Figure 9. The horizontal axis represents the number of each type of the labeled target samples corresponding to different experiments, and the vertical axis represents the overall accuracy rate. As can be seen from Figure 9, the accuracy of LCI and S^3VM is better than SVM. With more and more labeled samples, the accuracy of LCI and S^3VM is almost the same.

We know that the SVM, as a supervised learning method, requires a large number of labeled samples. Because the DP algorithm cannot provide enough labeled samples for the SVM, the SVM classification results are poor. For S^3VM and LCI, as a semi-supervised method, when the DP clustering outcome is reliable, they collect enough labeled samples to improve the recognition performance. Figure 10 shows the clustering results of the DP on the learning set when the number of the labeled samples is 21. In Figure 10(a), red circle, green circle, and blue circle are the cluster centers selected by the DP. The DP algorithm recommends that the learning set be divided into 3 categories, which is consistent with the actual situation. Figure 10(b) shows the clustering results of the DP. • are the cluster cores, ◦ are clustered halos, and ⊗

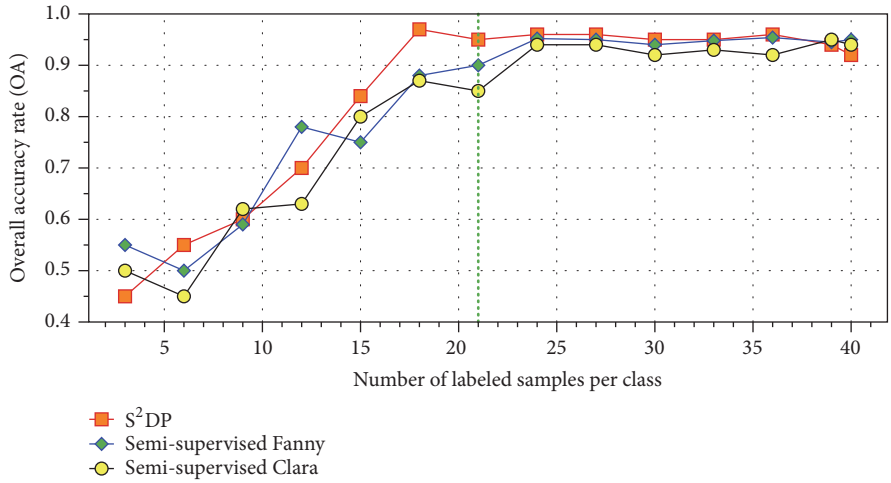


Figure 11: The OA trend chart of the S²DP, semi-supervised Clara, and semi-supervised Fanny with the number of labeled samples changes in SOC experiment.

are clustering errors. From Figure 10(b), the DP clustering has only minor errors and the result is quite accurate, further demonstrating that the recognition accuracy of the S³VM and LCI is equivalent. In addition, these errors are located in the cluster halos. In the LCI algorithm, the cluster halos and the cross-clustering samples will be deleted to ensure that the final labeled samples are reliable. Therefore, compared to the S³VM, the LCI recognition results are more consistent.

4.1.3. Verifying the Recognition Performance of S²DP. In order to verify the recognition performance of S²DP, S²DP is compared with its similar semi-supervised methods. These similar semi-supervised methods are the semi-supervised algorithms that replace DP in S²DP with other classical clustering algorithms: Clara [25] and Fanny [26], namely, semi-supervised Clara and semi-supervised Fanny. The experimental data configuration is shown in Table 1. With the changing numbers of the labeled samples, the OA trend chart of three methods is obtained, as shown in Figure 11. The horizontal axis represents the number of each type of labeled target samples corresponding to different experiments, and the vertical axis represents the overall accuracy rate.

By comparing the S²DP with the semi-supervised Clara and semi-supervised Fanny, the classification results of the different methods are greatly influenced by the number of the labeled samples. When the number of the labeled samples is less than 24, the overall accuracy of the three methods is continuously improved with the increase of the labeled samples. For the curve smoothness, the curve of the S²DP looks consistent over the curves of the other two methods. When the number of the labeled samples reaches 15, the recognition accuracy of S²DP is higher than that of the other two methods. When the number of the labeled samples reaches 24, the three methods have the same recognition accuracy and the curve trend is stable, but the S²DP is still better than the other two methods. Therefore, the S²DP is superior to the other two algorithms in terms of stability and classification accuracy.

When the labeled samples are very few, the DP clustering results of S²DP are too divergent to represent the unlabeled samples. Only few labeled samples are generated from the cluster core samples. In the end, the classification accuracy of the S²DP will not be high. As the number of the labeled samples increases, more and more labeled samples are generated by the cluster cores, which are also quite reliable. The S²DP classification accuracy is greatly improved. The other two methods are similar. However, as the number of the labeled samples increases, it is difficult for semi-supervised Clara and semi-supervised Fanny to guarantee the reliability of the labeled samples from the unlabeled samples during the clustering process. Therefore, their stability is not as good as that of S²DP. Figure 12 shows the three algorithms generate labeled samples from the learning set when the number of the labeled samples is 21. ⊗ are clustering errors. Obviously, the labeled samples generated by the S²DP algorithm are more reliable than the other two methods.

The Sections 4.1.1–4.1.3 experimental results show the relationship between the number of the labeled samples and the S²DP, verifying the validity of the WKFDA, LCI, and DP as the key step in the S²DP. It shows that the S²DP, compared with the other two methods, can achieve the best classification result when the initial labeled samples reach a certain number. But when the labeled samples are too few, its classification precision decreases. Therefore, the ability of the modified IS²DP to query semi-supervised samples needs to be verified.

4.1.4. Verifying the IS²DP in Ability to Query the Semi-Labeled Samples. When the labeled samples are few, the IS²DP can select the semi-labeled samples from the unlabeled samples as the labeled samples. In the Section 1, we know that PS³VM-D is also a semi-supervised method, which considers reliable incremental samples as semi-supervised samples by sample similarity. Therefore, PS³VM-D is selected as a comparative semi-supervised algorithm. They are all based on the extracted features by WKFDA. The experimental data

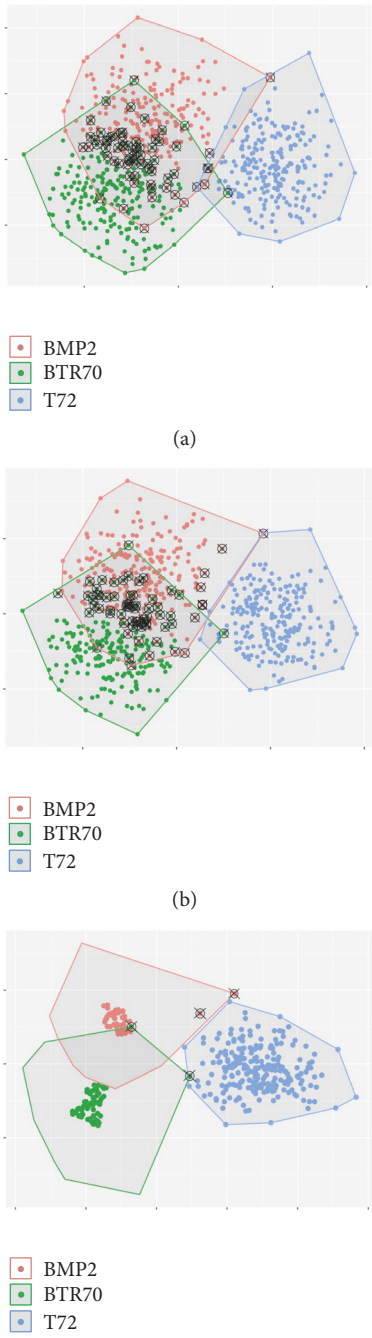


Figure 12: The semi-supervised Clara, semi-supervised Fanny, and S^2DP generate reliable labeled samples from the learning set when the number of the labeled samples is 21. ⊗ are clustering errors. (a) The Clara clustering results; (b) the Fanny clustering results; (c) the DP cluster cores.

configuration is shown in Table 1. With the change of the number of the labeled samples, the OA trend chart of two methods is obtained, as shown in Figure 13. The horizontal axis represents the number of each type of the labeled target samples corresponding to different experiments and the vertical axis represents the overall accuracy rate.

When the number of the labeled samples is less than 25, the classification accuracy of PS^{3}VM-D is obviously lower than that of IS^{2}DP, indicating that IS^{2}DP is more suitable for the case of too few labeled samples. When the number of the labeled samples is more than 25, the IS^{2}DP and PS^{3}VM-D have the same accuracy. It shows that the PS^{3}VM-D also gets enough labeled sample information, and the classification accuracy is improved.

We know that the core of PS^{3}VM-D is SVM. The optimal classification surface of PS^{3}VM-D is mainly influenced by SVM. The PS^{3}VM-D relies heavily on the labeled samples. It needs enough quantity to obtain a universal classification surface. Therefore, its classification performance varies significantly with the number of labeled samples and cannot remain stable until the labeled samples are sufficient. The classification performance of the IS^{2}DP is largely determined by the DP and WKFDA, which makes the IS^{2}DP more stable and accurate when the labeled samples are very few due to the sample description ability of the DP and the effective use of the labeled samples by the WKFDA.

4.2. Evaluation of Generalization Ability. The following will verify the generalization capabilities of S^{2}DP and IS^{2}DP under the EOC_1.

4.2.1. Verifying the S^{2}DP Generalization Capabilities. In Section 4.1.2, the comparison between LCI and S^{3}VM algorithm is actually the comparison of S^{2}DP with S^{3}VM based on WKFDA and DP (WKFDA+DP+S^{3}VM). The recognition accuracy of S^{2}DP and WKFDA+DP+S^{3}VM is equivalent in the SOC experiments. Here, we continue to compare the S^{2}DP and WKFDA+DP+S^{3}VM. The experimental data configuration is shown in Table 2. With the change of the number of labeled samples, the OA trend chart of two methods is obtained, as shown in Figure 14. The horizontal axis represents the number of each type of the labeled target samples corresponding to different experiments, and the vertical axis represents the overall accuracy rate.

In Figure 14, the recognition accuracy of S^{2}DP and WKFDA+DP+S^{3}VM algorithms increases with the increasing number of the labeled samples, and their final accuracy is equivalent. However, the curve of the S^{2}DP is relatively smooth. This shows that our method is stable and robust.

To verify this conclusion, we perform visual analysis of the key steps of the two methods, when the number of samples is 21. Figure 15(a) shows the features of the training and learning sets after the WKFDA processing. • represents the learning samples and ∗ represents the initial labeled sample. Figure 15(b) is the actual classification map of the WKFDA features after the DP clustering has been achieved. • represents the clustering core and ◦ represents the clustering halo. Figure 15(c) is the true classification map of Figure 15(b). • represents the clustering core, ⊗ represents the clustering error sample, and **?** represents the sample of the next step of the algorithm to be identified. As can be seen from Figure 15(a), the three types of targets are more confused at the boundary, which means that, in the future, they will affect the performance of the recognition algorithm if these

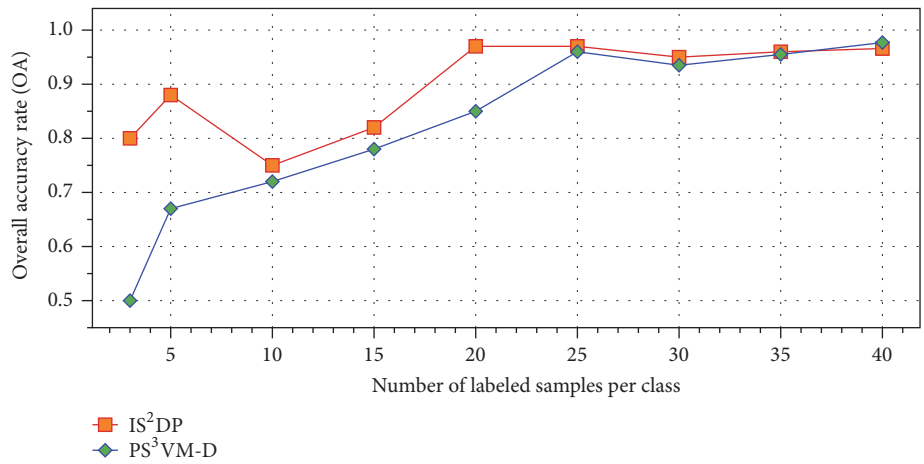


Figure 13: The OA trend chart of the IS^{2}DP and PS^{3}VM-D as the number of the labeled samples changes in the SOC experiment.

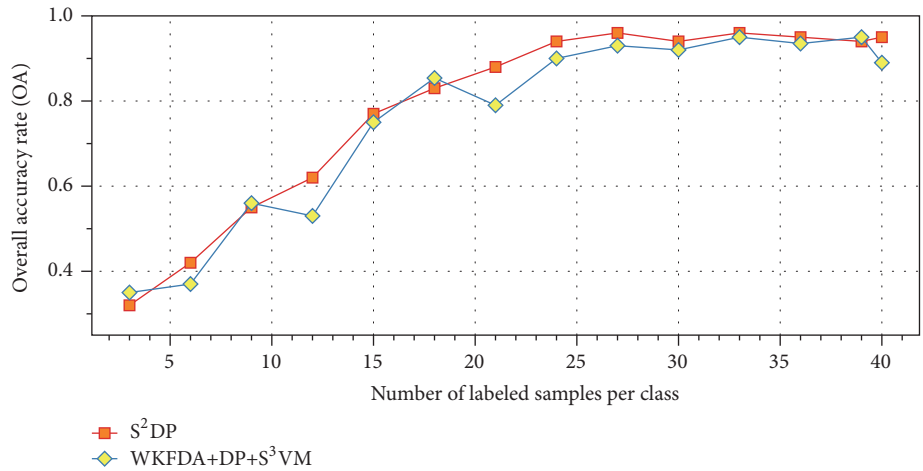


Figure 14: The OA trend chart of the S^{2}DP and WKFDA+DP+S^{3}VM as the number of labeled samples changes in the EOC_1 experiment.

samples are not cleared. As can be seen from Figure 15(b), the DP algorithm divides the WKFDA features into 5 clusters. Among these 5 clusters, clusters 1 and 3 have cluster haloes, and clusters 2, 4, and 5 are all clustered cores. As can be seen from Figure 15(c), the initial labeled samples ($*$ samples) are not included in clusters 4 and 5 and, therefore, the samples of clusters 4 and 5 need to wait for the next step of the algorithm to identify and label. Clusters 1, 2, and 3 contain initial labeled samples ($*$ samples), so they get the same label as the initial labeled sample. In the clustering halos of clusters 1 and 3, there are many clustering error samples ($\otimes$ samples) caused by the confused samples shown in Figure 15(a). This means that, in the future, they will affect the performance of the recognition algorithm if these $\otimes$ samples are not cleared.

For the WKFDA+DP+S^{3}VM algorithm, in the S^{3}VM training process, for one thing, the S^{3}VM cannot clear the $\otimes$ samples in Figure 15(c). And for another, for the **?** samples in Figure 15(c), the S^{3}VM can only identify them by traversing the samples. Therefore, the WKFDA+DP+S^{3}VM algorithm is unstable and inefficient. For the S^{2}DP algorithm, once the features of Figure 15(c) are input into the LCI, the LCI algorithm removes the unreliable features such as the clustering halos and cross-clustering features and makes full use of the cluster cores as reliable samples. Once the labeled samples are included in the cluster core, the other unlabeled samples are labeled with the labels of the labeled samples. For cluster cores that do not contain labeled samples, only the clustering center is identified, and the label of the whole cluster core can be obtained, which greatly improves the recognition efficiency. Figure 16 is a sequence diagram showing the recognition of the DP clustering result of Figure 15(b) by the LCI in the S^{2}DP algorithm. Figure 16(a) is the visualization of the DP clustering results after removing the interference samples. Figure 16(b) is the result diagram of LCI's final recognition of the DP clustering. As can be seen from Figure 16(a), both the confusing sample in Figure 15(a) and the $\otimes$ sample in Figure 15(c) are removed, greatly improving the reliability of sample identification. As can be seen from Figure 16(b), clusters 4 and 5 are correctly identified, and at the same time, only 5 samples with incorrect identification are in the

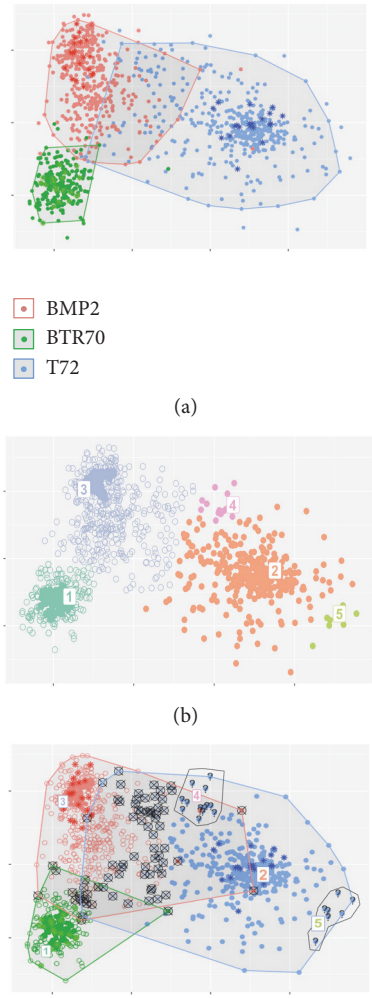


Figure 15: In EOC_1 experiment, the visualization results of the first two steps of the S^{2}DP and WKFDA+DP+S^{3}VM algorithms when the number of the samples is 21. (a) WKFDA features: • represents the learning set and ∗ represents the initial labeled samples; (b) actual classification map of WKFDA features after DP clustering has been carried: • represents the clustering core and ◦ represents the clustering halo; (c) true classification map of (b): • represents the clustering core, ⊗ represents the clustering error sample, and **?** represents the sample of the next step of the algorithm to be identified.

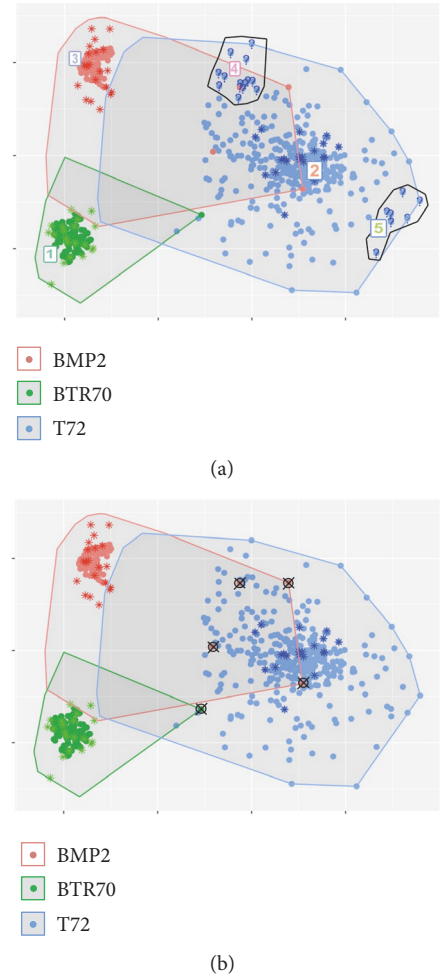


Figure 16: The sequence diagram of the LCI algorithm for identifying DP clustering results in Figure 15(b). (a) Visualization of DP clustering results after removing the interference samples; (b) LCI's final recognition visualization of DP clustering. • represents the clustering core; ∗ represents the initial labeled sample; **?** represents the sample of the next step of the algorithm to be identified; ⊗ represents the sample for labeling errors.

expanded labeled samples. Thus, S^{2}DP is quite reliable. In this way, the above conclusions are verified.

4.2.2. Verifying the IS^{2}DP Generalization Capabilities. In Section 4.2.1, the S^{2}DP is relatively stable, but its recognition accuracy is relatively low when the number of the labeled samples is less than 21. Therefore, the IS^{2}DP is required to generate a large number of the labeled samples to improve the recognition accuracy of S^{2}DP. Here, we compare the IS^{2}DP+S^{2}DP and S^{2}DP. The experimental data configuration is shown in Table 2 with the change of the number of the labeled samples, and the OA trend chart of the two methods is obtained, as shown in Figure 17. The horizontal axis represents the number of each type of target labeled samples corresponding to different experiments; the vertical axis represents the overall accuracy rate.

From Figure 17, we can see that when the number of the labeled samples is less than 21, the recognition performance of IS^{2}DP+S^{2}DP is 10% higher than that of S^{2}DP. With the number of labeled samples larger than 21, their classification accuracy is equivalent. To verify this conclusion, we apply 100 iterations onto IS^{2}DP when the number of labeled samples is 15. The labeled samples generated by IS^{2}DP are counted, as shown in Table 4. The accuracy rate of labeled samples generated from learning set is over 85%.

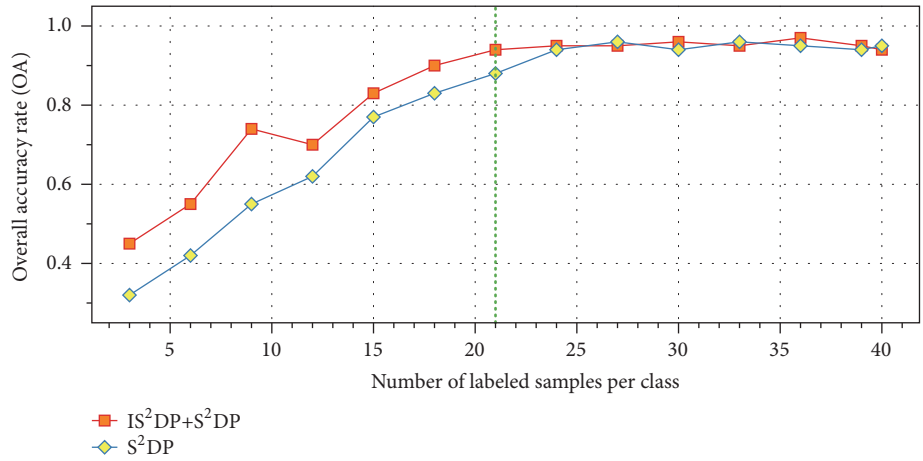


Figure 17: The OA trend charts of the IS^{2}DP+S^{2}DP and S^{2}DP with the number of labeled samples changes in EOC_1 experiment.

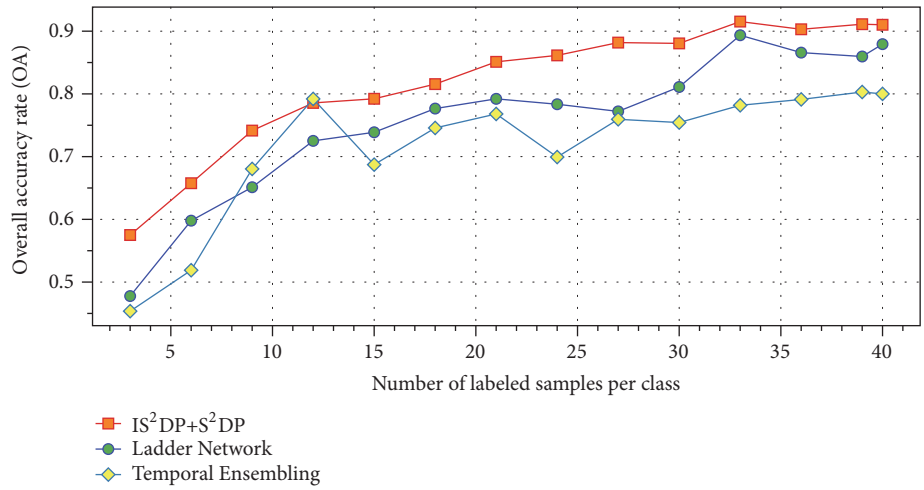


Figure 18: The OA trend charts of the IS^{2}DP+S^{2}DP, Ladder Network, and Temporal Ensembling with the number of labeled samples changes in EOC_2 experiment.

4.3. Comparison with Semi-Supervised Deep Learning. The semi-supervised deep learning algorithms, Ladder Network [7] and Temporal Ensembling [8], which contain a supervised and unsupervised learning process, similar to our algorithm. Therefore, we choose these two methods to compare with the IS^{2}DP-based S^{2}DP algorithm (IS^{2}DP+S^{2}DP). In addition to using SAR images as experimental data, we also use a set of publicly available optical image data to verify the effectiveness of our algorithm.

4.3.1. Testing with SAR Images. The experimental data configuration is shown in Table 3, and with the change of the number of the labeled samples, the OA trend chart of three methods is obtained, as shown in Figure 18. The horizontal axis represents the number of each type of labeled target samples corresponding to different experiments, and the vertical axis represents the overall accuracy rate.

In Figure 18, the recognition accuracy of the three methods is increasing with the increase of the labeled samples. From the curve smoothing, the accuracy curves of the Ladder Network and the Temporal Ensembling are fluctuating, especially the Temporal Ensembling. Comparing them, the accuracy curve of the IS^{2}DP+S^{2}DP is relatively consistent. From the classification accuracy, when the number of the labeled samples is less than 33, the results obtained by Ladder Network and Temporal Ensembling are not much different, but significantly lower than that of IS^{2}DP+S^{2}DP. When the number of the labeled samples reaches 33, the classification accuracy of IS^{2}DP+S^{2}DP is slightly better than that of Ladder Network. These results indicate that the learning set containing the interference samples has a great influence on the recognition performance of the Ladder Network and Temporal Ensembling. Because the Ladder Network and Temporal Ensembling were unable to remove these interference samples during the training process, their recognition accuracy was unstable and not high. Different from them, the IS^{2}DP+S^{2}DP can select reliable unlabeled samples and remove those interference samples, so its recognition performance is relatively stable and the accuracy is improved. When the number of the labeled samples is equal

Table 4: The IS^{2}DP generates labeled samples from the learning set when the number of labeled samples is 15 in EOC_1 experiment.

Target	Learning set	Generate labeled samples		Reject	Accuracy of each type of target(%)
		Correct	Error		
T72	190	167	11	12	87.89
BMP2	190	163	13	14	85.79
BTR70	190	172	8	10	90.53
Overall accuracy					88.07

Table 5: Under EOC_2, when labeled samples number 21, recognition results (confusion matrix) of the learning set by trained Temporal Ensembling.

	T72	BMP2	BTR70	Accuracy of each type of target(%)
T72	65	3	122	34.21
BMP2	0	174	16	91.58
BTR70	0	0	190	100
SLICY	0	0	190	0
Overall accuracy				56.45

to 21, we will analyze the use of the learning set by the three methods below.

In the Temporal Ensembling algorithm, one neutral network conducts two different works, supervised learning and unsupervised learning. Figures 19(a)–19(c), respectively, show losses in these two processes and in the whole method. Observed from the curve fluctuation, supervised learning loss in Figure 19(a) is the most stable while unsupervised learning in Figure 19(b) fluctuates significantly. It demonstrates that neural network performs well in learning labeled samples, but is still unstable to handle the learning set, thus resulting in unstable overall loss as shown in Figure 19(c). Finally, the Temporal Ensembling algorithm utilizes the learning set by 56.45% only, which is calculated based on the trained neutral network's recognition of the learning set. Recognition results of the learning set by the Temporal Ensembling algorithm is displayed in Table 5 (confusion matrix). Observed from the confusion matrix, the remaining 43.55% disturbs the learning process, for instance, by misrecognizing SLICY as BTR70 targets.

Similar to temporal ensembling algorithm, the neutral network in the Ladder Network algorithm consists of supervised learning and unsupervised learning as well. Figures 20(a)–20(c), respectively, show losses in these two processes and by the whole method. Observed from the curve fluctuation shown in Figure 20(a), supervised learning loss significantly fluctuates, probably because of inadequate labeled samples; in Figure 20(b), unsupervised learning performs stably, probably resulting from unsupervised learning (Autoencoder) embedded in the Ladder Network algorithm which could learn and recognize unlabeled samples and reduce certain interference. Thus, the overall loss shown in Figure 20(c) performs stably. Therefore, comparing with temporal ensembling, Ladder Network improves the utilization of the learning set to 68.42% (as shown in Table 6 confusion matrix), enhancing its recognition performance as well.

Differing from temporal ensembling and ladder network, the IS^{2}DP+S^{2}DP algorithm identifies reliable unlabeled samples by iterations before implementing feature learning, instead of directly learning features from the unlabeled samples. Here we employ 300 iterations on the IS^{2}DP+S^{2}DP algorithm for fair comparison. Figures 21(a)–21(c) show the screening of the reliable samples in the learning set during one iteration: (a) projection of the WKFDA algorithm on the learning set; (b) DP clustering result; (c) reliable samples labeled by LCI. In Figure 21, red circle, green circle, light blue circle, and blue circle represent BMP2, BTR70, T72, and SLICY target samples, respectively, and $*$ represents the labeled samples. Confused by SLICY interference targets, the WKFDA algorithm has some issue in projecting the learning set but performs well in dividing different samples during the DP clustering, and successfully identify SLICY during the LCI labeling process. Finally, IS^{2}DP+S^{2}DP improves the utilization of the learning set to 82.76% (as shown in Table 7). As 28.95% unreliable sample rejecting recognition will be deleted, only 10% false samples affects the performance; thus IS^{2}DP+S^{2}DP's recognition performance can be improved.

4.3.2. Testing with Optical Images. To verify the effectiveness of the proposed method on other data sets, we use optical image data to test IS^{2}DP+S^{2}DP. These optical image data come from some publicly available databases, and the detailed data configuration is shown in Table 8. The images of cats and dogs are from the database of the Kaggle competition platform [33]; the images of panda are from the ImageNet database [34]; the images of airplanes, motorbike, and faces are from the caltech101 database [35].

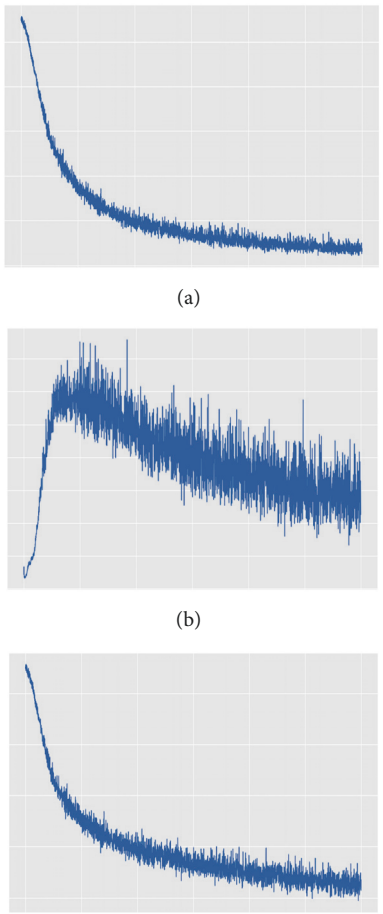
In Table 8, we set more stringent conditions than EOC_2 for SAR images, which is closer to the reality. Specifically, our interested targets are cats, dogs and panda. However, our learning set contains not only unlabeled interested targets,

Table 6: Under EOC_2, when the labeled samples' number is 21, the recognition results (confusion matrix) of the learning set by the trained ladder network.

	T72	BMP2	BTR70	Accuracy of each type of target(%)
T72	166	10	4	87.37
BMP2	12	178	0	93.68
BTR70	2	12	176	92.63
SLICY	67	53	70	0
Overall accuracy				68.42

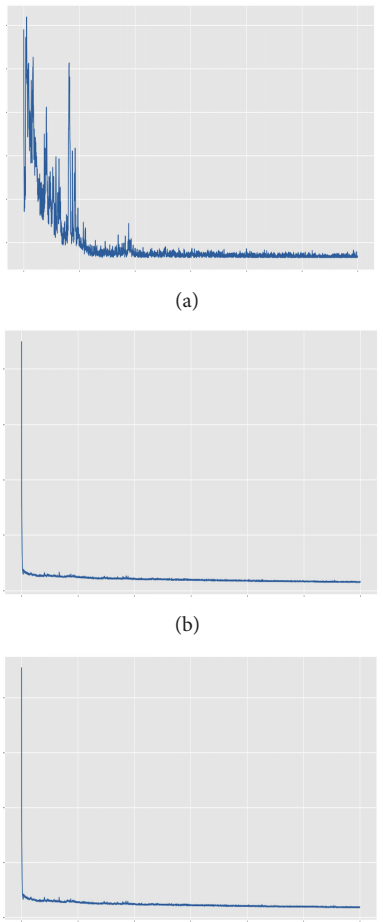
Table 7: Under EOC_2, when labeled samples number 21, recognition results of the learning set by IS^{2}DP+S^{2}DP after 300 iterations.

Target	Learning set	Generate labeled samples		Reject	Accuracy of each type of target(%)
		Correct	Error		
T72	190	155	12	23	81.58
BMP2	190	162	17	11	85.26
BTR70	190	147	22	21	77.37
SLICY	190	—	25	165	86.84
Overall accuracy					82.76



(a) (b) (c)

Figure 19: Under EOC_2, when labeled samples number 21, Temporal Ensembling losses during training: (a) supervised learning loss; (b)unsupervised learning loss; (c) overall loss.



(a) (b) (c)

Figure 20: Under EOC_2, when labeled samples number 21, Ladder Network losses during training: (a) supervised learning loss; (b) unsupervised method loss; (c) overall loss.

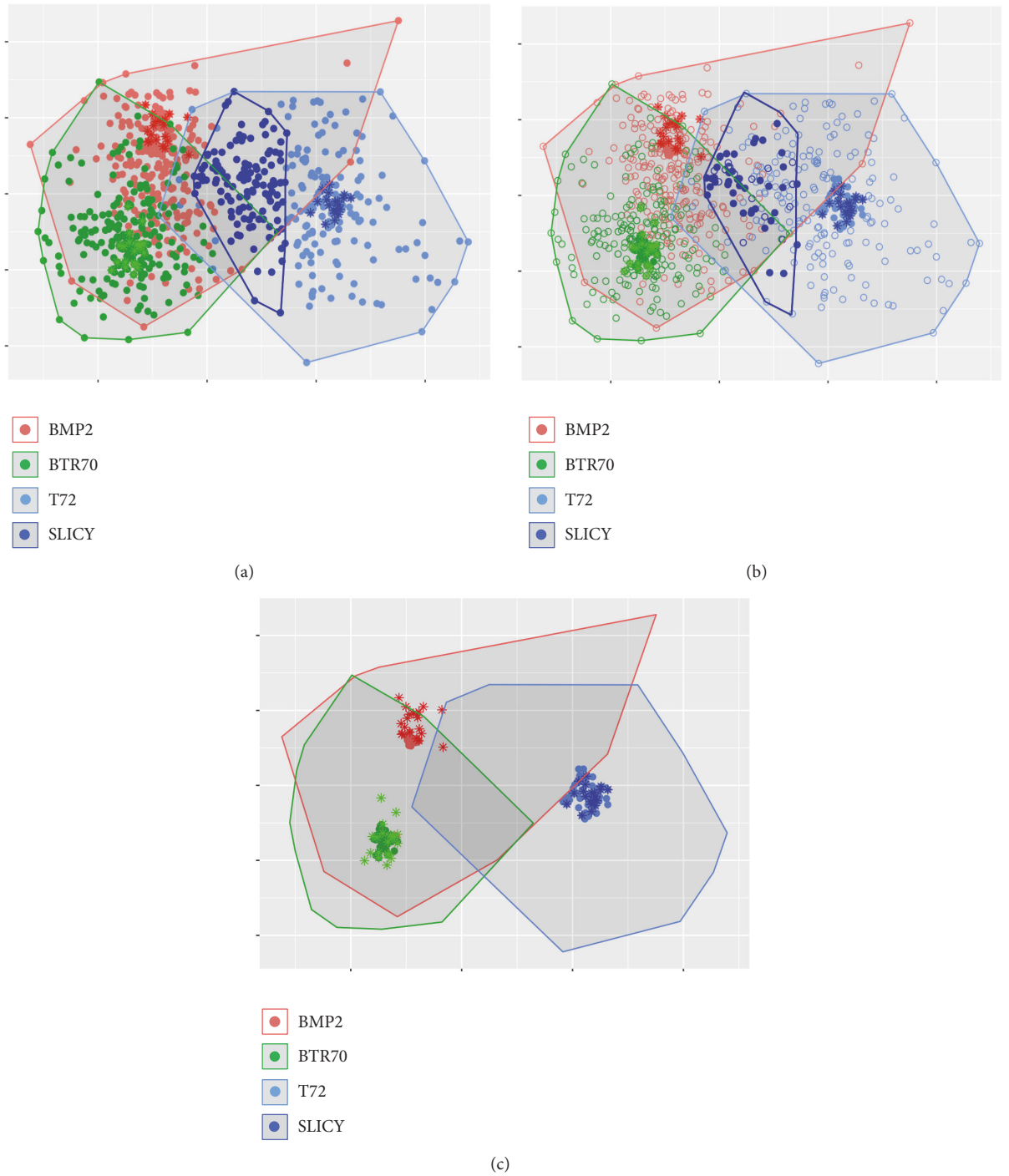


Figure 21: Under EOC_2, when the labeled samples number is 21, IS^{2}DP+S^{2}DP's outcomes of the learning set: (a) WKFDA's projection of the learning set; (b) DP clustering result; (c) reliable samples selected and labeled by LCI. Red circle, green circle, light blue, and blue circle represent BMP2, BTR70, T72, and SLICY target samples, respectively, and $*$ represents labeled sample.

but also other 3 types of interference targets (airplanes, motorbike, and faces) with the same number of unlabeled interested targets. Under such conditions, the Ladder Network, Temporal Ensembling, and IS^{2}DP+S^{2}DP are tested and compared. With the change of the number of the labeled samples, the OA trend chart of three methods is obtained, as shown in Figure 22. The horizontal axis represents the number of each type of target labeled samples corresponding to different experiments; the vertical axis represents the overall accuracy rate.

In Figure 22, the identification accuracy of our method IS^{2}DP+S^{2}DP is significantly better than Ladder Network and

Table 8: Optical image data set-up of the EOC.

		Data set						Testing set		
		Training set (labeled samples)			Learning set (unlabeled samples)					
Interested	Target	cats	dogs	panda	cats	dogs	panda	cats	dogs	panda
	Quantity	3~40	3~40	3~40	200	200	200	200	200	200
Interferential	Target	—	—	—	airplanes	motorbike	faces	—	—	—
	Quantity	—	—	—	200	200	200	—	—	—

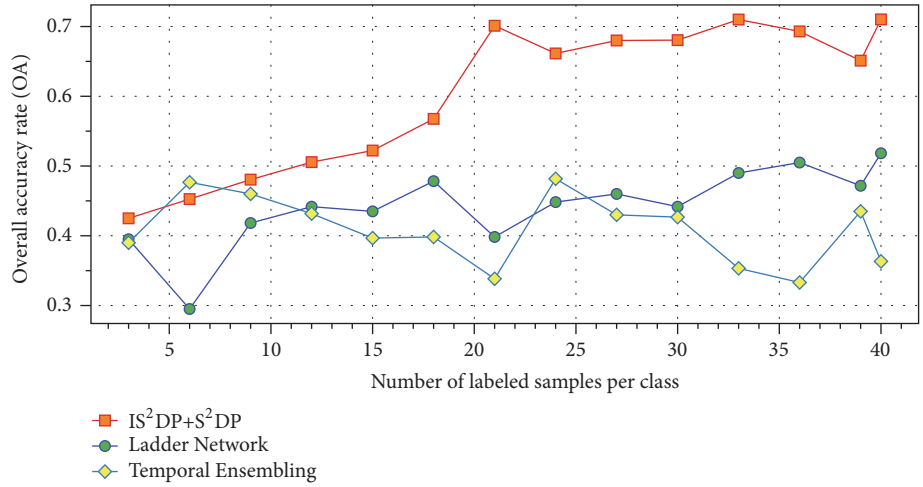


Figure 22: The OA trend charts of the IS^{2}DP+S^{2}DP, Ladder Network, and Temporal Ensembling with the number of the labeled samples changes in the optical images testing experiment.

Temporal Ensembling. From the OA trend, the recognition accuracy of Ladder Network and Temporal Ensembling does not increase significantly with the increase of the number of the labeled samples, while IS^{2}DP+S^{2}DP is significantly improved. Compared with the results of the SAR image test (Figure 18), the results of the three algorithms in the optical image test are significantly lower. This may be because in the learning set, we both increase the numbers of the target types and the number of the confusion targets, which leads to the less satisfactory results in learning the target features. From Figure 22, the Ladder Network and Temporal Ensembling algorithms are subject to more serious interference, and their average recognition accuracy is about 45%, respectively. Our algorithm IS^{2}DP+S^{2}DP is also subject to certain interference, but when the number of samples per class reaches 21, its average recognition accuracy is about 70%, which is significantly higher than the Ladder Network and Temporal Ensembling algorithms. When the number of the labeled samples is equal to 21, we will analyze the use of the learning set by the three methods.

Figure 23 shows the use of the learning set by the Ladder Network in the last 270 iterations during 1000 iterations of training. Figure 23(a) is the recognition accuracy of the learning set by Ladder Network; Figure 23(b) is the Ladder Network's loss value, where the blue line with square is the overall loss, the black line with circle is the supervised loss, and the green line with diamond is the unsupervised loss. From Figure 23(a), we know that the recognition accuracy is very low, about 33%. From Figure 23(b), we know that the supervised loss is low, while the unsupervised loss is high, which makes the overall loss difficult to reduce. Ladder Network is a complex network, which is intertwined by many components, but its core part mainly includes adding noise to samples, reconstructing samples and "skip connection" [36]. It first augments the unlabeled samples by adding noise to obtain a wider range of generalization information, secondly retains the sample information as much as possible by reconstructing the unlabeled samples in a regularized manner, and finally combines unsupervised learning with supervised learning to form semi-supervised learning by skip connection. Compared with supervised learning, unsupervised learning is more important in Ladder Network. Therefore, although Ladder Network has been well learned in the labeled samples, it has not been well learned in using the unlabeled samples, resulting in the whole algorithm has not been well trained. Finally, the Ladder Network recognition accuracy is neither stable nor high.

Figure 24 shows the use of the learning set by the Temporal Ensembling in the last 270 iterations during 1000 iterations of training. Compared with Figure 23, the recognition accuracy of Temporal Ensembling for the learning set is increased, about 48%, but it is still relatively low. Different from the Ladder Network, Temporal Ensembling adds noise to all the samples, which makes the labeled samples augmented. At the same time, in the initial stage of the training, the Temporal Ensembling's supervised learning plays an important role

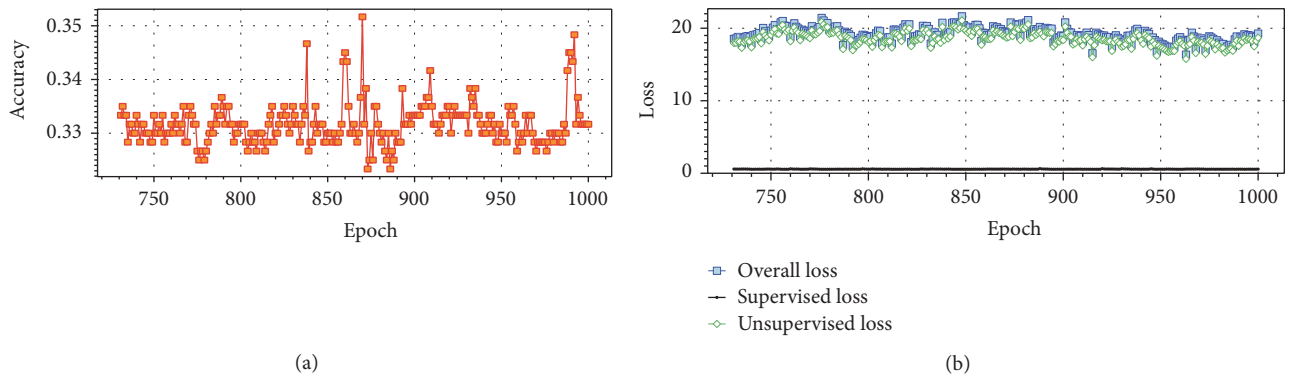


FIGURE 23: The use of the learning set by the Ladder Network in the last 270 iterations during 1000 iterations of training. (a) Recognition accuracy of learning sets and (b) Ladder Network loss value.

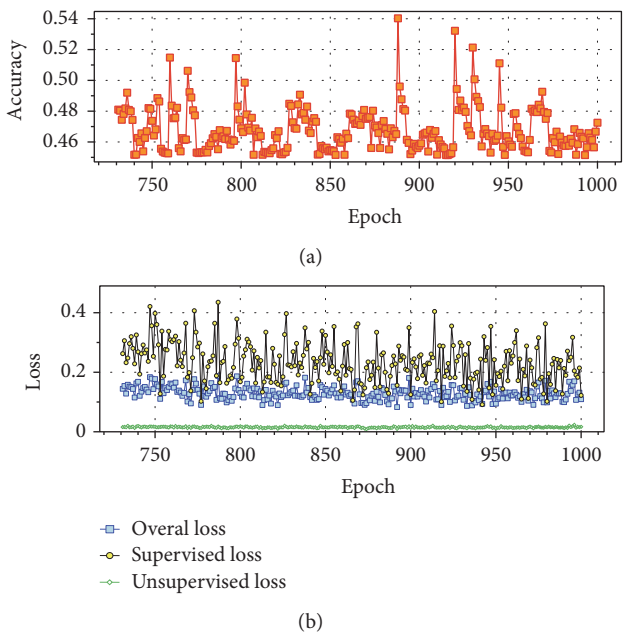


FIGURE 24: The use of the learning set by the Temporal Ensembling in the last 270 iterations during 1000 iterations of training. (a) Recognition accuracy of learning sets and (b) Temporal Ensembling loss value.

because of the small value of the unsupervised loss weighting function [8]. Therefore, the Temporal Ensembling is well trained to some extent. As the value of the loss weighting function increases, the unsupervised learning gradually plays an important role in Temporal Ensembling. Although the unsupervised loss is very low, Temporal Ensembling has not been well trained in learning interested target features because of the large number of unreliable samples in the learning set. Finally, Temporal Ensembling still has low recognition accuracy for interested targets.

Unlike the Ladder Network and the Temporal Ensembling algorithms, the IS^{2}DP+S^{2}DP algorithm first removes the interference samples in the process of using the learning set and then learns the selected reliable samples. Table 9 shows the recognition results of IS^{2}DP+S^{2}DP algorithm for learning set after 300 iterations. The average accuracy is 80%, which is significantly higher than that of Ladder Network and Temporal Ensembling algorithms. Compared with Table 7, the average accuracy of Table 9 is lower. However, the correct rate of rejection of the 3 types interference target samples has not been reduced, and these correct rates have reached more than 80%. In addition, the rejection error rate of the IS^{2}DP+S^{2}DP algorithm for the target samples is quite low; for example, cats is 18/200 = 0.09; dogs is 15/200 = 0.075; panda is 20/200 = 0.1. These experimental results show that the proposed algorithm is effective in optical image testing.

5. Conclusions

In order to accurately identify remote sensing images when there are few labeled samples, two new semi-supervised learning algorithms have been proposed in this paper: S^{2}DP and IS^{2}DP. They use labeled sample information to filter out reliable unlabeled samples to improve the performance of the semi-supervised algorithms.

The novelty of this paper lies in the following: (a) the WKFDA has been derived to explore the features of the images; (b) based on the clustering information of the DP, the labeling method LCI has been designed to query reliable unlabeled samples and accurately classify the unlabeled samples; (c) in IS^{2}DP, the unlabeled training set is divided into different subsets, which suppresses the deterioration of the algorithm by too many unreliable unlabeled samples in the learning process. Moreover, IS^{2}DP uses S^{3}VM twice to ensure reliable semi-supervised samples.

In the experiments for the actual SAR images recognition from the MSTAR database, the S^{2}DP has made a significant improvement in terms of the classification accuracy and the stability in comparison with other existing methods. In addition, the IS^{2}DP is effective and has applicable values to query the semi-labeled samples and is more suitable to deal with the situation where it lacks labeled samples.

How to make full use of remote sensing images to improve the performance of recognition algorithm has always been an open problem. Although the semi-supervised deep learning

Table 9: Recognition results of optical image learning set obtained after 300 iterations of the IS^{2}DP+S^{2}DP algorithm when the number of samples equals 21.

Target	Learning Set	Generate labeled samples		Reject	Accuracy of each type of target(%)
		Correct	Error		
cats	200	143	39	18	71.50
dogs	200	154	31	15	77.00
panda	200	165	15	20	82.50
airplanes	200	—	32	168	84.00
motorbike	200	—	40	160	80.00
faces	200	—	27	173	86.50
Overall accuracy					80.00

algorithm is susceptible to interfering samples, it has strong feature learning capabilities once the interfering samples have been removed. In the near future, we will try to further improve the feature learning ability of the S^{2}DP and IS^{2}DP algorithms by virtue of the semi-supervised deep learning.

Abbreviations

The following abbreviations are used in this manuscript:

DP:	Clustering by fast search and find of density peaks
EOC:	Extended operating conditions
IS^{2}DP:	Iterative S^{2}DP
KFDA:	Kernel Fisher discriminant analysis
KLFDA:	Kernel local Fisher discriminant analysis
KPCA:	Kernel principal component analysis
LCI:	Labeling method based on the DP clustering information
MSTAR:	Moving and Stationary Target Acquisition and Recognition database
OA:	Overall accuracy rate
PS^{3}VM-D:	Progressive semi-supervised SVM with diversity
SAR:	Synthetic aperture radar
Semi-KLFDA:	Semi-semisupervised KLFDA
S^{2}DP:	Semi-supervised learning method based on DP
SOC:	Standard operating conditions
SVM:	Support vector machine
S^{3}VM:	Semi-supervised SVM
S^{4}VM:	Safe S^{3}VM
WKFDA:	Weighted Kernel Fisher discriminant analysis.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61771027; 61071139; 61471019; 61171122; 61501011; 61671035), the Scientific Research Foundation of Guangxi Education Department (KY2015LX444), the Scientific Research and Technology Development Project of Wuzhou, Guangxi, China (201402205), the Guangxi Science and Technology Project (Guike AB16380273), and the Research and Practice on Teaching Reform of Web Page Making and Design Based on the Platform of "E-Commerce Pioneer Park" (Guijiao Zhicheng [2014]41). Professor A. Hussain was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) Grant no. EP/M026981/1. E. Yang was supported in part under the RSE-NNSFC Joint Project (2017-2019), grant number 6161101383, with China University of Petroleum (Huadong). H. Zhou was supported by UK EPSRC under Grant EP/N011074/1, Royal Society-Newton Advanced Fellowship under Grant NA160342, and European Union's Horizon 2020 research and innovation program under the Marie-Sklodowska-Curie Grant agreement no. 720325.

References

[1] M. Kang, K. Ji, X. Leng, X. Xing, and H. Zou, "Synthetic aperture radar target recognition with feature fusion based on a stacked autoencoder," *Sensors*, vol. 17, no. 1, p. 192, 2017.

[2] F. Gao, T. Huang, J. Sun, J. Wang, A. Hussain, and E. Yang, "A New Algorithm of SAR Image Target Recognition Based on," *Cognitive Computation*, pp. 1–16, 2018.

[3] F. Zhang, X. Yao, H. Tang, Q. Yin, Y. Hu, and B. Lei, "Multiple Mode SAR Raw Data Simulation and Parallel Acceleration for Gaofen-3 Mission," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pp. 1–12, 2018.

[4] A. Verma, H. Qassim, and D. Feinzimer, "Residual squeeze CNDS deep learning CNN model for very large scale places image recognition," in *Proceedings of the 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, pp. 463–469, New York, NY, USA, October 2017.

[5] A. Popella, G. Eiben, W. Geile, and S. Meltzer, "Knowledge-Based Interpretation of SAR Imagery," in *International Archives of Photogrammetry and Remote Sensing*, vol. 29, pp. 52-52, 1993.

[6] F. W. Rohde, P. F. Chen, and R. A. Hevenor, "Automated radar image analysis research in support of military needs,"

Automated Radar Image Analysis Research in Support of Military Needs, p. 87, 1986.

[7] A. Rasmus, H. Valpola, M. Honkala, M. Berglund, and T. Raiko, "Semi-Supervised Learning with Ladder Networks," *Computer Science*, vol. 9, supplement 1, p. 1, 2015.

[8] S. Laine and T. Aila, "Temporal Ensembling for Semi-Supervised Learning," 2016, https://arxiv.org/abs/1610.02242.

[9] G. Wang, S. Tan, C. Guan, N. Wang, and Z. Liu, "Multiple model particle filter track-before-detect for range ambiguous radar," *Chinese Journal of Aeronautics*, vol. 26, no. 6, pp. 1477–1487, 2013.

[10] Z. Huang, Z. Pan, and B. Lei, "Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data," *Remote Sensing*, vol. 9, no. 9, p. 907, 2017.

[11] T. Yang and C. E. Priebe, "The effect of model misspecification on semi-supervised classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 2093–2103, 2011.

[12] H. Gan, Z. Li, Y. Fan, and Z. Luo, "Dual Learning-Based Safe Semi-Supervised Learning," *IEEE Access*, vol. 6, pp. 2615–2621, 2017.

[13] Y.-F. Li and Z.-H. Zhou, "Improving semi-supervised support vector machines through unlabeled instances selection," in *Proceedings of the 25th AAAI Conference on Artificial Intelligence and the 23rd Innovative Applications of Artificial Intelligence Conference, AAAI-11 / IAAI-11*, pp. 386–391, USA, August 2011.

[14] K. P. Bennett and A. Demiriz, "Semi-supervised support vector machines," in *Proceedings of the 12th Annual Conference on Neural Information Processing Systems, NIPS 1998*, pp. 368–374, USA, December 1998.

[15] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and Their Applications*, vol. 13, no. 4, pp. 18–28, 1998.

[16] Y.-F. Li and Z.-H. Zhou, "Towards making unlabeled data never hurt," in *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, pp. 1081–1088, USA, July 2011.

[17] Y. Wang and S. Chen, "Safety-aware semi-supervised classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 11, pp. 1763–1772, 2013.

[18] H. Gan, Z. Luo, Y. Sun, X. Xi, N. Sang, and R. Huang, "Towards designing risk-based safe Laplacian Regularized Least Squares," *Expert Systems with Applications*, vol. 45, pp. 1–7, 2016.

[19] C. Persello and L. Bruzzone, "Active and semisupervised learning for the classification of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 6937–6956, 2014.

[20] D. Wang, F. Nie, and H. Huang, "Large-scale adaptive semi-supervised learning via unified inductive and transductive model," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2014*, pp. 482–491, USA, August 2014.

[21] N. Sokolovska, O. Cappé, and F. Yvon, "The asymptotics of semi-supervised learning in discriminative probabilistic models," in *Proceedings of the 25th International Conference on Machine Learning*, pp. 984–991, New York, NY, USA, July 2008.

[22] M. Kawakita and J. Takeuchi, "Safe semi-supervised learning based on weighted likelihood," *Neural Networks*, vol. 53, pp. 146–164, 2014.

[23] Y. M. Zhang, Y. Zhang, D. Y. Yeung, C. L. Liu, and X. Hou, "Transductive Learning on Adaptive Graphs," in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI*, Georgia, Atlanta , USA, July 2010.

[24] A. Laio and A. Rodriguez, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[25] L. Kaufman and P. J. Rousseeuw, *Clustering Large Applications (Program CLARA)*, John Wiley & Sons, 2008.

[26] A. Jatram and B. Biswas, "Dimension reduction using spectral methods in FANNY for fuzzy clustering of graphs," in *Proceedings of the 8th International Conference on Contemporary Computing, IC3 2015*, pp. 93–96, India, August 2015.

[27] T. Wen, J. Yan, D. Huang et al., "Feature extraction of electronic nose signals using QPSO-based multiple KFDA signal processing," *Sensors*, vol. 18, no. 2, p. 388, 2018.

[28] J.-M. Yin, M. Yang, and J.-W. Wan, "A kernel fisher linear discriminant analysis approach aiming at imbalanced data set," *Moshi Shibie yu Rengong Zhineng/Pattern Recognition and Artificial Intelligence*, vol. 23, no. 3, pp. 414–420, 2010.

[29] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Muller, "Fisher discriminant analysis with kernels," in *Proceedings of the 9th IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing (NNSP '99)*, pp. 41–48, Madison, Wis, USA, August 1999.

[30] Z. Shi and J. Hu, "A kernel approach to implementation of local linear discriminant analysis for face recognition," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 12, no. 1, pp. 62–70, 2017.

[31] M. Sugiyama, T. Ide, S. Nakajima, and J. Sese, "Semi-supervised local Fisher discriminant analysis for dimensionality reduction," *Machine Learning*, vol. 78, no. 1-2, pp. 35–61, 2010.

[32] Q. Wang, *Kernel Principal Component Analysis and its Applications in Face Recognition and Active Shape Models*, Computer Science, 2012.

[33] L. Smith and Y. Gal, "Understanding Measures of Uncertainty for Adversarial Example Detection," 2018, https://arxiv.org/abs/1803.08533.

[34] O. Russakovsky, J. Deng, H. Su et al., "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[35] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories," *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 59–70, 2007.

[36] M. Pezeshki, L. Fan, P. Brakel, A. Courville, and Y. Bengio, "Deconstructing the ladder network architecture," in *Proceedings of the 33rd International Conference on Machine Learning, ICML 2016*, pp. 3527–3539, USA, June 2016.